

CASE STUDY

AI Voice Agent for a National Treasury's Peak-Season Contact Operations

CGS Nexus deployed a production-grade AI voice assistant for a South American national treasury, managing high-volume citizen inquiries during tax season with a multi-model AI architecture purpose-built for local accent fidelity, regulatory accuracy, and zero service failures at peak load.



Key results

Production performance in peak tax season deployment

30–40%

Call deflection rate

Contained without follow-up call

60–70%

Fully resolved by AI

No human intervention required

85–90%+

AI accuracy

↑ from 75–80% at pilot

0

Critical failures

Full uptime at peak load

<48

Regulatory update SLA

Deployed to production

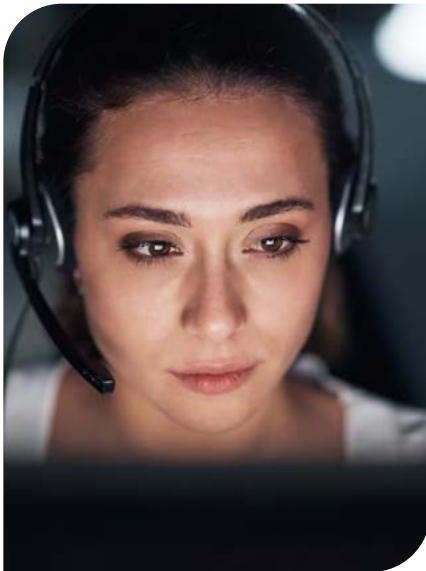


Wait times & call-backs

Significant reduction vs. baseline

The challenge

High-volume,
high-stakes contact
operations under peak
seasonal pressure



The National Treasury, the country's central body for tax collection, debt recovery, and financial disbursement, operates a high-volume citizen contact center that faces severe seasonal demand spikes during filing periods. Most inbound calls involve repetitive, high-frequency inquiries like payment deadlines, procedures, available payment methods, and regulatory requirements.

These predictable but numerous contacts consumed agent capacity, extended wait times, and drove call-back rates, all creating both a citizen experience problem and an operational cost challenge.

- **Peak-season volume surge** Tax filing deadlines generated sharp, predictable inbound spikes that overwhelmed traditional staffing models, spiking abandonment rates, and queue times.
- **Accent & language fidelity** Standard STT/TTS (speech-to-text/text-to-speech) engines underperformed on Chilean Spanish phonology, colloquialisms, and real-world call conditions, making off-the-shelf solutions insufficient.
- **Repetitive Tier-1 burden** 50+ high-frequency inquiry types like deadlines, procedures, and payment options were fully resolvable without agent involvement, yet they consumed most of the call volume.
- **Regulatory change velocity** Tax rules, deadlines, and procedures change frequently. The system required a rapid update pipeline to maintain accuracy and avoid citizen misinformation in a regulated context.

Solution achitecture

Multi-model AI architecture, purpose-built for government scale

The BPO operator designed and deployed a multi-model AI architecture integrated into the client's existing contact center infrastructure. Each layer was purpose-selected for performance and delivered under a white-labeled abstraction in the citizen-facing product.

01

Telephony & routing

- Contact Center platform integration
- IVR orchestration layer
- Real-time call routing

02

Voice AI engine

- Chilean Spanish STT/TTS
- Natural voice synthesis
- Noise-resilient processing

03

NLP / intelligence

- Large language model core
- Intent classification
- Low-confidence detection
- Context management

04

Knowledge & data

- 50+ tax use case knowledge base
- Regulatory updates in <48 hrs
- CRM integration

05

Analytics

- Deflection & containment dashboards
- Accuracy monitoring
- Inquiry trend analysis



Smart escalation

Human handoff with full context transfer

The AI escalates to a live agent when any of four conditions are detected. Full conversation context is transferred. Agents never restart the interaction, eliminating re-verification friction and reducing handle time on transferred calls.

- Low confidence score detected
- Repeated clarification loops
- Personalized / account-specific request
- Explicit caller request to speak with an agent

Scalability & next steps

A reusable platform, not a one-off build

The multi-model design is modular and not tied to a single government agency. The same architecture is deployable to other public services, multilingual contexts, and variable seasonal demand profiles, making it a reusable platform asset rather than a single-client build.

- **Replicable by agency**
Modular architecture can be deployed to any government service with a high-frequency inquiry profile: tax, social security, municipal services.
- **Multilingual ready**
STT/TTS layer is swappable — tested for Chilean Spanish; extensible to other dialects, languages, and regional accent variations.
- **Elastic capacity**
Scales instantly to seasonal demand without headcount changes. Peak load carries zero operational risk with AI-first handling.

Why CGS Nexus

AI in production — not in pilots

CGS Nexus has 20 AI deployments in live production across government, retail, healthcare, infrastructure, education, financial services, postal, and humanitarian sectors. Below is a selection of recent deployments. This is our track record, not our roadmap.

20

Production AI
deployments

8+

Industries
served

4

Solution types
deployed

2+

Government
agencies live

A selection of recent deployments

GOVERNMENT

National Treasury

Voice AI Agent

60–70% AI resolution rate

GOVERNMENT

National Procurement Platform

Voice AI + Auto-QA

27% AHT reduction
91.6% answer rate

RETAIL

Major National Retail Chain

Voice AI Agent

35% containment
16+ use cases

INFRASTRUCTURE

Highway Concession Operator

Voice AI Agent

AI-first deflection

INFRASTRUCTURE

Highway Infrastructure Group

Voice AI Agent

Multi-concession deployment

HEALTHCARE

National Health Network

Chat AI Agent

Patient self-service chatbot

HEALTHCARE

Health Network – Portal

Chat AI + Agent Assist

Agent-assist in contact center

EDUCATION

Technical University

Chat AI Agent

24/7 student self-service

Solution types in production

VOICE AI AGENTS • CHAT AI AGENTS
• COPILOT / AGENT ASSIST • AI ANALYTICS
/ AUTO-QA • RPA & AUTOMATION

If you're responsible for high-volume, high-stakes contact operations and want AI that is tested in production — not stuck in pilot — let's talk about how CGS Nexus can design and deploy a voice automation platform for your tax, social services, or public infrastructure use cases.

Let's start the conversation.

Customersuccess@cgsnexus.com
www.cgsnexus.com

