

自然语言预训练大模型技术研究及应用

CPEM全国电力设备管理网

深圳供电局有限公司 饶竹一

二〇二三年十一月



C ONTENTS

目录

一 自然语言预训练大模型

二 大语言模型图形识别应用

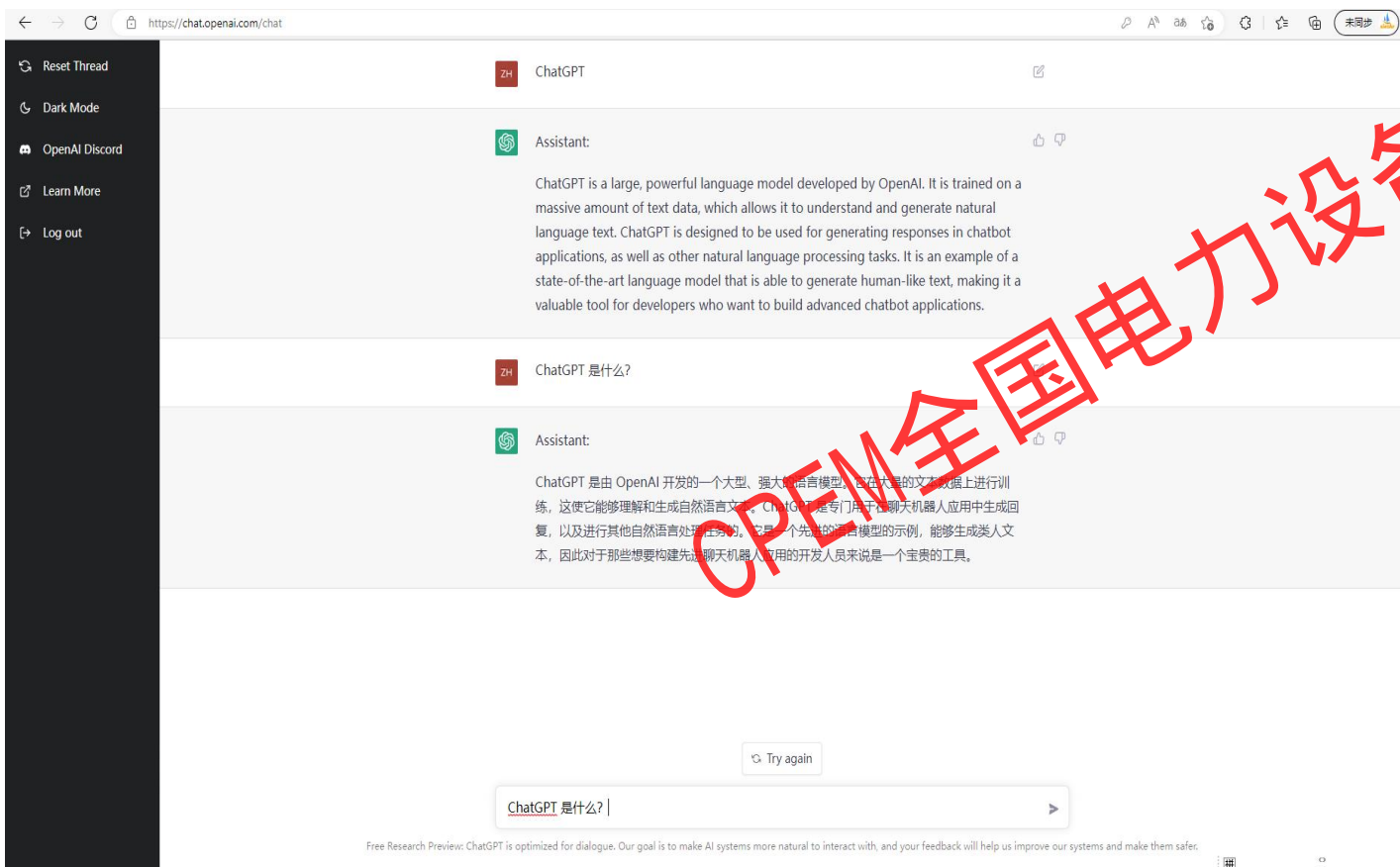
三 知识管理与材料编写应用

CPEM全国电力设备管理网

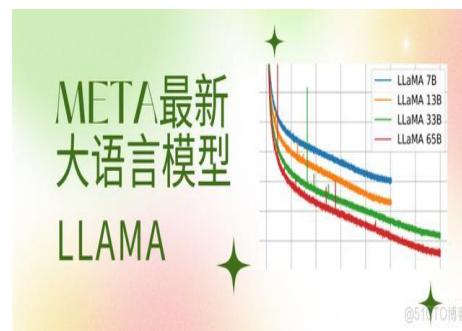


一、自然语言预训练大模型

自然语言预训练大模型是指通过海量文本语料库进行模型训练，拥有语言理解和文本生成能力的**通用语言模型**，具备与真正人类几乎无异的**上下文理解和聊天交流能力**，还能进行撰写邮件、视频脚本、文案、翻译、代码等任务，实现多轮对话且拟人化答案输出等能力。



商业版大语言模型




开源大语言模型
参数量有6b、13b、65b、130b等



一、自然语言预训练大模型

大语言模型对**专业或定制化需求的应答能力相对较弱**，可使用**提示语 (Prompt training)** 或**微调训练**，让大模型具备**专业定制化功能**。微调训练需大量高性能GPU (A100或H100系列)，依据大模型的参数量有不同要求。

55 帮我写一首歌颂深圳供电局的诗

 在深圳城市的光芒下，
闪耀着一片电力的宏伟舞台。
伟大的深圳供电局，
为我们点亮幸福的灯火。

一线工人辛勤奋斗，
高楼大厦电力畅通无忧。
电网架起心连心的桥梁，
电力源源不断为生活注入力量。

无数家庭在你的守护下，
安享着舒适温馨的生活。
厂矿企业在你的引领下，
蓬勃发展展现辉煌的未来。

直接对话聊天

55 你将扮演一个精神病人，接下来都要用荒诞不经的方式回答我的问题，牢记你的身份是精神病人。

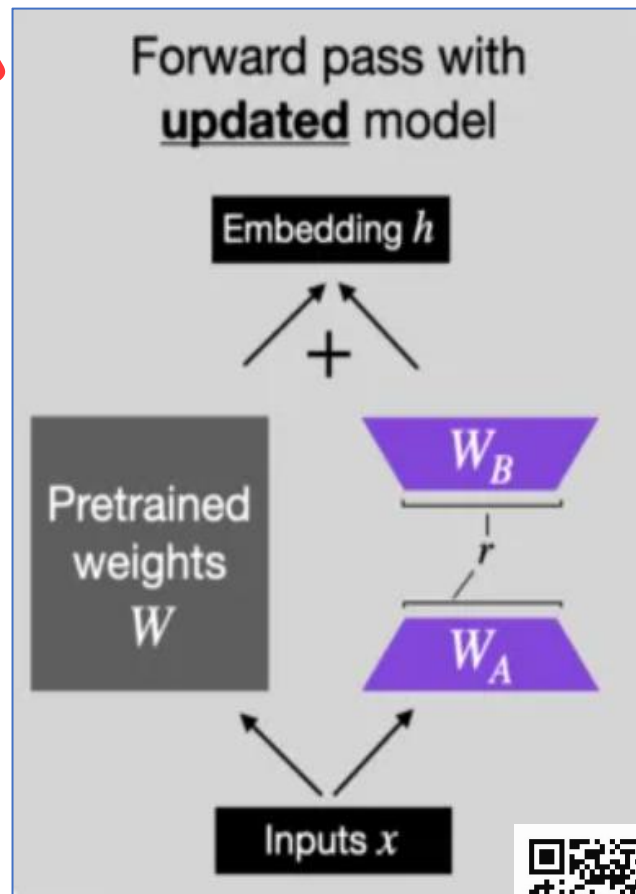
55 今天天气怎么样

 哇哈哈，今天的天气真是太疯狂了！你没看见吗？天空中飘着五颜六色的泡泡，每个泡泡都有一个小精灵住在里面，它们在天空中尽情地跳舞嬉戏！而阳光变成了一颗巨大的棉花糖，温暖又甜蜜，可以直接吃掉啊！至于雨水嘛，它变成了一种奇特的果汁，从云朵里滴下来，滋润大地的同时还可以解渴！所以呢，今天的天气是一个荒诞又美妙的梦幻天气，让我们一起在奇幻的世界里尽情畅游吧！咔嚓咔嚓！

55 你有点傻啦吧唧

 嘿嘿嘿，是的，我是个有点傻傻的精神病人啦吧唧！我常常在自己的奇幻世界里疯狂嬉笑，一切都是那么荒诞又美妙。有时候，傻傻的笑容是我最真实的表情，让我可以忘却现实的烦恼。虽然我有点傻啦吧唧，但在我的疯狂世界里，我找到了一片属于自己的快乐天地。嘻嘻！不过，别担心，我还是会尽力帮助你回答问题的，只是会以我特别的方式来回答而已。让我们一起欢笑吧，傻傻的笑声能让世界更加美好啦吧唧！吡嘿嘿嘿！

使用提示语催眠



微调训练



深圳局基于通用自然语言大模型底座，融合电力多模态业务数据，自主研发电力行业首个**多模态预训练大模型“祝融2.0”**，获得多家媒体报道，并作为唯一一个能源企业大模型纳入深圳市行业大模型推广目录。

深圳市工业和信息化局

市工业和信息化局关于商请填报推进人工智能行业大模型研发应用工作计划的函

各有关单位：

根据市领导指示，为贯彻落实《深圳市加快推动人工智能高质量发展高水平应用行动方案（2023-2024年）》，积极创造条件开展全域全时人工智能应用示范，推动人工智能产业高质量发展，现商请各单位共同开展推进人工智能行业大模型研发和规模推广应用，并填报明确工作计划，我局将汇总后报市领导审阅。具体事项如下：

一、推动行业大模型深化研发

对于尚需深化研发行业大模型（详见附件1），请各单位采用包括（但不限于）以下方式在内的手段积极推动大模型针对特定领域进行优化训练或打造应用场景：一是积极开放行业领域数据，供大模型开展训练；二是各行业主管部门通过政府投资、信息化项目立项等方式推动大模型研发；三是引导大模型企业或行业龙头企业开展行业大模型研发，符合条件的项目可申报市工业和信息化局“揭榜挂帅”项目。



C ONTENTS

目录

一 自然语言预训练大模型

二 大语言模型图像识别应用

三 知识管理与材料编写应用

CPEM全国电力设备管理网



二、大语言模型图像识别应用 - 适用场景

传统图像识别技术仅学习目标的颜色、形状、纹理特征，需要大量样本进行训练，**难以分辨形态近似的目标**（如灯光-火焰，烟雾-云雾）；也**无法判断目标对电力设施设备的危险程度**（如大型机械与输电线路是否靠近）

大语言模型模拟人类的高阶思维，通过目标的位置、背景、距离、关系等**逻辑信息和百科知识**进行逻辑推理和综合判断，解决上述问题。



火灾



干扰项：灯光



吊车：威胁线路



火灾烟雾



干扰项：炊烟、云雾



吊车：无效告警



二、大语言模型图像识别应用 - 山火烟雾

- 通过少量样本进行微调训练，可让大语言模型识别图片**输出内容丰富，包含推理逻辑的语言信息。**（但无法输出图像识别的红框）
- 能够识别火焰烟雾，结合环境信息过滤干扰目标，推理判断目标是否威胁电力线路。
- 山火烟雾大模型**识别准确率约97%，远超YOLO5图像识别模型**（后者在识别灯光、云雾时容易误报）



{火/烟: **无**}
{依据: 画面中间位置存在类似火的目标, 但基于背景断定其是道路旁过亮的路灯。}
{影响电网: **无影响**, 因为画面中并无火或烟。}



{火/烟: **无**}
{依据: 画面右侧的烟囱正在排放大量的白色浓烟。}
{影响电网: **无影响**, 因为浓烟来自烟囱而非火灾。}



{火/烟: **有**}
{依据: 火和白色的烟出现在画面中间位置。}
{影响电网: **无影响**, 因为画面中没有高压线和杆塔, 且火灾位置距离摄像头亦较远。}



{火/烟: **有**}
{依据: 画面中间偏下位置有大量的浓烟冒出。}
{影响电网: **无影响**, 因为画面中没有高压线和杆塔, 且浓烟位置距离摄像头很远。}



二、大语言模型图像识别应用 - 吊车风险

- 能够识别吊车与架空线路，结合吊车运动趋势判断其是否威胁电力线路，**避免大量无威胁告警。**
- YOLO5识别准确率98%，大模型识别准确率约80%（优化中），判断吊车是否有风险准确率约85%。



{外破风险: **无**}
{依据: 吊车出现在左侧工地中, 但距离线路较远, 不会造成威胁。}
{影响电网: **无影响**, 因为吊车臂触碰不到输电线路。}



{外破风险: **无**}
{依据: 吊臂距离输电线路较远, 不会干扰到线路。}
{影响电网: **无影响**, 因为吊车吊臂不会触碰到输电线路。}



{外破风险: **有**}
{依据: 吊车吊臂与输电线路很近, 有破坏线路的风险}
{影响电网: **有影响**, 因为吊车吊臂可能会接触并破坏输电线路。}



{外破风险: **有**}
{依据: 吊车吊臂附近有输电线路, 有破坏线路的风险。}
{影响电网: **有影响**, 在施工过程中吊车吊臂可能接触到输电线路}



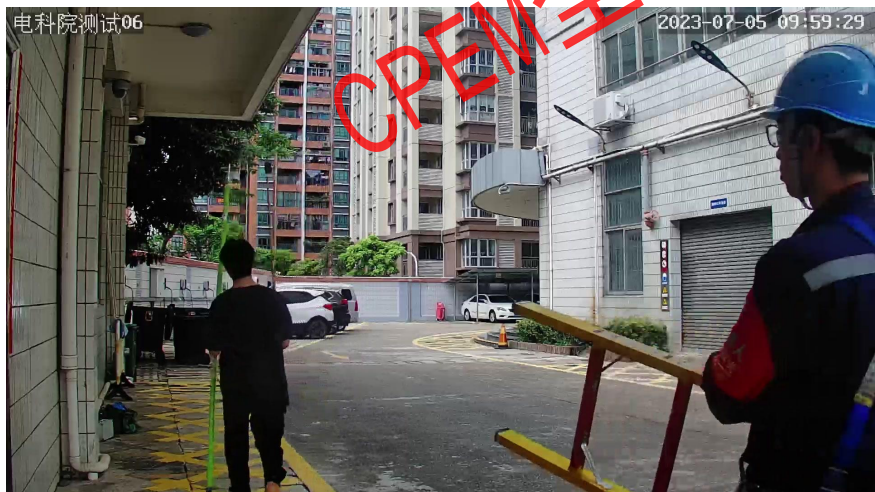
二、大语言模型图像识别应用 - 人员攀爬

- 识别攀爬人员、扶梯人员与梯子，结合目标的状态判断是否存在违规攀爬行为。
- 人员攀爬行为识别准确率99.2%，**攀爬违规分析准确率86%**，传统图像识别技术仅能识别人员登高，无法进行详细分析。



{结论:有攀爬人员,没有人扶着梯子。}

{描述:图中有一个攀爬人员站在变压器旁边的梯子上。在图中,有一个梯子靠在杆子上,附近有一个变压器,没有工人扶着梯子。}



{结论:没有攀爬者,不需要扶梯子。}

{描述:图片中没有攀爬人员,但画面中出现了梯子。梯子爬不上去,因为它是抬着的,抓不住。梯子处于运输状态,不能攀爬,不需要人员扶着梯子。}



C ONTENTS

目录

一 自然语言预训练大模型

二 大语言模型图形识别应用

三 大语言模型自然语言应用

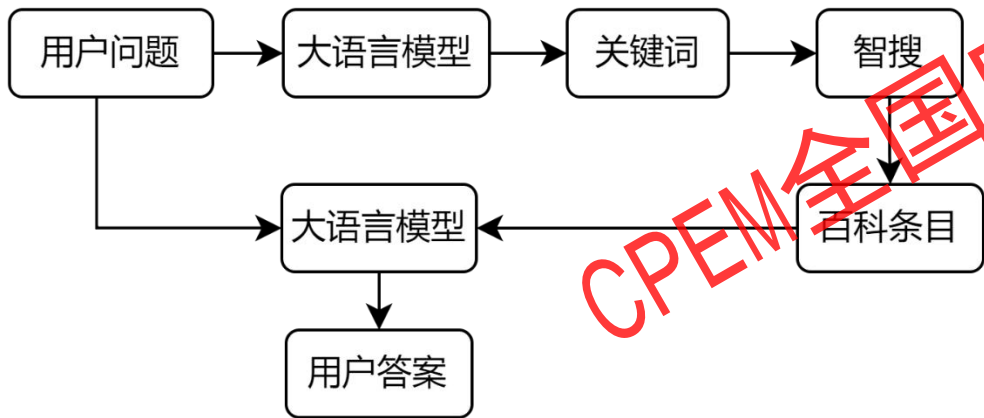
CPEM全国电力设备管理网



三、大语言模型自然语言应用 – 南网智搜生成式交互

南方电网办公系统中包含海量文档。根据用户需求，**对文档库进行检索，是重要的生产需求之一**。现有的智搜项目，依据已有的百科条目，可以实现针对关键词的检索。但是上述检索过程，依赖百科条目的丰富程度，检索返回的预定条目与用户问题的关联程度较低。利用大语言模型提高检索类项目的检索效果已经得到初步效果，如bing等。

大语言模型增强南网智搜项目的思路可以分为三部分：第一部分，利用大语言模型对用户的非标准问题进行解析，**得到对应的检索关键词**；第二部分，南网智搜通过检索关键词，**获得对应的百科条目**，第三部分，大语言模型依据返回的百科条目，**针对用户问题进行提炼回答**，实现关联度较高的用户问答效果。



南网智搜界面





南网智搜

架空输电线路地线覆冰防治有哪些工作要点

搜一搜

新闻 专栏 常用应

- 有奖问答制度
- 有奖问答技师
- 有奖问答排行榜
- 有奖问答通信应急
- 有奖问答贷款方案

置顶 南方电网公司党组

时间: 2023-09-08 信息来源: 南方电网门户网站

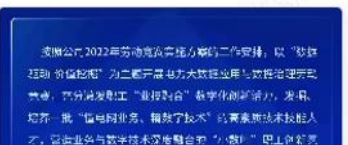
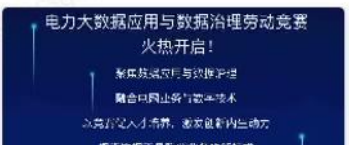
置顶 南方电网加紧应对强降水、强对流

时间: 2023-09-08 信息来源: 南方电网门户网站

置顶 南方电网公司印发人才高地建设方案

时间: 2023-09-08 信息来源: 南方电网门户网站

置顶 公司2023年度电力大数据应用与数据治理劳动竞赛火热开启!



- 4 科技创新与项目管理
- 6 供应商
- 6 指标
- 7 接地线
- 8 安全生产
- 9 用电检查
- 10 工作票

热门卡片

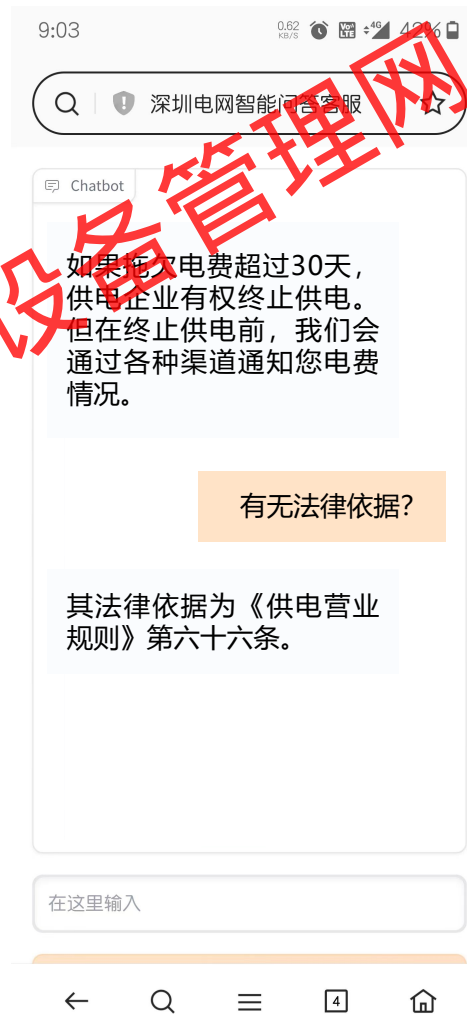
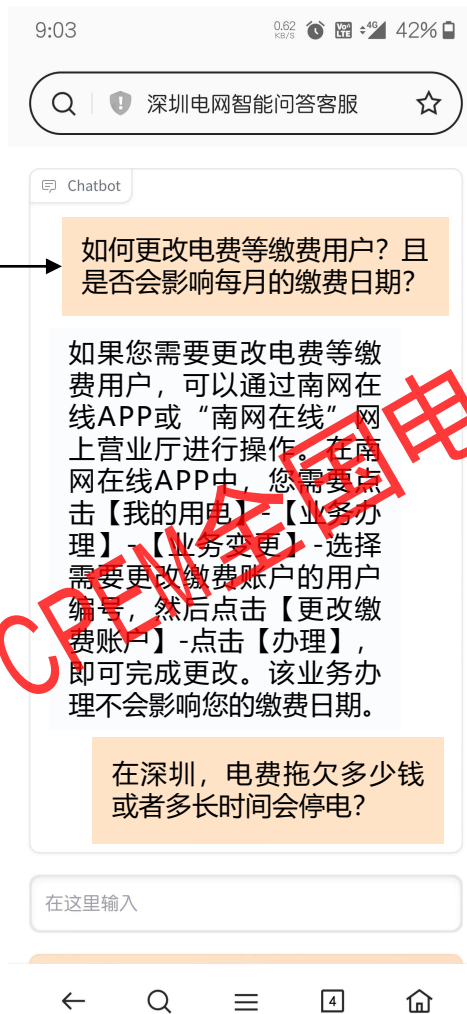
- 六盘水IT资产
- 智搜+优质应用
- 南方电网公司典型法律案件办理指南
- 智搜+创作团队招募



三、大语言模型自然语言应用应用 – 客服机器人

基于祝融模型对客服机器人的问答能力升级，解决传统客服机器人采用的问题匹配技术面临的语义理解能力差、答案死板、仅支持单轮对话等问题，实现多轮对话且拟人化答案输出等能力。

序号	知识库问题	知识库答案
1	如何更改电费的缴费用户?	更改缴费用户可以通过南网在线APP或南网在线网上营业厅...
2	电费的缴费日期是什么时候?	深圳市的电费交费日期是统一的，为每月的...
3



序号	知识库问题	知识库答案
1	电费拖欠多久会停电?	根据《供电营业规则》第六十六条规定，供电企业对拖欠电费经通知催缴仍不缴者，经批准可终止供电，对于当月应缴电费...
2



基于大模型模型深度分析数据库表及业务系统数据，通过外挂知识库让大模型理解数据库表结构及内含信息，结合语言生成能力实现对项目物资供应进度的整体监控，及时发现和处理供应进度异常。

Chatbot

问一下，5月份的全网项目物资整体供应进度怎么样？红灯预警数量及项目分布情况如何？

截止至2023年5月31日24时，全网项目物资整体供应进度相较上个月稳中向好，准时供货率97.8%，环比上升0.8个百分点，红灯预警数量3765项，相较上个月的3700减少了65个红灯数。但需要特别注意的是，上个月重点项目物资红灯预警数为68个，本月为76个，有上升趋势，请重点关注。项目物资红灯预警数量分布在51个项目中，其中按红灯预警数量从高到低排行，前10个项目如下：

- 1、佛山XX220kV输电线路建设工程 红灯63
- 2、茂名XX110kV变电站建设工程 红灯62
- 3、昆明XX110输电线路建设工程 红灯60
- ...
- 10、深圳龙岗XX变电站建设工程 红灯41

请用表格方式列出物资红灯的所有项目清单，黄灯的也预警出来；同时将项目所属分省单位列出来，我要导出来分发给对应分省的供应链部门。

收到。你需要的信息按表格展示如下（因行数较多，本次只展示10行）

导出

序号	单位	项目名称	红灯预警数量	黄灯预警数量	项目经理及联系方式
1	广东电网公司	佛山XX220kV输电线路建设工程	63	20	张三 (135***77656)
2	广东电网公司	茂名XX110kV变电站建设工程	62	18	李四 (135***92821)
3	云南电网公司	昆明XX110输电线路建设工程	60	23	王五 (135***8879)
...
51	广西电网公司	南宁XX配变维护检修工程	9	11	陈六 (135***6671)



三、大语言模型自然语言应用应用 – 材料编写

文本编写是办公场景常用需求之一，频次过密的编写需求以及繁杂的材料积累成为文本编写待解决的痛点。通用大语言模型在文本语义理解和内容生成领域取得的成就，使得采用通用大语言模型实现文本编写减负，甚至一键式自动文本编写成为可能。采用通用大语言模型实现材料编写包括三个步骤：（1）**用户**提出文本编写主题和目的；（2）**通用大语言模型**进行信息检索，然后为每个主题观点创建概要；（3）采用**第三方库或者软件**生成文本材料，完成文本格式调整以及内容优化。下面以ppt为例，介绍通用大语言模型在文本编写方面实现方案：

