

## Hypothesis Evidence Landscape and Temporal Dynamics

The LLM-based extraction pipeline produced a total of 3{,}684 assertions across the eight tracked hypotheses, drawn from the full corpus of  $N = 1208$  papers. The distribution of assertion types and the resulting citation-weighted scores reveal a differentiated evidence landscape:

Hypothesis	Score	Supports	Neutral	Contradict	Total	Character
H4: Predictive Coding	+0.59	837	417	0	1{,}254	Strong consensus
H5: Scalability	+0.62	142	110	0	252	Strong consensus
H6: Clinical Utility	+0.41	16	29	0	45	Moderate, growing
H8:	+0.39	54	96	0	150	Moderate,