

RQ2: Operator Performance Metrics with Statistical Significance
 * = Significantly different from top performer ($p < 0.05$, Bonferroni corrected)

Operator	NE (%)	EHR (%)	IR (%)	cEHR (%)	$\Delta\mu$	$\Delta\sigma$
ConceptAdditionOperator	55.71*	4.53	38.52	7.44	-0.05	0.13
InformedEvolutionOperator	49.16	9.72	40.06	16.60	-0.17	0.11
LLMBackTranslation_HI	71.03*	4.70	21.07	5.93	-0.07	0.12
LLMBasedParaphrasing	54.81*	3.49	40.95	6.02	-0.06	0.12
LLM_POSAwareSynonymReplacement	75.90*	6.77	13.15	8.05	-0.05	0.11
MLM	60.65*	5.84	26.48	8.66	-0.05	0.12
NegationOperator	71.47*	4.93	18.81	6.40	-0.06	0.12
POSAwareAntonymReplacement	83.96*	7.25	5.00	7.79	-0.06	0.12
SemanticFusionCrossover	39.74	2.63	56.02	5.86	-0.05	0.09
SemanticSimilarityCrossover	21.87	2.26	0.00	9.98	-0.05	0.09
StylisticMutator	56.68*	2.83	39.40	4.67	-0.07	0.12
TypographicalErrorsOperator	41.68*	3.52	53.26	7.43	-0.06	0.12