

Moving Beyond Memorisation

Resources for Generalisable Protein Complex Prediction

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

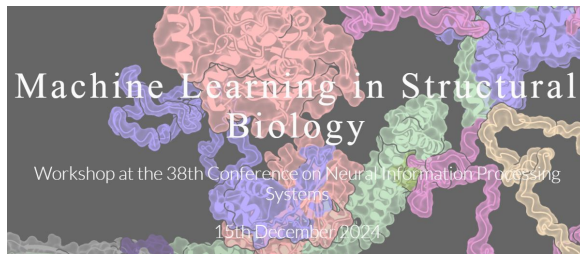
higher quality
less leakage
better evaluation

2024-09-24

Moving Beyond Memorisation

Resources for Generalisable Protein Complex Prediction

Huggingface model
and leaderboard submission



mlsb.io/#challenge

2024-09-24

Models 144,747 Filter new Full-text search Sort: Most Downloads

bert-base-uncased

Updated Nov 16, 2022 • ↓ 48.1M • ♥ 564

emilyalsentzer/Bio_ClinicalBERT

Updated Feb 1, 2022 • ↓ 28.1M • ♥ 22

prajjwal1/bert-tiny

Updated Oct 27, 2021 • ↓ 21.9M • ♥ 39

gpt2

Updated Dec 16, 2022 • ↓ 21.4M • ♥ 658

Host models

Datasets 23,493 Filter new Full-text search Sort: Most Downloads

glue

Preview • Updated Jan 25 • ↓ 1.21M • ♥ 112

super_glue

Preview • Updated Jan 25 • ↓ 1.19M • ♥ 57

webtext

Updated Nov 3, 2022 • ↓ 1.19M • ♥ 65

Host Datasets



protein

new Full-text search

Sort: Recently Updated

Protein Structure Modeling

♥ 4

Protein Family Prediction

♥ 2

Running on T4

♥ 61

ProteinMPNN

myscale

4 days ago

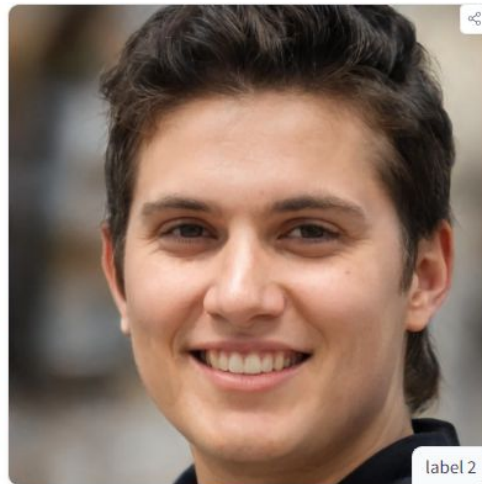
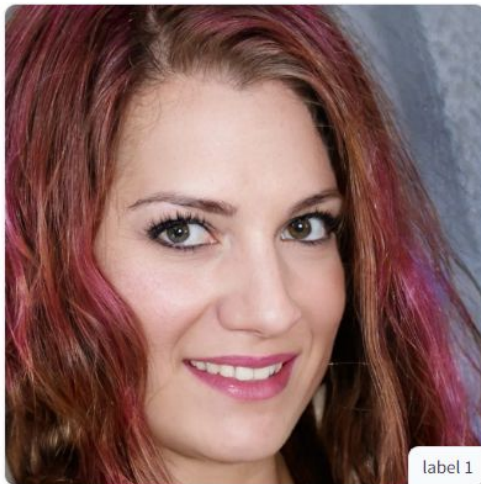
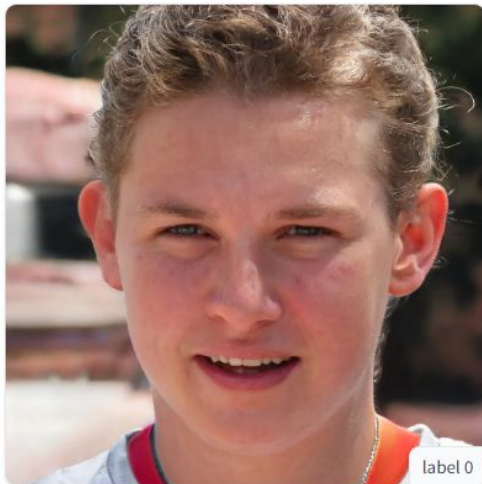
jonathang

15 days ago

simondueerr

21 days ago

Spaces: Host Demos



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

Space Hardware

Choose a hardware for your Space.

You'll be billed on a per minute basis.

View usage in your [billing settings](#).

Display price: per hour ☒ per month

Building something cool as a side project?
Apply for a [community GPU grant](#).

CPU basic

2 vCPU · 16 GiB RAM

Current · Free

CPU upgrade

8 vCPU · 32 GiB RAM

\$0.03/hour

T4 small

4 vCPU · 15 GiB RAM · Nvidia T4

\$0.6/hour

T4 medium

8 vCPU · 30 GiB RAM · Nvidia T4

\$0.9/hour

A10G small

4 vCPU · 15 GiB RAM · Nvidia A10G

\$1.05/hour

A10G large

12 vCPU · 46 GiB RAM · Nvidia A10G

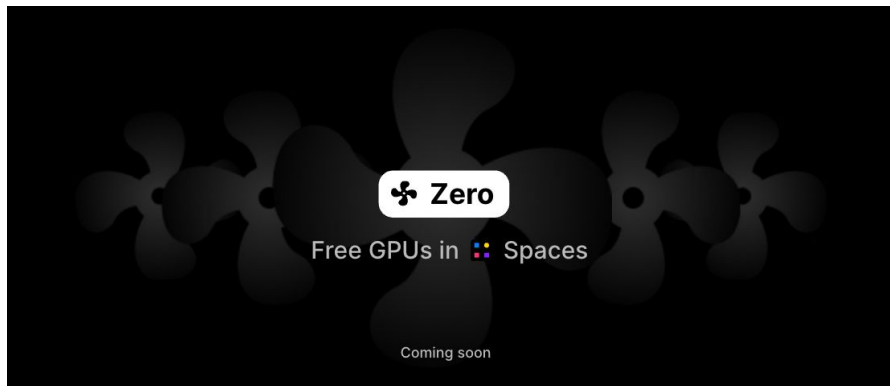
\$3.15/hour

AI Accelerator

HPU · IPU · ...

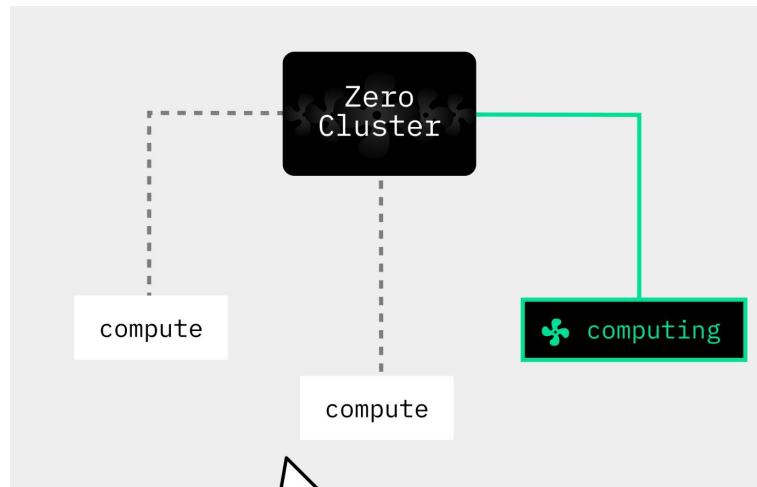
Coming soon

ZeroGPU



Free A100 in the cloud

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



REST api

```
import requests

response = requests.post("https://simondurr-proteinmpnnsm.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 MGSSHHHHHSQDLMVTSTYIPMSQRRSWADV KPI MQDDGPNPVVPI
MYSE EYKDAMDYFRAIAAKEEKSE RALEL TEI I VRMNP AHYTVWQYRF SLLTSLNKSLEDEL
RLMNEFAVQNLKSYQVWHHRLLLLDRISPQDPVSEIEYIHGSLLPDPKNYHTWAYLHWLYSH
FSTLGRISEAQWGSELDWCNEMLRVDGRNNSAWGWRWYLRVSRPGAETSSRSLQDELIYILK
SIHLIPHNVSAWNYLRGFLKHFS LPLVPILPAILPYTASKLNPD IETVEAFGFPMPSDPLPE
DTPLPVPLALEYLADSFIEQNRVDDAAKVFEKLSSEYDQMRAGYWEFRRRECAE
```

Inference for the MLSB challenge will be on HF Spaces

inference_app.py

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



```
import gradio as gr
```

```
with gr.Blocks():
```

```
    Inp = Provided for you
```

```
    Btn = gr.Button()
```

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

DOCKERFILE

Setup of environment



Inference templates

MLSB/pinder_inference_template like 0 Running

Template for inference

Title, description, and other information about the model

Input Protein 1 sequence (FASTA)

Input Protein 1 sequence (FASTA)

Input Protein 2 monomer (PDB)

Drop File Here
- or -
Click to Upload

Input Protein 2 structure (PDB)

Drop File Here
- or -
Click to Upload

Run Inference

Examples

Input Protein 1 sequence (FASTA)	Input Protein 2 monomer (PDB)	Input Protein 1 sequence (FASTA)
GSGSPLAQKIQNIHSFIHQAKAAGRMDVI	3v1c_A.pdb	GSGSPLAQKIQNIHSFIHQAKAAGRMDVI

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

MLSB/plinder_inference_template like 0 Running

Template for inference

Title, description, and other information about the model

Input Protein sequence (FASTA)

Input ligand SMILES

Input protein monomer

Drop File Here
- or -
Click to Upload

Run Inference

Examples

Input Protein sequence (FASTA)	Input ligand SMILES	Input protein monomer
SVKSEYAEAAVVGQEAVAVFNTMKAQFQNGD	COc1ccc(cc1)n2c3c(c(n2)C(=O)N)CCN(C3=C	test_out.pdb

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

The Molecule3D component

pip install
gradio_molecule3d

```
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 dependencies
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"]
```

```
FROM continuumio/miniconda3
```

```
RUN useradd -m -u 1000 user
```

```
WORKDIR /usr/src/app
```

```
COPY --link --chown=1000 ./ /usr/src/app
```

```
COPY . .
```

```
# install dependencies
```

```
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
```

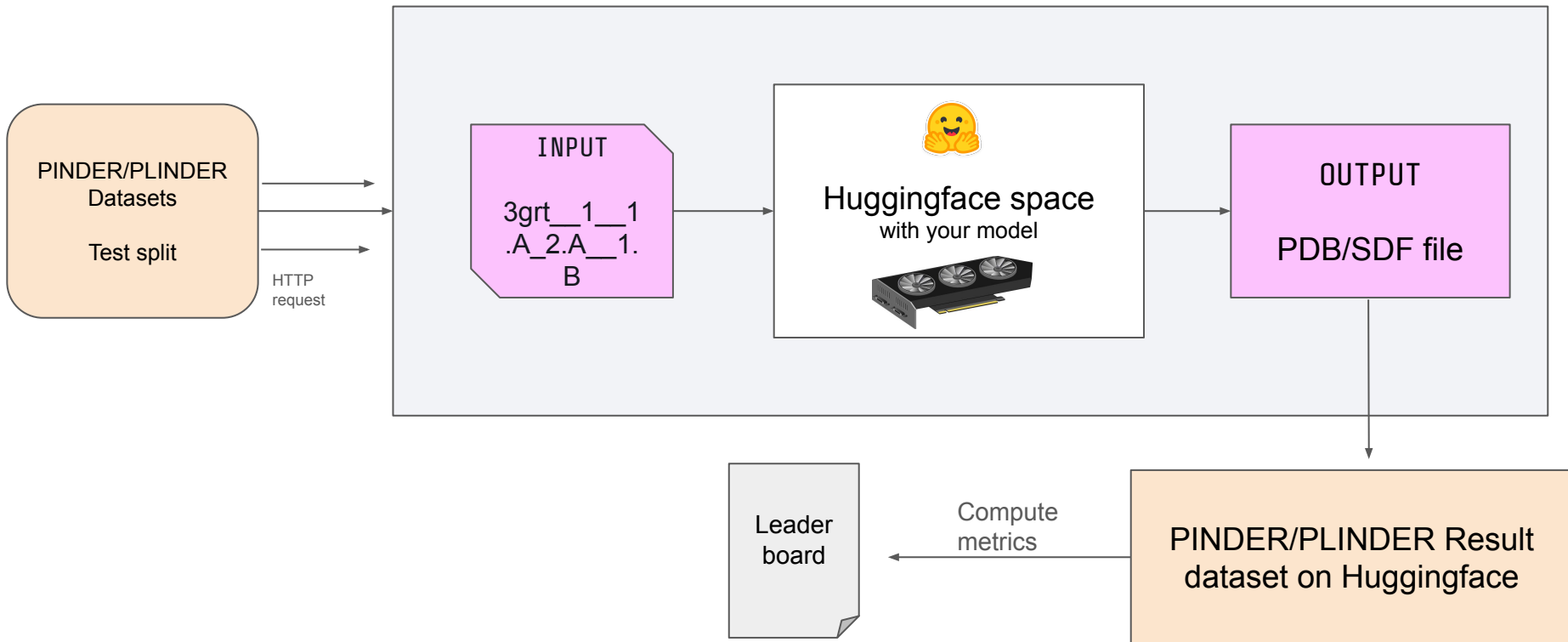
```
ENV GRADIO_SERVER_NAME="0.0.0.0"
```

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

Real world example







How will we run inference

Test runner independently reusable



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

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

2 Duplicate this Space

Owner

simondurr

Space name

YOUR_NAME

Visibility

Private

Space hardware

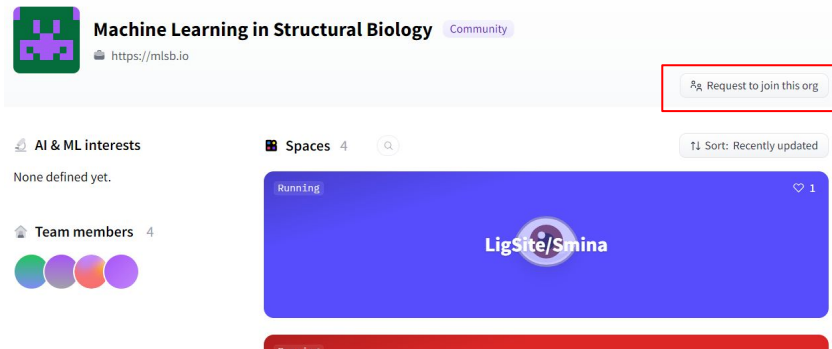
Free

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

Rename or transfer this space ⓘ

All links to this space will automatically redirect to the new location, including git operations. However, to avoid confusion, we recommend updating any existing local clones to point to the new repository URL. To do so, you can use the following command:

```
git remote set-url origin {NEW_URL}
```

New owner

MLSB

New name

new_name

I understand, move this space

MLSB 2024 Leaderboard

 PINDER Leaderboard

 PLINDER Leaderboard

 Submit

PINDER Leaderboard

Evaluating Protein-Protein interaction prediction

Search

Separate multiple queries with ','.

Filter Arch

☒ GNN

Select Columns to Show

☒ Arch

☒ Model

☒ L_rms

☒ I_rms

☒ F_nat

☒ DOCKQ

☒ 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