

程中荣

求职意向: 数据工程师 | 期望城市: 北京

男 | 27 岁 | wechat: cheng_marc | +86 156-5240-1103

chengmarc.com | linkedin.com/in/chengmarc | github.com/chengmarc

教育背景

多伦多大学 (University of Toronto)

2019 年 9 月 - 2023 年 6 月

荣誉理学学士 — 数学与应用 (概率/统计 Specialist) + 数学 (Major)

加拿大, 多伦多

- GPA 3.6/4.0, 高优毕业 (With High Distinction), 院长嘉奖名单: 2021、2022、2023
- 数学课程: 实分析、复变函数、常微分方程、偏微分方程、数论、组合数学、非线性优化
- 统计课程: 概率论、数理统计、时间序列分析、多元数据分析、数据分析方法、计算机科学导论

工作经历

前端开发工程师

2025 年 6 月 - 至今

DoubleTrends

加拿大, 多伦多

- 独立开发并部署 [doubletrends.com](#), 技术栈为 Tailwind CSS + Node.js, 实现完整响应式界面。
- 使用 Cloudflare Pages 完成部署闭环; 集成 Stripe 支付接口, 支持付费功能与用户交易。

产品数据工程师

2023 年 6 月 - 2025 年 5 月

蓝格赛电气有限公司 (Rexel Canada)

加拿大, 密西沙加

- 主导新品 PIM 系统 (Informatica P360) 端到端实施, 实现 20 万 + SKU 跨系统数据一致性管理, SKU 产品数据的数据建模与 workflow 设计, 数据链路覆盖 AS400 → TIBCO → P360 → SAP Hybris。
- 担任 SAP Hybris、ERP 系统 (IBM AS400) 与外部数据源之间的主要数据接口, 负责跨系统数据一致性管理。
- 使用 AWS CLI 与 Bash 开发 ETL 自动化质量控制管道, 实现日常数据验证全链路的批处理自动化。
- 设计并执行 UAT 测试框架与数据验证规则, 协调加拿大、印度、法国三地团队完成系统上线验收。
- 参与企业内部 LLM 项目 (RexelGPT) beta 测试与性能优化。

数据分析师

2021 年 7 月 - 2022 年 4 月

德勤 (Deloitte)

加拿大, 多伦多

- 为加拿大某一级银行执行 AML 反洗钱模型验证, 评估模型假设、特征选择与统计稳健性, 完整交付 AML 模型验证报告及风控模型合规评估。
- 在 SR 11-7 模型风险管理框架下, 对 AML 反洗钱分类模型 (KNN、K-Means、XGBoost) 进行验证, 核查大规模交易数据集 (100 万至 2000 万条, Oracle 数据库) 的特征一致性与管道逻辑。
- 使用 Python、R、SQL 进行模型逻辑核查与数据管道验证, 识别理论设计与生产行为之间的不一致性。
- 使用 SSMS 对 LoyaltyOne 用户忠诚度模型进行数据管道核查。

数据质检员 (实习)

2019 年 6 月 - 2019 年 7 月

北京尚德科技有限公司

中国, 北京

- 通过录音系统监控销售与客服记录; 整理并分析约 5 万条 SQL 客户数据, 制作数据透视表, 评估 6 个子部门绩效。
- 发现公司激励系统存在可被员工利用的结构漏洞, 主导 5 人小组讨论并提出处理方向, 获部门主任认可。

项目经历

PaySim 金融交易风控离线数仓 | PySpark, Hive, Hadoop | GitHub

2026 年 3 月

- 基于 PySpark + Hive + Hadoop 构建离线数仓, 处理 630 万条 PaySim 金融交易记录, 覆盖完整 ETL 链路 (Ingest → Transform → Aggregate)。
- 数仓分 ODS / DWD / ADS 三层: ODS 层采用 EXTERNAL TABLE 设计, DROP 操作仅清除元数据, HDFS 原始数据保持不变, 从架构层避免删库风险。
- DWD 使用 ORC + Snappy 存储, 以 (tx_type, tx_day) 双分区, tx_type 作顶层分区键, 保证 ORC 谓词下推在目录根节点完成分区剪枝。

Stoic AI - MAMBA 状态空间模型 | PyTorch, RunPod | GitHub

2024 年 12 月 - 2025 年 3 月

- 主导 MAMBA 状态空间模型的架构设计与工程实现 (选择性 SSM, 双分支投影, 深度卷积, SiLU 门控), 应用于多特征时间序列预测任务。
- 设计去趋势预处理管道: 对数趋势提取 → 残差预测 → 逆变换重建, 从源头消除金融时序预测 lag-1 退化问题。
- 引入 6 维链上特征, 对比 SegRNN、LSTM、Seq2Seq LSTM、Attention LSTM、Transformer 五个基线; MambaSSM 取得最优预测性能。
- 项目原为个人独立开发, 架构被 Stoic AI 团队引用并集成; 协助建设量化交易预测基础设施, 搭建基于 RunPod H100 的云端训练与快速实验迭代环境。

- 从零基于 PyTorch 复现 GPT-2: 1.62 亿参数, 768 维嵌入, 12 头因果自注意力, 12 层 Transformer。
- 实现全部核心组件: pre-LayerNorm、GELU (tanh 近似)、因果掩码 ($-\infty$ pre-softmax)、BPE 分词 (tiktoken); 使用 AdamW 训练。

技术栈

大数据与数仓: Hadoop、Hive、Spark、PySpark、HDFS、YARN

数据库: Oracle、MySQL、IBM DB2 (AS400)、SQL Server、BigQuery

云与基础设施: AWS、Azure、RunPod、Cloudflare

数据科学与机器学习: PyTorch、scikit-learn、pandas、numpy、scipy、matplotlib、seaborn、jupyter

企业系统: Informatica P360、SAP Hybris、Tableau、Power BI

编程语言与工具: Python、R、JavaScript、Node.js、Bash、PowerShell、Git、LaTeX、Markdown

自然语言: 英语 (母语级)、法语 (专业级)、普通话 (母语)

学术论文

Zhongmang (Marc) Cheng. "On the Impossibility of Complete Interpretability in Neural Networks: A Measure-Theoretic Analysis." *Zenodo preprint*, Feb. 2026. DOI: [10.5281/zenodo.18678067](https://doi.org/10.5281/zenodo.18678067)

奖项

多伦多大学 Dean's List: 2021、2022、2023

拉瓦尔大学 Concours Opti-math: 全省第 5 名 (2015)、全省第 7 名 (2014)

滑铁卢大学 Gauss 数学竞赛: 优异奖 (2013)