Domains	LLM-based Approaches								Symb. Baselines			
	PDDL2NL				Tpl		PDDL					
	Bas	CoT	Act	ReA	Bas	Act	Bas	Act	rnd	BrFS	lmc	ff
Block.	13	15	17	**18**	8	12	13	14	1	20	20	20
↳Valm23	15	14	17	18								
Logistics	5	10	16	**19**	2	7	6	15	0	20	20	20
↳Valm23	3	5	17	12								
Depot	0	5	5	13	0	5	0	3	0	20	20	20
↳Valm23	3	6	6	**15**								
Ferry	7	12	14	**18**	0	13	8	17	0	20	20	20
Floortile	0	0	0	0	0	0	0	0	0	18	20	20
Goldm.	1	1	3	1	1	3	1	**4**	0	20	20	20
Grid	8	6	16	**18**	1	12	6	12	0	20	20	20
Grippers	9	17	17	**20**	10	**20**	12	19	0	20	20	20
Movie	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**	3	20	20	20
Rovers	0	0	**18**	18	1	17	1	11	1	20	20	20
Satellite	14	16	**20**	**20**	11	18	14	18	0	20	20	20
Visitall	19	19	**20**	**20**	18	**20**	20	20	8	20	20	20
Σ (240)	96	121	166	**185**	72	147	101	153	13	238	240	240
Further scaled selected domains:												
Block.	3	3	12	**14**	0	6	1	4	0	12	19	20
Ferry	0	0	7	15	0	9	0	**17**	0	8	13	20
Gripper	17	12	**20**	19	16	**20**	16	**20**	0	10	8	20
Visitall	9	2	16	**18**	7	17	14	16	1	10	18	20
Σ (80)	29	17	55	**66**	23	52	31	57	1	40	58	80

Table 2: Number of solved tasks out of 20 per domain. "Valm23" rows show results of the manual encodings by Valmeekam et al. (2023a) in the respective domains. We show in **bold** the best LLM-based method. LLM-based results are averaged over 5 seeds, rnd over 10 seeds.