



英特尔®至强®可扩展处理器

Database Firewall

基于英特尔® 至强® 处理器的数据库防火墙系统



头条



大模型会毁了初级程序员 —— 对话图灵奖得主 Joseph Sifakis | 新程序员

从 ChatGPT 引发百模大战，GPTs 让人都能用自然语言构建 GPT，全球范围内对于 AGI 通用人工智能的探索日渐深入，而计算机领域关于“超智能”的神话也愈演愈烈，其中一个广泛传播的观点是...

一个 24 年 IT 老兵的码农生涯！

“技术为王”走不通晋升 CTO、创业这...

最赚钱的编程语言！Java 垫底，...

年薪最高可达 75 万美元

微软重磅推出 Windows App

硬件上的围墙，微软正在用软件去逐步...

张量命名争议后续：大模型极限竞...

由零一万物 Yi 大模型引发的争议

亚马逊被曝正在自研一款基于 Lin...

摒弃 Android！

热点

Sam Altman或回归OpenAI

Vite 5.0正式发布 | 极客头条

AI 时代的 GPU 生存工具包

每个开发人员必须了解的基本知识

Xiaomi Vela 全面开源

微软推全新“Windows App” | 极客头条

OpenAI 暂停 ChatGPT Plus 新用...

李开复再度回应争议：取之开源也贡献...

我又爱上了「将死」的 Rails

Ruby on Rails 从兴盛到衰落



昇腾AI 向上的力量

每位开发者都了不起



大模型长上下文运行的关键问题

上下文长度的增加是 LLM 的一个显著发展趋势。过去一年，几种长上下文语言模型陆续问世，包括 GPT-4 (32k上下文)、MosaicML 的 MPT (65k上下文)、Anthropic 的 Claude (100k上下文) 等。然而，扩大 Transformer 的上下文长度是一个挑...

1 赞 0 踩 作者: OneFlow_Official



语言大模型的进化轨迹

ChatGPT的发布是语言大模型（LLM）发展史的转折点，它让人们意识到LLM的潜力，并引发了“AI竞赛”，世界上主要人工智能实验室和初创公司都参与其中。在这之后，基于LLM的聊天机器人层出不穷。ChatGPT及相关LLM模型让我们共同见证了AI的历...

1 赞 0 踩 作者: OneFlow_Official



突发，OpenAI 深夜变天，Sam Altman 被踢出局，原 CTO 暂代临时 CEO

【CSDN 编者按】11月18日凌晨，OpenAI 突然发布一则官方声明，宣布 Sam Altman 经过慎重的审查程序后离开公司，首席技术官 Mira Murati 作为临时 CEO 暂代工作。作者 | yggs出品 | CSDN (ID: CSDNnews) 一觉醒来发生了两件大事：一是 Open...

2 赞 0 踩 作者: CSDN资讯



30 位学者合力发表 Nature 综述，10 年回顾解构 AI 如何重塑科研范式

科学发现是一个复杂过程，涉及到几个相互关联的阶段，包括形成假设、实验设计、数据收集及分析。近年来，AI 与基础科研的融合日益加深，借助 AI，科学家得以加速科研进度、促进科研成果的落地。

0 赞 0 踩 作者: HyperAI



关于语言大模型的八大论断

近几个月来，语言大模型（LLM）的广泛公开部署引起了倡导者、政策制定者和许多领域学者们的新一轮关注和参与。Anthropic 技术团队成员、纽约大学助理教授 Samuel R. Bowman 总结了八个可能引发思考的观点，并讨论了 LLM 还存在的局限性。即便没...

1 赞 0 踩 作者: OneFlow_Official



LoRA和QLoRA微调语言大模型：数百次实验后的见解

LoRA是用于训练自定义LLM的高效参数微调技术。本文作者Sebastian Raschka通过成百上千次实验，他为使用LoRA和QLoRA对LLM进行微调提供了实用见解，包括节省内存、选择最佳配置等。Sebastia是威斯康星大学麦迪逊分校的统计学助理教授，也是...

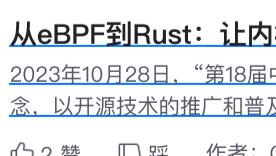
2 赞 0 踩 作者: OneFlow_Official



使用 Redis 构建轻量的向量数据库应用：图片搜索引擎（二）

本篇文章我们来继续聊聊轻量的向量数据库方案：Redis，如何完成整个图片搜索引擎功能。

0 赞 0 踩 作者: soulteary



从eBPF到Rust：让内核可观测、可编程和更安全

2023年10月28日，“第18届中国Linux内核开发者大会（CLK）”在深圳开幕。作为中国Linux内核领域最具有影响力的峰会之一，CLK以“自由、协作、创新”为理念，以开源技术的推广和普及为使命，促进Linux内核爱好者相互交流与进步。本届大会设置了主论坛，以及内存管理、云和服务器、Arch&虚拟化&I/O、...

2 赞 0 踩 作者: CSDN资讯



揭秘编码器与解码器语言模型

Transformer架构的问世标志着现代语言大模型时代的开启。自2018年以来，各类语言大模型层出不穷。通过LLM进化树（github.com/Mooler0410/LLMsPracticalGuide）来看，这些语言模型主要分为三类：一是“仅编码器”，该类语言模型擅长文...

向你推荐

所有活动 >

CSDN 指数

「新手期成长挑战」创作获流量

云原生系列在线峰会

「开源」CSDN数据地图

英特尔开发人员专区

Eolink开发者社区

华为开发者联盟HarmonyOS专区

ZEGO 即构开发者社区

松山湖开发者村综合服务平台

腾讯云开发者社区

百度AI Studio社区

华为云开发者联盟

Authing身份云

Unity官方开发者社区

兴智开发者社区

LAVAL社区

学院课程

开放原子开发者工作坊

直播

更多 >

直播预约



筑基赋能·智向未来--九...

11/21 14:00

直播预约



AI绘画火爆后，如何利用...

AIGC 抓住下一个内容风...

11/23 14:00

直播预约



2023 IDEA大会

About IDEA Beyond Ideas

11/22 09:40

直播预约



汽车行业软件代码能安全与信息安全的最...

11/29 20:00

1 赞 0 踩 作者: OneFlow_Official



英伟达的AI霸主地位会持久吗?

目前英伟达因其 GPU 芯片在 AI 革命中扮演着核心角色，使其成为 AI 时代最赚钱的公司。本文作者 Pete Warden 总结了铸就英伟达 AI 霸主地位的四点优势：几乎无人运行大规模机器学习应用；英伟达的替代品都很糟糕；研究人员掌握着硬件采购的风向舵；训...

1 赞 0 踩 作者: OneFlow_Official



SQL拦截：想要限制每次查询的结果集不能超过10000行，该如何实现？

对于一些SaaS化软件，当某个租户在执行查询SQL时，如果查询条件出现了BUG，导致去查了所有租户的数据，这种情况是非常严重的，此时就需要在架构层面做限制，禁止一些特殊SQL的执行，另外，为了保护数据库，也可能会限制某些查询语句不要查...

1 赞 0 踩 作者: 爱读源码的大都督



开源软件 FFmpeg 生成模型使用图片数据集

本篇文章聊聊，成就了无数视频软件公司、无数在线视频网站、无数 CDN 云服务厂商的开源软件 ffmpeg。分享下如何使用它将各种视频或电影文件，转换成上万张图片数据集、壁纸集合，来让下一篇文章中的模型程序“有米下锅”，这个方法特别适合宫崎...

1 赞 0 踩 作者: soulteary



语言大模型推理性能工程：最佳实践

在这篇文章中，MosaicML 工程师团队分享了如何在生产环境中充分利用流行开源语言大模型（LLM）的最佳实践。此外，他们还提供了围绕模型部署推理服务的指南，以帮助用户更好地选择模型和部署硬件。他们在生产环境中使用了多个基于 PyTorch 的后...

1 赞 0 踩 作者: OneFlow_Official



有趣的 TCP 抢带宽行为

总有人说不受控的 udp 要比 tcp 快，其实一个优秀的 udp-based 协议并不比 tcp 快，它至少把 tcp 那些东西重新在 udp 上实现了一遍，比如 quic，最后就成了 yet-another tcp 了。如果放宽算法的公平性约束，抢带宽，让带宽就自然多了，非常像高...

1 赞 0 踩 作者: dog250

云原生周刊：KubeSphere 3.4.1 发布 | 2023.11.13

它在 Kubernetes 集群中管理 eBPF 程序的打包、部署和执行，包括许多基于 BCC 工具的程序，以及一些专为在 Inspektor Gadget 中使用而开发的程序。对于 CPU 资源的调整，则不需要重启容器。原地 Pod 资源调整是 Kubernetes 中的一个新功能，允许在不重启 Pod 的情况下调整分配给 Pod 的 CPU 和内存资源...

0 赞 0 踩 作者: KubeSphere 云原生



为什么开源语言大模型很重要？

在LLM领域，开源究竟意味着什么？假设开源社区拥有了真正的开源LLM，其权重、数据集、代码和基础设施都可公开获取，我们又将从中获得哪些重要收益？本文作者为Vikram Sreekanti和Joseph E. Gonzalez，前者是Aqueduct的联合创始人及CEO，后...

7 赞 0 踩 作者: OneFlow_Official



网易有道上线“易魔声”开源语音合成引擎 用户可免费下载使用

共建良好技术生态，让AI延续声音温度。

3 赞 0 踩 作者: 有道AI情报局



GPU架构与计算入门指南

大多数工程师对CPU和顺序编程都十分熟悉，这是因为自从他们开始编写CPU代码以来，就与之密切接触。然而，对于GPU的内部工作原理及其独特之处，他们的了解则相对较少。过去十年，由于GPU在深度学习中得到广泛应用而变得极为重要。因此，每位...

3 赞 0 踩 作者: OneFlow_Official



“AI教父”Geoffrey Hinton：智能进化的下一个阶段

ChatGPT等大模型带来的震撼技术革新，让Geoffrey Hinton突然改变了自己的一个想法。这位75岁的“人工智能教父”意识到，数字智能优于生物智能的进程无法避免，超级智能很快就会到来，他必须要对其风险发出警示，而人类需要找到一种方法来控制AI...

1 赞 0 踩 作者: OneFlow_Official

活动日历

更多 >

- 21 北京 筑基赋能·智向未来--九章云极Data
11月 aCanvas大模型系列成果发布会
- 25 上海 开放原子开发者工作坊 第四期
11月
- 26 上海 2023腾讯技术创作特训营 | 技术
11月 创作如何赋能技术人的职场内与外?
- 8 北京 Flink Forward 峰会 | Flink Forwar
12月 d Asia 2023
- 10 线下 Google DevFest 2023 上海站
12月

竞赛平台

更多 >

亚马逊云科技

『云上探索实验室』Amazon Codewhisperer 挑战周赛
解锁Codewhisperer，码上实战等你来
① 2023年10月7日-10月13日

立即报名

【云上探索实验室】Amazon CodeWhisperer 挑战周赛
主办方: 亚马逊云科技 协办方: CSDN

会员精选

更多 >

VIP Java8编程实战

VIP 面试之排序算法

VIP 实用数据分析：数据分析师从小白到精通

VIP Java之路

VIP 韦东山嵌入式Linux第一期视频

VIP 编程可以这样学

推荐专题

更多 >

浪潮信息云峦服务器操作
系统征文挑战赛

阿里云 免费领取阿里云
云服务器ECS
部署Java Web环境
赢大奖

星火培训【感知专项营】

以网强算，中国移动提出
“算力网络”全新技术方案

Qualcomm 开发者

英特尔开发人员专区



BSV区块链中文开发者社区



ConsenSys Web3社区



上海城市开发者社区
原创 999+ 粉丝 3395



深圳城市开发者社区
原创 999+ 粉丝 4589



长沙城市开发者社区
原创 999+ 粉丝 376



武汉城市开发者社区
原创 999+ 粉丝 4196



重庆城市开发者社区
原创 999+ 粉丝 13258



合肥城市开发者社区
原创 999+ 粉丝 2262



视觉中国

BOSS直聘

看准网

优设网

View UI

DevPress社区

CSDN文库

Edge插件

您愿意向朋友推荐“CSDN首页”吗?



联系我们 (工作时间: 8:30-22:00)

400-660-0108 kefu@csdn.net 在线客服

关于我们 | 招贤纳士 | 商务合作 | 寻求报道

京ICP备19004658号 经营性网站备案信息

公安备案号11010502030143 营业执照

北京互联网违法和不良信息举报中心 家长监护

中国互联网举报中心 网络110报警服务

Chrome商店下载 账号管理规范

版权与免责声明 版权申诉 出版物许可证

©1999-2023北京创新乐知网络技术有限公司

