

WEBCON

AI Proxy Self-hosted

1.0.0.235

history of changes for version 1.0.0.235

1. New features

- Released **AI Proxy Self-hosted** — a new component that enables customers to use AI functionality in WEBCON based on their own infrastructure, completely bypassing WEBCON infrastructure. The solution is available under the Enterprise subscription license and the perpetual WEBCON Agentic AI Framework license.

AI Proxy Self-hosted can be installed in the customer's environment (on-premise or private cloud), with the option to connect and configure the customer's own AI models — hosted in Google Vertex, OpenAI, or Azure AI Foundry. This gives customers full control over the data and infrastructure used for AI processing, making it easier to meet security requirements, privacy policies, and industry regulations.

A dedicated AI Proxy Self-hosted mode has been introduced, along with a new configuration node for this mode, added to the `appsettings.json` configuration files used to run AI Proxy, including:

- the option to enable Azure Key Vault for storing sensitive data,
- a variable storing the path to a `.pem` or `.pfx` certificate along with its password.

An alternative class implementation has also been prepared which, in this mode, allows Azure Key Vault to be replaced with environment variables.

A new, unified AI Proxy Self-hosted provider-configuration model based on priorities has been designed and released. It allows defining multiple providers, automatically switching to lower-priority providers, and assigning different AI models to specific methods (text, image, and audio models). The configuration is defined in the `aiconfiguration.json` file, which is loaded when the AI Proxy application starts.

The ability to select the application's operating mode in the configuration file has been added, along with support for local certificates (`.pem/.pfx`) used to sign and verify access tokens — without the need to communicate with external authorization services (including LAS) or databases. In AI Proxy Self-hosted mode, dependencies on external infrastructure components have been reduced: the application can run without connections to licensing services, databases, or cloud data stores, and the required mechanisms are automatically selected based on configuration.