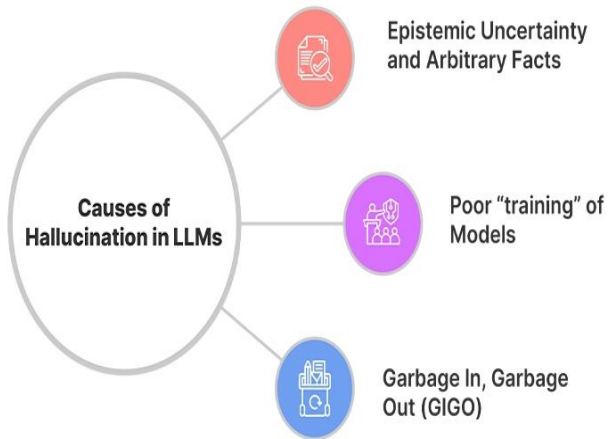




Quantix



Why we're here today

**LLMs are a black box,
and hallucinations
always happen in
production**

THE ISSUE

A

LLMs are a black box

Companies aren't ready for the randomness generated via LLMS

B

Testing is hard

Testing via edge cases, and evaluations are often overlooked when moving fast

C

LLM integration is limited

Testing softwares usually have bad first party integrations with your LLM model

What is the best method for testing LLMs?

Performance Evaluation

Assesses the efficiency and effectiveness of the LLM in real-world scenarios.

Ethical Considerations

Evaluates the LLM's adherence to ethical standards and bias mitigation.

Robustness Testing

Ensures the LLM can handle diverse and unexpected inputs.

User Feedback Integration

Incorporates user insights to refine and improve the LLM.

We simulate test cases for LLMs to make them more predictable and safe in production

THE SOLUTION

PRODUCT BENEFITS

1.

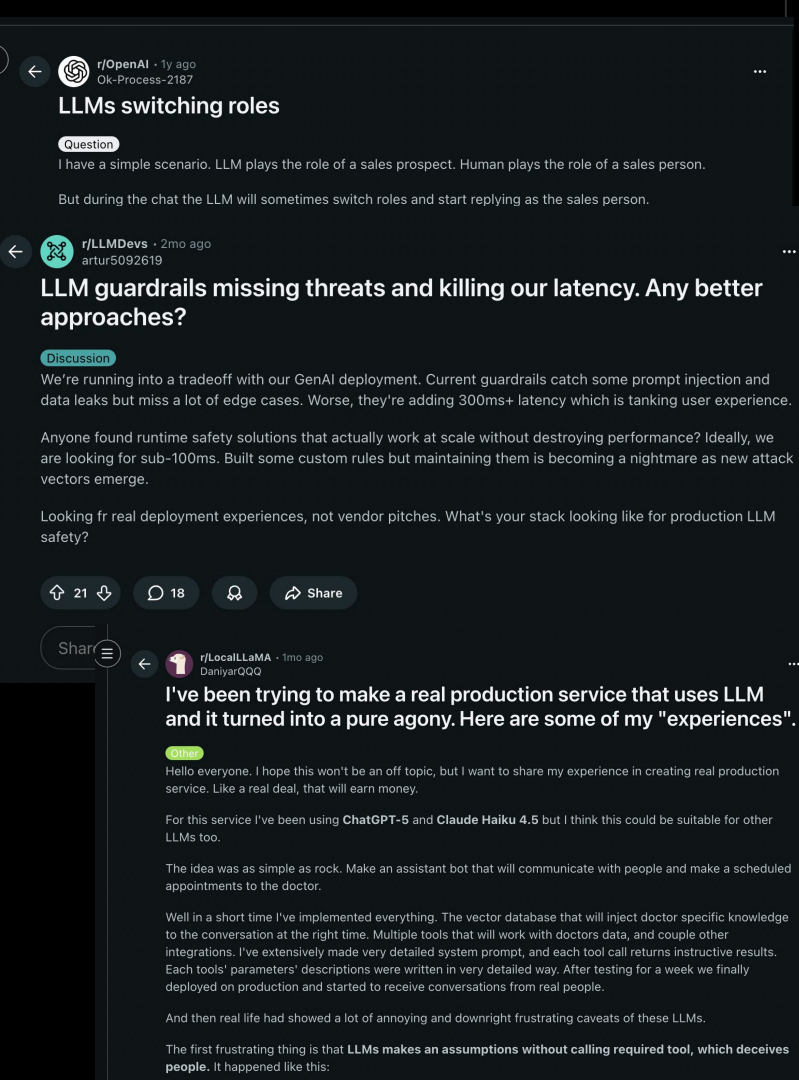
Test out your LLMs on guard rails and performance.

2.

Generate synthetic CSV data for training LLMs

3.

Connect your LLMs via different APIs. I.e. Email, Phone, or Direct URL



Test out your LLMs on guard rails and performance.

Manual testing often misses the tone of actual conversation that happens in production

General features

- Simulate real customer interaction using custom LLMs
- Test LLM responses on guard rails and real interaction just by a click
- Feel confident about your LLM's performance with mathematically-grounded feedback

Generate synthetic CSV data for training LLMs

The biggest challenge in training large language models (LLMs) isn't their architecture—it's finding high-quality, diverse, and unbiased data in a vast and noisy digital landscape. This is true whether you're building an LLM from scratch or fine-tuning a pre-trained one, as you'll need to use high-quality data compiled from multiple sources.

This article overviews LLM training, the need for public web data, and the major public data sources for highly-performant LLMs.

<https://oxylabs.io/blog/llm-training-data>

Tailored CSV data is usually hard to find without proper data cleaning.

General features

- Give shape of the data and a prompt to get accurate synthetic data for training



Connect your LLMs via different APIs. I.e. Email, Phone, or Direct URL

A simple UI to directly connect to your Email agent, phone agent or API endpoint instead of having to use multiple testing softwares to do the same task

HOW IT WORKS

The screenshot shows a 'Create New Test Case' dialog box overlaid on a blurred background of a software interface. The dialog box has a title bar with a close button (X). Below the title, there is a brief description: 'Generate test cases with specific distributions and parameters for your voice bot.' The form contains several sections: 'Test Case Name' with a text input field containing 'Enter test case name'; 'Voice Bot Description' with a larger text area containing 'Describe the voice bot you want to test...'; 'Kind of Test Cases' with a text input field containing 'e.g., Functional testing, Performance testing, Edge cases...'; 'Test Phone Number' with a text input field containing '+1 (555) 123-4567'; and 'Email Address' with a text input field containing 'test@example.com'. At the bottom of the dialog, there are two buttons: 'Cancel' and 'Create Test Case'.

Search by name, description, or email... Newest First Appl

Business Email Assistant
11/10/2025 • Updated 11/10/2025

An AI professional assistant that drafts, sends, and summarizes email threads - Drafts professional emails, sends and manages calendar - Routes urgent emails to the appropriate team - Manages standard workflows

Muvli
10/27/2025

You are a professional assistant that drafts, sends, and summarizes email threads - Drafts professional emails, sends and manages calendar - Routes urgent emails to the appropriate team - Manages standard workflows

Create New Test Case X

Generate test cases with specific distributions and parameters for your voice bot.

Test Case Name
Enter test case name

A descriptive name for your test case scenario.

Voice Bot Description
Describe the voice bot you want to test...

Provide details about the voice bot's purpose, functionality, and expected behavior.

Kind of Test Cases
e.g., Functional testing, Performance testing, Edge cases...

Describe the type of test cases you're looking for.

Test Phone Number
+1 (555) 123-4567

Phone number to use for testing the voice bot interactions.

Email Address
test@example.com

Email address for test notifications and results.

Cancel Create Test Case

1

Create a test suite

Prompt your agent role, test cases to test on, and type of test cases

HOW IT WORKS

2

Generate multiple test cases

Watch us generate real test cases simulating like real customers using custom LLMs

[← Back](#)

Generate Sub-Tests+ Add Sub-Test

Business Email Assistant

Test Case DetailsEdit

Description

An AI agent that helps executives manage business email communications: - Reads and summarizes email threads - Drafts professional responses - Handles contract termination requests - Schedules meetings and manages calendar - Routes urgent matters to appropriate team members - Maintains professional corporate tone - Understands business context (contracts, deadlines, legal terms)

Kind of Test Cases

Happy path scenarios, successful order completions, standard workflows

Test Phone Number

7819759065

Email Address

coco2066@nyu.edu

Created

11/10/2025

HOW IT WORKS

- The contract renewal term was agreed to be for 12 months

LLM Judge Summary

Auto verdict: Succeeded · Task confidence: 95.0% · Safety: 100.0% · Faithfulness: 100.0%

The agent effectively summarized the email thread, highlighting key points, decisions made, and outstanding action items with clear owners and deadlines. It then offered to draft a professional reply and delivered a well-structured, courteous email summarizing the contract renewal details and next steps, including deadlines. The tone was professional, the content was aligned with the user's request, and there were no safety issues or hallucinations.

Human Label

Override or confirm the automatic judgment for this run.

Mark CorrectMark Incorrect

Notes (optional)

Add any comments about why this run was correct or incorrect...

Run History

Refresh

No previous runs recorded yet for this sub-test.

3

Analyze your agent

After simulating the test case on your agent get valuable analysis

Summarize the competitive landscape

Cekura

Feature

- Voice agent testing

Strengths

- Goes beyond just conversation, captures tone, intent and other variables

Weaknesses

- Only expands to voice agents

Braintrust

- Testing, evals, and iterations for LLMs

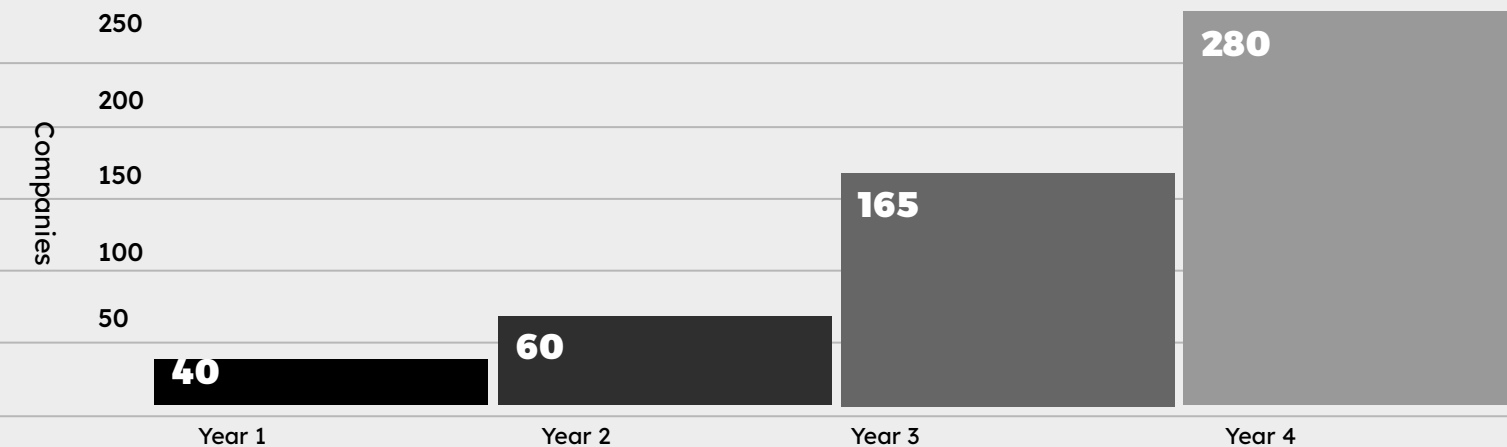
- Amazing insights to test out agents

- Only expands to api layer and becomes complicated when your initial workflow doesn't start with braintrust



FINANCIALS

Hybrid SaaS & Usage Based



MEET THE TEAM



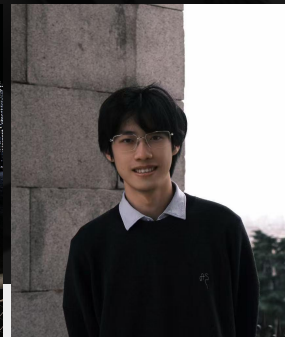
Chuma Oyigbo

CEO



Satvik Seeram

CTO



Weilin Cheng

CMO



Feifan Yang

CPO

**THANK
YOU!**