Table 1: ID and OOD test grid accuracy with SEM (Standard Error of the Mean) as error bars for the *CompGen* experiments part of settings C1-C3 and for the Grid-ViT and LLaDA models.

		Grid-ViT		**LLaDA**	
Setting	**Experiment**	**ID**	**OOD**	**ID**	**OOD**
C1	experiment-1	43.5 ± 16.9	0.0 ± 0.0	27.7 ± 1.4	0.0 ± 0.0
	experiment-2	47.2 ± 18.6	0.0 ± 0.0	67.0 ± 0.4	0.0 ± 0.0
	experiment-3	56.7 ± 7.7	0.0 ± 0.0	33.3 ± 2.5	0.0 ± 0.0
	experiment-4	98.6 ± 0.1	95.6 ± 1.2	87.7 ± 1.6	70.1 ± 6.1
	experiment-5	59.7 ± 11.9	0.0 ± 0.0	32.6 ± 7.5	0.0 ± 0.0
C2	experiment-1	56.2 ± 5.8	0.0 ± 0.0	31.2 ± 0.8	0.0 ± 0.0
	experiment-2	94.3 ± 1.5	0.0 ± 0.0	72.0 ± 1.2	0.0 ± 0.0
	experiment-3	56.6 ± 5.8	0.0 ± 0.0	27.2 ± 1.2	0.0 ± 0.0
	experiment-4	98.2 ± 0.4	95.0 ± 1.1	89.6 ± 0.5	75.1 ± 1.8
	experiment-5	42.9 ± 9.6	0.0 ± 0.0	32.2 ± 3.6	0.0 ± 0.0
C3	experiment-1	40.5 ± 18.8	1.2 ± 0.5	91.7 ± 3.3	1.2 ± 0.1
	experiment-2	68.1 ± 29.5	15.0 ± 7.5	98.4 ± 0.6	24.5 ± 1.1
	experiment-3	51.4 ± 1.2	2.6 ± 0.4	82.7 ± 5.9	5.5 ± 0.8
	experiment-4	67.7 ± 6.5	0.4 ± 0.1	51.2 ± 15.0	0.3 ± 0.0
	experiment-5	66.7 ± 14.1	0.6 ± 0.3	84.3 ± 9.9	1.8 ± 0.7