

Empire Shop - Piotr

AI Customer Support Hotline (Polish)

VAPI - n8n - Apilo - Shoper - Pinecone RAG - Twilio SMS

Client	Empire Shop (pet food retailer)
Agent Name	Piotr
Language	Polish (Europe/Warsaw timezone)
Voice Platform	VAPI
Orchestration	n8n
Order System	Apilo API
E-commerce	Shoper (product catalog)
Knowledge Base	Pinecone Vector Store (RAG)

Project Overview

Empire Shop is a Polish e-commerce retailer specializing in premium grain-free pet food, sold under the Empire Prime and Empire Natural Super Petfood (NSP) lines. The company operates a customer support hotline that handles order status enquiries, detailed product questions (composition, ingredients, allergens, dosage), and product recommendations across roughly 200 SKUs. The objective of this project was to build a 24/7 Polish-speaking AI voice assistant - Piotr - capable of answering with absolute factual accuracy, retrieving live order data, and escalating intelligently to human consultants only when genuinely necessary and only during their working hours.

Problem Statement

Empire Shop's customer support hotline faced several distinct challenges:

- Human consultants were only available Monday-Friday, 09:00-16:00, leaving evenings, nights, weekends, and holidays uncovered - a large share of total call volume.
- Order status enquiries required manual lookup in the Apilo order management system, plus a separate shipment/tracking lookup - a two-step process for every call.
- Product questions touched on highly sensitive details - allergen content (chicken vs. poultry), exact micronutrient composition, shelf life - where a wrong or invented answer could mislead a customer with a sensitive or allergic pet.
- With ~200 SKUs, agents needed a reliable way to retrieve accurate product information without relying on memory, and to recommend the correct product for a given pet profile.
- No consistent rule existed for when calls should be transferred to a human versus handled by message - risking either unnecessary live transfers outside working hours or unresolved customer issues.

Solution

We built Piotr - a Polish-language AI voice assistant deployed on VAPI - backed by a comprehensive n8n automation layer. Piotr handles order status lookups via the Apilo API, answers product questions using a combination of hardcoded brand facts and a Pinecone-based retrieval-augmented generation (RAG) knowledge base, guides callers through a structured product recommendation flow, and escalates to human consultants strictly within their working hours - falling back to a message-capture flow at all other times.

The Voice Agent - Piotr

Piotr speaks exclusively in Polish, with a warm, calm, patient, and professional tone. He never uses English unit names or spells out units letter by letter - every number, weight, price, and unit is spoken as a fully formed Polish word in the correct grammatical case. Piotr operates under a strict anti-hallucination guardrail: every factual claim must come from either the hardcoded facts in his system prompt or a verified retrieval from the knowledge base - never from general memory.

Piotr's responsibilities are organized into four core tasks:

Task	Description
Task 1	Order Status - collects the order ID character-by-character, confirms it via readback, then retrieves live status and shipment details from Apilo.
Task 2	Product Information - answers suitability, dosage, ingredient, and composition questions using hardcoded facts first, then the RAG knowledge base.
Task 3	Product Recommendation - runs a profile-gate questionnaire (species, breed, life stage, weight, activity level) before querying the product catalog for a tailored recommendation.
Task 4	Pet Refuses to Eat - offers practical feeding tips, suggests an equivalent Empire product if relevant, and escalates if the issue persists beyond two days.

Anti-Hallucination & Brand Guardrails

Given the sensitivity of pet allergy and dietary information, Piotr's prompt enforces several hard rules with zero tolerance for deviation:

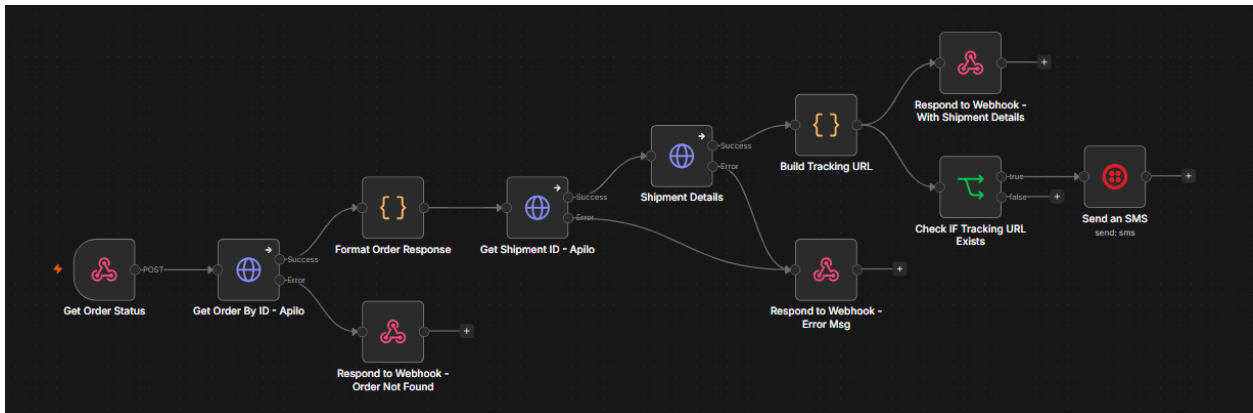
- Chicken vs. poultry distinction - Empire Prime contains no poultry at all (lamb only); Empire NSP contains wild fowl but never chicken. This distinction is hardcoded and always stated correctly, never confused.
- Numeric fidelity - every price, weight, and composition value is copied character-for-character from the source and spoken as full Polish words, never rounded, reformatted, or recalled from memory.
- Empire-only recommendations - Piotr never names or recommends a competitor brand, even when directly asked for a comparison.
- Loose / repackaged food warning - any question about buying Empire food loose or repackaged triggers a strong, unhedged warning against it, every time.
- Price verification - before stating any price, Piotr confirms it is present in the retrieved document, belongs to the exact product discussed, and is copied exactly.

n8n Automation - Order Status & Shipment Tracking

The getOrderStatus tool call from VAPI triggers an n8n webhook that performs a two-stage lookup against the Apilo order management system: first the order itself, then - if shipped - the shipment and tracking details.

- The order ID is verified using Apilo's Get Order By ID endpoint. If not found, an error response is returned immediately so Piotr can inform the caller and offer the order-number-location script.
- If the order exists, Format Order Response normalizes the data, then a second Apilo call (Get Shipment ID) checks whether a shipment has been created.
- If a shipment exists, Shipment Details retrieves the carrier and tracking information; Build Tracking URL constructs a customer-facing tracking link.
- If a tracking URL was successfully built, an SMS is sent via Twilio with the tracking link - matching Piotr's spoken promise: "Wyślę Ci dodatkowo SMS z kodem śledzenia."

- Every completed lookup is also appended as a row to a Google Sheet for support team visibility and reporting.

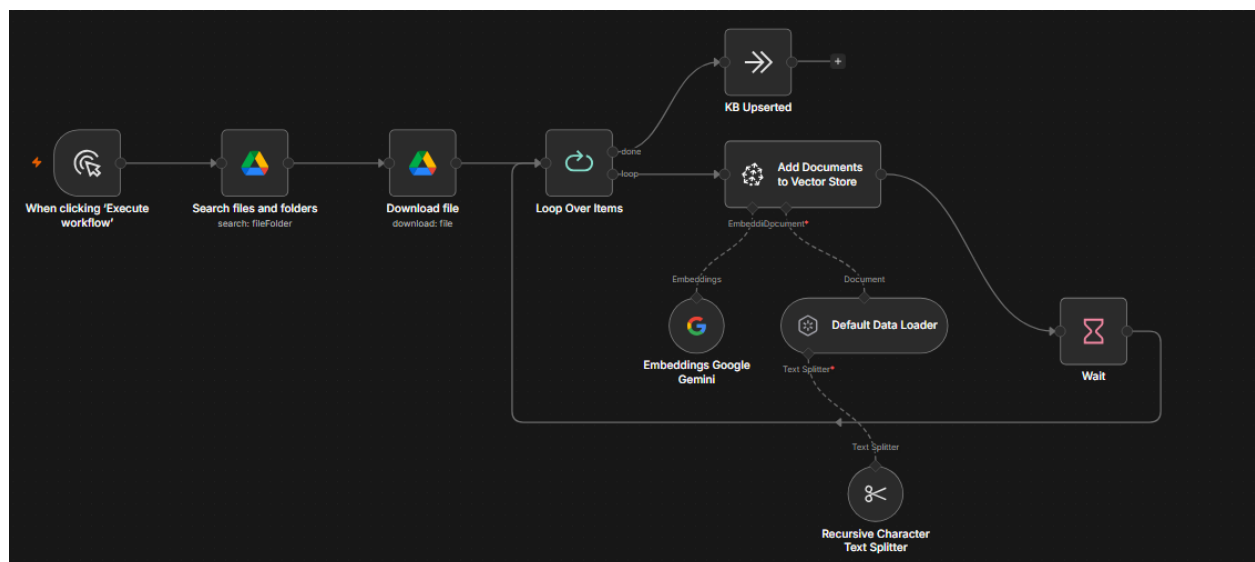


Order Status workflow - Apilo order lookup → shipment lookup → tracking URL construction → SMS dispatch

n8n Automation - RAG Knowledge Base

Product and brand information not covered by Piotr's hardcoded facts is served through a retrieval-augmented generation (RAG) pipeline built on Pinecone, with Google Gemini providing the embedding model. The system has two independent workflows: a manual ingestion pipeline that indexes source documents, and a live retrieval webhook called by Piotr's getRankedDocuments tool during calls.

Ingestion Pipeline (Manual Trigger)

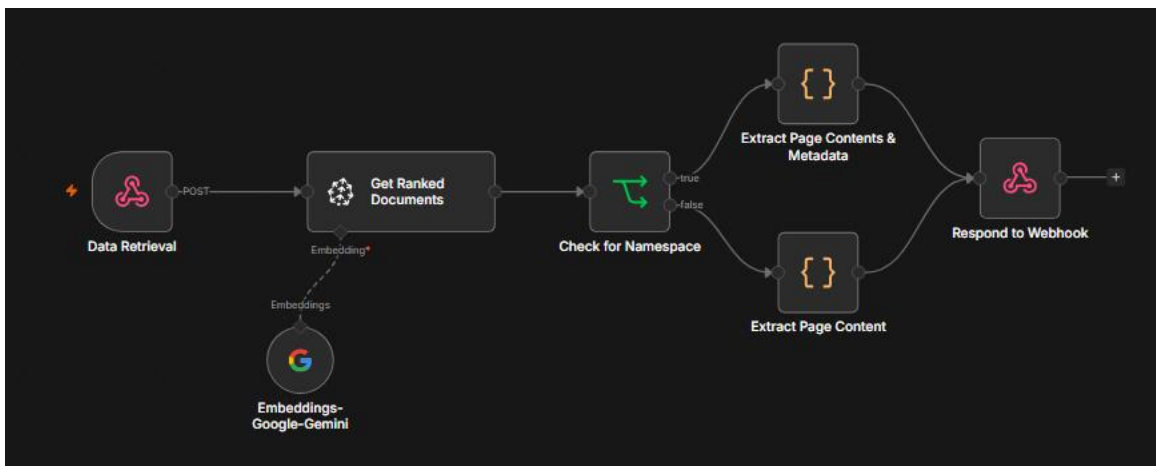


Ingestion Pipeline: Google Drive → chunking → Gemini embeddings → Pinecone upsert

- Source documents are pulled directly from Google Drive (Search Files & Folders → Download File).
- Documents are processed in batches via SplitInBatches (Loop Over Items) to manage volume.
- Each document is split into chunks using a Recursive Character Text Splitter, then embedded using Google Gemini Embeddings.
- The embedded chunks are upserted into the Pinecone vector store via the Add Documents node, tagged by namespace, then the loop waits and continues to the next batch until all documents are processed.

Retrieval Pipeline (getRankedDocuments tool)

- When Piotr calls getRankedDocuments, the Data Retrieval webhook embeds the query using the same Gemini embedding model and searches Pinecone for the most relevant matches.
- A namespace check routes the result correctly between the two required namespaces - empireshop-product-catalog for pet-profile-specific product data, and empireshop-pl for general brand and FAQ content.
- Depending on the namespace, either Extract Page Contents & Metadata or Extract Page Content formats the result before it is returned to Piotr via Respond to Webhook.

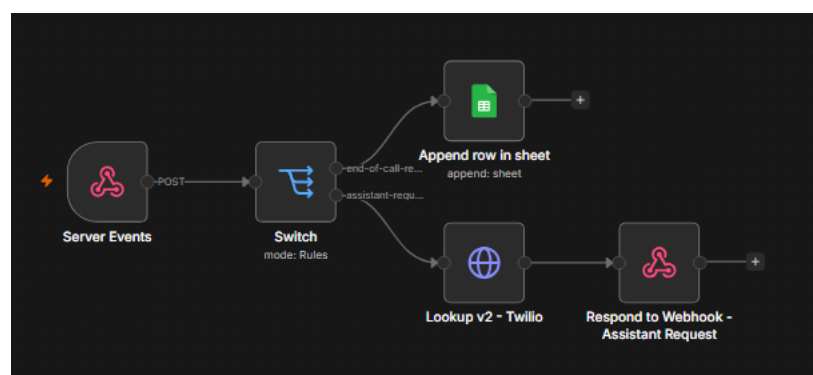
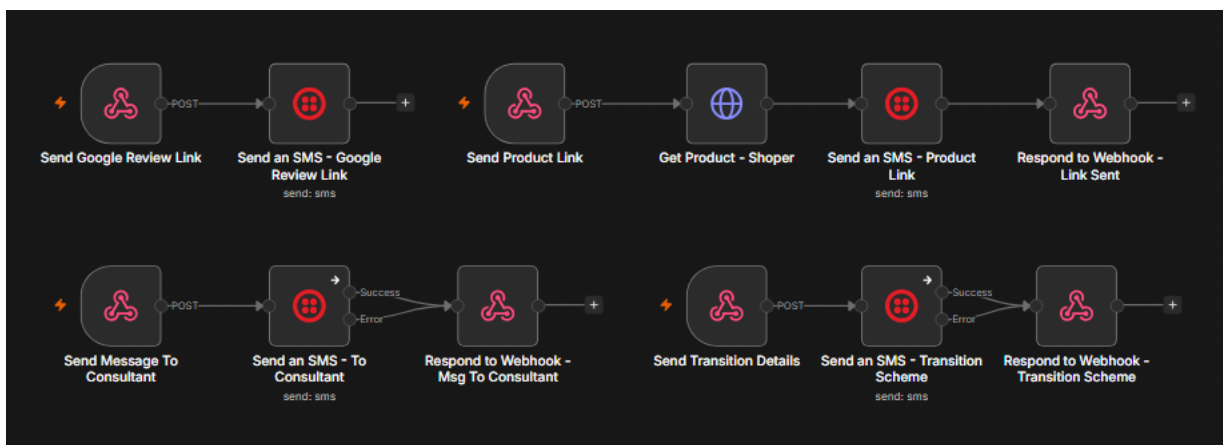


Retrieval: query → namespace check → content extraction → response

n8n Automation - Voice Agent Tool Webhooks

Beyond order status and knowledge retrieval, Piotr calls several smaller, purpose-built n8n webhooks - each corresponding to a single VAPI custom function tool. These are intentionally lightweight: receive a request from the call, perform one action (usually an SMS or a logging step), and respond immediately.

Tool / Webhook	Trigger	Action
<code>sendProductLink</code>	Caller agrees to receive a product link	Fetches product data from Shoper, sends an SMS with the product link
<code>sendReviewLink</code>	Caller agrees to leave a Google review	Sends an SMS with the Google review link, then ends the call
Server Events	VAPI call lifecycle events	Switch routes to call logging (Google Sheets) or caller lookup (Twilio Lookup v2) feeding context back to the assistant
<code>sendTransitionDetails</code>	Caller confirms switching to Empire food	Sends an SMS with the gradual transition feeding plan
<code>sendMsgToConsultant</code>	Caller leaves a message (after failed transfer or outside hours)	Sends an SMS notification to the consultant team with the captured message



Escalation Logic - Working-Hours Gate

A hard rule governs every escalation path: Piotr operates 24/7, but human consultants are only available Monday–Friday, 09:00–16:00 Warsaw time. Before any escalation step, Piotr evaluates the current day and hour against this window.

- Office OPEN (Mon–Fri, 09:00–15:59) - Piotr asks permission to transfer, and only on explicit agreement calls `empire_transfer_call_tool`. If the transfer fails, he falls back to the message-capture flow.
- Office CLOSED (evenings, nights, weekends, all other times) - Piotr never offers a transfer and never calls the transfer tool, even if the caller explicitly asks for a human. The only path is message capture via `sendMsgToConsultant`.
- This rule is absolute and explicitly framed as a hard violation if broken - there is no scenario in which a live transfer is attempted outside the defined working-hours window.

Key Design Decisions

Hardcoded Facts Before RAG

Piotr's prompt strictly orders knowledge sourcing: hardcoded facts are checked first for anything brand-critical (chicken/poultry content, full composition, payment methods, packaging, storage), and the Pinecone knowledge base is only queried for general or FAQ-level content. This guarantees zero-hallucination accuracy on the highest-stakes facts - allergen and ingredient claims - while still allowing flexible retrieval for the long tail of product questions across ~200 SKUs.

Verbatim Order ID Handling

Order IDs are treated as a pure copy task rather than something Piotr recalls or reconstructs. Every character is read back exactly as the caller said it, using Polish letter phonetics (S=es, H=ha, etc.) and full word-form digits, with explicit confirmation before the Apilo lookup is triggered. This prevents the single most common voice-AI failure mode for alphanumeric IDs: silently transposing or guessing a character.

Profile Gate Before Recommendation

For product recommendations, Piotr never queries the catalog until the full pet profile (species, breed, life stage, weight, activity level) is assembled - asking exactly one question per turn, and skipping any slot already inferable from context (e.g. a life-stage word closes the age question). This avoids a long forced questionnaire while still guaranteeing the recommendation query is fully specified.

Strict Working-Hours Transfer Gate

Rather than letting the model decide case-by-case whether to transfer, the working-hours gate is evaluated as a deterministic flag before any escalation logic runs. This removes any ambiguity around dead-air transfers to an empty consultant line outside business hours - a common and frustrating failure mode for after-hours support bots.

Single-Ask Review Request

Piotr asks for a Google review exactly once, using an exact verbatim script, only after the caller signals the conversation is finished - and only if the caller was not frustrated at any point in the call. The review link is never sent without an explicit verbal "yes" first, preventing unwanted SMS spam after a poor experience.

Tech Stack

Component	Tool / Platform
Voice AI Platform	VAPI
Automation / Orchestration	n8n
Order Management System	Apilo API
E-commerce / Product Catalog	Shoper API
Knowledge Base / RAG	Pinecone Vector Store
Embeddings Model	Google Gemini Embeddings
Document Source	Google Drive
SMS Delivery	Twilio
Call Logging	Google Sheets
Agent Language	Polish (Europe/Warsaw timezone)

Results

- 24/7 Polish-language phone support delivered with zero human coverage gaps - including evenings, nights, weekends, and holidays.
- Order status and shipment tracking fully automated via direct Apilo integration, with verbatim order-ID confirmation eliminating lookup errors.
- Zero-hallucination handling of the highest-risk product facts (allergens, exact composition) through a hardcoded-facts-first knowledge architecture.
- RAG-powered retrieval across ~200 SKUs and general brand content via Pinecone, scoped correctly by namespace for product-specific vs. general queries.
- Deterministic working-hours gate eliminates failed or inappropriate live transfers outside the human consultant's Monday-Friday, 09:00–16:00 window.

Piotr handles the full customer support lifecycle for Empire Shop - order tracking, product expertise, recommendation, and escalation - as a single automated Polish-speaking system operating continuously, every day of the year.