

**Способен ли искусственный интеллект быть личностью? Аргументация
Александра Прусса**

Научный руководитель – Кузнецов Антон Викторович

Михайлов Иван Сергеевич

Студент (бакалавр)

Московский государственный университет имени М.В.Ломоносова, Философский
факультет, Москва, Россия

E-mail: ivansergeevich99@gmail.com

В дискуссии по проблеме тождества личности и связи этой проблемы с искусственным интеллектом существуют различные позиции, в том числе и обосновывающая то, что искусственный интеллект не способен быть личностью. Одним из представителей данного "лагеря" можно считать канадского философа Александра Прусса. Объектом моего исследования главным образом являются способы построения системы аргументов в его работе "Искусственный интеллект и тождество личности"[2].

Аргументация данной работы Прусса строится на нескольких глобальных предпосылках. Вот главные из них:

- допущение способности искусственного интеллекта быть личностью;
- отрицание различия между понятиями «компьютер», «робот», «искусственный интеллект».

В своей работе Прусс также рассматривает несколько вариантов постановки вопроса о проблеме тождества личности. Для систематизации исследования его аргументации следует выделить основные из них:

- вопрос о тождестве во времени;
- вопрос о количестве личностей.

Центральным звеном постановки вопроса о тождестве во времени является сконструированная ситуация наличия некоего робота. Вокруг этой ситуации строятся дальнейшие умозрительные эксперименты, завязанные на переходе робота от включенного состояния к выключенному и наоборот.

Для ответа на вопрос о количестве личностей Пруссом выделяется два подхода: аппаратный и программный. Общим звеном этих подходов является их цель: с помощью выяснения количества возможных личностей в условиях того или иного умозрительного эксперимента попытаться ответить на вопрос о возможности существования личности в заданных условиях в принципе.

Центральным же звеном всей работы можно считать метод *reductio ad absurdum* (доведение до абсурда), применяемый Пруссом в построении аргументации. Работает это по такому алгоритму:

- 1) допущение способности искусственного интеллекта быть личностью; 2) выбор подхода к проблеме тождества личности; 3) постановка вопроса о способности ИИ быть личностью в рамках выбранного подхода; 4) приведение аргументов, подтверждающих возможность наличия у ИИ выше обозначенной способности; 5) приведение аргументов, опровергающих эту возможность; 6) за получением в итоге ситуации неразрешённого противоречия между приведёнными аргументами признание подобной ситуации абсурдной; 7) избавление от абсурда путём отрицания исходной посылки, то есть исходного допущения о способности ИИ быть личностью; 8) вывод о том, что ИИ не способен быть личностью.

В принципе, получение подобного вывода фактически можно считать и основной целью работы «Искусственный интеллект и тождество личности». Однако в своём исследовании я собираюсь сосредоточиться больше не на мотивационной составляющей работы, а на анализе аргументации Прусса, выявлении её особенностей, их дальнейшей классификации и оценке, то есть рассмотрении этих особенностей уже как слабых и сильных сторон.

Одним из главных спорных моментов в исследуемой мною работе является наличие скрытых предпосылок при построении аргументации. Самой масштабной из них стоит признать существование такой субстанции как душа. «Искусственный интеллект и тождество личности» Прусса не содержит на своих страницах ни чётко оформленного понятия души, ни ссылок на его определение в каких-либо других работах его или чьего-либо ещё авторства. При этом большую часть приводимых Пруссом аргументов можно принять за действующие, лишь исходя из данной предпосылки. Лишь в примечаниях к своему труду автор оставил хоть какую-то зацепку, отметив схожесть построения одного из своих аргументов с построением аргумента в пользу существования души Ричарда Суинберна [4]. Поэтому в своём исследовании я также обозначу свою точку зрения на то, насколько исчерпывающей можно считать данную ремарку, не говоря уже о проверке оставшейся аргументации.

Источники и литература

- 1) Джерт Б. Систематический подход
- 2) Alexander R. Pruss. Artificial Intelligence and Personal Identity // International Conference, Baylor University, Waco, Texas, U.S.A., 2008
- 3) Сингер П. Мораль Истории: Антология Этики Через Литературу и Этику биологических исследований
- 4) Richard Swinburne. Substance Dualism // International Conference, Baylor University, Waco, Texas, U.S.A., 2008