

Секция «Философия техники (совместно с кафедрой философии техники Дрезденского технического университета, Германия)»

## **Сильный искусственный интеллект: проблема субъектности / Strong Artificial Intelligence: The Problem of Subjectivity**

**Научный руководитель – Яковлев Алексей Александрович**

*Голова Варвара Алексеевна*

*Студент (бакалавр)*

Российский университет дружбы народов, Факультет физико-математических и естественных наук, Москва, Россия

*E-mail: varya-g@mail.ru*

Каждому современному человеку знакомо словосочетание «Искусственный интеллект». Но что такое сильный искусственный интеллект? И сможет ли ИИ когда-нибудь мыслить и понимать все так же, как человек, или же все-таки у будущего искусственного интеллекта существует проблема субъектности?

Начнем с актуальности данной темы. Искусственный интеллект на данный момент одна из наиболее изучаемых тем. Все прекрасно понимают, что в том или ином виде он будет все больше появляться в нашей обыденной жизни. К примеру, уже сейчас существует «Умный дом», в некоторых городах используется беспилотный транспорт, например, в Стокгольме - беспилотные автобусы, в Нюрнберге - беспилотное метро, а также в России тестируется беспилотный грузовой автомобиль КамАЗа.

Существует гипотеза о сильном искусственном интеллекте, но для начала хотелось бы напомнить, что такое искусственный интеллект. По сути это свойство интеллектуальных систем выполнять творческие функции, то есть функции, которые может выполнять только человек. ИИ должен научиться быть похожим на человека.

Один из самых известных экспериментов на тему искусственного интеллекта - тест Тьюринга. Этот тест был придуман английским математиком Аланом Тьюрингом в 1950 году. Главной задачей теста Тьюринга являлся ответ на вопрос - может ли искусственный интеллект имитировать человека.

Американский философ Джон Сёрл в 1980 г. сформулировал мысленный эксперимент «Китайская комната». Он опровергает тест Тьюринга и даже гипотезу «сильного искусственного интеллекта». Этот эксперимент показывает, что хотя, возможно, машина сможет себя повести как человек, она, скорее всего, не будет осознавать, что она делает, так как будет действовать просто по инструкции, то есть никаких человеческих качеств машина в таком случае не обретет.

Ну что же, теперь перейдем уже к самой гипотезе о сильном искусственном интеллекте. Эта гипотеза заключается в том, что компьютеры могут мыслить и осознавать себя как отдельную самостоятельную личность, от части понимать свои мысли. Сам способ мышления у искусственного интеллекта может отличаться от человеческого. На данный момент уже есть пример проявления самосознания у роботов - в 2015 году был проведен эксперимент, в ходе которого роботы проявили базовые признаки самосознания.

Сильный искусственный интеллект должен уметь принимать решения и действовать в нестандартных ситуациях, иметь знания и представление о реальности, строить планы, быть способным к обучению и уметь общаться на естественном языке. Для лучшего понимания также приведу понятие слабого искусственного интеллекта.

Слабый искусственный интеллект по сути отвергает возможность создания искусственного интеллекта, который был бы способен мыслить, как человек, то есть слабый искусственный интеллект допускает решение задач по инструкции, к примеру определять голос или распознавать изображение на картинке.

Необходимо также поговорить о нейронных сетях. Искусственная нейронная сеть - это [https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\\_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C), а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования [https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F\\_%D1%81%D0%B5%D1%82%D1%8C](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C) — сетей <https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD> живого организма. Нейронная сеть является одним из способов реализации ИИ. Машинное обучение включает в себя методы искусственного интеллекта, которые стремятся к тому, чтобы ИИ в процессе решения задачи мог обучиться, как решать задачи, на основе решений схожих задач. Таким образом, с помощью нейронных сетей, которые дают способность к самообучению, может работать машинное обучение, чтобы ИИ работало на основе уже полученного опыта.

Теперь перейдем к субъектности искусственного интеллекта. Так почему же может возникнуть проблема субъектности? На любую ситуацию ИИ может выдать только такие реакции, которые изначально в нем заложены, он не может придумать что-то новое. Из-за этого могут возникнуть проблемы, если произойдет какая-то нестандартная ситуация, ведь машина сможет выдать только такую реакцию, которая была бы выдана как реакция на наиболее схожую ситуацию или не выдать реакцию вообще. Наиболее вероятный путь избежать этого, это использовать машинное обучение для усовершенствования исходных знаний, заложенных в ИИ при создании.

Таким образом, на данный момент вопрос, сможет ли ИИ сравниться с человеком остается открытым. Но я считаю, что в ближайшем будущем, ИИ будет продолжать активно развиваться, однако вопросы, которые решаются человеком через эмоции, будет очень сложно решаться и на данный момент даже невозможно, но максимального приближения к этому человечество сможет достигнуть.