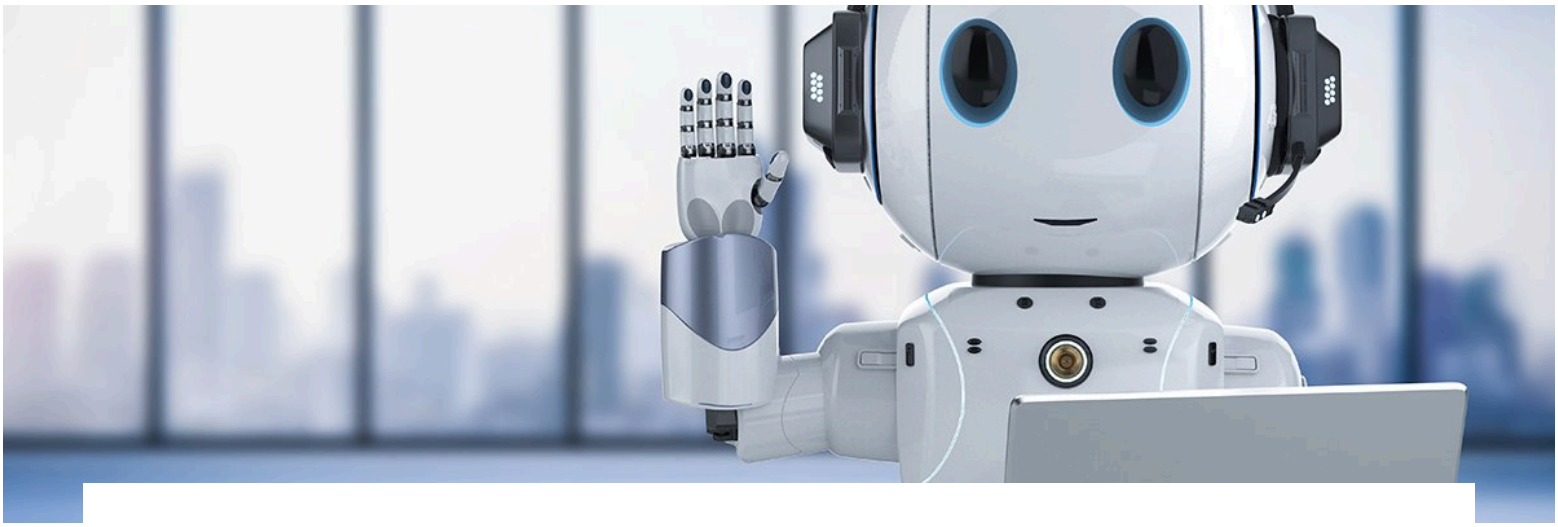
GETTY IMAGES/ISTOCKPHOTO



FEATURE

Building an internal AI call simulator: Lessons for CIOs

LegalZoom built an in-house AI call simulator that improves agent training and automates grading, showing how GenAI can offer real operational gains.

By **Tim Murphy,** Site editor

Published: **09 Oct 2025**

When legal services company LegalZoom set out to improve how agents handle customer calls, they didn't start with a vendor tool: they started with an experiment.

Using ChatGPT's Advanced Voice capabilities, the training team prototyped an AI-powered role-play tool that mimicked real customer interactions --and immediately saw agents perform better.

That early success prompted a deeper collaboration between LegalZoom's operations and engineering teams to build a production-ready version in-house. Rather than purchasing a vendor tool specifically for agent training, they developed their own training and grading system using OpenAI's Realtime APIs for the language model, alongside FastAPI and Kubernetes.

The simulator acts as a virtual customer and grades each training session. During evaluation, the model compares agent responses against knowledge base content to identify inaccuracies or missed opportunities. This offers data-driven feedback that would normally require hours of manual review. For CIOs and technology leaders, the project presents a strategic blueprint: how to transform grassroots generative AI (GenAI) experimentation into a scalable internal application that enhances outcomes.

In the following interview, Noah VanValkenburg, LegalZoom's head of central operations, and Emilio Esposito, director of data and platform engineering, shared how they partnered across teams to design and deploy the tool.

Editor's note: *This transcript was edited for length and clarity.*

What pain points in training led you to pursue the AI call simulator?

Noah VanValkenburg: In the past, we had an expert trainer sit down with agents, role-play with them and score their performance -- which would take multiple days. When ChatGPT became a thing, we started thinking about the different ways to use GenAI.

Within our legal team's guidelines, we started using ChatGPT's Advanced Voice feature to say, "ChatGPT, imagine that you're a LegalZoom customer, and role-play with me. I'm a sales agent." We tried that in a couple of classes, and the feedback we received was unbelievable.

So, I asked one of our legal counsels if we could use some private information. They said we would have to run it on our own servers and suggested that I talk to Emilio in data engineering.

Can you walk us through the deployment cycle?

Emilio Esposito: As this AI revolution has been unfolding, one of the things I saw coming was the concept of internal applications having a much larger presence in the enterprise. Sure, you can get a vendor tool for lots of use cases, but it ends up being tons of different vendors. [illegible]s onboarding costs -- time, money, resources -- and you lack customization.

We realized that there were many advantages to bringing the call simulator in-house. I had already done some of the legwork with our DevOps and data platform partners to get a baseline framework going. We have a monolithic repository -- a single repository that holds all our code -- where we can spin up a new AI service in one day, complete with a back-end API on FastAPI and a front end.

We already had the Kubernetes infrastructure, the FastAPI framework and clearance to use OpenAI's Realtime APIs. The bulk of the tech work for this app was done in one day, though it builds on prior platform investments that made integration fast and efficient.

Can you walk me through the tech stack?

Esposito: The underlying large language model was from OpenAI. It offers a human-like feel of conversation out of the box. We had to do some tuning to make sure it didn't pick up call center background noise, but we were able to create a framework that lets call center managers spin up their own training scenario.

For deployment, the language is all within Python. FastAPI is one of the most well-respected API frameworks in Python. On the front end, there's an open-source package called Chainlit that we use. We use PostgreSQL on the back end for data persistence. For security, we use Azure OAuth, which is our internal single sign-on system. Then we hooked it up with Google Sheets and Google Forms because our end users were already familiar with those applications.

Did you train the tool on any internal data?

Esposito: Noah's team started with some initial prompts, and I did some prompt engineering work on them and on the grading. Since this scenario is for customer role-playing, it's not particularly proprietary.

Where it does get a little bit more into internal data is on the grading piece. The models can hit our internal knowledge bases. We have knowledge base articles for our care and fulfillment specialists. Those are synced daily to a vector database that we host on AWS -- Bedrock's open server vector databases. When it enters grading mode, the system queries the knowledge base to identify any statements that contradict the information stored in our internal article center data.

role did operations play versus engineering in the creation of the simulator?

VanValkenburg: In operations, we initially used our legal team's framework. When we felt that those guardrails weren't serving us, we started to loop in Emilio's engineering team. Our training manager wrote a great two-page brief outlining what we're doing, our challenges, our goals and our results. Emilio was able to take that brief and then work the engineering magic to make it much better than that initial pilot.

Esposito: I recently created a document template that functions like an internal request for proposal. The template outlines key elements, such as the business value you're trying to drive and key success metrics. This helps us clearly define the inputs and outputs, both in the current state and the envisioned AI-driven state.

How does this project align with your company's overall business strategy?

VanValkenburg: There's a huge nexus with this program and overall company strategy. Our CEO talks about how to bolster human expertise. He talks about how to deliver our customers AI tools when they want them and human expertise when they need it. The simulator lets us get that expertise in front of customers faster and better.

What business benefits did you see?

VanValkenburg: The first class to use it reached 94% of a tenured agent's performance within their first week of training. By the end of that month, they were almost at 300% of what we had initially scoped for them as their monthly goal. We don't typically see changes like that in training and enablement. Ordinarily, we might see a positive change of 10%, and we'd be happy about that.

What were the hardest technical or organizational challenges throughout this project?

Esposito: Before this project started, we had to convince our engineering and DevOps partners that we needed a framework for quickly launching things internally.

On a technical level, AI results can be non-deterministic, which makes setting up testing infrastructure and automated testing more challenging and interesting. Even when writing a unit test or an integration test, you must approach it differently. While the industry is still figuring out the best patterns for this, we focused on finding strategies that work -- like collecting data to monitor business quality metrics over time. This was crucial, especially for meeting the legal approval bar.

What advice do you have for other IT leaders looking to deploy GenAI internally?

Esposito: GenAI's best use cases are cases where R&D is expensive and validation is cheap. There's a good parallel here with AI system coding, which is great because it's very easy to check the output, validate and test results--just like any other pull request. It drastically accelerates launching applications. Similarly, in legal services, AI can handle the bulk of R&D work, and an expert can do the last-mile assessment to give it a stamp of approval, ensuring confidence in the output.

VanValkenburg: As an operations guy, I encourage more technical folks to include the operations team and build trust between both groups. If an operations team tries OpenAI or Gemini, there's a good chance that one of the organization's engineers knows how to make that a real business application.

Tim Murphy is site editor for Informa TechTarget's IT Strategy group.