

The small language models (SLMs) that power Arcee Orchestra.

MODEL NAME	WHAT IT EXCELS AT	DESCRIPTION	TOP USE CASES
Virtuoso Large	General Purpose	Our most powerful and versatile general-purpose model, designed to excel at handling complex and varied tasks across domains. With state-of-the-art performance, it offers unparalleled capability for nuanced understanding, contextual adaptability, and high accuracy.	<ul style="list-style-type: none"> Advanced content creation, such as technical writing and creative storytelling Data summarization and report generation for cross-functional domains Detailed knowledge synthesis and deep-dive insights from diverse datasets Multilingual support for international operations and communications
Virtuoso Medium	General Purpose	A versatile and powerful model, capable of handling complex and varied tasks with precision and adaptability across multiple domains. Ideal for dynamic use cases requiring significant computational power.	<ul style="list-style-type: none"> Content generation Knowledge retrieval Advanced language understanding Comprehensive data interpretation
Virtuoso Small	General Purpose	A streamlined version of Virtuoso, maintaining robust capabilities for handling complex tasks across domains while offering enhanced cost-efficiency and quicker response times.	<ul style="list-style-type: none"> General-purpose task handling Business communication Automated document processing for mid-scale applications
Caller Large	Tool & Function Calls	Engineered for seamless integrations, Caller-Large is a robust model optimized for managing complex tool-based interactions and API function calls. Its strength lies in precise execution, intelligent orchestration, and effective communication between systems, making it indispensable for sophisticated automation pipelines.	<ul style="list-style-type: none"> Managing integrations between CRMs, ERPs, and other enterprise systems Running multi-step workflows with intelligent condition handling Orchestrating external tool interactions like calendar scheduling, email parsing, or data extraction Real-time monitoring and diagnostics in IoT or SaaS environments
Coder Large	Coding	A high-performance model tailored for intricate programming tasks, Coder-Large thrives in software development environments. With its focus on efficiency, reliability, and adaptability, it supports developers in crafting, debugging, and refining code for complex systems.	<ul style="list-style-type: none"> Writing modular, reusable code across various programming languages Debugging and optimizing performance in large-scale applications Generating efficient algorithms for computationally intensive tasks Supporting DevOps processes, such as script automation and CI/CD pipelines
Coder Small	Coding	A compact, high-performance coding model designed for efficient programming tasks, including generating code, debugging, and optimizing scripts for smaller projects.	<ul style="list-style-type: none"> Lightweight development tasks Automated code reviews Generating templates or prototypes quickly, code completion
Spotlight	Vision-Language	Spotlight is a specialized vision-language model designed to bridge the gap between visual and textual data. It excels at interpreting images, generating insights, and combining multimodal data to make complex information more accessible.	<ul style="list-style-type: none"> Analyzing visual data for industry-specific applications like medical imaging, architecture, or fashion Multimodal storytelling through synchronized text and visuals Automated annotation of large-scale image datasets for AI training Extracting insights from charts, graphs, and infographics for business analytics
Maestro	Reasoning Engine	Maestro is the pinnacle of reasoning and analytical AI, designed to tackle the most complex problem-solving scenarios with unparalleled precision and depth. It excels in handling multifactorial decision-making, abstract reasoning, and scenario modeling, delivering insights that rival expert-level human analysis.	<ul style="list-style-type: none"> Strategic decision-making for high-stakes scenarios, such as financial modeling or geopolitical analysis Comprehensive diagnostics and root-cause analysis across technical and non-technical domains Advanced simulation modeling for predicting outcomes or evaluating "what-if" scenarios Thought leadership assistance for crafting well-reasoned arguments, strategies, and long-term plans

And Featuring

Metronome	Intelligent Router	A small yet highly efficient 300M parameter model trained to route messages to the most appropriate model in the Arcee Orchestra lineup, based on the context of the input.	<ul style="list-style-type: none"> Query classification Task delegation Operational efficiency in hybrid workflows involving multiple models
-----------	--------------------	---	---