

Performance of AI tools across specialties

Specialty	Authors	AI tool	Accuracy(number)
Anesthesiology	Lin et al.	ChatGPT 4	92.3%(n=39)
Cardiology	Haze et al.	ChatGPT 3.5	58.5% (n=41)
	Ozeri et al.	ChatGPT 3.5	33.3%(n=12)
	Siebielec et al.	ChatGPT-3.5	57.8%(n.a.)
	Knoedler et al.	ChatGPT 3.5	44.7%(n=208)
	Lin et al.	ChatGPT 4	76.3%(n=38)
	Haze et al.	ChatGPT 4	87.8% (n=41)
	Wójcik et al.	ChatGPT 4	67.1%(n=120)
Dermatology	Knoedler et al.	ChatGPT 3.5	66.7%(n=72)
	Lin et al.	ChatGPT 4	85.7%(n=42)
Endocrinology	Siebielec et al.	ChatGPT-3.5	78.6%(n.a.)
	Knoedler et al.	ChatGPT-3.5	45.3%(n=137)
	Haze et al.	ChatGPT-3.5	56.9% (n=51)
	Ozeri et al.	ChatGPT-3.5	35.7%(n=14)
	Haze et al.	ChatGPT-4	80.4% (n=51)
	Lin et al.	ChatGPT 4	87.5%(n=32)
Emergency medicine	Sadeq et al.	Bard	80.0%(n=10)
	Smith et al.	Bard – PaLM	63.8%(n=360)
	Sadeq et al.	Bing	100%(n=10)
	Akhter	Bing	66.7%(n=50)
	Smith et al.	Bing	68.3%(n=360)
	Sadeq et al.	Claude	80.0%(n=10)
	Siebielec et al.	ChatGPT-3.5	64.2%(n.a.)
	Akhter	ChatGPT 3.5	68.7% (n=50)
	Flores-Cohaila et al.	ChatGPT 3.5	100%(n=16)
	Sadeq et al.	ChatGPT 3.5	90.0%(n=10)
	Smith et al.	ChatGPT 3.5	48.8%(n=360)
	Suwała et al.	ChatGPT 3.5	58.5%(n=20)
	Haze et al.	ChatGPT 3.5	61.8% (n=55)
	Lin et al.	ChatGPT 4	88%(n=25)
	Smith et al.	ChatGPT 4	75.8%(n=360)
	Haze et al.	ChatGPT 4	83.6% (n=55)
	Flores-Cohaila et al.	ChatGPT 4	87.5%(n=16)
	Sadeq et al.	ChatGPT 4	100%(n=10)
	Sadeq et al.	Perplexity	70.0%(n=10)
Family medicine	Siebielec et al.	ChatGPT-3.5	62.9%(n.a.)
	Suwała et al.	ChatGPT 3.5	55.8%(n=20)
	Mousavi et al.	ChatGPT 3.5	73.6%(n=77)

	Lin et al.	ChatGPT 4	92.5%(n=40)
	Mousavi et al.	ChatGPT 4	80.5%(n=77)
Gastroenterology and hepatology	Siebielec et al.	ChatGPT 3.5	54.3% (n.a.)
	Haze et al.	ChatGPT 3.5	55.8% (n=52)
	Knoedler et al.	ChatGPT 3.5	54.5% (n=191)
	Lin et al.	ChatGPT 4	77.1% (n=35)
	Haze et al.	ChatGPT 4	78.8% (n=52)
General Surgery	Sadeq et al.	Bard	44.0% (n=25)
	Torres-Zegarra et al.	Bard	74.1% (n=27)
	Moglia et al.	Bard	79.5% (n=44)
	Sadeq et al.	Bing	44.0% (n=25)
	Torres-Zegarra et al.	Bing	81.5% (n=27)
	Sadeq et al.	Claude	68.0% (n=25)
	Torres-Zegarra et al.	Claude	62.9% (n=27)
	Siebielec et al.	ChatGPT 3.5	53.8% (n.a.)
	Sadeq et al.	ChatGPT 3.5	56.0% (n=25)
	Moglia et al.	ChatGPT 3.5	71.4% (n=44)
	Meo et al.	ChatGPT 3.5	60.0%(n=10)
	Meyer et al.	ChatGPT 3.5	50.0% (n=44)
	Rojas et al.	ChatGPT 3.5	62.2% (n=60)
	Torres-Zegarra et al.	ChatGPT 3.5	70.3% (n=27)
	Suwała et al.	ChatGPT 3.5	52.9% (n=27)
	Flores-Cohaila et al.	ChatGPT 3.5	84.8% (n=33)
	Gencer & Aydin	ChatGPT 3.5	90.5% (n=105)
	Sadeq et al.	ChatGPT 4	60.0% (n=25)
	Ebrahimian et al.	ChatGPT 4	75.0% (n=24)
	Meyer et al.	ChatGPT 4	93.2% (n=44)
	Rojas et al.	ChatGPT 4	90.0% (n=60)
	Torres-Zegarra et al.	ChatGPT 4	96.2% (n=27)
	Moglia et al.	ChatGPT 4	91.5% (n=44)
	Lin et al.	ChatGPT 4	75.4% (n=122)
	Flores-Cohaila et al.	ChatGPT 4	81.8% (n=33)
	Gencer & Aydin	ChatGPT 4	93.3% (n=105)
	Ghanem et al.	ChatGPT 4	65.8%(n=120)
	Long et al.	ChatGPT 4	73.5%(n=21)
	Rojas et al.	ChatGPT 4V	86.1% (n=60)
	Sadeq et al.	Perplexity	52.0 % (n=25)
Gynecology and obstetrics	Torres-Zegarra et al.	Bard	48.5% (n=35)
	Sadeq	Bard	54.5% (n=11)
	Torres-Zegarra et al.	Bing	77.1% (n=35)
	Sadeq et al.	Bing	45.5% (n=11)
	Sadeq et al.	Claude	54.5% (n=11)
	Torres-Zegarra et al.	Claude	48.5% (n=35)

	Siebielec et al.	ChatGPT 3.5	57.4% (n.a.)
	Knoedler et al.	ChatGPT 3.5	56.8%(n=132)
	Meyer et al.	ChatGPT 3.5	72.1% (n=43)
	Torres-Zegarra et al.	ChatGPT 3.5	60.0% (n=35)
	Meo et al.	ChatGPT 3.5	70.0% (n=10)
	Rojas et al.	ChatGPT 3.5	56.7% (n=87)
	Sadeq et al.	ChatGPT 3.5	63.6% (n=11)
	Suwała et al.	ChatGPT 3.5	57.4% (n= 26)
	Haze et al.	ChatGPT 3.5	54.8% (n=62)
	Flores-Cohaila et al.	ChatGPT 3.5	56.7% (n=30)
	Cohen et al.	ChatGPT 3.5	38.7% (n=150)
	Rojas et al.	ChatGPT 4	77.8% (n=87)
	Sadeq et al.	ChatGPT 4	63.6% (n=11)
	Torres-Zegarra et al.	ChatGPT 4	68.5% (n=35)
	Meyer et al.	ChatGPT 4	83.7% (n=43)
	Lin et al.	ChatGPT 4	77.6% (n=125)
	Haze et al.	ChatGPT 4	75.8% (n=62)
	Flores-Cohaila et al.	ChatGPT 4	76.7% (n=30)
	Rojas et al.	ChatGPT 4V	74.3% (n=87)
	Sadeq et al.	Perplexity	18.2% (n=11)
Healthcare	Haze et al.	ChatGPT 3.5	54.7% (n=106)
	Haze et al.	ChatGPT 4	73.6% (n=106)
Internal Medicine	Sadeq et al.	Bard	62.2% (n=127)
	Torres-Zegarra et al.	Bard	69.3% (n=75)
	Sadeq et al.	Bing	64.6% (n=127)
	Torres-Zegarra et al., 2	Bing	90.0% (n=75)
	Sadeq et al.	Claude	66.9% (n=127)
	Torres-Zegarra et al., 2	Claude	78.6% (n=75)
	Siebielec et al.	ChatGPT 3.5	55.7% (n.a.)
	Sadeq et al.	ChatGPT 3.5	66.1%(n=127)
	Meyer et al.	ChatGPT 3.5	59.7% (n=176)
	Rojas et al.	ChatGPT 3.5	52.7% (n=201)
	Torres-Zegarra et al., 2	ChatGPT 3.5	73.3% (n=75)
	Suwała et al.	ChatGPT 3.5	59.6% (n=39)
	Flores-Cohaila et al.	ChatGPT 3.5	77.1% (n=48)
	Ebrahimian et al.	ChatGPT 4	69.6% (n=46)
	Flores-Cohaila et al.	ChatGPT 4	89.6% (n=48)
	Lin et al.	ChatGPT 4	84.9%(n=33)
	Meyer et al.	ChatGPT 4	91.5% (n=176)
	Rojas et al.	ChatGPT 4	75.6% (n=201)
	Sadeq et al.	ChatGPT 4	85.8% (n=127)
	Torres-Zegarra et al., 2	ChatGPT 4	90.0% (n=75)
	Rojas et al.	ChatGPT 4V	78.1% (n=201)
	Sadeq et al.	Perplexity	60.6% (n=127)

Male Reproductive System	Knoedler et al.	ChatGPT 3.5	57.9%(n=38)
Microbiology and Infectiology	Meyer et al.	ChatGPT 3.5	67.6% (n=74)
	Haze et al.	ChatGPT 3.5	59.6% (n=47)
	Ozeri et al.	ChatGPT 3.5	38.5% (n=26)
	Meyer et al.	ChatGPT 4	91.9% (n=74)
	Lin et al.	ChatGPT 4	92.3% (n=39)
	Haze et al.	ChatGPT 4	83.0% (n=47)
Multisystem Processes and Dis	Knoedler et al.	ChatGPT 3.5	46.9%(n=175)
Nephrology and urology	Noda et al.	Bard	32.3% (n=99)
	Siebielec et al.	ChatGPT 3.5	60.0% (n.a.)
	Noda et al.	ChatGPT 3.5	31.3% (n=99)
	Nicikowski et al.	ChatGPT 3.5	45.7% (n=1556)
	Haze et al.	ChatGPT 3.5	51.0% (n=49)
	Knoedler et al.	ChatGPT 3.5	48.4% (n=128)
	Lin et al.	ChatGPT 4	75.7% (n=37)
	Noda et al.	ChatGPT 4	54.5% (n=99)
	Nicikowski et al.	ChatGPT 4	69.8% (n=1556)
	Haze et al.	ChatGPT 4	85.7% (n=49)
Neurology	Stengel et al.	Bard	69%(n=75)
	Cheong et al.	Bard	45.5% (n=301)
	Stengel et al.	Bing	60%(n=75)
	Knoedler et al.	ChatGPT 3.5	60.8%(n=263)
	Haze et al.	ChatGPT 3.5	60% (n=45)
	Ali et al.	ChatGPT 3.5	73.4%(n=500)
	Stengel et al.	ChatGPT 3.5	65%(n=75)
	Meyer et al.	ChatGPT 3.5	54.5%(n=112)
	Cheong et al.	ChatGPT 3.5	46.8% (n=301)
	Giannos	ChatGPT 3.5	56.5%(n=69)
	Giannos	ChatGPT 3.5	42.0%(n=69)
	Giannos	ChatGPT 4	63.8%(n=69)
	Lin et al.	ChatGPT 4	87.5%(n=72)
	Meyer et al.	ChatGPT 4	89.3%(n=112)
	Haze et al.	ChatGPT 4	77.8% (n=45)
	Guillen-Grima et al.	ChatGPT 4	90.0%(n=10)
	Ali et al.	ChatGPT 4	83.4%(n=500)
	Cheong et al.	ChatGPT 4	68.1% (n=301)
Oncology	Siebielec et al.	ChatGPT-3.5	68.1%(n.a.)
	Huang et al.	ChatGPT 3.5	63.1% (n=293)
	Ozeri et al.	ChatGPT 3.5	40.0% (n=10)
	Haze et al.	ChatGPT 3.5	39.3% (n=28)

	Knoedler et al.	ChatGPT-3.5	45.5%(n=154)
	Haze et al.	ChatGPT 4	75.0% (n=28)
	Huang et al., 2023 (49)	ChatGPT 4	74.1% (n=293)
	Lin et al.	ChatGPT 4	97.4% (n=38)
Ophthalmology and otorhinolary	Sadeq et al.	Bard	58.3%(n=36)
	Sadeq et al.	Bing	63.9%(n=36)
	Sadeq et al.	Claude	72.2%(n=36)
	Haze et al.	ChatGPT 3.5	43.5% (n=23)
	Meo et al.	ChatGPT 3.5	75%(n=20)
	Sadeq et al.	ChatGPT 3.5	55.6%(n=36)
	Lin et al.	ChatGPT 4	71.1% (n=38)
	Haze et al.	ChatGPT 4	73.9% (n=23)
	Panthier & Gatinel	ChatGPT 4	91.2%(n=6785)
	Sadeq et al.	ChatGPT 4	77.8%(n=36)
	Sadeq et al.	Perplexity	55.6%(n=36)
Osteology	Knoedler et al.	ChatGPT 3.5	57.4%(n=108)
	Fiedler et al.	ChatGPT 3.5	60.8%(n=51)
	Kim et al.	ChatGPT 3.5	37.5%(n=160)
	Haze et al.	ChatGPT 3.5	70.8% (n=24)
	Guillen-Grima et al.	ChatGPT 4	100% (n=10)
	Kim et al.	ChatGPT 4	60%(n=160)
	Fiedler et al.	ChatGPT 4	66.7%(n=51)
	Lin et al.	ChatGPT 4	75%(n=36)
	Haze et al.	ChatGPT 4	91.7% (n=24)
	Saad et al.	ChatGPT4	67.5(n=240)
Paediatrics	Sadeq et al.	Bard	58.1%(n=93)
	Torres-Zegarra et al.	Bard	61.1%(n=18)
	Sadeq et al.	Bing	81.7%(n=93)
	Torres-Zegarra et al.	Bing	61.1%(n=18)
	Sadeq et al.	Claude	63.4%(n=93)
	Torres-Zegarra et al.	Claude	61.1%(n=18)
	Siebielec et al.	ChatGPT 3.5	58.7% (n.a.)
	Meyer et al.	ChatGPT 3.5	58.1%(n=62)
	Rojas et al.	ChatGPT 3.5	57.9%(n=87)
	Haze et al.	ChatGPT 3.5	47.2% (n=53)
	Flores-Cohaila et al.	ChatGPT 3.5	82.1%(n=28)
	Sadeq et al.	ChatGPT 3.5	55.9%(n=93)
	Suwała et al.	ChatGPT 3.5	54.7%(n=29)
	Gritti et al.	ChatGPT 3.5	37.5%(n=88)
	Torres-Zegarra et al.	GPT-3.5	55.5%(n=18)
	Gritti et al.	ChatGPT 4	65.9%(n=88)
	Ebrahimian et al.	ChatGPT 4	66.7%(n=24)
	Sadeq et al.	ChatGPT 4	76.3%(n=93)

	Meyer et al.	ChatGPT 4	82.3%(n=62)
	Rojas et al.	ChatGPT 4	81.6%(n=87)
	Flores-Cohaila et al.	ChatGPT 4	92.9%(n=28)
	Lin et al.	ChatGPT 4	81.0%(n=137)
	Haze et al.	ChatGPT 4	84.9% (n=53)
	Torres-Zegarra et al.	GPT-4	77.7%(n=18)
	Guillen-Grima et al.	ChatGPT 4	90.0%(n=10)
	Rojas et al.	ChatGPT 4V	78.9%(n=87)
	Sadeq et al.	Perplexity	59.1%(n=93)
Pharmacology	Siebielec et al.	ChatGPT-3.5	63.8%(n.a.)
	Haze et al.	ChatGPT 3.5	54.5%(n=11)
	Haze et al.	ChatGPT 4	54.5%(n=11)
Plastic surgery	Lin et al.	ChatGPT 4	73.3%(n=15)
Psychiatry	Siebielec et al.	ChatGPT-3.5	70.3%(n.a.)
	Suwała et al.	ChatGPT 3.5	77.2%(n=14)
	Meyer et al.	ChatGPT 3.5	79.6%(n=54)
	Knoedler et al.	ChatGPT 3.5	77.9%(n=95)
	Haze et al.	ChatGPT 3.5	64.3% (n=42)
	Rojas et al.	ChatGPT 3.5	69.8%(n=42)
	Rojas et al.	ChatGPT 4	82.5%(n=42)
	Lin et al.	ChatGPT 4	87.3%(n=63)
	Haze et al.	ChatGPT 4	83.3% (n=42)
	Herrmann-Werner et al	ChatGPT4	92.5%(n=307)
	Meyer et al.	ChatGPT 4	90.7%(n=54)
	Rojas et al.	ChatGPT 4V	82.5%(n=42)
Public health	Torres-Zegarra et al.	Bard	80.9%(n=21)
	Torres-Zegarra et al.	Bing	76.1%(n=21)
	Torres-Zegarra et al.	Claude	85.7%(n=21)
	Siebielec et al.	ChatGPT-3.5	75.6%(n.a.)
	Rojas et al.	ChatGPT 3.5	64.2%(n=27)
	Torres-Zegarra et al.	GPT-3.5	71.4%(n=21)
	Flores-Cohaila et al.	ChatGPT 3.5	77.8%(n=18)
	Rojas et al.	ChatGPT 4	82.7%(n=27)
	Torres-Zegarra et al.	GPT-4	85.7%(n=21)
	Flores-Cohaila et al.	ChatGPT 4	83.3%(n=18)
	Rojas et al.	ChatGPT 4V	74.1%(n=27)
Radiology and Nuclear Medicine	Sadeq et al.	Bard	20.0%(n=10)
	Sadeq et al.	Bing	40.0%(n=10)
	Sadeq et al.	Claude	10.0%(n=10)
	Kufel et al.	ChatGPT 3.5	56.4% (n=117)
	Sood et al.	ChatGPT 3.5	31.7%(n=890)

	Sadeq et al.	ChatGPT 3.5	20.0%(n=10)
	Sadeq et al.	ChatGPT 4	40.0%(n=10)
	Sood et al.	ChatGPT 4	70.7%(n=890)
	Hirano et al.	ChatGPT 4 Tui	41.0%(n=139)
	Hirano et al.	ChatGPT 4 Tui	44.6%(n=139)
	Sadeq et al.	Perplexity	0%(n=10)
Respirology	Haze et al.	ChatGPT 3.5	52.9% (n=34)
	Siebielec et al.	ChatGPT 3.5	64.3% (n.a.)
	Knoedler et al.	ChatGPT 3.5	63.0% (n=146)
	Haze et al.	ChatGPT 4	73.5% (n=34)
Rheumatology and immunology	Knoedler et al.	ChatGPT 3.5	60.9%(n=110)
	Haze et al.	ChatGPT 3.5	52.0% (n=25)
	Ozeri et al.	ChatGPT 3.5	41.7% (n=12)
	Lin et al.	ChatGPT 4	81.8% (n=22)
	Haze et al.	ChatGPT 4	92.0% (n=25)