KAIST DeepAuto.ai

30th ANNIVERSARY EMNLP 2025
Suzhou, China | 中国苏州
November 4-9 | 11月4日–9日

MONAQ 👑

Multi-Objective Neural Architecture Querying for Time-Series Analysis on Resource-Constrained Devices

Patara Trirat[1] Jae-Gil Lee[2]

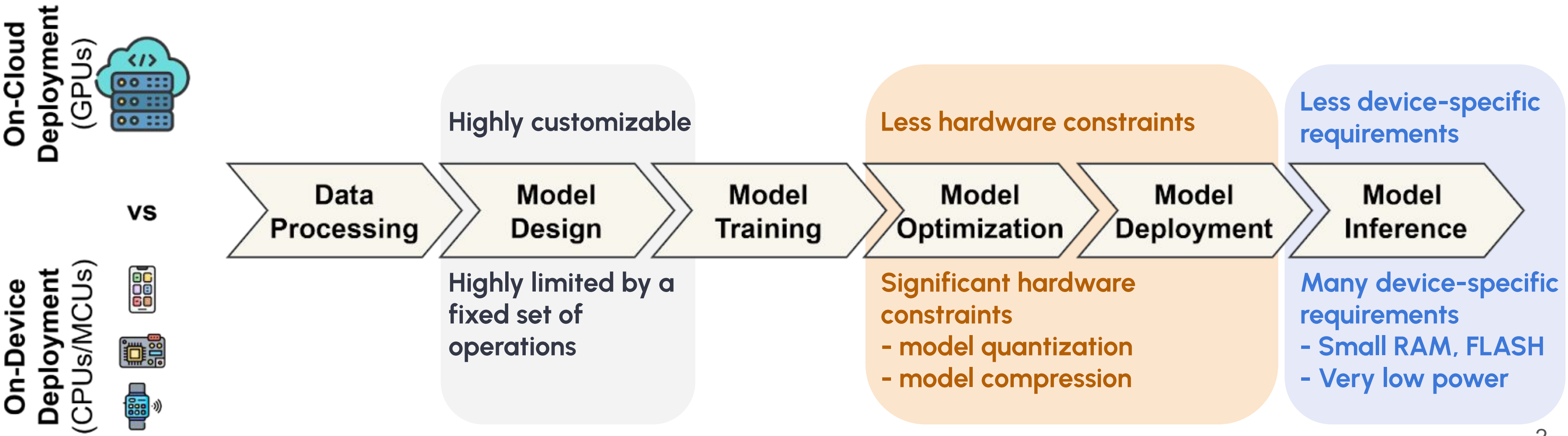
[1]DeepAuto.ai [2]KAIST

https://github.com/kaist-dmlab/MONAQ

2992-Findings

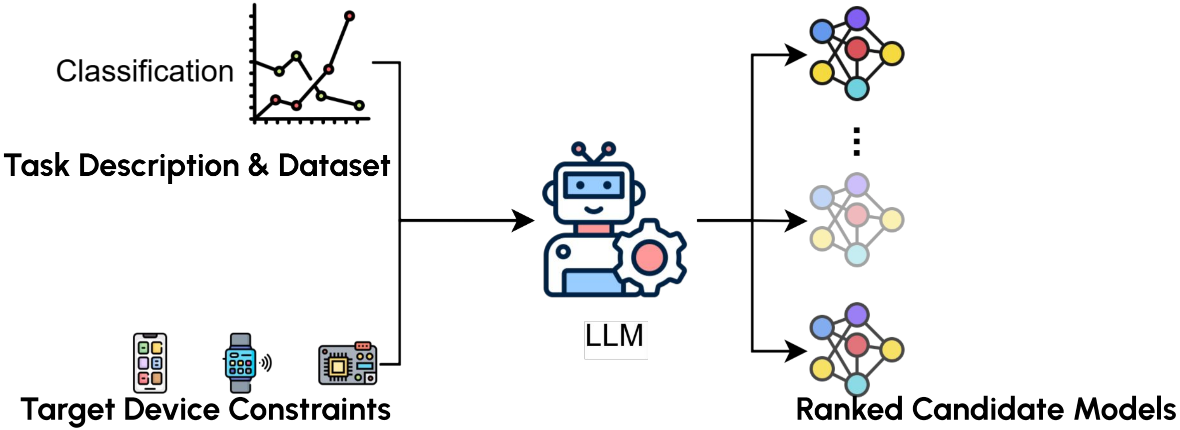
Motivation — Time-Series Analysis on Edge Devices

- Smartphones and IoT devices that are ubiquitous nowadays provide new potential applications in the resource-constrained environments, such as edge intelligence.
- Efficient models runnable on resource-constrained devices often require a more complicated design process than traditional models run on GPUs.



LLMs as Multi-Objective Optimizers

- LLMs have extensive pre-trained knowledge relevant to neural architecture design.
- LLMs can identify relationships between neural architectures, model complexity, and the resource constraints of specific devices.
- LLMs have been shown to estimate the potential performance of neural networks for a given task.



Related Work

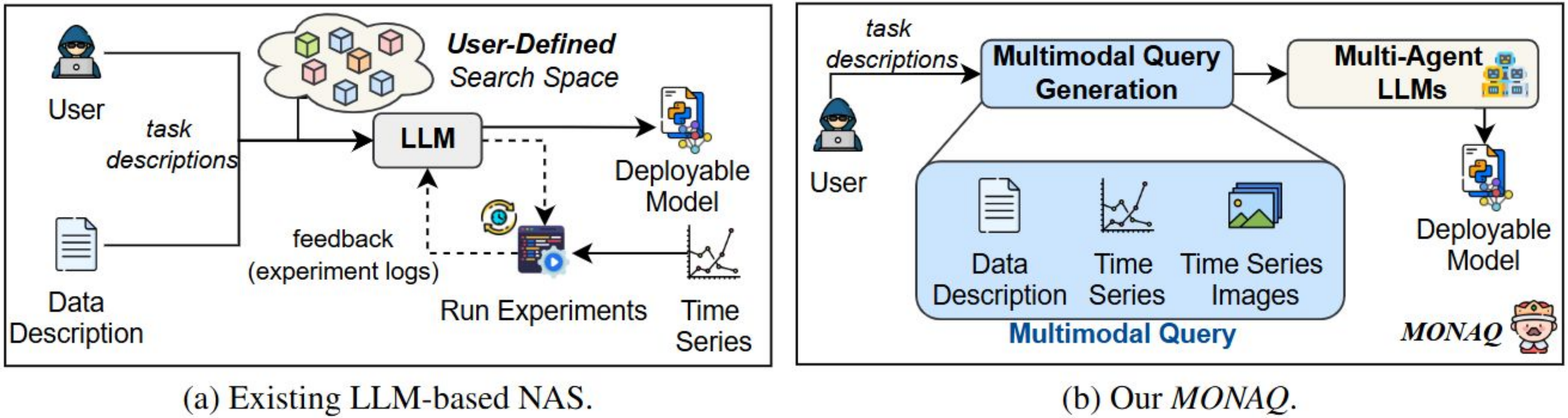


Figure 1: Comparison between (a) existing LLM-based NAS and (b) our proposed *MONAQ* framework.

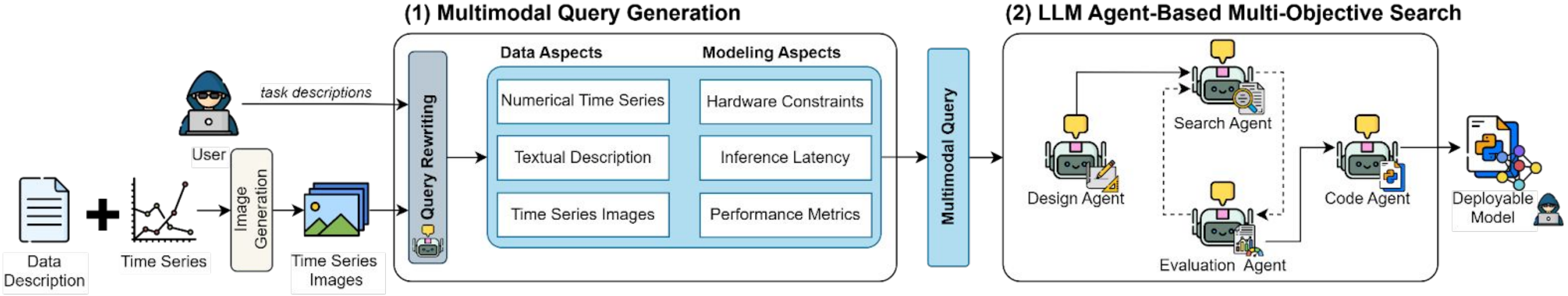
Challenges in LLMs with Time Series

***RQ1**. How to find high-performing architectures without user-defined search spaces?*

***RQ2**. How to make LLM agents accurately understand time-series data and user requirements?*

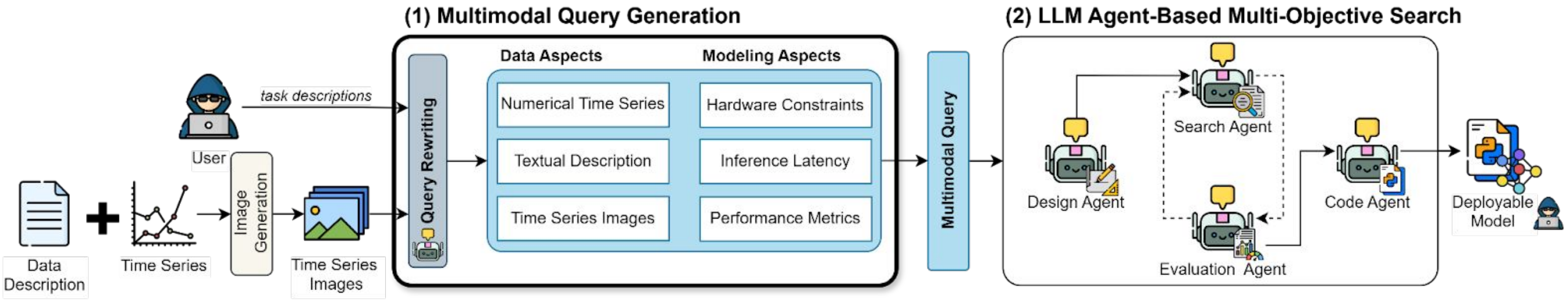
MONAQ👑: Multimodal Neural Architecture Querying

We propose a novel LLM-based multi-agent NAS framework with an open-ended search space for time-series analysis on resource-constrained devices.



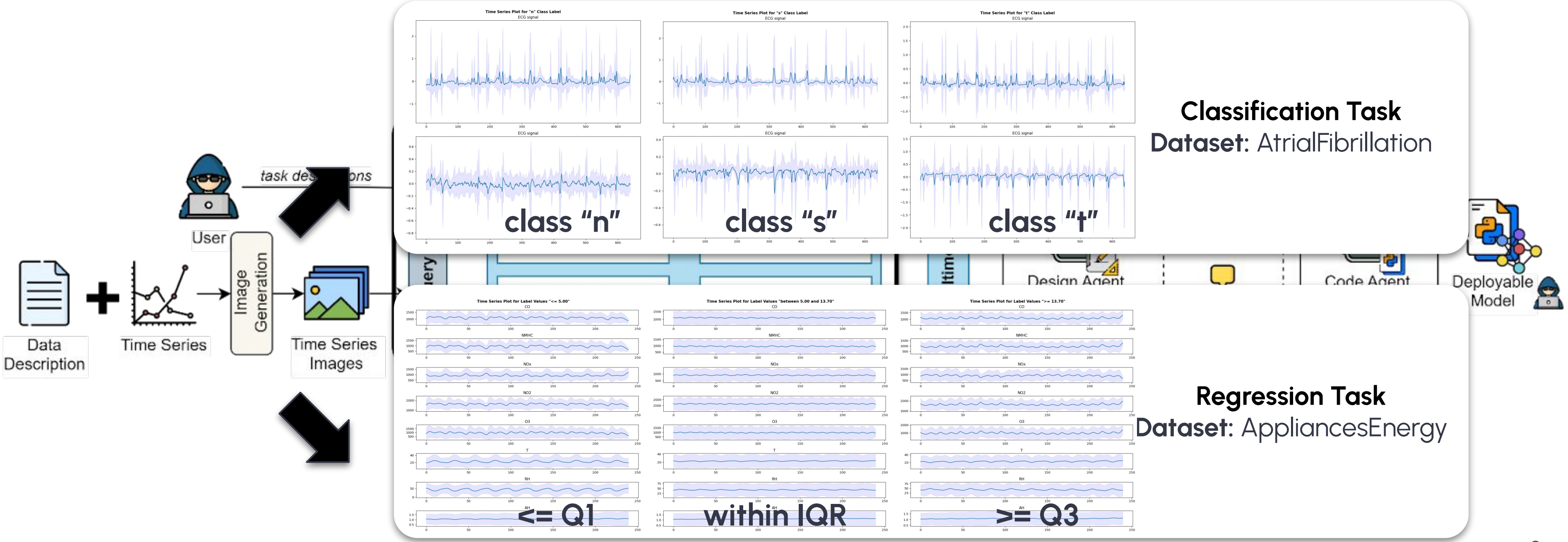
MONAQ👑: Multimodal Query Generation

This module processes the input time series and natural language queries with constraints, then outputs a multimodal data, representing the original user's query from data and modeling perspectives with different aspects of the input time series.



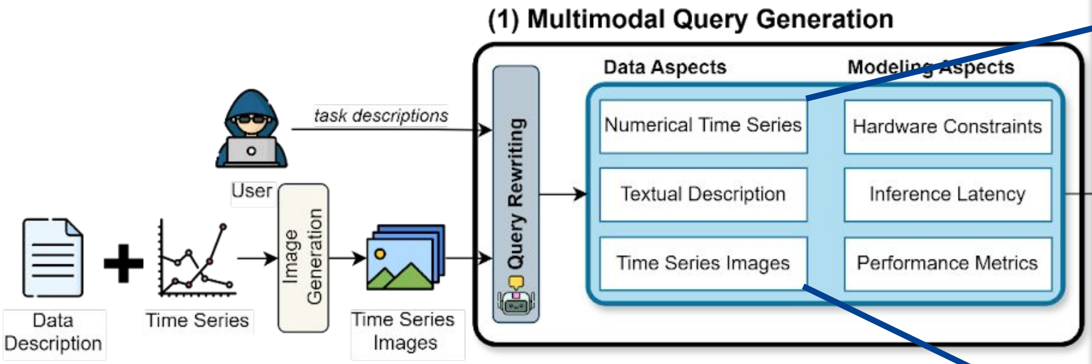
MONAQ👑: Multimodal Query Generation

We generate the representative time series from the training set. For example, one image per class (classification), or one image per range (regression).



MONAQ👑: Data Aspect Query

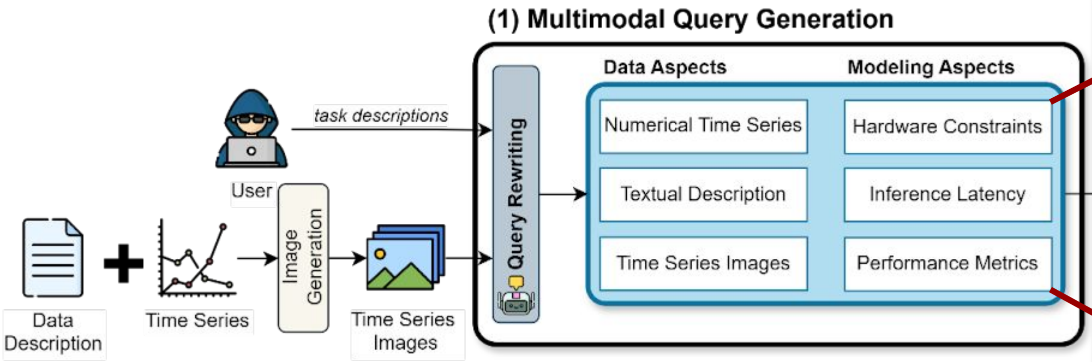
The rewritten query of the data aspect will be used for the subsequent search process.




```
{
    "task_description": "The user wants to build a classification model to categorize ECG signals into three types of atrial fibrillation (AF) for deployment on wearable devices like Fitbit trackers.",
    "data_aspects": {
        "name": "PhysioNet ECG Dataset",
        "description": "This dataset consists of two-channel ECG recordings used in the Computers in Cardiology Challenge 2004. It includes 5-second segments of atrial fibrillation, sampled at 128 samples per second, with class labels 'n', 's', and 't'.",
        "features": "The dataset contains 1-D ECG signals from two channels. Class 'n' represents non-terminating AF, class 's' represents self-terminating AF after at least one minute, and class 't' represents AF terminating immediately.",
        "context": "The dataset was created for an open competition aimed at developing automated methods for predicting spontaneous termination of AF.",
        "patterns": "The time series plots show distinct patterns for each class, with variations in amplitude and frequency that can be leveraged for classification.",
    },
    "model_aspects": {...},
}
```

MONAQ👑: Model Aspect Query

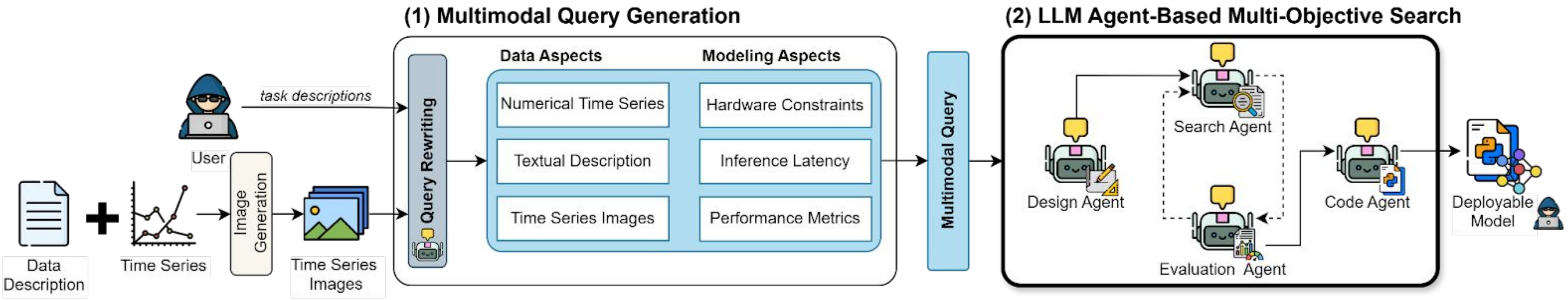
The rewritten query of the model aspect focuses on the key aspects of building efficient models for resource-constrained environments.




```
{...,
    },
    "model_aspects": {
        "name": "ECG Classification Model",
        "hardware_specs": {
            "device_name": "Fitbit Tracker",
            "ram": "512000",
            "flash": "2000000",
        },
        "MAC": "1000000",
        "parameters": "50000",
        "latency": "100",
        "performance": "High accuracy for classifying ECG
signals into the specified AF types.",
    },
}
```

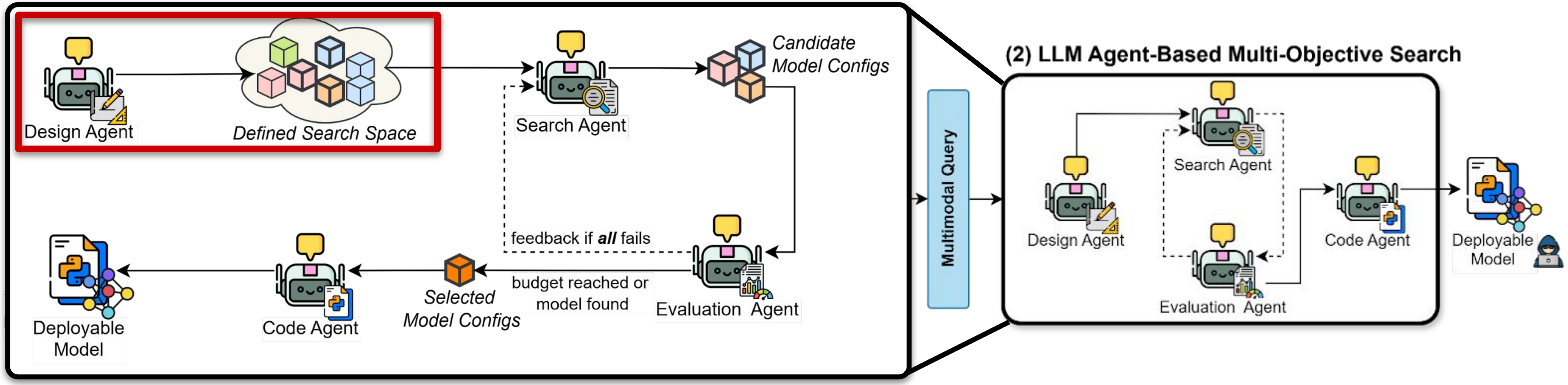
MONAQ👑: LLM Agent-Based Multi-Objective Search

We introduce an LLM agent-based multi-objective search module, enabling LLMs to autonomously interact based on the extracted multimodal query.



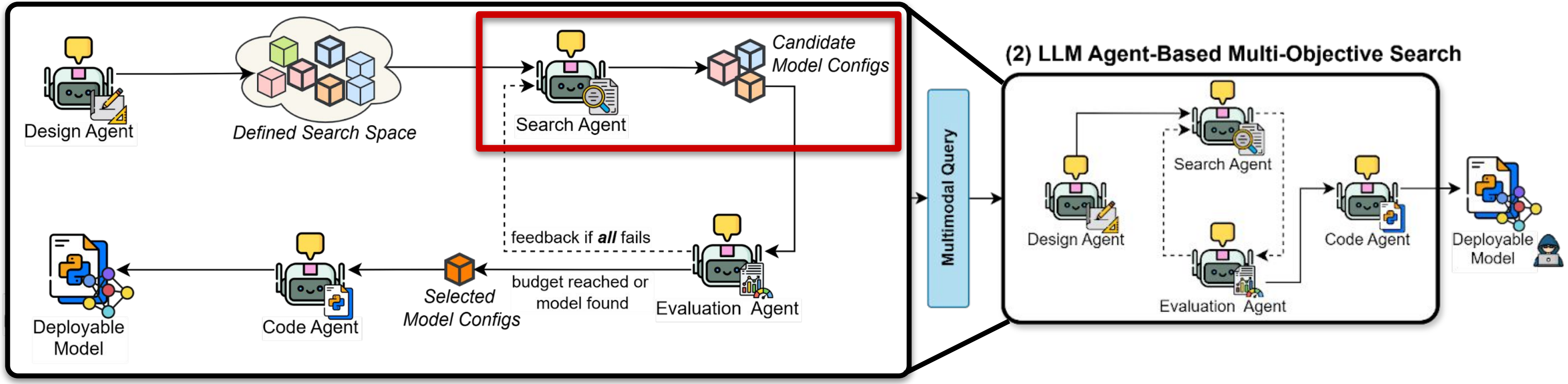
MONAQ👑: LLM Agent-Based Multi-Objective Search

Design Agent takes the organized multimodal query and generated images as input and design the search space.



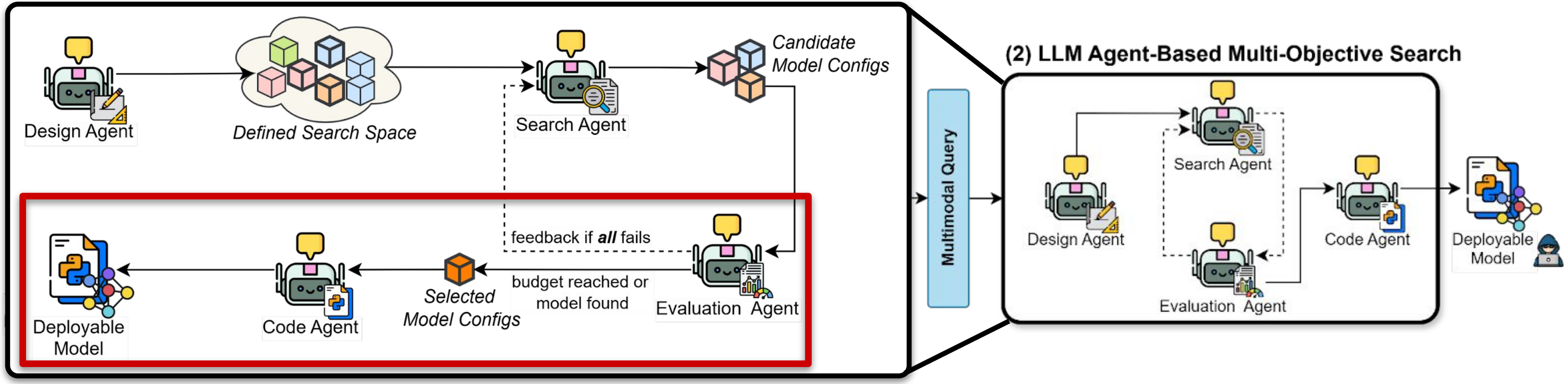
MONAQ👑: LLM Agent-Based Multi-Objective Search

Search Agent generates (a set) of candidate models.



MONAQ👑: LLM Agent-Based Multi-Objective Search

If any of the suggested candidates pass the evaluation, the selected network is forwarded to the Code Agent, which generates code to produce a deployable model for the user.

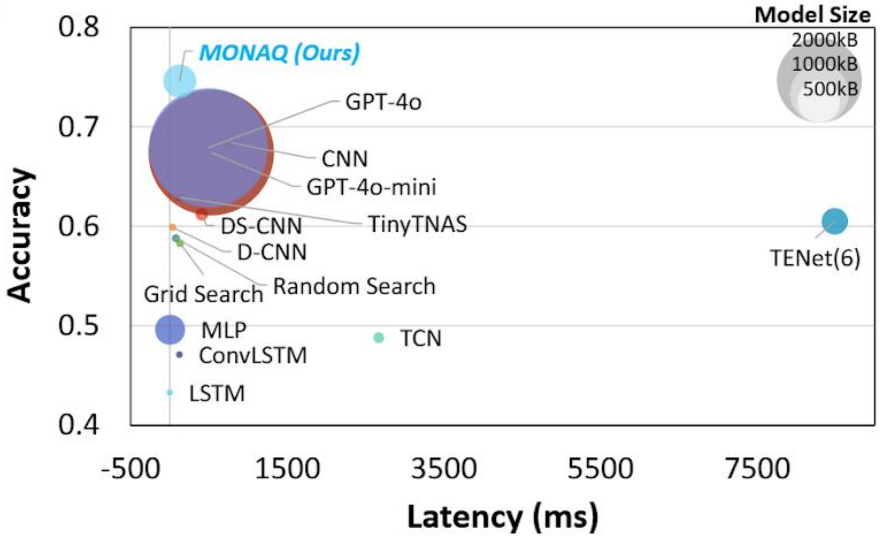


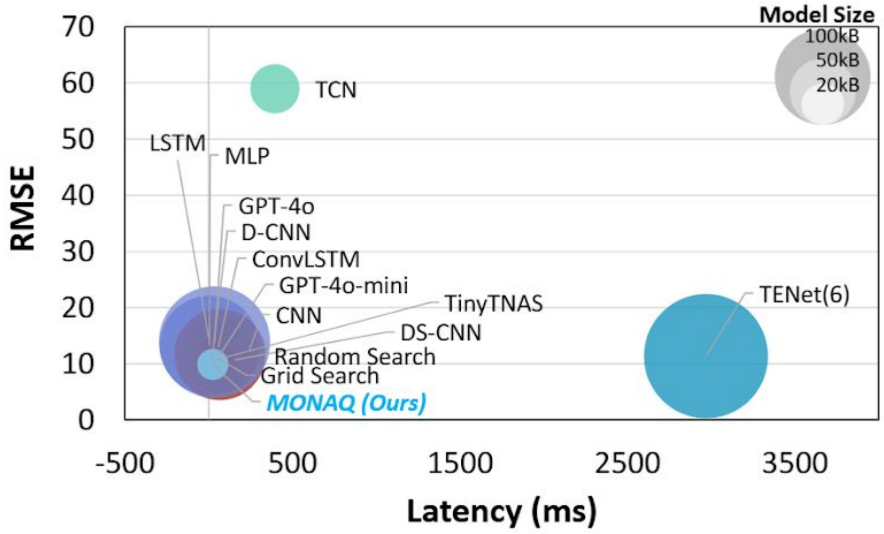
Experimental Setup

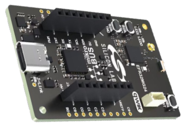
Datasets	Length	Feature Dims (# Sensors)	# Train	# Test	# Classes	Application Domain	Missing Values
Classification (Bagnall et al., 2018; Li et al., 2023)							
AtrialFibrillation	640	2	15	15	3	Health Monitoring	No
BinaryHeartbeat	18530	1	204	205	2	Health Monitoring	No
Cricket	1197	6	108	72	12	Human Activity Recognition	No
Fault Detection (A)	5120	1	10912	2728	3	Industrial System Monitoring	No
UCI-HAR	206	3	7352	2947	6	Human Activity Recognition	No
P12	233	36	9590	2398	2	Health Monitoring	**Yes**
P19	401	34	31042	7761	2	Health Monitoring	**Yes**
PAMAP2	4048	17	4266	1067	8	Human Activity Recognition	**Yes**
Regression (Tan et al., 2021)							
AppliancesEnergy	144	24	96	42		Energy Monitoring	No
BenzeneConcentration	240	8	3433	5445		Environment Monitoring	**Yes**
BIDMC32SpO2	4000	2	5550	2399		Health Monitoring	No
FloodModeling	266	1	471	202	N/A	Environment Monitoring	No
LiveFuelMoistureContent	365	7	3493	1510		Environment Monitoring	No
HouseholdPowerConsumption1	1440	5	746	694		Energy Monitoring	**Yes**
HouseholdPowerConsumption2	1440	5	746	694		Energy Monitoring	**Yes**

Table 1: Summary of benchmark datasets.

Downstream Task Performance







Results are based on a simulation on EFR32xG24 at 78MHz (1536 kB flash and 256 kB RAM) using MLTK toolkit.

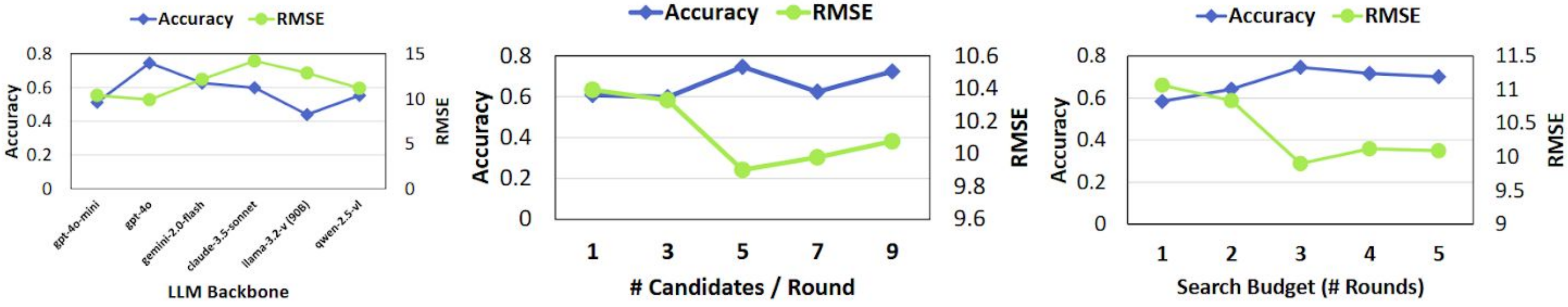
Ablation Study

Agents	Numerical Time Series	Textual Descriptions	Time Series Images	Classification Latency (ms)	Classification Accuracy	Classification FLASH (kB)	Regression Latency (ms)	Regression RMSE	Regression FLASH (kB)
Single (GPT-4o Backbone)	✓			519.159	0.679	3349.453	23.797	13.944	125.445
		✓		1017.267	0.665	4126.024	42.779	13.227	134.901
			✓	593.541	0.690	5971.792	35.859	12.562	193.926
	✓	✓		807.459	0.628	4926.157	40.485	13.556	137.581
	✓	✓	✓	557.665	0.629	3871.910	22.726	12.681	90.398
Multiple (GPT-4o Backbone)	✓			149.320	0.434	**12.066**	54.270	12.284	10.611
		✓		170.461	0.440	15.198	110.751	12.084	12.560
			✓	280.198	0.661	15.638	**13.661**	11.653	**7.885**
	✓	✓		205.623	0.517	16.035	28.049	13.207	13.875
	✓	✓	✓	**127.260**	**0.746**	257.742	24.729	**9.902**	10.582

Variations	Classification Latency (ms)	Classification Accuracy	Classification FLASH (kB)	Regression Latency (ms)	Regression RMSE	Regression FLASH (kB)
MONAQ	127.260	**0.746**	257.742	24.729	**9.902**	10.582
w/o Query Rewriting	206.871	0.651	17.186	14.623	11.994	10.155
w/o $\mathcal{A}_{design}$	863.358	0.647	518.762	95.654	13.512	33.243
w/o $\mathcal{A}_{eval}$	540.411	0.641	4775.661	26.335	12.783	109.627
w/o $\mathcal{A}_{eval}$ & $\mathcal{A}_{search}$	601.313	0.643	5907.363	188.123	11.261	665.110
Only $\mathcal{A}_{code}$	579.876	0.612	4158.205	21.530	12.274	99.638

- Without query rewriting → −10% accuracy
- Without multi-agent setup → slower and less efficient
- Full MONAQ = best trade-off

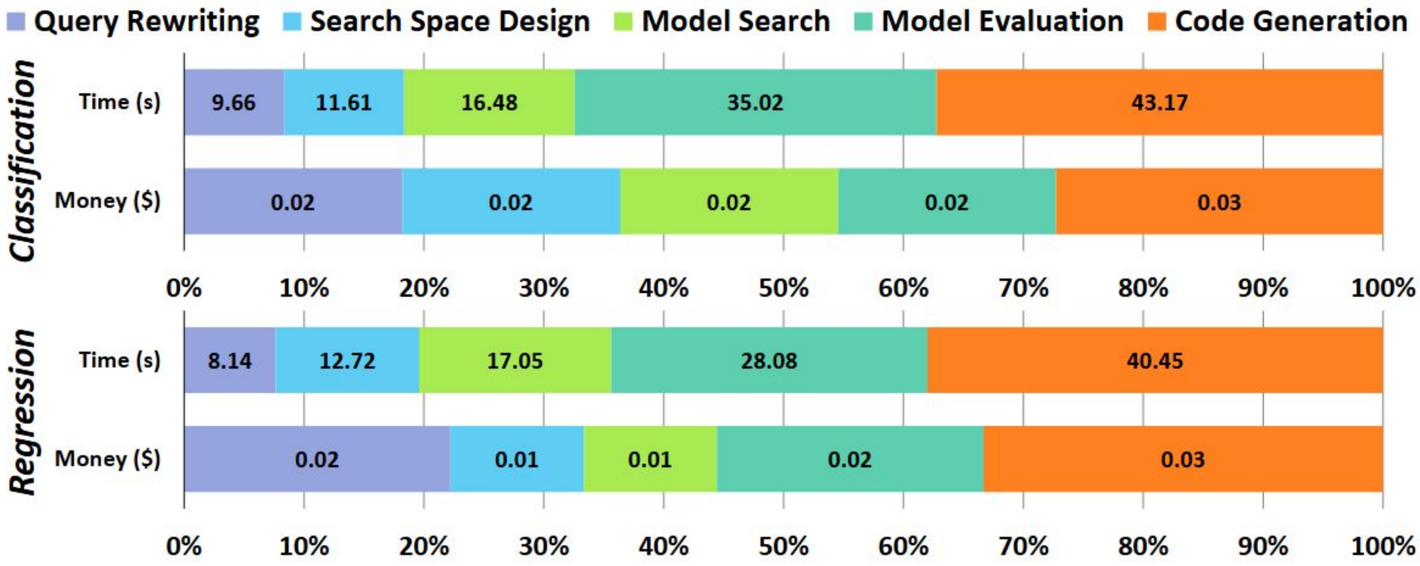
Hyperparameter Study



(a) LLM Backbones. (b) Number of Candidates. (c) Search Budgets.

Figure 7: Comparison between (a) LLM backbones, (b) number of candidates per round, and (c) search budget on model performance, as measured by accuracy (higher is better) and RMSE (lower is better).

Cost Breakdown



Classification
(115.94 seconds, $0.11 / dataset)

Regression
(106.45 seconds, $0.09 / dataset)

Conclusions

Key Contributions:

- Reformulated NAS as Neural Architecture Querying
- Introduced Multimodal Query Generation for time-series understanding
- Developed LLM Agent-Based Multi-Objective Search

Results:

- State-of-the-art accuracy under strict device limits
- Fast, low-cost, training-free search

Implications:

- New paradigm for LLM-driven AutoML
- Potential extension to vision, speech, and sensor fusion domains

Thank you!

GitHub: https://github.com/kaist-dmlab/MONAQ
Email: patara@deepauto.ai