

# Espressive Barista Large Language Model (LLM) Gateway™

Integrating large language models, like ChatGPT, with virtual agents, like Espressive Barista, holds the promise of delivering an even faster time-to-value with an improved employee experience. At the same time, organizations are concerned about potential data privacy and confidentiality breaches from allowing their employees access to LLMs. The Espressive Barista LLM Gateway solves that problem by providing safe and responsible access to LLMs.

## Our Large Language Model – The Employee Language Cloud

Espressive started building our domain-specific large language model (LLM), the Barista Employee Language Cloud™, over 6 years ago. Today it is trained on over 4B phrases across 15 enterprise departments in over 100 languages. The Employee Language Cloud is specifically focused on the employee self-help use case. It incorporates technologies that enable industry-leading accuracy and deflection rates with a layer that is dynamically tuned to understand the nuances of an enterprise's lexicon. The Employee Language Cloud also includes automations, integrations, and curated knowledge to deliver the best possible employee experience.

Despite the size of our Employee Language Cloud, the convenience of Barista means there are times when employees will ask questions not typically asked of a service desk. That is where generalized LLMs come into play as they can provide responses to a wealth of technical and non-technical questions outside of the typical employee domain. We also leverage the generative AI capabilities of LLMs to create responses to service desk questions, when appropriate. By enabling integration with generalized LLMs like ChatGPT 3.5, Espressive gives customers the best of all worlds — a domain-specific LLM that is focused on automating resolution of employee questions combined with the massive knowledge and generative capabilities of a generalized LLM..

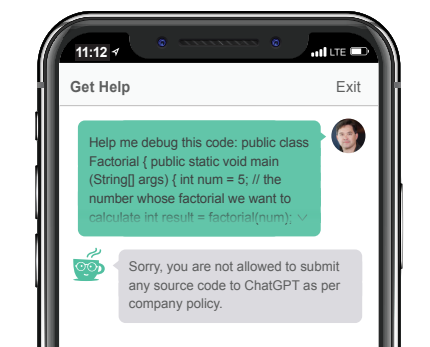
## A Safe And Responsible Approach To Integrating LLMs With A Virtual Support Agent

While LLMs like ChatGPT are extremely promising, Wall Street banks including JPMorgan Chase, Goldman Sachs, Citigroup, and others are banning or limiting use due to perceived risks. These organizations are considering the potential data privacy and confidentiality breaches as well as the accuracy and reliability of the output.

That is why Espressive is not simply enabling integration to these LLMs. Instead, Espressive Barista is acting as a gateway to LLMs, enabling corporations to enforce their policies, such as not permitting the sharing of source code or PII, as well as providing the ability to "turn off" LLM access for areas such as Epic Systems in healthcare organizations, or HR questions. For instance, an employee can ask about the best restaurants in the area, but not be able to ask how to best diagnose a patient.

Safeguards include:

- Verify policy compliance (e.g., no source code, no PII, no data with customer names)
- Disable access to LLMs for specific content areas (e.g., Epic or HR)
- Restrict questions to only those that are work related
- Ensure employees understand that responses from LLMs may not be accurate

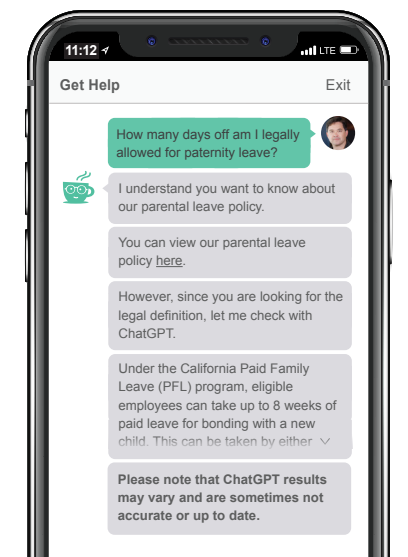
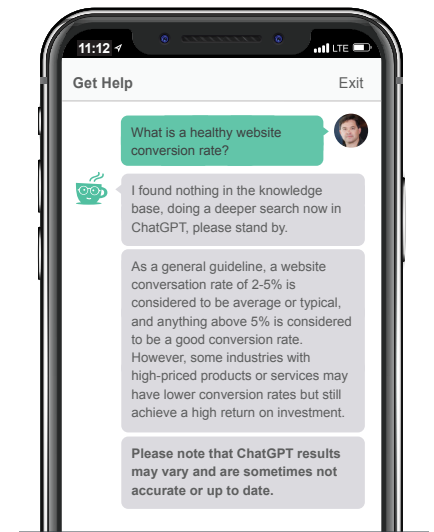


In addition, customers receive analytics on usage which is helpful in many ways. First, they can understand what employees are asking, which is indicative of what they care about. Next, they can see where content gaps exist so they can focus their teams on content that matters most. Also, they can see who the top users of the LLM are to monitor where cost is being allocated. Optionally, by notifying employees that there are analytics, they could discourage them from overuse and asking inappropriate questions.

## Integration With Experience Selector™

The Experience Selector within Espressive Barista uniquely considers relevance and possible outcomes to provide the best experience to employees. It prioritizes automating resolution versus providing instructions for an employee to read. It identifies if content or automation will result in a less than optimal outcome for employees, and when that occurs it uses Smart Ticketing™ to create, attribute, triage, and route a ticket to the right service team, minimizing mean time to resolution (MTTR).

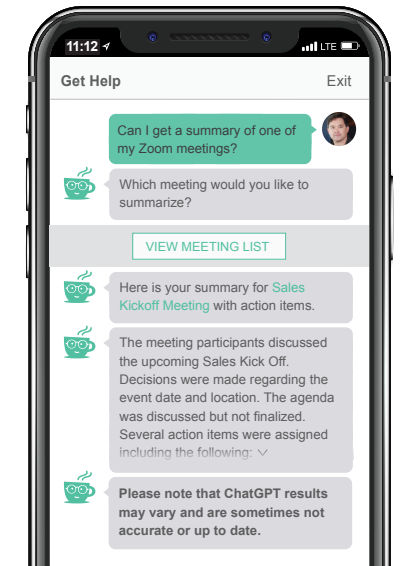
With the Barista LLM Gateway, the Experience Selector can now choose a response from an LLM like ChatGPT. The Experience Selector will always prioritize automation or a customer's content versus displaying a generative response from an LLM. This prioritization ensures employees follow the best practices recommended by their corporations for compliance. However, when there is no understanding on how to act on an employee question, if enabled, the employee will be notified that Barista is doing a deeper search in ChatGPT. Barista will then present the results while letting the employee know that the results may vary and are sometimes inaccurate.



## Barista Control Center™ Hides The Complexity Of Prompt Engineering

The Barista Control Center combines conversational AI, employee context, automation, and prompt engineering to ensure it unlocks the full power of ChatGPT and other LLMs. With LLMs the response is only as good as the question, and getting the best outcome requires employees to become experts in prompt engineering. With Espressive, customers are assured we will leverage known information and integrations to deliver the best possible outcome to their employees, without the complexity.

In this example, an employee is asking how many days off he is "legally" allowed for paternity leave without providing any context regarding which laws he is referring to. Because of the prompt engineering via the Control Center, Barista adds his location (i.e., San Francisco) to the question before asking ChatGPT, so that the answer will be more accurate.



## Unlocking New Capabilities In Barista

Historically, Barista has helped employees around the world get assistance with work related questions or issues across 15 different departments from IT to Payroll and beyond. For example, an employee might ask, "How do I get access to DocuSign," or, "There is an issue with my paycheck." By integrating Barista into an LLM like ChatGPT, Espressive is now unlocking a whole world of new workplace assistance and employee productivity capabilities, such as gaining tips on making hybrid meetings more effective. And with Barista integrations to third-party systems such as Zoom, employees could request a list of Zoom meetings, even requesting meeting summaries and action items. This drives greater employee adoption and satisfaction, which in turn delivers an even higher ROI.

## Over 100 Languages, Each Leveraging The Best Possible Translation Engine

Espressive was the first to support translations in over 100 languages, and with the Barista LLM Gateway, can now leverage LLMs based on the best possible translation engine for a given topic in a given language. For instance, Barista could use Lionbridge for Latin based languages, ChatGPT for Japanese, and Baidu for Chinese.



**ESPRESSIVE** is the pioneer in automating digital workplace assistance, redefining how employees get help by delivering exceptional employee experiences. Barista, our VSA, takes on the role of the service agent, bringing the best of human experience with the best of artificial intelligence. Barista automates resolution of employee questions with personalized experiences that result in employee adoption of 80 to 85% and reduced help desk call volume of 50 to 70% . Visit us at [Espressive.com](https://www.espressive.com) to learn more and [request a demo](#).