

# 字节跳动OmniHuman-1 开源项目解读

📅 2025年2月11日 ⌚ 1 分钟阅读

#人像视频生成

#字节跳动

#OmniHuman-1

#开源

#AI

字节跳动开源的OmniHuman-1项目，并对其技术原理、功能特点、应用前景和伦理风险进行了详细解读。

OmniHuman-1是字节跳动于2025年2月5日正式发布的端到端多模态人类视频生成框架。以下是关于它的相关信息：

## 技术原理

**基于扩散变换器架构：**以扩散变换器（Diffusion Transformer, DiT）为基础，结合3D因果变分自编码器（3D Causal VAE）和流匹配（Flow Matching）技术，在潜在空间对视频进行去噪生成。

**多模态信号融合：**并行处理文本、图像、音频和姿态数据，将运动信息压缩为紧凑格式并逐步精炼为视频输出。

**动态比例控制：**训练中对较弱条件赋予更高比例，避免模型过度依赖强条件，提升泛化能力。

## 技术亮点

**多模态输入与“全条件”训练：**整合文本、音频、图像和姿态信号作为输入条件，采用创新的“全条件”训练方法，使模型能够从更大规模、更多样化的数据中学习。

**自适应输入处理系统：**支持任意纵横比的图像输入，包括纵向、半身及全身图像，通过可变形卷积网络实现不同场景下的特征自适应对齐。

**兼容多样化风格：**能处理真实人像、卡通、动物等多种风格的输入，保持风格化运动特征。

## 功能特点

**全身体动画生成：**突破传统AI模型局限，可从面部特写、半身像到全身像进行全方位动态生成，人物有自然唇音同步、流畅

## 目录

## 文章信息

字数

阅读时间

发布时间

更新时间

## 标签

#人像视频生成

#字节跳动

#OmniHuman-1

#开源

#AI

手势和肢体动作，还能处理人与物体交互。

**精准的动作与音频同步：**能确保生成视频中的人物手势、面部表情与输入音频精准同步，如让人物实现演讲、唱歌、乐器演奏等动作与音频的完美匹配。

**适应不同图像风格和质量：**无论是高分辨率 portrait、低质量快照还是 stylized illustration，都能智能适配，生成流畅、逼真的动态视频。

## 应用前景

**虚拟演讲：**可用于生成虚拟人物演讲视频，为线上会议、培训等提供便利。

**教育内容制作：**能够制作乐器演奏演示等教育视频，以更生动的方式辅助教学。

**影视特效预演：**帮助影视制作团队快速生成特效预演视频，提高制作效率和创意展示效果。

**社交媒体与娱乐：**在社交媒体平台上，用户可利用该技术轻松创作个性化的AI视频内容，如生成自己的数字分身进行表演等。

## 伦理风险

**深度伪造风险：**可能被用于制造虚假政治演讲、金融诈骗内容等，引发严重的社会和安全问题。

**身份盗窃与隐私问题：**能通过一张照片和音频生成逼真视频，可能导致个人身份被冒用，侵犯隐私。

**传播虚假信息：**容易被用于制作误导性的新闻、广告等内容，扰乱信息传播秩序，影响公众判断。

## 相关链接

[官网](#)

[论文](#)

分享这篇



## 文章

### 相关文章推荐

#### UI-TARS Desktop -..

字节跳动UI-TARS  
桌面应用项目

#### 计算机使用 代理

计算机使用代理

#### DeepSeek R1 Paper...

A comprehensive  
review of t...