

Supplemental Table 1. Comparison of GPT-4 prompting strategy accuracy by subspecialty.

Version	Section	Diagnosis			Next Step		
		Mean	95% C.I.	p-value	Mean	95% C.I.	p-value
ZS	Cornea	43.6 %	[29.1 %, 59.3 %]	1.00	71.8 %	[55.9 %, 83.6 %]	0.56
ZS PS+		43.6 %	[29.1 %, 59.3 %]		74.4 %	[58.6 %, 85.6 %]	
ZS	General Medicine	50.0 %	[5.9 %, 94.1 %]	1.00	100.0 %	[100 %, 100 %]	1.00
ZS PS+		50.0 %	[5.9 %, 94.1 %]		100.0 %	[100 %, 100 %]	
ZS	Glaucoma	66.7 %	[37.6 %, 86.9 %]	1.00	58.3 %	[30.8 %, 81.5 %]	0.30
ZS PS+		66.7 %	[37.6 %, 86.9 %]		66.7 %	[37.6 %, 86.9 %]	
ZS	Lens and Cataract	44.4 %	[17.7 %, 74.9 %]	1.00	55.6 %	[25.1 %, 82.3 %]	0.29
ZS PS+		44.4 %	[17.7 %, 74.9 %]		66.7 %	[33.3 %, 88.9 %]	
ZS	Neuro-ophthalmology	28.4 %	[18.9 %, 40.2 %]	0.086	56.7 %	[44.7 %, 68.0 %]	0.80
ZS PS+		38.8 %	[28.0 %, 50.9 %]		58.2 %	[46.2 %, 69.4 %]	
ZS	Oculoplastics	38.7 %	[23.5 %, 56.5 %]	1.00	74.2 %	[56.3 %, 86.5 %]	1.00
ZS PS+		38.7 %	[23.5 %, 56.5 %]		74.2 %	[56.3 %, 86.5 %]	
ZS	Pathology and Tumors	58.1 %	[43.1 %, 71.8 %]	0.027	62.8 %	[47.6 %, 75.8 %]	0.48
ZS PS+		72.1 %	[57.0 %, 83.4 %]		67.4 %	[52.3 %, 79.7 %]	
ZS	Pediatrics	44.6 %	[32.3 %, 57.7 %]	0.19	62.5 %	[49.2 %, 74.1 %]	1.00
ZS PS+		53.6 %	[40.6 %, 66.1 %]		62.5 %	[49.2 %, 74.1 %]	
ZS	Retina and Vitreous	38.5 %	[29.4 %, 48.6 %]	0.043	55.2 %	[45.2 %, 64.8 %]	0.23
ZS PS+		49.3 %	[39.4 %, 59.2 %]		60.7 %	[50.6 %, 69.9 %]	
ZS	Uveitis	40.3 %	[29.3 %, 52.4 %]	0.81	55.2 %	[43.2 %, 66.6 %]	1.00
ZS PS+		38.8 %	[28.0 %, 50.9 %]		55.2 %	[43.2 %, 66.6 %]	

ZS Traditional Zero-Shot prompting strategy; ZS PS+ Zero-Shot Plan-and-Solve + Prompting Strategy