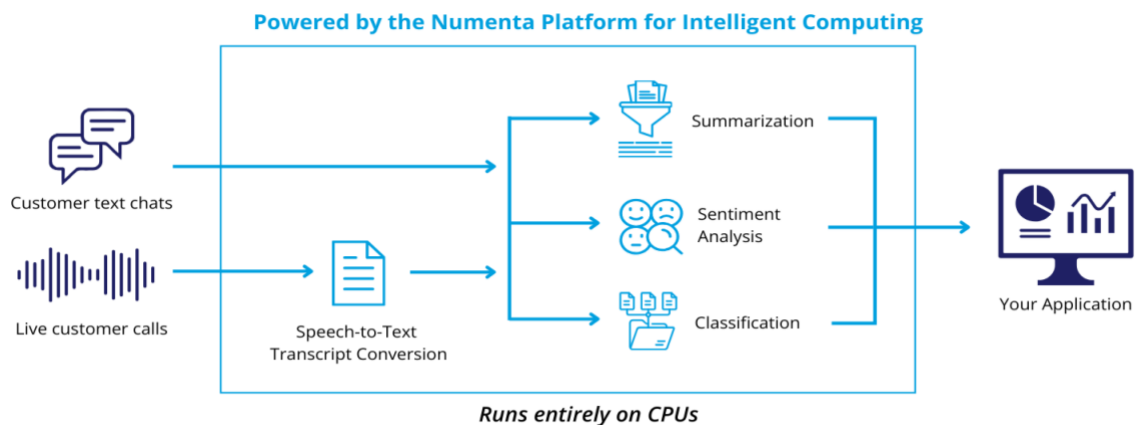


Customer Support – Powered by NuPIC™

In today's fast-paced digital landscape, providing efficient and personalized customer support across multiple channels is challenging. The large volume of support tickets translates to response delays and low quality, posing a risk to overall customer satisfaction. Businesses need an efficient solution that can quickly identify and address customer concerns.

The Numenta Platform for Intelligent Computing (NuPIC™) is a neuroscience-based AI software platform that enables the deployment of Large Language Models (LLMs) on CPUs. It dramatically streamlines customer support operations, enabling faster response times, more accurate issue resolution and higher customer satisfaction.



EXAMPLE IMPLEMENTATION

A business with thousands of daily customer support calls can leverage NuPIC to transcribe calls into searchable text instantly. They can then use NuPIC BERT models for sentiment analysis to categorize calls according to their emotional tone, directing escalations to the relevant support staff. Subsequently, NuPIC GPT models can provide the Operations team with concise summaries of the flagged tickets, allowing them to handle tickets more efficiently and focus on complex cases.

NuPIC Capabilities	Description
Sentiment Analysis	Classify customer sentiments to prioritize and route tickets to the appropriate support agent.
Classification	Categorize tickets based on predefined criteria such as complaints, questions, or requests.
Summarization	Summarize and maintain context of previous communications across multiple channels.
Speech-to-Text	Convert customer calls into accurate, searchable text transcripts instantly.

NUPIC BENEFITS

PERFORMANCE IMPROVEMENTS OVER GPUS

- **High Throughput:** NuPIC models are 7-17x faster for inference on Intel Xeon CPUs compared to GPUs.
- **Low Latency:** NuPIC does not require batching for high performance, making it ideal for real-time low-latency applications.
- **Run Multiple Models on the Same Server:** You can run dozens of different models concurrently and asynchronously on a single CPU server.

PLUG-AND-PLAY SCALABILITY

- **Manage Multiple Clients:** NuPIC can handle a large number of asynchronous, separate client requests while maintaining high throughput.

- **Seamless Integration:** NuPIC easily integrates into standard MLOps solutions such as Kubernetes.

SAVINGS AND PRIVACY

- **Delivered via Docker:** You can install NuPIC on any CPU-based server in your infrastructure, on premise, private cloud, or even on your laptop.
- **No GPUs Required:** NuPIC accelerates AI inference on CPUs, leading to advantages in inferences per dollar and inferences per watt.
- **Run in Your Private Network:** Your data and models reside completely on your private storage systems, giving you complete control over updates and data governance policies.

HOW IT WORKS

With a library of production-ready pretrained models that can be customized to a variety of natural language processing use cases, you can directly run the models in the highly-optimized NuPIC Inference Server. You can also fine-tune the models on your data with the NuPIC Training Module, then deploy your custom fine-tuned model to the NuPIC Model Library and run it in the inference server.

TECHNICAL CONFIGURATIONS

Models:	NuPIC BERT, NuPIC GPT (similar to LLAMA2)	Operating System:	Recent version of Linux (Ubuntu 22.04) is required to ensure kernel support for AMX
Minimum Processor:	Any Intel-compatible server with AVX512/AVX2/AMX instruction set support	Software:	Docker, Python 3.8 or later
Recommended Processor:	AMX enabled server such as AWS m7i.4xlarge	Training:	GPU with at least 12GB RAM
Memory & Storage:	16GB RAM, 200GB of storage space		

GET STARTED TODAY

Discover the power of AI-driven customer support with NuPIC. Contact us for a demo to discuss how we can tailor our solutions to meet your business needs: numenta.com/demo



889 Winslow Street, 4th Floor
Redwood City, CA 94063

info@numenta.com

As a world leader in deploying large AI models on CPUs, Numenta has mapped its neuroscience-based advances to modern CPU architectures to redefine what's possible in AI. Numenta's AI platform NuPIC, the Numenta Platform for Intelligent Computing, helps businesses leverage the flexibility of CPUs to build robust AI applications that are efficient, scalable, and secure.