P(L)INDER training workshop

plinder
The Protein-Ligand Interaction Dataset and Evaluation Resource

more data
more diversity
more flexible

pinder
The Protein Interaction Dataset and Evaluation Resource

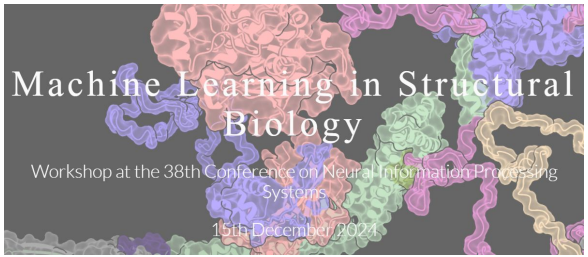
higher quality
less leakage
better evaluation

2024-09-24

Moving Beyond Memorisation

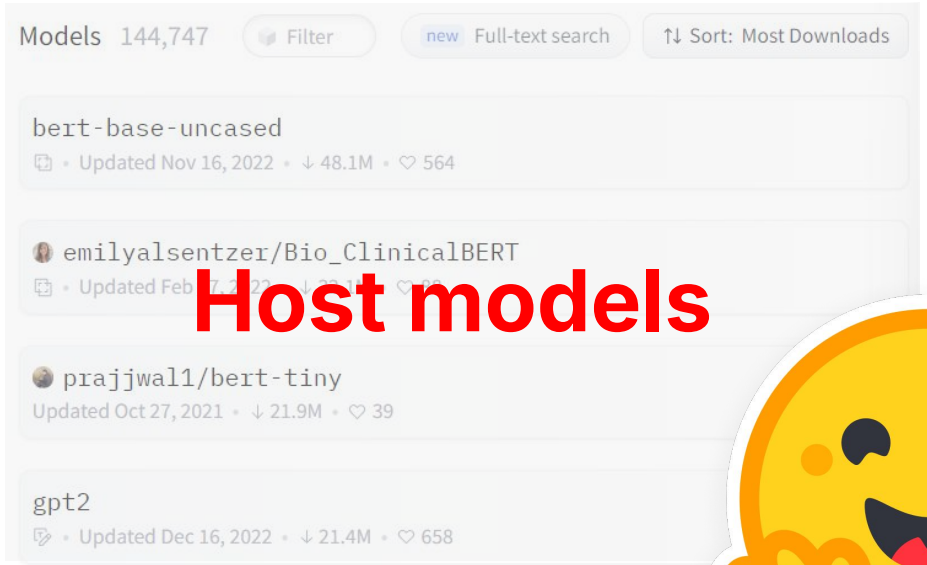
Resources for Generalisable Protein Complex Prediction

Huggingface model and leaderboard submission

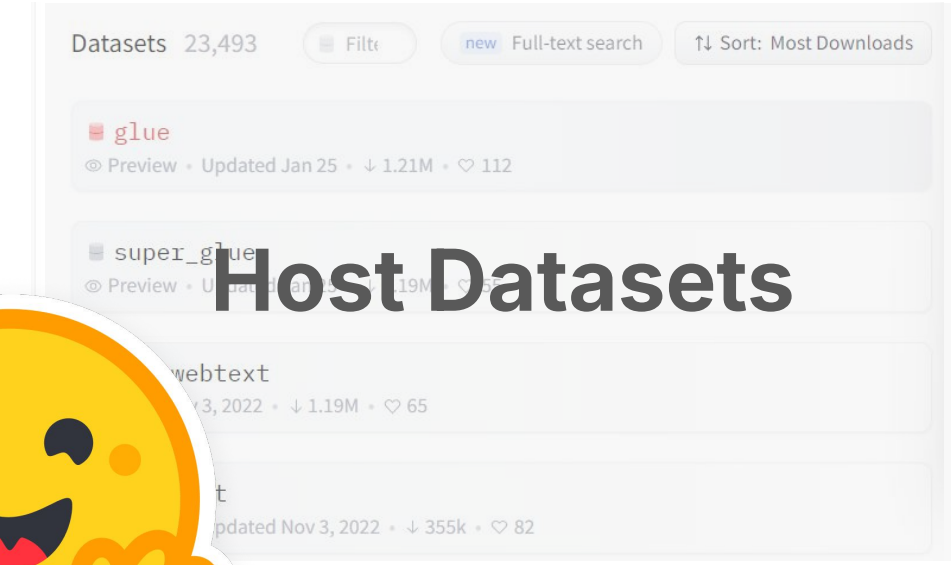


mlsb.io/#challenge

2024-09-24



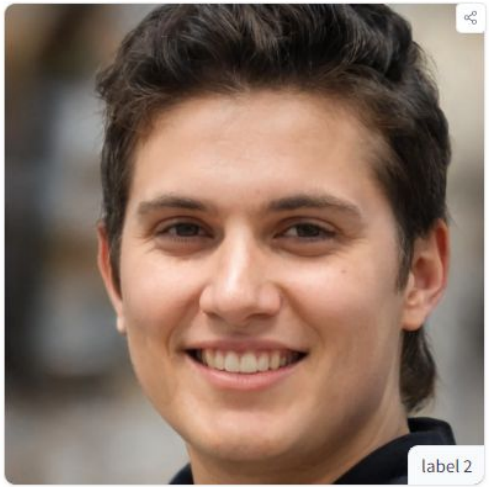
Host models



Host Datasets

Spaces: Host Demos

gradio



Generate images

Fast, easy setup

Gradio can be **installed with pip**. Creating a Gradio interface only requires adding a couple lines of code to your project.

Seamlessly use any python library on your computer. If you can write a python function, gradio can run it.

Present and share

Gradio can be embedded in **Python notebooks** or presented as a **webpage**.

A Gradio interface can automatically generate a public link you can share with colleagues that lets them interact with the model on your computer remotely from their own devices.

Permanent hosting

Once you've created an interface, you can permanently host it on Hugging Face.

Hugging Face Spaces will host the interface on its servers and provide you with a link you can share.

github.com/gradio-app/gradio

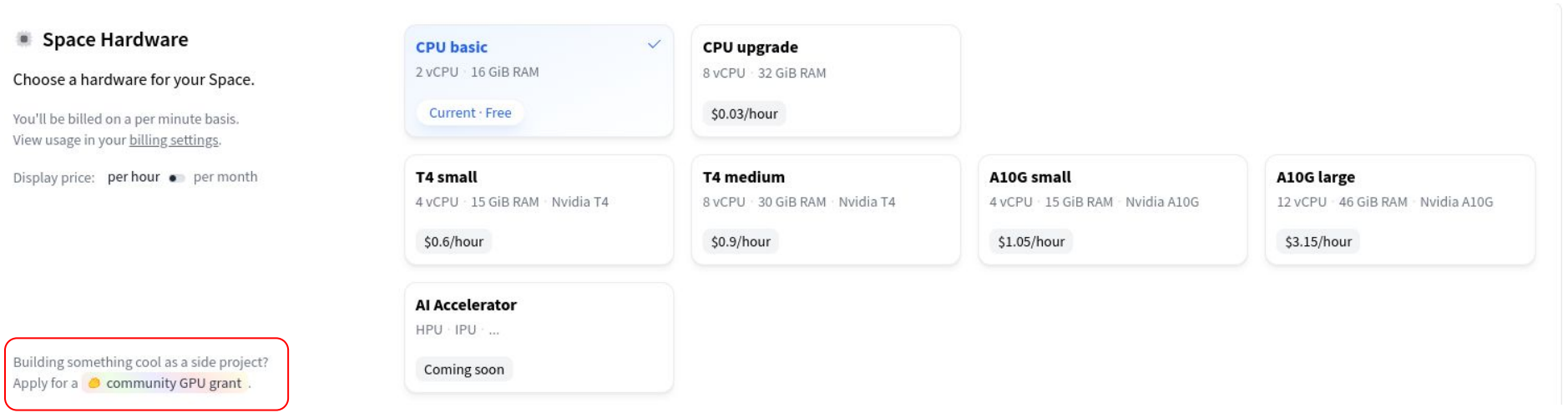
Apache 2.0 license

Develop gradio app locally or on Google Colab

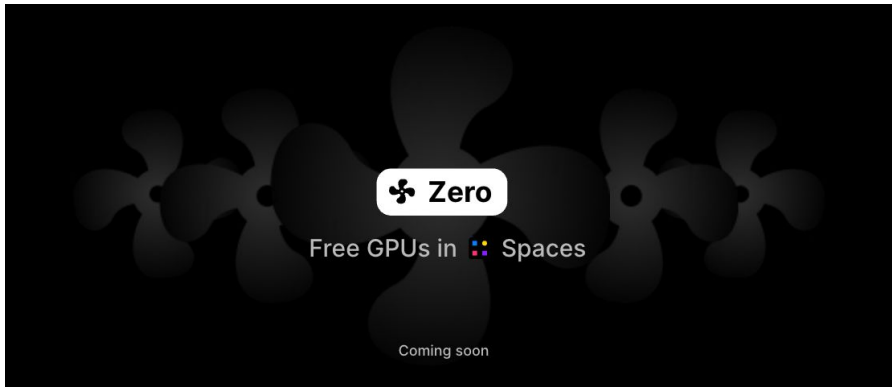
Run on remote machine (workstation, cluster, huggingface) and access UI from your laptop

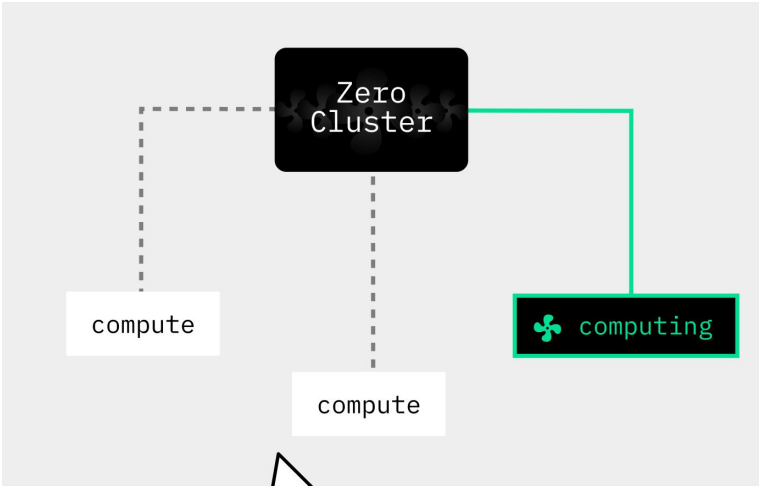
```
127.0.0.1 AND aosdokoe23as.gradio.live
```

Resources on HuggingFace



ZeroGPU





Free A100 in the cloud

https://huggingface.co/zero-gpu-explorers

gradio

Use via API · Built with Gradio

REST api

```python
import requests

response = requests.post("https://simonduerr-proteinmpnnesm.hf.space/run/proteinmpnn",
json={
  "data": [
    "1PGA", // pdb code
    None, // alternative, file as base64
    "A", // designed chain
    "B", // fixed chain
    False, // homomer?
    10, // num of sequences
    "0.1", // sampling temp
    "v_48_002", // model
    "0", // backbone noise
    "", // fixed residues
]}).json()

data = response["data"]
```

Why is a REST API useful?

Example ChimeraX Commands

This prediction of farnesyltransferase subunit alpha takes about 20 seconds and uses the sequence from an experimental structure PDB 7t0a chain A.

```
open 7T0A
esmfold predict /A
```

You can also make the prediction the sequence with no experimental structure.

```
esmfold predict MGSSHHHHHHSQDLMVTSTYIPMSQRRSWADVKPIMQDDGPNPVVPI
MYSEEYKDAMDYFRAIAAKEEKSERALELTEIIVRMNPAHYTVWQYRFSLLTSLNKSLEDEL
RLMNEFAVQNLKSYQVWHHRLLLLDRISPQDPVSEIEYIHGSLLPDPKNYHTWAYLHWLYSH
FSTLGRISEAQWGSELDWCNEMLRVDGRNNSAWGWRWYLRVSRPGAETSSRSLQDELIYILK
SIHLIPHNVSAWNYLRGFLKHFSLPLVPILPAILPYTASKLNPDIETVEAFGFPMPSDPLPE
DTPLPVPLALEYLADSFIEQNRVDDAAKVFEKLSSEYDQMRAGYWEFRRRECAE
```

Inference for the MLSB challenge will be on HF Spaces

inference_app.py

```
def predict:
    #add your code here
```

gradio

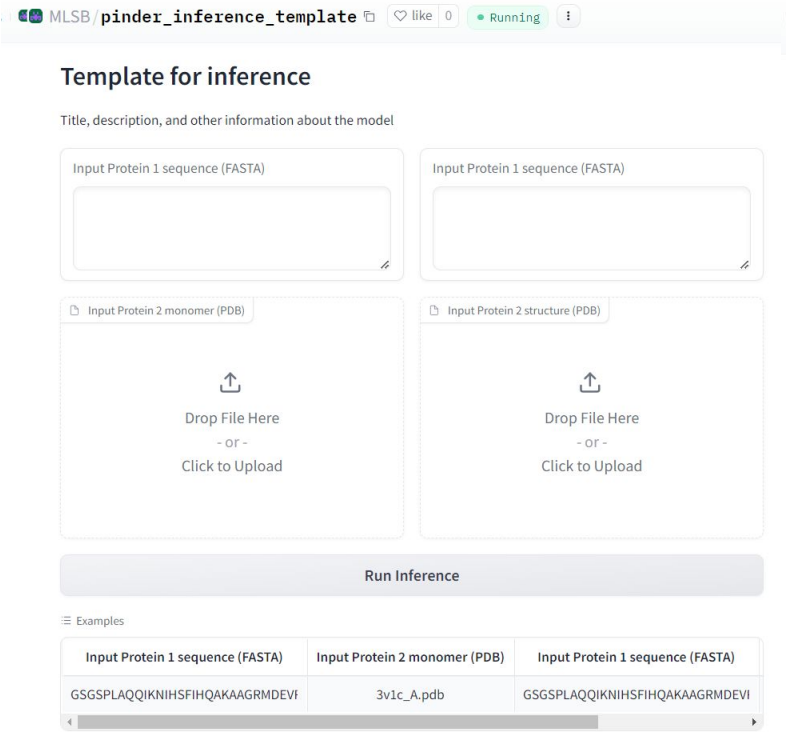
```
import gradio as gr

with gr.Blocks():
    Inp =
    Btn = gr.Button()

   btn.click(predict, inputs, outputs)
```

Provided for you

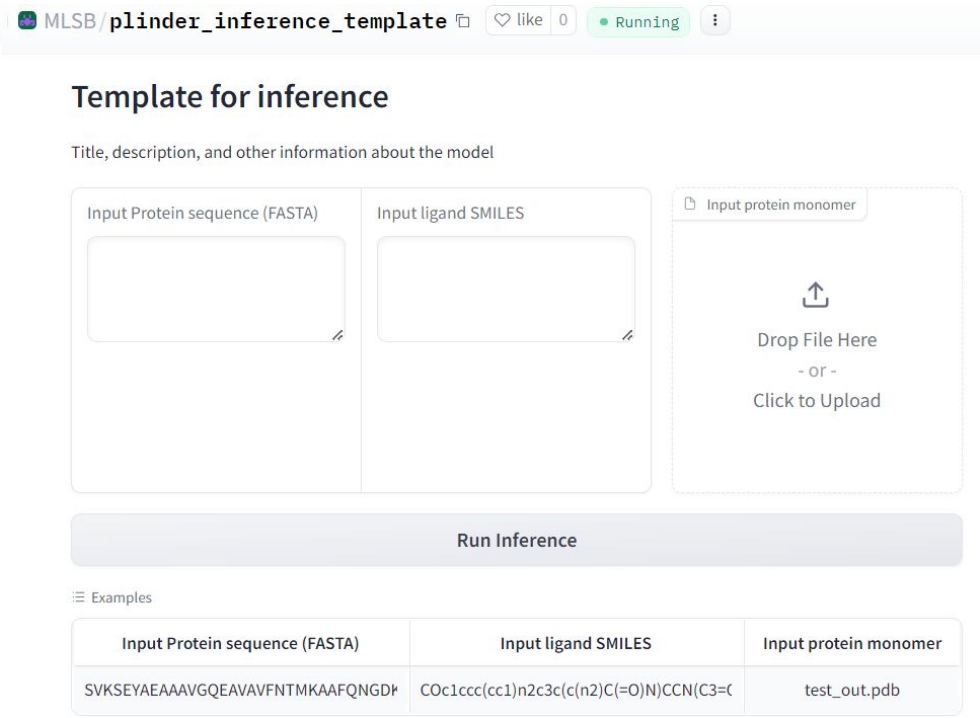
DOCKERFILE

Setup of environment

Inference templates



https://huggingface.co/spaces/MLSB/pinder_inference_template



https://huggingface.co/spaces/MLSB/plinder_inference_template

The Molecule3D component

pip install gradio_molecule3d

```python
import gradio as gr
from gradio_molecule3d import Molecule3D

reps =    [
    {
      "model": 0,
      "style": "stick",
      "color": "whiteCarbon",
      "around": 0,
    }
  ]

def predict(x):
    print("predict function", x)
    print(x.name)
    return x

with gr.Blocks() as demo:
    gr.Markdown("# Molecule3D")
    inp = Molecule3D(label="Molecule3D", reps=reps)
    out = Molecule3D(label="Output", reps=reps)

    btn = gr.Button("Predict")
    btn.click(predict, inputs=inp, outputs=out)


if __name__ == '__main__':
  demo.launch()
```

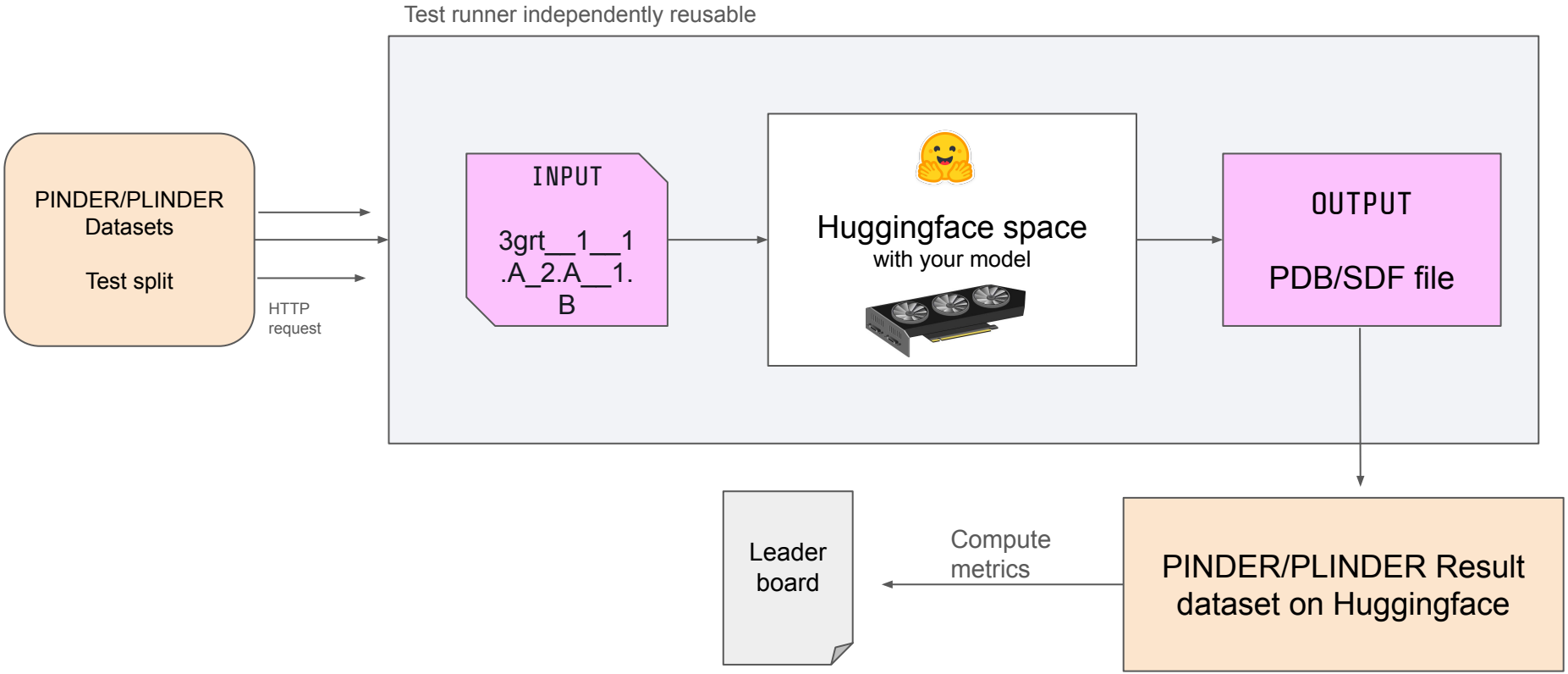
DOCKERFILE

```
FROM continuumio/miniconda3

WORKDIR /usr/src/app
COPY . .
# install dependcies
RUN conda install -y pandas numpy scikit-learn
RUN pip install --no-cache-dir -r requirements.txt
EXPOSE 7860
ENV GRADIO_SERVER_NAME="0.0.0.0"

CMD ["python", "inference_app.py"]
```

Real world example

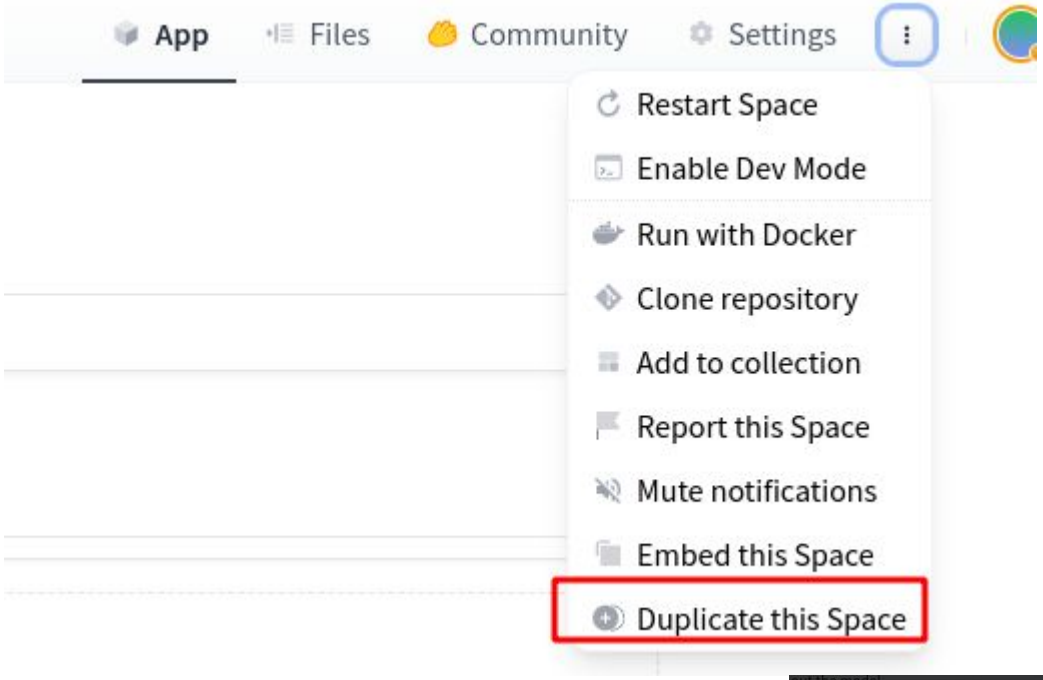
```
FROM continuumio/miniconda3

RUN useradd -m -u 1000 user
WORKDIR /usr/src/app
COPY --link --chown=1000 ./ /usr/src/app
COPY . .
# install dependcies
RUN apt-get update
RUN apt-get -y install gcc
RUN conda install -y -c conda-forge  -c openbabel rdkit  openbabel
RUN pip install mdtraj cython ipython
RUN pip install git+https://github.com/bowman-lab/enspara
RUN pip install biopython
RUN pip install pdb-tools
RUN pip install gradio gradio_molecule3d
USER user
RUN wget https://sourceforge.net/projects/smina/files/smina.static/download -O smina.static
RUN chmod u+x smina.static
# don't change below
EXPOSE 7860
ENV GRADIO_SERVER_NAME="0.0.0.0"

CMD ["python", "inference_app.py"]
```

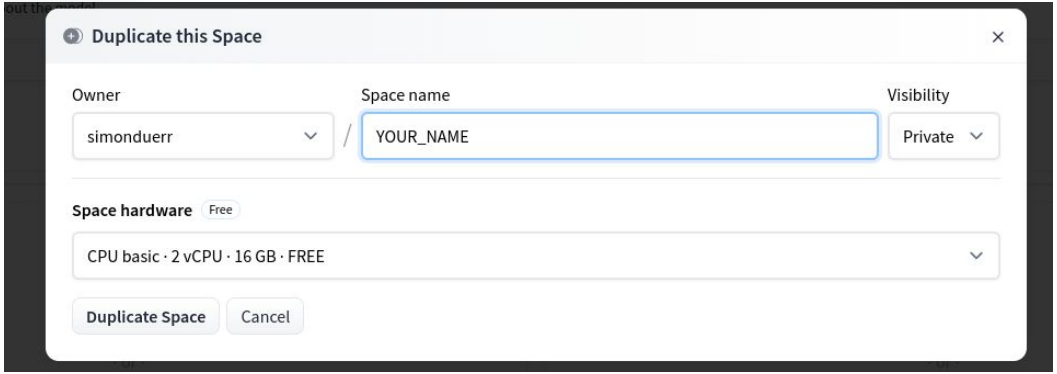
How will we run inference



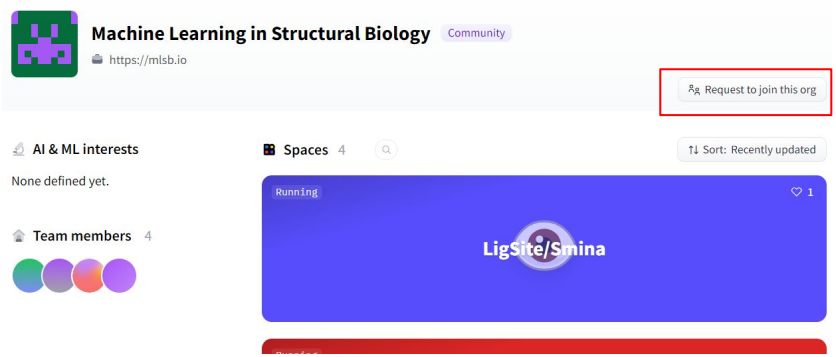
How to submit?

- Duplicate HuggingFace template for challenge in your personal HF account
- Develop and test your space locally and set up DOCKERFILE to match your local environment
- Push your local space to Huggingface and verify it builds correctly (GPUs won’t be assigned yet so inference might be slow)
- Once everything is ready:
  - Transfer space to MLSB organization
  - Fill submission form: https://huggingface.co/spaces/MLSB/leaderboard2024
- We will grant the GPU once we start running inference, your space will remain accessible after the challenge as well


App
Files
Community
Settings
Restart Space
Enable Dev Mode
Run with Docker
Clone repository
Add to collection
Report this Space
Mute notifications
Embed this Space
Duplicate this Space

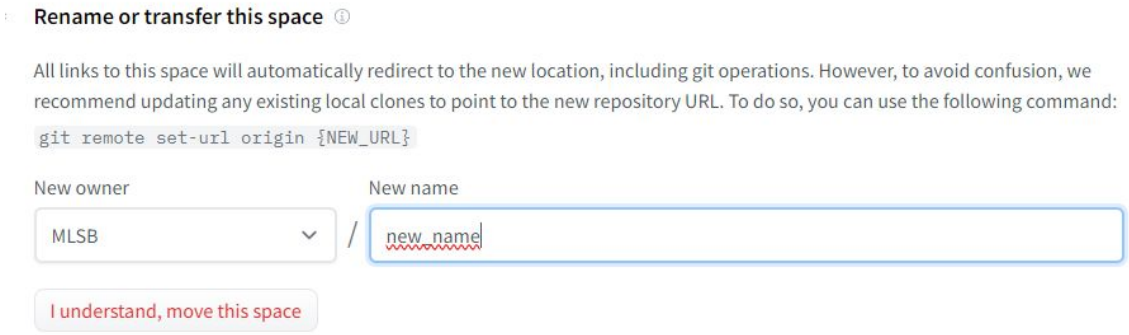

Duplicate this Space
Owner
Space name
Visibility
simonduerr
/
YOUR_NAME
Private
Space hardware
Free
CPU basic · 2 vCPU · 16 GB · FREE
Duplicate Space
Cancel

How to transfer space to MLSB organization



Join MLSB organization

Go to your personal space and transfer it to MLSB



MLSB 2024 Leaderboard

PINDER Leaderboard | PLINDER Leaderboard | Submit

PINDER Leaderboard

Evaluating Protein-Protein interaction prediction

Search

Separate multiple queries with ';'.

Filter Arch

- [x] GNN

Select Columns to Show

- [x] Arch
- [x] Model
- [x] L_rms
- [x] I_rms
- [x] F_nat
- [x] DOCKQ
- [x] CAPRI
- [ ] Runtime
- [ ] Hub License
- [ ] #Params (B)
- [ ] Model sha

Arch	Model	L_rms	I_rms	F_nat	DOCKQ	CAPRI
GNN	example/example	81.22	79.78	91.15	77.95	74.5

Timeline

Leaderboard Opens: October 9th, 2024 (following acceptance notifications for MLSB).
Leaderboard Closes: November 9th, 2024
Winner Notification: Wednesday, November 27th, 2024

Resources to get started

Discord server for questions: https://discord.gg/NFVgccusBj

Templates:

https://huggingface.co/spaces/MLSB/pinder_inference_template

https://huggingface.co/spaces/MLSB/plinder_inference_template