

Healthcare tasks, NLP and NLU tasks, and dimensions of evaluation across 519 studies

Source: RAISE Health, 2025 | Chart: 2025 AI Index report

	Dimensions of evaluation							
	Accuracy	Comprehensiveness	Factuality	Robustness	Fairness, bias, and toxicity evaluation	Deployment metrics	Calibration and uncertainty	
Task	Enhancing medical knowledge	222	91	44	33	16	10	3
	Making diagnoses	100	38	11	11	14	4	0
	Educating patients	88	68	32	22	18	3	2
	Making treatment recommendations	47	22	9	8	3	1	0
	Communicating with patients	35	29	8	15	22	1	0
	Care coordination and planning	36	24	5	5	7	1	0
	Triaging patients	24	7	5	2	8	8	0
	Carrying out a literature review	18	7	3	2	2	2	0
	Synthesizing data for research	16	7	2	3	2	2	0
	Generating medical reports	8	8	2	0	3	0	0
	Conducting medical research	8	7	3	3	3	0	0
	Providing asynchronous care	8	5	3	3	1	1	0
	Managing clinical knowledge	5	5	1	1	0	0	0
	Clinical note-taking	6	2	1	1	0	0	1
	Generating clinical referrals	3	0	0	0	0	0	0
	Enhancing surgical operations	3	3	1	1	0	0	0
	Biomedical data mining	2	0	0	0	0	0	0
	Generating billing codes	1	0	0	0	0	0	0
	Writing prescriptions	1	0	0	0	0	0	0
	Question answering*	398	194	71	61	54	14	5
Text classification*	29	10	6	5	10	2	0	
Information extraction*	29	12	8	5	4	6	0	
Summarization*	29	21	7	3	8	0	1	
Conversational dialogue*	6	6	1	1	5	1	0	
Translation*	5	1	2	2	1	2	0	