Prevalence of Error Categories by Condition

Categories sorted by mean error rate across all conditions

	Gold (Baseline)	Gemini 2.5 Pro	Gemini 2.5 Flash
Actions & Runners	0.46	0.32	0.44
Expressions & Conditions	0.23	0.32	0.36
Syntax & Structure	0.10	0.30	0.42
Workflow Triggers	0.04	0.10	0.08
Permissions & Security	0.06	0.03	0.08
Other	0.01	0.02	0.02
Shell / Execution	0.00	0.01	0.01

Error Rate

- 100%
- 75%
- 50%
- 25%
- 0%