≡ 〰 **Nesa**   🔍

AIVM Interface for Model Interaction

The AIVM frontend interface serves as the gateway for users to interact with the on-chain model and AIVM repository, facilitating the uploading, managing, and monitoring of AI models and their corresponding virtual machines.

The frontend interface caters to both novice users with Nesa's Minimal Code Suite which offers presets for smaller or non-proprietary models, as well as complete customization ability for experienced developers.

In this section, we outline the features and functionalities of the interface that facilitate user communication with Nesa. The overall flow of our system for managing the AIVM is shown in below.
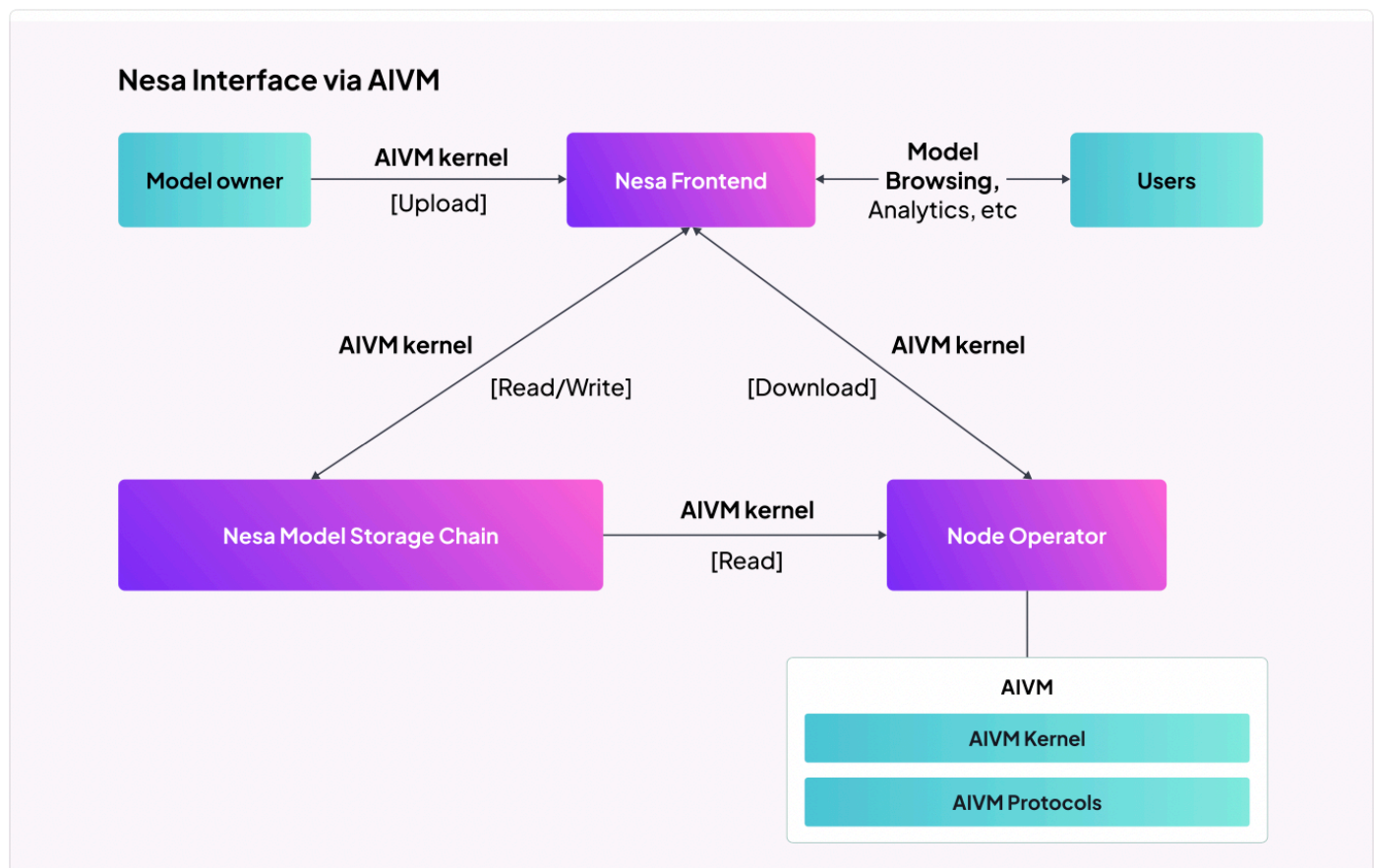
Nesa inference and management via AIVM. AI Model owners can upload and manage AIVM kernels by using the Nesa frontend. The kernels are stored on a dedicated decentralized storage chain connected to the Nesa network. Node operators can either download kernels from the Nesa frontend or directly

from the chain. AI dApp users can browse the frontend to obtain rich information about various models supported by our community.

Model Browsing and Discovery

Users explore the repository through the front-end interface, which provides a catalog of available AIVM kernels with rich associated model information. The interface supports granular search and filtering to find models that suit a user's specific needs based on criteria such as model type, complexity, and performance metrics.

Detailed Model Information

The interface presents detailed metadata for each AI model uploaded to Nesa's AIVM repository, including authorship, version history, and performance benchmarks, that can be reviewed when selecting a model to deploy for inference.

Model Upload and Registration

For users and developers looking to contribute new models to the platform, the interface provides a guided upload process. This includes the submission of model parameters, the AIVM configuration file, and the inference code, along with any necessary documentation. The upload and registration process ensures that all models on Nesa adhere to the required standards for AIVM compatibility.

Model Deployment and Execution

The interface facilitates turnkey deployment of AI models from repository selection to inference task initiation. Users can input their data, configure execution parameters, and submit their requests directly through the interface. Nesa handles the execution of the model within the AIVM and returns the results to the user.

Security and Privacy Controls

When dealing with private or proprietary models, the frontend interface provides tools for managing encryption keys and access controls. Users can specify which nodes are authorized to execute their models and distribute decryption keys through secure means.

Real-time Monitoring and Analytics

The interface provides real-time monitor- ing of model performance and usage statistics, letting users track the activity of their deployed models, analyze usage patterns, and gain insights into the performance of their inference tasks.

The AIVM interface democratizes access to AI by allowing users from diverse backgrounds to participate in the sharing, creation, deployment, management, and collective evaluation of AI models.

In the future, this will become a full-fledged AI model marketplace on-chain, complete with version control, reviews, and lookbacks for model performance and example output, at the discretion of the model creator and model's users.

The AIVM has been engineered to cater to the full spectrum of developer skills, from beginner developers who are seeking default presets that allow them to upload their model and then plug and play, through to advanced developers who seek complete customization ability of the end-to-end system, from control over their kernel configuration, to the type of inference, security, consensus, and output that they seek.

Nesa's future expansion plans for AIVM include further interoperability of the system with other blockchain networks to drive universal adoption of the platform, and a roadmap of new modular features that will give creators and users more control in how they build, present, and implement AIVM kernels.

For a detailed look at these plans, please see this article's Addendum.

Previous
On-Chain Model and AIVM Repository

Next
Decentralized Inference

Last updated 1 month ago