

**УДК 16:007**

## **НЕЯВНАЯ ПРЕДПОСЫЛКА РАССУЖДЕНИЙ О ТЕХНОЛОГИЧЕСКОЙ СИНГУЛЯРНОСТИ**

*Степанов В.В.*

*В данной статье рассматривается явление технологической сингулярности, история изучения данной проблемы и философские перспективы её исследования. Выявляется имплицитно содержащееся, но невысказанное утверждение о разуме, которое является краеугольным камнем выстраиваемых систем технологической сингулярности.*

**Ключевые слова:** *искусственный интеллект, технологическая сингулярность, разум.*

Искусственный интеллект (ИИ) изучался в течение десятилетий, и по-прежнему остаётся одним из самых неуловимых предметов в компьютерных науках. Этот предмет огромен и туманен, отчего, частично, он всё ещё плохо исследован. Предметная область исследования ИИ раскинулась от машины, действительно способной мыслить, до поиска алгоритмов, используемых в настольных играх. Он используется почти во всех сферах общества, в которых применяются компьютеры.

Гомер упоминает о самодвижущихся треножниках (373) [2]. С тех пор мысль о механических помощниках сопутствует человечеству на протяжении всей истории. Гёте в Фаусте устами учёного-книжника Вагнера, который посвятил себя открытию тайн жизни посредством наук и магии, в сцене создания гомункула сообщает о великом человеческом чаянии:

«Нам говорят «безумец» и «фантаст»,  
Но выйдя из зависимости грустной,  
С годами мозг мыслителя искусный  
Мыслителя искусственного создаст» [1, стр. 282].

Термин «искусственный интеллект» был введён Джоном Маккарти в 1956 году на первой проведённой им академической конференции, посвящённой данной теме. Но история самой идеи искусственного интеллекта может быть прослежена гораздо раньше. Некоторые исследователи указывают на несколько источников: философия, математика, компьютерные технологии и т.д. Соответственно, разные даты могут быть взяты как точки отсчёта. Хотя, те предпосылки, которые находят в философии, а именно в работах Аристотеля, Декарта, Гоббса, Паскаля и Лейбница – по большей части касаются разработанных ими методов формального мышления или указаний на схожесть процессов мышления с процессами вычислений. Философские предпосылки представляют собой скорее разработки касательно разума вообще. Соответственно современные исследователи искусственного интеллекта могут

указывать на них как на некоторые исторические варианты того, как рассматриваемую ими проблему решали философы прошлого [14], [5], [11, стр. 3-9]. Так, сюда относят силлогизмы Аристотеля и декартовские правила ума, индукцию как основу познания у рационалистов, или вычислительную процедуру Карнапа для получения знаний из элементарных опытов.

Вэннвар Буш в основополагающей работе «Как мы можем мыслить» [6] предложил систему, которая укрепляет собственные знания людей. Предложенная гипертекстовая система Метех повлияла на целые поколения кибернетиков, которые питали вдохновение в его видении будущего. Пять лет спустя Алан Тьюринг написал статью с упоминанием машин, которые могут симулировать человеческое существо и способность к интеллектуальным действиям, таким как игра в шахматы [15].

Никто не опровергает возможность компьютеров работать с логическими процессами. Неизвестно, может ли машина думать. Так как данный тезис встретил сильное аргументированное сопротивление, становится важным точное определение того, что значит думать. Классический пример – аргумент комнаты Сёрля [4]. Представим, что некто заперт в комнате. Ему посылают записки на китайском языке, а также инструкции, что делать с китайскими иероглифами при поступлении других команд. Инструкций достаточно, чтобы выполнить поступающие команды. Этот некто выполняет команды и переставляет китайские иероглифы в полном соответствии с инструкциями так, что на выходе получается вполне адекватное сообщение, понятное носителю языка. Но запертый в комнате не знает языка. Для него работа с иероглифами ничем не отличается от любой другой работы с ничего не значащим набором символов. Сёрль по пунктно отстаивает позицию, заключающуюся в том, что подобных действий недостаточно, чтобы утверждать, что данная система (закрытый человек, или робот, следующий определённым инструкциям) обладает сознанием.

«Статья Сёрля о Китайской комнате не только поставила новые вопросы философии искусственного интеллекта, но и в некотором плане притормозила развитие этой области аналитической философии сознания. Позиции ученых разделились, по крайней мере, на два направления: 1) философы продолжили изучать отношения синтаксиса и семантики, сознания и критериев интеллектуальности; 2) инженеры забросили «тесты Тьюринга» и принялись за создание роботов, имитирующих основные функции человеческого организма» [3, стр. 22].

Тема искусственного интеллекта напрямую связана с понятием «технологической сингулярности». Технологическая сингулярность, в свою очередь, напрямую связана с «интеллектуальным взрывом» и «взрывом скоростей». Аргумент интеллектуального взрыва был сформулирован Ирвингом Джоном Гудом в его статье 1965 года «Размышления о первой ультраинтеллектуальной машине»:

«Определим ультраинтеллектуальную машину как машину, которая способна значительно превзойти все интеллектуальные действия любого человека, как бы умён тот ни был. Поскольку способность разработать такую машину также является одним из этих интеллектуальных действий, ультраинтеллектуальная машина может построить даже более совершенные машины. За этим, несомненно, последует "интеллектуальный взрыв", и разум человека намного отстанет от искусственного...

Таким образом, первая ультраинтеллектуальная машина – последнее изобретение, которое надо сделать человеку, при условии, что машина будет достаточно покорна и поведает нам, как держать ее под контролем " [9, стр. 33].

Аргумент «взрыва скоростей» (speed explosion) был сформулирован исследователем ИИ Рэем Соломоновым (Ray Solomonoff) в его работе 1985 года «Временная шкала искусственного интеллекта» (The Time Scale of Artificial Intelligence) [13]. Элизер Юджовски даёт краткую версию этого аргумента в статье 1996 года «Вглядываясь в Сингулярность» [18]:

«Скорость вычислений удваивается каждые два субъективных года исследований. Через два года после того, как искусственный интеллект достигнет человеческого уровня, его скорость удвоится. Через один год, его скорость удвоится опять. Шесть месяцев – три месяца – полтора месяца ... Сингулярность».

Термин сингулярность в таком смысле был представлен в 1983 году писателем научной фантастики Вернором Винджем (Vinge) [16], распространен его влиятельной статьёй 1993 года «Приближающаяся технологическая сингулярность» [17]. Популярность термин приобрёл благодаря известной работе футуриста и изобретателя Рэя Куртзвейла «Сингулярность уже близко» [10]. На практике «сингулярность» употребляется в разных смыслах. Нас интересует именно «технологическая сингулярность», то есть последствия «взрыва скоростей» и «интеллектуального взрыва».

Можно предположить, что технологическая сингулярность представляет огромный интерес для академических философов, учёных, занимающихся когнитивными науками и искусственным интеллектом. В действительности, как утверждает Дэвид Чалмерс, дело обстоит иначе. Данная тема, за некоторыми исключениями, представляет интерес в основном для неакадемических кругов, включающих в себя: интернет форумы, популярные книги, а также мастерские, организованные Независимым Институтом Сингулярности [7; стр. 3].

Дэвид Чалмерс считает, что, несмотря на пренебрежение идеей технологической сингулярности в академических кругах, следует отнестись к ней с должным вниманием, так как она имеет большое значение для науки и философии. Среди философских проблем, на которые указывает Чалмерс в связи с технологической сингулярностью, можно выделить следующие:

Необходимость переосмыслить явление разума как такового и ментальных возможностей искусственных машин;

Потенциальные последствия интеллектуального взрыва заставляют пересмотреть ценности, нравственность, сознание и персональную идентичность.

Практический вопрос можно обозначить так: «могут ли машины иметь интеллект или нет, и если да, то это хорошо или плохо?». Конечно, важнейшая проблема, которая появляется вместе с невероятным развитием машин: жизнь или смерть. Человечество, чьи ментальные способности уже преодолены, как дальше оно будет жить, выживать, существовать? Если для дальнейшего выживания необходимо улучшить тело, мозг, каков будет результат таких изменений? Кого считать сознающим себя человеком, останется ли личная идентичность? [7; стр. 4]

Чалмерс предпринимает попытку вывести обсуждение технологической сингулярности на академический уровень. Для этого он предлагает следовать строгому определению и даёт несколько подробных определений технологической

сингулярности через силлогизмы. У него их несколько, но на деле они лишь уточняют первый. Итак, следуя за Чалмерсом, пусть ИИ – это искусственный интеллект уровня человека или выше (который, по крайней мере, равен среднему человеческому интеллекту). Пусть ИИ+ будет искусственным интеллектом, превышающим человеческие возможности (которые являются возможностями наиболее умного человека). Пусть ИИ++ (или сверхинтеллект) – это искусственный интеллект значительно выше человеческого уровня (допустим, настолько же выше, насколько высочайший человеческий интеллект выше мышинового). Тогда можно представить следующее доказательство интеллектуального взрыва:

1. ИИ+ будет.
2. Если есть (будет) ИИ+, будет и ИИ++.
3. ИИ++ будет.

Всё доказательство строится на том, что в качестве истинного принимается предпосылка: есть такая вещь как интеллект, и один интеллект может быть выше (лучше, развитее) другого. Иначе ИИ+ и ИИ++ не будут иметь смысла. Доказательство, таким образом, будет невозможным [7, стр. 5].

Тема ИИ напрямую связана с проблемой определения и измерения интеллекта. Причём измерение интеллекта становится краеугольным камнем. Для того, чтобы мы вообще могли говорить об интеллектуальном взрыве как о предпосылке сингулярности, должны существовать показатель или группа показателей, которые создавали бы разницу между двумя интеллектами. Эта разница должна выражаться так, чтобы в конечном итоге можно было сделать примерно следующее заключение: интеллект  $x$  больше интеллекта  $y$  по показателю  $z$ . Принятие этой посылки позволит производить дальнейшие умозаключения в сфере ИИ. Только если возможна разница в интеллектах, возможно создание «ультраинтеллектуальной» машины, о которой писал И. Дж. Гуд. Только тогда возможны дискуссии о том, плохо это или хорошо.

Как утверждает Чалмерс, можно обойтись без такой мерки интеллекта. В качестве замены может выступить мера способностей к программированию или определённая способность мыслить логически [7, стр. 17].

Однако это невозможно. Чалмерс продолжает размышление дальше и повторяет тем самым других авторов, пишущих о технологической сингулярности. Здесь необходимо сделать остановку и детально рассмотреть данный вопрос.

Во-первых, можно заметить, что Виндж, Гуд, а за ними и Чалмерс делают одно и то же. Они обходят определение интеллекта. Оставляя на интуитивное понимание то, что следует понимать под словами «разум», «интеллект», они расставляют ловушку, похожую на софизм. Рассуждение, построенное на пропущенном шаге, следует для справедливости продолжать, мысленно ставя перед каждым последующим предложением фразу: «Если, мир выглядит так, как представлено здесь, а именно: интеллект – А, разум – В, то...».

Предполагается, что ещё предстоит выяснить, чем именно является интеллект, разум, и как работает мозг. Однако рассуждения о технологической сингулярности строятся так, будто уже имеется определённая модель понимания того, что это такое. В частности, такие понятия, как «загрузка данных», «извлечение памяти», «улучшение способности мыслить», «считать, решать различного рода задачи», могут быть основаны только на определённом ряде возможных моделей того, что

такое интеллект и как работает мозг. Таким образом, вопрос об интеллекте фактически снят подразумеваемым наличием имплицитной его модели. Выявление этой модели или моделей у различных авторов, пишущих о технологической сингулярности, требует специального исследования. В данном случае, интересны те особенности, которые можно обнаружить в уже упоминавшихся источниках.

Во-первых, важным качеством предполагаемого интеллекта становится его способность к увеличению. Предполагается, что суперинтеллектуальная машина мощнее в интеллектуальном плане, чем человек и компьютеры новейшего поколения. Во-вторых, увеличение, если и имеет ограничения, то весьма неопределённые. В-третьих, увеличение интеллекта есть улучшение всех связанных с ним способностей в примерно одинаковой пропорции. Так, некий интеллект  $X$  лучше интеллекта  $Y$  в отношении всех способностей, которые предположительно закреплены за интеллектом и не может быть такой ситуации, когда интеллект  $X$  лучше интеллекта  $Y$  в одной способности, но хуже или равен в другой.

Снимается вопрос о сущности интеллекта, мозга, но вместе с ними снимается и вопрос об источниках, из которых берётся «интуитивное» понимание, что это такое. Можно предположить, что используются некие общие знания биологии, эволюции, генетики и информатики, так как упоминается, что конкретные пути прихода к технологической сингулярности могут быть различными [7, стр. 12]. Однако экстраполяция общих сведений не очевидна и имеет не одну возможность понимания сущности рассматриваемого явления. Более того, следует учитывать, что представления об одном и том же объекте в науке могут меняться в связи с получением новых данных и проведением новых исследований. В исследованиях мозга точку пока не поставили. Об этом свидетельствует, например, не так давно произошедший пересмотр локализации психических функций в определённых частях мозга и новые исследования синаптической пластичности [12]. Если за основу интеллекта берётся мозг, (как некий физический субстрат интеллекта), то почему не рассматриваются более подробно другие имеющие значение функции мозга: эмоции, воля? В упомянутых уже вариантах рассмотрения технологической сингулярности вопрос эмоций также снимается. Имплицитно подразумевается, что будущий сверхинтеллект либо будет без эмоций, либо с эмоциями, но с такими, какие пожелает человек. Такое утверждение не является очевидным и следующим с необходимостью из любых данных о мозге [8].

**Выводы.** Рассуждения о технологической сингулярности не заканчиваются на простом указании причин, которые, возможно, к ней приведут. Хотя само понятие «сингулярность» подразумевает, что невозможно предсказать, что будет в постсингулярном мире, совершаются попытки обрисовать определённые варианты путей прихода к ней и её результатов. Однако, кроме допущения о самой технологической сингулярности, совершается ещё одно неявное допущение. Оно касается сознания, творческих способностей, того, что такое мозг и интеллект.

Эта пропущенная посылка является ключевым звеном в любом размышлении на тему технологической сингулярности, которое обычно опускается и не высказывается. Очевидно, без данной посылки рассуждение невозможно вовсе. Следует чётко обозначать понимание (даже если это совершается на уровне

допущения) того, что такое мозг, разум, интеллект. Построенное таким образом рассуждение будет более ясным и, возможно, более непротиворечивым.

### Список литературы

1. Гёте И. В. Фауст [Текст] / Иоганн Вольфганг Гёте. – М. : Правда, 1975. – 480 с.
2. Гомер. Илиада [Текст] / Гомер. – М.-Л. : Гослитиздат, 1949. – 551 с.
3. Нечаев. С. Ю. Китайская комната Дж. Р. Серля в контексте проблем философии [Текст] / С.Ю. Нечаев // Известия Саратовского университета, сер. Философия. Психология. Педагогика. – 2010. – Т. 10, вып. 4. – С. 19-23
4. Серль Дж. Р. Сознание, мозг и программы [Текст] / Дж.Р. Серль // Аналитическая философия: Становление и развитие: Антология / Общ. ред. и сост. А.Ф. Грязнов. М., 1998. – С. 376-401.
5. Buchanan B.G. A (Very) Brief History of Artificial Intelligence [Электронный ресурс] / Bruce G. Buchanan // Электрон. текстовые дан. – AI AI Magazine, 2005 – Vol. 26 Number 4. – pp.53-60 – Режим доступа: <http://aitopics.net/assets/PDF/AIMag26-04-016.pdf>
6. Bush, Vannevar. As We May Think [Электронный ресурс] / Vannevar Bush // Электрон. текстовые дан. – The Atlantic Monthly. – July 1945. – pp.112-124 – Режим доступа: <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
7. Chalmers, D.J. The Singularity: A Philosophical Analysis [Электронный ресурс] / David J. Chalmers // Электрон. текстовые дан. – Journal of Consciousness Studies. – 2010. – Vol. 17. – Режим доступа: <http://consc.net/papers/singularity.pdf>
8. Dalgleish T. The emotional brain / Dalgleish T. // Nat. Rev. Neurosci., 2004. - July, Vol. 5(7) – 583 p. – 9 p.
9. Good, I.J. Speculations concerning the first ultraintelligent machine [Электронный ресурс] / I.J. Good // F. Alt & M. Rubino, eds. – Электрон. текстовые дан. – Advances in Computers. – 1965. – Vol 6. – pp. 31-88 – Режим доступа: [http://www.stat.vt.edu/tech\\_reports/2005/GoodTechReport.pdf](http://www.stat.vt.edu/tech_reports/2005/GoodTechReport.pdf)
10. Kurzweil, R. The Singularity is Near [Электронный ресурс] : When Humans Transcend Biology / R. Kurzweil – Электрон. текстовые дан. – New-York, 2005. – Режим доступа: Singularity is Near.pdf
11. Legg Sh. Machine Super Intelligence [Электронный ресурс] / Shane Legg. –Электрон. текстовые дан. – 2008. – Режим доступа: [http://www.vetta.org/documents/Machine\\_Super\\_Intelligence.pdf](http://www.vetta.org/documents/Machine_Super_Intelligence.pdf)
12. Schwartz J. M., Begley Sh. The Mind and the Brain: Neuroplasticity and the Power of Mental Force / Schwartz J. M., Begley Sh. - New York: HarperCollins, 2002. – 432 p.
13. Solomonoff, R.J. The time scale of artificial intelligence: Reflections on social effects [Электронный ресурс] / Ray J. Solomonoff. // Электрон. текстовые дан. –North-Holland Human Systems Management 5, 1986. – pp.149-153. – Режим доступа: <http://world.std.com/~rjs/timesc.pdf>
14. The History of Artificial Intelligence [Электронный ресурс] / Chris Smith, Brian McGuire, Ting Huang, Gary Yang. – Электрон. текстовые дан. – University of Vashington, 2006. – 27p. – Режим доступа: <http://www.cs.washington.edu/education/courses/csep590/06au/projects/history-ai.pdf>
15. Turing, Alan. Computing Machinery and Intelligence [Электронный ресурс] / Alan M. Turing // Электрон. текстовые дан. – Mind. – 1950. – Vol. 49. – pp.433-460. – Режим доступа: <http://www.csee.umbc.edu/courses/471/papers/turing.pdf>
16. Vinge, V. First word [Текст] / V. Vinge // Omni. – January 1983. – p. 10.
17. Vinge, V. The coming technological singularity [Электронный ресурс] : How to survive in the post-human era / V. Vinge // Электрон. текстовые дан. – Whole Earth Review. – winter 1993.
18. Yudkowsky, E. Staring at the singularity [Электронный ресурс] / E. Yudkowsky – Электрон. текстовые дан. – 1996. – Режим доступа: <http://yudkowsky.net/obsolete/singularity.html>

**Степанов В.В.** *Неявна передумова міркувань щодо технологічної сингулярності* // Вчені записки Таврійського національного університету ім. В. І. Вернадського. Серія: Філософія. Культурологія. Політологія. Соціологія. – 2013. – Т. 26 (65). – № 4. – С. 64–70.

У даній статті розглядається явище технологічної сингулярності, історія вивчення даної проблеми і філософські перспективи її дослідження. Виявляється таке твердження про розум, що міститься імпліцитно, але не висловлюється, яке є наріжним каменем систем технологічної сингулярності.

**Ключові слова:** Девід Чалмерс, штучний інтелект, сингулярність

**Stepanov V. The missing premise in technological singularity reasoning** // Scientific Notes of Taurida National V.I. Vernadsky University. Series: Philosophy. Culturology. Political sciences. Sociology. – 2013. – Vol. 26 (65). – № 4. – P. 64–70.

The article examines the phenomenon of technological singularity, the history of the problem and some of its philosophical implications. It is found out that works in the field mainly describe postsingular world, a place of a person in it, problems with identity and problems of human-AI communication. Reasoning about technological singularity cannot be made without mentioning what is intellect, brain, etc. Thus authors simply omit the issue on the reason that science still doesn't have a proper and sufficient knowledge in the field. But whole postsingular systems are based on some implicit understanding of what intelligence is and how brain works. This premise is always avoided and there is no straight way to find out the sources of author's vision on such concepts as mind, brain, intelligence. Therefore it is crucial to make the always avoided premise clear, so that the whole reasoning would be as clear and comprehensive.

**Key words:** artificial intelligence, technological singularity.