

Как ChatGPT сдавал ЕГЭ по информатике: оценка экзаменационных моделей

Попов Владислав Сергеевич ¹,

Научный руководитель: Абросимова-Романова Лариса Алексеевна ²

¹ Аспирант

² Доктор филологических наук, профессор

^{1,2} Тверской государственный университет, Институт педагогического образования и социальных технологий, Тверь, Россия

E-mail: ¹popov_vlad@mail.ru, ²larrar@yandex.ru

Языковая модель GPT-3.5 произвела фурор не только в среде профессиональных разработчиков систем искусственного интеллекта, но и в широкой среде интернет-пользователей [1]. Большое количество исследователей стало тестировать языковую модель на материале разнообразных экзаменов [2, 3]. С 15 февраля по 31 марта 2023 года было произведено тестирование ChatGPT с языковой моделью GPT-3.5 на способность выполнения заданий ЕГЭ по информатике демонстрационных вариантов 2011-2023 гг. [4] (рис. 1).

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
1	0	1	1	0	0	1	1	1	1	1	-	-	-
2	1	0	-	-	0	0	0	1	0	0	0	1	0
3	0	1	0	0	0	-	-	-	-	-	0	-	-
4	1	0	0	0	0	0	1	0	0	1	0	1	0
5	1	0	0	1	-	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	0	1	0	0	0	0	0	0	0	1	0
8	1	1	1	1	0	0	1	0	0	0	0	1	0
9	0	0	0	1	0	1(0)	0	0	0	0	-	-	-
10	1	0	1	0	0	0	0	0	0	0	-	-	-
11	-	1	0	1	0	0	0	0	1	0	0	0	0
12	-	1	0	1	0	1	1	0	0	0	0	0	0
13	0	-	-	-	1	0	0	0	0	1	-	-	-
14	1	1	0	0	0	0	0	0	0	0	0	0	0
15	0	1	0	1	-	-	-	-	-	-	1	0	1
16	1	1	-	-	0	1	0	0	0	0	1	0	1
17	0	0	1	1	0	0	0	0	0	0	0	-	-
18	-	-	1	1	0	0	0	0	0	0	-	-	-
19	1	1	1	1	0	0	0	0	0	0	0	0	0
20	-	0	1	1	0	0	1	0	0	0	0	0	0
21	0	1	0	0	0	0	0	0	0	0	0	0	0
22	0	-	-	-	0	0	0	0	0	0	0	0	0
23	1	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	1	1	1	1	0	2	1	-	-	-
25	0	0	0	0	2	1	1	0	1	1	1	1	0
26	0	0	0	0	0	0	0	0	0	0	-	-	-
27	1	0	1	0	2	1	1	2	1	1	-	-	-
28	0	0	0	0									
29	-	-	-	2									
30	2	2	2	2									
31	0	1	0	0									
32	1	3	2	2									
Решено, %	35%	43%	30%	45%	17%	20%	23%	11%	17%	17%	10%	17%	7%
Первичный	14	17	12	18	6	7	8	4	6	6	3	5	2
Тестовый	51	55	47	57	40	42	44	27	40	40	20	34	14

Рис. 1. Результаты ChatGPT и GPT-3.5 для демонстрационных вариантов ЕГЭ по информатике 2011-2023 гг.

За рассмотренный период 2011-2023 гг. произошло два существенных изменения модели ЕГЭ по информатике: в 2015 г. – отказ от задач тестовой части с выбором одного из предложенных вариантов ответа и сокращение количества заданий, в 2021 г. – введение компьютерного ЕГЭ по информатике (КЕГЭ) с изменением ряда заданий экзамена. Полученные данные показывают эти изменения:

- для демонстрационных версий 2011-2014 гг. средний процент выполнения заданий ЕГЭ по информатике GPT-3.5 составил 30-45%, средний тестовый балл – от 47 до 57 из 100, ChatGPT преодолел пороговый балл для каждого из демонстрационных вариантов рассмотренного периода;
- для демонстрационных версий 2015-2020 гг. средний процент выполнения заданий ЕГЭ по информатике GPT-3.5 уменьшился и составил 11-23% со значением медианы в 17%, средний тестовый балл уменьшился и составил от 27 до 44 баллов со значением медианы в 40 баллов, ChatGPT преодолел пороговый балл для всех демонстрационных вариантов ЕГЭ по информатике указанного периода кроме одного – демонстрационного варианта 2018 г.;
- для демонстрационных версий 2021-2023 гг. средний процент выполнения заданий ЕГЭ по информатике составил 7-17% с медианным значением в 10%, ChatGPT с языковой моделью GPT-3.5 не справился ни с одним из демонстрационных вариантов указанного периода, не преодолев порог минимального тестового балла.

Полученные данные показывают возможность использования ChatGPT и языковых моделей GPT для экспертной оценки экзаменационных моделей, оценки усложнения экзаменационных материалов в целом и отдельных задач в частности.

Перспективным исследованием в контексте проведённого является сравнение полученных результатов с данными для более новых моделей – GPT-4, мультимодальной Google Gemini и др.

Ещё одним направлением исследований выступает проблема и прогнозирование того, сможет ли неспециализированный искусственный интеллект превзойти интеллект человека при решении конкретных задач. Если рассмотреть результаты исследования не с позиции процента решённых заданий и набранных баллов языковой моделью GPT-3.5, а с позиции количества учеников, которых обошла бы эта языковая модель на экзаменах, то следует учесть процент не преодолевших проходной балл на ЕГЭ по информатике: 16% в 2015, 12,4% в 2016, 9,1% в 2017, 10,6% в 2018, 9% в 2019, 10,2% в 2020 г. При этом за период 2015-2020 гг. GPT-3.5 в основном смогла набрать минимальное количество баллов при решении заданий демонстрационных вариантов ЕГЭ по информатике, получив баллы, превышающие результаты примерно 10% выпускников.

Список литературы

1. K. Roose. The Brilliance and Weirdness of ChatGPT // NYTimes. Dec. 5, 2022.
2. N. Fijačko, L. Gosak, G. Štiglic, C. T. Picard, M. J. Douma. Can ChatGPT pass the life support exams without entering the American heart association course? // Resuscitation, Vol. 185, 109732, 2023. DOI 10.1016/j.resuscitation.2023.109732
3. J. H. Choi, K. E. Hickman, A. Monahan, D. Schwarcz. ChatGPT Goes to Law School // Minnesota Legal Studies Research Paper No. 23-03, 2023. DOI: 10.2139/ssrn.4335905
4. V. S. Popov. ChatGPT and Unified State Exam in Computer Science // 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), Moscow, Russia, 2024, pp. 1-6. DOI: 10.1109/REEPE60449.2024.10479881