Ray, below is the step-by-step approach that we can take to complete the project:

PHASE 1) Defining the scope of the Digital Twin:

a. **Keeping it knowledge focused:** Creating a thorough Q&A system that answers like the professor would, based on their published work, lectures, and teaching style
b. **Making it persona focused:** Mimicking the personality, speech style and other intricate details of the professor. Ray, do you think we should head in this direction? Or this part is not that important as of now?

PHASE 2) Data Collection:

a. **Lectures / Talks**: Transcripts + video/audio to capture his teaching style.
b. **Research Papers / Books**: To embed his domain expertise.
c. **Course Material**: Slides, quizzes, rubrics to replicate his teaching pattern.

Ray, we can train our GPT on even more data like any of the course material that we might have of professor Morgan, I am currently searching if I can find more of his research papers and books, do you know if we have any of him course material or recorded lectures that we can use?

PHASE 3) Modelling Approaches:

1. No Code Approaches:
    a. Knowledge-based chatbots: Like the one that is already created using tools like ChatGPT Custom GPTs, Poe bots and Perplexity Pages.
    b. No coding needed — but limited control.
2. Coding Approaches:
    a. Knowledge-based chatbots using RAG (Retrieval-Augmented Generation):
        i. Using a Vector DB to like Pinecone, Weaviate and Chroma to load the professor's lectures/papers.
        ii. Connect it to GPT/Claude/Mistral via an API.
        iii. Requires Python + LangChain/LlamaIndex.

    This seems to be a good approach to take, what RAG essentially does is that when a user asks a question, the system retrieves relevant documents (professor's lectures/papers) from a vector db and feeds them along with the user's question to the LLM and LLM answers using only the given context. This is fully controllable and scalable to millions of documents. This will help us mitigate the "Black Box Issue" that we were discussing about AI tools.

PHASE 4) Voice and Avatar:

a. Voice Cloning: Using tools like ElevenLabs, OpenVoice
b. Realtime Avatar: Using tools like HeyGen Realtime, Synthesia, Unreal Metahuman.

So, this is the phase-wise approach that I am planning to take, and some of the opinions that I have, please let me know any opinions that you have on this.