

亚马逊宣布推出新一代基础模型Amazon Nova

亚马逊的新一代基础模型提供前沿智能和行业领先的性价比

Amazon Nova模型扩展了Amazon Bedrock中针对企业客户提供的丰富且强大的基础模型选择

北京2024年12月4日 /美通社/ -- 亚马逊在2024 re:Invent全球大会上，宣布推出新一代基础模型Amazon Nova，这些模型在多种任务上展现出优秀智能，且具备行业领先的性价比。Amazon Nova模型将在Amazon Bedrock中提供，包括超快速文本生成模型Amazon Nova Micro，以及能够处理文本、图像和视频并生成文本的多模态模型Amazon Nova Lite、Amazon Nova Pro和Amazon Nova Premier。此外，亚马逊还推出了两个全新模型——用于生成高质量图像的Amazon Nova Canvas和用于生成高质量视频的Amazon Nova Reel。

亚马逊通用人工智能高级副总裁Rohit Prasad表示：“在亚马逊内部，我们约1000个生成式AI应用正在进行中，因此可以全面了解开发者所面临的挑战。我们的新一代Amazon Nova模型旨在帮助内外部开发者应对这些挑战，提供强大的智能和内容生成功能，并在延迟、成本效益、定制化、检索增强生成（RAG）和智能体能力等方面取得显著进展。”

Amazon Nova理解模型展现出卓越的智能、能力与速度

Amazon Nova包含四款先进的模型。Amazon Nova Micro是一款仅处理文本的模型，能够以极低的延迟和成本提供响应。Amazon Nova Lite是一款极具成本效益的多模态模型，能够快速处理图像、视频和文本输入；Amazon Nova Pro是一款功能强大的多模态模型，可在准确性、速度和成本之间实现绝佳平衡，适合多种任务；Amazon Nova Premier是亚马逊的顶级多模态模型，专为复杂推理任务而生，并可用于作为“教师模型”来蒸馏定制模型。目前，Amazon Nova Micro、Amazon Nova Lite和Amazon Nova Pro已全部正式可用，Amazon Nova Premier预计将在2025年第一季度可用。

亚马逊对Amazon Nova模型进行了广泛的行业基准测试。结果表明，Amazon Nova Micro、Amazon Nova Lite和Amazon Nova Pro在各自类别中与最佳模型相比均表现得相当有竞争力。

Amazon Nova Micro在所有11项适用基准测试中，其表现与Meta的LLaMa 3.1 8B相当或更优，而在所有12项适用基准测试中，其表现也与Google Gemini 1.5 Flash-8B持平或更优。凭借每秒210个输出token的行业领先速度，Amazon Nova Micro非常适合需要快速响应的应用场景。

Amazon Nova Lite在同类模型中同样具备强劲竞争力。与OpenAI的GPT-4o mini相比，它在19项基准测试中有17项表现持平或更优，与Google的Gemini 1.5 Flash-8B相比，在21项基准测试中也有17项表现相当或更优，与Anthropic的Claude Haiku 3.5相比，则在12项基准测试中有10项表现相当或更优。除了在文本基准测试中展现出色的准确性外，Amazon Nova Lite在视频、图表和文档理解方面也表现突出，这一点通过VATEX、ChartQA和DocVQA等基准测试得到了证明。此外，Amazon Nova Lite在智能体工作流上也表现优异，如在Berkeley Function Calling Leaderboard（伯克利函数调用排行榜）上的函数调用能力，以及在理解视觉元素以在浏览器和计算机屏幕上执行操作的核心能力方面表现出色，这一点在VisualWebBench（网页浏览器操作基准测试）和Mind2Web（通用多模态代理基准测试）得到证明。

Amazon Nova Pro在与OpenAI的GPT-4o的20项基准测试中，有17项表现相当或更优；在与Google的Gemini 1.5 Pro的21项基准测试中，有16项表现持平或更优；在与Anthropic的Claude Sonnet 3.5v2的20项基准测试中，有9项表现相当或更优。除了在文本和视觉智能基准测试中的准确性外，Amazon Nova Pro在遵循指令和多模态智能体工作流方面也表现出色，相关测评包括综合RAG（CRAG）、Berkeley Function Calling Leaderboard（伯克利函数调用排行榜）和Mind2Web。

多语言和多模态支持，具备长上下文处理能力

Amazon Nova Micro、Lite和Pro支持200多种语言。Amazon Nova Micro支持128K输入token的上下文长度，而Amazon

Nova Lite和Pro的上下文长度可达300K token，处理视频时长可达30分钟。到2025年初，Amazon将支持超过200万输入token的上下文长度。

快速且具有成本效益

Amazon Nova模型快速，具有成本效益，且设计上易于与客户的系统和数据无缝集成。Amazon Nova Micro、Lite和Pro在各自智能类别中，成本比Amazon Bedrock中表现最佳的模型至少便宜75%，同时也是Amazon Bedrock中对应类别速度最快的模型。

与Amazon Bedrock的无缝集成

所有Amazon Nova模型均已与Amazon Bedrock集成。Amazon Bedrock是一个完全托管的服务，通过单一API提供来自领先AI公司的高性能模型。客户可以轻松试用和评估Amazon Nova模型及其他模型，找到最适合其应用的模型。

支持微调以提升准确性

这些模型还支持自定义微调，允许客户根据自身专有数据中的已标记示例来提升准确性。Amazon Nova模型能够从客户的数据（包括文本、图像和视频）中学习最重要的信息，随后Amazon Bedrock会训练一个私有的微调模型，以提供定制化响应。

蒸馏用于训练更小、更高效的模型

除了支持微调外，这些模型还支持蒸馏，可以将规模更大、实力更强的“教师模型”中的特定知识转移到更小的模型上，从而实现更高的精确度、更快的响应速度以及更低的运营成本。

基于数据响应的检索增强生成（RAG）

Amazon Nova模型与Amazon Bedrock知识库紧密集成，专注于检索增强生成（RAG）方法，使客户能够通过利用自身的数据确保生成内容的高度准确性和相关性。

针对智能体应用进行优化

Amazon

Nova模型经过优化，便于智能体应用在执行多步骤任务时通过多个API与企业的专有系统和数据进行交互。

借助Nova创意内容生成模型，访问生产级视觉内容

Amazon Nova Canvas是一款先进的图像生成模型，能够根据输入的文本或图像生成专业级的视觉内容。它提供了便捷的编辑功能，用户可以通过文本轻松调整图像，控制颜色方案和布局。该模型内置的安全控制措施确保AI的安全和负责任使用，包括水印功能，能够追溯每张图像的来源；以及内容审核功能，限制可能产生的有害内容。根据第三方的对比评估，Amazon Nova Canvas在性能上优于OpenAI的DALL-E 3和Stable Diffusion，并在关键的自动化指标上表现出色。

Amazon Nova Reel是一款先进的视频生成模型，能够帮助客户轻松从文本和图像创建高质量视频，适用于广告、营销和培训等内容创作。客户可以使用自然语言提示控制视觉风格和节奏，包括相机运动、旋转和缩放。根据第三方的评估，Amazon Nova Reel在质量和一致性上优于同类模型，客户更倾向于选择由其生成的视频，而非Runway的Gen-3 Alpha生成的视频。与Amazon Nova Canvas一样，Amazon Nova Reel也具备内置的安全控制措施，包括水印和内容审核。目前，Amazon Nova Reel支持生成六秒的视频，未来几个月将支持最长为两分钟的视频生成。

下一步计划：语音到语音和多模态到多模态模型

亚马逊计划在2025年第一季度推出Amazon Nova语音到语音模型。该模型旨在通过理解自然语言的流式语音输入，解释语言和非语言信号（如语调和节奏），提供流畅的类人交互，彻底改变对话式AI应用，确保低延迟的双向交流。

此外，Amazon还在开发一种新模型，能够接受文本、图像、音频和视频作为输入，并以任意模态生成输出。这款具备本地多模态到多模态——即“任意到任意”模态能力的Amazon Nova模型预计将在2025年中旬推出。该模型将简化应用开发，使同一模型能够执行多种任务，如内容模态转换、内容编辑，以及驱动能够理解和生成所有模态的AI代理。

亚马逊云科技的合作伙伴和客户已开始使用Amazon Nova模型，从其强大功能和超高性价比中获益

亚马逊云科技的战略合作伙伴SAP正在将Amazon Nova模型集成到SAP AI Core生成式AI中心支持的大语言模型（LLMs）家族中。这一举措使开发者能够为SAP的AI助手Joule创建新技能，并利用SAP数据中完整的业务上下文，安全地构建AI驱动的解决方案，从而实现自动化、个性化以及供应链规划等高级功能。

德勤作为亚马逊云科技的战略合作伙伴，致力于为全球各行业提供一流的生成式AI服务。德勤深知，没有一种AI解决方案和基础模型能解决所有问题，认为Amazon Nova模型的高级定制能力和增强安全性将推动创新，为全球客户带来卓越的价值。

Amazon致力于开发负责任的人工智能

Amazon Nova模型内置安全和保护措施。亚马逊还推出了Amazon AI服务卡，提供有关使用场景、局限性和负责任AI实践的透明信息。

原文地址：<http://www.china-nengyuan.com/news/218487.html>