

Beyond Generic Models and Datasets: Using MLAssetSelection to Build Robust ML-Enabled Systems

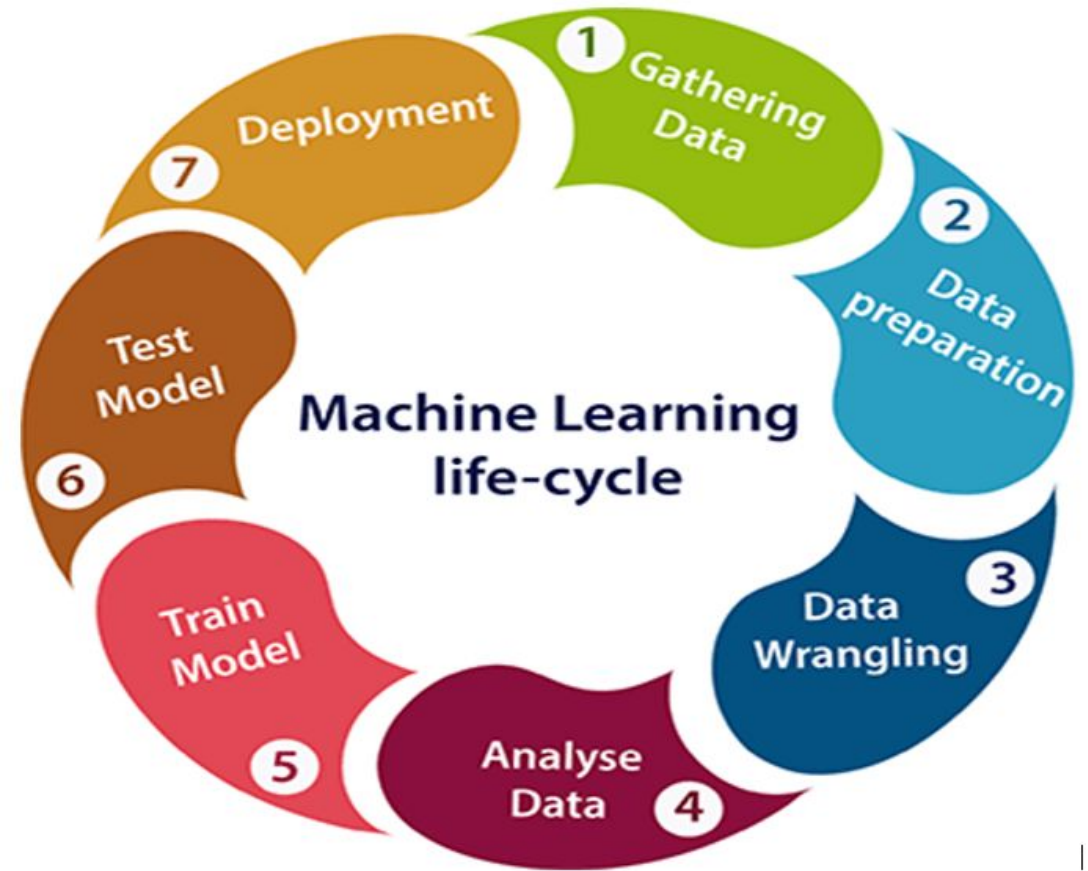
Alexandra González

Universitat Politècnica de Catalunya – BarcelonaTech (UPC)

Beyond the Model Meeting (BTMM) | July 10th, 2026

Introduction

- Machine Learning (ML) assets: high-value artifacts contributing to ML-driven workflows such as **models** and **datasets**.



Motivation

- **Model Reuse:** Similar to software reuse, with ML-specific challenges.
- **Developer Mandates:** Choosing the right ML asset is not only a matter of performance. Developers must also:
 - Ensure system compatibility
 - Avoid integration problems
 - Manage privacy risks, bias, and ethical concerns.
- **Discovery Challenge:** Systematically discovering SE assets remains challenging, making the process time-consuming and error-prone.

Hugging Face (HF)

- Popular open-source registry for Machine Learning (ML) assets
- One new repository is created every **15 seconds!**
- **2024** average:
 - **2,199** new models/day
 - **+6M** downloads/day on HF

+2.8M

models
(as of July 2026)



Hugging Face (HF)

Current situation:

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation
- Zero-Shot Object Detection
- Text-to-3D
- Image-to-3D
- Image Feature Extraction
- Keypoint Detection
- Video-to-Video

Natural Language Processing

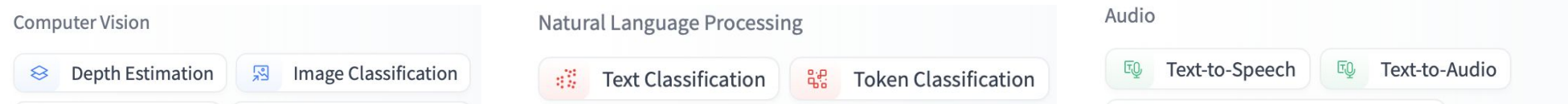
- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Feature Extraction
- Text Generation
- Text2Text Generation
- Fill-Mask
- Sentence Similarity
- Text Ranking

Audio

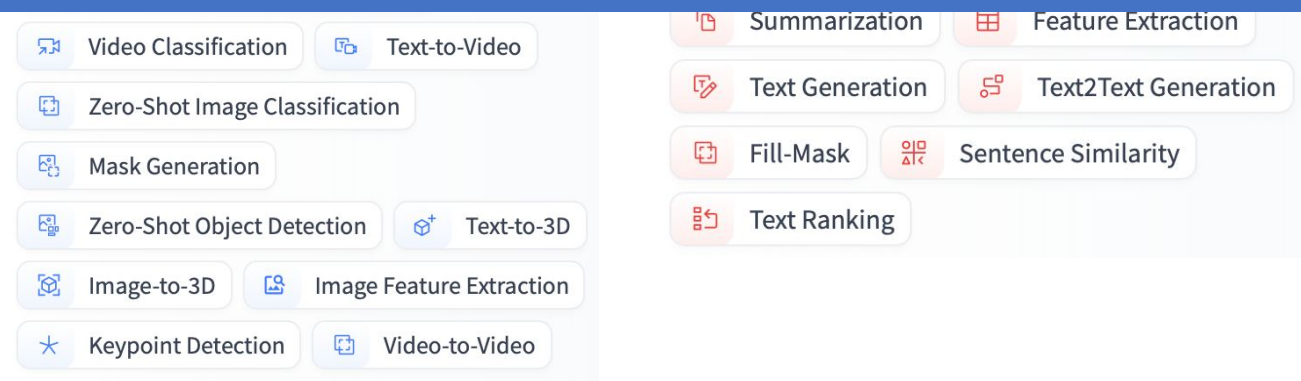
- Text-to-Speech
- Text-to-Audio
- Automatic Speech Recognition
- Audio-to-Audio
- Audio Classification
- Voice Activity Detection

The Gap: Lack of SE Focus

Current situation:



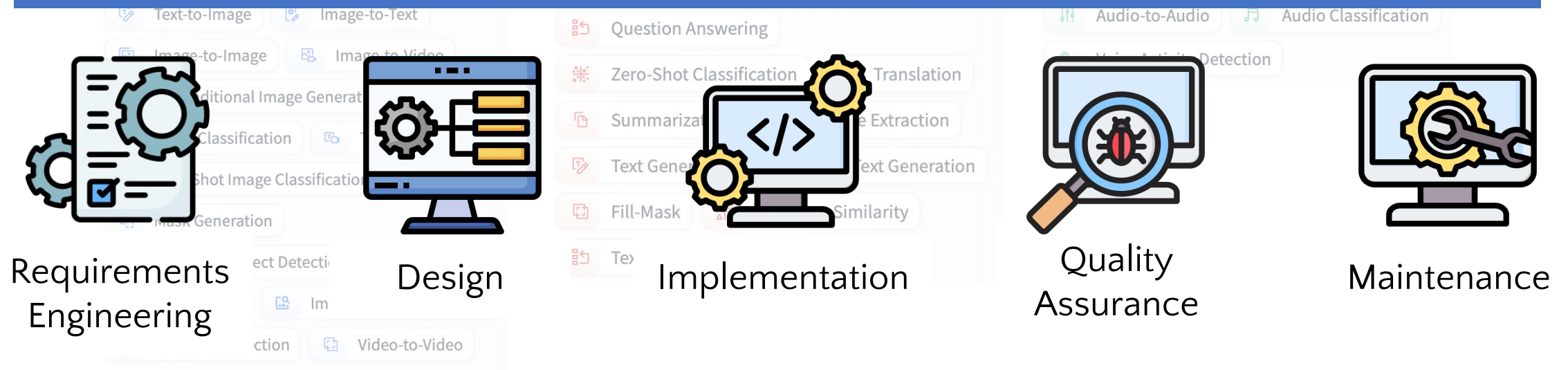
Are there Software Engineering (SE) models?



The Gap: Lack of SE Focus

Current situation:

Are there Software Engineering (SE) models?



Related Work – ML Registries

HFCommunity

- Periodic relational database dumps of HF assets.
- No updates since October 2024.



Related Work – ML Registries

HFCommunity

- Periodic relational database dumps of HF assets.
- No updates since October 2024.



Linked Papers with Code

ML research as an RDF knowledge graph, linking tasks, datasets, methods, and evaluations from PWC.



LPWC

Related Work – ML Registries

HFCommunity

- Periodic relational database dumps of HF assets.
- No updates since October 2024.



Linked Papers with Code

ML research as an RDF knowledge graph, linking tasks, datasets, methods, and evaluations from PWC.



LPWC

However, they are designed for ML purposes: **no SE organization!**

Related Work – Repository Mining

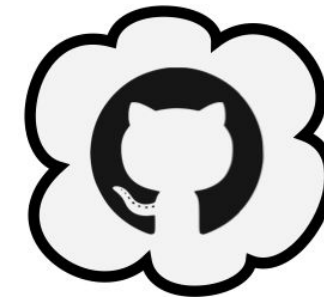
SEART Data Hub

Web application to build and pre-process datasets of code mined from public GitHub repositories.



GitHub Search

Web & dataset with 25 characteristics (e.g., commits, license) for over 735k GitHub repositories



They are not designed for SE researchers and practitioners that reuse ML assets!

Related Work – Leaderboards

🏆 EvalPlus Leaderboard 🏆

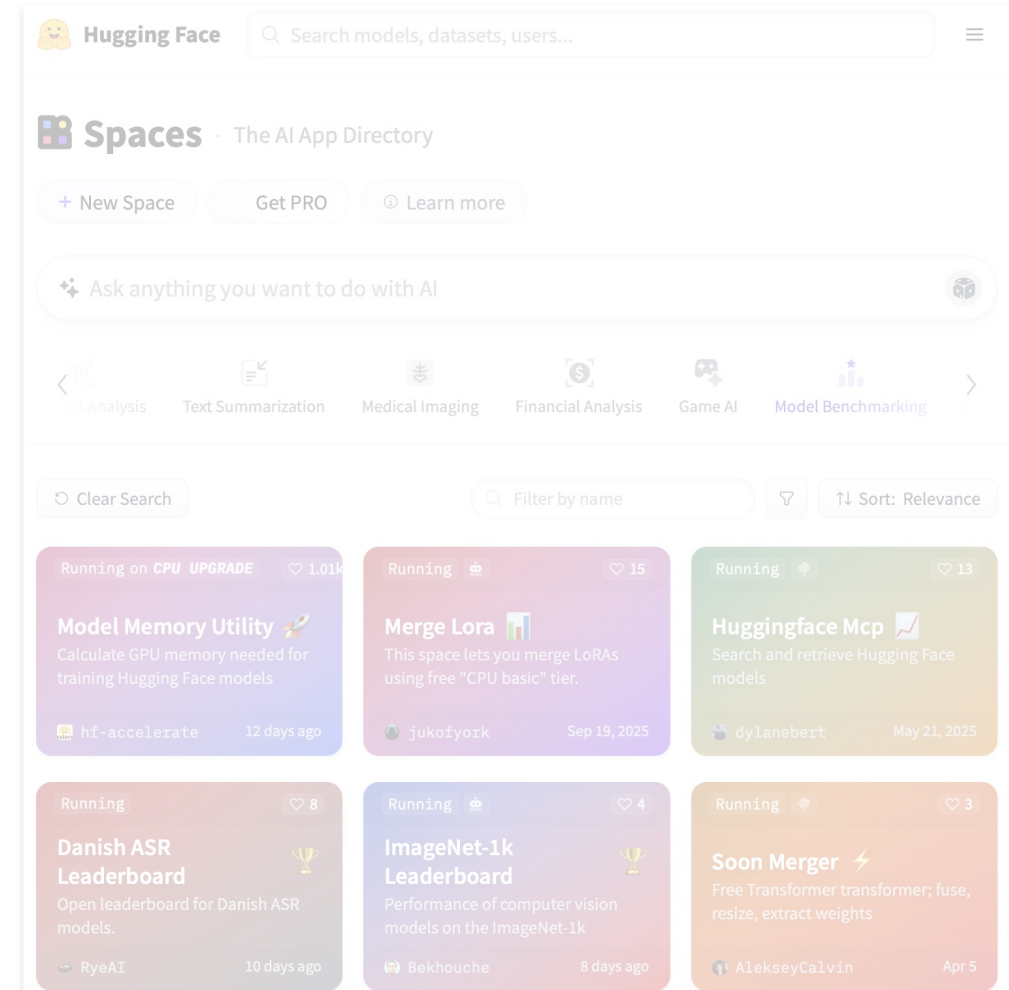
EvalPlus evaluates AI Coders with rigorous tests.

📢 News: Beyond correctness, how's their code efficiency? Checkout [EvalPerf!](#)

OSQ Benchmark (Evaluating LLMs with OSQs and MCQs)

[Website](#) | [GitHub](#) | [Paper](#) | [X1](#) | [X2](#)

Open-LLM-Leaderboard, for evaluating large language models (LLMs) by transitioning from multiple-choice questions (MCQs) to open-style questions.



Related Work – Leaderboards



EvalPlus Leaderboard

EvalPlus evaluates AI Coders with rigorous tests.

News: Beyond correctness, how's their code efficiency? Checkout [EvalPerf!](#)

OSQ Benchmark (Evaluating LLMs with OSQs and MCQs)

[Website](#) |
 [GitHub](#) |
 [Paper](#) |
 [X1](#) |
 [X2](#)

Open-LLM-Leaderboard, for evaluating large language models (LLMs) by transitioning from multiple-choice questions (MCQs) to open-style questions.



SWE-bench


Hugging Face



Spaces

The AI App Directory

[+ New Space](#)
[Get PRO](#)
[Learn more](#)

Ask anything you want to do with AI

[Text Analysis](#)
[Text Summarization](#)
[Medical Imaging](#)
[Financial Analysis](#)
[Game AI](#)
[Model Benchmarking](#)

Running on CPU **UPGRADE**  1.01k

Model Memory Utility

Calculate GPU memory needed for training Hugging Face models

hf-accelerate 12 days ago

Running  15

Merge Lora

This space lets you merge LoRAs using free "CPU basic" tier.

jukofyork Sep 19, 2025

Running  13

Huggingface Mcp

Search and retrieve Hugging Face models

dylanebert May 21, 2025

Running  8

Danish ASR Leaderboard

Open leaderboard for Danish ASR models.

RyeAI 10 days ago

Running  4

ImageNet-1k Leaderboard

Performance of computer vision models on the ImageNet-1k

Bekhouche 8 days ago

Running  3

Soon Merger

Free Transformer transformer; fuse, resize, extract weights

AlekseyCalvin Apr 5

Related Work – Leaderboards

🏆 EvalPlus Leaderboard 🏆

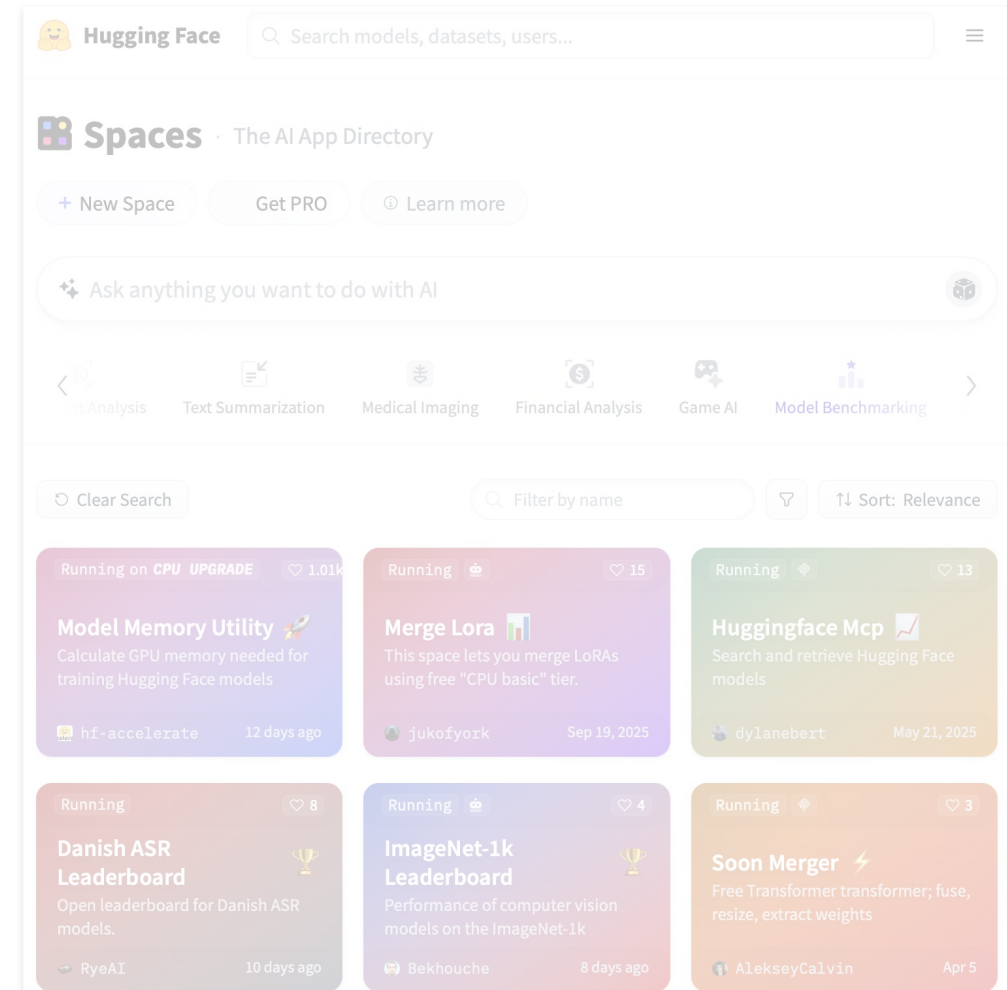
EvalPlus evaluates AI Coders with rigorous tests.

📢 News: Beyond correctness, how's their code efficiency? Checkout [EvalPerf!](#)

OSQ Benchmark (Evaluating LLMs with OSQs and MCQs)

[Website](#) | [GitHub](#) | [Paper](#) | [X1](#) | [X2](#)

Open-LLM-Leaderboard, for evaluating large language models (LLMs) by transitioning from multiple-choice questions (MCQs) to open-style questions.



Related Work – Leaderboards

🏆 EvalPlus Leaderboard 🏆

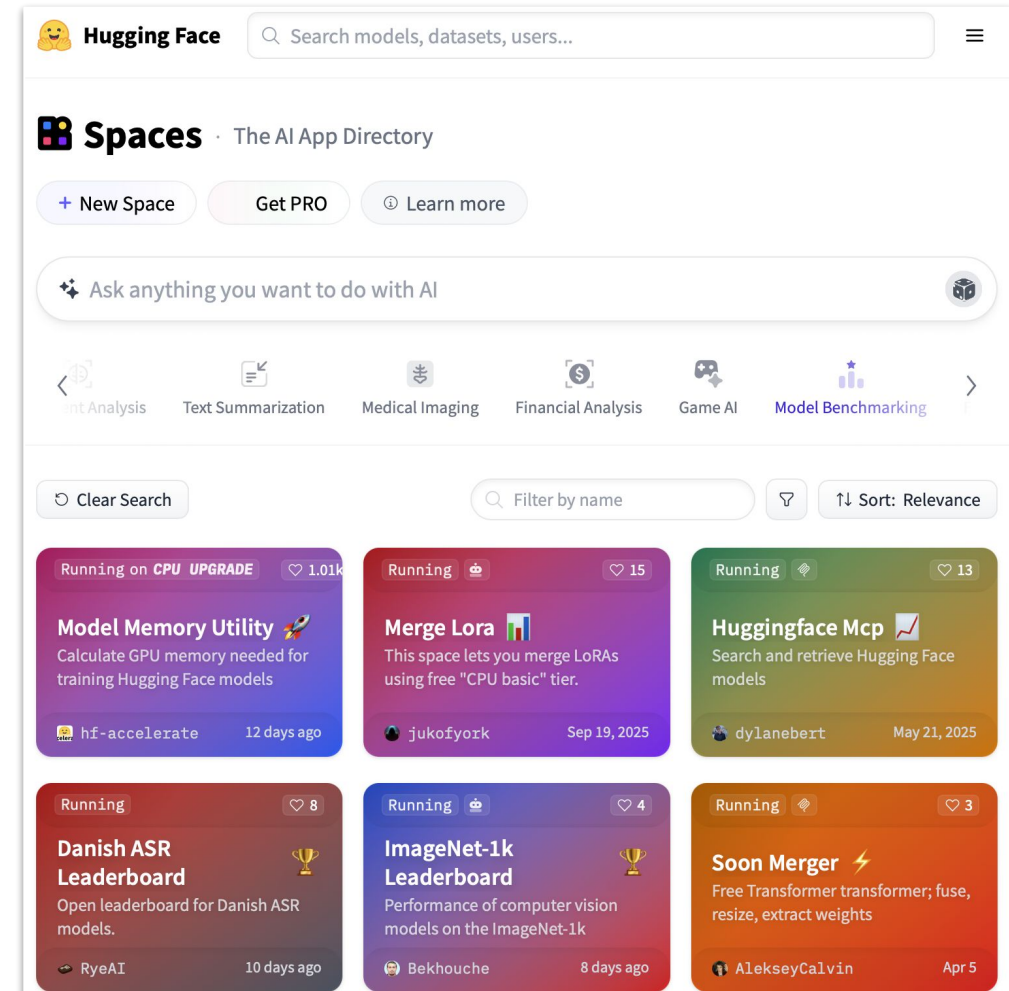
EvalPlus evaluates AI Coders with rigorous tests.

📢 News: Beyond correctness, how's their code efficiency? Checkout [EvalPerf!](#)

OSQ Benchmark (Evaluating LLMs with OSQs and MCQs)

[Website](#) | [GitHub](#) | [Paper](#) | [X1](#) | [X2](#)

Open-LLM-Leaderboard, for evaluating large language models (LLMs) by transitioning from multiple-choice questions (MCQs) to open-style questions.





MLAssetSelection

A comprehensive platform for discovering, evaluating and selecting the best Machine Learning assets for Software Engineering projects

 Explore Models

 View Leaderboard

 Browse Datasets

Hugging Face vs. MLAssetSelection



Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Hugging Face vs. MLAssetSelection



Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Hugging Face vs. MLAssetSelection



Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Hugging Face vs. MLAssetSelection



Tool



Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Hugging Face vs. MLAssetSelection



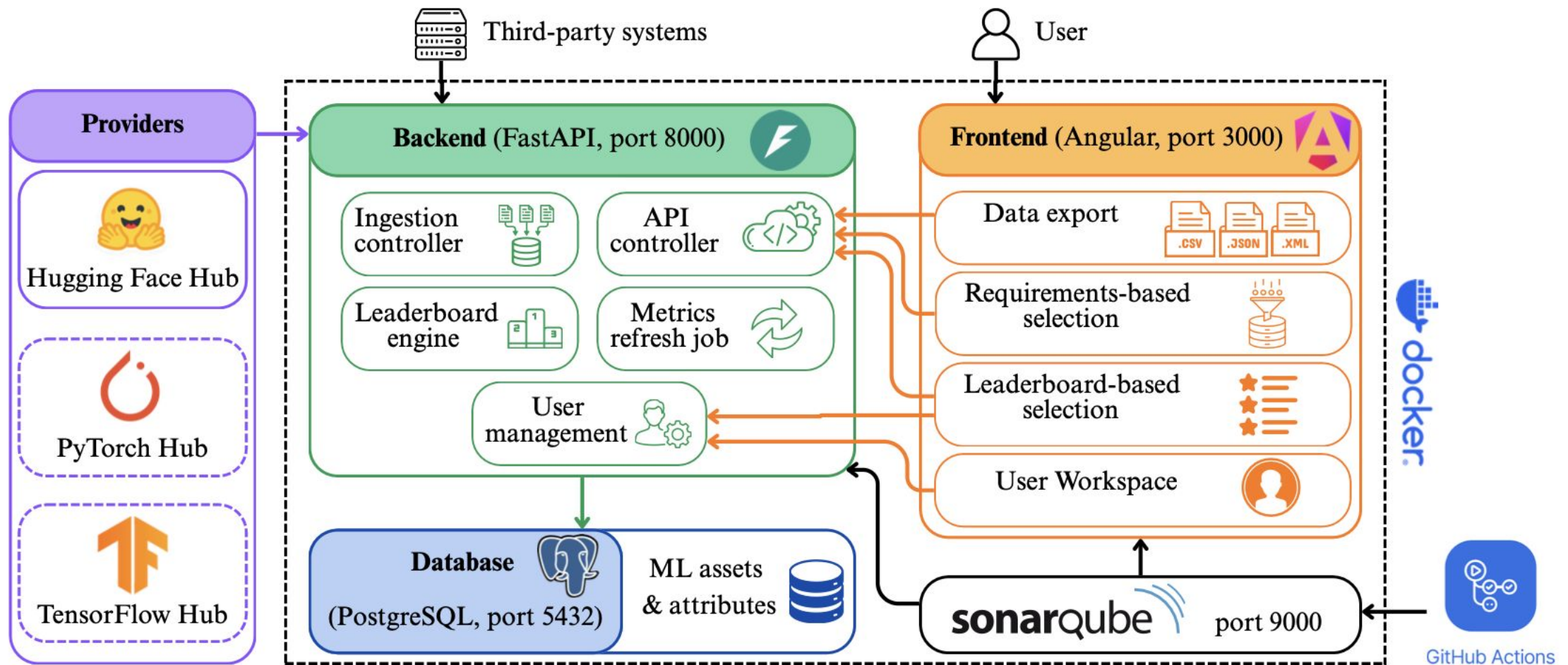
Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Hugging Face vs. MLAssetSelection



Filters	Hugging Face	MLAssetSelection
SE task	✗ Not supported	Explicit SE task filters
Metadata	Licenses, libraries, languages, ML tasks	Licenses, libraries, languages, ML tasks, creation date
Popularity (likes, downloads)	✗ No filtering Sort by most only	Filterable by range Sortable ascending/descending
Activity (commits, contributors)	✗ Not supported	Filterable by range Sortable ascending/descending
Model-specific	Inference providers, number of parameters, apps	Size, region, training datasets, inference providers, evaluation results
Dataset-specific	Size, format, modalities	Size, format, modalities, disciplines

Tool Architecture



MLAssetSelection – Frontend



Leaderboard

1. **Benchmark** (e.g., HumanEval)
2. **Implementation** (e.g., Explain)
3. **Programming Language** (e.g., C++)
4. **Metric** (e.g., Pass@1)
5. **Metric configuration** (e.g., threshold 0.2)

Requirements-based Selection

- Asset identifier
- SE tasks
- Model-specific attributes
- Dataset-specific attributes
- Metadata
- Popularity
- Activity

Data Export

CSV
JSON
XML

User Workspace

- Save ML assets into customizable lists
- Define preferences (e.g., benchmarks or selection criteria)
- Receive alerts when newly ingested models match those preferences

MLAssetSelection – Frontend



Leaderboard

1. **Benchmark** (e.g., HumanEval)
2. **Implementation** (e.g., Explain)
3. **Programming Language** (e.g., C++)
4. **Metric** (e.g., Pass@1)
5. **Metric configuration** (e.g., threshold 0.2)

Requirements-based Selection

- Asset identifier
- SE tasks
- Model-specific attributes
- Dataset-specific attributes
- Metadata
- Popularity
- Activity

Data Export

CSV
JSON
XML

User Workspace

- Save ML assets into customizable lists
- Define preferences (e.g., benchmarks or selection criteria)
- Receive alerts when newly ingested models match those preferences

MLAssetSelection – Frontend



Leaderboard

1. **Benchmark** (e.g., HumanEval)
2. **Implementation** (e.g., Explain)
3. **Programming Language** (e.g., C++)
4. **Metric** (e.g., Pass@1)
5. **Metric configuration** (e.g., threshold 0.2)

Requirements-based Selection

- Asset identifier
- SE tasks
- Model-specific attributes
- Dataset-specific attributes
- Metadata
- Popularity
- Activity

Data Export

CSV
JSON
XML

User Workspace

- Save ML assets into customizable lists
- Define preferences (e.g., benchmarks or selection criteria)
- Receive alerts when newly ingested models match those preferences

MLAssetSelection – Frontend



Leaderboard

1. **Benchmark** (e.g., HumanEval)
2. **Implementation** (e.g., Explain)
3. **Programming Language** (e.g., C++)
4. **Metric** (e.g., Pass@1)
5. **Metric configuration** (e.g., threshold 0.2)

Requirements-based Selection

- Asset identifier
- SE tasks
- Model-specific attributes
- Dataset-specific attributes
- Metadata
- Popularity
- Activity

Data Export

CSV
JSON
XML

User Workspace

- Save ML assets into customizable lists
- Define preferences (e.g., benchmarks or selection criteria)
- Receive alerts when newly ingested models match those preferences

MLOAssetSelection – Backend



Ingestion Controller

Ingests assets from HF by applying the cataloguing pipeline to models and datasets.

API Controller

- Search, filter, and retrieve ML assets
- 3rd party API access

Metrics Refresh

- Daily ingestion of newly uploaded assets
- 12-hour updates of dynamic metrics

Leaderboard Engine

- Extracts & processes model performance
- Generates leaderboards from model cards

User Management

- Authentication & profile management
- 2FA email verification for enhanced security

MLOAssetSelection – Backend



Tool

Ingestion Controller

Ingests assets from HF by applying the cataloguing pipeline to models and datasets.

API Controller

- Search, filter, and retrieve ML assets
- 3rd party API access

Metrics Refresh

- Daily ingestion of newly uploaded assets
- 12-hour updates of dynamic metrics

Leaderboard Engine

- Extracts & processes model performance
- Generates leaderboards from model cards

User Management

- Authentication & profile management
- 2FA email verification for enhanced security

MLAssetSelection – Backend



Tool

Ingestion Controller

Ingests assets from HF by applying the cataloguing pipeline to models and datasets.

API Controller

- Search, filter, and retrieve ML assets
- 3rd party API access

Metrics Refresh

- Daily ingestion of newly uploaded assets
- 12-hour updates of dynamic metrics

Leaderboard Engine

- Extracts & processes model performance
- Generates leaderboards from model cards

User Management

- Authentication & profile management
- 2FA email verification for enhanced security

MLOAssetSelection – Backend



Tool

Ingestion Controller

Ingests assets from HF by applying the cataloguing pipeline to models and datasets.

API Controller

- Search, filter, and retrieve ML assets
- 3rd party API access

Metrics Refresh

- Daily ingestion of newly uploaded assets
- 12-hour updates of dynamic metrics

Leaderboard Engine

- Extracts & processes model performance
- Generates leaderboards from model cards

User Management

- Authentication & profile management
- 2FA email verification for enhanced security

MLOAssetSelection – Backend



Tool

Ingestion Controller

Ingests assets from HF by applying the cataloguing pipeline to models and datasets.

API Controller

- Search, filter, and retrieve ML assets
- 3rd party API access

Metrics Refresh

- Daily ingestion of newly uploaded assets
- 12-hour updates of dynamic metrics

Leaderboard Engine

- Extracts & processes model performance
- Generates leaderboards from model cards

User Management

- Authentication & profile management
- 2FA email verification for enhanced security

Demo



Tool



MLAssetSelection Home Models Leaderboard **Datasets** Docs About Log In Sign Up

Datasets

Arrakark / scaled_guidelines

Created: 12/7/25, 11:06 PM • 14 • 1 • 5 commits • 1 contributors Software Engineering Activities and Tasks >

Formats: csv
Natural Languages: English
Libraries: Croissant, Datasets, polars, pandas
Licenses: gpl-3.0
Modalities: Text
Other: digital, design, verifaciton, systemverilog, verilog
Regions: US
Size Categories: 1K - 10K

Asma / UStAI

Created: 8/18/25, 4:59 AM • 13 • 0 • 8 commits • 1 contributors Software Engineering Activities and Tasks >

MLAssetSelection Home Models Leaderboard **Datasets** Docs About Log In Sign Up

Models

QuantFactory /

gemma-2b-aps-it-GGUF

Created: 10/17/24, 10:05 AM • 193 • 1 • 22,17 GB • 16 commits • 1 contributors Software Engineering Activities and Tasks >

Libraries: Transformers, GGUF
Licenses: gemma
Other: Inference Endpoints
ML Tasks: Text Generation
Regions: US
Size Categories: 10GB - 100GB

RichardErkhov /

<https://mlassetselection.essi.upc.edu/>

Planned Evaluation



We plan to *analyze* the MLAssetSelection tool *for the purpose of* assessment *with respect to* ease of use and usefulness *from the point of* software engineers and researchers *in the context of* ML assets applied to SE.

Short term

- Focus groups with SE practitioners and researchers
- TAM-based evaluation (usefulness, ease of use, adoption)

Long term

- Controlled experiment vs. direct provider interaction
- Measure efficiency, selection quality, and user satisfaction

MLOps Course

@UPC

Learning Objectives

1. Interpret the basic concepts of **Software Engineering for ML systems**, especially in relation to the use and exploitation of **MLOps** practices.
2. Apply and analyze **good software engineering practices related to data science and machine learning projects**.
3. Apply and analyze **MLOps practices to build ML models**, fostering reproducibility and quality assurance.
4. Apply and analyze **MLOps practices to deploy ML models**, fostering API development.

Organization

- Students will build **teams of approximately 4 members**
- During the semester, each team will be responsible for taking part in **the process of building an ML component of an ML-based system**, discussing the data analysis, documenting and presenting their results:

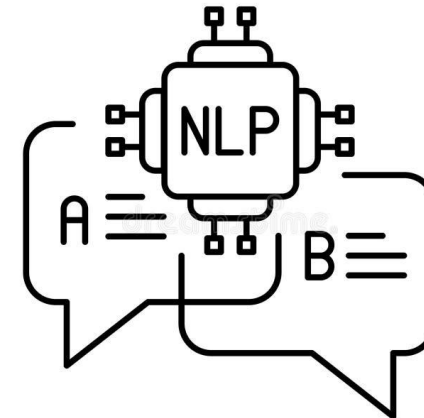
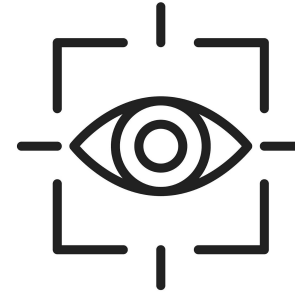
Datasets suggestions (current)

- **Computer vision** datasets:

- [German Traffic Sign Recognition Benchmark \(GTSRB\)](#)
- [MNIST](#)
- [Cifar-10](#)

- **Natural language processing** datasets:

- [Sentiment140](#)
- [Amazon Reviews](#)
- [IMDB Movie Reviews](#)



Datasets suggestions (future)



Tool



MLAssetSelection

Home Models Leaderboard **Datasets** Docs About

Log In Sign Up

Datasets

Search

Arrakark / scaled_guidelines

Created: 12/7/25, 11:06 PM • 14 • 1 • 5 commits • 1 contributors **Software Engineering Activities and Tasks**

Formats: csv

Natural Languages: English

Libraries: Croissant, Datasets, polars, pandas

Licenses: gpl-3.0

Modalities: Text

Other: digital, design, verifcaiton, systemverilog, verilog

Regions: US

Size Categories: 1K - 10K

Asma / UStAI

Created: 8/18/25, 4:59 AM • 13 • 0 • 8 commits • 1 contributors **Software Engineering Activities and Tasks**

Conclusions & Future Work



Conclusions

- MLAssetSelection: a web for **selecting ML assets across the SDLC**.
- It combines **SE filtering** with a **standardized leaderboard**.
- Has already attracted **registered users** from research & practice.

Future Work

- Assets from additional open-source providers and marketplaces
- Incorporating model derivation knowledge

Thank you!

- Questions? Comments?
- Contact: alexandra.gonzalez.alvarez@upc.edu



Access the tool!

This work was supported by the FI-STEP predoctoral fellowship (grant no. 2025 STEP 00407) awarded by AGAUR (Government of Catalonia) and co-funded by the European Social Fund Plus (FSE+) of Catalonia 2021–2027, and by Grant PID2024-156019OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

The authors thank **Antoni Roca** for initial development activities.



**Generalitat
de Catalunya**