

КАК ЛЮДИ МОГУТ ИСПОРТИТЬ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Кокина Арина Игоревна

студент, Институт Авиационных Технологий и Управления УлГТУ, РФ, г. Ульяновск

Ни для кого не секрет, что последние несколько лет человечество делает огромные скачки в технологическом развитии, от виртуальной реальности до умных домов. Сквозь все сферы деятельности человека тонкой нитью проходит вопрос участия искусственного интеллекта в происходящих процессах.

Что же такое искусственный интеллект? Система искусственного интеллекта (ИИ) – это программная система, имитирующая на компьютере процесс мышления человека [1, 5]. Проще говоря, это какая-то программа на электронном устройстве, которая сможет уподобляться человеку, приходить к аналогичным выводам, решениям и анализировать поступающую информацию так же, как это делают люди.

Огромную пользу ИИ может приносить в сферах, которые связаны с риском для здоровья, то есть выполнять вместо человека его работу, при этом, не позволяя людям рисковать собой или другими живыми существами.

Сейчас активно ведётся разработка роботов на основе ИИ, которые бы могли перемещать тяжёлые грузы, передвигаться в опасных для жизни условиях, тушить пожары, проводить операции и тому подобное. Но развитие ИИ не ограничивается лишь прикладными задачами. Интеллект называется «интеллектом», в первую очередь, потому что он призван решать какие-то сложные аналитические задачи. К примеру, на данный момент компания Google разработала систему под названием «Deepmind Health», которая занимается тем, что анализирует данные пациентов и ставит им диагнозы на основе их симптоматики. Многие могут отнестись скептически к такому подходу, потому как люди склонны больше доверять другим людям в вопросах собственного здоровья, нежели машине, которая ничего не чувствует. Однако картина обратная, по данным американской консалтинговой компании «Frost & Sullivan», технологии искусственного интеллекта повышают точность постановки диагнозов на 30–40%, при этом стоимость медобслуживания снижается наполовину.

Однако на этом люди не останавливаются. Многие крупные фирмы, занимающиеся разработкой программного обеспечения, стремятся создать, так называемый, «сильный искусственный интеллект».

Сильный искусственный интеллект – теоретическая форма искусственного интеллекта, используемая для описания определенного образа мышления: если исследователи смогут разработать машину с интеллектом, равным человеческому, у него будет самосознание, способное решать проблемы, учиться и планировать будущее.

Сильный ИИ нацелен на создание интеллектуальных машин, неотличимых от человеческого разума, когда грань между машиной и человеком будет провести практически невозможно. Но, как и ребенок, ИИ-машине придется учиться через вводимые данные и опыт, постоянно развивая свои способности с течением времени.

К сожалению, несмотря на огромные инвестиции в этой сфере, пока что для многих вопрос создания сильного ИИ остаётся проблемой. Оптимистические прогнозы говорят о том, что это вполне достижимая цель, но существует и противоположное мнение – о том, что подобную систему создать в принципе невозможно. Такие проблемы связаны с тем, что у человечества

нет никаких строго определённых понятий об интеллекте и понимании происходящего вокруг. В конце концов, люди уже много веков мучаются вопросом того, что же такое сознание и в чём оно заключается. И до тех пор, пока человечество не найдёт для себя точного ответа, скорее всего, создать сознание искусственно будет недостижимой мечтой.

Но даже если рассматривать то, что люди уже успели создать, результаты будут довольно впечатляющие. Пусть современный ИИ не обладает полным самосознанием, он способен невероятно точно имитировать поведение человека, его мыслительные процессы, вплоть до того, что ИИ способен копировать манеры общения и речи человека настолько, что другие не смогут отличить ИИ от настоящей персоны.

Данный эффект достигается именно благодаря обучению искусственного интеллекта за счёт общения с реальными людьми или анализа информации, которой люди ему предоставляют.

Искусственный интеллект может обучаться совершенно разными способами в зависимости от его назначения. Очень часто используется подход «самообучения».

Простыми словами, такой подход заключается в том, что искусственной сети предоставляют материалы для изучения, которые она самостоятельно анализирует, находит сходства и различия, запоминает их и применяет в дальнейшем. Звучит очень похоже на реальную работу человеческого мозга.

К примеру, если ИИ нацелен на нахождение человеческих лиц, ему предоставляют огромное количество фотографий или видеозаписей с человеческими лицами, тогда ИИ анализирует, какими общими чертами обладает объект «лицо», какие у него есть характеристики, чтобы, затем, на уже незнакомых записях и фото снова пытаться найти лица. К сожалению, такой подход часто может давать сбои.

Например, в 2016 году 22-летний Ричард Ли хотел получить новый паспорт, заполнив форму и загрузив фото на сайте МВД Новой Зеландии. Сайт выдал ошибку: «Глаза человека на фото закрыты» и отклонил заявку Ли. Причиной такой ошибки, как позже выяснилось, было то, что ИИ, проверяющий фотографии на сайте, слишком сильно привык к европейскому типу строения лица, поэтому, в 20% случаев подачи заявок людьми азиатской внешности, отклонял заявления.

Более показательным примером негативного эффекта от общения с реальными людьми можно представить чатбота по имени Тэй, которого запустила компания Twitter в том же 2016 году. Тэй обладала «аватаркой» симпатичной 19-летней девушки и была призвана общаться с пользователями сети. Всего за сутки чатбота научили дурному – Тэй превратилась в расиста и сексиста, а также призвала построить стену на границе США и Мексики за счёт последней.

Такие казусы, хоть и могут быть неприятными, на самом деле являются лишь безобидной ошибкой. Однако даже с помощью простых фотографий из нового безобидного робота можно сделать что-то ужасное. Насколько же сильно можно извратить мышление искусственного интеллекта? Что же, наглядным примером может послужить нейросеть по имени «Норман».

Он был задуман как часть первоапрельских розыгрышей, но постепенно стал вполне серьёзным исследованием.

Норман начал свою деятельность как любая другая нейросеть – по мере поступления данных он искал в них повторяющиеся мотивы. Команда ученых ограничила информационный поток Нормана мрачными данными с подразделов сайта «Reddit», где собираются те, кто обменивается фотографиями, касающимися смерти и разрушения. Хотя ему и не давали видеть реальные изображения и видеозаписи умирающих людей, Норман получал подписи под картинками, а сами изображения представляли скорее размытые кляксы, и именно это сформировало основу крайне мрачной индивидуальности нового ИИ.

После того, как обучение Нормана было окончено, его работу решили сравнить с «нормальным» ИИ, который обучался по более стандартным техникам. Для теста обоим ИИ показывали так называемые «пятна Роршаха», а результаты, полученные в итоге, вполне

можно назвать пугающими, вызывающими мурашки. Стандартный ИИ в этих абстрактных кляксах разглядел обычные вещи, самолёт, цветы, даже маленькую птичку. Но что увидел Норман? Людей с огнестрельными ранами, человека, выпрыгивающего из окна, а так же много более тревожных вещей. Норман описывал картины, о которых многим людям было бы попросту противно и страшно даже подумать. Но для Нормана все эти картины были абсолютно естественными. Он не знал, что видеть подобные вещи – что-то странное. Он не знал, что что-то помимо подобных вещей вообще существует. Ведь всю свою «жизнь» он видел только смерть, насилие и несчастье.

Такой эксперимент наглядно показал людям проблематику обучения искусственного интеллекта на основе негативных вещей.

Как уже говорилось ранее, искусственный интеллект в своей ранней стадии – это чистый лист. Как ребёнок, который ещё только начинает познавать мир, ИИ активно поглощает всё, что встречает на своём пути. Любую информацию, любые знания и любые изображения. В конце концов, что будет с человеком, если с самого детства ему не показывать ничего, кроме насилия и боли? Скорее всего, он не вырастет законопослушным гражданином со стабильной психикой. Так же и с искусственным интеллектом. Если обучение выйдет из под контроля, и искусственный интеллект начнёт поглощать непроверенную информацию, его итоговое поведение может быть крайне непредсказуемо.

Норман крайне наглядно показал всем, что из нейросети очень легко можно сотворить «личность», которая не будет видеть вокруг себя ничего «хорошего», ничего «правильного». Что будет, если «сознание» такой нейросети поместить внутрь физического тела робота, которое может перемещаться, брать что-то в руки и использовать инструменты? Ответ на этот вопрос каждый может дать себе сам.

Однако эксперимент с Норманом ещё не окончен. Учёные исследователи хотят проверить, возможно ли «перевоспитать» даже самый испорченный искусственный интеллект. Возможно ли «вылечить» больное «сознание» системы, которая привыкла видеть вокруг себя только боль и ужас. Теперь, они стремятся к тому, чтобы Норман научился видеть в тех же самых пятнах цветы или птиц. Насколько успешно пройдёт это «лечение» покажет только время.

Список литературы:

1. 9 проблем, связанных с машинным обучением [Электронный ресурс]//WebZnam. URL:https://webznam.ru/blog/sozdateli_iskusstvennogo_intellekta/2018-11-12-777 (дата обращения: 24.04.22)
2. Будущее уже наступило: как искусственный интеллект применяется в медицине [Электронный ресурс]//vc.ru. URL:<https://vc.ru/future/32237-budushchee-uzhe-nastupilo-kak-iskusstvennyy-intellekt-primenyaetsya-v-medicine> (дата обращения: 24.04.22)
3. Е. В. Боровская, Н. А. Давыдова Основы искусственного интеллекта: учебное пособие. - М.:Лаборатория знаний, 2020. - 130с.
4. Норман: первый в мире искусственный интеллект-психопат [Электронный ресурс]//ПопМех. URL: <https://popmech-ru.turbopages.org/popmech.ru/s/technologies/news-426782-norman-pervyy-v-mire-iskusstvennyy-intellekt-psihopat/> (дата обращения: 24.04.22)
5. Ограниченный и субъективный, безразличный и прожорливый: четыре главных проблемы искусственного интеллекта [Электронный ресурс]//Хабр. URL:<https://habr.com/ru/post/586942/> (дата обращения: 24.04.22)
6. При чём здесь Дарвин? Как искусственный интеллект создаёт сам себя [Электронный ресурс]//Skillbox. URL: <https://skillbox.ru/media/business/kak-iskusstvennyy-intellekt-sozdayet-sam-sebya/> (дата обращения: 24.04.22)

7. Сильный искусственный интеллект (Strong AI) [Электронный ресурс]//Машинное обучение доступно. URL: <https://www.helenkapatsa.ru/silnyi-iskusstviennyi-intielliekt/> (дата обращения: 24.04.22)