ФИЛОСОФСКИЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Титов А.В.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Ратникова И.М. – кандидат философских наук, доцент

Работа посвящена рассмотрению вопросов о природе и сущности искусственного интеллекта, его атрибутивных свойствах и границах применения.

Введение. В данный момент активно развивается индустрия искусственного интеллекта (далее ИИ). В этой связи обсуждение философских проблем, связанных с ИИ, является важнейшим элементом безопасного развития технологий, переосмысления человеческой природы и сохранения её идентичности. ИИ принято разделять на два вида: слабый и сильный. Слабый ИИ – системы, предназначенные для решения узкоспециализированных задач. Сильный ИИ – гипотетические системы, способные воспроизводить все аспекты человеческого интеллекта, включая сознание и свободу воли. Основная философская проблема, связанная с искусственным интеллектом: возможно ли воссоздать человеческий разум. Если ответ отрицательный, то все остальные проблемы по этой теме не имеют смысла. В данной статье основные проблемы будут рассмотрены в контексте концепции сильного ИИ, так как именно здесь поднимаются наиболее сложные вопросы о природе разума, морали и будущего человечества.

ИИ и сознание. В научной литературе до сих пор нет единого подхода к исследованию и определению сознания, что значительно усложняет поставленную перед нами задачу. Естественный интеллект (далее ЕИ) – это сложнейшая система, природа которой и сегодня является предметом острых дискуссий. Можно выделить два основных подхода к его пониманию. Сторонники ЕИ как сложного механизма рассматривают человеческий разум как высокоорганизованную биологическую систему обработки информации. По словам Дэниела Деннета, сознание – это иллюзия, создаваемая параллельной обработкой информации в мозге [1]. И нельзя отрицать, что, если получится понять, каким образом работает человеческий мозг, не составит труда применить полученный алгоритм и для машин. С другой стороны Карл Поппер утверждает, что сознание – это не просто процесс обработки информации, а более неуловимый и таинственный феномен [2]. Нет никаких гарантий, что, следуя “волшебному алгоритму разума”, ИИ не окажется жертвой эксперимента “Китайской комнаты” Джона Сёрля, в котором он лишь имитирует понимание, но не обладает внутренним опытом [3].

В данном случае ответ будет зависеть от того, как мы определяем сознание и какие критерии считаем необходимыми для его возникновения. Ведь разум - это не только о решении поставленных задач, но и о чувственном опыте, оказывающем непосредственно влияние на наше сознание.

С одной стороны, ИИ может имитировать эмоции, распознавая настроение пользователей и реагируя соответствующим образом, но это не означает, что он действительно их испытывает. Эмоции связаны с биологическими процессами, такими как выброс гормонов и активность нейронов, которые отсутствуют у машин.

С другой стороны, Марвин Мински считает, что эмоции – это не только биологический феномен, но и результат обработки информации [4]. Если ИИ сможет воспроизвести сложные процессы, лежащие в основе эмоций, возможно, он сможет испытывать что-то похожее. Но даже в этом случае остается вопрос: будут ли эти “эмоции” субъективными переживаниями или просто алгоритмическими реакциями?

ИИ и свобода воли. Свобода воли в широком смысле определяется как способность индивида принимать решения, которые не предопределены внешними обстоятельствами или внутренними состояниями. В то же время эксперименты Бенджамина Либета продемонстрировали, что нейронные процессы в мозге инициируют действия за сотни миллисекунд до их осознанного принятия, что ставит под сомнение саму концепцию свободного выбора даже у человека. Эти данные получили развитие в работах Роберта Сапольски [5], который утверждает, что все человеческие решения, в конечном счете, определяются сложным переплетением генетических факторов, влияния среды и нейрофизиологических процессов, оставляя мало места для традиционного понимания свободы воли. В контексте ИИ эта проблема приобретает особую остроту – современные ИИ-системы принимают решения исключительно на основе предопределенных алгоритмов и обучающих данных, что делает их поведение принципиально детерминированным.

Даже если у ИИ будет способность к рефлексии и возможность менять свой внутренний код, это может приблизить его к понятию свободы воли, так как он анализирует свои действия и опыт, на основе которого может принять решение об изменении своего поведения или состояния.

Однако даже в этом случае остается вопрос: является ли такая “свобода” настоящей или же это просто более сложная форма детерминизма? Ведь способность к рефлексии и изменению

кода тоже будет запрограммирована разработчиками. Таким образом, свобода воли ИИ, даже с учетом рефлексии и самоизменения, может быть ограничена рамками его изначальной архитектуры и целей.

ИИ и мораль. Говоря о этических проблемах, имеется в виду несоответствие поведения искусственного интеллекта нормам морали. Сложность заключается в том, что не так-то просто провести грань между добром и злом, что является правильным, а что нет. Если рассматривать три золотых правила робототехники, сформулированные Айзеком Азимовым, то возникнет проблема, к примеру, как перевести на любой из известных языков программирования термин “причинить вред” или “допустить”. Эти понятия, кажущиеся интуитивно понятными для человека, оказываются крайне сложными для формализации в машинном коде. Например, что считать “вредом”? Физический ущерб, эмоциональную травму или даже косвенные последствия, такие как экономический ущерб? И даже если эти термины получится как-то переформулировать, важно будет проверить, как сама система ИИ их поняла.

Ещё одной этической проблемой является интерпретация моральных норм. Мораль – это не статический набор правил, а динамическая система, которая может значительно различаться в зависимости от культуры, социального положения и истории общества. То, что будет нормой для одних, может быть неприемлемым для других. Поэтому важным вопросом будет обсуждение того, какие нормы нужно закладывать в его алгоритмы.

Кроме того, возникает вопрос о гибкости моральных решений. В реальной жизни люди часто сталкиваются с дилеммами, где нет однозначно правильного ответа. В той же всем известной проблеме вагонетки перед искусственным интеллектом будет стоять непростая задача выбора о направлении движения вагонетки. Может ли он считать что-то более важным, чем человеческая жизнь, и чем он объяснит такое решение.

Также проблемой будет обучении ИИ моральным нормам. Современные системы ИИ обучаются на данных, созданных людьми, и поэтому они могут перенять человеческие предрассудки и ошибки. Например, если ИИ обучается на исторических данных, которые отражают дискриминацию и несправедливость, он может воспроизводить эти проблемы в своих решениях. Одним из способов решения этой предвзятости может служить обучение на репрезентативных и разнообразных данных, охватывающих различные расовые, гендерные и культурные группы. Другим способом является обеспечение того, чтобы алгоритмы обучения были разработаны таким образом, чтобы обнаруживать предвзятости и исправлять их. Но даже в этих случаях придется согласовать и провести тщательный анализ, что можно считать предвзятостью, а что нет.

При этом возникает противоречие: если нормы морали будут определены человечеством, то нельзя будет говорить о таком понятии для ИИ, как свобода воли. Ведь все его действия будут основываться на общепринятых и оговоренных стандартах. В данном случае будет играть роль то, чему мы отдаем приоритет: контроль и безопасность, чтобы использовать его как инструмент, или же свобода и риск, ведь создав что-то независимое, мы не сможем это подчинить и управлять им, давая ему права выбора и самостоятельность.

Вывод. В статье были рассмотрены основные философские проблемы искусственного интеллекта. Главный вопрос заключается в том, возможно ли воссоздать человеческий разум в искусственном интеллекте, и если да, то насколько это безопасно. Опасность заключается в том, что ИИ со свободой воли может выйти из-под контроля, если его действия не будут ограничены нормами морали. Однако, если свобода воли ИИ будет жестко регламентирована, возникает парадокс: можно ли считать это истинной свободой воли, или это будет лишь иллюзия автономии, ограниченная рамками программирования? Таким образом, ключевая дилемма состоит в поиске баланса между свободой ИИ и безопасностью для человечества.

Список использованных источников:

1. Dennet, D.C. Are we explaining consciousness yet?– Cognition, 79 (1-2). – P. 221-237.
2. Поппер К., Экклс Дж. “Я и его мозг”.– М.: Прогресс, 2008. – 512 с.
3. Сёрль Дж. “Китайская комната” // Философия сознания. – М.: Идея-Пресс, 2002. – С. 96-188.
4. Мински М. “Общество разума”. – М.: ИД ЯСК, 2018. – 456 с.
5. Сапольски Р. “Биология добра и зла”. – М.: Фльпина нон-фикшн, 2020. – 776 с.