

机器之心报道编辑：杜伟、泽南大模型的未来时刻，已经来了？速度太快了。商汤一下子把多模态大模型的发展进度条，快进到了落地阶段。商汤的大模型体系「日日新 SenseNova」今天刚刚发布了 4.0 版，不论语言能力还是文生图能力都有全面升级，还自带低门槛的落地工具。新一代 SenseNova 不仅在大语言模型、文生图模型等方面进行了重大升级，部分垂直领域能力超越 GPT-4，还发布了全新多模态大模型，并面向数据分析、医疗等场景提供了全新版本，让大模型通用能力适配到了更多领域。API 申请网址：<https://platform.sensenova.cn/>

与此同时，商汤还推出了日日新·商量大语言模型 Function call \u0026 Assistants

API 版本，除了对话能力外，还支持多种内置工具，包括图片生成（文生图）、智能识图（图生文）、数据分析（代码解释器）、在线检索。这是全球首个支持了文生图、图生文，并可支持不同模态工具调用的工具，跑在了 OpenAI 的前面。这一系列发布，从技术进步到落地「两翼齐飞」，可谓把通用大模型技术卷上了天。看来在技术竞争中，国内科技公司逐渐有了反超的趋势。最高支持 128k 长窗口商量 SenseChat 测试全方位比肩 GPT-4 自 ChatGPT 出现以来，大模型成为了 AI 赛道的主力军。商汤的大模型体系正在「大模型 + 大装置」的战略布局下快速迭代。去年 4 月，商汤公布了「日日新 SenseNova」大模型体系，一上来就在自然语言处理、文生图创作、数字人生成、3D 场景和物体生成，自动化数据标注、自定义模型训练等多个领域全面发力。与此同时，商汤还直接提供图片生成、自然语言对话、视觉推理和标注服务的 API 接口。此后，该大模型体系持续推陈出新，在基础能力、API 服务、模型应用等多个方面不断进步，给用户和开发者们带来了越来越好用的技术。如今，近 10 个月过去了，商汤新一代「日日新 SenseNova 4.0」在 2024 年的新春之际与大家见面了，不仅对已有多个大模型进行全方位升级，还有一些「新面孔」。升级之后，日日新在长文本理解、综合推理（包括数字推理）、代码生成、多模态交互等整体表现上「更上一层楼」，不仅全面超越了 GPT-3.5，并且大部分接近甚至超越了 GPT-4 系列模型。用下面一组核心数据说话，SenseNova 4.0 的：推理能力：达到 GPT-4 Turbo 的 99%；代码能力：在 HumanEval 代码生成基准测试上准确率达到 75.6，超越 GPT-4 (74.4)；多模态能力：在 MMBench 多模态大语言模型综合评估基准上的整体性能超越了 GPT-4V (84.4 vs 74.4)；数据分析能力：正确率 (85.71%) 超越 GPT-4 (84.62%)；在部分垂直领域能力超越 GPT-4 Turbo。而日日新全维度、无死角的能力飙升，首要归功于商量大语言模型 SenseChat 的重大升级。此次发布的商量大语言模型-通用版本 (SenseChat V4) 在整体能力比肩 GPT-4，并相较于 GPT-3.5 实现显著超越。如下两图为 SenseChat V4 与 GPT-3.5、GPT-4 在整体、考试、语言、知识、推理、数理、编程等数据集上的性能比较数据。至于为何能有如此明显的性能提升，SenseChat 4.0 在以下多个方面获得了加强。首先是更全面的知识覆盖，新增了包括业务通用数据、数学能力数据、K12 考试数据、文学期刊数据等在内约 600B tokens 的中英文预训练语料，这样理解多领域内容更加得心应手。同时，模型质量也通过数据清洗和增强得到进一步提高。其次推理能力变得更加可靠。从初始 1.0 版本以来，前后四次超强预训练的积累让模型在阅读理解、综合推理、代码能力等多项任务上实现了 5%-10% 的定向性提升。最后也是此次 4.0 版本升级的重点——更强的长文本理解分析能力，更新了 3 种不同上下文窗口的全新模型，即 SenseChat-4K、SenseChat-32k 和 SenseChat-128k，不仅使得模型理解上下文的能力迎来史诗级加强，还提升了

模型的适应能力，拓宽了应用范围，为用户提供根据需求自由选择模型的机会。在与 GPT-3.5、GPT-4 的多任务较量中，我们直观地看到了 SenseChat 不同上下文窗口版本的真正实力。其中，SenseChat-4K 虽然支持最少的 4k tokens（约 4000 中文字）的输入和输出，但仍然在写作总结、知识问答、闲聊娱乐、专业技能、安全测试等主客观题和安全性能上超越了 GPT-4。另外，新增的引文功能还可以返回在线搜索的知识来源。SenseChat-32k 则能够处理 32k tokens（约 3 万中文字）的长文本总结，总能力平均得分达到了同等上下文窗口 GPT-4-32k 能力的 90% 以上水平，中文理解能力则超越了后者。铺开来讲，SenseChat-32k 在平均考试能力和理解能力、以及 HellaSwag、C3、LAMBADA、CHID 等推理和理解类测试集中超越 GPT-4-32k；在 LongBench 长文本理解测试基准以及 tpo、multidocqa、scientificqa、PassageRetrieval-zh 等长文本测试集上均超越了 GPT-4-32k。对于支持最长 128k tokens（约 12 万以上中文字）长文本的 SenseChat-128k，它的中文理解能力也超过了 GPT-4 的水平。下表 1 和 2 分别为 SenseChat 三个版本模型与 GPT 系列在长文本理解和推理等测试集上的平均得分比较。表 1：Normalbench v1-4 万题对比结果。表 2：长文本 Leval 和 Longbench 测试集对比结果。看起来，SenseChat V4 不仅在主客观题方面达到了 GPT-4 的水平，更在长文本理解和推理能力上实现了全面超越。作为商汤「日日新 SenseNova」大模型体系的通用基础模型，SenseChat V4 的大幅度升级使得人们在使用模型处理多样化语言任务时更高效、更准确，让国产大模型拥有不输于 GPT-4 的使用体验。对于更多人来说，未来在商量 SenseChat 大语言模型的基础上开展学术研究、技术创新、商业应用也有了更多机会。填补行业空缺，打造专用大模型首家开放支持多模态的 Assistants API 基础模型之外，商汤也希望能通过高效融合垂直领域知识，帮助人们构建各类专业大模型，降低大模型的下游应用成本和门槛。多模态是人工智能大模型重要的技术演进方向，新一代「日日新 SenseNova」推出了拥有 300 亿参数的日日新·商量多模态大模型

（SenseChat-Vision V4），其图文感知能力处于全球领先水平，在权威评测基准测试集 MME Benchmark 上综合得分排名首位。目前，该模型可以支持智能驾驶、智能车舱、电力行业等多个实际场景的应用。与常规的 OCR 能力不同，它不仅可以理解图中的文字和物体，并且可以根据逻辑进行推理，实现了一定程度的认知能力。在办公与数据分析领域，商汤推出了日日新·商量语言大模型-数据分析版本（SenseChat-DataAnalysisCode V4），它可以通过自然语言输入，结合商汤大模型的意图识别、逻辑理解与代码解释器的能力，自动将数据转化为有意义的分析和可视化结果。目前，该工具已经支持 xls、xlsx、csv、txt、json 等格式的文件和表格处理。就实际效果而言，办公小浣熊在 1000+ 测试集精度上略胜于 GPT-4。 体验入口：<https://raccoon.sensetime.com/office> 在医疗健康领域，大语言模型的医疗版本也有全新升级，日日新·商量语言大模型-医疗版本“大医”（SenseChat-Medical V4）在本次更新后可以有效实现专业医学问答及复杂医学任务推理，并支持更多模态医学文件的智能解读和交互问答。据介绍，“大医”在两项行业权威评测——2023 年职业药剂师考试大模型评测和中文医疗大语言模型开放评测平台 MedBench 中，均实现综合评分排名第二，性能接近 GPT-4。商汤自研的日日新-秒画文生图大模型（SenseMirage V4）较此前版本，参数量提升至百亿量级，通过 Mixture of text experts、Spatial-aware CFG 等算法优化，语义理解能力与图像质感细节表现显著增强，可达成电影级海报生成水平。同时结合 Adversarial Distillation 算法，秒画 SenseMirage-Turbo V4 也对外发布，

相较于基础版本，可达到 10 倍推理加速效果。秒画一键生成电影海报级的精美图像再进一步，商汤还把调用不同模态的能力，做到了一个端口上，这就是全球首个支持调用不同模态的 Assistants API。去年 11 月，OpenAI 在其首届开发者大会上推出专门构建的 AI 工具——Assistants API，通过代码解释器、检索和函数调用等新功能帮助开发者构建高质量的 AI 应用。不过，至今这个工具也没有支持构建视觉相关的多模态应用。商汤提出的 Assistants API 填补了这一空缺。作为一个基于商量大语言模型构建的、具有状态的多轮对话接口，它不仅首次支持了文生图、图生文的不同模态工具调用，还内置数据分析、搜索引擎工具。如果把大模型看作是大脑，Assistants API 相当于给 AI 增加了眼睛和手，能够自主理解人类下达的任务，并做出正确规划，使用合适的资源和工具。Assistants API 提供了一个桥梁，将先进的大模型与各类应用服务工具连接起来，支持图文结合的多模态交互和代码执行结果的直观呈现，可以帮助人们快速解决复杂的问题。目前，商汤的大模型体系已经在全面落地。在全行业层面上，自发布以来已经拥有了超过 3000 家企业用户，累积调用量已达近 9000 万次，服务的行业包含互联网娱乐、游戏、文娱、教育、医疗健康、金融、编程等方面。结语还记得去年的「百模大战」吗？现在，科技领域的大模型军备竞赛形势已经有了改变，竞争不再是单纯的模型技术，而变成了拼体系——除了模型技术的升级改进，各家厂商正在整合与调优基础底座，开放的趋势也在催生出逐渐繁荣的生态。如今，战火已经燃烧到了多模态技术的落地上。能够睁开眼睛看世界的大模型，为我们带来了更多的想象力。而为了在千行百业中用好它们，真正实现「重做所有产品」，一套完整的体系势必能让我们事半功倍。在这一方面，商汤已经做到了更好。© THE END 转载请联系本公众号获得授权投稿或寻求报道：

content@jiqizhixin.com