Supplementary Materials for the paper entitled: Harnessing Pre-trained Generalist Agents for Software Engineering Tasks

In the following, we provide supplementary materials supporting our findings on the paper entitled: Harnessing Pre-trained Generalist Agents for Software Engineering Tasks.

1 Hyperparameters used to fine-tune the generalist agents

As mentioned in Section **3.2.7**, Tables 1, 2 3, and 4 summarize the hyperparameters we used on each generalist agents, as well as the ones we used for each online training algorithm considered in this paper (DQN, PPO, MAENT, and V_TRACE).

2 Results of the bug localization task

As mentioned in Section **3.2.6**, of the paper Table 5 reports the number of bug reports on each repository. In the following we describe each table reporting our findings in the bug localization task.

- Table 6, 10 report the performance of the pre-trained generalist agents across all fine-tuning data budgets on AspectJ and Tomcat projects respectively and the baseline specialist agent w.r.t our evaluation metrics. We observe that for all metrics except top@10, at least one configuration of the generalist agents performs better than the specialist agents.
- Table 9 report the performance of the pre-trained generalist agents across all fine-tuning data budgets on JDT project and the baseline specialist agent w.r.t our evaluation metrics. We observe that for all metrics except MRR, at least one configuration of the generalist agents performs better than the specialist agents.
- Table 8, 11 report the performance of the pre-trained generalist agents across all fine-tuning data budgets on Birt and Eclipse projects respectively and the baseline specialist agent w.r.t our evaluation metrics. We

Table 1: The common hyperparameters that are used to fine-tune the MGDT.

Hyperparameters	Value
Number of layers	10
Number of attention heads	20
Embedding dimension	1280
Nonlinearity function	TanH
Batch size	32
Updates between rollouts	300
Target entropy	-dim(Action)
Buffer size	10000
Weight decay	0.0005
Learning rate	0.0001

Table 2: The common hyperparameters that are used to fine-tune IMPALA

Hyperparameters	Value
Number of actors	48
Nonlinearity function	ReLu
Batch size	32
Use of lstm	True
RMSProp smoothing constant	0.99
RMSProp epsilon	0.01
entropy cost	0.0006
Learning rate	0.00048

observe that for all metrics, at least one configuration of the generalist agents performs better than the specialist agents.

- Table 7 report the performance of the pre-trained generalist agents across all fine-tuning data budgets on SWT project and the baseline specialist agent w.r.t our evaluation metrics. We observe that for all metrics except top@1, at least one configuration of the generalist agents performs better than the specialist agents.

3 Results of the task-based scheduling

Regarding the task-based scheduling, we collect the makespan, the training time, the testing time, as well as the average cumulative reward earned and report our findings in the following:

- Tables 22, 26 and 19 report the performance of the fine-tuned generalist agents on the PDR-based scheduling at zero-shot, 1% and 2% data budget and the baseline specialist in terms of makespan time across the 30×20

Table 3: The hyperparameters that we use to fine-tune the MGDT for each online algorithm

Online algorithms	epsilon start	epsilon end	epsilon clip	surrogate loss epochs
MAENT	NA	NA	NA	NA
PPO	NA	NA	0.2	15 (Blockmaze), 15 (PDR), 3 (MsPacman)
DQN	0.99	0.05	NA	NA

Table 4: The hyperparameters that we use to fine-tune IMPALA for each online algorithm

Online algorithms	epsilon clip	surrogate loss epochs	baseline cost
V-TRACE	NA	NA	0.5
PPO	0.2	15 (Blockmaze), 15 (PDR), 3 (MsPacman)	NA

and 6×6, with the generalist agents achieving greater average performance compared to the baseline specialist.

- Tables 25, 29 and 21 report the results of cumulative reward, training, and testing times of the fine-tuned generalist agents on the PDR-based scheduling at zero-shot, 1% and 2% data budget and the baseline specialist. Across studied instances, the generalist agents achieve greater performance regarding training time and cumulative reward metrics compared to the baseline specialist.

- Table 33 reports the results of the post-hoc test analysis for various generalist agents and the baseline on the 30×20 instance of the task-based scheduling in terms of training time. IMPALA agents significantly outperform the baseline specialist at 1% and 2% fine-tuning data budgets.

- Tables 31 and 32 report the results of the post-hoc test analysis for various generalist agents and the baseline on the 30×20 instance of the task-based scheduling in terms of makespan and cumulative reward. All generalist agent configurations significantly outperform the baseline specialist agent across all fine-tuning data budgets.

4 Results of the playtesting in games task

Regarding the Blockmaze game, we collected the time to find bugs, the training time, the testing time, and the average cumulative reward earned, and report our findings in the following:

- Table 17 reports the performance of the pre-trained generalist agents at 2% fine-tuning data budget on the Blockmaze game and the baseline in terms of time to find bugs, with at least one generalist agent configuration achieving greater average performance.

- Table 18 reports the results of cumulative reward, training, and testing times of the fine-tuned generalist agents on the Blockmaze game at 2%

Table 5: Benchmark statistics of software projects on the bug localization task.

Project	Bugs reported
AspectJ	593
Birt	4,178
Eclipse	6,495
SWT	4,151
Tomcat	1,056

Table 6: Performance of the pre-trained generalist agents across all fine-tuning data budget on AspectJ project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
Baseline	RLO_Ent	NA	2.23e+2	1.22e+8	3.70e-1	4.70e-2	4.35e-2	2.30e-1	4.75e-1
	RLO	NA	2.22e+2	8.53e+7	4.44e-1	4.88e-2	4.52e-2	2.63e-1	**5.01e-1**
zero-shot	IMP_VTRACE	-6.77e+0	3.97e+3	NA	2.94e-1	1.96e-2	2.50e-2	8.36e-2	1.21e-1
	MAENT	**-2.77e-1**	4.79e+2	NA	6.08e-1	4.71e-2	**5.22e-2**	2.26e-1	4.52e-1
1%	MAENT	-2.74e-1	4.81e+2	**6.12e+2**	6.08e-1	4.64e-2	**5.22e-2**	2.17e-1	4.43e-1
	DQN	**-2.77e-1**	5.09e+2	5.98e+3	6.08e-1	4.70e-2	**5.22e-2**	2.26e-1	4.52e-1
	PPO	-3.98e-1	6.13e+2	3.70e+3	**6.09e-1**	**5.05e-2**	3.48e-2	**2.96e-1**	4.96e-1
	IMP_PPO	-6.21e+0	5.76e+1	1.12e+4	2.96e-1	1.87e-2	1.57e-2	8.00e-2	1.22e-1
	IMP_VTRACE	-6.65e+0	4.99e+1	6.28e+3	2.35e-1	1.89e-2	1.77e-2	9.21e-2	1.26e-1
2%	MAENT	-3.24e-1	4.85e+2	1.09e+3	5.99e-1	4.65e-2	**5.22e-2**	2.17e-1	4.43e-1
	DQN	**-2.77e-1**	4.77e+2	1.17e+4	6.08e-1	4.71e-2	**5.22e-2**	2.26e-1	4.52e-1
	PPO	-3.98e-1	4.96e+2	7.25e+3	**6.09e-1**	**5.05e-2**	3.48e-2	**2.96e-1**	4.96e-1
	IMP_PPO	-6.21e+0	**5.41e+1**	2.16e+4	2.79e-1	1.55e-2	1.59e-2	5.32e-2	9.05e-2
	IMP_VTRACE	-6.06e+0	1.90e+2	9.77e+3	2.72e-1	2.11e-2	2.14e-2	8.56e-2	1.11e-1

data budget and the baseline. Generalist agents achieve greater performance compared to the baseline specialist in terms of training time.

- Tables 15 and 16 report the performance of the pre-trained generalist agents at zero-shot and 1% fine-tuning data budgets on the Blockmaze game and the baseline in terms of time to find bugs. At zero-shot fine-tuning data budget, the baseline specialist is faster than the generalist agents at finding bugs.

- Table 24 and 28 report the results of cumulative reward, training and testing times performance of the pre-trained generalist agents at zero-shot and 1% fine-tuning data budget on the Blockmaze game and the baseline specialist agents. The MGDT agents achieve greater reward performance compared to the IMPALA agents.

- Table 30 report the results of post-hoc tests for testing time of the baseline specialist and the MGDT generalist agent on the Blockmaze game. The baseline specialist agent achieves the lowest testing compared to the generalist agents.

Regarding the MsPacman game, we report the results of our experiments in the following:

- Tables 13, 20, 12, 23, 14 and 27.

Table 7: Performance of the pre-trained generalist agents across all fine-tuning data budget on SWT project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
Baseline	RLO	NA	2.91e+3	1.54e+8	2.87e-1	4.49e-2	4.67e-2	2.17e-1	4.43e-1
	RLO_Ent	NA	3.38e+3	1.90e+8	1.98e-1	4.47e-2	**4.99e-2**	2.31e-1	4.35e-1
zero-shot	IMPALA	-5.99e+0	2.85e+3	NA	2.86e-1	4.34e-2	4.11e-2	1.97e-1	2.85e-1
	MGDT	0.00e+0	5.02e+3	NA	1.74e-1	4.59e-2	4.27e-2	2.11e-1	4.64e-1
1%	IMP_VTRACE	-5.99e+0	2.76e+3	9.26e+3	2.21e-1	**4.71e-2**	4.93e-2	2.00e-1	2.87e-1
	MAENT	**0.00e+0**	4.98e+3	**8.17e+2**	1.74e-1	4.56e-2	4.27e-2	2.07e-1	4.59e-1
	DQN	**0.00e+0**	5.03e+3	6.28e+3	1.74e-1	4.59e-2	4.27e-2	2.11e-1	4.64e-1
	PPO	**0.00e+0**	5.15e+3	3.81e+3	1.72e-1	4.71e-2	4.40e-2	2.39e-1	4.71e-1
	IMP_PPO	-5.54e+0	2.75e+3	1.77e+4	**3.52e-1**	4.19e-2	3.63e-2	1.92e-1	2.78e-1
2%	MAENT	**0.00e+0**	4.98e+3	1.39e+3	1.74e-1	4.58e-2	4.27e-2	2.08e-1	4.61e-1
	DQN	**0.00e+0**	5.05e+3	1.23e+4	1.74e-1	4.59e-2	4.27e-2	2.11e-1	4.64e-1
	PPO	**0.00e+0**	5.06e+3	7.39e+3	1.72e-1	**4.71e-2**	4.40e-2	**2.39e-1**	**4.71e-1**
	IMP_PPO	-5.99e+0	**1.35e+3**	3.78e+4	2.27e-1	4.39e-2	4.59e-2	1.94e-1	2.85e-1
	IMP_VTRACE	-5.99e+0	4.09e+2	2.23e+4	2.41e-1	4.49e-2	4.43e-2	2.07e-1	2.93e-1

Table 8: Performance of the pre-trained generalist agents across all fine-tuning data budget on Birt project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
Baseline	RLO	NA	8.39e+2	1.16e+8	2.68e-1	5.12e-2	5.26e-2	2.39e-1	5.07e-1
	RLO_Ent	NA	8.35e+2	1.87e+8	3.49e-1	5.23e-2	5.66e-2	2.54e-1	5.23e-1
zero-shot	IMPALA	-5.66e+0	7.43e+2	NA	3.10e-1	5.11e-2	5.54e-2	2.30e-1	3.31e-1
	MGDT	7.38e-1	1.60e+3	NA	**5.50e-1**	5.44e-2	7.43e-2	**2.86e-1**	5.40e-1
1%	MAENT	7.43e-1	1.62e+3	**6.57e+2**	5.48e-1	5.43e-2	7.43e-2	**2.86e-1**	5.34e-1
	DQN	7.38e-1	1.59e+3	6.03e+3	5.50e-1	5.44e-2	**7.43e-2**	**2.86e-1**	5.40e-1
	PPO	**7.51e-1**	1.64e+3	3.72e+3	5.42e-1	5.38e-2	5.71e-2	2.51e-1	**5.69e-1**
	IMP_PPO	-5.39e+0	4.95e+2	1.65e+4	3.76e-1	4.94e-2	4.46e-2	2.19e-1	3.10e-1
	IMP_VTRACE	-5.38e+0	7.40e+2	7.81e+3	3.22e-1	5.70e-2	6.00e-2	2.48e-1	3.48e-1
2%	MAENT	7.50e-1	1.61e+3	1.17e+3	**5.50e-1**	5.42e-2	**7.43e-2**	**2.86e-1**	5.37e-1
	DQN	7.38e-1	1.60e+3	1.19e+4	**5.50e-1**	5.44e-2	**7.43e-2**	**2.86e-1**	5.40e-1
	PPO	**7.51e-1**	1.63e+3	7.31e+3	5.42e-1	5.38e-2	5.71e-2	2.51e-1	**5.69e-1**
	IMP_PPO	-5.64e+0	3.84e+2	3.24e+4	2.11e-1	5.06e-2	5.66e-2	2.25e-1	3.19e-1
	IMP_VTRACE	-5.84e+0	**1.34e+2**	1.66e+4	3.14e-1	**5.72e-2**	5.71e-2	1.89e-1	2.54e-1

Table 9: Performance of the pre-trained generalist agents across all fine-tuning data budget on JDT project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
zero-shot	IMPALA	-5.99e+0	1.46e+4	NA	2.22e-1	4.79e-2	5.29e-2	2.12e-1	2.96e-1
	MGDT	**0.00e+0**	2.06e+3	NA	1.70e-1	4.80e-2	**6.37e-2**	2.51e-1	4.79e-1
Baseline	RLO	NA	1.07e+3	1.35e+8	**4.75e-1**	4.61e-2	3.91e-2	2.26e-1	4.65e-1
	RLO_Ent	NA	1.88e+3	1.25e+8	1.20e-1	4.66e-2	4.77e-2	2.33e-1	4.67e-1
1%	MAENT	**0.00e+0**	2.09e+3	**6.39e+2**	1.70e-1	4.80e-2	6.37e-2	**2.51e-1**	4.79e-1
	DQN	**0.00e+0**	2.08e+3	5.85e+3	1.70e-1	**4.81e-2**	6.37e-2	2.51e-1	4.81e-1
	PPO	**0.00e+0**	2.10e+3	3.60e+3	1.66e-1	4.72e-2	5.27e-2	2.48e-1	4.90e-1
	IMP_PPO	-5.42e+0	**2.22e+2**	1.31e+4	2.62e-1	4.54e-2	4.31e-2	2.00e-1	2.97e-1
	IMP_VTRACE	-5.99e+0	2.30e+2	4.11e+3	1.72e-1	4.47e-2	4.79e-2	1.89e-1	2.83e-1
2%	MAENT	**0.00e+0**	2.32e+3	1.14e+3	1.69e-1	4.81e-2	**6.37e-2**	2.53e-1	4.77e-1
	DQN	**0.00e+0**	2.06e+3	1.15e+4	1.70e-1	**4.81e-2**	**6.37e-2**	**2.51e-1**	4.81e-1
	PPO	**0.00e+0**	2.10e+3	7.01e+3	1.66e-1	4.72e-2	5.27e-2	2.48e-1	**4.90e-1**
	IMP_PPO	-5.70e+0	2.28e+2	2.65e+4	2.67e-1	4.63e-2	4.22e-2	2.09e-1	3.02e-1
	IMP_VTRACE	-5.42e+0	3.74e+2	1.01e+4	2.87e-1	4.58e-2	4.53e-2	1.99e-1	2.90e-1

Table 10: Performance of the pre-trained generalist agents across all fine-tuning data budget on Tomcat project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
Baseline	RLO	NA	4.12e+2	9.26e+7	2.49e-1	4.76e-2	4.07e-2	2.41e-1	**4.83e-1**
	RLO_Ent	NA	4.07e+2	1.03e+8	3.46e-1	4.40e-2	3.59e-2	2.04e-1	4.50e-1
zero-shot	IMPALA	-5.73e+0	4.20e+2	NA	2.45e-1	**4.91e-2**	4.67e-2	2.18e-1	3.00e-1
	MGDT	7.27e-1	7.81e+2	NA	7.57e-1	4.42e-2	2.99e-2	2.16e-1	4.49e-1
1%	MAENT	7.21e-1	7.75e+2	**6.71e+2**	7.54e-1	4.46e-2	2.99e-2	2.22e-1	4.55e-1
	DQN	7.27e-1	7.69e+2	5.89e+3	7.57e-1	4.42e-2	2.99e-2	2.16e-1	4.49e-1
	PPO	**8.00e-1**	7.83e+2	3.61e+3	**7.71e-1**	4.28e-2	2.40e-2	2.04e-1	4.31e-1
	IMP_PPO	-5.61e+0	3.99e+2	1.54e+4	1.84e-1	4.49e-2	4.67e-2	1.98e-1	2.78e-1
	IMP_VTRACE	-5.72e+0	4.18e+2	6.83e+3	2.34e-1	4.74e-2	**5.03e-2**	1.97e-1	2.78e-1
2%	MAENT	7.13e-1	7.79e+2	1.16e+3	7.57e-1	4.50e-2	2.99e-2	**2.28e-1**	4.67e-1
	DQN	7.27e-1	7.67e+2	1.16e+4	7.57e-1	4.42e-2	2.99e-2	2.16e-1	4.49e-1
	PPO	**8.00e-1**	7.80e+2	7.05e+3	**7.71e-1**	4.28e-2	2.40e-2	2.04e-1	4.31e-1
	IMP_PPO	-5.82e+0	1.40e+2	2.76e+4	2.10e-1	4.16e-2	4.92e-2	1.89e-1	2.67e-1
	IMP_VTRACE	-5.47e+0	**7.63e+1**	1.25e+4	2.13e-1	4.34e-2	4.43e-2	1.91e-1	2.75e-1

Table 11: Performance of the pre-trained generalist agents across all fine-tuning data budget on Eclipse project and the baseline in (in bold are the values of generalist agents configurations with greater performance).

Data Budgets	Algorithms	Reward	TE time	TA time	MRR	MAP	Top 1	Top 5	Top 10
Baseline	RLO	NA	1.37e+3	1.22e+8	2.23e-1	4.39e-2	4.16e-2	2.21e-1	4.33e-1
	RLO_Ent	NA	1.37e+3	1.11e+8	4.33e-1	4.49e-2	4.47e-2	2.36e-1	4.51e-1
zero-shot	IMPALA	-5.50e+0	1.12e+3	NA	2.51e-1	4.14e-2	3.80e-2	1.80e-1	2.64e-1
	MGDT	7.33e-1	2.49e+3	NA	4.46e-1	4.56e-2	5.29e-2	2.31e-1	4.49e-1
1%	MAENT	7.31e-1	2.51e+3	**6.76e+2**	4.48e-1	4.54e-2	5.10e-2	2.27e-1	4.49e-1
	DQN	7.33e-1	2.50e+3	5.79e+3	4.46e-1	4.56e-2	5.29e-2	2.31e-1	4.49e-1
	PPO	**7.45e-1**	2.54e+3	3.70e+3	4.78e-1	**4.73e-2**	**6.67e-2**	**2.37e-1**	**4.67e-1**
	IMP_PPO	-5.72e+0	1.10e+3	1.57e+4	2.30e-1	4.39e-2	4.16e-2	1.94e-1	2.88e-1
	IMP_VTRACE	-5.28e+0	1.07e+3	8.44e+3	3.14e-1	4.60e-2	5.06e-2	1.95e-1	2.85e-1
2%	MAENT	7.29e-1	2.50e+3	1.20e+3	4.51e-1	4.52e-2	5.10e-2	2.29e-1	4.47e-1
	DQN	7.33e-1	2.49e+3	1.14e+4	4.46e-1	4.56e-2	5.29e-2	2.31e-1	4.49e-1
	PPO	**7.45e-1**	2.60e+3	7.26e+3	**4.78e-1**	**4.73e-2**	**6.67e-2**	**2.37e-1**	**4.67e-1**
	IMP_PPO	-5.71e+0	2.67e+2	3.50e+4	2.81e-1	4.10e-2	3.96e-2	1.80e-1	2.66e-1
	IMP_VTRACE	-5.99e+0	**2.64e+2**	1.94e+4	2.02e-1	4.33e-2	4.08e-2	1.93e-1	2.87e-1

Table 12: Performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on MsPacman game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environment			MsPacman			
Metrics			Type 1	Type 2	Type 3	Type 4
Baseline agent		mean	**3.05e+3**	**2.59e+3**	**5.43e+3**	**6.04e+3**
		std	**1.53e+3**	**1.07e+3**	**2.72e+3**	**3.30e+3**
		median	**3.37e+3**	**2.75e+3**	**4.43e+3**	**4.43e+3**
MGDT	MAENT	mean	0.0	0.0	4.10e+4	3.06e+4
		std	0.0	0.0	3.34e+4	2.75e+4
		median	0.0	0.0	2.28e+4	1.15e+4
	DQN	mean	0.0	0.0	1.91e+4	5.77e+4
		std	0.0	0.0	1.98e+4	5.34e+4
		median	0.0	0.0	9.68e+3	5.67e+4
	PPO	mean	0.0	0.0	0.0	0.0
		std	0.0	0.0	0.0	0.0
		median	0.0	0.0	0.0	0.0
IMPALA	V_TRACE	mean	8.40e+3	8.45e+3	1.02e+4	1.04e+4
		std	4.92e+3	6.40e+3	4.34e+3	4.78e+3
		median	6.95e+3	6.29e+3	8.71e+3	9.33e+3
	PPO	mean	5.20e+3	5.55e+3	8.48e+3	8.58e+3
		std	5.36e+3	5.79e+3	6.73e+3	6.69e+3
		median	3.41e+3	3.79e+3	6.95e+3	6.94e+3

- Table 34 reports the results of post-hoc test analysis for the baseline and the generalist agents on the MsPacman game involving their training time performance. At 1%fine-tuning data budget, MGDT agents achieve greater performance compared to the generalist agents.

Table 13: Performance of the pre-trained generalist agents at 2% fine-tuning data budget on MsPacman game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environment			MsPacman			
Metrics			Type 1	Type 2	Type 3	Type 4
Baseline agent		mean	**3.05e+3**	**2.59e+3**	**5.43e+3**	**6.04e+3**
		std	**1.53e+3**	**1.07e+3**	**2.72e+3**	**3.30e+3**
		median	**3.37e+3**	**2.75e+3**	**4.43e+3**	**4.43e+3**
MGDT	MAENT	mean	0.0	0.0	3.89e+4	3.99e+4
		std	0.0	0.0	3.01e+4	3.08e+4
		median	0.0	0.0	2.53e+4	2.77e+4
	DQN	mean	0.0	0.0	3.28e+4	3.28e+4
		std	0.0	0.0	2.30e+4	2.30e+4
		median	0.0	0.0	1.66e+4	1.66e+4
	PPO	mean	0.0	0.0	0.0	0.0
		std	0.0	0.0	0.0	0.0
		median	0.0	0.0	0.0	0.0
IMPALA	V_TRACE	mean	5.46e+3	6.24e+3	8.97e+3	8.56e+3
		std	5.39e+3	5.56e+3	7.31e+3	7.42e+3
		median	4.30e+3	4.95e+3	6.99e+3	6.16e+3
	PPO	mean	5.26e+3	5.71e+3	9.18e+3	9.10e+3
		std	5.11e+3	5.53e+3	7.70e+3	7.38e+3
		median	3.48e+3	3.90e+3	7.18e+3	7.16e+3

Table 14: Performance of the pre-trained generalist agents at 1% fine-tuning data budget on MsPacman game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environment			MsPacman			
Metrics			Type 1	Type 2	Type 3	Type 4
Baseline agent		mean	**3.05e+3**	**2.59e+3**	**5.43e+3**	**6.04e+3**
		std	**1.53e+3**	**1.07e+3**	**2.72e+3**	**3.30e+3**
		median	**3.37e+3**	**2.75e+3**	**4.43e+3**	**4.43e+3**
MGDT	MAENT	mean	0.0	0.0	1.07e+4	4.28e+4
		std	0.0	0.0	7.45e+3	4.11e+4
		median	0.0	0.0	1.07e+4	3.50e+4
	DQN	mean	0.0	0.0	2.51e+4	2.51e+4
		std	0.0	0.0	8.06e+3	8.06e+3
		median	0.0	0.0	2.51e+4	2.51e+4
	PPO	mean	0.0	0.0	0.0	0.0
		std	0.0	0.0	0.0	0.0
		median	0.0	0.0	0.0	0.0
IMPALA	V_TRACE	mean	6.18e+3	6.67e+3	9.44e+3	9.34e+3
		std	5.38e+3	5.37e+3	7.56e+3	7.30e+3
		median	4.46e+3	4.95e+3	6.98e+3	6.98e+3
	PPO	mean	5.90e+3	6.44e+3	9.43e+3	9.85e+3
		std	5.34e+3	5.77e+3	7.28e+3	7.27e+3
		median	4.23e+3	4.51e+3	8.06e+3	8.44e+3

Table 15: Performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on the Blockmaze game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environment			Blockmaze	
Metrics			Type 1	Type 2
Baseline agent		mean	**1.46e+1**	**1.22e+1**
		std	**2.26e+1**	**2.67e+1**
		median	**5.88e+0**	**3.22e+0**
MGDT	MAENT	mean	3.00e+3	3.53e+3
		std	2.81e+3	3.28e+3
		median	2.07e+3	3.54e+3
	DQN	mean	3.00e+3	3.53e+3
		std	2.81e+3	3.28e+3
		median	2.07e+3	3.54e+3
	PPO	mean	3.00e+3	3.53e+3
		std	2.81e+3	3.28e+3
		median	2.07e+3	3.54e+3
IMPALA	V_TRACE	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0
	PPO	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0

Table 16: Performance of the pre-trained generalist agents at 1% fine-tuning data budget on the Blockmaze game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environments			Blockmaze	
Metrics			Type 1	Type 2
Baseline agent		mean	**1.46e+1**	**1.22e+1**
		std	**2.26e+1**	**2.67e+1**
		median	**5.88e+0**	**3.22e+0**
MGDT	MAENT	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0
	DQN	mean	1.04e+2	7.46e+1
		std	7.84e+1	1.01e+2
		median	9.18e+1	3.91e+1
	PPO	mean	2.97e+1	4.29e+1
		std	6.37e+0	3.19e+1
		median	3.01e+1	4.00e+1
IMPALA	V_TRACE	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0
	PPO	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0

Table 17: Performance of the pre-trained generalist agents at 2% fine-tuning data budget on the Blockmaze game and the baseline in terms of time to detect bugs (in bold are the values of generalist agents configurations with greater performance).

Environment			Blockmaze	
Metrics			Type 1	Type 2
Baseline agent		mean	1.46e+1	1.22e+1
		std	2.26e+1	2.67e+1
		median	5.88e+0	3.22e+0
MGDT	**MAENT**	mean	0.0	**1.13e+1**
		std	0.0	**9.24e-1**
		median	0.0	**1.19e+1**
	DQN	mean	**1.42e+1**	4.78e+1
		std	**4.31e+0**	6.90e+1
		median	**1.41e+1**	1.66e+1
	PPO	mean	2.83e+1	3.99e+1
		std	5.66e+0	2.87e+1
		median	2.84e+1	3.92e+1
IMPALA	V_TRACE	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0
	PPO	mean	0.0	0.0
		std	0.0	0.0
		median	0.0	0.0

Table 18: Cumulative reward, training, and testing times of the fine-tuned generalist agents on the Blockmaze game on 2% data budget and the baseline (in bold are the values of generalist agents configurations with greater performance).

Environment			Blockmaze		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	4.32e+4	**8.82e+2**	NA
		std	0.0	**2.76e+2**	NA
		median	4.32e+4	**9.60e+2**	NA
MGDT	**MAENT**	mean	**8.64e+2**	7.91e+3	-2.97e+2
		std	**0.0**	2.13e+2	0.0
		median	**8.64e+2**	7.99e+3	-2.97e+2
	DQN	mean	**8.64e+2**	2.19e+3	-1.49e+2
		std	**0.0**	8.95e+1	0.0
		median	**8.64e+2**	2.17e+3	-1.49e+2
	PPO	mean	**8.64e+2**	8.68e+3	**-1.22e+2**
		std	**0.0**	1.59e+2	**1.42e-14**
		median	**8.64e+2**	8.60e+3	**-1.22e+2**
IMPALA	**V_TRACE**	mean	**8.64e+2**	6.62e+3	-3.96e+2
		std	**0.0**	1.48e+2	3.99e+0
		median	**8.64e+2**	6.59e+3	-3.96e+2
	PPO	mean	**8.64e+2**	7.38e+3	-4.00e+2
		std	**0.0**	2.87e+2	0.0
		median	**8.64e+2**	7.38e+3	-4.00e+2

TA time and TE time refer to training time and testing time respectively.

Table 19: Performance of the fine-tuned generalist agents on the PDR-based scheduling on 2% data budget and the baseline in terms of makespan time (in bold are the values of generalist agent configurations with greater performance).

Environment			PDR	
Metrics			(6 x 6)	(30 x 20)
Baseline agent		mean	5.73e+2	2.48e+3
		std	3.27e+0	1.07e+0
		median	5.74e+2	2.47e+3
MGDT	**MAENT**	mean	**5.00e+2**	2.01e+3
		std	**0.0**	5.93e+0
		median	**5.00e+2**	2.01e+3
	DQN	mean	5.02e+2	2.01e+3
		std	5.68e-14	5.06e+0
		median	5.02e+2	2.01e+3
	PPO	mean	5.01e+2	2.01e+3
		std	0.0	4.23e+0
		median	5.01e+2	2.01e+3
IMPALA	V_TRACE	mean	5.01e+2	2.01e+3
		std	3.73e+0	1.70e+0
		median	5.04e+2	2.01e+3
	PPO	mean	5.03e+2	**2.01e+3**
		std	4.98e+0	**1.11e+0**
		median	5.02e+2	**2.01e+3**

Table 20: Cumulative reward, training, and testing times of the fine-tuned generalist agents on MsPacman game on 2% data budget and the baseline (in bold are the values of generalist agents configurations with greater performance).

Environment			MsPacman		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	**8.14e+3**	**2.57e+4**	NA
		std	**3.57e+2**	**1.08e+3**	NA
		median	**7.99e+3**	**2.55e+4**	NA
MGDT	**MAENT**	mean	8.70e+3	3.92e+5	**2.00e+2**
		std	3.50e+3	5.52e+4	**4.33e+1**
		median	9.19e+3	3.69e+5	**1.88e+2**
	DQN	mean	9.09e+3	3.66e+5	1.50e+2
		std	2.47e+3	1.62e+4	1.91e+0
		median	1.10e+4	3.62e+5	1.51e+2
	PPO	mean	1.27e+4	5.99e+5	4.50e+1
		std	3.29e+3	1.54e+4	0.0
		median	1.11e+4	6.05e+5	4.50e+1
IMPALA	V_TRACE	mean	9.92e+3	3.21e+4	8.90e+1
		std	4.16e+2	3.92e+3	1.86e+0
		median	1.00e+4	3.36e+4	8.88e+1
	PPO	mean	1.02e+4	3.24e+4	1.82e+2
		std	7.44e+2	2.36e+3	1.76e+1
		median	1.00e+4	3.19e+4	1.73e+2

TA time and TE time refer to training time and testing time respectively.

Table 21: Cumulative reward, training, and testing times of the fine-tuned generalist agents on the PDR-based scheduling on 2% data budget and the baseline (in bold are the values of generalist agents configurations with greater performance).

Environment			PDR					
			(6 x 6)			(30 x 20)		
Metrics			TA time	TE time	Cumulative reward	TA time	TE time	Cumulative reward
Baseline agent		mean	3.96e+3	**1.05e+1**	-5.73e+2	7.42e+4	**1.73e+2**	-2.48e+3
		std	1.37e+3	**1.65e+0**	3.27e+0	2.42e+4	**4.70e+0**	1.07e+0
		median	4.58e+3	**9.64e+0**	-5.74e+2	8.79e+4	**1.73e+2**	-2.47e+3
MGDT	**MAENT**	mean	1.30e+3	1.44e+2	**-3.91e+2**	4.20e+4	2.57e+3	**-1.26e+3**
		std	1.30e+1	6.52e+0	**0.0**	1.31e+4	8.56e+1	**1.20e+0**
		median	1.31e+3	1.40e+2	**-3.91e+2**	4.77e+4	2.53e+3	**-1.26e+3**
	DQN	mean	4.57e+3	1.41e+2	-3.92e+2	1.43e+5	2.50e+3	-1.26e+3
		std	3.63e+1	2.12e+0	0.0	2.04e+4	5.64e+1	1.71e-1
		median	4.56e+3	1.42e+2	-3.92e+2	1.40e+5	2.53e+3	-1.26e+3
	PPO	mean	1.56e+4	1.46e+2	-3.93e+2	2.77e+5	2.06e+3	-1.26e+3
		std	7.15e+1	6.49e-1	0.0	8.26e+4	5.30e+2	3.47e-1
		median	1.56e+4	1.46e+2	-3.93e+2	3.35e+5	2.46e+3	-1.26e+3
IMPALA	**V_TRACE**	mean	1.20e+2	4.94e+1	-3.92e+2	**2.60e+3**	9.73e+3	-1.43e+3
		std	6.07e+0	9.60e-1	6.27e-1	**9.99e+2**	3.71e+2	3.00e+2
		median	1.22e+2	4.99e+1	-3.92e+2	**2.05e+3**	9.75e+3	-1.26e+3
	PPO	mean	**1.10e+2**	4.85e+1	-3.92e+2	2.87e+3	9.50e+3	-1.26e+3
		std	**5.67e+0**	1.15e+0	4.26e-1	2.49e+2	4.45e+2	2.55e-1
		median	**1.07e+2**	4.88e+1	-3.92e+2	2.76e+3	9.75e+3	-1.26e+3

TA time and TE time refer to training time and testing time respectively.

Table 22: Performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on the PDR-based scheduling and the baseline in terms of makespan time (in bold are the values of generalist agents configurations with greater performance).

Environment			PDR	
Metrics			(6 x 6)	(30 x 20)
Baseline agent		mean	5.73e+2	2.48e+3
		std	3.27e+0	1.07e+0
		median	5.74e+2	2.47e+3
MGDT	MAENT	mean	**5.00e+2**	2.02e+3
		std	**5.68e-14**	0.0
		median	**5.00e+2**	2.02e+3
	DQN	mean	**5.00e+2**	2.02e+3
		std	**5.68e-14**	0.0
		median	**5.00e+2**	2.02e+3
	PPO	mean	**5.00e+2**	2.02e+3
		std	**5.68e-14**	0.0
		median	**5.00e+2**	2.02e+3
IMPALA	**V_TRACE**	mean	5.04e+2	**2.01e+3**
		std	4.63e+0	**3.10e+0**
		median	5.04e+2	**2.01e+3**
	PPO	mean	5.04e+2	**2.01e+3**
		std	4.63e+0	**3.10e+0**
		median	5.04e+2	**2.01e+3**

Table 23: Cumulative reward, training and testing times performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on MsPacman game and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			MsPacman		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	8.14e+3	**2.57e+4**	NA
		std	3.57e+2	**1.08e+3**	NA
		median	7.99e+3	**2.55e+4**	NA
MGDT	MAENT	mean	NA	3.51e+5	1.93e+2
		std	NA	1.84e+4	4.21e+1
		median	NA	3.57e+5	2.11e+2
	DQN	mean	NA	3.90e+5	1.52e+2
		std	NA	1.29e+4	3.52e+0
		median	NA	4.00e+5	1.52e+2
	PPO	mean	NA	6.77e+5	4.50e+1
		std	NA	2.29e+4	0.0
		median	NA	6.76e+5	4.50e+1
IMPALA	**V_TRACE**	mean	NA	3.89e+4	**1.96e+2**
		std	NA	6.09e+3	**3.59e+1**
		median	NA	3.79e+4	**1.84e+2**
	PPO	mean	NA	2.96e+4	8.84e+1
		std	NA	3.60e+3	7.13e-1
		median	NA	2.73e+4	8.83e+1

TA time and TE time refer to training time and testing time respectively.

Table 24: Cumulative reward, training and testing times performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on the Blockmaze game and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			Blockmaze		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	4.32e+4	**8.82e+2**	NA
		std	0.0	**2.76e+2**	NA
		median	4.32e+4	**9.60e+2**	NA
MGDT	**MAENT**	mean	NA	2.15e+3	**-1.49e+2**
		std	NA	7.78e+1	**0.0**
		median	NA	2.16e+3	**-1.49e+2**
	DQN	mean	NA	2.15e+3	**-1.49e+2**
		std	NA	7.78e+1	**0.0**
		median	NA	2.16e+3	**-1.49e+2**
	PPO	mean	NA	2.15e+3	**-1.49e+2**
		std	NA	7.78e+1	**0.0**
		median	NA	2.16e+3	**-1.49e+2**
IMPALA	V_TRACE	mean	NA	7.77e+3	-4.00e+2
		std	NA	1.69e+3	0.0
		median	NA	6.79e+3	-4.00e+2
	PPO	mean	NA	7.77e+3	-4.00e+2
		std	NA	1.69e+3	0.0
		median	NA	6.79e+3	-4.00e+2

TA time and TE time refer to training time and testing time respectively.

Table 25: Cumulative reward, training and testing times performance of the pre-trained generalist agents at zero-shot fine-tuning data budget on the PDR-based scheduling and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			PDR					
			(6 x 6)			(30 x 20)		
Metrics			TA time	TE time	Cumulative reward	TA time	TE time	Cumulative reward
Baseline agent		mean	3.96e+3	**1.05e+1**	-5.73e+2	7.42e+4	**1.73e+2**	-2.48e+3
		std	1.37e+3	**1.65e+0**	3.27e+0	2.42e+4	**4.70e+0**	1.07e+0
		median	4.58e+3	**9.64e+0**	-5.74e+2	8.79e+4	**1.73e+2**	-2.47e+3
MGDT	**MAENT**	mean	NA	1.59e+2	**-3.91e+2**	NA	2.70e+3	**-1.26e+3**
		std	NA	6.50e+0	**0.0**	NA	1.58e+2	**0.0**
		median	NA	1.59e+2	**-3.91e+2**	NA	2.66e+3	**-1.26e+3**
	DQN	mean	NA	1.59e+2	**-3.91e+2**	NA	2.70e+3	**-1.26e+3**
		std	NA	6.50e+0	**0.0**	NA	1.58e+2	**0.0**
		median	NA	1.59e+2	**-3.91e+2**	NA	2.66e+3	**-1.26e+3**
	PPO	mean	NA	1.59e+2	**-3.91e+2**	NA	2.70e+3	**-1.26e+3**
		std	NA	6.50e+0	**0.0**	NA	1.58e+2	**0.0**
		median	NA	1.59e+2	**-3.91e+2**	NA	2.66e+3	**-1.26e+3**
IMPALA	V_TRACE	mean	NA	5.13e+1	-3.92e+2	NA	1.05e+4	-1.26e+3
		std	NA	5.46e+0	4.78e-1	NA	9.81e+2	1.08e+0
		median	NA	4.88e+1	-3.92e+2	NA	1.01e+4	-1.26e+3
	PPO	mean	NA	5.13e+1	-3.92e+2	NA	1.05e+4	-1.26e+3
		std	NA	5.46e+0	4.78e-1	NA	9.81e+2	1.08e+0
		median	NA	4.88e+1	-3.92e+2	NA	1.01e+4	-1.26e+3

TA time and TE time refer to training time and testing time respectively.

Table 26: Performance of the fine-tuned generalist agents on the PDR-based scheduling on 1% data budget and the baseline in terms of makespan time (in bold are the values of generalist agent configurations with greater performance).

Environment			PDR	
Metrics			(6 x 6)	(30 x 20)
Baseline agent		mean	5.73e+2	2.48e+3
		std	3.27e+0	1.07e+0
		median	5.74e+2	2.47e+3
MGDT	**MAENT**	mean	**5.00e+2**	**2.01e+3**
		std	**0.0**	**0.0**
		median	**5.00e+2**	**2.01e+3**
	DQN	mean	5.06e+2	2.01e+3
		std	5.68e-14	3.03e+0
		median	5.06e+2	2.01e+3
	PPO	mean	5.05e+2	2.01e+3
		std	0.0	6.86e-1
		median	5.05e+2	2.01e+3
IMPALA	V_TRACE	mean	5.04e+2	2.01e+3
		std	3.46e+0	2.47e+0
		median	5.04e+2	2.01e+3
	PPO	mean	5.04e+2	2.01e+3
		std	3.28e+0	3.83e+0
		median	5.04e+2	2.01e+3

Table 27: Cumulative reward, training, and testing times performance of the fine-tuned generalist agents on MsPacman game on 1% data budget and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			MsPacman		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	8.14e+3	**2.57e+4**	NA
		std	3.57e+2	**1.08e+3**	NA
		median	7.99e+3	**2.55e+4**	NA
MGDT	**MAENT**	mean	4.87e+3	4.23e+5	**1.69e+2**
		std	2.67e+2	6.37e+4	**2.93e+1**
		median	4.84e+3	4.41e+5	**1.53e+2**
	DQN	mean	**4.06e+3**	3.50e+5	1.51e+2
		std	**1.09e+3**	3.68e+4	2.63e+0
		median	**3.69e+3**	3.23e+5	1.53e+2
	PPO	mean	5.46e+3	5.74e+5	4.50e+1
		std	6.16e+1	5.93e+3	0.0
		median	5.48e+3	5.76e+5	4.50e+1
IMPALA	V_TRACE	mean	1.02e+4	3.01e+4	8.79e+1
		std	2.33e+2	1.65e+3	9.66e-1
		median	1.03e+4	2.99e+4	8.80e+1
	PPO	mean	1.09e+4	3.01e+4	8.84e+1
		std	1.41e+3	1.68e+3	9.31e-1
		median	1.04e+4	3.08e+4	8.83e+1

TA time and TE time refer to training time and testing time respectively.

Table 28: Cumulative reward, training and testing times performance of the fine-tuned generalist agents on the Blockmaze game on 1% data budget and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			Blockmaze		
Metrics			TA time	TE time	Cumulative reward
Baseline agent		mean	4.32e+4	**8.82e+2**	NA
		std	0.0	**2.76e+2**	**NA**
		median	4.32e+4	**9.60e+2**	NA
MGDT	**MAENT**	mean	**4.32e+2**	8.11e+3	-1.42e+2
		std	**0.0**	2.20e+2	7.76e+1
		median	**4.32e+2**	8.05e+3	-1.03e+2
	DQN	mean	**4.32e+2**	2.17e+3	-1.49e+2
		std	**0.0**	6.87e+1	3.30e-1
		median	**4.32e+2**	2.18e+3	-1.49e+2
	PPO	mean	**4.32e+2**	9.13e+3	**-1.22e+2**
		std	**0.0**	7.17e+2	**1.07e-1**
		median	**4.32e+2**	8.68e+3	**-1.22e+2**
IMPALA	**V_TRACE**	mean	**4.32e+2**	6.58e+3	-3.96e+2
		std	**0.0**	1.64e+2	3.99e+0
		median	**4.32e+2**	6.50e+3	-3.99e+2
	PPO	mean	**4.32e+2**	6.66e+3	-4.00e+2
		std	**0.0**	4.93e+2	0.0
		median	**4.32e+2**	6.73e+3	-4.00e+2

TA time and TE time refer to training time and testing time respectively.

Table 29: Cumulative reward, training and testing times performance of the fine-tuned generalist agents on the PDR-based scheduling on 1% data budget and the baseline (in bold are the values of generalist agent configurations with greater performance).

Environment			PDR					
			(6 x 6)			(30 x 20)		
Metrics			TA time	TE time	Cumulative reward	TA time	TE time	Cumulative reward
Baseline agent		mean	3.96e+3	**1.05e+1**	-5.73e+2	7.42e+4	**1.73e+2**	-2.48e+3
		std	1.37e+3	**1.65e+0**	3.27e+0	2.42e+4	**4.70e+0**	1.07e+0
		median	4.58e+3	**9.64e+0**	-5.74e+2	8.79e+4	**1.73e+2**	-2.47e+3
MGDT	**MAENT**	mean	6.42e+2	1.38e+2	**-3.91e+2**	1.80e+4	3.01e+3	**-1.26e+3**
		std	9.04e+0	6.63e-1	**0.0**	9.47e+2	7.29e+2	**0.0**
		median	6.48e+2	1.38e+2	**-3.91e+2**	1.82e+4	2.69e+3	**-1.26e+3**
	DQN	mean	2.28e+3	1.40e+2	**-3.91e+2**	6.77e+4	2.47e+3	-1.26e+3
		std	7.92e+0	2.49e+0	**0.0**	3.26e+3	4.65e+1	9.26e-1
		median	2.28e+3	1.39e+2	**-3.91e+2**	6.66e+4	2.44e+3	-1.26e+3
	PPO	mean	7.74e+3	1.41e+2	-3.93e+2	1.59e+5	2.47e+3	-1.26e+3
		std	4.76e+1	1.91e+0	0.0	5.69e+3	1.81e+1	1.37e+0
		median	7.72e+3	1.41e+2	-3.93e+2	1.60e+5	2.46e+3	-1.26e+3
IMPALA	**V_TRACE**	mean	**1.11e+2**	4.86e+1	-3.91e+2	**1.31e+3**	1.02e+4	-1.44e+3
		std	**2.31e+0**	7.93e-1	5.60e-1	**4.31e+2**	3.82e+2	3.04e+2
		median	**1.09e+2**	4.84e+1	-3.91e+2	**1.07e+3**	1.02e+4	-1.26e+3
	PPO	mean	1.12e+2	4.76e+1	-3.92e+2	1.37e+3	9.35e+3	-1.26e+3
		std	4.38e+0	2.00e+0	2.08e-1	3.55e+1	4.54e+2	7.16e-1
		median	1.12e+2	4.67e+1	-3.92e+2	1.38e+3	9.47e+3	-1.26e+3

TA time and TE time refer to training time and testing time respectively.

Table 30: Results of post-hoc tests analysis of testing time performance by the baseline specialist and the MGDT generalist agent on the Blockmaze game (in bold are DRL configurations where p-value is < 0.05 and have greater performance w.r.t the effect size).

Data budgets	A	B	mean(A)	mean(B)	pval	CLES
zero-shot	**Baseline**	MGDT-DQN	**8.82e+2**	2.15e+3	**1.67e-3**	**3.79e-5**
	Baseline	MGDT-MAENT	**8.82e+2**	2.15e+3	**1.67e-3**	**3.79e-5**
	Baseline	MGDT-PPO	**8.82e+2**	2.15e+3	**1.67e-3**	**3.79e-5**
	MGDT-DQN	MGDT-MAENT	2.15e+3	2.15e+3	1.00e+0	5.00e-1
	MGDT-DQN	MGDT-PPO	2.15e+3	2.15e+3	1.00e+0	5.00e-1
	MGDT-MAENT	MGDT-PPO	2.15e+3	2.15e+3	1.00e+0	5.00e-1
1%	**Baseline**	MGDT-DQN	**8.82e+2**	2.17e+3	**2.11e-3**	**1.74e-4**
	Baseline	MGDT-MAENT	**8.82e+2**	8.11e+3	**1.53e-9**	**2.13e-75**
	Baseline	MGDT-PPO	**8.82e+2**	9.13e+3	**1.20e-5**	**3.87e-22**
	MGDT-DQN	MGDT-MAENT	**2.17e+3**	8.11e+3	**1.02e-6**	**8.60e-94**
	MGDT-DQN	MGDT-PPO	**2.17e+3**	9.13e+3	**1.40e-4**	**3.17e-14**
	MGDT-MAENT	MGDT-PPO	**8.11e+3**	9.13e+3	**1.43e-1**	**1.12e-1**
2%	**Baseline**	MGDT-DQN	**8.82e+2**	2.19e+3	**1.27e-3**	**2.77e-5**
	Baseline	MGDT-MAENT	**8.82e+2**	7.91e+3	**2.17e-9**	**6.68e-73**
	Baseline	MGDT-PPO	**8.82e+2**	8.68e+3	**7.76e-9**	**1.97e-106**
	MGDT-DQN	MGDT-MAENT	**2.19e+3**	7.91e+3	**9.61e-8**	**8.53e-109**
	MGDT-DQN	MGDT-PPO	**2.19e+3**	8.68e+3	**9.43e-10**	**9.86e-222**
	MGDT-MAENT	MGDT-PPO	**7.91e+3**	8.68e+3	**2.37e-3**	**4.78e-3**

mean(A) and mean(B) refer to testing time values.

Table 31: Results of post-hoc tests analysis of makespan performance by the baseline and generalist agents on the 30×20 instance (in bold are DRL configurations where p-value is < 0.05 and have greater performance w.r.t the effect size).

Data budgets	A	B	mean(A)	mean(B)	pval	CLES
zero-shot	Baseline	**IMPALA-PPO**	2.48e+3	**2.01e+3**	**7.31e-11**	**1.00e+0**
	Baseline	**IMPALA-V_TRACE**	2.48e+3	**2.01e+3**	**7.31e-11**	**1.00e+0**
	Baseline	**MGDT-DQN**	2.48e+3	**2.02e+3**	**2.51e-11**	**1.00e+0**
	Baseline	**MGDT-MAENT**	2.48e+3	**2.02e+3**	**2.51e-11**	**1.00e+0**
	Baseline	**MGDT-PPO**	2.48e+3	**2.02e+3**	**2.51e-11**	**1.00e+0**
	IMPALA-PPO	IMPALA-V_TRACE	2.01e+3	2.01e+3	1.00e+0	5.00e-1
	IMPALA-PPO	MGDT-DQN	2.01e+3	2.02e+3	9.61e-2	5.33e-2
	IMPALA-PPO	MGDT-MAENT	2.01e+3	2.02e+3	9.61e-2	5.33e-2
	IMPALA-V_TRACE	MGDT-DQN	2.01e+3	2.02e+3	9.61e-2	5.33e-2
	IMPALA-V_TRACE	MGDT-MAENT	2.01e+3	2.02e+3	9.61e-2	5.33e-2
	MGDT-DQN	MGDT-MAENT	2.02e+3	2.02e+3	1.00e+0	5.00e-1
	MGDT-DQN	MGDT-PPO	2.02e+3	2.02e+3	1.00e+0	5.00e-1
	MGDT-MAENT	MGDT-PPO	2.02e+3	2.02e+3	1.00e+0	5.00e-1
1%	Baseline	**IMPALA-PPO**	2.48e+3	**2.01e+3**	**9.18e-10**	**1.00e+0**
	Baseline	**IMPALA-V_TRACE**	2.48e+3	**2.01e+3**	**8.12e-9**	**1.00e+0**
	Baseline	**MGDT-DQN**	2.48e+3	**2.01e+3**	**7.16e-11**	**1.00e+0**
	Baseline	**MGDT-MAENT**	2.48e+3	**2.01e+3**	**2.64e-11**	**1.00e+0**
	Baseline	**MGDT-PPO**	2.48e+3	**2.01e+3**	**5.55e-16**	**1.00e+0**
	IMPALA-PPO	IMPALA-V_TRACE	2.01e+3	2.01e+3	1.00e+0	4.61e-1
	IMPALA-PPO	MGDT-DQN	2.01e+3	2.01e+3	9.77e-1	3.78e-1
	IMPALA-PPO	MGDT-MAENT	2.01e+3	2.01e+3	9.20e-1	6.61e-1
	IMPALA-PPO	MGDT-PPO	2.01e+3	2.01e+3	8.37e-1	3.02e-1
	IMPALA-V_TRACE	MGDT-DQN	2.01e+3	2.01e+3	9.89e-1	3.92e-1
	IMPALA-V_TRACE	MGDT-MAENT	2.01e+3	2.01e+3	6.51e-1	7.75e-1
	IMPALA-V_TRACE	MGDT-PPO	2.01e+3	2.01e+3	8.17e-1	2.84e-1
	MGDT-DQN	MGDT-MAENT	2.01e+3	2.01e+3	3.61e-1	8.48e-1
	MGDT-DQN	MGDT-PPO	2.01e+3	2.01e+3	9.99e-1	4.36e-1
	MGDT-MAENT	MGDT-PPO	**2.01e+3**	2.01e+3	**1.80e-3**	**7.19e-8**
2%	Baseline	**IMPALA-PPO**	2.48e+3	**2.01e+3**	**0.0**	**1.00e+0**
	Baseline	**IMPALA-V_TRACE**	2.48e+3	**2.01e+3**	**2.71e-11**	**1.00e+0**
	Baseline	**MGDT-DQN**	2.48e+3	**2.01e+3**	**8.67e-9**	**1.00e+0**
	Baseline	**MGDT-MAENT**	2.48e+3	**2.01e+3**	**2.52e-6**	**1.00e+0**
	Baseline	**MGDT-PPO**	2.48e+3	**2.01e+3**	**2.24e-9**	**1.00e+0**
	IMPALA-PPO	IMPALA-V_TRACE	2.01e+3	2.01e+3	2.18e-1	9.58e-2
	IMPALA-PPO	MGDT-DQN	2.01e+3	2.01e+3	9.95e-1	4.15e-1
	IMPALA-PPO	MGDT-MAENT	2.01e+3	2.01e+3	9.96e-1	5.80e-1
	IMPALA-PPO	MGDT-PPO	2.01e+3	2.01e+3	8.52e-1	3.07e-1
	IMPALA-V_TRACE	MGDT-DQN	2.01e+3	2.01e+3	9.76e-1	6.28e-1
	IMPALA-V_TRACE	MGDT-MAENT	2.01e+3	2.01e+3	7.81e-1	7.40e-1
	IMPALA-V_TRACE	MGDT-PPO	2.01e+3	2.01e+3	1.00e+0	5.52e-1
	MGDT-DQN	MGDT-MAENT	2.01e+3	2.01e+3	9.83e-1	6.19e-1
	MGDT-DQN	MGDT-PPO	2.01e+3	2.01e+3	9.99e-1	4.34e-1
	MGDT-MAENT	MGDT-PPO	2.01e+3	2.01e+3	9.07e-1	3.21e-1

mean(A) and mean(B) refer to makespan values.

Table 32: Results of post-hoc tests analysis of cumulative reward performance by the baseline and generalist agents on the PDR task on the 30×20 instance (in bold are DRL configurations where p-value is < 0.05 and have greater performance w.r.t the effect size).

Data budgets	A	B	mean(A)	mean(B)	pval	CLES
zero-shot	Baseline	**IMPALA-PPO**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	Baseline	**IMPALA-V_TRACE**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	Baseline	**MGDT-DQN**	-2.48e+3	**-1.26e+3**	**4.21e-12**	**0.0**
	Baseline	**MGDT-MAENT**	-2.48e+3	**-1.26e+3**	**4.21e-12**	**0.0**
	Baseline	**MGDT-PPO**	-2.48e+3	**-1.26e+3**	**4.21e-12**	**0.0**
	IMPALA-PPO	IMPALA-V_TRACE	-1.26e+3	-1.26e+3	1.00e+0	5.00e-1
	IMPALA-PPO	MGDT-DQN	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	IMPALA-PPO	MGDT-MAENT	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	IMPALA-PPO	MGDT-PPO	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	IMPALA-V_TRACE	MGDT-DQN	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	IMPALA-V_TRACE	MGDT-MAENT	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	IMPALA-V_TRACE	MGDT-PPO	-1.26e+3	-1.26e+3	1.00e+0	4.68e-1
	MGDT-DQN	MGDT-MAENT	-1.26e+3	-1.26e+3	1.00e+0	5.00e-1
	MGDT-DQN	MGDT-PPO	-1.26e+3	-1.26e+3	1.00e+0	5.00e-1
	MGDT-MAENT	MGDT-PPO	-1.26e+3	-1.26e+3	1.00e+0	5.00e-1
1%	Baseline	**IMPALA-PPO**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	Baseline	**IMPALA-V_TRACE**	-2.48e+3	**-1.44e+3**	**4.50e-2**	**2.53e-3**
	Baseline	**MGDT-DQN**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	Baseline	**MGDT-MAENT**	-2.48e+3	**-1.26e+3**	**4.19e-12**	**0.0**
	Baseline	**MGDT-PPO**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	IMPALA-PPO	IMPALA-V_TRACE	-1.26e+3	-1.44e+3	8.90e-1	6.84e-1
	IMPALA-PPO	MGDT-DQN	-1.26e+3	-1.26e+3	4.19e-1	8.15e-1
	IMPALA-PPO	MGDT-MAENT	-1.26e+3	-1.26e+3	5.62e-1	2.15e-1
	IMPALA-PPO	MGDT-PPO	-1.26e+3	-1.26e+3	8.35e-1	7.00e-1
	IMPALA-V_TRACE	MGDT-DQN	-1.44e+3	-1.26e+3	8.92e-1	3.17e-1
	IMPALA-V_TRACE	MGDT-MAENT	-1.44e+3	-1.26e+3	8.88e-1	3.15e-1
	IMPALA-V_TRACE	MGDT-PPO	-1.44e+3	-1.26e+3	8.91e-1	3.17e-1
	MGDT-DQN	MGDT-MAENT	-1.26e+3	-1.26e+3	9.37e-2	4.08e-2
	MGDT-DQN	MGDT-PPO	-1.26e+3	-1.26e+3	9.99e-1	4.43e-1
	MGDT-MAENT	MGDT-PPO	-1.26e+3	-1.26e+3	3.76e-1	8.43e-1
2%	Baseline	**IMPALA-PPO**	-2.48e+3	-1.26e+3	5.88e-14	0.0
	Baseline	**IMPALA-V_TRACE**	-2.48e+3	**-1.43e+3**	**4.27e-2**	**2.14e-3**
	Baseline	**MGDT-DQN**	-2.48e+3	**-1.26e+3**	**3.51e-13**	**0.0**
	Baseline	**MGDT-MAENT**	-2.48e+3	**-1.26e+3**	**0.0**	**0.0**
	Baseline	**MGDT-PPO**	-2.48e+3	**-1.26e+3**	**1.63e-14**	**0.0**
	IMPALA-PPO	IMPALA-V_TRACE	-1.26e+3	-1.43e+3	8.90e-1	6.84e-1
	IMPALA-PPO	MGDT-DQN	-1.26e+3	-1.26e+3	5.63e-1	2.22e-1
	IMPALA-PPO	MGDT-MAENT	-1.26e+3	-1.26e+3	9.61e-1	6.39e-1
	IMPALA-PPO	MGDT-PPO	-1.26e+3	-1.26e+3	7.70e-1	7.22e-1
	IMPALA-V_TRACE	MGDT-DQN	-1.43e+3	-1.26e+3	8.90e-1	3.16e-1
	IMPALA-V_TRACE	MGDT-MAENT	-1.43e+3	-1.26e+3	8.91e-1	3.08e-1
	IMPALA-V_TRACE	MGDT-PPO	-1.43e+3	-1.26e+3	8.91e-1	3.17e-1
	MGDT-DQN	MGDT-MAENT	-1.26e+3	-1.26e+3	8.45e-1	7.04e-1
	MGDT-DQN	MGDT-PPO	-1.26e+3	-1.26e+3	1.80e-1	8.97e-1
	MGDT-MAENT	MGDT-PPO	-1.26e+3	-1.26e+3	9.99e-1	4.36e-1

mean(A) and mean(B) refer to cumulative reward values.

Table 33: Results of post-hoc tests analysis of training time performance by the baseline and generalist agents on PDR task on the 30×20 instance (in bold are DRL configurations where p-value is < 0.05 and have greater performance w.r.t the effect size).

Data budgets	A	B	mean(A)	mean(B)	pval	CLES
1%	Baseline	**IMPALA-PPO**	7.42e+4	**1.37e+3**	**2.91e-2**	**9.94e-1**
	Baseline	**IMPALA-V_TRACE**	7.42e+4	**1.31e+3**	**2.90e-2**	**9.96e-1**
	Baseline	MGDT-DQN	7.42e+4	6.77e+4	9.94e-1	5.86e-1
	Baseline	MGDT-MAENT	7.42e+4	1.80e+4	6.90e-2	9.72e-1
	Baseline	MGDT-PPO	**7.42e+4**	1.59e+5	**1.46e-2**	**2.40e-3**
	IMPALA-PPO	IMPALA-V_TRACE	1.37e+3	1.31e+3	1.00e+0	5.43e-1
	IMPALA-PPO	MGDT-DQN	**1.37e+3**	6.77e+4	**1.31e-5**	**1.29e-74**
	IMPALA-PPO	MGDT-MAENT	**1.37e+3**	1.80e+4	**2.35e-5**	**1.02e-55**
	IMPALA-PPO	MGDT-PPO	**1.37e+3**	1.59e+5	**3.91e-6**	**2.63e-135**
	IMPALA-V_TRACE	MGDT-DQN	**1.31e+3**	6.77e+4	**8.78e-6**	**5.86e-82**
	IMPALA-V_TRACE	MGDT-MAENT	**1.31e+3**	1.80e+4	**7.11e-7**	**9.24e-50**
	IMPALA-V_TRACE	MGDT-PPO	**1.31e+3**	1.59e+5	**3.35e-6**	**6.59e-151**
	MGDT-DQN	**MGDT-MAENT**	6.77e+4	**1.80e+4**	**1.21e-5**	**1.00e+0**
	MGDT-DQN	MGDT-PPO	6.77e+4	1.59e+5	5.99e-7	1.27e-35
	MGDT-MAENT	MGDT-PPO	1.80e+4	1.59e+5	3.62e-6	7.92e-106
2%	Baseline	**IMPALA-PPO**	7.42e+4	**2.87e+3**	**3.13e-2**	**9.93e-1**
	Baseline	**IMPALA-V_TRACE**	7.42e+4	**2.60e+3**	**3.07e-2**	**9.95e-1**
	Baseline	MGDT-DQN	**7.42e+4**	1.43e+5	**2.85e-2**	**3.14e-2**
	Baseline	**MGDT-MAENT**	7.42e+4	**4.20e+4**	**3.56e-1**	**8.37e-1**
	Baseline	MGDT-PPO	**7.42e+4**	2.77e+5	**3.75e-2**	**1.81e-2**
	IMPALA-PPO	IMPALA-V_TRACE	2.87e+3	2.60e+3	9.95e-1	5.87e-1
	IMPALA-PPO	MGDT-DQN	**2.87e+3**	1.43e+5	**9.71e-4**	**3.67e-10**
	IMPALA-PPO	MGDT-MAENT	**2.87e+3**	4.20e+4	**2.30e-2**	**3.92e-3**
	IMPALA-PPO	MGDT-PPO	**2.87e+3**	2.77e+5	**1.55e-2**	**1.49e-3**
	IMPALA-V_TRACE	MGDT-DQN	**2.60e+3**	1.43e+5	**9.37e-4**	**3.18e-11**
	IMPALA-V_TRACE	MGDT-MAENT	**2.60e+3**	4.20e+4	**2.19e-2**	**2.33e-3**
	IMPALA-V_TRACE	MGDT-PPO	**2.60e+3**	2.77e+5	**1.54e-2**	**8.06e-4**
	MGDT-DQN	**MGDT-MAENT**	1.43e+5	**4.20e+4**	**6.79e-4**	**1.00e+0**
	MGDT-DQN	MGDT-PPO	1.43e+5	2.77e+5	1.57e-1	7.95e-2
	MGDT-MAENT	MGDT-PPO	**4.20e+4**	2.77e+5	**2.50e-2**	**5.94e-3**

mean(A) and mean(B) refer to training time values.

Table 34: Results of post-hoc tests analysis of training time performance by the baseline and the generalist agents on MsPacman game (in bold are DRL configurations where p-value is < 0.05 and have greater performance w.r.t the effect size).

Data budgets	A	B	mean(A)	mean(B)	pval	CLES
1%	**BASELINE**	IMPALA-PPO	**8.14e+3**	1.09e+4	**1.81e-1**	**5.93e-2**
	BASELINE	IMPALA-V_TRACE	**8.14e+3**	1.02e+4	**2.25e-4**	**4.99e-6**
	BASELINE	**MGDT-DQN**	8.14e+3	**4.06e+3**	**6.32e-3**	**9.99e-1**
	BASELINE	**MGDT-MAENT**	8.14e+3	**4.87e+3**	**9.07e-6**	**1.00e+0**
	BASELINE	**MGDT-PPO**	8.14e+3	**5.46e+3**	**5.20e-4**	**1.00e+0**
	IMPALA-PPO	IMPALA-V_TRACE	1.09e+4	1.02e+4	9.54e-1	6.45e-1
	IMPALA-PPO	**MGDT-DQN**	1.09e+4	**4.06e+3**	**4.78e-3**	**1.00e+0**
	IMPALA-PPO	**MGDT-MAENT**	1.09e+4	**4.87e+3**	**2.21e-2**	**1.00e+0**
	IMPALA-PPO	**MGDT-PPO**	1.09e+4	**5.46e+3**	**3.23e-2**	**9.99e-1**
	IMPALA-V_TRACE	**MGDT-DQN**	1.02e+4	**4.06e+3**	**1.45e-3**	**1.00e+0**
	IMPALA-V_TRACE	**MGDT-MAENT**	1.02e+4	**4.87e+3**	**1.96e-8**	**1.00e+0**
	IMPALA-V_TRACE	**MGDT-PPO**	1.02e+4	**5.46e+3**	**3.91e-6**	**1.00e+0**
	MGDT-DQN	MGDT-MAENT	4.06e+3	4.87e+3	6.99e-1	2.56e-1
	MGDT-DQN	MGDT-PPO	4.06e+3	5.46e+3	2.81e-1	1.25e-1
	MGDT-MAENT	MGDT-PPO	4.87e+3	5.46e+3	6.08e-2	2.84e-2
2%	**BASELINE**	IMPALA-PPO	**8.14e+3**	1.02e+4	**1.86e-2**	**1.18e-2**
	BASELINE	IMPALA-V_TRACE	**8.14e+3**	9.92e+3	**1.90e-3**	**1.86e-3**
	BASELINE	MGDT-DQN	**8.14e+3**	9.09e+3	**9.62e-1**	**3.67e-1**
	BASELINE	MGDT-MAENT	**8.14e+3**	8.70e+3	**9.99e-1**	**4.43e-1**
	BASELINE	MGDT-PPO	**8.14e+3**	1.27e+4	**2.38e-1**	**1.09e-1**
	IMPALA-PPO	IMPALA-V_TRACE	1.02e+4	9.92e+3	9.70e-1	6.28e-1
	IMPALA-PPO	MGDT-DQN	1.02e+4	9.09e+3	9.35e-1	6.54e-1
	IMPALA-PPO	MGDT-MAENT	1.02e+4	8.70e+3	9.41e-1	6.49e-1
	IMPALA-PPO	MGDT-PPO	1.02e+4	1.27e+4	6.97e-1	2.56e-1
	IMPALA-V_TRACE	MGDT-DQN	9.92e+3	9.09e+3	9.78e-1	6.16e-1
	IMPALA-V_TRACE	MGDT-MAENT	9.92e+3	8.70e+3	9.74e-1	6.21e-1
	IMPALA-V_TRACE	MGDT-PPO	9.92e+3	1.27e+4	5.98e-1	2.26e-1
	MGDT-DQN	MGDT-MAENT	9.09e+3	8.70e+3	1.00e+0	5.32e-1
	MGDT-DQN	MGDT-PPO	9.09e+3	1.27e+4	5.38e-1	2.16e-1
	MGDT-MAENT	MGDT-PPO	8.70e+3	1.27e+4	5.83e-1	2.28e-1

mean(A) and mean(B) refer to training values.