So if I told you to think about the people who created social media, you'd probably think about a certain type of guy.
But what if that's not the whole story?
We mythologized these Silicon Valley men, but actually as my reporting shows, they constantly don't know what they're talking about and they don't know what they're doing.
And a lot of the, like, all these, like, breakthrough features on social media, by the way, were actually created by users.
The untold history of social media.
This week on Intuit from Vulture and New York Magazine.
At PropG Media, we attempt to stay up on the latest technologies.
And one way we do this is to actually use these technologies ourselves.
In 2023, that means using, wait for it, AI tools.
We've been experimenting with translating our podcasts into other languages, creating short videos for social media, and now we've developed an AI tool of our own that we'd like to share with you.
It's called PropG.AI.
I know, that sounds scary.
It's a chatbot similar to ChatGPT only with a twist.
Instead of chatting with open AI servers, you're chatting with a digital version of me.
The catalyst here is C above, we want to learn about technologies, but also I receive dozens of emails each day from thoughtful people asking for advice.
And as much as I'd like to respond, I can't.
And so we tasked the team with coming up with a generative AI that could sound very similar and provide responses that felt sort of on point.
This is a bit eerie because we took many of the office hours questions that we received, put them into PropG AI, and found that the responses were pretty similar to what I would have said or how I would have responded.
Anyways, with that, and here to explain how we actually made this tool and some of what we learned about the market along the way is PropG Media's editor-in-chief, Jason Stowers.
When ChatGPT came out late last year, one of the first things we did at PropG Media was ask it to imitate Scott.
We spent a lot of our time working with Scott on scripts and articles and other writing, and it would be incredible to have a digital Scott available to us 24 seven.
Plus, we thought it'd be fun.
As Scott talked about last week on markets, open AI used some of his books to train GPT.
So the bot can make an effort to imitate him, but you get a pretty generic vague version.
We thought we could do better, so we built our own.
Building a chatbot requires two primary components.
The artificial intelligence portion that does the heavy lifting is what's called a large language model or LLM.
These are enormous statistical engines that take in a string of text and then predict what the most likely next string of text is going to be.
That's a narrow skill, but as everyone who's used these tools has seen, it turns out to be a very powerful one.

These systems are quite good at predicting the right answer to a question or how that answer might sound in the style of a pirate or written in computer code.
However, because LLMs are just statistical engines, they can be a bit finicky to work with.
That's where the second component comes in.
The chatbot itself is an extra layer of software that sits between the user and the LLM.
It's not really a translator since the LLM knows every language.
The chatbot is more like a diplomat.
It takes the user's questions and instructions, and it gives the LLM more context for how it should respond.
For example, most chatbots insert text before the user's message along the lines of, you are a helpful AI assistant that politely and accurately responds to user messages.
The idea is to increase the statistical likelihood of the ideal response.
So to make a digital Scott, we needed an LLM and we needed a chatbot that could provide the LLM with enough context about how Scott thinks and writes that the LLM could accurately predict how he would respond to any question.
To accomplish this, we turned to a London startup called spirito.ai.
Spirito was founded by two engineers who left Meta just a few weeks before we met them, Dennis and Alice.
I've asked them to join me and explain how we made a digital Scott.
Dennis, one of the first decisions we had to make was which LLM we wanted to use.
There's quite a few options available in the marketplace, right?
Yeah, definitely.
And it feels like there are new ones every week.
Some are trained on specific knowledge domains like Google's Med-Home, which is specific to the medical field.
Some are more general purpose like GPT-4 from OpenAI.
Some are open source like LLM from Meta.
So there's a bunch of different services out there and there's a lot of variety in industry.
Okay.
So we decided to go with OpenAI's GPT.
Why did that work for us?
Yeah.
So there are two main criteria that we're looking at here.
So one is production capability and then the other is basically model performance.
So on production capability, what we're really concerned about is basically like, can we even use this LLM at scale?
A lot of the LLMs that exist out there are primarily for research purposes or academic purposes or haven't yet been released or they don't have the infrastructure to basically support what we're trying to do, which is build a large scale consumer application.
And on model performance, what we're really trying to evaluate is basically how good is the LLM at this specific use case of building digital versions of creators.
And ultimately we felt that OpenAI's products basically were best at addressing both these criteria and there were a couple other sort of bonus features as well like fine tuning.

So can you explain a bit more about what fine tuning is and how it's helpful to us?
So basically fine tuning is where you take a general purpose model like GPT-4 and train it to better perform at specific tasks.
You kind of show the LLM how to respond by giving it a bunch of data and then later it will use that data to basically like help itself improve on those sets of tasks.
So in our case what we did, we took GPT-3.5, we gave it a bunch of questions and then we gave it a bunch of responses in terms of how we would want the ideal stock bot to respond.
And in the end we have signed two models that perform better than base GPT-3.5.
Now that's the LLM piece of the equation, but then we needed a chat bot that could provide the LLM with the context it would need to capture Scott.
And there we had a great advantage because Scott has been writing prolifically for years and he's recorded hundreds of hours of podcasts.
So what we needed the chat bot to do was to provide the LLM with just the portions of all that writing that would help it respond to each user question.
Alice, can you explain how we went about that?
One can only process a certain amount of tokens at a time and spot prolific writing is definitely more than the limit there.
So we use a strategy of chunking, embeddings, and similarity search to find the relevant text when someone asks a question.
So let's go over each of these.
Embedding is basically dividing the checks into smaller pieces, which we can then embed and store in a database.
Embeddings are important because in the next step we use the embeddings to run a similarity search to find the chunks that are similar to a question, let's say, that is put into the chat bot.
This helps us find the right slivers of information when someone asks the chat bot a specific topic.
And then how does the chat bot coach the LLM to use that material and sound like Scott?
And we want our chat bot to sound like Scott.
We can use a system prompt, which essentially guides the LLM into how to approach answering a question.
And we want to balance style, such as tone, key phrases, maybe Scottisms, instructions essentially act like a chat bot embodying Scott.
And we're also going to feed in some extra context and how to handle that extra context that's passed in.
So we need to be careful about all of these because adding too much information can cause the LLM to forget instructions while adding too little information can cause it to perform suboptimally.
So it's really a part science, but also part art.
Thanks Alice.
Thanks Dennis.
We've made our chat bot available at profg.ai where you can check it out.
It's an experiment and still in the early stages, it handles some questions better than others and it will get better as it answers more questions.

So some fine print, one, I'm sure some of this will be wrong and then again I'm wrong quite a bit, but I'm sure some of this will not hit the mark and we're open to feedback for how we make it better.
And two, and most importantly, this cannot replace human relationships.
And our hope is that this not only provides insight and guidance to people who I otherwise couldn't get back to, but that you use this information as a catalyst for reaching out to potential friends, potential mentors to increase your dialogue, your intimacy and your contact with other people.
Every digital analog of your life is a shittier version of your life.
The digital facsimiles of relationships are just that, they're facsimiles.
And mentors, discuss this with friends.