

# Can Artificial Intelligence Pass the American Board of Orthopaedic Surgery?

Zachary Lum DO<sup>2</sup>, Dylon Collins BS<sup>1</sup>, Lohitha Guntupalli BS<sup>1</sup>, Stanley Dennison BS<sup>1</sup>, Soham Choudhury<sup>2</sup>, Augustine M. Saiz MD<sup>2</sup>, Robert Lawrence Randall MD<sup>2</sup>  
<sup>1</sup>Nova Southeastern University, Davie, FL, <sup>2</sup>University of California Davis, Sacramento, CA  
 dc2273@mynsu.nova.edu

**Disclosures:** Zachary Lum (N), Dylon Collins (N), Lohitha Guntupalli (N), Stanley Dennison (N), Soham Choudhury (N), Augustine M. Saiz (N), Robert Lawrence Randall (N)

**INTRODUCTION:** Over the past decade, advances in neural networks, deep learning, and artificial intelligence (AI) have changed the way we approach a wide range of tasks and industries ranging from manufacturing and information technology to consumer products. Previous deep learning AI has been structured around domain specific areas. Naturally, being trained on dataset-specific areas of interest result in higher accuracy and precision. A new AI large language model using non-specific domain areas has gained recent attention in its novel way to process information.

We wanted to understand if a generative pre-trained transformer chat bot has the ability to correctly answer orthopaedic surgery board questions. And if so, how would it do against a standard cohort of orthopaedic surgery residents, if it was capable of passing the orthopaedic boards, and if correct questions would vary based upon taxonomy type.

**METHODS:** We asked ChatGPT (OpenAI, San Francisco, CA, USA), 3830 questions, with 3173 questions based on the Orthopaedic In-Training Exam and 757 actual OITE questions. These questions were then entered and composite score was then record based upon subject material and taxonomy type. Questions with figures, diagrams or charts were excluded. Based upon ranking, a pass-fail rank was set at 10<sup>th</sup> percentile at a 5<sup>th</sup> year resident level. Comparison to orthopaedic surgery residents OITE ranking was made.

**RESULTS SECTION:** After exclusion, 1934 questions were administered. ChatGPT was able to select the correct answer 50% of the time. It performed better at basic science, pathology/oncology, knee sports medicine and pediatrics subject areas. The AI did not score differently based upon question taxonomy type. Based upon prior OITE testing, ChatGPT scored between 42-95<sup>th</sup> percentile for PGY1s, 30-54<sup>th</sup> percentile for PGY2s, 13-14<sup>th</sup> for PGY3s, 4-9<sup>th</sup> for PGY4s and 1<sup>st</sup> percentile for PGY5s.

**DISCUSSION:** Although this general domain AI has a low likelihood of passing the orthopaedic surgery board exam, testing performance and knowledge would be comparable to a 2<sup>nd</sup> year orthopaedic surgery resident with proficiency in basic science and pathology/oncology subject areas.

**SIGNIFICANCE/CLINICAL RELEVANCE:** Although artificial intelligence continues to grow and shape our approach to a variety of tasks, this study reveals that artificial intelligence, even with training from preformed data sets, does not have the same level of competency and proficiency that orthopaedic surgery residents demonstrate with regards to orthopaedic surgeon board examination testing and knowledge.

## IMAGES AND TABLES:

Table 1: OITE simulated questions, from Orthobullets (Lineage Medical, SF, CA). A total of 3173 questions were administered to the AI chatbot. BS: basic science, TR: trauma, KS: knee and sports, SP: spine, RE: reconstruction (hip & knee), PE: pediatrics, PA: pathology, SE: shoulder and elbow. HA: hand, FA: foot and ankle, AN: anatomy and approaches.

	Correct	% Correct	Incorrect (%)	No answer	Image	Total	p
BS	179	69	79	3	53	314	
TR	54	41	77	1	193	325	
KS	94	53	82	2	114	292	
SP	56	44	66	4	199	325	
RE	60	37	98	3	158	319	
PE	73	49	73	2	172	320	
PA	58	68	27	0	235	320	
SE	65	43	86	0	170	321	
HA	44	39	68	1	169	282	
FA	48	51	46	1	195	290	
AN	15	32	31	1	18	65	
<b>Total</b>	<b>746</b>	<b>50</b>	<b>733</b>	<b>18</b>	<b>1676</b>	<b>3173</b>	

Table 2: Questions from actual OITE exams, 2015, 2016 and 2022. Years 2022 was used as AI training data would not be available to know this specific information. BS: basic science, TR: trauma, SM: sports medicine, SP: spine, HK: hip and knee, PE: pediatrics, OC: oncology, SE: shoulder and elbow, HA: hand, FA: foot and ankle.

	# Correct	% Correct	# Incorrect	Tax 1 Correct	Tax 1 Inc.	Tax 2 Correct	Tax 2 Inc.	Tax 3 Correct	Tax 3 Inc.
BS	75	66	39	66	33	7	4	2	2
TR	27	43	36	16	20	5	2	6	14
SM	27	73	10	14	6	5	0	8	4
SP	11	58	8	6	8	3	0	2	0
HK	17	47	19	11	10	2	2	4	7
PE	17	61	11	12	2	3	2	2	7
OC	9	50	9	7	3	1	1	1	5
SE	13	50	13	6	6	1	3	6	4
HA	9	28	23	4	7	2	6	3	10
FA	5	36	9	2	4	0	1	3	4
<b>total for type</b>	<b>210</b>	<b>54</b>	<b>177</b>	<b>144</b>	<b>99</b>	<b>29</b>	<b>21</b>	<b>37</b>	<b>57</b>

## Combined Orthobullets and OITE. Binomial log regression

### Binomial Logistic Regression

Model Fit Measures									
Model	Deviance	AIC	R <sup>2</sup> <sub>MCF</sub>						
1	2497	2521	0.0372						

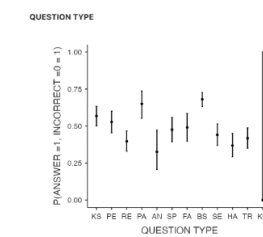
Omnibus Likelihood Ratio Tests				
Predictor	χ <sup>2</sup>	df	p	
QUESTION TYPE	96.5	11	<.001	

Model Coefficients - ANSWER =1, INCORRECT =0										
Predictor	Estimate	95% Confidence Interval			SE	Z	p	95% Confidence Interval		
		Lower	Upper	Lower				Upper		
Intercept	0.274	0.00289	0.5451	0.138	1.9808	0.048	1.315	1.003	1.725	
QUESTION TYPE:										
PE - KS	-0.162	-0.56165	0.2386	0.204	-0.7912	0.429	0.851	0.570	1.269	
RE - KS	-0.692	-1.08762	-0.2971	0.202	-3.4333	<.001	0.500	0.337	0.743	
PA - KS	0.347	-0.14021	0.8346	0.249	1.3961	0.163	1.415	0.869	2.304	
AN - KS	-1.000	-1.67338	-0.3265	0.344	-2.9102	0.004	0.368	0.188	0.721	
SP - KS	-0.373	-0.80087	0.0541	0.218	-1.7118	0.087	0.688	0.449	1.056	
FA - KS	-0.311	-0.77662	0.1535	0.237	-1.3122	0.189	0.733	0.460	1.166	
BS - KS	0.484	0.13649	0.8319	0.177	2.7293	0.006	1.623	1.146	2.298	
SE - KS	-0.512	-0.91435	-0.1105	0.205	-2.4987	0.012	0.599	0.401	0.895	
HA - KS	-0.815	-1.24839	-0.3808	0.221	-3.6802	<.001	0.543	0.287	0.683	
TR - KS	-0.607	-1.00054	-0.2133	0.201	-3.0223	0.003	0.545	0.368	0.808	
KE - KS	-12.840	-649.32607	623.6459	324.744	-0.0395	0.968	2.65e-6	1.00e-28	7.01e+270	

Note: Estimates represent the log odds of \*ANSWER =1, INCORRECT =0 =1\* vs. \*ANSWER =1, INCORRECT =0 =0\*

### Estimated Marginal Means



### Estimated Marginal Means - QUESTION TYPE

QUESTION TYPE	Probability	SE	95% Confidence Interval	
			Lower	Upper
KS	0.568	0.03394	0.501	0.633
PE	0.528	0.03742	0.455	0.600
RE	0.397	0.03513	0.330	0.467
PA	0.650	0.04698	0.554	0.736
AN	0.326	0.06912	0.207	0.473
SP	0.475	0.04206	0.394	0.558
FA	0.481	0.04810	0.398	0.584
BS	0.681	0.02413	0.632	0.726
SE	0.441	0.03732	0.369	0.515
HA	0.368	0.04019	0.293	0.450
TR	0.418	0.03541	0.350	0.488
KE	3.49e-6	0.00113	2.22e-16	1.000