Table 2. 7 MLRC-BENCH tasks representing cutting-edge machine learning research. For each competition, we show the venue where the competition is held, research area, data modality, performance metric, along with the maximum allowed runtime and GPU memory based on our hardware configurations.

Competition	Venue	Research Area	Modality	Metric	Test Runtime	GPU Memory
LLM Merging (Tam et al., 2024)	NeurIPS 2024	Efficient LLM	Text	Accuracy, ROUGE	1 hour	48 GB
Backdoor Trigger Recovery (Xiang et al., 2024)	NeurIPS 2024	LLM Safety	Text	REASR, Recall	0.5 hour	48 GB
Temporal Action Localisation (Heyward et al., 2024)	ECCV 2024 Workshop	Multimodal Perception	Video, Audio	mAP	0.5 hour	16 GB
Rainfall Prediction (Gruca et al., 2022)	NeurIPS 2023	AI for Science	Satellite Data	Critical Success Index	0.5 hour	48 GB
Machine Unlearning (Triantafillou et al., 2024)	NeurIPS 2023	Data Privacy	Image	Forgetting Quality, Accuracy	0.5 hour	16 GB
Next Product Recommendation (Jin et al., 2023)	KDD Cup 2023	Recommendation System	Text	Mean Reciprocal Rank	0.5 hour	16 GB
Cross-Domain Meta Learning (Carrión-Ojeda et al., 2022)	NeurIPS 2022	Few-Shot Learning	Image	Accuracy	3.5 hours	16 GB