SONY

Vision and Sensing Application SDK AI Model and PPL Deploy Functional Specifications

Version 0.2.0

2023 - 1 - 30

TOC

1. Change history 1
2. Terms/Abbreviations 2
3. Reference materials 3
4. Expected use case 4
5. Functional overview/Algorithm 5
6. User interface specifications(Deploy AI model) 19
7. User interface specifications(Cancel AI model deployment state) 25
8. User interface specifications(Delete deployment configuration) 28
9. User interface specifications(Deploy PPL) 31
10. Target performances/Impact on performances 35
11. Assumption/Restriction 36
12. Remarks 37
13. Unconfirmed items 38

1. Change history

Date	What/Why
2023/01/30	Initial draft

2. Terms/Abbreviations

Terms/Abbreviations	Meaning
Deployment configuration	Used to deploy AI models. Specify the AI model to deploy, etc. and register it in Console for AITRIOS. Run a deployment by specifying a registered deployment configuration. Note that it is separate from the `configuration.json` to configure notebook runtime settings.
PPL	A module that processes the output of the AI model(Output Tensor) of edge AI devices

3. Reference materials

- Reference/Related documents
  - API Reference
    - https://developer.aitrios.sony-semicon.com/development-guides/reference/api-references/
  - Console User Manual
    - https://developer.aitrios.sony-semicon.com/file/download/console-developer-edition-ui-manual

4. Expected use case

- You want to deploy an AI model imported into Console for AITRIOS to edge AI devices
- You want to deploy a PPL imported into Console for AITRIOS to edge AI devices
- You want to check AI model or PPL deployment state

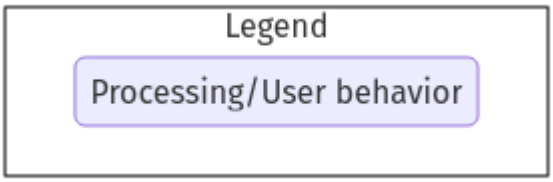
5. Functional overview/Algorithm

Functional overview

- Users can use Console Access Library in the SDK's Dev Container (Local PC or Codespaces)
  - Users can do the following through the Console Access Library:
    - Get device information for deployment
    - Register a new deployment configuration to deploy AI model
    - Get a list of registered deployment configurations
    - Delete a registered deployment configuration
    - Deploy an AI model to edge AI devices
    - Deploy a PPL to edge AI devices

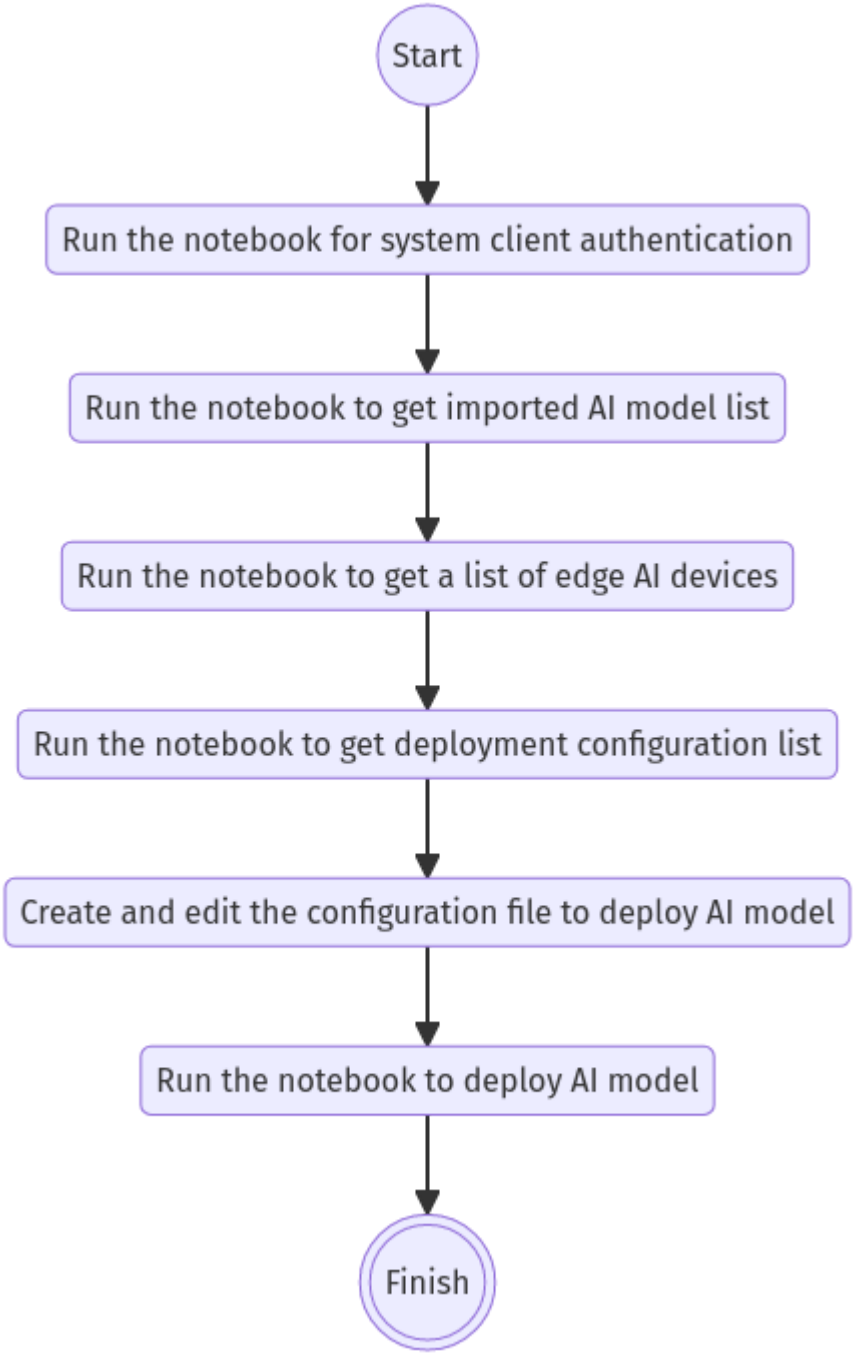
To run inference after deploying an AI model,
set the model ID using command parameter of the Console for AITRIOS.
These are beyond the scope of this book.
See *Console User Manual* for details.

Flow



Deploy AI model

- Flow

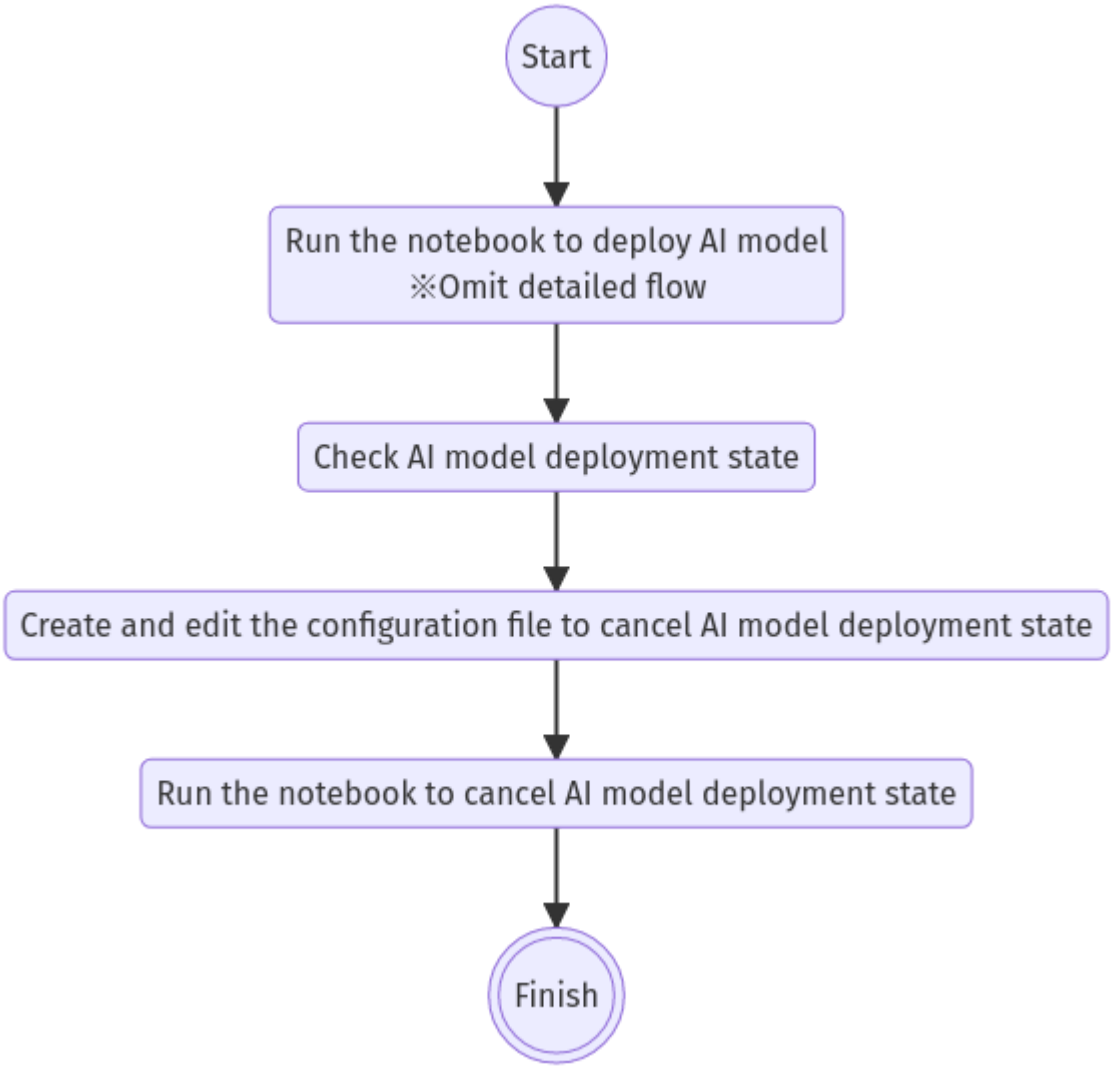


- Flow details
    1. Run the notebook for system client authentication
    2. Run the notebook to get imported AI model list
        - Run the notebook to get a list of AI models that have been imported into Console for AITRIOS, and get settings in the configuration file, **`model_id`**.
            - See *configuration.json* for details.
    3. Run the notebook to get a list of edge AI devices
        - Run the notebook to get a list of edge AI devices registered in Console for AITRIOS, and get settings in the configuration file, **`device_id`**, **`model_id`**, and **`model_version`**.
            - See *configuration.json* for details.
    4. Run the notebook to get deployment configuration list
        - Get deployment configuration to deploy AI model
            - Run the notebook to get a list of deployment configurations registered in Console for AITRIOS, and get settings in the configuration file, **`config_id`**.
              See *configuration.json* for details.
    5. Create and edit the configuration file to deploy AI model
        - Create and edit the configuration file *configuration.json* to configure notebook runtime settings
    6. Run the notebook to deploy AI model

Cancel AI model deployment state

- Flow

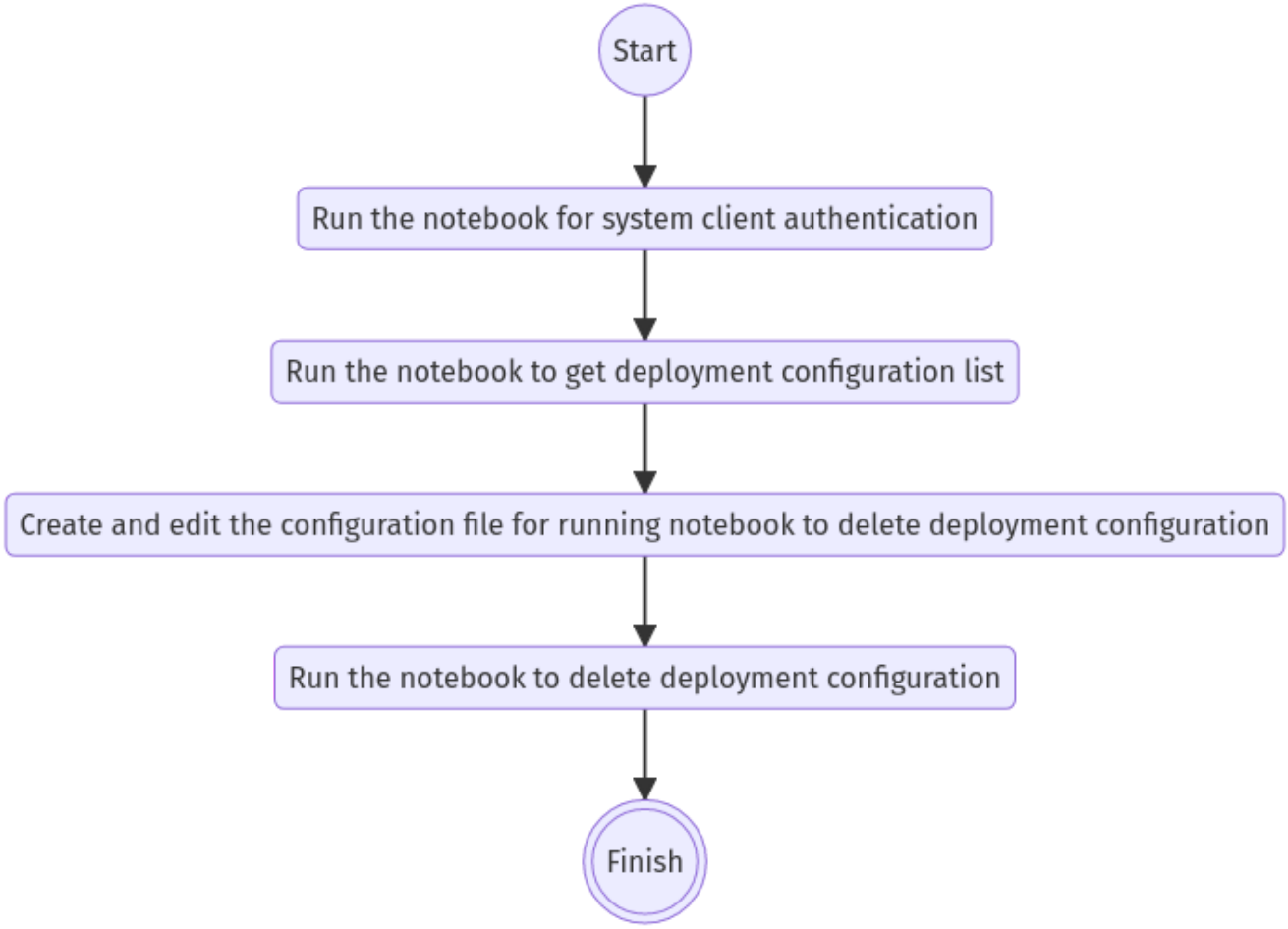
> "Cancel AI model deployment state" is to reset state on the database.
> Use when a edge AI device stops responding after deploying AI model, leaving deployment state "running" on database.
> You can't rerun deployment in this state and must cancel.
> (Do not rerun the notebook to deploy AI model in this state.)
> You can't cancel deployment to edge AI devices.
> You can't recover that edge AI device stops responding by SDK.
> Restart or reset by other means.



- Flow details
  1. Run the notebook to deploy AI model
     - See *flow* for details
  2. Check AI model deployment state
     - Run the notebook to deploy AI model and check the deployment results
  3. Create and edit the configuration file to cancel AI model deployment state
     - Create and edit the configuration file *configuration.json* to configure notebook runtime settings
  4. Run the notebook to cancel AI model deployment state

Delete deployment configuration

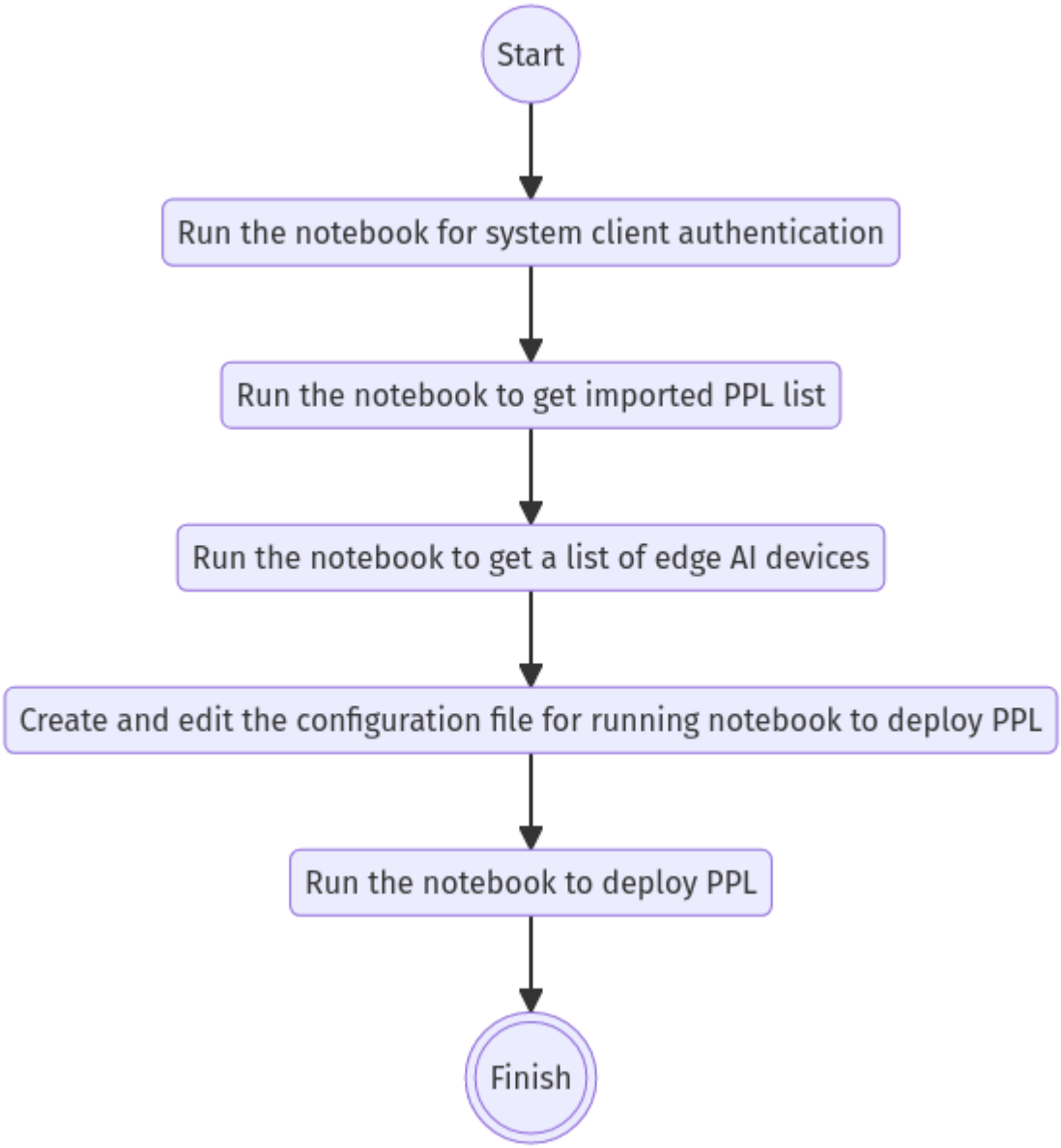
- Flow



- Flow details
  1. Run the notebook for system client authentication
  2. Run the notebook to get deployment configuration list
     - Run the notebook to get a list of deployment configurations registered in Console for AITRIOS, and get settings in the configuration file, `config_id`.
  3. Create and edit the configuration file for running notebook to delete deployment configuration
     - Create and edit the configuration file *configuration.json* to configure notebook runtime settings
  4. Run the notebook to delete deployment configuration
     - Run the notebook to delete deployment configuration specified in the configuration file from Console for AITRIOS

Deploy PPL

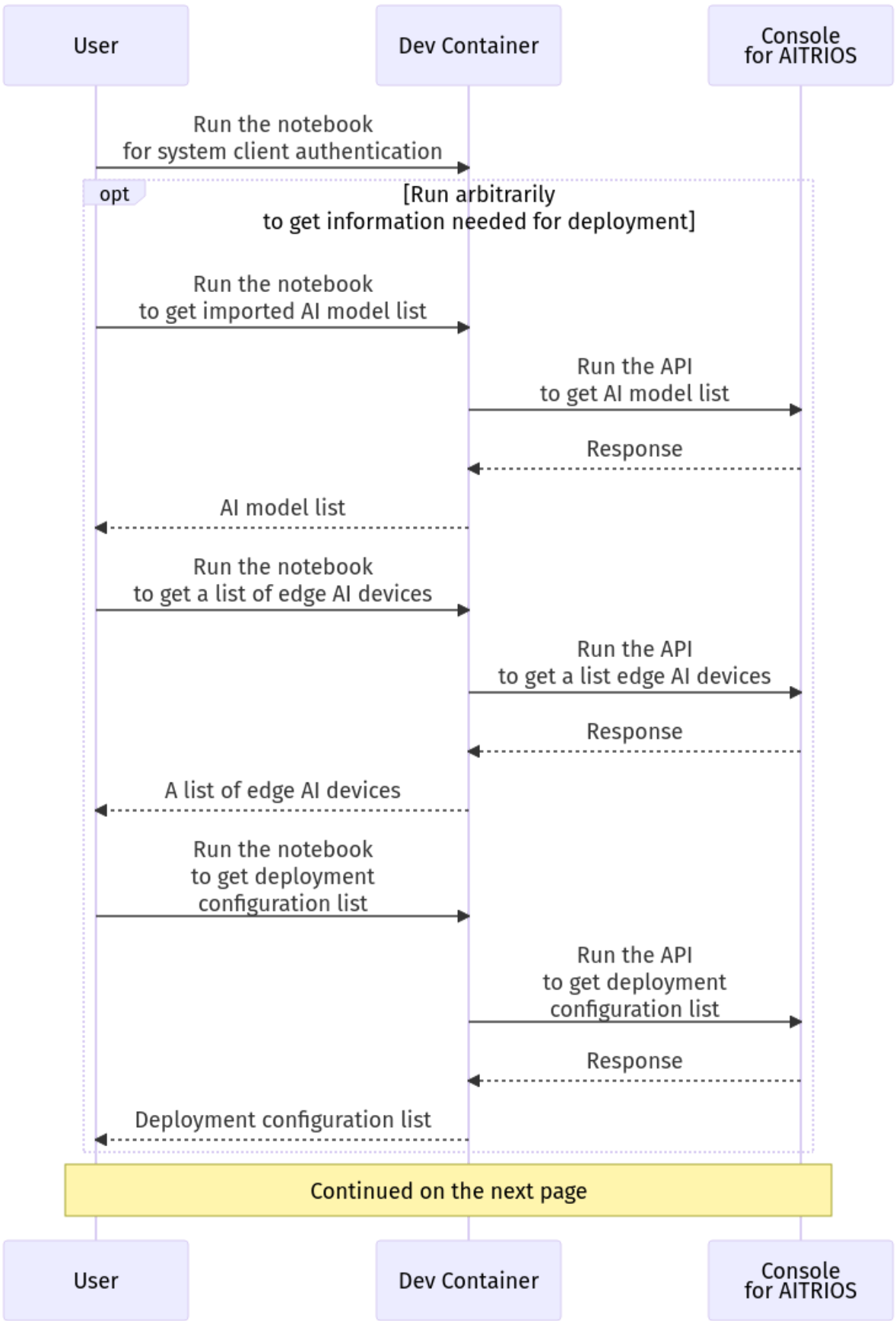
- Flow

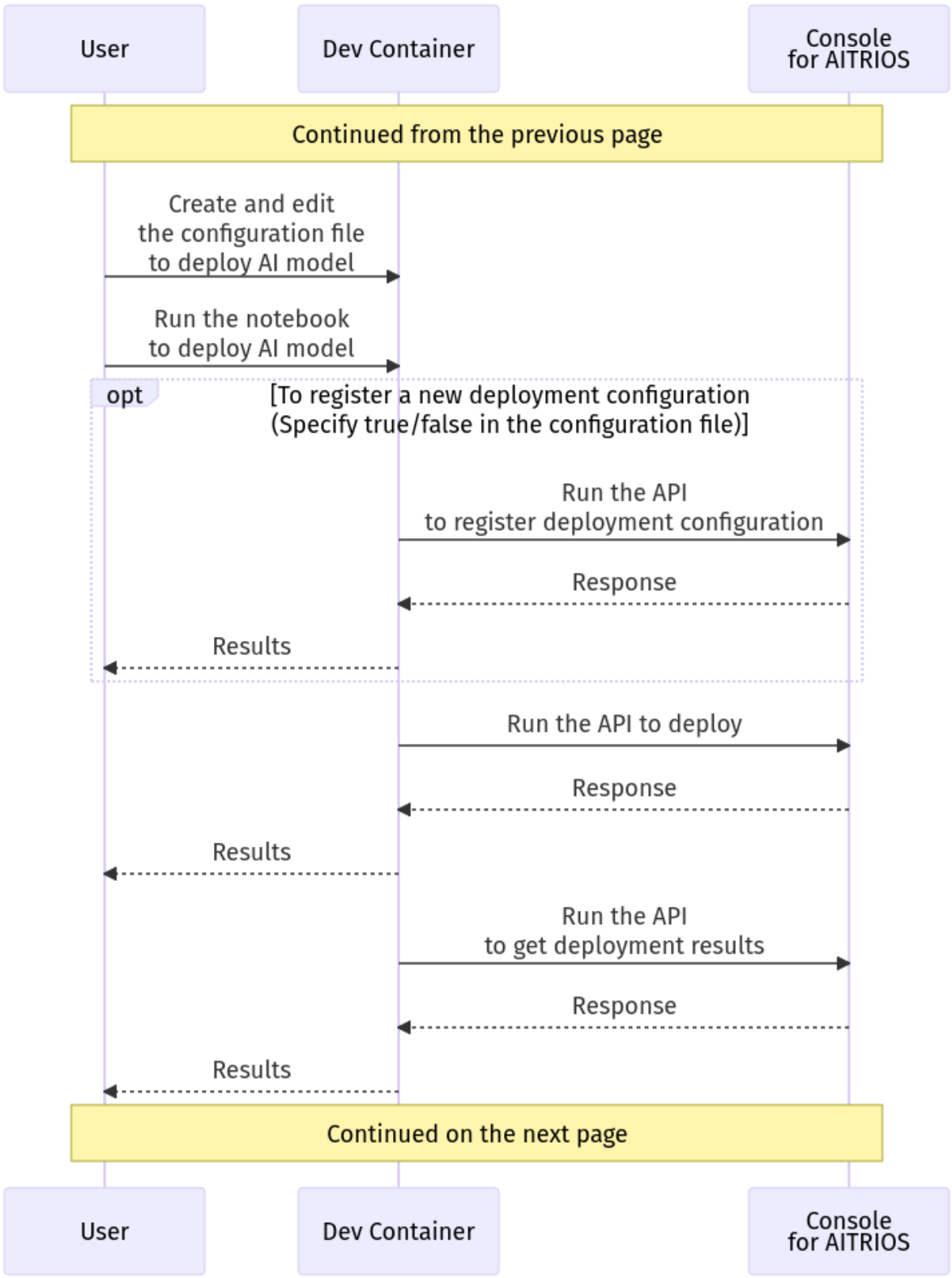

Start
Run the notebook for system client authentication
Run the notebook to get imported PPL list
Run the notebook to get a list of edge AI devices
Create and edit the configuration file for running notebook to deploy PPL
Run the notebook to deploy PPL
Finish

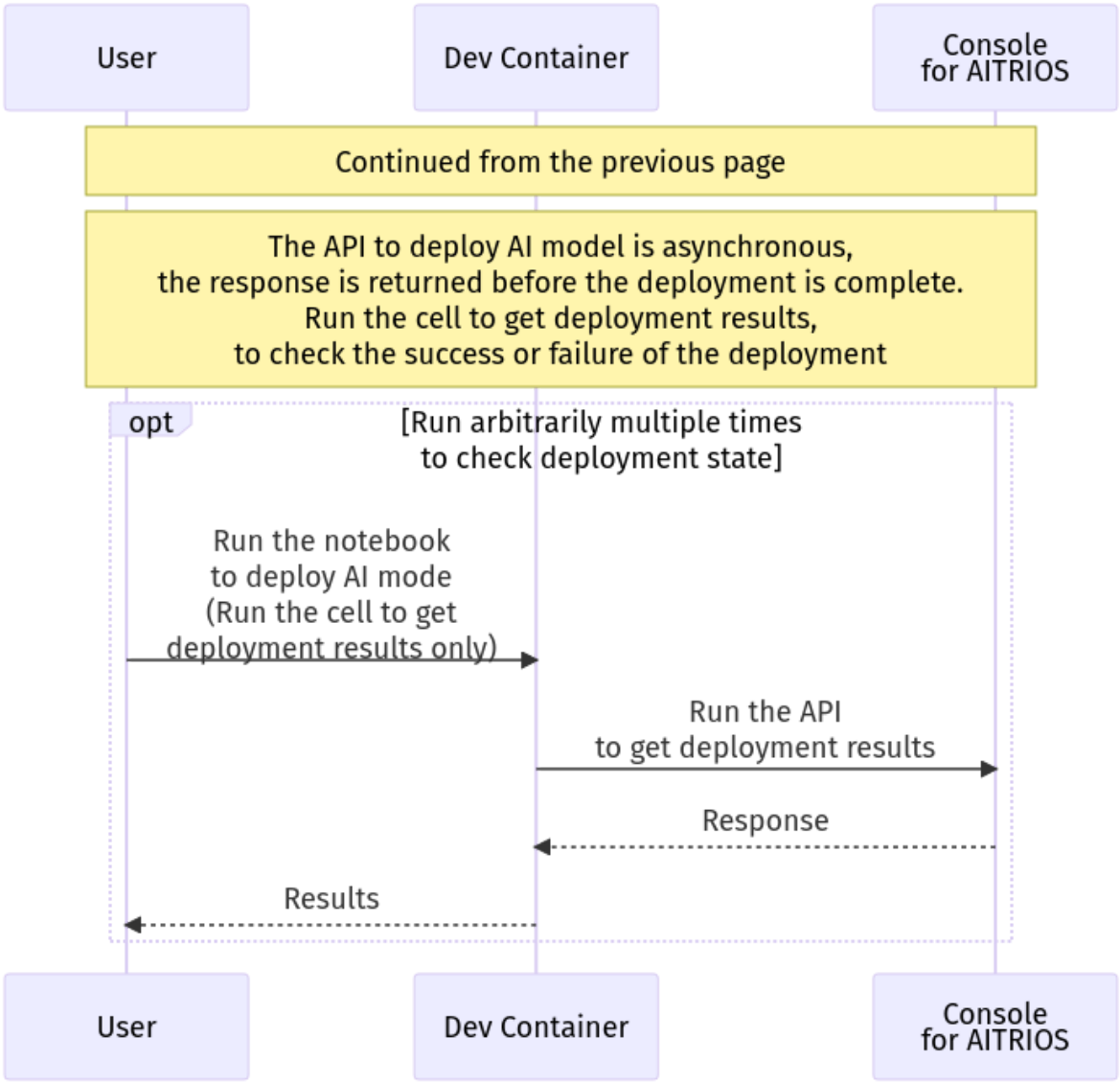
- Flow details
    1. Run the notebook for system client authentication
    2. Run the notebook to get imported PPL list
        - Run the notebook to get a list of PPL that have been imported into Console for AITRIOS, and get settings in the configuration file, **`app_name`** and **`version_number`**.
            - See *configuration.json* for details.
    3. Run the notebook to get a list of edge AI devices
        - Run the notebook to get a list of edge AI devices registered in Console for AITRIOS, and get settings in the configuration file, **`device_id`**.
            - See *configuration.json* for details.
    4. Create and edit the configuration file for running notebook to deploy PPL
        - Create and edit the configuration file *configuration.json* to configure notebook runtime settings
    5. Run the notebook to deploy PPL

Sequence

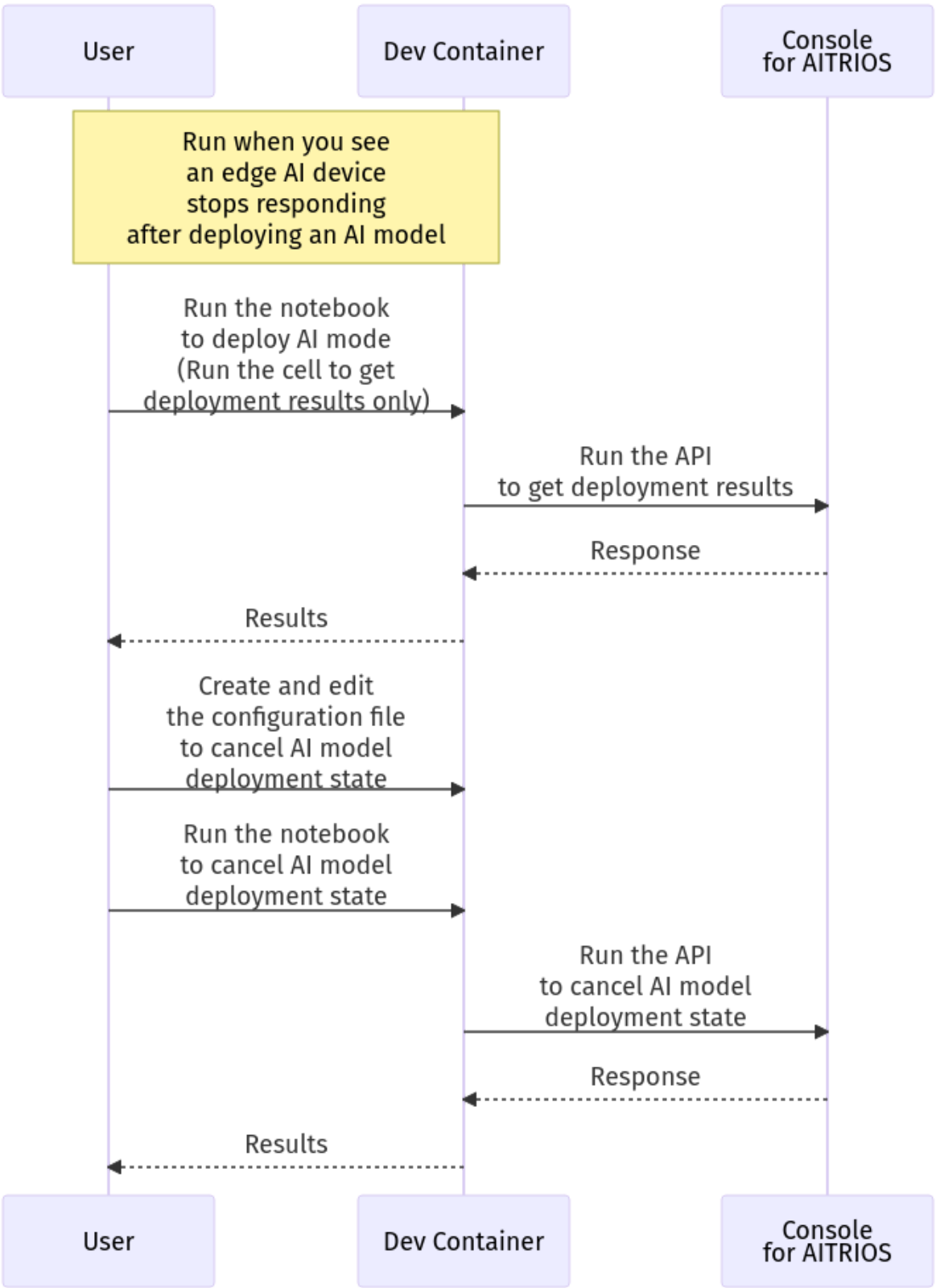
Deploy AI model



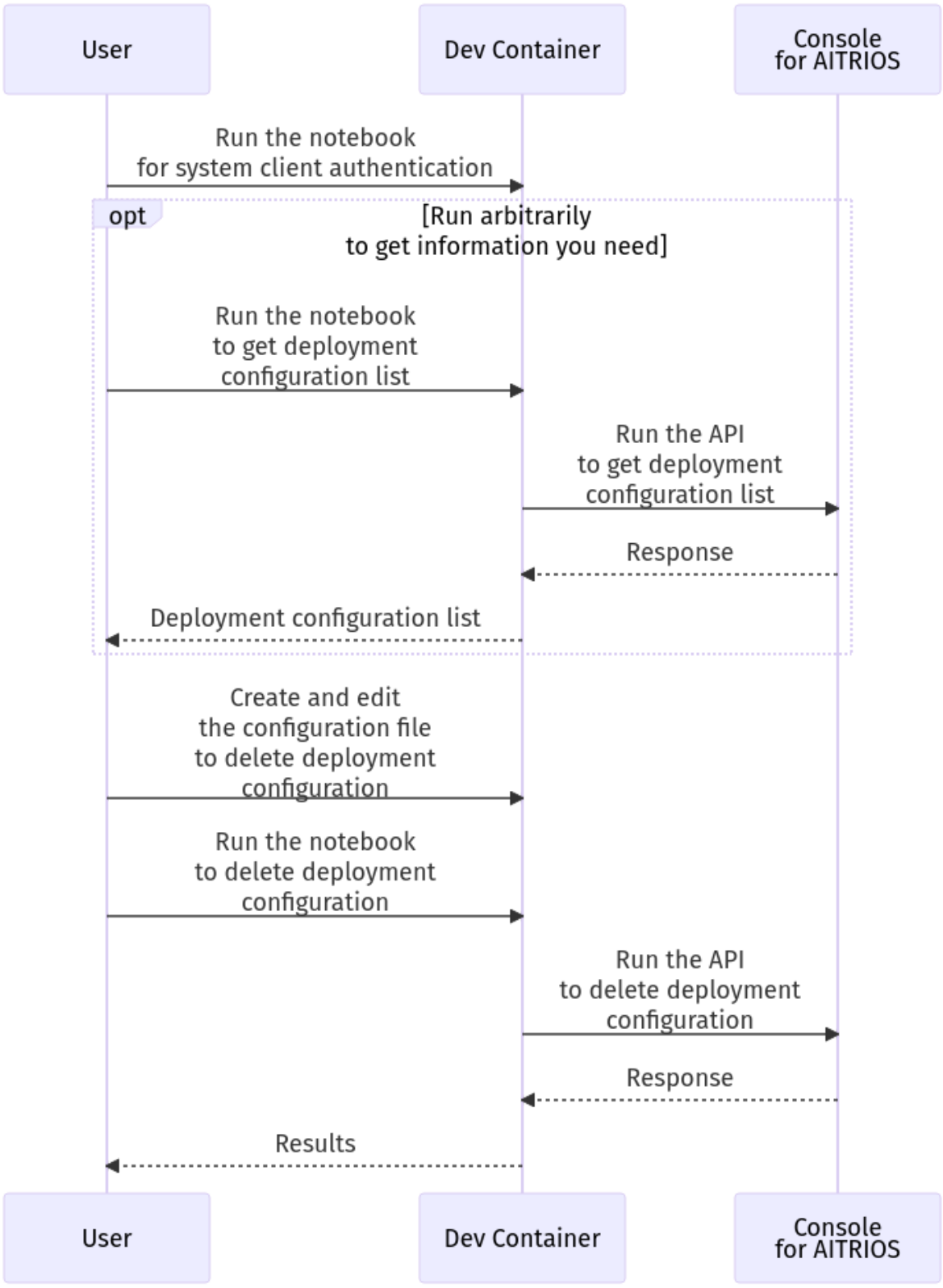

User
Dev Container
Console for AITRIOS
Continued from the previous page
Create and edit the configuration file to deploy AI model
Run the notebook to deploy AI model
opt
[To register a new deployment configuration (Specify true/false in the configuration file)]
Run the API to register deployment configuration
Response
Results
Run the API to deploy
Response
Results
Run the API to get deployment results
Response
Results
Continued on the next page
User
Dev Container
Console for AITRIOS


User
Dev Container
Console for AITRIOS
Continued from the previous page
The API to deploy AI model is asynchronous,
the response is returned before the deployment is complete.
Run the cell to get deployment results,
to check the success or failure of the deployment
opt
[Run arbitrarily multiple times
to check deployment state]
Run the notebook
to deploy AI mode
(Run the cell to get
deployment results only)
Run the API
to get deployment results
Response
Results
User
Dev Container
Console for AITRIOS

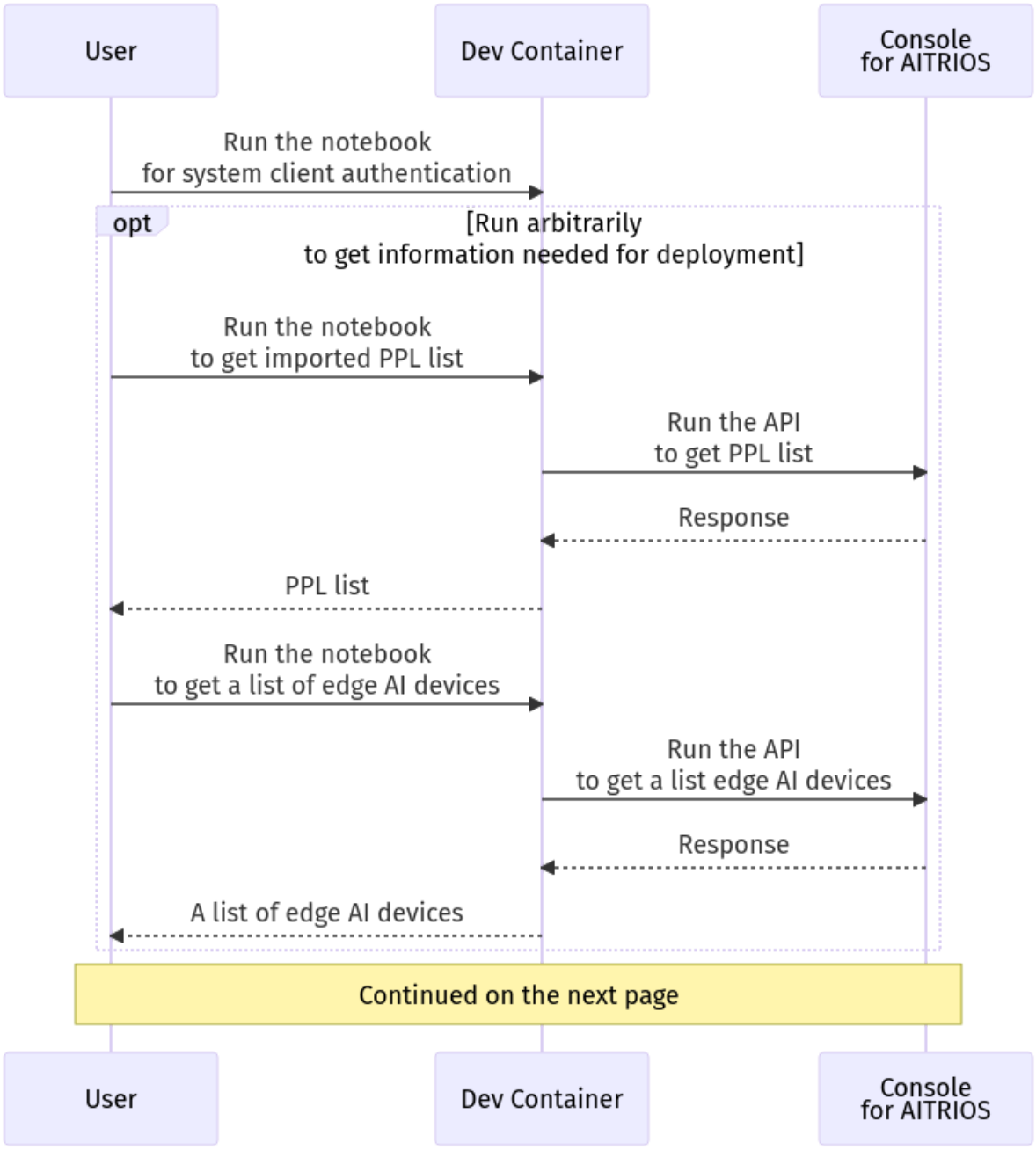
Cancel AI model deployment state

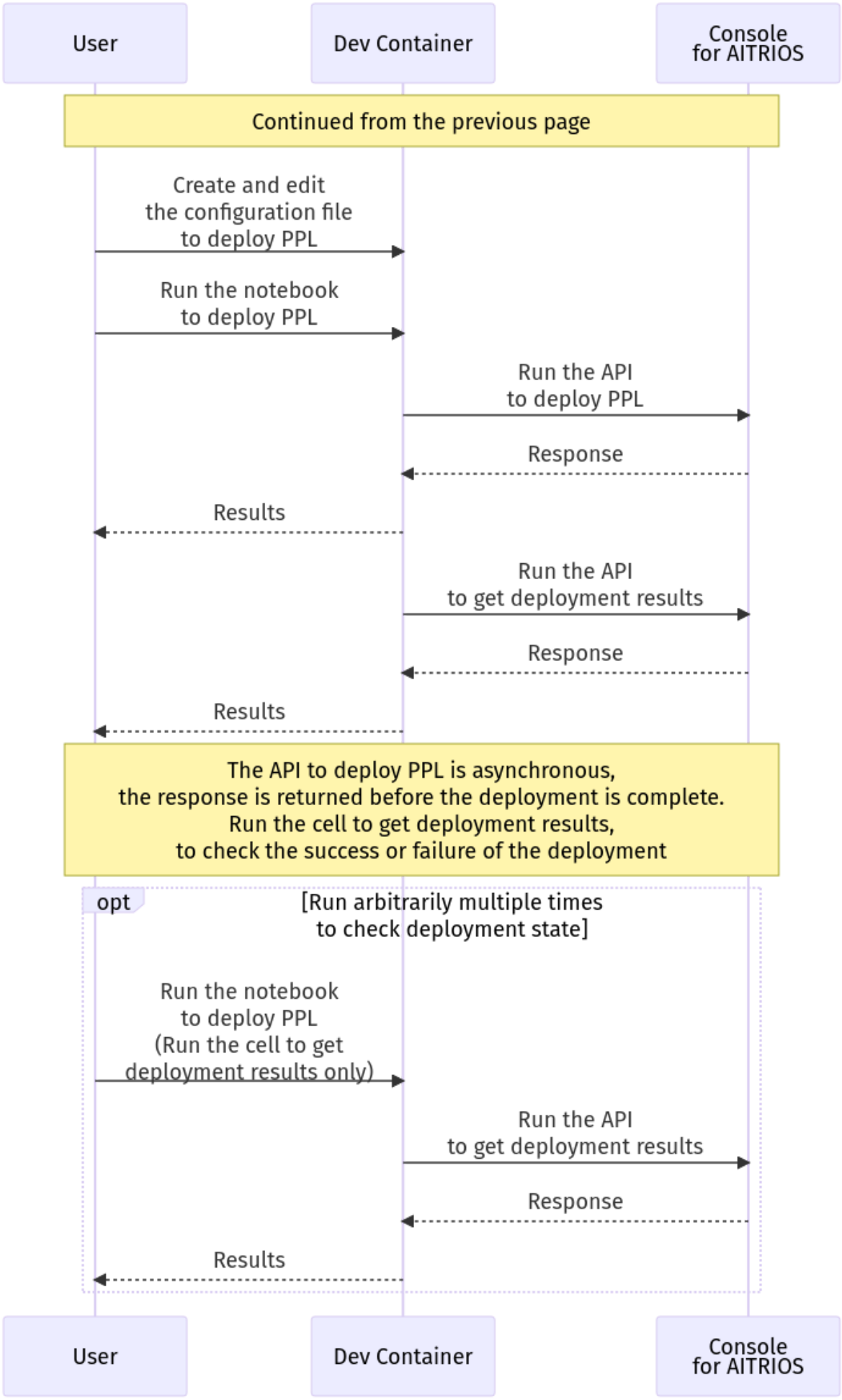


Delete deployment configuration



Deploy PPL


User
Dev Container
Console for AITRIOS
Run the notebook for system client authentication
opt
[Run arbitrarily to get information needed for deployment]
Run the notebook to get imported PPL list
Run the API to get PPL list
Response
PPL list
Run the notebook to get a list of edge AI devices
Run the API to get a list edge AI devices
Response
A list of edge AI devices
Continued on the next page
User
Dev Container
Console for AITRIOS


User
Dev Container
Console for AITRIOS
Continued from the previous page
Create and edit the configuration file to deploy PPL
Run the notebook to deploy PPL
Run the API to deploy PPL
Response
Results
Run the API to get deployment results
Response
Results
The API to deploy PPL is asynchronous, the response is returned before the deployment is complete. Run the cell to get deployment results, to check the success or failure of the deployment
opt
[Run arbitrarily multiple times to check deployment state]
Run the notebook to deploy PPL (Run the cell to get deployment results only)
Run the API to get deployment results
Response
Results
User
Dev Container
Console for AITRIOS

6. User interface specifications(Deploy AI model)

Prerequisite

- You have registered as a user through Portal for AITRIOS and participated in the AITRIOS project
- You have uploaded an AI model to the Console for AITRIOS

How to start each function

1. Launch the SDK environment and preview the **`README.md`** in the top directory
2. Jump to the **`README.md`** in the **`tutorials`** directory from the hyperlink in the SDK environment top directory
3. Jump to the **`README.md`** in the **`3_prepare_model`** directory from the hyperlink in the **`README.md`** in the **`tutorials`** directory
4. Jump to the **`README.md`** in the **`develop_on_sdk`** directory from the hyperlink in the **`README.md`** in the **`3_prepare_model`** directory
5. Jump to the **`README.md`** in the **`4_deploy_to_device`** directory from the hyperlink in the **`README.md`** in the **`develop_on_sdk`** directory
6. Jump to the **`README.md`** in the **`deploy_to_device`** directory from the hyperlink in the **`README.md`** in the **`4_deploy_to_device`** directory
7. Jump to each feature from each file in the **`deploy_to_device`** directory

Run the notebook for system client authentication

1. Jump to the **`README.md`** in the **`set_up_console_client`** directory from the hyperlink in the **`README.md`** in the **`deploy_to_device`** directory
2. Open the notebook for system client authentication, **.ipynb*, in the **`set_up_console_client`** directory, and run the python scripts in it

Run the notebook to get imported AI model list

1. Jump to the **`README.md`** in the **`get_model_list`** directory from the hyperlink in the **`README.md`** in the **`deploy_to_device`** directory

2. Open the notebook to get AI model list, **.ipynb*, in the **`get_model_list`** directory, and run the python scripts in it
    - If successful, information about the AI models imported into Console for AITRIOS, such as model ID, version, etc., is displayed in the notebook

Run the notebook to get a list of edge AI devices

1. Jump to the **`README.md`** in the **`get_device_list`** directory from the hyperlink in the **`README.md`** in the **`deploy_to_device`** directory
2. Open the notebook to get a list of edge AI devices, **.ipynb*, in the **`get_device_list`** directory, and run the python scripts in it
    - If successful, information about the edge AI devices registered in Console for AITRIOS, such as device ID, deployed model ID, etc., is displayed in the notebook

Run the notebook to get deployment configuration list

1. Jump to the **`README.md`** in the **`get_deploy_config`** directory from the hyperlink in the **`README.md`** in the **`deploy_to_device`** directory
2. Open the notebook to get deployment configuration list, **.ipynb*, in the **`get_deploy_config`** directory, and run the python scripts in it
    - If successful, information about the deployment configurations registered in Console for AITRIOS, such as config ID, etc., is displayed in the notebook

Create and edit the configuration file to deploy AI model

All parameters are required, unless otherwise indicated.

All values are case sensitive, unless otherwise indicated.

The parameters passed to the Console Access Library API are as specified in the *Console Access Library API*.

1. Create and edit the configuration file, **`configuration.json`**, in the **`deploy_to_device`** directory.

Configuration	Meaning	Range	Remarks
should_create_deploy_config	Whether to register new deployment configuration	true or false true:New registration false:Use registered	Don't abbreviate
config_id	ID of the deployment configuration ・Specify any character string for new registration ・If using registered, specify its ID	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API ・**deployment.deployment.Deployment.create_deploy_configuration** ・**deployment.deployment.Deployment.deploy_by_configuration**

Configuration		Meaning	Range	Remarks
create_config	**comment**	Description of the newly registered deployment configuration	String Details follow the Console Access Library API specification.	Optional ・Use to register a new deployment configuration. Used for the following Console Access Library API. ・ **deployment.deployment.Deployment.create_deploy_configuration**
	model_id	ID of the AI model to deploy Specify the ID of an imported AI model	String Details follow the Console Access Library API specification.	Optional. But don't abbreviate this to register a new deployment configuration. ・Use to register a new deployment configuration. Used for the following Console Access Library API. ・ **deployment.deployment.Deployment.create_deploy_configuration**
	model_version_number	Version of the AI model to deploy Specify the version of an imported AI model	String Details follow the Console Access Library API specification.	Optional ・Use to register a new deployment configuration. Used for the following Console Access Library API. ・ **deployment.deployment.Deployment.create_deploy_configuration**
device_ids		ID of the edge AI devices to deploy AI model	List of strings	Don't abbreviate Used for the following Console Access Library API. ・ **deployment.deployment.Deployment.deploy_by_configuration**

Configuration	Meaning	Range	Remarks
`replace_model_id`	ID of the AI model to be replaced Specify the ID of the AI model to replace (overwrite) among the models deployed on the device	String Details follow the Console Access Library API specification.	Optional if you don't replace the AI model If not specified when the number of models deployed on the edge AI device has reached the limit, an error occurs. Used for the following Console Access Library API ・ **`deployment.deployment.Deployment.deploy_by_configuration`**
`comment`	Deployment description	String Details follow the Console Access Library API specification.	Optional Used for the following Console Access Library API ・ **`deployment.deployment.Deployment.deploy_by_configuration`**

Run the notebook to deploy AI model

1. Open the notebook, **`deploy_to_device.ipynb`**, in the **`deploy_to_device`** directory, and run the python scripts in it
    - The script does the following:
        - Checks that *configuration.json* exists in the **`deploy_to_device`** directory
            - If an error occurs, the error description is displayed and running is interrupted.
        - Checks the contents of *configuration.json*
            - If an error occurs, the error description is displayed and running is interrupted.
        - Checks the contents of **`configuration.json`** for **`should_create_deploy_config`**
            - If true, run the API to register deployment configuration
                - If the deployment configuration is successfully registered, **`deploy_to_device.ipynb`** displays a successful message
                - If an error occurs, the error description is displayed and running is interrupted.
        - Run the API to deploy AI model
            - If API execution is successful, **`deploy_to_device.ipynb`** displays a successful message
            - If an error occurs, the error description is displayed and running is interrupted.
        - Run the API to get AI model deployment results
            - If results are gotten successfully, **`deploy_to_device.ipynb`** displays a successful message and deployment results
            - If an error occurs, the error description is displayed and running is interrupted.
            - See Cloud SDK Console Access Library(Python) Functional Specifications for details on errors and response times

7. User interface specifications(Cancel AI model deployment state)

Use when a edge AI device stops responding after an AI model deployment and the deployment state on the database remains "running".
Operation is not guaranteed when this function is executed under normal conditions.

Prerequisite

- You have registered as a user through Portal for AITRIOS and participated in the AITRIOS project
- After deploying an AI model, check the deployment state to determine whether to cancel the deployment state

How to start each function

1. Launch the SDK environment and preview the `README.md` in the top directory
2. Jump to the `README.md` in the `tutorials` directory from the hyperlink in the SDK environment top directory
3. Jump to the `README.md` in the `3_prepare_model` directory from the hyperlink in the `README.md` in the `tutorials` directory
4. Jump to the `README.md` in the `develop_on_sdk` directory from the hyperlink in the `README.md` in the `3_prepare_model` directory
5. Jump to the `README.md` in the `4_deploy_to_device` directory from the hyperlink in the `README.md` in the `develop_on_sdk` directory
6. Jump to the `README.md` in the `cancel_deploy_state` directory from the hyperlink in the `README.md` in the `4_deploy_to_device` directory
7. Jump to each feature from each file in the `cancel_deploy_state` directory

Create and edit the configuration file to cancel AI model deployment state

All parameters are required, unless otherwise indicated.

The parameters passed to the Console Access Library API are as specified in the *Console Access Library API*.

1. Create and edit the configuration file, **`configuration.json`**, in the **`cancel_deploy_state`** directory.

Configuration	Meaning	Range	Remarks
`device_id`	ID of the edge AI device to cancel deployment state	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API ・ **`deployment.deployment.Deployment.cancel_deployment`**
`deploy_id`	Deployment ID to cancel deployment state	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API ・ **`deployment.deployment.Deployment.cancel_deployment`**

After running the notebook to deploy AI model, deployment results and state are displayed. Get settings in the configuration file, **`device_id`** and **`deploy_id`** from them.

Run the notebook to cancel AI model deployment state

1. Open the notebook, **`cancel_deploy_state.ipynb`**, in the **`cancel_deploy_state`** directory, and run the python scripts in it
    - The script does the following:
        - Checks that *configuration.json* exists in the **`cancel_deploy_state`** directory
            - If an error occurs, the error description is displayed and running is interrupted.
        - Checks the contents of *configuration.json*
            - If an error occurs, the error description is displayed and running is interrupted.
        - Run the API to cancel AI model deployment state
            - If API execution is successful, **`cancel_deploy_state.ipynb`** displays a successful message
            - If an error occurs, the error description is displayed and running is interrupted.

When the API is executed, the deployment state on the Console transitions from "Running" to "Canceled".

8. User interface specifications(Delete deployment configuration)

Prerequisite

- You have registered as a user through Portal for AITRIOS and participated in the AITRIOS project
- You have registered a deployment configuration in the Console for AITRIOS

How to start each function

1. Launch the SDK environment and preview the **`README.md`** in the top directory
2. Jump to the **`README.md`** in the **`tutorials`** directory from the hyperlink in the SDK environment top directory
3. Jump to the **`README.md`** in the **`3_prepare_model`** directory from the hyperlink in the **`README.md`** in the **`tutorials`** directory
4. Jump to the **`README.md`** in the **`develop_on_sdk`** directory from the hyperlink in the **`README.md`** in the **`3_prepare_model`** directory
5. Jump to the **`README.md`** in the **`4_deploy_to_device`** directory from the hyperlink in the **`README.md`** in the **`develop_on_sdk`** directory
6. Jump to the **`README.md`** in the **`delete_deploy_config`** directory from the hyperlink in the **`README.md`** in the **`4_deploy_to_device`** directory
7. Jump to each feature from each file in the **`delete_deploy_config`** directory

Run the notebook for system client authentication

1. Jump to the **`README.md`** in the **`set_up_console_client`** directory from the hyperlink in the **`README.md`** in the **`delete_deploy_config`** directory
2. Open the notebook for system client authentication, **.ipynb*, in the **`set_up_console_client`** directory, and run the python scripts in it

Run the notebook to get deployment configuration list

1. Jump to the **`README.md`** in the **`get_deploy_config`** directory from the hyperlink in the **`README.md`** in the **`delete_deploy_config`** directory

2. Open the notebook to get deployment configuration list, **.ipynb*, in the `get_deploy_config` directory, and run the python scripts in it
    - If successful, information about the deployment configurations registered in Console for AITRIOS, such as config ID, etc., is displayed in the notebook

Create and edit the configuration file for running notebook to delete deployment configuration

All parameters are required, unless otherwise indicated.

The parameters passed to the Console Access Library API are as specified in the *Console Access Library API*.

1. Create and edit the configuration file, `configuration.json`, in the `delete_deploy_config` directory.

Configuration	Meaning	Range	Remarks
`config_id`	ID of the deployment configuration to delete	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API ・ `deployment.deployment.Deployment.delete_deploy_configuration`

Run the notebook to delete deployment configuration

1. Open the notebook, **`delete_deploy_config.ipynb`**, in the **`delete_deploy_config`** directory, and run the python scripts in it
    - The script does the following:
        - Checks that *configuration.json* exists in the **`delete_deploy_config`** directory
            - If an error occurs, the error description is displayed and running is interrupted.
        - Checks the contents of *configuration.json*
            - If an error occurs, the error description is displayed and running is interrupted.
        - Runs the API to delete deployment configuration
            - If deletion is successful, **`delete_deploy_config.ipynb`** displays a successful message
    - If an error occurs, the error description is displayed in the **`delete_deploy_config.ipynb`** and running is interrupted.
        - See Cloud SDK Console Access Library(Python) Functional Specifications for details on errors and response times

9. User interface specifications(Deploy PPL)

Prerequisite

- You have registered as a user through Portal for AITRIOS and participated in the AITRIOS project
- You have uploaded a PPL to the Console for AITRIOS

How to start each function

1. Launch the SDK environment and preview the **`README.md`** in the top directory
2. Jump to the **`README.md`** in the **`tutorials`** directory from the hyperlink in the SDK environment top directory
3. Jump to the **`4_prepare_application`** directory from the hyperlink in the **`README.md`** in the **`tutorials`** directory
4. Jump to the **`README.md`** in the **`3_deploy_to_device`** directory from the hyperlink in the **`README.md`** in the **`4_prepare_application`** directory
5. Jump to each feature from each file in the **`3_deploy_to_device`** directory

Run the notebook for system client authentication

1. Jump to the **`README.md`** in the **`set_up_console_client`** directory from the hyperlink in the **`README.md`** in the **`3_deploy_to_device`** directory
2. Open the notebook for system client authentication, **.ipynb*, in the **`set_up_console_client`** directory, and run the python scripts in it

Run the notebook to get imported PPL list

1. Jump to the **`README.md`** in the **`get_application_list`** directory from the hyperlink in the **`README.md`** in the **`3_deploy_to_device`** directory
2. Open the notebook to get PPL information list, **.ipynb*, in the **`get_application_list`** directory, and run the python scripts in it
    - If successful, information about the PPL imported into Console for AITRIOS, such as application name, version, etc., is displayed in the notebook

Run the notebook to get a list of edge AI devices

1. Jump to the **`README.md`** in the **`get_device_list`** directory from the hyperlink in the **`README.md`** in the **`3_deploy_to_device`** directory
2. Open the notebook to get a list of edge AI devices, **.ipynb*, in the **`get_device_list`** directory, and run the python scripts in it
    - If successful, information about the edge AI devices registered in Console for AITRIOS, such as device ID, etc., is displayed in the notebook

Create and edit the configuration file for running notebook to deploy PPL

All parameters are required, unless otherwise indicated.

The parameters passed to the Console Access Library API are as specified in the *Console Access Library API*.

1. Create and edit the configuration file, **`configuration.json`**, in the **`3_deploy_to_device`** directory.

Configuration	Meaning	Range	Remarks
app_name	Name of the PPL to deploy	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API ・**`deployment.deployment.Deployment.deploy_device_app`** ・**`deployment.deployment.Deployment.get_device_app_deploys`**

Configuration	Meaning	Range	Remarks
`version_number`	Version of the PPL to deploy	String Details follow the Console Access Library API specification.	Don't abbreviate Used for the following Console Access Library API · **`deployment.deployment.Deployment.deploy_device_app`** · **`deployment.deployment.Deployment.get_device_app_deploys`**
`device_ids`	ID of edge AI device to deploy PPL	List of strings	Don't abbreviate Used for the following Console Access Library API · **`deployment.deployment.Deployment.deploy_device_app`**
`comment`	PPL deployment description	String Details follow the Console Access Library API specification.	Optional Used for the following Console Access Library API · **`deployment.deployment.Deployment.deploy_device_app`**

Run the notebook to deploy PPL

1. Open the notebook, **`deploy_to_device.ipynb`**, in the **`3_deploy_to_device`** directory, and run the python scripts in it
    - The script does the following:
        - Checks that *configuration.json* exists in the **`3_deploy_to_device`** directory
            - If an error occurs, the error description is displayed and running is interrupted.
        - Checks the contents of *configuration.json*
            - If an error occurs, the error description is displayed and running is interrupted.
        - Run the API to deploy PPL
            - If API execution is successful, **`deploy_to_device.ipynb`** displays a successful message
            - If an error occurs, the error description is displayed and running is interrupted.
        - Run the API to get PPL deployment results
            - If results are gotten successfully, **`deploy_to_device.ipynb`** displays a successful message and deployment results
            - If an error occurs, the error description is displayed and running is interrupted.
            - See Cloud SDK Console Access Library(Python) Functional Specifications for details on errors and response times

10. Target performances/Impact on performances

- Usability
  - When the SDK environment is built, AI models and PPL can be deployed from Console for AITRIOS to edge AI devices without any additional installation steps
  - UI response time of 1.2 seconds or less
  - If processing takes more than 5 seconds, then the display during processing can be updated sequentially
  - Provides users with documentation of usage tools and version information

11. Assumption/Restriction

- You can’t cancel deployment or deletion of deployment configuration halfway
- If you cancel and restart notebooks, start each process from the beginning instead of resuming in the middle

12. Remarks

- None

13. Unconfirmed items

- None