

Demonstration

Run the program

1. backend: import data

- Show how we can initialize the database by importing the json file
- `demo.json` is the previous `db_with_pswd.json`

```
python database/load_data.py
```

2. frontend: run program

```
streamlit run frontend/app.py
```

1. user types

- i. common user login

ii. common user register and login

New User Registration

Username

Lewis

Password

admin

Confirm Password

.....

Register

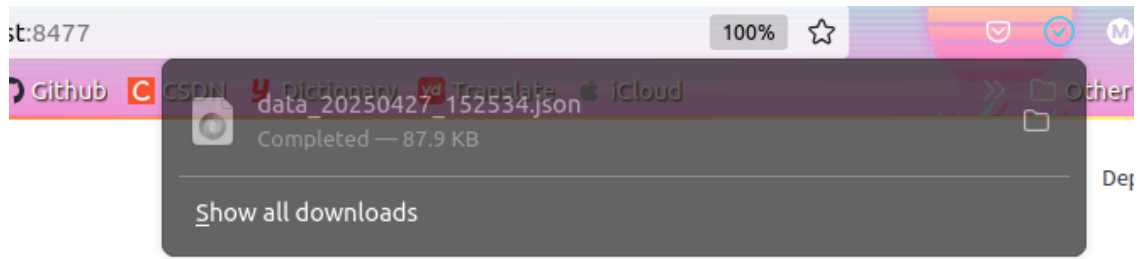
iii. admin login: has some pages that common users don't have

type	user	admin
sidebar	<div><div>Open Model Hub</div><div>Welcome, Lewis!</div><div>Log Out</div><div>Navigation Menu<ul style="list-style-type: none">HomeModel RepositoryDatasets</div></div>	<div><div>Open Model Hub</div><div>Welcome, admin!</div><div>Log Out</div><div>Navigation Menu<ul style="list-style-type: none">HomeModel RepositoryDatasetsUser Managementdata insight</div></div>

2. page types

i. Home

- export and download data



Platform Overview

Total Models	Total Datasets	Registered Users	Today's Downloads
92	100	30	2543

Export All Data

ii. Model Repository

- search with/without specifying the instance

Model Repository

Search Models

Enter natural language query

Select Field

model_id

Enter Query Value

Search

Select Type

models

all

models

datasets

users

Upload New Model

ResNet-50

Architecture: CNN | Media Type: image | Parameters: 23,500,000

View Details

EfficientNet-B0

Architecture: CNN | Media Type: image | Parameters: 5,300,000

View Details

VGG16

Architecture: CNN | Media Type: image | Parameters: 138,357,544

View Details

- LLM assisted search
- upload model

Model Repository

Search Models

Select Type

Enter natural language query

m▼

Select Field

model_id▼

Enter Query Value

Search

Upload New Model

Model Name*

Deepseek-V3

Parameter Count

700000000-+

Architecture Type*?

TRANSFORMER▼

Training Type*?

PRETRAIN▼

Media Type*?

TEXT▼

Task Types*

CLASSIFICATION x

GENERATION x

SEGMENTATION x

Select Model File*

Drag and drop file here

Limit 200MB per file • PT, PTH, CKPT, TXT

Browse files

checkpoint-0000001.txt

294.0B

x

Submit

-
- view details; paging; **view details: show 2 tables**

Open Model Hub

Welcome, admin!

Log Out

Navigation Menu

- Home
- Model Repository
- Datasets
- User Management
- data insight

Architecture: RNN | Media Type: text | Parameters: 28,000,000

SimpleRNN-TextGen-5

Architecture: RNN | Media Type: text | Parameters: 30,000,000

View Details

2/10 ← →

Model Details - SimpleRNN-TextGen-5

Basic Information:

	Property	Value
0	Model ID	20
1	Architecture Type	RNN
2	Parameter Count	30,000,000
3	Training Type	Trainname:FINETUNE

Related Information:

	Type	Name
0	Supported Tasks	generation
1	Authors	Eve, Grace, Katherine
2	Related Datasets	yelp_review_full

- iii. Datasets
- iv. (Admin Privilege) User Management

- create user

User Management

Search Users

Select Type

Enter natural language query

u: ▼

Select Field

Enter Query Value

user_id ▼

Search


+ Add New User

^

Username*

Rain

Password*

..... 

☒ Admin Privileges

Organization

The Chinese University of Hong Kong, Shenzhen

Create User

User ID	Username	Organization	Admin Status
1	Bradley	The Chinese University of Hong Kong, Shenzhen	✓
2	Rosalind	The Chinese University of Hong Kong, Shenzhen	✓
3	Alice	Tsinghua University	✗
4	Bob	Tsinghua University	✗
5	Charlie	Peking University	✗
6	David	Peking University	✗
7	Eve	Fudan University	✗
8	Frank	Fudan University	✗
9	Grace	Shanghai Jiaotong University	✗
10	Hannah	Shanghai Jiaotong University	✗
			✗

Admin features

Select Field

user_id



Enter Query Value

Search

- edit user

v. data insight

- (presenter: Linyong Gan)

LLM assisted search

1. search according to architecture: transformer models

- result in the table: ...

Model Repository

Search Models

find all transformer models

Select Type

models

Select Field

model_id

Enter Query Value

Search

📄

🔍

⌵

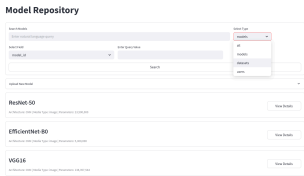
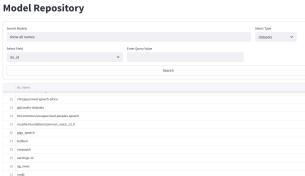
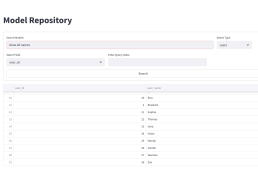
	model_id	model_name	param_num	media_type	arch_name	trainname	param
0	41	BERT	110000000	TEXT	TRANSFORMER	Trainname.FINETUNE	0,0,0,0,0,0,0,0,0
1	42	RoBERTa	125000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
2	43	DistilBERT	66000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
3	44	XLNet	117000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
4	45	ALBERT	120000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
5	46	GPT-2	117000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
6	47	GPT-3	175000000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
7	48	T5	220000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
8	49	BART	139000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0
9	50	DeBERTa	140000000	TEXT	TRANSFORMER	Trainname.PRETRAIN	0,0,0,0,0,0,0,0,0

- sql query is also available:

Query Details

```
▼ { 
  "Natural Language Query" : "find all transformer models"
  "Generated SQL" : "SELECT * FROM model
                     WHERE arch_name = 'Transformer';"
  "Error Code" : 0
  "Has Results" : true
  "Error Message" : NULL
  ▼ "Query Results" : [ 
    ▼ 0 : {
      "model_id" : 41
      "model_name" : "BERT"
      "param_num" : 110000000
      "media_type" : "TEXT"
      "arch_name" : "TRANSFORMER"
      "trainname" : "Trainname.FINETUNE"
      "param" : "b'\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00'"
    }
    ▼ 1 : { 
      "model_id" : 42
      "model_name" : "RoBERTa" 
      "param_num" : 125000000
      "media_type" : "TEXT"
      "arch_name" : "TRANSFORMER"
      "trainname" : "Trainname.PRETRAIN"
      "param" : "b'\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00'"
    }
    ▼ 2 : {
```

2. compare classified and not classified

instance type	model	dataset	user
query: show all names			

3. more complicated search: ranking

- query: top 10 users with the most published datasets

Model Repository

Search Models

top 10 users with the most published datasets

Select Type

u: ▼

Select Field

user_id ▼

Enter Query Value

Search

	user_name	dataset_count
0	Grace	16
1	Marselo	15
2	Parvage	14
3	Ivy	13
4	Eve	13
5	Wendy	12
6	Jack	12
7	Bob	12
8	Yasmine	11
9	Rosalind	11

Query Details

```
▼ {
  "Natural Language Query" :
  "top 10 users with the most published datasets"
  "Generated SQL" :
  "SELECT u.user_name, COUNT(ud.ds_id) AS dataset_count
  FROM user u
  JOIN user_ds ud ON u.user_id = ud.user_id
  GROUP BY u.user_id
  ORDER BY dataset_count DESC"
```

```
LIMIT 10;"
```

```
"Error Code" : 0
```

todo list

urgent

- ☒ translate all code
- ☐ "LLM refined schema"
- ☐ powerpoint
- ☐ code demo

not urgent

- ☐ report + each persons' part
- ☐ report zip, pdf