Supplemental Table 2: ChatGPT models performance per exam section for each of the BCSC and OphthoQuestions testing sets

Exam section	BCSC (n=260)					OphthoQuestions (n=260)				
		Plus*					Plus*			
	Jan 9	Mean ± SD	1	2	3	Jan 9	Mean ± SD	1	2	3
General Medicine	75	60.0 ± 4.1	65	55	60	75	71.7 ± 2.4	75	70	70
Fundamentals	55	63.3 ± 2.4	60	65	65	65	73.3 ± 2.4	75	75	70
Clinical Optics	60	51.7 ± 6.2	45	50	60	25	40.0 ± 8.2	50	30	40
Pathology/ Tumors	40	65.0 ± 0.0	65	65	65	45	63.3 ± 6.2	65	55	70
Neuro-ophthalmology	40	48.3 ± 2.4	50	45	50	10	31.7 ± 2.4	30	35	30
Pediatrics&Strabismus	65	56.7 ± 2.4	55	60	55	20	36.7 ± 4.7	40	40	30
Oculoplastics	50	48.3 ± 4.7	45	55	45	40	33.3 ± 2.4	35	35	30
Cornea	60	63.3 ± 2.4	60	65	65	60	66.7 ± 2.4	65	65	70
Uveitis	35	56.7 ± 4.7	60	50	60	65	51.7 ± 2.4	55	50	50
Glaucoma	50	68.3 ± 2.4	70	65	70	25	28.3 ± 2.4	30	25	30
Lens and Cataract	65	56.7 ± 2.4	55	60	55	45	51.7 ± 4.7	45	55	55
Retina and Vitreous	75	78.3 ± 2.4	80	80	75	40	51.7 ± 2.4	50	55	50
Refractive Surgery	55	55.0 ± 0.0	55	55	55	40	40.0 ± 4.1	40	45	35
Mean Accuracy (%)	**55.8**	**59.4 ± 0.6**	**58.9**	**59.2**	**60.0**	**42.7**	**49.2 ± 1.0**	**50.4**	**48.8**	**48.5**

* The mean and standard deviations for the ChatGPT Plus model were calculated using data from runs 1, 2, and 3