拟公示算法机制机理内容

算法名称	OptimaTalk 深锶数字人同步说话生成算法
算法基本原理	OptimaTalk 深锶数字人同步说话生成算法是一种基于 StyleGAN2 的 GAN 模型，用于生成高保真、时序稳定的说话头部视频。其核心创新包括优化的音频-视觉同步技术、Whisper 音频嵌入、风格融合的跨模态条件生成以及抖动判别器。通过改进数据预处理、模型架构和损失函数，实现了高分辨率、精准唇语同步和时序连贯的视频生成。
算法运行机制	1. 视频预处理：该阶段首先上传视频文件，随后进行视频合规性检查，验证视频是否合规。通过检查后，使用 S3FD 算法对面部区域进行检测和裁剪，得到符合算法需求的视频面部数据； 2. 音频预处理：该阶段首先上传音频文件，接着进行音频合规性检查，确认音频合规。检查通过后，采用 whisper 模型对音频内容进行编码处理，将声音信号转化为可供算法处理的音频特征数据； 3. 生成目标视频：该阶段将前两个阶段处理后的视频面部数据和音频特征数据同时输入 OptimaTalk 算法，通过算法内部的合成处理，最终生成口型与语音同步的输出视频。
算法应用场景	一、应用产品 深锶赛奇 WebAPP 短视频制作工具、深锶 Metaluna 数字人全息舱智能交互硬件 二、应用场景 向用户提供高清晰度、高稳定性、口型真实正确的数字人视频生成能力。

算法目的意图	算法目的意图是仅通过输入一段 1~3 分钟的视频和对应的音频，即可生成逼真同步的数字人视频或直播流，实现在多种应用场景下的数字人智能交互。这种技术可以让用户在直播、短视频制作、智能硬件等多领域基于数字人制作内容。
算法公示情况（选填）	