ChatGPT VisionTalk: A Proposal for an AI-Generated Video Explanation System

Author: Ian Peixoto
Date: November 7, 2025

Overview

This document proposes a new multimodal feature for ChatGPT that would allow the model to automatically generate AI-created educational videos based on its text explanations.
The system would combine natural voice synthesis, visual design, and dynamic scene generation to produce clear, engaging, and accessible video responses.

The feature would be especially valuable for users who struggle with reading comprehension or prefer audiovisual learning formats.

Problem Statement

While ChatGPT currently delivers high-quality text explanations, many users (including those with limited literacy, attention deficits, or learning disabilities) face difficulties understanding long written responses.
Text-only outputs limit accessibility and engagement, particularly in educational or inclusive environments.

Proposed Solution

Introduce a new mode called “Video Explanation Mode” that transforms ChatGPT’s textual answers into short, AI-generated videos.
Each video would include:

- ➢ Natural AI-generated narration (using OpenAI’s text-to-speech models);
- ➢ Automatically generated visuals, diagrams, or animations illustrating the concepts;
- ➢ Optional subtitles and translations;
- ➢ Style customization (educational, professional, storytelling, etc.).

This multimodal approach would make ChatGPT explanations more interactive, inclusive, and appealing to a broader audience.

Technical Feasibility

This concept leverages existing OpenAI technologies:

- ➢ Text comprehension and summarization: GPT-5 core;
- ➢ Voice generation: current TTS/Whisper models;
- ➢ Visual generation: DALL·E and Sora (text-to-video model).

By integrating these tools under a unified interface, ChatGPT could automatically convert explanations into short educational videos within seconds.

User Control and Community Feedback

The video explanation feature would be optional, not mandatory.
Users could freely choose whether they prefer to read the text, watch the AI-generated video, or experience both.
This ensures that the feature remains inclusive, enhancing accessibility without replacing traditional text-based learning.

To maintain high accuracy and continuous improvement, each video would include a simple feedback tool below the playback area, inviting users to report issues such as:

“If you noticed any error, select the time in the video and briefly describe what happened.”

This collaborative system would allow the community to help refine the generated content, ensuring that AI explanations remain clear, precise, and trustworthy while benefiting all types of learners, visual, auditory, and textual.

Educational and Social Impact

The feature would:

- ➢ Improve accessibility for users with reading difficulties;
- ➢ Expand the use of ChatGPT in classrooms and online learning;
- ➢ Support teachers, students, and self-learners globally;
- ➢ Make AI knowledge dissemination more inclusive and equitable.

This innovation aligns strongly with OpenAI’s mission: to ensure that artificial intelligence benefits all of humanity.

Potential Monetization

The feature could be introduced as part of a premium tier (e.g., GPT-5 Plus or GPT-5 Ultra).
Users would gain the privilege of converting their chats into high-quality videos, a compelling “reward” for subscribers and educators who value audiovisual explanations.

Closing Statement

I am submitting this proposal as an independent idea contributor to inspire future developments at OpenAI.
I believe this feature could significantly expand ChatGPT’s accessibility, usability, and social impact.

Respectfully submitted,
Ian Peixoto