

# 李佳阳 -java开发实习生

👤 李佳阳/男/22岁  
☎ 18710258940 (微信同号)

✉ lijiaoyang9@126.com

## 教育背景

北京石油化工学院 - 数据科学与大数据技术 2023.09 - 2026.06  
北京石油化工学院 - 自动化 2021.09 - 2023.06  
出于对计算机技术的职业热情，于2023年6月由自动化专业转入。为夯实专业基础，选择延长一年学制以系统学习核心课程。

## 技能

### Java后端：

- 掌握Java基础语法，对集合、IO流、泛型、反射等核心知识有扎实的理解；了解JUC并发、JVM调优等；
- 熟悉SSM、Spring Boot、Spring Cloud等主流开发框架；

### 编程语言：

- 熟悉 Python 及常用库，有良好的编程习惯；熟悉FastAPI、Flask、Django等框架；
- 了解 C/C++ 基础，了解 STL 标准库；

### 大数据技术：

- 熟悉 Hadoop、Hive、PySpark 等大数据组件的使用与配置；
- 熟悉 Linux 常用命令和开发环境。了解 Docker、Kubernetes 等容器化技术；

### 数据库：

- 掌握关系型数据库MySQL的使用，并熟悉NoSQL数据库Redis的核心原理与主要应用场景；
- 了解MongoDB等 其他NoSQL 数据库的基本操作。

### 人工智能：

- 深入理解 Transformer、BERT 等主流模型，熟练掌握命名实体识别 (NER) 等NLP任务；
- 了解 DDPM (扩散模型)、CLIP (多模态) 等前沿模型的基本原理。

工具：熟悉Git、Maven等项目管理及构建工具。

## 项目

### 基于Hadoop的海域水质大数据分析系统

Selenium Hadoop pyspark Hive Mysql pyecharts

- **项目描述：**独立开发并部署了一个完整的大数据分析平台，旨在对中国四大海域的水质数据进行端到端的采集、处理、分析与可视化展示。
- **工作内容：**
  - 负责数据全链路开发，使用Selenium爬取水质数据，通过Sqoop将其由MySQL导入Hive数仓并完成数据清洗。
  - 使用PySpark对数据进行多维度聚合分析，并运用Pyecharts 将分析结果转化为动态交互式图表进行可视化。
  - 使用Flask框架搭建Web应用并整合所有图表，最终完成在Linux 系统上的项目部署与维护。

### 基于XLM-RoBERTa的多语言命名实体识别系统

XLM-RoBERTa PyTorch Hugging Face Flask BiLSTM-CRF

- **项目描述：**  
设计并实现了一个完整的多语言命名实体识别（NER）系统。该系统基于XLM-RoBERTa模型，能够识别处理中文、英文、德文和西班牙文四种语言的33类细粒度实体，并提供Web可视化界面进行交互式分析。
- **工作内容：**
  - 设计并实现了两种模型架构：基于XLMRobertaForTokenClassification的标准模型，以及融合了BiLSTM和CRF层的增强模型。
  - 开发了完整的数据预处理流水线，解决了多语言数据集（MultiCoNER v2）的子词切分、标签对齐和序列标准化问题。
  - 使用Flask构建后端推理服务，封装模型接口，实现了对用户输入文本和批量文件的识别处理。
  - 采用HTML/CSS/JavaScript技术栈构建前端交互界面，实现了实体高亮、结果统计和动态饼图等可视化功能，提升了系统的易用性和分析效率。