Supplemental Table 6: ChatGPT Plus performance per exam section compared to human performances from the BCSC and OphthoQuestions sets

Exam section	BCSC (n=260)		OphthoQuestions (n=260)	
	Humans*	ChatGPT Plus	Humans**	ChatGPT Plus
General Medicine	73	60.0 ± 4.1	67.3 ± 0.6	71.7 ± 2.4
Fundamentals	69	63.3 ± 2.4	67.3 ± 4.7	73.3 ± 2.4
Clinical Optics	64	51.7 ± 6.2	53.7 ± 4.9	40.0 ± 8.2
Pathology/ Tumors	71	65.0 ± 0.0	57.3 ± 3.8	63.3 ± 6.2
Neuro-ophthalmology	73	48.3 ± 2.4	64.0 ± 5.3	31.7 ± 2.4
Pediatrics&Strabismus	74	56.7 ± 2.4	58.0 ± 5.3	36.7 ± 4.7
Oculoplastics	75	48.3 ± 4.7	72.0 ± 4.6	33.3 ± 2.4
Cornea	78	63.3 ± 2.4	61.7 ± 4.9	66.7 ± 2.4
Uveitis	75	56.7 ± 4.7	63.3 ± 5.0	51.7 ± 2.4
Glaucoma	73	68.3 ± 2.4	64.7 ± 4.2	28.3 ± 2.4
Lens and Cataract	77	56.7 ± 2.4	67.7 ± 5.9	51.7 ± 4.7
Retina and Vitreous	75	78.3 ± 2.4	59.7 ± 5.9	51.7 ± 2.4
Refractive Surgery	76	55.0 ± 0.0	62.3 ± 4.6	40.0 ± 4.1
Mean Accuracy (%)*	**73.3**	**59.4 ± 0.6**	**63.0 ± 4.5**	**49.2 ± 1.0**

* The BCSC Self-Assessment Program provides average peer scores without breakdown per year of training or any historical data

** OphthoQuestions provides historical data that corresponds to the user's year of training. To calculate the mean accuracies and standard deviations, we used data from 3 consecutive years of training: 1st year (2019 – 2020), 2nd year (2020 – 2021) and 3rd year (2021 – 2022)

*** The mean accuracies per section and overall are provided solely for general comparison purposes, and they reflect the average performance on the BCSC and OphothoQuestions testing sets, which comprise around 4,458 and 4,539 questions, respectively. It should be noted that these figures do not reflect the mean accuracies for the 520 question sample exams that were used to assess ChatGPT's performance, as this specific data cannot be obtained