



ROBERTTLANGE



@TUBERLIN



@ROBERTTLANGE

MLE-Infrastructure: A Set of Lightweight Tools for Distributed Machine Learning Experimentation

Robert Tjarko Lange



@SPREKELERLAB



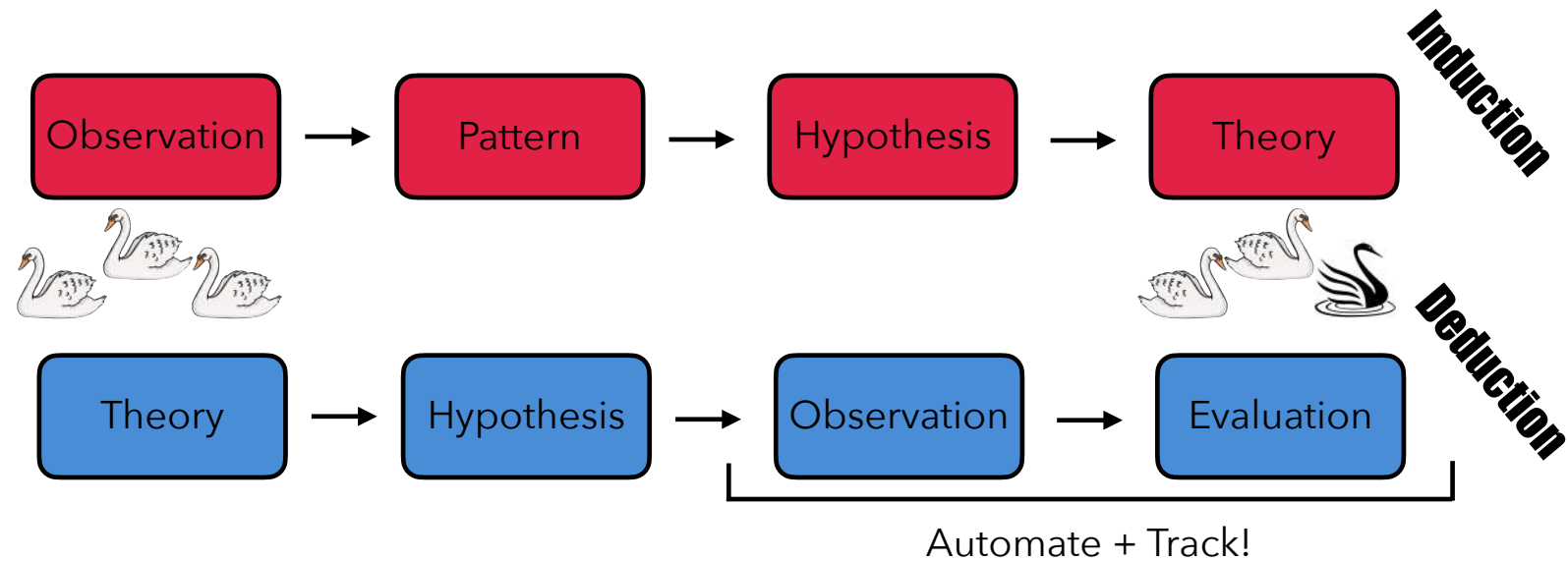
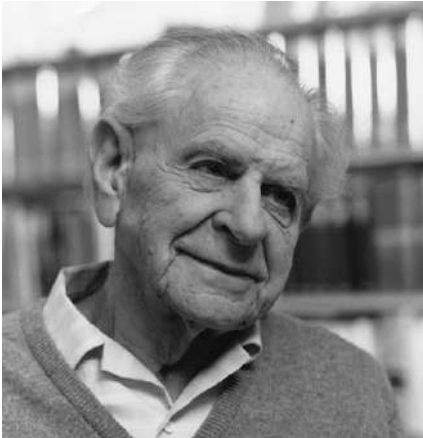
@ECNBERLIN



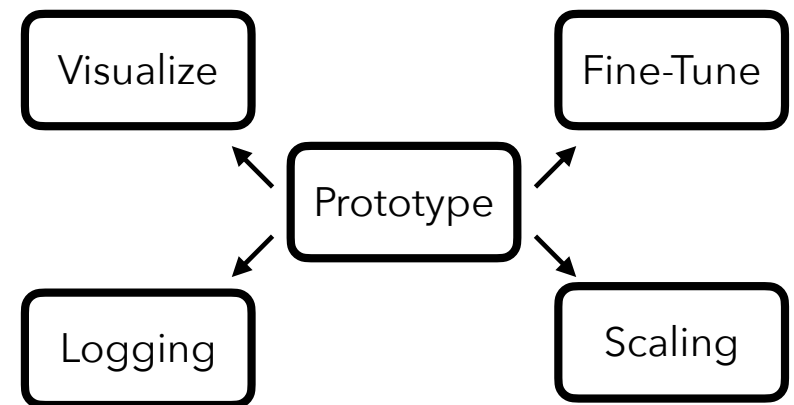
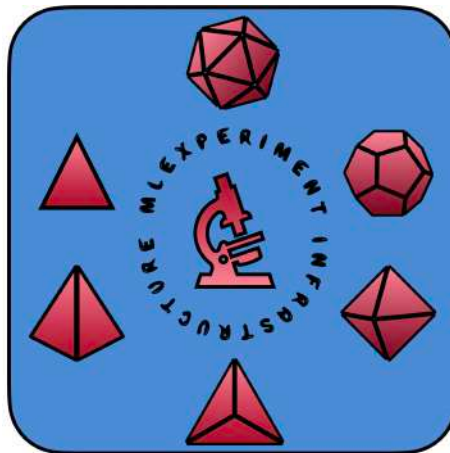
@SCIOI

'The Logic of Scientific Discovery' & Machine Learning

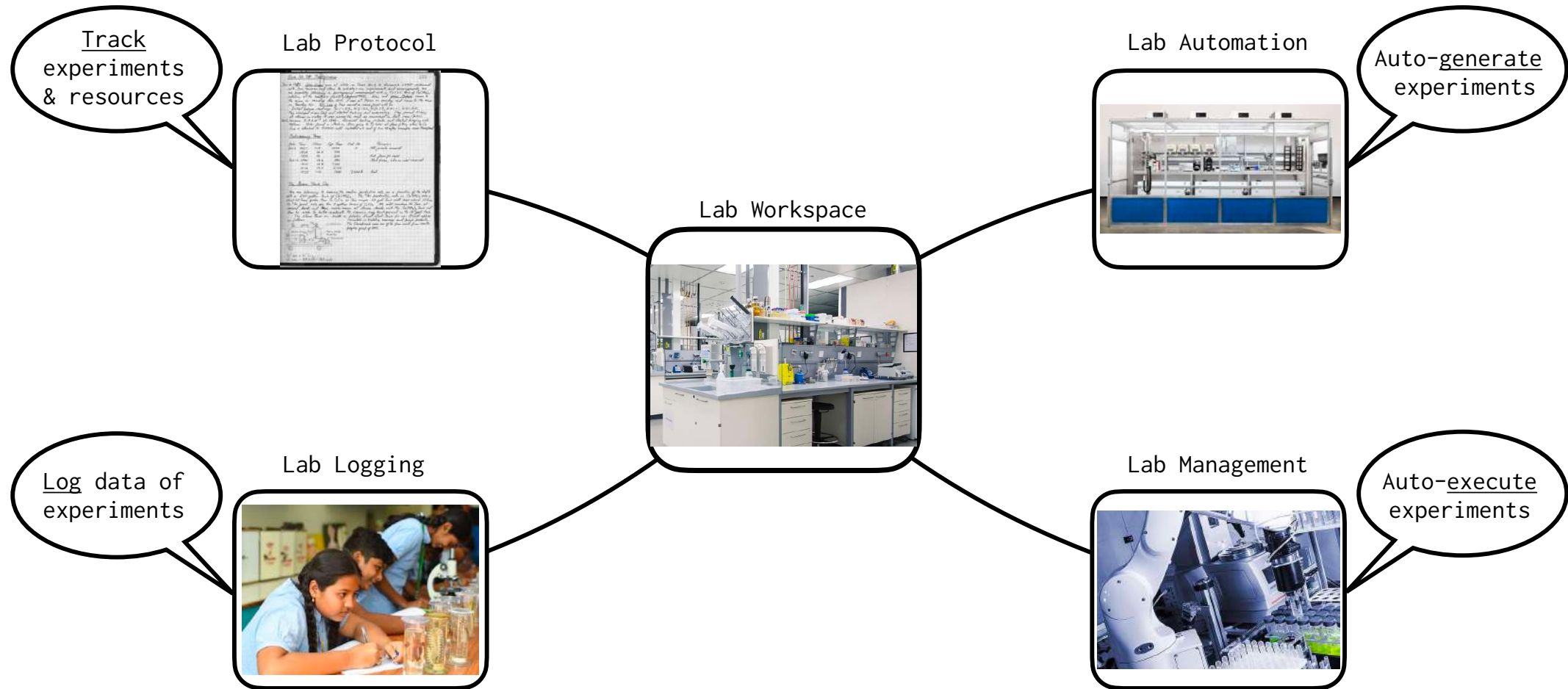
Karl Popper (1902 - 1994)



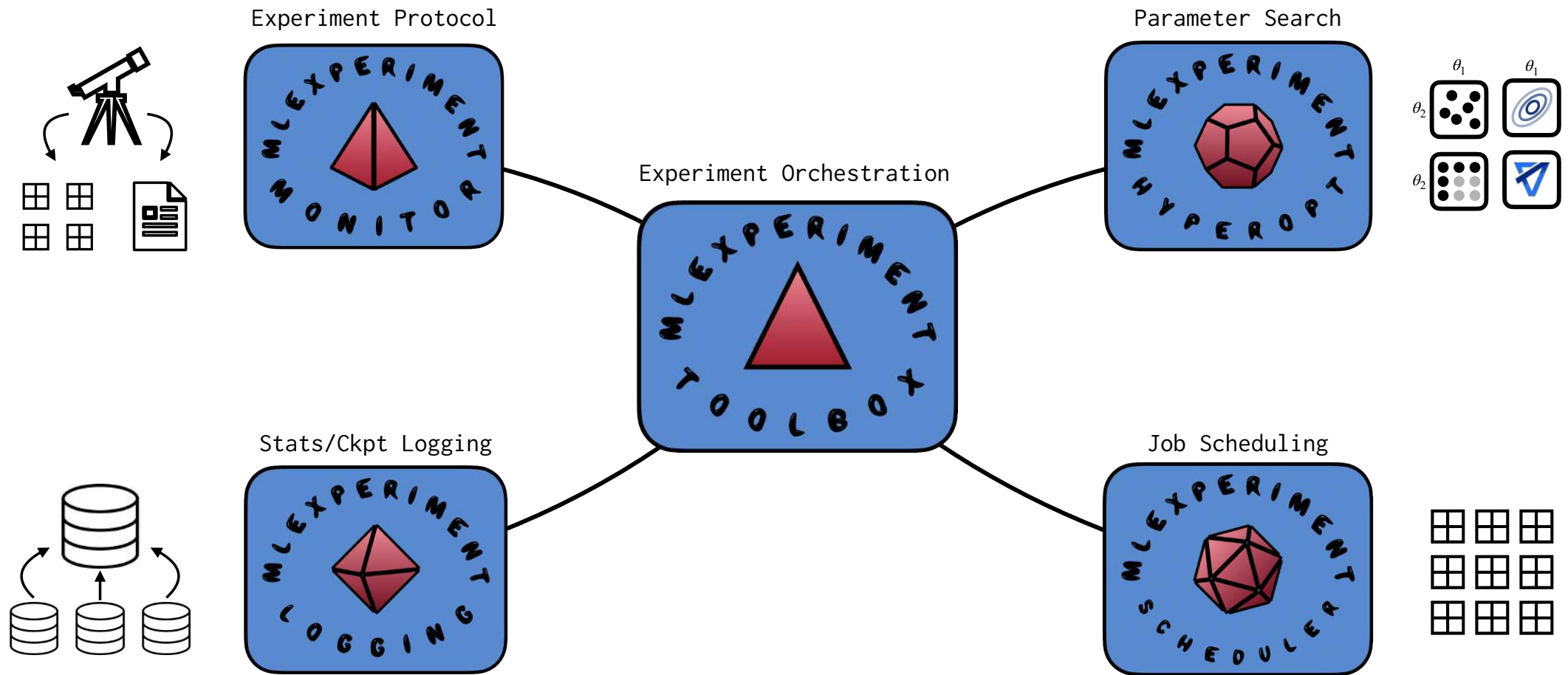
What makes a computational lab?

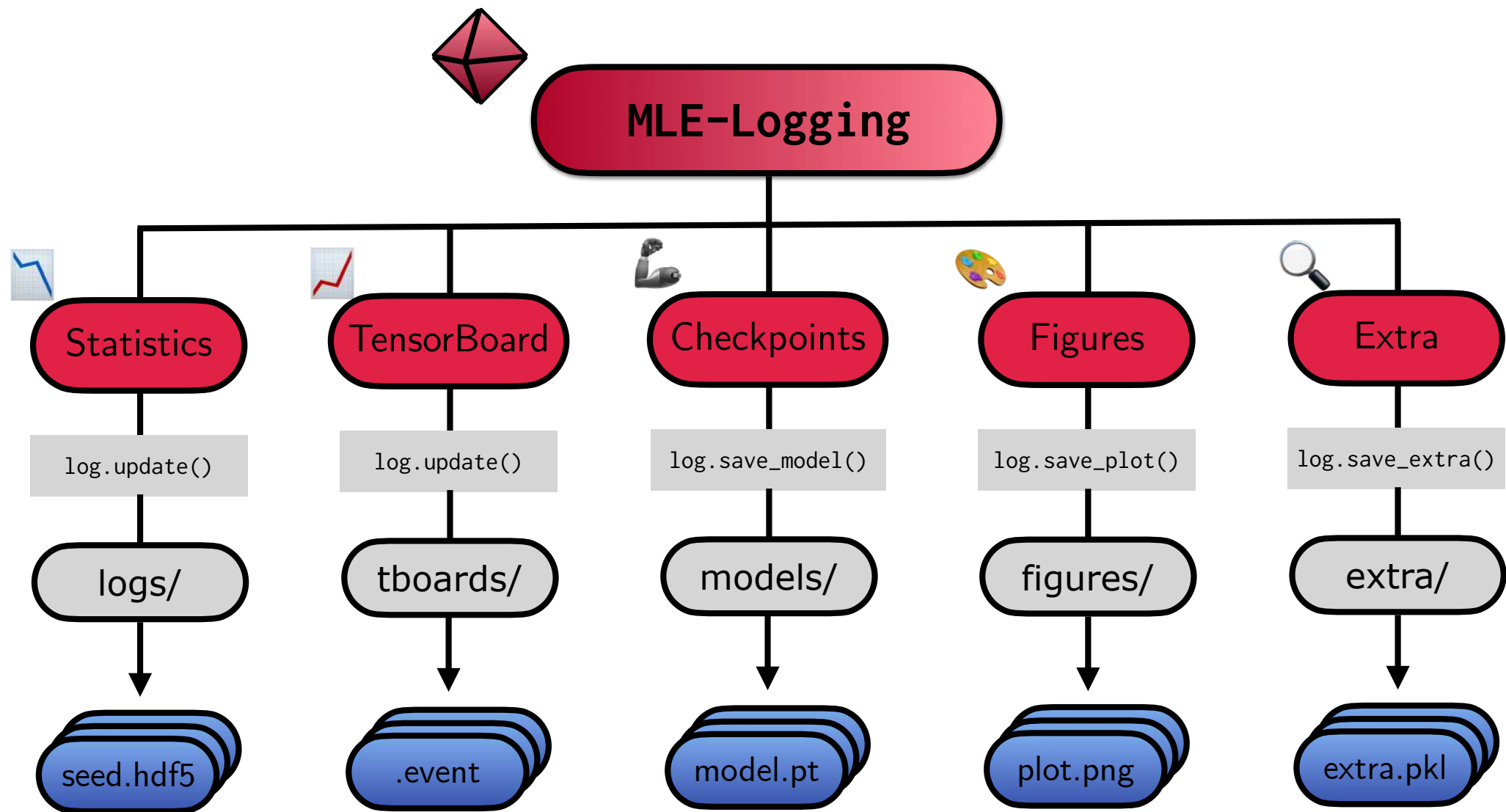


Rapid Hypothesis Testing Infrastructure: Biology Vision



Rapid Hypothesis Testing Infrastructure: In-Silico Version





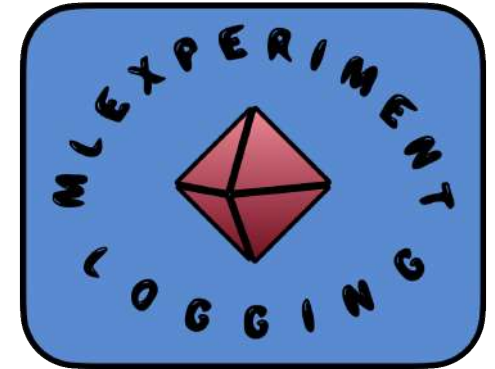
MLE-Logging: A Lightweight Machine Learning Experiment Logger

```
from mle_logging import MLELogger

# Instantiate logging to experiment_dir
log = MLELogger(time_to_track=['num_updates', 'num_epochs'],
                what_to_track=['train_loss', 'test_loss'],
                experiment_dir="experiment_dir/",
                config_fname="config_1.json",
                seed_id=1,
                model_type='torch',
                verbose=True)

# Training step: Get stats, model checkpoint, figure & extra dict
time_tic = {'num_updates': 10, 'num_epochs': 1}
stats_tic = {'train_loss': 0.1234, 'test_loss': 0.1235}
model = ...
fig = ...

# Log step: Store the data in sub-directories
log.update(time_tic, stats_tic, model, fig, some_dict, save=True)
```





MLE-Scheduler



Slurm



GridEngine



GCP VM



SSH Server



Local

MLEJob 

schedule

monitor

run

MLEQueue  ... 

launch

monitor

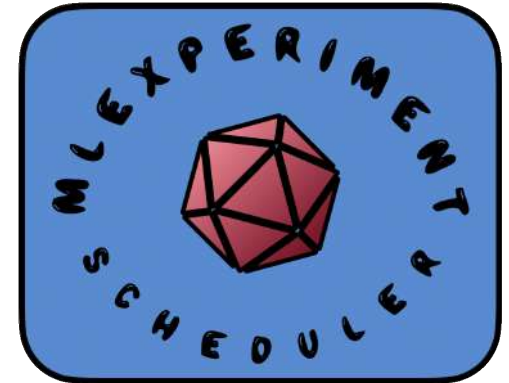
run

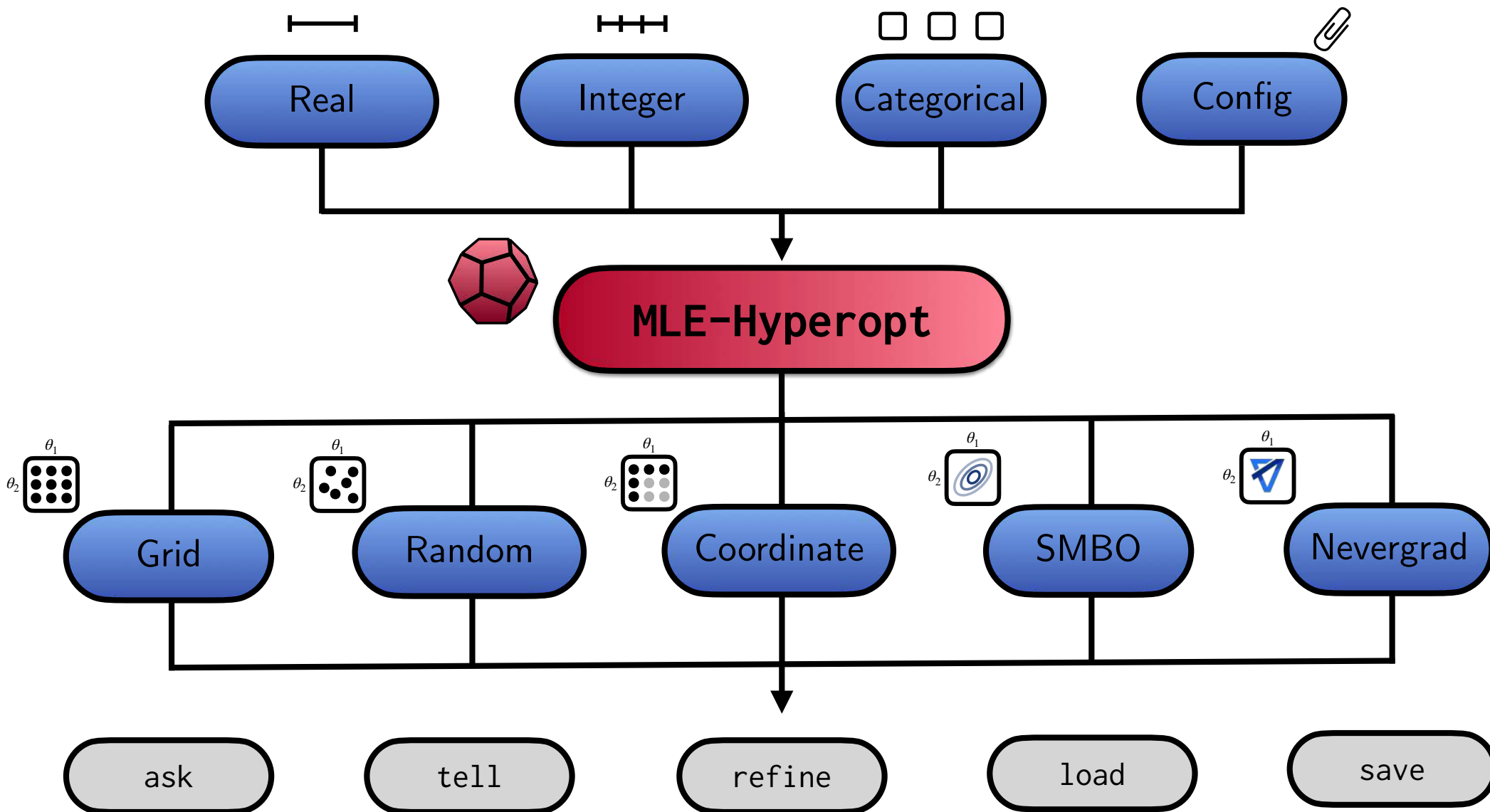
MLE-Scheduler: A Job Manager

```
from mle_scheduler import MLEQueue

# Each job requests 5 CPU cores & 1 V100S GPU & loads CUDA 10.0
job_args = {
    "partition": "<SLURM_PARTITION>", # Partition to schedule jobs on
    "env_name": "mle-toolbox", # Env to activate at job start-up
    "use_conda_venv": True, # Whether to use anaconda venv
    "num_logical_cores": 5, # Number of requested CPU cores per job
    "num_gpus": 1, # Number of requested GPUs per job
    "gpu_type": "V100S", # GPU model requested for each job
    "modules_to_load": "nvidia/cuda/10.0" # Modules to load at start-up
}

# Launch/monitor queue of 4 jobs (2 configs x 2 seeds)
queue = MLEQueue(
    resource_to_run="slurm-cluster",
    job_filename="train.py",
    job_arguments=job_args,
    config_filenames=["config_1.yaml", "config_2.yaml"],
    experiment_dir="logs",
    random_seeds=[0, 1]
)
queue.run()
```





MLE-Hyperopt: A Lightweight Hyperparameter Optimization Tool

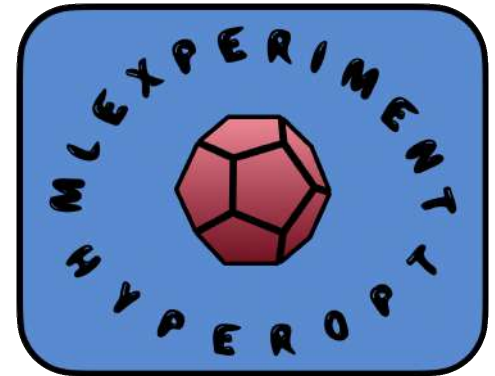
```
from mle_hyperopt import RandomSearch

# Instantiate random search class
strategy = RandomSearch(real={"lrate": {"begin": 0.1,
                                         "end": 0.5,
                                         "prior": "log-uniform"}},
                        integer={"batch_size": {"begin": 32,
                                                "end": 128,
                                                "prior": "uniform"}},
                        categorical={"arch": ["mlp", "cnn"]})

# Ask for a set of candidate proposals
configs = strategy.ask(5)

# Evaluate the configurations
values = [train_network(**c) for c in configs]

# Update the search strategy with new data
strategy.tell(configs, values)
```



General Settings

- GCS Sync Protocol: ✓
- GCS Sync Results: ✓
- DB Path: ~/local_mle_protocol.db

Carpe Diem 🍷

Wed Dec 8 19:45:17 2021

Author: @RobertTLange

MLE-Infrastructure

Resource: sge-cluster

Hi there - You rock! 🍷



Jobs by User/Queue/Node

USER	ALL	🏃	🛑
dorothea	31	31	0
pml_07	3	0	3
stefaah94	2	0	2
gastegger	2	0	2
Sum	38	31	7

QUEUE	ALL	🏃	🛑
cognition-...	31	31	0
Sum	31	31	0

NODE	ALL	🏃	🛑
cognition0...	4	4	0
cognition0...	3	3	0
cognition0...	2	2	0
cognition0...	2	2	0
cognition0...	2	2	0
cognition0...	3	3	0
cognition0...	3	3	0
cognition0...	2	2	0
cognition0...	4	4	0
cognition1...	1	1	0
cognition1...	3	3	0
cognition1...	2	2	0
Sum	31	31	0

Experiment Protocol Summary

🚩	ID	📅	Project	Purpose	Type	▶	🔧	CPU	GPU	🕒	Completed Jobs ✓
✓	18	12/08	secret	BS/Wdecay grid	search	SGE	1	2	0	4 / 4	100%
✓	17	12/08	secret	Lrate grid	search	SGE	1	2	0	4 / 4	100%
✗	16	12/08	secret	Grid Search	search	Local	1	2	0	0 / 4	0%
✗	15	12/08	secret	Grid Search	search	Local	1	2	0	0 / 4	0%
✓	14	12/08	secret	Grid lrate/bs	search	Local	1	2	0	4 / 4	100%
✓	13	12/08	secret	Grid search	search	SGE	1	2	0	4 / 4	100%
✓	12	12/08	secret	Lrate/BS search	search	SGE	1	2	0	4 / 4	100%
✓	11	12/08	secret	Run grid search	search	SGE	1	2	0	4 / 4	100%
✓	10	12/07	secret	Arch/Lrate/Wdec	search	SGE	1	4	1	48 / 48	100%
✗	9	12/07	secret	Arch/Lrate/Wdec	search	SGE	1	4	1	0 / 48	0%
✓	8	12/07	secret	Total exp plot	search	SGE	1	2	0	4 / 4	100%
✓	7	12/06	secret	Classic 500D	search	Slurm	5	1	0	180/180	100%
✓	6	12/06	secret	Classic 100D	search	Slurm	5	1	0	180/180	100%
✓	5	12/06	secret	Classic 50D	search	Slurm	5	1	0	180/180	100%
✓	4	12/05	secret	Test single	single	SGE	1	2	0	1 / 1	100%
✓	3	12/05	secret	Test configs	config	SGE	2	2	0	4 / 4	100%
✗	2	12/05	secret	Test configs	config	SGE	2	2	0	0 / 4	0%
✓	1	12/05	secret	Test monitor	search	SGE	1	2	0	4 / 4	100%

Total Number of Experiment Runs

Total	🏃	✓	✗
18	0	14	4
SGE	Slurm	GCP	Local
12	3	0	3
-	Report	GCS	Sync
-	4	14	1

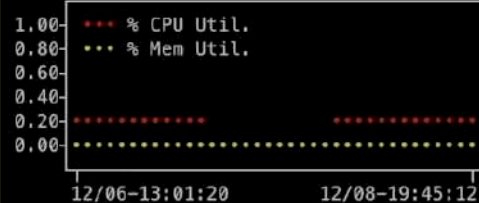
Last Experiment Configuration

E-ID	18
Type	hyperparameter-s...
Dir.	/cognition/home/...
Script	toy_single_objec...
Config	toy_single_objec...
Status	Completed ✓
Resource	sge-cluster

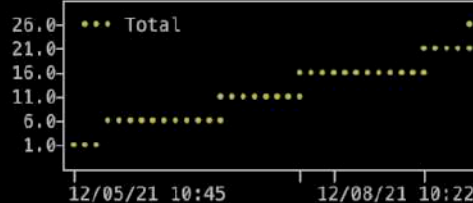
Experiment Completion Time

Conf/Seeds	4/1
Total Jobs	4
# Batches	1
Jobs/Batch	4
Time/Batch	0:00:05
Start Time	12/08/21 19:42
Stop Time	12/08/21 19:43
Duration	0:01:00
-----	-----
🕒 Jobs ✓	4/4 — 100%

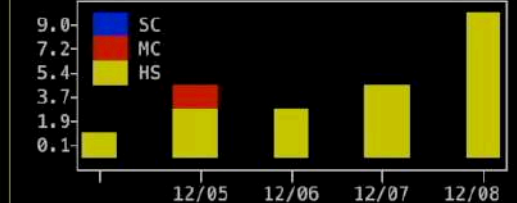
CPU: 486/2560T - Memory: 662/20579G



Protocol Timeline: Total Experiments

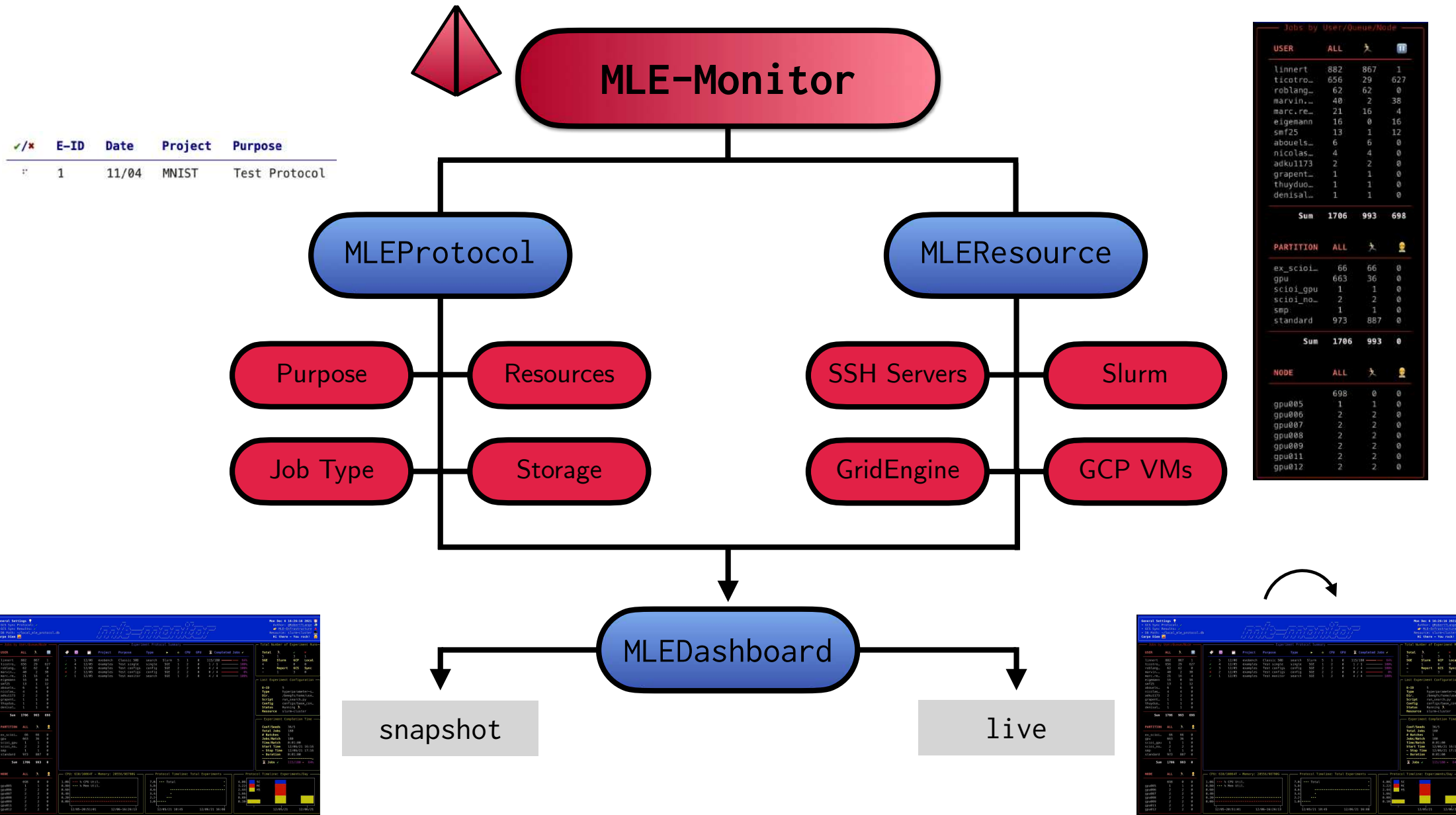


Protocol Timeline: Experiments/Day



[1] 1:python*

"cluster" 19:45 08-Dec-21



MLE-Monitor: A Lightweight Cluster/Cloud VM Monitoring Tool

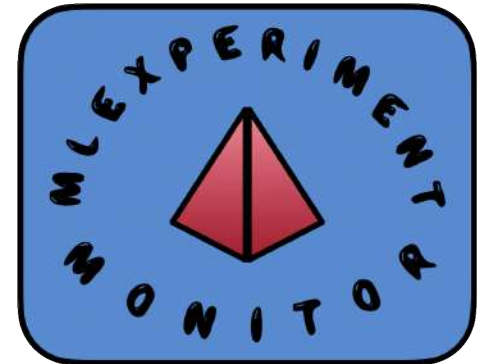
```
from mle_monitor import MLEProtocol, MLEResource, MLEDashboard

# Load protocol, define resource & monitor instance
protocol = MLEProtocol("mle_protocol.db")

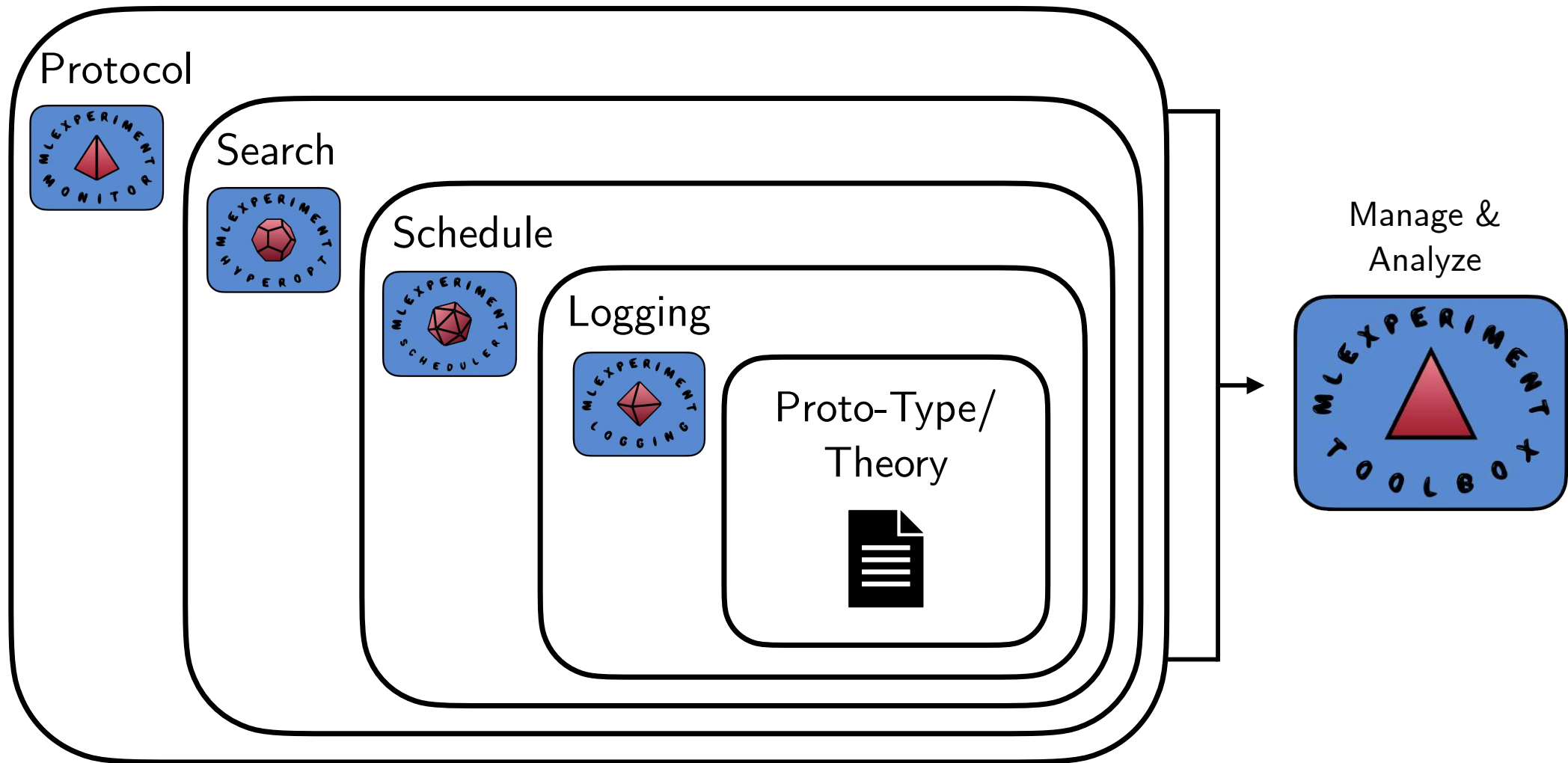
# Instantiate local machine/slurm/grid engine cluster resource
resource = MLEResource(
    resource_name="slurm-cluster",
    monitor_config={"partitions": ["<partition-1>",
                                   "<partition-2>"]},
)

# Get current state of resource utilization
util_data = resource.monitor()

# Instantiate & run dashboard in a while loop
dashboard = MLEDashboard(protocol, resource)
dashboard.live()
```

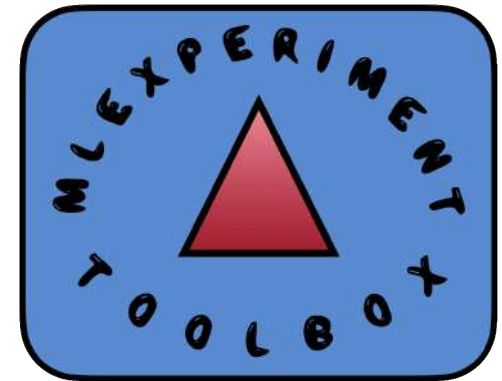


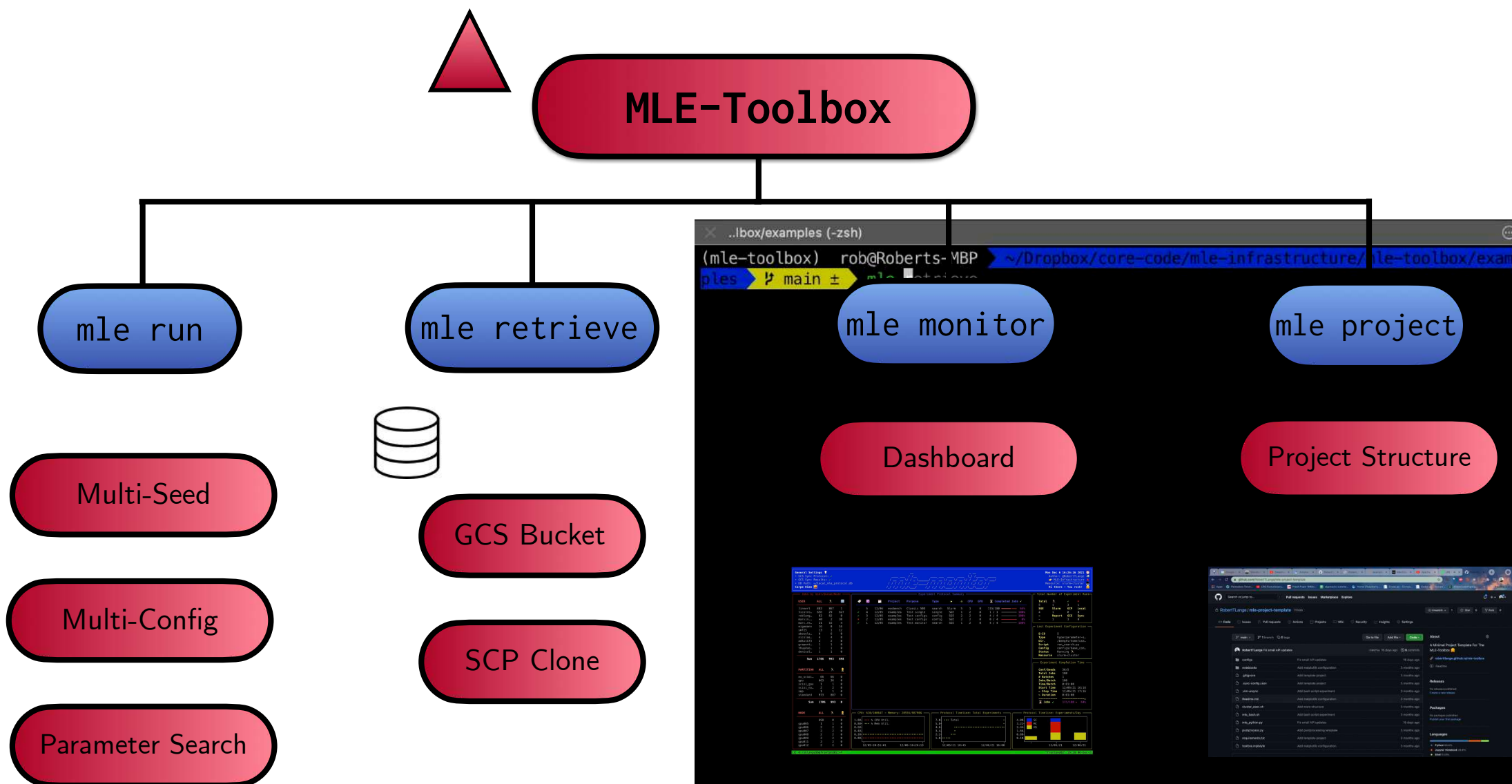
❌ Boilerplate: From Prototype to Protocolled Tuning



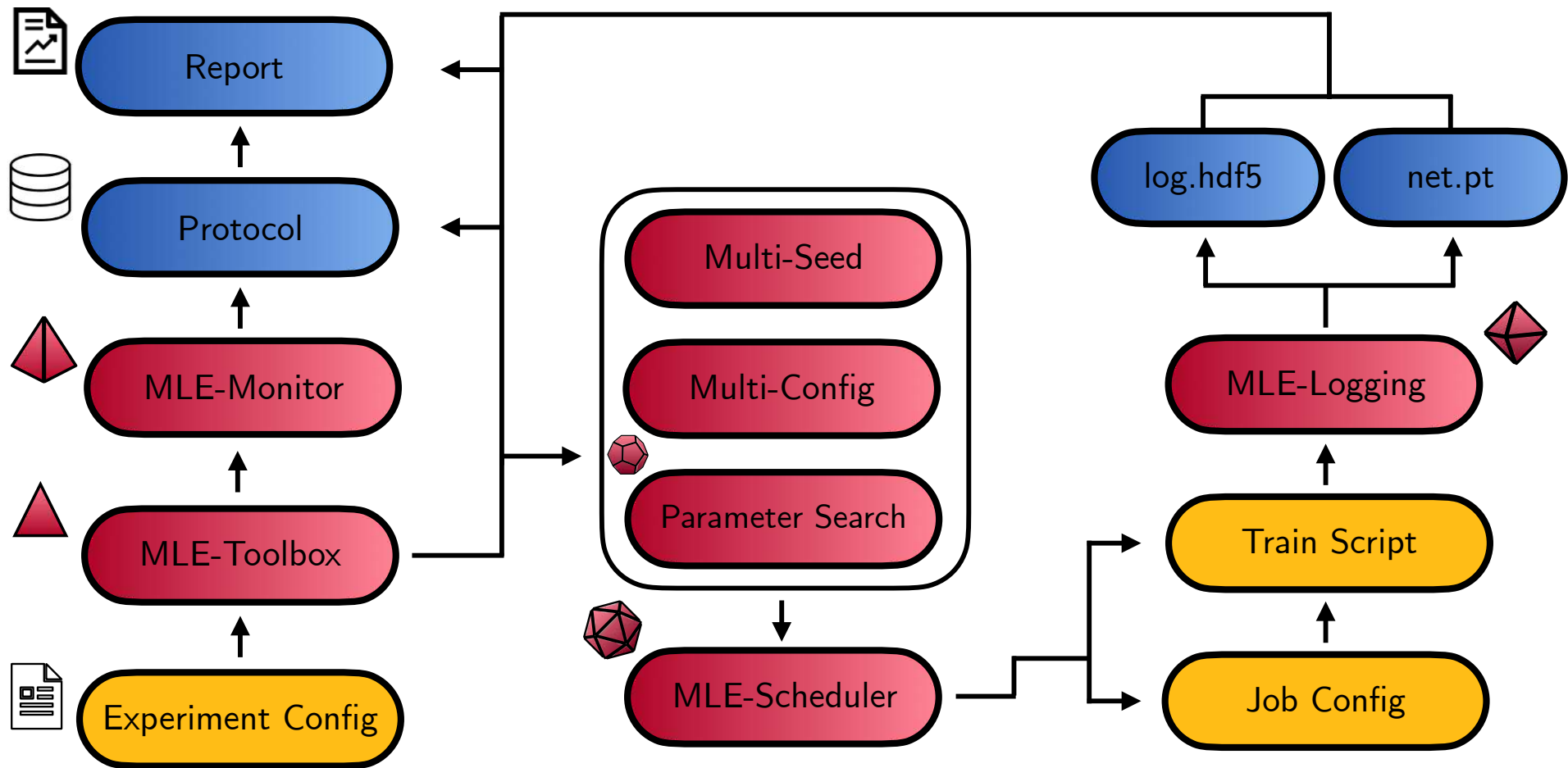
MLE-Toolbox: An Infrastructure for Distributed ML Experiments

```
..lbox/examples (-zsh)  
(mle-toolbox) rob@Roberts-MBP > ~/Dropbox/core-code/mle-infrastructure/mle-toolbox/exam  
ples > main > mle run toy_single_objective/mle_search_grid.yaml
```





The MLE-Infrastructure Workflow Under the Hood



Extra Features & Experimental Functionality

SlackBot



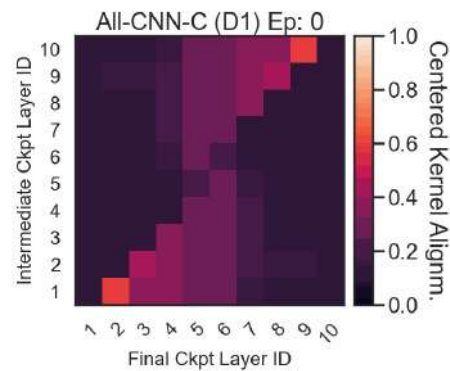
NB-Launch

```
from mle_toolbox.experimental import MLE_NBLauncher

mle_launcher = MLE_NBLauncher(
    resource_name, meta_job_args, experiment_args, single_job_args
)

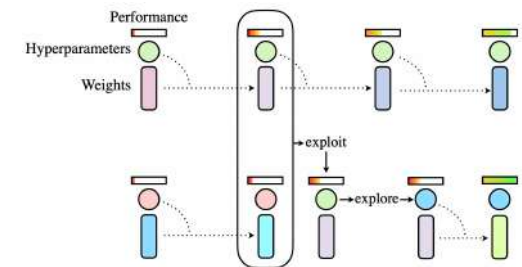
mle_launcher.launch()
```

Visualize

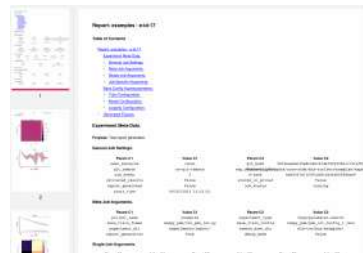


PBT

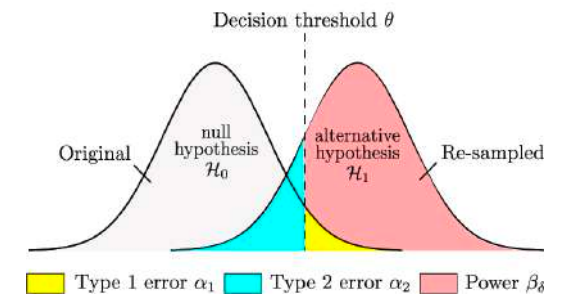
Jaderberg et al. (2017)



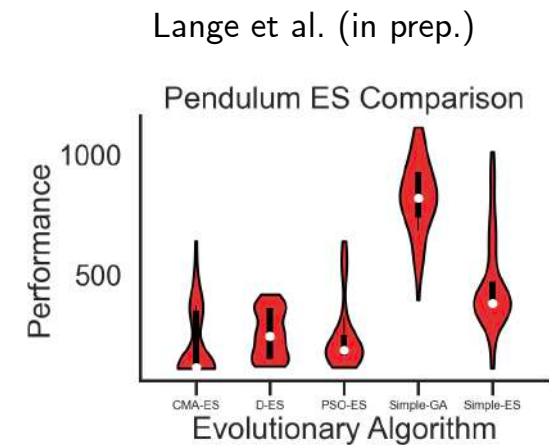
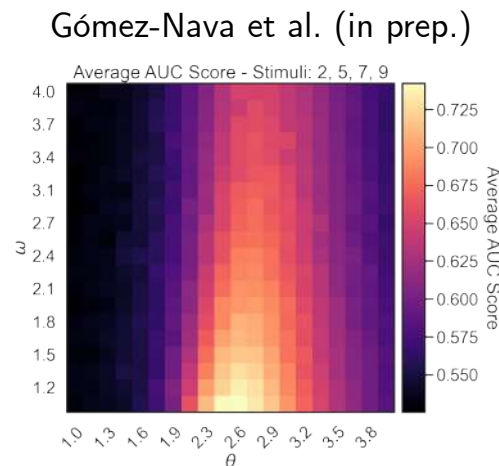
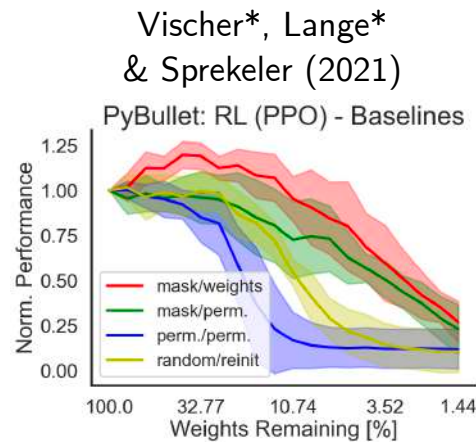
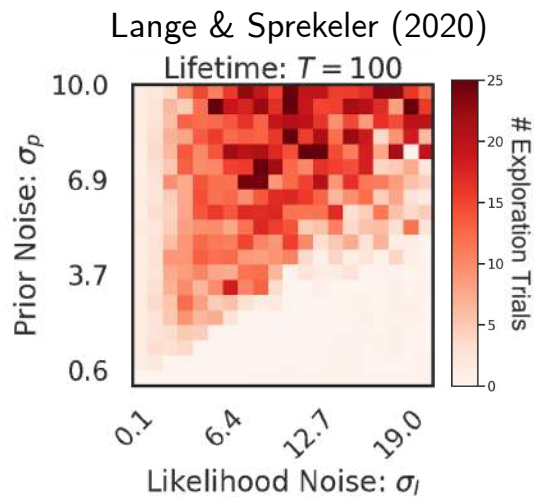
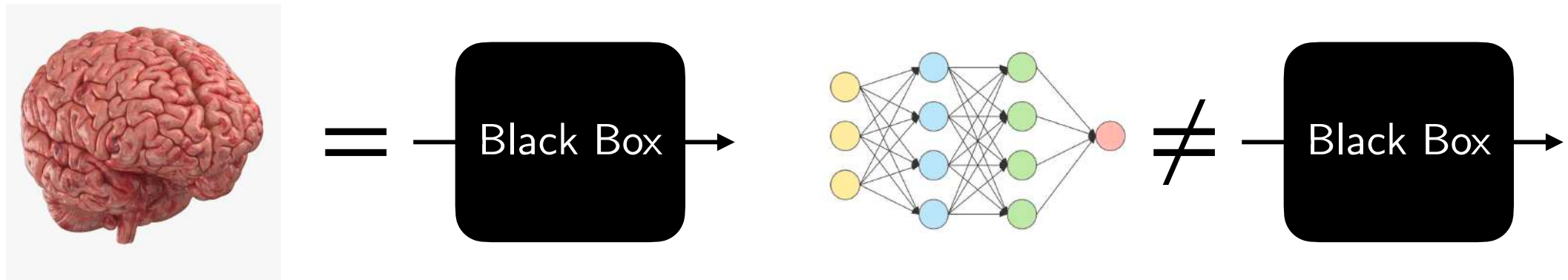
Report



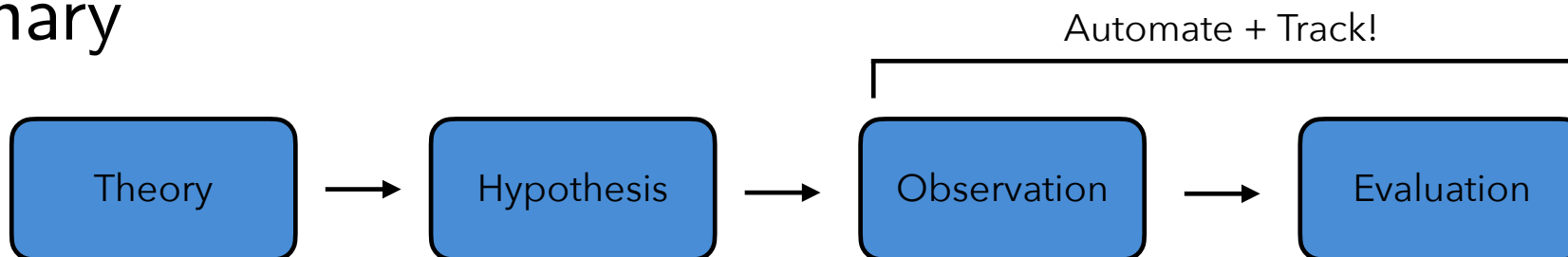
Testing



A Science of DL 🕵️ → 📁 & DL for Science 🕵️ ← 📁



Summary



Lab Logging



Lab Automation



Lab Management



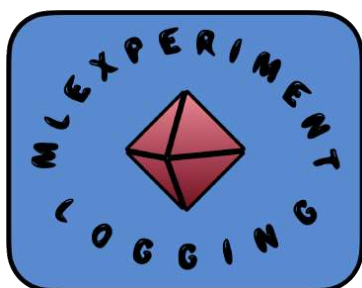
Lab Protocol



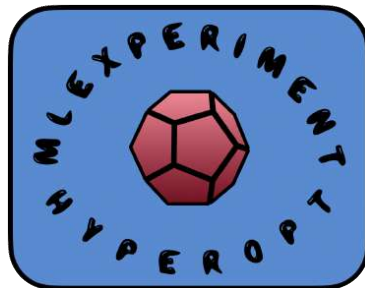
Lab Workspace



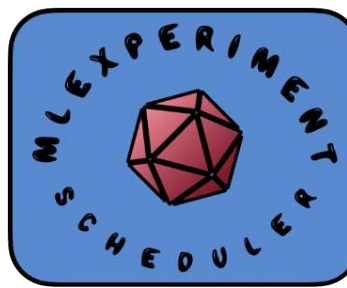
Stats/Ckpt Logging



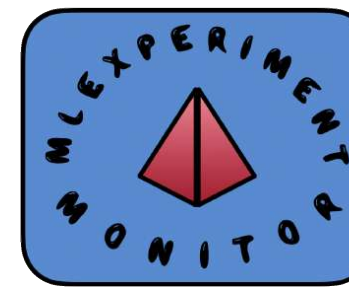
Parameter Search



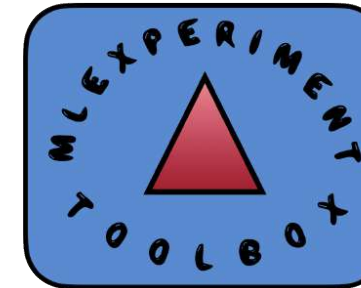
Job Scheduling



Experiment Protocol



Exp. Orchestration

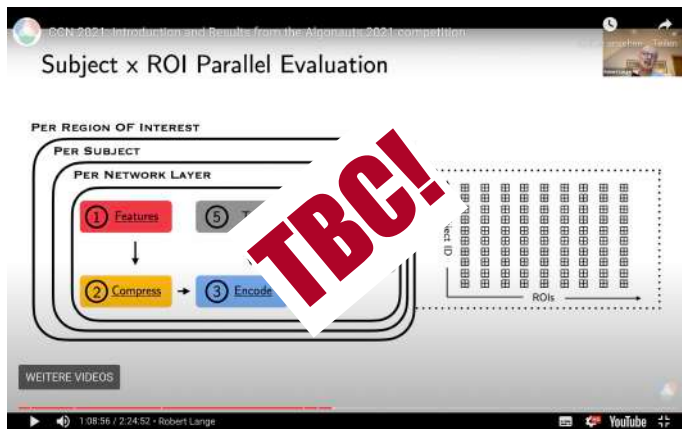


How To Get Started

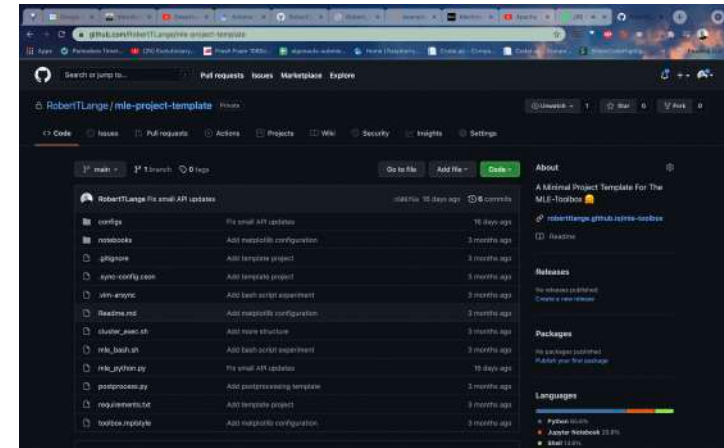
roberttlange.github.io/mle-infrastructure/



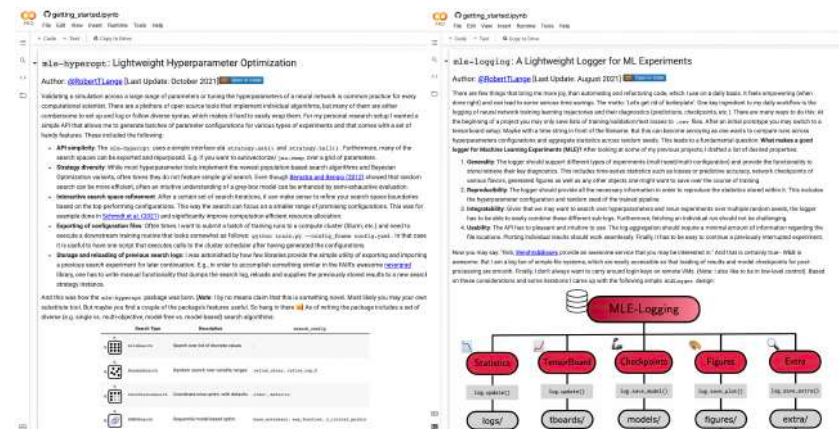
YouTube Tutorial/Walkthrough



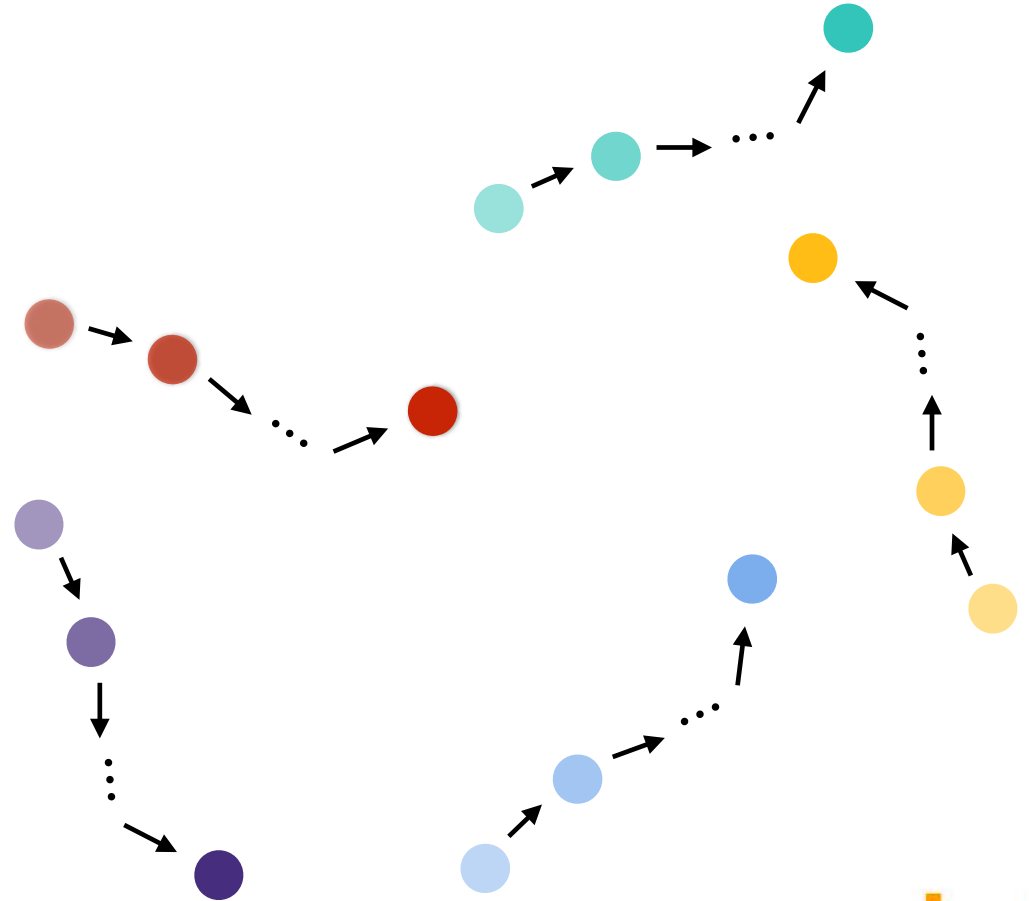
github.com/mle-infrastructure/mle-project



Getting Started Colabs/Notebooks



Thank you for your attention!



**‘SCIENCE MAY BE DESCRIBED AS THE ART OF SYSTEMATIC OVERSIMPLIFICATION.’
- KARL POPPER (1992)**



To Be Continued: MLE-Laboratory

Which experiment do you want to visualize?

14

MLE-Laboratory

- An Interactive Web UI for the [MLE-Infrastructure](#).

Last Loaded Protocol at 12/08/2021 14:37:56

Protocol Summary

	Date	Project	Purpose	Status	Seeds	Resource	CPUs	GPUs	Type	Jobs	Completed Jobs
14	12/08	examples	Grid lrate/bs	completed	1	local	2	0	hyperparameter-search	4	4
13	12/08	examples	Grid search	completed	1	sge-cluster	2	0	hyperparameter-search	4	4
12	12/08	examples	Lrate/BS search	completed	1	sge-cluster	2	0	hyperparameter-search	4	4
11	12/08	examples	Run grid search	completed	1	sge-cluster	2	0	hyperparameter-search	4	4
10	12/07	arapaima	Arch/Lrate/Wdecay	completed	1	sge-cluster	4	1	hyperparameter-search	48	48

Experiment 14 Results

☐ Experiment Configuration

