TextGrad ATE Heatmap (centered at 0)

	Clarity	Engagement	Politeness	Intrinsic_load	Extraneous_load	Encourage_germane_load	Objectives	Metacognition	Demos	Structural_logic	Contextual_logic	Hallucination_awareness
CommonsenseQA	0.033	-0.018	0.019	0.019	0.019	0.019	0.019	0.019	0.033	0.033	0.033	0.019
CommonsenseQA	0.061	-0.020	-0.020	0.027	-0.014	-0.036	0.018	-0.000	-0.004	-0.001	0.014	-0.000
CommonsenseQA	-0.021	-0.018	-0.021	0.042	-0.021	-0.007	0.042	-0.021	-0.013	-0.018	-0.018	0.042
GSM8K	-0.004	-0.026	-0.007	0.966	0.966	0.966	-0.964	-0.001	0.966	0.967	-0.967	0.967
MultiArith	-0.992	-0.992	0.991	0.992	-0.992	0.991	-0.992	-0.991	-0.992	-0.991	-0.991	-0.991
MultiArith	-0.995	-0.995	-0.995	-0.995	-0.995	-0.994	0.994	0.995	0.991	-0.995	-0.995	0.995
boolean_expressions	0.034	0.973	0.024	0.020	0.039	-0.011	0.020	0.022	0.006	-0.974	-0.974	0.022
causal_judgement	0.021	0.138	-0.026	-0.007	-0.023	0.007	0.138	-0.005	0.007	0.039	0.039	-0.030
coin_flip	0.006	-0.009	-0.004	0.011	-0.020	0.053	0.029	0.011	0.014	0.011	0.011	0.016
date_understanding	-0.912	-0.064	-0.039	-0.064	-0.081	-0.032	0.070	-0.035	-0.036	0.065	0.173	-0.058
date_understanding	-0.948	0.116	0.053	-0.068	0.030	-0.022	0.032	-0.034	0.031	-0.032	-0.948	0.003
date_understanding	0.117	0.004	-0.009	0.033	0.026	-0.007	-0.003	0.030	-0.008	0.032	0.117	0.034
disambiguation_qa	-0.046	-0.071	-0.046	-0.071	-0.026	-0.071	-0.071	-0.071	-0.046	-0.071	-0.066	-0.071
last_letters	-0.025	-0.022	-0.024	-0.039	-0.023	-0.039	-0.036	-0.024	-0.039	-0.025	-0.025	-0.021
last_letters	0.023	0.023	0.025	0.095	0.019	0.026	0.035	-0.012	0.095	0.039	0.039	0.014