

УДК 101

ЭТИЧЕСКИЕ ПРОБЛЕМЫ ПРИ СОЗДАНИИ СИСТЕМ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

Аксенов И.Ю., студент гр. ИТм-191, I курс
Научный руководитель: Баумгартэн М.И., к.ф.-м.н., доцент
Кузбасский государственный технический университет
имени Т.Ф. Горбачева
г. Кемерово

Ошибка многих разработчиков искусственного интеллекта состоит в том, что они пытаются запрограммировать компьютеры таким образом, чтобы они функционировали подобно человеческому разуму. Они хотят достичь поставленной цели, обойдя вниманием вопрос о сути разума, о том, что означает слово «понимать» [1].

Несомненно, новые технологии меняют нашу обыденную, повседневную жизнь, а также приносят с собой большое количество вопросов, в том числе и этические, которых раньше не возникало. Изменения в жизни человечества, которые может принести и уже приносит искусственный интеллект, трудно сравнить с тем, что появлялось ранее. Человечество может избавиться от большинства известных профессий, а потенциально и создать новую форму жизни.

Задействование искусственного интеллекта может быть применимо практически в любой сфере деятельности. Поиск и обнаружение неточностей и мошенничества, проведение серьезных исследований, организация сложных систем управления: “умные” компьютеризированные системы делают нашу жизнь лучше. Чем более продвинутыми они становятся, тем проще становится человеку, а значит и весь мир работает продуктивнее.

Во многих случаях это такой же новый рубеж для этики и оценки рисков, как и для новых технологий. Так какие же проблемы влекут за собой создание и совершенствование систем с искусственным интеллектом?

Иерархия труда преимущественно связана с автоматизацией. Изобретая новые способы автоматизации работы, мы можем дать место новым, более сложным профессиям, переходя от физической работы к умственному труду, который наиболее характерен для стратегической работы. Если нам удастся совершить успешный переход, однажды сможем оглянуться и подумать, какими же примитивными вещами мы занимались, тратя на это огромные временные ресурсы.

Появление интеллекта происходит благодаря процессу обучения, не зависимо от того, человек вы или машина. Все системы искусственного интеллекта проходят через фазу тренировки, где они “обучаются” находить корректные паттерны и действовать на основе введенных данных. После

таких “тренировок” система может перейти в фазу тестов, где на основе различных примеров будет изучаться поведение системы и оцениваться то, как она справляется. Конечно же, фаза тренировки, не может покрыть все возможные варианты событий реального мира, с которыми может столкнуться система. Здесь можно сказать, что такую систему можно будет обмануть таким образом, которым не получится обмануть человека.

При помощи абстрактного набора точек, можно заставить систему “видеть” то, чего на самом деле нет. Поскольку мы полагаемся на тот факт, что искусственный интеллект приведет нас к более совершенному труду, эффективному и безопасному, мы должны быть уверены, что системы ведут себя корректно и что другие не смогут получить доступ к контролю над такой системой для использования в личных целях.

Так как такие технологии будут являться очень могущественными, существует вероятность, что ее будут использовать в низких целях. Здесь можно сказать как о роботах, которые будут заменять солдат (людей) на полях сражений, или автономном оружии, так и о системах искусственного интеллекта, призванных навредить. Поскольку битвы в полях с использованием таких технологий не единственная проблема, то кибербезопасность будет очень важным аспектом таких систем.

Даже в том случае, если создатели алгоритма машинного обучения не преследуют цель убивать людей и приносить иной вред, они, зачастую, хотят извлечь выгоду. То есть, можно сказать, что даже если алгоритм работает на благо общества, скорее всего он, таким образом, приносит большую выгоду создателю.

Это можно проследить даже сейчас в области медицины, где важнее всего не вылечить пациента, а порекомендовать ему побольше различных препаратов, зачастую, даже особо и не нужных. Да, зачастую такие препараты не навредят, но и польза от них минимальна. В таком случае, если машинное обучение советует что-либо платное – с большой долей вероятности, такой алгоритм направлен на заработок.

Бывают и случаи, когда само общество, заинтересовано в том, чтобы полученный алгоритм был образцом морали. К примеру, существует зависимость между скоростью движения транспортных средств и смертностью на дорогах. Можно было бы ограничить скорость движения до 20 км/ч, и, тем самым, снизить смертность до минимума, но тогда в крупных городах жизнь стала бы существенно труднее.

Но беспокоится стоит не только о тех, кто заинтересован, чтобы использовать ИИ в низких целях. Что будет в том случае, если сам искусственный интеллект повернется против нас? Это не в прямом смысле означает то, что он станет враждебным к нам. Не голливудское представление “злого” ИИ. Скорее здесь можно привести пример “джина из лампы”, который будет исполнять наши желания. Но последствия, которые последуют после такого желания, могут быть чудовищными, хотя зачастую мы и сами того не желали.

В случае ИИ, это не будет проявлением какой-то “злости”, а просто будет являться недостатком понимания контекста, загаданного нами желанием. Представим систему с ИИ, которой поставили цель искоренить какой-либо вирус. После большого количества вычислений, система выдает формулу, которая действительно искореняет данный вирус, убив все живое на планете. Алгоритм достиг поставленной цели «искоренить вирус», и это эффективный способ, только вот совершенно не желательный и не гуманный.

Представим такую ситуацию: мы даем задачу алгоритму сверстать бюджет страны, ставя следующие основные цели - максимизировать продолжительность жизни, ВВП, а так же производительность труда. При постановке такой задачи мы не ставим никаких этических ограничений. При анализе алгоритм приходит к выводу, что на защиту окружающей среды, детские дома, дома престарелых, не нужно выделять денег, ведь это не увеличивает ВВП.

Вдумайтесь, это всего лишь бюджет, порученный алгоритму, а если задача будет стоять масштабнее, то выйдет так, что все неработоспособное население будет “выгоднее” попросту убить, для того, чтобы повысить производительность труда. Это означает то, что при создании таких систем и алгоритмов, этические вопросы, должны быть изначально, как одна из целей.

Причина, по которой люди находятся на вершине пищевой цепочки не в том, что у нас острые зубы или сильные мышцы. Доминирование человечества строится почти полностью на нашей изобретательности и интеллекте. Мы можем получить лучшее от больших, быстрых и более сильных животных, потому что мы можем создать инструменты для их контроля: физические, вроде клеток и оружия, и когнитивные, вроде дрессировки.

Это ставит серьезный вопрос об искусственном интеллекте: сможет ли он однажды получить такое же преимущество над нами? Мы не можем полагаться только на вытягивание вилки из розетки, потому что достаточно продвинутая машина сможет предвидеть этот шаг и защитить себя. Это будет тот самый момент, когда человеческие существа перестанут быть наиболее умными на Земле.

Некоторые этические вопросы касаются смягчения стараний, некоторые — риска негативных последствий. Пока мы рассматриваем эти риски, мы также должны держать в голове то, что в целом этот технологический прогресс означает лучшую жизнь для всех. Искусственный интеллект имеет огромный потенциал, и его ответственное использование зависит от нас.

Список литературы:

1. Джефф Хокинс Об интеллекте – Times Books, 2004. – 272 с. - URL: <https://web.archive.org/web/20060218002137/http://ebooks.palm.com/product/detail/17826>