

个人信息

姓名：于先生

学历：博士

性别：男

专业：计算机软件

民族：汉族

住址：北京市海淀区中关村

github <https://github.com/facebook>

电话（微信）：**138-1071-7899**

邮箱：**ZiyueYu@163.com**

技能描述

- 熟悉单模态/多模态 LLM 技术栈（底座模型训练、增强学习、大模型工程化），深入理解 GPT/Transformer/LLaVA/Clip/BLIP2 架构，大模型应用模式 RAG/Agent；
- 熟悉大模型 LLM Pre-Training：基于 Megatron-DeepSpeed 集群训练，Post-Training:SFT、RLHF。熟悉深度学习框架 Pytorch, Huggingface，有项目实践经验。
- 熟悉 NLP 自然语言处理，基于 Transformer, vector/token/embeddings 等算法实现机器翻译，人机问答和摘要生成，文章生成等，并能采用 attention 机制提升一些场景的预测精度。
- 带领团队从事基于大模型、自然语言处理，人工智能服务于企业数字化转型，基础创新研究，应用研究，产品孵化，支撑用户服务，企业管理以及面向各类数据资产的云端跨平台应用。

教育背景

北京航空航天大学	计算机软件	博士	2003.9-2008.10
	研究方向：人工智能及应用		
北京航空航天大学	计算机应用	硕士	1996.9-1999.7
北京航空航天大学	计算机科学	学士	1992.9-1996.7

工作经历

2021.10—至今

中防融创

技术总监

研发百亿参数大模型 LLMs，推进成果转化，和多家央企合作，针对重点行业应用和场景，进行数据标注和大模型微调，提升大模型垂直行业性能，解决用户的大模型新技术、业务赋能、提质增效需求。

1.大模型/多模态 LLM 训练。针对文本生成、内容理解场景和多模态文本图像场景，设计和优化大模型的架构（类 GPT 编码器），标注训练数据、组织 GPU 加速卡千卡集群，基于 Pytorch 和 Megatron-DeepSpeed 并行框架，进行大模型的准备、调试、训练和性能调试。（合作单位：国家电网，北航重点实验室）

2. 大模型/多模态/人工智能行业应用。基于大模型和多模态大模型，针对重点行业的业务需求，融合多模态场景，设计输出行业模型二次训练、调优算法方案，并输出相应的基础设施建设方案。（ToB ToG 应用：央企中石化、国家电网等）

（1）基于人工智能大模型，为企业整体规划、设计“业务+人工智能分析平台”应用体系架构，系统梳理了典型应用场景及应用模式，形成架构统一、多模态场景丰富、机制长效、体系完善的解决方案。

（2）能源行业应用，基于能源行业数据，精调大模型，形成能源行业年鉴。

（3）行业信息决策助手，基于应用场景，生成相关关键信息，支撑用户决策。

3.算力中心规划实施和算力调度。规划北京、山西 1000P（CPU+GPU 异构算力）人工智能算力中心，满足人工智能、数字经济等广泛领域应用，采用先进液冷技术，满足绿色双碳数据中心标准，构建算力中心智能云生态，提供智能 IaaS,PaaS,SaaS 服务，基于多云模式研发远程算力中心调度系统，满足用户远程算力调度需求。（1）针对 AI 基础大模型训练、智算中心、行业 AI 等场景，设计、输出 AI 基础设施解决方案的技术架构。（2）负责 AI 大模型基础设施方案、关键特性设计，输出 AI 基础设施智算中心的版本设计材料，进行版本落地。

2017.06 —2021.09

国家电网集团一级部门 首席专家

1、国家电网数字孪生应用与规划。联系北航、清华大学等共同研究前沿技术，推动科研成果转化，

深化数字孪生电网顶层设计，围绕源网荷储全要素协同互动等重点方向，加强数据驱动，加快构建数字孪生模型，因地制宜探索开展实用化应用，助力传统电网向能源互联网升级。

2、国家电网人工智能深化应用顶层规划，全面提升人工智能“两库一平台”两级贯通、云边协同能力，深化人工智能通用组件在各专业、各单位应用，加快人工智能技术在设备等相关专业规模化推广，健全人工智能运营和共享机制，打造人工智能平台、应用、服务良性运营生态。

3、国家电网 2019-2020 年人工智能应用顶层规划和实施，依托人工智能框架，规划人脸识别，OCR，语音识别，RPA，知识图谱，NLP 六个人工智能组件，在国网总部、省公司和直属单位等多个业务系统推广应用。

4、国家电网大数据中心建设顶层规划和实施，数据中心规划，数据中台选型规划，数据平台建设、数据归集、数据治理、数据应用，国网数据分析对接。

5、国家电网泛在电力物联网顶层规划，电力泛在物联网体系结构规划，系统实施路线，三期目标规划，国网总部和省公司业务对接。

6、国家电网知识库管理顶层规划，在调研国网公司现状和相关头部企业如 CNKI 的基础上形成国家电网只知识管理框架、标准和内容，交由相关公司具体实施；

7、国家电网数据管理顶层规划和实施，对国网集团主要业务部门 1700 多套业务系统数据进行梳理，形成国家电网数据管理现状报告，基于现状规划下一步数据管理工作重点内容；

8、国家电网集团数字化转型顶层规划，数字化转型团队组件，赴相关企业华为、百度、阿里、东航、施耐德、西门子企业调研，完成调研报告、数字化转型现状、数字化转型路径报告。

● 国家电网 95598 客服数据应用分析 2020.5 – 2022.5

工作职责：算法设计，系统架构，团队管理

项目描述：95598 服务 11 亿用户，作为国家电网联系政府、企业、事业单位和家庭用户的窗口，在电网业务改进、服务优化方面，蕴含重要价值。

针对 95598 用户，进行用户画像，分类管理，进一步提升客户的服务水平。

基于数据中台，针对 95598 工单，梳理数据分析主题，重点开展意见类工单和迎峰度夏工单主题分析。意见类工单主要进行主旨话题分析，分析客户情感，关联投诉工单分析，预警热点，敏感话题，降低投诉发生概率，进一步提升客服满意度，提升数据支撑。

迎峰度夏抢修工单，主要对故障时长，故障类型，典型日故障开展数据分析，掌握趋势规律，预测抢修峰值，服务业务部门优化抢修资源分配，提高迎峰度夏抢修效率。

主要技术栈：机器学习、自然语言处理，python/java/c++，SKLearn，Tensorflow，Hadoop，Spark，

● 基于深度学习的企业舆情检测和经营风险分析系统 2018.1 - 2019.6

工作职责：算法设计，系统架构，团队管理

项目描述：

应用事件抽取和情感分析等 NLP 技术，对诸如新闻媒体和社交平台上的信息按照情感、业务关联风险和重要程度进行分类整理和清晰呈现。提早发现关注公司的正负面新闻，利用先发优势辅助决策交易。

海量 Web（微博、论坛、新闻等）多区域、跨渠道外部信息的数据结构化内容提取技术；文档主题生成模型和概率潜在语义分析技术，获取关于公司的主题结构；基于自然语言处理的 Web 短文本立场判定技术；基于深度学习的舆情传播网络溯源分析技术；

基于深度学习的企业内部经营管理风险智能识别技术：深度学习的经营管理风险特征编码技术；基于半监督学习的经营管理风险识别技术；基于特征值扰动的风险成因解释技术。

● 企业图像处理智能云 2017.1 - 2018.5

工作职责：系统架构，算法设计，团队管理

项目描述：

构建云、端人工智能处理系统，实现从自动化、智能化数据处理流水线，提升准确度和效率；

建立精细的识别类别体系,实现目标部件的分类,实现部件的缺陷识别;

在当前几种典型的目标检测识别方法的基础上,利用 CNN 等深度学习算法提升小部件识别的准确性和效率,分析了不同参数对检测结果的影响;

利用特定的 GPU 计算单元可以实时目标检测和识别,为无人机巡检图像的智能化处理提供核心支撑;

2014.01 — 2017.06

IBM

大数据和机器学习系统架构师

● 业务系统的大数据分析和挖掘项目

2015.6-2017.5

工作职责: 系统架构师, 团队管理, 架构设计, 算法实现

负责 IBM 公司与大数据、云计算相关的大型企业信息系统的软件架构设计及核心代码编写, 参与基于 Cloud Foundry 的大数据处理平台 BigInsights 的架构设计与开发, 从事中国东方航空公司基于 BigInsights 云平台的财务一体化大数据分析平台的项目管理、解决方案、架构设计及核心代码编写工作。

参与苏宁易购大型电子商务网站的系统架构设计的系统架构设计、数据库设计以及分布式云计算和大数据平台的技术构建, 参与系统重构、性能优化以及部分核心代码编写。平台前期采用 IBM 的 WCS 电子商务解决方案 WebSphereCommerce, 应用服务器采用 WebSphere, 数据库采用 DB2, 服务器采用 AIX 操作系统。后期为解决 WCS 在缓存系统、数据库集群技术、读写分离等方面技术的缺陷, 采用基于 Spring, Struts, JQuery, iBatis Freemarker, Redis Cache 等开源技术架构重构优化平台框架并基于 Hadoop、HBase 等大数据技术进行改造; 参与电商云平台底层 IaaS 和 PaaS 层的建设, 采用 OpenStack 构造 IaaS 层, 在底层通过对 KVM、Xen 虚拟化技术的综合使用和优化, 深度融合 Linux 系统, 为电商系统提供对云主机的计算、存储和网络资源的深度优化; 在 PaaS 层采用基于 Cloud Foundry 的 BigInsights 平台, 并针对电商系统进行专门优化。

● 互联网海量数据复杂实体关联分析系统研究与实现

2014.3-2015.5

工作职责: 系统架构师, 团队管理, 架构设计, 算法实现

项目描述: 基于系统采集的互联网数据, 对新闻、论坛、博客、微博等多元数据中的人物、组织等实体进行深入的关联分析, 从而为全局性网络舆情分析提供支持。

a. 负责实体知识库的构建, 数据规模达到百万级, 提出基于全局特征的感知器模型, 从无结构文本中抽取实体属性; b. 作为项目架构师, 研究实体关联分析的关键技术, 协调系统软件各个模块的联合部署、调试与测试; c. 面向大规模网络文本抽取特定领域的实体关系, 准确率达到 92% 以上; d. 面向领域分类的热点实体发现及推荐。

2009.01 — 2013.10

中国国铁集团

软件架构师

● 系统数据挖掘和分析项目

2012.9-2013.10

工作职责: 系统架构师

● 票务系统管理运营中心 SOC 研发

2009.6-2011.6

工作职责: 系统架构师

● 高铁 12306 票务系统

2009.1-2011.2

工作职责: 系统架构

英语能力

● 英语等级: CET-6;

● 较强的英语读写能力, 较好的英语口语和听力。