

亚马逊云科技上线Meta Llama 3.2模型

为构建生成式AI应用提供更多选择

北京2024年10月10日 /美通社/ -- 亚马逊云科技宣布，Meta的新一代模型Llama

3.2，包括其首款多模态模型，现已在Amazon Bedrock和Amazon SageMaker中正式可用。客户也可以在基于Amazon Trainium和Amazon Inferentia的Amazon Elastic Cloud Compute（Amazon EC2）实例中部署这些模型。

Llama 3.2系列模型保留了此前的成功经验，并进行了全新升级，提供高度差异化的功能更新，包括支持图像推理的小型 and 中型视觉大语言模型，以及针对设备端优化过的轻量级纯文本模型。这些新模型旨在提高使用的便捷性和效率，同时特别注重负责任的创新 and 安全性。此外，针对内容安全分类而微调的Llama Guard 3 Vision模型，现在也已在Amazon SageMaker JumpStart中可用。

Llama 3.2系列模型优势：

Meta推出的首款多模态视觉模型：Llama 3.2 11B Vision和Llama 3.2 90B Vision

Llama 3.2系列中最大的模型。支持图像理解和视觉推理的使用场景。

擅长分析视觉数据，如图表和图形，提供更精确的答案和洞察。适用于图像标注、视觉问答、图像与文本检索、文档处理、多模态聊天机器人，以及长篇文本生成、多语言翻译、编程、数学和高级推理。

为边缘和移动设备量身定制：Llama 3.2 1B和Llama 3.2 3B

轻量级，纯文本模型。既可在云端又可在本地进行数据处理，响应速度极快。非常适合高度个性化的应用，如文本生成与总结、情感分析、客户服务应用、文本润色、多语言知识检索和移动AI写作助手。

针对内容安全分类进行微调：Llama Guard 3 11B Vision

仅在Amazon SageMaker JumpStart中可用 Llama Guard 3 11B

Vision可以对大型语言模型的输入（即提示词分类）和输出（即响应分类）进行内容安全保护 专为支持图像推理场景设计，提升了对提示词输入中的有害多模态内容（文本和图像）的检测能力，以及模型输出的文本内容检测。

Meta表示，Llama 3.2模型已在超过150个基准数据集上进行评估，展现出与领先基础模型相媲美的竞争力。与Llama 3.1类似，所有Llama 3.2模型均支持128K的上下文长度，并支持涵盖八种语言的多语言对话使用场景，包括英语、德语、法语、意大利语、葡萄牙语、印地语、西班牙语和泰语。

原文地址：<http://www.china-nengyuan.com/news/216204.html>