AI 技术每日分析：电力企业开始参与超大规模人工智能基础设施的客户开发

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 11 月 28 日



摘要

日本发电企业 JERA 被报道正在与多家美国人工智能及云计算公司接洽数据中心合作，电源、机房、服务器和最终用户之间的商业协同正在加深。美国财经媒体关注 OpenAI 和 Anthropic 扩大驻场工程师队伍的做法，模型供应商的企业服务正在延伸至具体业务流程。腾讯发布面向专业三维资产生产的 Hy3D 3.6，英伟达补充公开文档解析模型的训练数据与微调方法，俄罗斯 GigaChat 团队的旗舰模型技术资料也受到关注。围绕算力供给、应用交付和专业模型的不同技术路线，人工智能产业呈现更细致的工程分工。

一、JERA 接洽 Anthropic、Meta 等客户，算力项目向发电资产附近布局

路透社于世界协调时 11 月 26 日 22 时 13 分发布报道，对应北京时间 11 月 27 日 6 时 13 分。报道援引《日本经济新闻》称，日本发电企业 JERA 正在同 Anthropic、Meta 等美国科技公司讨论在东京周边发电设施附近建设数据中心，同时涉及亚马逊、微软与谷歌等潜在客户。JERA 发言人表示，合作方 RHAELM 负责市场开发，并告知其正在同多家大型云计算企业沟通。相关客户的具体采购合同尚未公布。

今年 10 月，JERA 已经与戴尔科技及英国人工智能基础设施开发商 RHAELM 签署合作备忘录，拟从千叶地区规划投资约 150 亿美元的数据中心项目起步，探索可推广的基础设施建设模式。此次报道新增的信息集中于潜在客户的接洽进展。随着单个训练集群的规模扩大，变电容量、供电可靠性、散热系统、光纤网络和服务器交付周期逐渐需要统一协调。发电企业拥有电源资产及电网接入经验，服务器企业能够提供整机柜工程能力，云计算或模型企业则能够提供长期业务需求。围绕特定场地形成的协作关系，有利于建设方更早掌握负荷曲线及分期建设安排。

二、OpenAI 和 Anthropic 扩大面向客户的驻场工程能力

《巴伦周刊》11 月 27 日的检索结果显示，该媒体发表了关于两家模型企业扩充驻场工程师的报道，公开网页确认了两家公司争取企业客户、向客户现场部署工程人员的方向。此类岗位在企业软件行业已有较长历史，尤其常见于数据平台和复杂分析系统的实施工作。人工智能企业采用这种方式，可以把模型能力同已有数据库、身份认证系统、业务软件和文档管理环境连接起来。

企业项目中的工作往往从流程梳理开始。以制造业订单处理为例，工

程人员需要明确订单的结构、产品编号、授权边界和异常处理方式，才能把模型调用组合成可运行的任务。若客户还要求在本地部署模型，服务器配置、网络隔离和软件版本管理也会成为工程内容。应用规模扩大以后，具有行业经验的实施团队与能够复用的连接器、测试工具及权限组件会形成互相促进的关系。该报道提供了理解模型公司企业业务组织方式的新视角。付费网页未完整开放，本文不采用检索摘要中的人员规模和营收数字。

三、腾讯发布 Hy3D 3.6，专业三维资产生成进入 API 工作流

11 月 27 日,腾讯混元公开新一代三维生成模型 Hy3D 3.6,重点改善复杂曲面、高精度网格和基于物理渲染的材质贴图。公开资料显示,几何和纹理由变分自编码器压缩至共享的稀疏潜在空间，并通过扩散 Transformer 生成，后训练结合监督微调与偏好优化。模型已接入腾讯云 Token Hub 以及 Miora、LAP 和 WorkRally 三款产品，创作引擎提供每天五次免费体验。对工业可视化而言，生成网格可用于设备外观、场景模型与培训资料制作，精确尺寸和装配关系则可继续由参数化设计软件管理。两类三维资产通过标准化交换格式进入同一展示场景，将形成不同精度层次的模型生产体系。

四、英伟达补充开放文档解析模型的训练材料

英伟达在 Hugging Face 发表的 11 月 26 日技术文章介绍 Nemotron 3.5 Parse 的训练及推理工程。该模型从文档页面图像提取文字、阅读顺序、版面类别及坐标信息，并将表格和图表整理为可供程序使用的结构。此次新增公开内容包括合成训练数据、微调方案和用于推理加速的 DFlash2 草稿模型。

文档解析的工程价值在工业企业中尤其明显。设备规格书、检验记录和工艺卡片常混有表格、示意图及多个栏目，普通文本转换容易破坏行列对应关系。具备版面理解能力的模型可以在提取结果中保留定位信息，使检索服务直接回到原始页面，减少技术人员在多个版本之间寻找出处的时间。合成数据可以利用预先生成的正确答案训练模型识别复杂排版，并为企业根据内部模板开展领域适配提供可复用的方法。公开页面未列出精确发布时间，本条作为 48 小时范围的技术背景处理，不进入严格当日博客选题。

五、GigaChat 4 技术资料显示混合注意力架构的发展方向

11 月 27 日的技术媒体汇总关注到俄罗斯 Sber 体系下 GigaChat 4 旗舰模型的技术资料。其 Hugging Face 官方模型卡给出的架构规模为 7080 亿总参数、每个词元约 410 亿激活参数，最大上下文为 1048576 词元。模型配置将线性注意力层与稀疏注意力层结合，并提供即时响应和推理两种调用方式，旗舰权重采用 MIT 许可。此前已有权重发布，当前检索不能确认其首次开放发生在本次 24 小时窗口，故只列入技术背景。

混合注意力设计试图在长文档处理和推理成本之间建立更合适的计算结构。线性递归状态承担部分序列信息处理，稀疏注意力则从历史词元中挑选相关信息。专家混合结构还使每次计算仅激活部分网络参数。对于企业模型部署而言，权重总规模直接关联存储和设备间通信，而激活参数影响计算开销，两类数据需要结合上下文缓存、并发请求和量化精度进行资源规划。模型卡及开放权重也为推理框架团队适配新架构提供了基础。

趋势观察

发电企业协调电源与机房选址、模型公司把工程师派到客户现场，三维生成和文档解析模型则进入专业工具链。几条消息指向同一件事：人工

智能项目的交付边界正在向电力设施、企业业务系统和行业数据延伸；算力规模与模型参数之外，场地、集成和适配能力也影响项目能否落地。

参考资料

1. Reuters 原文｜原稿来源
   https://www.reuters.com/world/asia-pacific/japans-jera-enters-talks-with-anthropic-meta-build-data-centres-nikkei-reports-2026-11-26/
2. Barron's 原文｜原稿来源
   https://www.barrons.com/articles/anthropic-claude-openai-chatgpt-ai-hiring-aa91b640
3. 腾讯混元发布稿（认证账号）｜原稿来源
   https://www.itsdw.cn/news/53218.html
4. IT 之家｜原稿来源
   https://www.ithome.com/1/033/282.htm
5. 腾讯混元 3D ｜原稿来源
   https://3d.hunyuan.tencent.com/
6. NVIDIA 技术文章｜原稿来源
   https://huggingface.co/blog/nvidia/building-on-nemotron-35-parse-open-data-fine-tunin
7. GigaChat 4 模型卡｜原稿来源
   https://huggingface.co/ai-sage/GigaChat4-708B-A41B-GGUF
8. 11 月 27 日技术新闻汇总｜原稿来源
   https://toolsift.ai/llm-news/2026-11-27/

机构介绍与关注

新质生产力工作委员会

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

工业智能算网

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

关注高促会新质生产力工委会公众号

关注工业智能算网平台

https://gyznsw.cn