DSpy MIPROv2 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 | 0.066 | -0.071 | 0.052 | -0.011 | 0.009 | -0.082 | 0.025 | -0.052 | 0.047 | 0.051 | 0.070 | -0.071 |
| CommonsenseQA | 0.003 | -0.080 | -0.022 | -0.074 | 0.011 | -0.081 | -0.066 | -0.081 | -0.058 | 0.011 | 0.011 | -0.036 |
| GSM8K | 0.031 | -0.073 | -0.098 | -0.104 | 0.018 | -0.045 | -0.138 | -0.142 | -0.027 | 0.082 | -0.095 | -0.108 |
| GSM8K | 0.143 | -0.103 | 0.129 | 0.136 | 0.121 | 0.140 | -0.131 | 0.139 | 0.107 | 0.040 | -0.137 | -0.122 |
| MultiArith | -0.457 | 0.104 | -0.366 | -0.095 | -0.308 | 0.104 | -0.261 | 0.103 | -0.214 | -0.312 | -0.085 | -0.256 |
| MultiArith | -0.047 | 0.181 | -0.044 | 0.509 | -0.365 | 0.570 | 0.321 | -0.107 | 0.096 | -0.005 | -0.110 | -0.057 |
| StrategyQA | -0.139 | 0.112 | -0.067 | -0.012 | -0.003 | 0.012 | -0.104 | -0.052 | -0.033 | -0.139 | 0.012 | -0.038 |
| StrategyQA | 0.103 | -0.078 | 0.100 | 0.099 | -0.077 | 0.112 | 0.099 | -0.105 | -0.083 | 0.101 | 0.063 | -0.112 |
| boolean_expressions | -0.051 | -0.164 | -0.128 | -0.017 | 0.123 | -0.165 | -0.015 | -0.138 | -0.248 | -0.101 | -0.082 | 0.050 |
| boolean_expressions | -0.008 | 0.023 | 0.033 | -0.021 | -0.004 | -0.034 | 0.966 | -0.023 | -0.055 | -0.004 | -0.018 | -0.021 |
| causal_judgement | -0.024 | 0.050 | 0.016 | 0.059 | -0.050 | 0.030 | -0.067 | 0.064 | 0.017 | -0.081 | 0.010 | 0.059 |
| causal_judgement | -0.070 | 0.044 | -0.010 | -0.115 | 0.056 | 0.061 | -0.021 | -0.047 | 0.024 | 0.044 | -0.002 | 0.010 |
| coin_flip | 0.981 | 0.295 | -0.021 | 0.022 | -0.295 | 0.005 | 0.020 | 0.005 | 0.019 | 0.007 | 0.012 | 0.005 |
| coin_flip | -0.010 | 0.018 | 0.974 | 0.016 | -0.011 | 0.018 | -0.010 | 0.016 | 0.015 | -0.011 | -0.037 | 0.018 |
| date_understanding | -0.927 | 0.138 | 0.042 | 0.017 | 0.033 | -0.071 | -0.055 | 0.077 | -0.050 | -0.037 | 0.039 | 0.062 |
| date_understanding | -0.081 | 0.096 | 0.109 | -0.094 | -0.041 | 0.070 | -0.026 | 0.127 | 0.175 | -0.048 | -0.119 | -0.078 |
| disambiguation_qa | -0.034 | 0.181 | 0.018 | 0.022 | -0.084 | 0.126 | 0.103 | 0.146 | 0.143 | -0.059 | 0.028 | 0.070 |
| disambiguation_qa | -0.006 | 0.022 | 0.009 | 0.048 | 0.041 | -0.050 | -0.036 | -0.068 | -0.067 | 0.025 | 0.012 | 0.009 |
| last_letters | -0.061 | 0.055 | 0.065 | 0.065 | -0.055 | 0.055 | 0.047 | 0.055 | 0.064 | 0.938 | -0.061 | 0.064 |
| last_letters | 0.054 | 0.011 | 0.223 | -0.001 | 0.065 | -0.007 | -0.265 | 0.008 | -0.210 | 0.223 | 0.178 | -0.034 |