

Критика искусственного интеллекта Х. Дрейфусом. Современный взгляд

Научный руководитель – Кузнецов Антон Викторович

Власова Дарья Александровна

Студент (магистр)

Московский государственный университет имени М.В.Ломоносова, Философский факультет, Кафедра истории зарубежной философии, Москва, Россия

E-mail: daalz@mail.ru

Научные исследования в области искусственного интеллекта (ИИ) являются одной из стремительно развивающихся областей прикладной науки. Некоторые футурологические прогнозы относительно успехов искусственного интеллекта, которые были сделаны более двадцати лет назад, воплощаются в жизнь. Например, Рэй Курцвейл предсказывал, что в 1998-м году компьютер победит человека в шахматы. Это произошло на год раньше 1997-м году Deep Blue (IBM) победил в шахматной игре Гарри Каспарова [3]. С этого времени разработки в области ИИ развиваются еще стремительнее. XXI век можно назвать веком высоких технологий, несомненно, искусственный интеллект стоит в их авангарде.

Однако стоит признать тот факт, что не все смелые прогнозы относительно создания искусственного интеллекта сильного типа сбылись. В конце XX века стремительное развитие компьютерных технологий стало основой для смелых прогнозов, несмотря на то, что со многими практическими проблемами разработчики программ искусственного интеллекта научились справляться, решение этих проблем, не приблизило нас к созданию искусственного интеллекта сильного типа. Различение «слабого» и «сильного» ИИ произвел Дж. Сёрл в 1980 году [2]. Под «слабым» ИИ можно понимать те программы, которые способны решать информационные задачи, оперировать символами и довольно большими объемами данных. Кроме того, существует представление и о «сильном» ИИ, предполагается, что такая программа будет не просто оперировать символами по определенным правилам, но понимать смысл производимых ею действий и смысл используемых символов, на подобие того, как действует человек. Такая программа будет сравнима с естественным интеллектом в способности понимать и рефлексировать.

Исследователей в области ИИ условно можно разделить на два лагеря: тех, кто допускает и даже уверен, что создание «сильного» ИИ возможно, и тех, кто принципиально отрицает такую возможность. Одним из тех мыслителей, кто отрицал возможность создания ИИ сильного типа, был Хьюберт Дрейфус (1929-2017 гг.). В его ключевой работе на тему критики ИИ «Чего не могут вычислительные машины: Критика искусственного разума» (1972) он представил пессимистическую оценку развития искусственного интеллекта и ряд ключевых аргументов против возможности создания «сильного» ИИ на базе существующих компьютерных технологий [1]. Хьюберт Дрейфус хотел понять, почему программисты так уверены в том, что смогут создать искусственный аналог человеческого разума, который не будет уступать в способностях. Он приходит к выводу, что они принимают некоторые необоснованные допущения. Цель философа - оспорить правомочность этих допущений:

1. Биологическое допущение - упрощение работы сознания. Представление о том, что на уровне нейронов процесс передачи информации носит дискретный характер и происходит по довольно определенной и простой схеме, следовательно, этот процесс можно воплотить на не биологическом носителе, например, электрической схеме.

2. Психологическое допущение - упрощение тех функций, которые осуществляет наше сознание. Представление о том, что наше мышление - это процесс обработки информации

согласно определенным формальным правилам. Так как всю информацию можно представить в виде двоичного кода, следовательно, точно такие же функции доступны и ЭВМ.

3. Эпистемологическое допущение - вся информация, которую «обрабатывает» человеческое сознание, может быть формализована и выражена в терминах логических отношений, точнее, в терминах булевых функций - логического исчисления, задающего правила обращения с информацией, заданной в двоичном коде.

4. Онтологическое допущение - вообще все, что может быть существенно для разумного поведения, может быть представлено в дискретных, четко определенных терминах.

Основные моменты критики Х. Дрейфусом этих допущений:

1. Биологическое допущение может иметь место, только если мозг действует как цифровая машина. То есть, при передаче и обработке информации, он связывает единицу информации с символом некоторого описательного языка. Хьюберт Дрейфус ссылаясь на данные нейробиологических исследований, отмечает, что наш мозг работает скорее, как аналоговая система, при чем, помимо электрических импульсов, в нейронах происходят и химические реакции.

2. Критика психологического допущения, строится на основе философских взглядов Х. Дрейфуса и данных нейро-наук. Х. Дрейфус считал, что если мы примем такое допущение, то нам грозит не пришествие сверхразумных машин, а появление неполноценно мыслящих человеческих существ. Человек не может жить по правилу потому, что правильный способ жизни придумать невозможно. Жизнь человека - настолько разнообразна и непредсказуема, что требует от нас гибкости ума.

3. Эпистемологическое допущение Х. Дрейфус критикует с позиции философии здравого смысла. Здесь он делает акцент на том, что даже если мы и можем описать некоторые процессы, происходящие в мире на формальном языке, это не значит, что именно они происходят в действительности. Реальное действие не может быть основано на формализации.

4. Онтологическое допущение строится на том, что мир и все события, происходящие в нем, можно проанализировать в рамках атомарных фактов, не связанных друг с другом. Х. Дрейфус, используя феноменологическую аргументацию и идеи Мерло-Понти, пытается доказать обратное: мир, континуален, а не атомарен, все события и вещи мира тесно взаимосвязаны, образуя единство жизненного пространства.

Критика ИИ Х. Дрейфусом во многом является философской. В ходе исследования было выявлено, что технические ограничения, связанные с развитием ИИ могут быть преодолены, что позволило вывести программы на качественно новый уровень. Многие из критических замечаний относительно развития ИИ оказались несостоятельны, тем не менее они стали стимулом к развитию технологий в рамках поиска альтернативных методов, кроме того, ключевые идеи в концепции Х. Дрейфуса актуальны и по сей день. Допущения, на которых основываются программисты так и не преодолены. Возможно, именно здесь надо искать объяснение трудностей на пути создания «сильного» ИИ.

Источники и литература

- 1) Дрейфус Х. Чего не могут вычислительные машины. М.: Прогресс, 1978.
- 2) Сёрл Дж. Р. Сознание, мозг и программы // Аналитическая философия: становление и развитие. М., 1998.
- 3) Каспаров против Deep Blue | Матч, изменивший историю [Электронный ресурс]. — Режим доступа: <https://www.chess.com/ru/article/view/kasparov-protiv-deep-blue-match-izmenivshii-istoriiu> (дата обращения 02.02.2020).