

DoorDash

AI driver support cut made-up answers by 90%

DoorDash built an AI support assistant for its couriers that answers from past cases and help articles, then checks every reply before a courier sees it.

Food delivery | Driver support | Retrieval-augmented AI | Two-layer guardrail

90%
fewer made-up answers (hallucinations)

99%
fewer potentially severe compliance issues

1,000s
of Dashers helped by the system every day

AT A GLANCE

COMPANY	WHO IT SERVES	AI MODELS NAMED	REPORTED
DoorDash	Dashers, the independent contractors who deliver orders	GPT-4 and Claude-3	DoorDash engineering blog, Sept. 17, 2024

01

The challenge

When Dashers hit a problem during a delivery, they reach out to DoorDash support. The existing automated system ran on pre-built flows, so only a small subset of issues could be resolved quickly.

Dashers could be connected to human agents, but self-help through the knowledge base had three problems the team named: it was difficult to find the relevant article, it took time to find useful information inside an article, and the articles were all in English while many Dashers prefer a different language.

Before building, the team manually reviewed thousands of chat transcripts to define what a good answer looks like.

3

problems with the help articles: hard to find, slow to read, English only

1,000s

of chat transcripts reviewed by hand to define quality

5

quality areas the team set for judging every answer

BEFORE

- × Pre-built flows resolved only a small subset of issues quickly
- × Help articles hard to find and slow to read
- × Articles available only in English
- × No automatic check on what an AI answer says

AFTER

- ✓ Answers built from similar resolved cases and help articles
- ✓ Every reply checked for accuracy and compliance
- ✓ Failed replies go to human agents instead
- ✓ Thousands of Dashers assisted each day

02

The solution

DoorDash built a retrieval-augmented (RAG) chatbot. It condenses the Dasher's conversation to the core problem, pulls the most similar resolved cases, and writes a tailored answer.

Every reply then passes through an LLM Guardrail, an online monitor that evaluates each output. A response must pass all guardrail tests to be shown to the Dasher; when it does not, human agents can take over.

Separately, an offline LLM Judge scores chatbot quality in five areas, calibrated against a human team that reviews random samples of transcripts.

HOW IT WORKS



STEP 01

Summarize the conversation

The system condenses the whole conversation to pinpoint the core problem.



STEP 02

Find similar resolved cases

It searches historical data for the top similar cases already resolved, tied to knowledge base articles.



STEP 03

Write a tailored answer

A large language model drafts the reply from what was retrieved.



STEP 04

Guardrail check 1: does it match the source?

A low-cost similarity check compares the reply with the knowledge base article segments.



STEP 05

Guardrail check 2: is it grounded and compliant?

An AI evaluator checks groundedness, coherence and compliance. Fail any test and the reply is not shown.



STEP 06

Judge and improve

An offline LLM Judge scores quality; prompt changes must pass a regression test suite before they ship.

BUILT WITH

Retrieval-augmented generation (RAG) · GPT-4 · Claude-3 · Vector store · LLM Guardrail · LLM Judge · Promptfoo regression tests

03

The results

The guardrail is what moved the numbers. DoorDash reports it cut overall hallucinations by 90% and potentially severe compliance issues by 99%, while the system assists thousands of Dashers on its own each day.

90%

reduction in overall hallucinations (made-up answers)

99%

reduction in potentially severe compliance issues

1,000s

of Dashers assisted autonomously each day

5

quality areas scored by the LLM Judge: retrieval, accuracy, language, coherence, relevance

REPORTED REDUCTIONS AFTER ADDING THE GUARDRAIL



Overall hallucinations



Potentially severe compliance issues

04

What made it work, and what to watch

> "Although these responses sound natural and legitimate, users may not realize that they are inaccurate."
>
> DoorDash engineering team, on why every AI answer is checked

1	**Check cheaply first, then deeply**	An early, more sophisticated guardrail model slowed responses and used so many tokens it was prohibitively expensive. A cheap similarity check now runs first.
2	**Know when to hand off**	Guardrail latency is a real drawback. Defaulting to human agents for the few problem replies keeps the experience good, and complex cases still need live agents.
3	**Fix the knowledge base too**	Answers are only as good as the articles behind them, so the team stresses complete, accurately phrased articles.
4	**Write prompts a model can follow**	Break complex prompts into smaller parts, avoid negative language, and use step-by-step reasoning.
5	**Test every change**	A suite of predefined tests runs on any prompt change and blocks prompts that fail.

IS THIS A FIT FOR YOUR BUSINESS?

- ✓ Do customers or field staff ask your team the same questions again and again?
- ✓ Do you already have help articles, policies or past tickets the answers should come from?
- ✓ Would a wrong or off-policy answer cause a compliance or customer problem?
- ✓ Do you support people in more than one language?

05

How Agent Victoria could do this for you

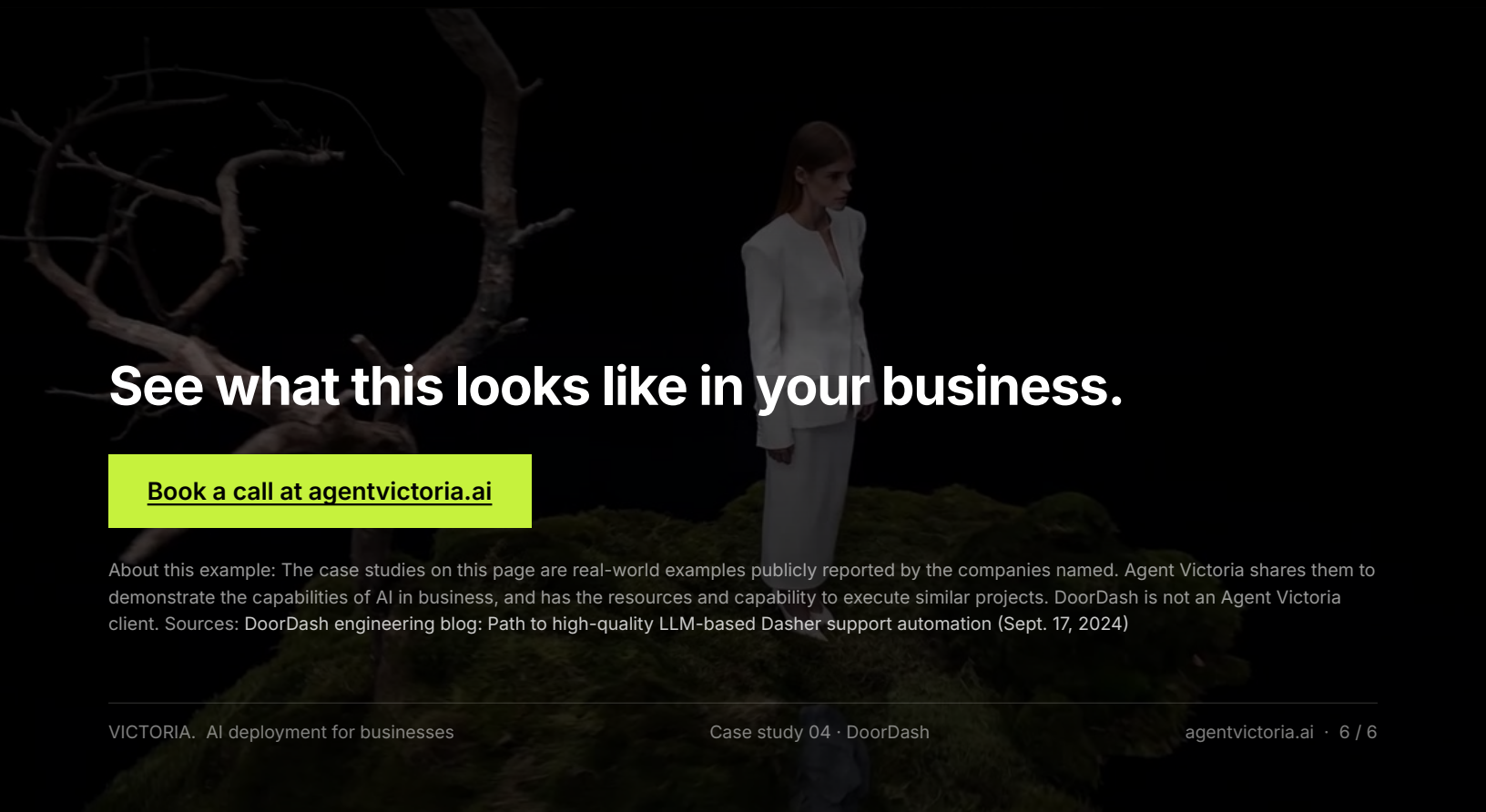
We could build a support assistant that answers from your own help articles and past cases, with checks that stop wrong or off-policy answers. Your customers or field staff get fast help, and you stay in control of what it says.

01
Connect your sources
Help articles, policies and past tickets you already have.

02
Answer from them
Replies built only from your approved material.

03
Check every reply
Automatic checks for accuracy and policy before anyone sees it.

04
Hand off to a person
A clear route to your team when a case needs one.



See what this looks like in your business.

Book a call at agentvictoria.ai

About this example: The case studies on this page are real-world examples publicly reported by the companies named. Agent Victoria shares them to demonstrate the capabilities of AI in business, and has the resources and capability to execute similar projects. DoorDash is not an Agent Victoria client. Sources: DoorDash engineering blog: Path to high-quality LLM-based Dasher support automation (Sept. 17, 2024)