

# НАУЧНЫЙ СОВЕТ РАН ПО МЕТОДОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (НСМИИ РАН)

---

119991, г. Москва, ул. Волхонка д. 14, Институт философии РАН URL: <http://iph.ras.ru/ai.htm>; e-mail: [secretary.scmair@philosophy.ru](mailto:secretary.scmair@philosophy.ru), тел.: (495) 697-9576 (Вт, Чт. 14.00-16.00).

---

«24» апреля 2013 г., г. Москва

---

**Уважаемые коллеги!**

Приглашаем Вас принять участие в работе НСМИИ РАН (59-е заседание Совета)

**24 апреля 2013 г., 17.00-17.30**

Сопредседатели Совета: академик С.Н. Васильев, проф. Д.И. Дубровский,  
академик В.А. Лекторский, академик В.Л. Макаров

Ученый секретарь: к.ф.н. Д.В. Иванов

## **Повестка дня**

1. Об итогах Первой Всероссийской конференции с международным участием «Глобальное будущее 2045: Антропологический кризис. Конвергентные технологии. Трансгуманистические проекты» (11 – 12 апреля 2013 г. Белгород)

***Д.И. ДУБРОВСКИЙ***

2. О пятой международной конференции по фундаментальным основам информационной науки FIS – 2013 (Москва. Московский гуманитарный университет, 21-24 мая 2013 г.)

***К.К. КОЛИН***

**60-е заседание междисциплинарного научно-теоретического семинара  
"Философско-методологические проблемы искусственного интеллекта".**

**24 апреля 2013 г., 17.30-20.00**

Обсуждение доклада

# «ПЕРСПЕКТИВЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ТЕОРЕМА ПЕНРОУЗА. ЧАСТЬ I. ОБЩАЯ КРИТИКА ТЕХНОЛОГИЧЕСКОЙ СИНГУЛЯРНОСТИ»

**Докладчик:** ПАНОВ Александр Дмитриевич, кандидат физико-математических наук, ведущий научный сотрудник Института ядерной физики МГУ

**Содокладчики-оппоненты:** ДУНИН-БАРКОВСКИЙ Виталий Львович, зав. отделом нейроинформатики Центра оптико-нейронных технологий Института системных исследований РАН, доктор физико-математических наук, профессор;

**РЕЖАБЕК Борис Георгиевич**, кандидат биологических наук, директор Института ноосферных исследований, зав. кафедрой нанобионики

---

**Место проведения семинара:** Институт философии РАН, 5-ый этаж, 524 ауд.  
(м. Кропоткинская, последний вагон из центра, перейти улицу Волхонка, вход через чугунные ворота во двор, на лифте подняться на пятый этаж).

## Тезисы доклада А. Д. ПАНОВА

### **Перспективы искусственного интеллекта и теорема Пенроуза. Часть I. Общая критика технологической сингулярности.**

Данный доклад представляет первую часть работы, которая состоит из двух взаимосвязанных, но относительно независимых частей. Как следует из названия, в первой части критически рассматриваются некоторые гипотезы, лежащие в основе представления о технологической сингулярности. Вторая часть должна быть посвящена, главным образом, критическому анализу следствий из теоремы Пенроуза об искусственном интеллекте (ИИ) и следующим отсюда представлениям о перспективах ИИ, но эту часть приходится рассматривать в отдельной лекции, так как весь материал невозможно охватить в одном выступлении, сохранив достаточно детальное изложение.

Под сильным ИИ понимается такой ИИ, который превосходит интеллект человека *во всех отношениях*. Понятие технологической сингулярности связано именно с таким сильным ИИ. Если предположить, что сильный ИИ когда-нибудь будет создан, то, очевидно, он не будет более нуждаться в участии человека, чтобы развивать сам себя. Более того, поскольку он будет не только качественно не слабее человеческого интеллекта, но, скорее всего, много быстрее, то люди будут просто неспособны уследить за автоэволюцией ИИ, поэтому машинная эволюция может полностью оторваться от человеческой и стать для людей совершенно непостижимой. Этот момент отрыва и называется технологической сингулярностью.

Одним из основных идеологов технологической сингулярности является Рэй Курцвейл, и критика технологической сингулярности в докладе строится с использованием его представлений в предположении, что они являются достаточно типичными и для других сторонников технологической сингулярности.

Основная идея Рэя Курцвейла состоит в том, что перспективы создания сильного ИИ определяются ростом мощности компьютеров, выраженной в операциях в секунду, или, в другой терминологии, удешевлением быстрогодействия компьютеров. Именно, как только быстроедействие коммерческих компьютеров превзойдет совокупное быстроедействие мозга всех людей вместе взятых, то сильный ИИ автоматически станет действительностью. На основе экстраполяции закона Мура Рэй Курцвейл дает прогноз, что это произойдет вблизи 2045 года.

Вывод Курцвейла основан на ряде допущений, некоторые из которых не являются очевидно верными, некоторые даже явно им не зафиксированы, а некоторые основаны на недоразумениях. В докладе рассматриваются следующие такие «слабые гипотезы»:

1. В оценке быстрогодействия мозга человека Курцвейл предполагает, что эта величина может быть вычислена просто как количество синаптических связей в мозге (порядка  $10^{14}$ ), помноженное на частоту переключения одного синапса (порядка 100 Гц). Это дает около  $10^{16}$  операций в секунду, что и рассматривается как базовая оценка быстрогодействия мозга. В этой оценке проигнорирован тот факт, что мозг демонстрирует два (как минимум) разных вида активности. Один вид активности называется быстрой активностью, это и есть нейронная синаптическая активность, учтенная в оценке Курцвейла. Другой вид активности называется медленной, и этот вид активности связан с модификацией структуры нейронной сети мозга – созданием одних синаптических связей, уничтожением других, но, возможно не только с этим. В оценке Курцвейла неявно предполагается, что медленная активность вообще не требует никаких ресурсов в смысле быстрогодействия, – нуль операций в секунду. Эта гипотеза, как минимум, требует обоснования и, в действительности, вовсе не выглядит правдоподобной. Имеются как косвенные, так и прямые свидетельства того, что существуют очень сложные внутринейронные процессы информационной природы, которые могут требовать огромных вычислительных ресурсов.

К косвенным свидетельствам относятся очень сложные формы поведения одноклеточных существ, вовсе лишенных нейронной системы, и, соответственно, быстрой нейронной активности. В таких организмах все управление происходит на субклеточном уровне, и по каналу консерватизма эволюции этот механизм управления должен быть унаследован нейронами, которые, по сути, являются не более чем сильно специализированными одноклеточными организмами. На основе этой идеи можно предсказать, что и отдельный изолированный нейрон способен демонстрировать сложные формы поведения, не имеющие отношения к его синаптической активности. И действительно, такое поведение наблюдается, например, в опытах Б. Г. Режабека с хвостовым нейроном речного рака (память изолированного нейрона и др.). Это прямое указание на существование сложных субнейронных информационных процессов.

Оценить дополнительную скорость обработки информации, связанную с такой субнейронной активностью, пока трудно. Однако обращает на себя внимание то, что в клетке потенциально существуют структуры, которые могут обеспечить очень высокую скорость переключений. Такой структурой, например, могут быть стенки микротрубочек цитоскелета, очень напоминающие клеточный автомат с двумя или даже большим количеством состояний ячеек. Такого рода «устройство» потенциально может обеспечить быстроедействие масштаба  $10^{15}$  операций в секунду на один нейрон, что на сто миллиардов нейронов мозга дает в общей сложности  $10^{26}$  операций в секунду. Тогда чисто синаптическая быстрая активность на этом фоне становится пренебрежимо малой величиной (одна десятиллиардная часть). И стенки микротрубочек – это далеко не единственное место в нейроне, где, в принципе, может вестись внутриклеточная

обработка информации. Все эти вопросы совершенно не ясны, они требуют детального изучения, которым должна заняться новая наука – «субнейрофизиология».

2. В оценках Курцвейла предполагается, что сравнение быстродействия компьютера и мозга вообще имеет смысл. Наши компьютеры в обычном понимании, независимо от их мощности, являются классическими (не квантовыми) устройствами, и сравнение мозга с компьютером по быстродействию утрачивает смысл в том случае, если на некотором глубоком уровне мозг ведет себя как квантовое вычислительное устройство. Аргументация Курцвейла против возможности существования квантовой обработки информации в мозге основана на том, что квантовые вычислительные устройства, создаваемые людьми, требуют сверхнизких температур и других экзотических условий, которых заведомо не может быть в мозге. Эта аргументация неверна, так как почти все искусственные квантовые вычислительные устройства пока основаны на управлении спинами, квантовые состояния которых очень хрупки и требуют для устойчивого существования сверхнизких температур, но природные квантовые компьютеры могут быть основаны на совсем другой «элементной базе». Например, на управлении конформациями макромолекул, на одноатомных «координатных» переключателях и т.д. Такая возможность обосновывается реальным существованием сложных макромолекул, устойчиво находящихся в спутанных квантовых состояниях ЭПР типа при комнатной, и даже при повышенной температуре, в жидкой или твердой фазе. Примерами таких объектов являются молекулы красителей. Так что квантовый характер обработки информации в мозге вовсе не исключен, и классический характер обработки информации, принимаемый Курцвейлом, вовсе не очевиден. Более того, теорема Пенроуза об ИИ (см. ниже) дает косвенное доказательство того, что квантовые процессы обработки информации в мозге действительно имеют место.

3. Рэй Курцвейл предполагает, что если мощность компьютерного «железа» превысит некоторый порог (совокупное быстродействие мозга всех людей), то сильный ИИ возникнет почти автоматически. Это далеко не очевидно, более того – скорее всего совсем не так. Практика показывает, что быстродействие и иные показатели мощности компьютеров не имеют прямого отношения к «интеллектуальности» компьютерных систем. Так, например, самая популярная издательская система Word по своей функциональности практически не изменилась за последние более чем 20 лет, несмотря на то, что мощности бытовых компьютеров возросли за это же время примерно в миллион раз. Никаких «интеллектуальных электронных секретарей» за это время не появилось, хотя вполне можно было бы ожидать. В докладе обсуждаются и другие примеры. Контраргументом к аргументации Курцвейла является также то, что до сих пор не удается удовлетворительно промоделировать нейронную систему нематоды *Caenorhabditis elegans*, состоящую всего из 302 нейронов, несмотря на предпринимаемые усилия. Совершенно очевидно, что источник проблем здесь кроется вовсе не в быстродействии компьютеров, которой, в соответствии с критериями Курцвейла, больше чем достаточно для решения этой задачи. Главная (но не единственная) проблема в том, что непонятны все тонкости работы этой простейшей нервной системы. Иными словами, трудность лежит в плоскости создания программного обеспечения для ИИ и в недостатке понимания, что именно нужно делать, но вовсе не только в быстродействии компьютеров. Но, как показывают наблюдения, программное обеспечение чрезвычайно консервативно, и вовсе не следует закону, подобному кривой Мура. Растущие мощности компьютеров в настоящее время почти на 100% тратятся на чисто техническое развитие пользовательского интерфейса и качество графики, на миниатюризацию, на средства коммуникации – что автоматически вовсе не ведет в сторону создания сильного ИИ.

4. Рэй Курцвейл полностью игнорирует аргументацию Рождера Пенроуза, показывающую, что устройство, способное моделировать все интеллектуальные способности человека, вообще не может быть создано на базе архитектуры конечного автомата – компьютера в современном понимании. Точнее говоря, Курцвейл путает

теорему Пенроуза об искусственном интеллекте с теоремой Геделя-Тьюринга о конечных автоматах, которая действительно ничего не говорит о сопоставлении возможностей человеческого мозга и компьютера. Этот круг вопросов в данном докладе упомянут, но подробно рассматриваться не будет, так как составляет содержание другого доклада (вторая часть данной работы)

В качестве выводов формулируются направления, на которые стоило бы обратить особое внимание в рамках программы «Аватар».