

## Проблема искусственного интеллекта

Философия активно обсуждает актуальные вопросы, в том числе и проблему искусственного интеллекта. Можно ли сделать копию сознания человека? Как опеределить, что перед нами искусственный интеллект? Можно ли вообще сделать искусственное сознание?

Аристотель, например, считал, что разум это свойство исключительно человеческое, и никто в этом не сомневался, вплоть до середины 20 века, когда появились первые компьютеры, способные вычислять. В 1956 году впервые вводится понятие «Искусственный интеллект». Тогда было дано первое определение: «ИИ — это процесс, в рамках которого машины выполняют те задачи, которые потребовали бы интеллектуальных усилий от человека». Поэтому минимальному определению, любой компьютер, телефон, или даже даже стиральная машина — будет искусственным интеллектом.

Но по мнению американского философа, Джона Сёрла, будет лишь слабым искусственным интеллектом. В 1980 году в одной из своих работ, он вводит два понятия, сильный и слабый ИИ. Слабый, всего лишь полезный инструмент человека, а сильный должен не просто решать интеллектуальные задачи, но и обладать сознанием, в человеческом понимании. И вот по поводу сознания, философия рассуждала, в течение всей своей истории.

Платон под понятием Душа понимал его в метафизическом смысле. Христианские богословы, утверждали, что душа это неповторимое творение бога. Декарт вводит понятие Когита, которое и есть в некотором смысле это метафизическое сознание, которое человек объединяет в себе вместе с физическим теле.

Лейбниц в одной из своих работ описал мысленный эксперимент, в котором пытается представить душу человека в виде механического объекта, мельницы. Если человек зайдет внутрь этой мельницы и начнет перечислять все детали из которых состоит этот механический объект, то среди них он в итоге не сможет найти саму душу. Что доказывает по мнению Лейбница невозможность искусственного создания души.

Но как уже упоминалось, в середине 20 века вопрос о сознании пришлось поставить заново. В 1950 году, знаменитый ученный Аллан Тьюринг, публикует в философском журнале статью, в которой ставит вопрос, актуальный по сей день. Может ли машина совершать действия неотличимые от осознанных. Именно тот самый тест Тьюринга должен ответить на поставленный вопрос. Во время теста испытуемый должен вести анонимную переписку с несколькими собеседниками, один из которых робот. По итогам переписки, испытуемому нужно указать, кто из собеседников настоящая личность. Если робот заставляет человека регулярно ошибаться, значит перед нами осознанный искусственный интеллект.

Но уже знакомый нам философ Джон Серл критикует тест и остается убежден, в том, что сознание неповторимо. В 1980 году в одной из своих работ он предлагает провести мысленный эксперимент, под названием «Китайская комната». Представим себе человека в коробке, который абсолютно не знает китайского языка. К этой коробке подходит китаец и бросает в нее иероглиф, в надежде получить ответ. Человек в коробке открывает книгу, в которой есть все исчерпывающий инструкции какими иероглифами отвечать на другие иероглифы. Таким образом, следуя инструкциям, он будет давать адекватные и понятные ответы китайцу, при этом не понимая смысла своих ответов. Джон Серл сравнивает положение человека в китайской комнате с алгоритмами компьютера, как бы доказывая нам, что искусственный интеллект не может осознать своих действий, ровно так же как не может осознать их человек в коробке.

Возвращаясь к тесту Тьюринга, относительно недавно, в 2014 году машины научились его проходить без особых проблем. Поэтому был предложен новый концепт «Схема винограда». Это тест для испытания предельных возможностей компьютера. Если машина сможет определить подразумеваемый субъект, основываясь на подсказках из контекста, значит машина использует рассуждения на основе здравого смысла, а следовательно может считаться разумной. Например такой вопрос: «Трофей не смог поместиться в чемодан, потому что он слишком большой». Человек с легкостью ответит на этот вопрос, потому что он знает как работают контейнеры, что большое не может поместиться в маленьком, и следовательно, поймет, что речь идет о трофее. Второй пример: «Я расстелил скатерть на столешнице, чтобы защитить ее» Человек знает, что такое скатерть и что такое столешница.

Американский философ, Дэниел Денед глубоко убежден, создать ИИ возможно. По его мнению человек слишком идеализирует сознание, не стоит относить его к области метафизического. На самом деле то, что мы называем сознанием, ни что иное как продукт человеческой эволюции, который просто является легкой для нас формой описания окружающего мира. Денед основывает эту позицию на своей теории о интернациональной

установке.

Интернациональность — свойство сознания быть направленным на что либо. Философ приводит простой пример на шахматной игре с компьютером. Во время игры нам стоит задача выиграть. Когда наступает наш ход, мы пытаемся рассчитать следующие ходы противника. Чтобы это сделать в случае с компьютером, мы можем вычислить ход на уровне устройства и строения шахматного компьютера, используя ученых со всего мира и обладая бесконечным запасом времени, но в условиях реальной игры и недостатка времени мы конечно будем действовать предписывая компьютеру интернациональность, а именно желания, опасения и надежды. Как мы обычно делаем и в жизни, общаясь с себе подобными, мы контактируем орудуя категориями желаний, опасения и надежд.

Так почему же тогда если мы волей неволей как бы одушевляем в компьютер во время шахматной игры, у нас есть вера в существование некоего особого сознания, которое якобы присуще только человеку? Может быть это все таки просто иллюзия?