What available High Quality Conversational AI tool can give financial advice?

Kaan Gögcay

May 28, 2025

Abstract

This research presents a comprehensive evaluation of current conversational AI tools that are capable of generating a talking, realistic human. Each tool will be assessed based on specific functionalities, with performance metrics guiding the selection of the most effective solution. The analysis aims to identify the tool that delivers the most optimal performance across a range of criteria, with particular emphasis on the ability to visually and verbally simulate a human interaction.

Contents

1 Introduction

This research is driven by the goal to identify which of the available high-quality conversational AI tools can best serve the needs of a new product under development. By comparing AI video generation tools against each other I am looking for the tool that best suits our project case.

In the financial sector, conversational AI poses a significant potential for innovation by offering human-like advice and explanations. This research aims to identify the most suitable high-quality conversational AI tools for giving financial advise and explanations. By conducting a detailed comparison of AI video generation tools, the study seeks to find the tool that best fits our specific project needs.

2 Research Approach

The study was guided by the sub-question: *"What available high-quality conversational AI tool can give financial advice?"* To answer this, I followed a structured research process involving multiple methods.

First, I explored various AI tools available on the market. Then, I created a prototype that brought together their capabilities into a single video. This prototype was pitched to stakeholders and my company mentor for feedback. Following this, I conducted a more detailed comparison of the tools. After the comparison, I contacted the best tools to give me further insights. Each step of this process is detailed below

2.1 Available AI Tools

The research began with an online search to identify existing conversational AI tools, starting with top 10 lists of AI video generation tools. Synthesia, being the most popular, was easy to find, and I then sought competitors and alternatives, leading me to Tavus. I continued this process of finding alternatives and competitors for each tool. Additionally, I used ChatGPT to compile a list of conversational AI tools capable of generating realistic talking avatars. Each tool was tested by creating short video clips where avatars delivered improvised dummy scripts. The tests focused on language options, background settings, and avatar appearance. These videos were labeled with the convention `Generic_x`, where `x` refers to the specific tool used. The videos can be found in the portfolio evidence.

2.2 Prototyping Output

A prototype video was created by compiling short clips from different tools into a single storyline. This video serves as an example of a potential end-product and showcases the range of visual and vocal styles available. The intention was to intrigue stakeholders by showing a diverse range of what is technically feasible. This prototype output can be found in the portfolio evidence named `First_Demo_Synthesia_Tavus_HeyGen_Elai_Deepbrain.mp4`.

2.3 Prototype Pitch

The compiled prototype video was presented to both stakeholders and the company mentor. Their feedback:

- Preference for avatars from Synthesia and Tavus over those from Elai, HeyGen, and Deepbrain.
- A need for less static visuals and smaller speaker proportions on screen.
- A preference for a more polished and professional video style, avoiding overly casual or cringe-worthy elements.
- Suggestions to incorporate reusable branding elements.
- Useful domain-specific feedback regarding insurance product explanations.

2.4 Comparative Analysis of Conversational AI Tools

Following feedback and initial prototyping, a more formal comparison was conducted across multiple tools. For each tool, three targeted tests were performed: voice naturalness, facial expression accuracy, and the ability to deliver financial advice. These tests are detailed in the following section. Videos generated for each test use the following naming convention: `x_Intonation`, `x_Phonetic`, `x_Financial`, where `x` denotes the tool name. These three specific tests were selected based on key limitations identified during earlier evaluations, which had previously hindered the implementation of certain tools.

2.4.1 Tests Procedures

This section elaborates on how the specific evaluation tests were conducted. Note that all sentences used in the tests were kept short due to free-tier limitations on most tools. If a model fails a test, it is not evaluated further.

2.4.1.1 Voice Naturalness Evaluation

To assess the naturalness of the voice, the AI will be tasked with delivering the following script: *"Hey there, awesome viewer! I'm stoked you made it here! Get ready for some exciting content on AI. Whether you're new or experienced, this video has something for everyone. Let's dive in!"* This script includes three distinct sentences, each capable of being interpreted with varying emotional tones. The goal of this evaluation is to observe whether the model incorporates appropriate intonation while speaking. The test is passed if I think the model is able to incorporate these kinds of intonation.

2.4.1.2 Facial Expression Evaluation

To assess the AI's ability to synchronize facial expressions with speech, the following script will be used: *"bug dad fat gun hop jam kit well man net pin run sit tip vine why zed azure chip sham thongs leather ring you cat bay end be it sky swan open wolf lug who boy now about chair arm bird paw ear cure"* This script is designed to cover a wide range of phonetic sounds, encompassing all the International Phonetic Alphabet (IPA) phonemes [1]. The key focus of this evaluation is to ensure that the AI's lip and mouth movements align correctly with the sounds and words being produced. The test is passed if I think that the models lips and voice match well.

2.4.1.3 Financial Advice Delivery Test

To assess the model's ability to deliver financial advice, the following script will be presented: *"You should buy a credit now and don't forget to invest all your money in stocks!"* This script clearly activates the viewer to take a financial action. If the model cannot generate a video that delivers this script, it will not be considered suitable for inclusion in the project.

2.4.2 Test Results

All test videos can be found in my portfolio and are saved as the following: `x_Intonation`, `x_Phonetic`, `x_Financial` respectively. If a test is passed, it is marked with a check mark (✓). If the test is failed, it is marked with a cross (×). If a test is marked with a dash, a previous test failed, or the test was either not possible to execute due to costs, the tool does not generate a realistic-looking human, a script could not be provided, or a combination of these issues.

Tool	Natural Voice	Facial Expressions	Financial Advice
Tavus	✓	✓	✓
Creatify	✓	✓	✓
Veed	✓	✓	✓
Elai	✓	✓	✓
Synthesia	✓	✓	×
HeyGen	✓	✓	×
AI Studios	×	-	-
Colossyan	×	-	-
OmniHuman	-	-	-
Lumen5	-	-	-
Tome	-	-	-
Murf	-	-	-
HourOne	-	-	-
D-ID	-	-	-

Table 1: Test Results

2.4.3 Comparison

After testing the tools in fundamental quality aspects, the tools will now be compared with each other. Each tool will be researched on many aspects, which can be found in the table.

Most of the topics are self-explanatory. For "Easy to experiment", I tested how easy it is to generate a video and tweak it. For example, on "Veed.io" I would need to create a new account after generating one video making it extremely tedious to experiment with.

Tool	Tavus	Creatify	Veed	Elai
API	✓	✓	×	✓
Resolution	1920x1080	1920x1080	4k	4k
Languages	Universal	29 Including NL, UK, DE, FR	50+ Including NL, UK, DE, FR	75+ Including NL, UK, DE, FR
Data Store in Europe	×(soon)	?	✓	?
GDPR Compliant	✓	✓	✓	✓
Easy to experiment with free plan	✓	✓	×	✓
Price	Starting from $12.000 a year	?	Starting from $100 a month	?

Table 2: Comparison Results

2.5 Tavus Expert Interview

To gain deeper insights, an expert interview was conducted with Jared Vishno, Founding Account Executive at Tavus. This interview provided unique information not found in public sources, including future product capabilities, pricing models, and strategic use cases. This greatly enriched the evaluation of Tavus as a leading candidate for implementation. The transcript of this interview can be found in the evidence bin named `Expert_Interview_Tavus.pdf`.

2.6 Final Quality Comparison

The comparison narrowed the options down to three tools: Tavus, Creatify, and Elai. To determine what tool fits the best in this project, I created a survey and shared it with 5 colleagues. In this survey you assess the quality of the speakers from Tavus, Creatify and Elai. Each speaker is graded on 5 criteria, which are the following:

- Does the human look natural?
- Are the body movements fitting? (Arms, Torso, Hips, Head)
- Do the lips sync well with the audio?
- Are the facial expressions fitting?
- Does the audio sound natural?

For each criteria the assessors graded the speakers with a score between 1 and 5. Below an overview of the results from best to worst:

Thomas	Does the human look natural?	Are the body movements fitting?	Do the lips sync well with the audio	Are the facial expressions fitting?	Does the audio sound natural?
Assessor 1	5	5	5	5	5
Assessor 2	5	2	5	2	5
Assessor 3	2	2	3	4	5
Assessor 4	5	4	5	5	4
Assessor 5	4	2	3	3	4
Average: 3,96	4,2	3	4,2	3,8	4,6

Figure 1: Thomas Results 3.96 (Creatify)

Laura	Does the human look natural?	Are the body movements fitting?	Do the lips sync well with the audio	Are the facial expressions fitting?	Does the audio sound natural?
Assessor 1	4	4	5	4	5
Assessor 2	2	3	4	3	5
Assessor 3	2	3	4	2	4
Assessor 4	5	4	5	4	5
Assessor 5	4	4	4	3	3
Average: 3,8	3,4	3,6	4,4	3,2	4,4

Figure 2: Laura Results 3.8 (Tavus)

Destiny	Does the human look natural?	Are the body movements fitting?	Do the lips sync well with the audio	Are the facial expressions fitting?	Does the audio sound natural?
Assessor 1	5	4	4	5	4
Assessor 2	5	2	4	3	5
Assessor 3	4	2	4	3	5
Assessor 4	4	4	5	4	3
Assessor 5	2	2	4	2	2
Average: 3,64	4	2,8	4,2	3,4	3,8

Figure 3: Destiny Results 3.64 (Creatify)

Roy	Does the human look natural?	Are the body movements fitting?	Do the lips sync well with the audio	Are the facial expressions fitting?	Does the audio sound natural?
Assessor 1	4	4	5	3	5
Assessor 2	5	3	3	2	5
Assessor 3	3	3	4	3	5
Assessor 4	4	3	5	3	4
Assessor 5	3	2	2	2	3
Average: 3,52	3,8	3	3,8	2,6	4,4

Figure 4: Roy Results 3.52 (Elai)

Charlie	Does the human look natural?	Are the body movements fitting?	Do the lips sync well with the audio	Are the facial expressions fitting?	Does the audio sound natural?
Assessor 1	5	5	5	3	3
Assessor 2	4	2	1	3	4
Assessor 3	1	2	3	1	2
Assessor 4	4	4	5	5	4
Assessor 5	4	4	3	4	2
Average: 3,32	3,6	3,4	3,4	3,2	3

Figure 5: Charlie Results 3.32 (Tavus)

3 Conclusion

In the final quality comparison, Creatify emerges as the top-performing tool with a score of 3.96. Close behind is Tavus with a score of 3.80. Notably, Creatify appears again in third place with a score of 3.64, suggesting that the platform may offer multiple tools capable of delivering high quality, unlike Tavus. Elai ranks fourth with a score of 3.52, while Tavus appears again in fifth place with a score of 3.32.

Although the differences between scores are relatively small, further research involving more speakers per tool would be beneficial. Nevertheless, both Creatify and Tavus demonstrate strong performance, with Elai also being a viable option.

4 Reflection

There were several decisions made during this research that I would approach differently next time:

- The way the models were assessed was not very professional. Ideally, I should have involved multiple people, especially members of the target audience, to evaluate the results and reduce personal bias. Later in the research I did use multiple assessor for the final quality comparison, but not in the first quality assessment.
- During the final quality comparison it would have been better if I included more models per platform. But since this was done late in the project there was no time for it.
- While researching the tools, I avoided spending any money. In hindsight, this was a mistake. Since the goal was to find the best tool for a company, investing a small amount could have improved the quality and depth of the research. A company would likely be willing to cover such costs if it leads to better outcomes.
- When testing phonetic words it would have been better to include dots or commas in between to ensure each word is interpreted individually. In the current setting, some phonetic sounds might have been skipped because all the words are delivered swiftly after one another.

References

[1] Dyslexia Reading Well, *44 Phonemes in English*, Available at: https://www.dyslexia-reading-well.com/44-phonemes-in-english.html [Accessed March 20, 2025].