

Pellerex Managed API Service

Context

Companies spend months and months of development time and effort to create scalable engineering environments and infrastructure with all the compliance, organisation's non-negotiables, engineering standards, DevOps and automation building blocks.

And that's exactly what Pellerex's Managed API infrastructure provides. An environment where you can provision APIs, and assert Identity and Security standards while accessing your APIs.

You could also use this capability to host and run different types of workload, such as running AI models at scale. Especially Deep Learning models that are resource-intensive and need a lot of processing capacity (GPU, CPU), memory, and storage, and if utilised for small number of models, it will not be cost-effective to run and maintain. Pellerex runs a lot of AI models, and hence has the ability to shrink the cost while providing the same level of performance as your dedicated Cloud infrastructure.

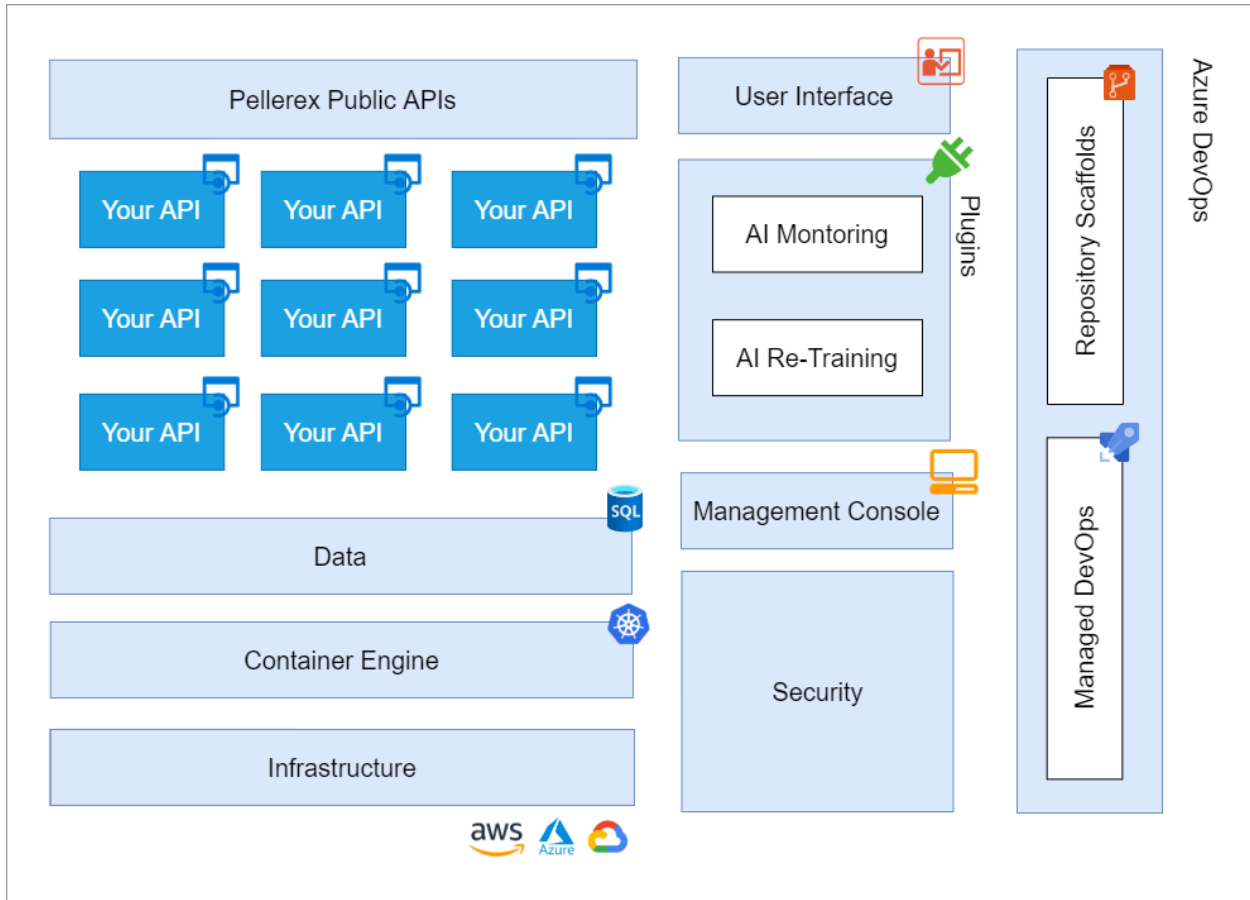
The end result is a public facing API that you could consume from your application anywhere at commercial scales.

Pellerex provides an ecosystem of functions (such as Identity, API, and Containerised Infrastructure) to run and share your work with the world. Hence you can be purely focused on what you are very good at (i.e. your product), and avoid the complexity of creating such engineering infrastructure to expose and share your work.

To make your life super easy, we have taken automation to the next level and hence, all you need to do is to push your code to the Git repository we provision for you, and your API will be up and running behind a public endpoint in less than 2 minutes.

Architecture

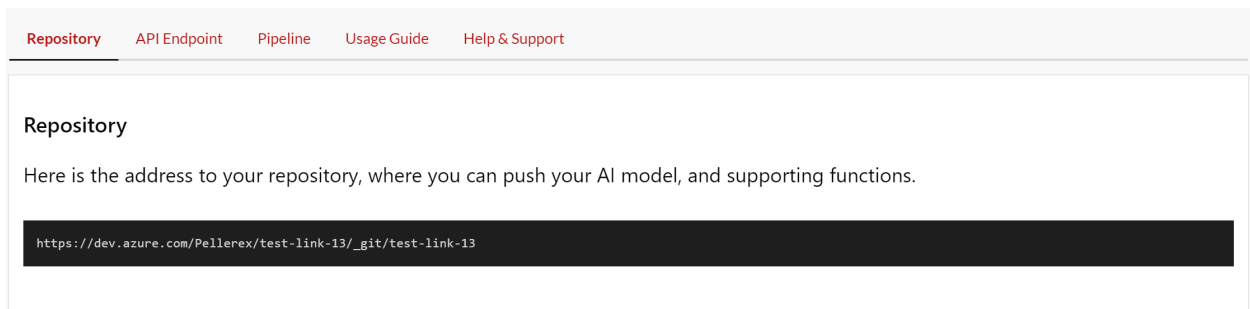
Looking under the hood, we have a containerised infrastructure that allows us to support every technology and stack your team needs to use with highest level of infrastructure utilisation. Once your API gets deployed, it is fully isolated from the rest of the workload other users deploy, and hence we fully protect your IP and confidentiality. Below diagram is a high level demonstration of our infrastructure and how it works.



Technical Walk-through

Once you subscribe to a license, you will have access to AI Hosting Service via your [dashboard](#).

After you go through the API provisioning process, you will receive an invite to Pellerex organisation on Azure DevOps, through which you will have access to your code repository, with a Python scaffold.



Navigate to your repository and you will see the Python scaffold repository to push your model into:

Files ✓ succeeded [Clone](#)

Contents History

Name ↑	Last change	Commits
Infrastructure	Oct 31	6da01058 Tokenise Pellerex.com
.gitattributes	Feb 15	40a831cf Init Pellerex.com
.gitignore	Feb 15	40a831cf Init Pellerex.com
PY app.py	Feb 15	40a831cf Init Pellerex.com
Dockerfile	Feb 15	40a831cf Init Pellerex.com
Python.SDK.sln	Oct 31	6da01058 Tokenise Pellerex.com
readme.md	Feb 15	40a831cf Init Pellerex.com
requirements.txt	Feb 15	40a831cf Init Pellerex.com

Python API Scaffold: [Pellerex.com](#)

This scaffold is used to accelerate development of AI powered APIs for Pellerex AI platform.

You should use this scaffold if you are a data scientist building AI models to be deployed using Pellerex Machine Learning as a Service infrastructure.

It supports the below features:

- Basic template to build APIs using [Flask-RESTful](#)
- [Docker file](#) to container your Python API and test it prior to deployment

Navigate to your app.py file in your repository, and create a route through which you'd like to access your model:

Once you are done with your changes, raise a pull request and merge the change to the master branch. You can then monitor the build/deployment status of your change through the Pipeline tab in your model's dashboard:

Repository [API Endpoint](#) **Pipeline** [Usage Guide](#) [Help & Support](#)

Pipelines

Here you can find Build and Deployment status of your model pipelines.

✓ **Build** Last Build: 31 Oct 22 04:59:22 UTC

✓ **Deployment** Last Deployment: 31 Oct 22 05:00:28 UTC

At the end, you can access call the API responsible for your model through the address in the Api Endpoint tab:

API Endpoint

Once your API is deployed, you can access your AI model functions through the below URL. Just append the path with the routes you have coded in your repo, and provide the payload your model requires.

```
https://api.pellerex.com/.../7af7a9dd-5bb4-4212-b640-7663a0bdc11
```