Секция «Философия и методология науки»

Влияние современных интерпретаций ИИ на формирование его проблемного поля.

Научный руководитель – Алексеев Андрей Юрьевич

Антипова Анна Викторовна
Студент (бакалавр)
Московский государственный университет имени М.В.Ломоносова, Философский факультет, Кафедра философии и методологии науки, Москва, Россия
E-mail: annaantipova1415@gmail.com

Анализируя современные тренды развития искусственного интеллекта, можно прийти к выводу о том, что ИИ вышел на плато, т.е. достиг того уровня развития, когда технология масштабируется и применяется в увеличивающемся количестве сфер, но при этом не изменяется качественно. С одной стороны, прогресс в данной области обеспечивается увеличивающейся скоростью обработки данных и тормозится в той же мере медленным развитием эффективных алгоритмов, с другой стороны, это может оказаться показателем принципиальных ограничений в развитии технологии. Системы на основе ИИ стали лучше анализировать данные, но отнюдь не показали себя более сознательными. Искусственный интеллект нашел применение в медицинской диагностике, искусстве, решении экологических проблем, государственном управлении, банковском секторе. Данные сферы становятся зависимыми от эффективной обработки данных и других инструментов ИИ. Также с каждым годом все более увеличиваются темпы автоматизации производства, которые гипотетически могут привести к массовой безработице. Глубокое обучение будет и дальше менять нашу жизнь, поэтому реакционно появляются пессимистические настроения, которые включают в себя мнения об угрозе ИИ и огромных этических проблемах, связанных с его применением.

В действительности, около 40 процентов статей об искусственном интеллекте в крупнейших мировых изданиях посвящены именно этическим вопросам или теме угрозы обществу. Тем не менее, есть две основные трактовки данных проблем: со стороны вероятного появления общего искусственного интеллекта (ОИИ) и со стороны неурегулированного использования данной технологии. Обе концепции имеют сторонников в современном философско-научном дискурсе, однако необходимо сделать выбор в пользу одной интерпретации, поскольку они могут противоречить друг другу в плане выбора методов и целей.

Намечающиеся контуры проблем относятся к проблемам регулирования и постановки целей. Сторонники концепции ОИИ считают, что, однажды неизбежно появившись, он уничтожит все человечество, поэтому необходимо сразу озаботиться условиями регулирования ИИ и создать что-то вроде универсальной кнопки отключения или изначально не допустить слишком большой зависимости от решений, который принимает искусственный интеллект. Главное, на ранних этапах осознать потенциальную опасность, исходящую от самого ИИ, и контролировать его развитие, возможно, начать разрабатывать права роботов.

Вторую точку зрения кратко можно выразить в следующей цитате из работы Н. Винера «Кибернетика и общество»: «эта опасность заключается в том, что подобные машины, хотя и безвредны сами по себе, могут быть использованы человеком или группой людей для усиления своего господства над остальной человеческой расой». Представители данной концепции отводят искусственному интеллекту место универсальной технологии и не более. Тем не менее, ее внедрение и распространение происходит слишком быстро и не

включает в себя рабочую модель человеческих ценностей. Данная модель рассматривает будущее ИИ не как появление человекоподобного коллеги или сознательного маньяка-убийцы, а интеллектуального инструмента.

Мы считаем, что основные недопонимания происходят из-за недостаточно разработанной концепции интеллекта, которая на текущий момент опирается на понятия мышления, воображения, воли и рефлексии, которые считаются специфично человеческими. Пока не существует убедительных исследований, проводящих взаимосвязь между количеством логических операций в секунду и появлением ментальных феноменов. Мы считаем, что следует обособить способность к быстрым логическим вычислениям и поиску корреляций в данных, поскольку искусственный интеллект по этим показателям превосходит естественный, тем не менее, никаких признаков сознательного поведения (как бы мы его ни определяли) не демонстрирует.

С этим также связан феномен «интенциональности», который описывает Дэниел Деннет как способность видеть в объектах, чье поведение мы хотим предсказать, рационального агента. Сложно культивировать в себе безжалостный скептический подход к андроидам, чат-ботам или голосовым помощникам, которые обучались имитировать человеческие модели поведения. Можно заметить, что именно это ставил целью пионер ИИ Джон МакКарти на Дартмутском семинаре, говоря, что «наше исследование основано на предположении, что любое свойство интеллекта может быть столь точно описано, что машина сможет его *симулировать*».

Однако данная модель отнюдь не должна приостанавливать развитие когнитивных наук, поскольку если мы хотим улучшить качества ИИ как интеллектуального инструмента и помощника, необходимо разработать модели ценностного выравнивания и человеческого поведения, которые помогут роботу выбрать оптимальную стратегию поведения, которая не искажала бы человеческие ценности, имплицитно в него заложенные. Заметим, это не означает, что мы предлагаем делать роботов с развитым моральным чувством, мы лишь увеличиваем объем данных, на которые необходимо обращать внимание при анализе определенной проблемы. Для этой же цели необходимо изучать человеческие способы обработки информации для поиска более эффективной структуры данных или нового алгоритма. Вторая модель также предлагает больше акцентировать внимание на опасностях, исходящих от людей, которые используют масштабную технологию в собственных интересах или намеренно искажают информацию, на которой обучается ИИ.

Стивен Пинкер справедливо полагает, что если отбросить фантазии о цифровом всеведении и контроле за каждой частицей во Вселенной, становится очевидным, что ИИ схож с любой другой технологией, «он разрабатывается поэтапно, предназначается для удовлетворения многочисленных условий, тестируется перед внедрением и постоянно дорабатывается в целях эффективности и безопасности». Таким образом, выбирая вторую модель интерпретации, мы предпочитаем насущное решение проблем в области искусственного интеллекта спекулятивным рассуждениям о его сознательном будущем.

Источники и литература

1) Искусственный интеллект – надежды и опасения : [сборник : перевод с английского В. Желнинова] / под ред. Джона Брокмана. – Москва : Издательство АСТ, 2020. – 384 с. – (Наука, идеи, ученые).

2) Искусственный интеллект: Пределы возможного / Мередит Бруссард ; Пер. с англ. — М. : Альпина нон-фикшн, 2020. — 362 с.

3) Цветкова Л. А. Технологии искусственного интеллекта как фактор цифровизации экономики России и мира // Экономика науки. 2017. №2.

4) Н. Винер. Кибернетика и общество. Перевод Е. Г. Панфилова / Общая редакция и предисловие Э. Я. Кольмана. - Издательство иностранной литературы, Москва, 1958

5) D.Dennett. Kinds of Minds: Towards an Understanding of Consciousness . - Weidenfeld & Nicolson, 1996