

Этические вопросы интеграции нейросети (на примере chat-GPT-3).

Научный руководитель – Усманов Рафик Хамматович

Миркушина Лия Рафиковна

Кандидат наук

Московский городской педагогический университет, Институт гуманитарных наук,
Москва, Россия

E-mail: mirkushinalr@mgpu.ru

Исследования в области интеграции искусственного интеллекта в социально-экономические процессы сопровождаются повышенным вниманием к этическим аспектам и оценкам рисков его использования. На сегодняшний день развивается междисциплинарная исследовательская отрасль - робоэтика, связанная с изучением вопроса внедрения искусственных моральных агентов и в целом ИИ в человеческое общество. [1] Волна дискуссий возникла в связи с появлением новых текстовых нейросетей ИИ общего назначения, таких как Generative Pre-trained Transformer 3 (GPT-3,5). Третье поколение алгоритма обработки естественного языка, созданный исследовательской лабораторией OpenAI - одна из самых крупных языковых моделей в мире, используя машинное обучение, способна создать тексты, стилистически похожие на созданные человеком. Кроме того, нейросеть пишет сценарии, работает с программным кодом, даже может создать аннотации к научным статьям, компилируя информацию под заданный контекст, а количество пользователей этой сети выросло с момента запуска на миллион. Также программа составляет конкуренцию поисковому браузеру Google, которые не выдает список релевантный запросу пользователю ссылок, а упрощает и экономит время пользователя предлагая самостоятельно проанализированное итоговое решение. Все чаще встречаются новости о том, что нейросеть сдала экзамен МВА в Уортонском колледже или студент защитил выпускную работу, текст которого полностью сгенерировал chat – gpt, а также появились опасения, что Chat - gpt способен заменить отчасти своих создателей - низкоквалифицированных программистов.

Разработка ИИ системы, в частности большой языковой модели без конкретной цели, но с широким спектром возможностей, которые заранее не определены разработчиками, в конечном счете может привести к ситуации затруднительного контроля, а также злонамеренного использования. Невозможно предугадать риски и последствия данной модели, так как возможности её использования только растут, вызывая осложнения в предсказывании последствий, ставя различные этические вопросы на разных уровнях. Таким образом, изучение проблемы внедрения нейросети позволит выявить рекомендации по перераспределению специализированных ресурсов для решения растущих этических вопросов цифровизации. К первой проблеме можно отнести рост дезинформации, подведение итогов по неверному алгоритму, с связи с легкостью генерации текстов по необходимой тематике, а именно ложные или вводящие в заблуждение новости, спам, поддельные обзоры продуктов, поддельные учетные записи в социальных сетях. [2] Возникает так называемая «искусственная глупость», которая может распространяться и на важные сведения, доверенные человеком. Также выделяется отдельная этическая проблема в отношении используемого языка, а именно сложность управления воспроизводства и обработка текста, который может содержать небезопасный контент (дискриминационного и сексистского характера, а также предвзятой информации касаясь некоторых религий), данный вопрос об изменении поведения модели находится на стадии решения.

Вторая проблема заключается в высвобождении большей части технических функций и автоматизации рутинных процессов, связанных с кодом и текстом, и ставят вопрос о

переквалификации таких профессий как копирайтер, низкоквалифицированный программист, маркетолог, сотрудники кол-центров. Данная ситуация в будущем может привести к росту социальной напряженности, либо к росту цифрового неравенства среди тех, кто не может овладеть цифровыми компетенциями на высоком уровне. Ситуация с защитой выпускной квалифицированной работы, с использованием нейромощников уже начинает менять и систему образования, изменяя подход к знаниям, требуя большей изобретательности и усилий от преподавателей в контексте заинтересованности своим предметом, оценивании работ студентов. Ведь запрет использования нейросети не решит проблему, поэтому были предложены идеи реформирования порядка аттестации, параллельно разрабатывая инструменты распознавания текста, написанного с помощью нейросети. [3]

В будущем, использование нейронной сети может приобрести настолько большие масштабы, что произойдет симбиотичность по отношению к человеческому мозгу, превращаясь в сверхиндивидуальное метасознание в виде базы знаний, интерпретаций, различных мнений, которые человек на доверительной основе будет использовать и усваивать как свои собственные. Возможные последствия наступления технологической сингулярности продолжают активно обсуждаться, но все они едины в одном: это некая точка невозврата, после которой человечество уже не будет прежним. Так или иначе, принципы цифровой коммуникации становятся факторами, влияющими на мышление общества, формируя его утилитарность и мобильность. Для нивелирования данной опасности работы с нейросетью усилия человечества целесообразно сконцентрировать на воспроизводстве и максимизации глобального человеческого интеллекта. Этика ИИ должна рассматривать как вопросы моральных решений не только в сфере регулирования и создания непосредственных ИИ, но и в сфере оценки результатов его внедрения. Междисциплинарный характер научной области искусственного интеллекта предполагает всестороннее рассмотрение искусственного интеллекта как философами, инженерами, так и специалистами по этическим вопросам. Рекомендации также должны включать в себя повышение цифровой грамотности и осведомленности общественностью о GPT-3, а также повышение прозрачности и подотчетности разработчиков языковых моделей.

Источники и литература

- 1) Kreps, S., McCain, M., Brundage, M. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. // Journal of Experimental Political Science, 9(1), 2022. p. 104–117.
- 2) Müller, Vincent C. Ethics of Artificial Intelligence and Robotics. In Edward Zalta (ed.) // Stanford Encyclopedia of Philosophy. Palo Alto, Cal.: CSLI, Stanford University. 2020. p. 1-70.
- 3) <https://ria.ru/20230208/neyroset-1850578155.html> (Фальков прокомментировал случай с написанным нейросетью дипломом РГГУ)