



ВЕРНОР

ВИНДЖ

СИНГУЛЯРНОСТЬ

*Книги, изменившие мир.
Писатели, объединившие
поколения.*

Э К С К Л Ю З И В Н А Я К Л А С С И К А

ЭКСКЛЮЗИВНАЯ КЛАССИКА (АСТ)

Вернор Виндж
Сингулярность

«Издательство АСТ»

1993, 2003, 2007

УДК 004.8
ББК 32.813

Виндж В. С.

Сингулярность / В. С. Виндж — «Издательство АСТ», 1993,
2003, 2007 — (Эксклюзивная классика (АСТ))

ISBN 978-5-17-114349-7

Создание интеллекта, превосходящего человеческий, произойдет в ближайшие тридцать лет. ...Это та самая точка, где наши прежние модели перестают работать, и в свои права вступает новая реальность. Как приближение Сингулярности повлияет на человеческое мировоззрение? И что случится в течение пары месяцев (или пары дней) после этого? В моем распоряжении есть только аналогия, на которую я могу указать: возникновение человечества. Мы окажемся в постчеловеческой эпохе... — это цитата из программной статьи Вернора Винджа «Грядущая технологическая сингулярность», одной из самых часто упоминаемых работ об искусственном интеллекте за последние 25 лет. В формате a4.pdf сохранен издательский макет книги.

УДК 004.8
ББК 32.813

ISBN 978-5-17-114349-7

© Виндж В. С., 1993, 2003, 2007
© Издательство АСТ, 1993, 2003, 2007

Содержание

Грядущая технологическая сингулярность. Как выжить в постчеловеческую эпоху	6
Резюме	7
Что такое Сингулярность?	8
Можно ли избежать Сингулярности?	11
Другие пути к Сингулярности: усиление интеллекта	14
Конец ознакомительного фрагмента.	15

Вернор Виндж Сингулярность

Vernor Vinge

The Coming Technological Singularity What If the Singularity Does Not Happen? The Cookie
Monster

© Vernor Vinge, 1993, 2003, 2007

© Перевод. М. Левин, 2019

© Перевод. В. Гришечкин, 2019

© Издание на русском языке AST Publishers, 2019

* * *

Грядущая технологическая сингулярность. Как выжить в постчеловеческую эпоху

Предлагаемая статья была написана для симпозиума VISION-21, спонсированного исследовательским центром НАСА Lewis Research Center и Аэрокосмическим институтом Огайо и проходившего 30–31 марта 1993 года. Ее можно также найти на сервере технических отчетов НАСА как часть документа NASA CP-10129. Слегка измененная версия была опубликована в зимнем выпуске 1993 года *Whole Earth Review*.

Резюме

В ближайшие тридцать лет у нас появятся технические средства для создания сверхчеловеческого интеллекта. Вскоре после этого эра человека закончится.

Можно ли избежать такого развития событий? И если нет, то можно ли направить эти события таким образом, чтобы у нас была возможность выжить? Этим вопросам посвящена данная статья, в которой представлены некоторые возможные ответы (и указаны некоторые дальнейшие угрозы).

Что такое Сингулярность?

Основной характеристикой текущего столетия было и остается ускорение технического прогресса. Я в данной работе утверждаю, что мы стоим на грани перемены, сравнимой с возникновением на Земле человека. Конкретная причина этой перемены – неизбежное создание с помощью техники сущностей, чей интеллект превзойдет человеческий. Средств, которыми наука может добиться этого прорыва, существует несколько (и это усиливает уверенность в его неизбежности):

- Развитие компьютеров, «проснувшихся» и сверхчеловечески умных. (До сих пор основные споры про ИИ вертелись вокруг вопроса, сможем ли мы создать эквивалент человека в виде машины. Если ответ будет положительным, то вне, всяких сомнений, вскоре после этого можно будет создать и более разумные существа.)

- Сотрудничество человека с компьютером может стать столь тесным, что пользователей вполне реально будет рассматривать как обладателей сверхчеловеческого интеллекта.

- Большие компьютерные сети (вместе с пользователями) могут «осознать себя» как сущности, обладающие сверхчеловеческим разумом.

- Биология может дать средства развития и совершенствования природного человеческого интеллекта.

Первые три возможности во многом зависят от развития аппаратного обеспечения компьютеров. График прогресса в этой отрасли в последние десятилетия был на удивление стабилен [16].

Основываясь на этой тенденции, я считаю, что создание интеллекта, превосходящего человеческий, произойдет в ближайшие тридцать лет. (Чарльз Платт [19] указывает, что энтузиасты ИИ говорят то же самое последние тридцать лет. Чтобы не прятаться за неоднозначностью относительного времени, скажу конкретнее: меня удивит, если это событие произойдет до 2005 или после 2030 года.)

Каковы следствия этого события? Прогресс, подхлестнутый интеллектом сильнее человеческого, пойдет намного быстрее. Действительно, мы не видим никаких причин, чтобы сам по себе прогресс не подразумевал возможности создания еще более интеллектуальных сущностей – и в еще более короткое время. Наилучшую аналогию этому я вижу в прошлой эволюции: животные умеют приспосабливаться к трудностям и «изобретать» способы их преодоления, но не быстрее, чем делает свою работу естественный отбор: в случае естественного отбора мир действует как симулятор самого себя. Мы, люди, обладаем возможностью строить модель мира у себя в голове и на ней прокручивать всяческие «что, если»; многие проблемы мы можем решать в тысячи раз быстрее естественного отбора. Теперь, создавая средства, еще сильнее ускоряющие этот процесс, мы входим в режим, который так же радикально отличается от нашего человеческого прошлого, как отличаемся мы сами от низших животных.

Человеческому взгляду эта перемена представится как выбрасывание на свалку всех прежних правил (вероятно, произойдет это в мгновение ока), как экспоненциальное разбегание из-под контроля без малейшей надежды его остановить. События, которые, как считалось ранее, могли случиться «где-то через миллион лет» (если вообще могли), с большой вероятностью произойдут в следующем веке. (Грег Бир в [4] рисует картину коренных перемен, происходящих в считанные часы.)

Я думаю, правильно будет назвать эту перемену сингулярностью (в данной же статье – Сингулярностью, с большой буквы). Это та самая точка, где наши прежние модели перестают работать и в свои права вступает новая реальность. Чем ближе мы подбираемся к этой точке, тем сильнее нависает эта угроза над всей человеческой жизнью, и упоминание о ней превращается уже в общее место. Но когда она все же осуществится, это может оказаться огромной

неожиданностью – и еще большей неизвестностью. Некоторые (очень немногие) видели это еще в 50-х годах XX века. Стэн Улам [27] так перефразировал Джона фон Неймана:

В центре нашего разговора были ускорение технологического прогресса и перемены в образе жизни людей, свидетельствующие о приближении существенной сингулярности в истории рода человеческого, такой сингулярности, после которой дела людские в том виде, в котором они нам известны, продолжаться уже не смогут.

Видите, фон Нейман даже использует термин «сингулярность», хотя, мне кажется, он все же думает об обычном прогрессе, а не о создании сверхчеловеческого интеллекта. (Для меня сутью Сингулярности является именно сверхчеловеческая ее природа. Без нее мы просто получим изобилие технической роскоши, толком так и не усвоенной (см. [24])).

В 1960-х годах уже были осознаны некоторые следствия появления сверхчеловеческого интеллекта. И. Дж. Гуд писал [10]:

Назовем ультраинтеллектуальной машину, далеко превосходящую в интеллектуальной деятельности любого человека, как бы умен он ни был. Поскольку проектирование машин также является такой деятельностью, то ультраинтеллектуальная машина сможет проектировать машины еще лучшие, что, без сомнения, станет «интеллектуальным взрывом», который оставит интеллект человека далеко позади. Таким образом, первая ультраинтеллектуальная машина будет последним изобретением, которое придется сделать человеку, – при условии, что эта машина будет достаточно любезна, чтобы рассказать нам, как удерживать ее под контролем. ...И скорее всего, в двадцатом столетии такая ультраинтеллектуальная машина будет построена и окажется последним изобретением, которое придется сделать человеку.

Гуд понял суть процесса, но не стал разрабатывать его наиболее тревожные последствия. Никакая интеллектуальная машина того сорта, что он описывает, не станет «орудием» человечества – точно так, как сами люди не являются орудиями кроликов, птиц или шимпанзе.

В шестидесятых-семидесятых-восьмидесятых предвидение этого катаклизма стало более распространенным [28], [1], [30], [4]. Вероятно, первыми поняли его конкретные проявления авторы научной фантастики. В конце концов, именно авторы «твердой» НФ пытаются писать произведения о том, что конкретно может с нами сделать технология. Все чаще эти авторы натываются на непрозрачную стену, отделяющую от нас будущее. Когда-то они могли относить подобные фантазии на миллионы лет вперед [23]. Сегодня же они увидели, что их самые тщательные экстраполяции обещают непознаваемое уже в ближайшем будущем. Когда-то постчеловеческую эпоху казалось правильным относить к временам галактических империй. Сейчас, к сожалению, ее можно отнести и к временам межпланетных.

Что можно сказать о девяностых, нулевых, десятых с точки зрения приближения к этой границе? Как приближение Сингулярности повлияет на человеческое мировоззрение?

Какое-то время вполне уважаемой точкой зрения будет скептицизм по отношению к самой возможности существования «машины сапиенс». В конце концов, глупо ведь думать, что мы сможем создать интеллект, эквивалентный человеческому (или даже превосходящий его), пока у нас не будет аппаратуры такой же мощной, как человеческий мозг. (Существует умозримая возможность того, что можно создать эквивалент человека на менее мощной аппаратной базе, если поступиться скоростью, удовлетвориться искусственным существом, тормозным в буквальном смысле слова [29]. Но почти наверняка разработка нужного программного обеспечения окажется непростым процессом, с множеством фальстартов, проб и ошибок. Если так,

то появление машин с самосознанием не произойдет до тех пор, пока не появится аппаратная база, существенно более мощная, чем естественная человеческая.)

Но с течением времени нельзя будет не обратить внимания на новые симптомы. Дилемма, ощущаемая авторами научной фантастики, станет существенной в других творческих работах. (Я слышал, что авторы комиксов беспокоятся о том, как создавать зрелищные эффекты, когда все видимое может быть воспроизведено обычными техническими средствами.) Мы увидим автоматизацию все более и более сложных работ и рабочих мест. Даже сейчас у нас есть инструменты (математические программы, автоматизация проектирования и производства), освобождающие нас почти от всей низкоуровневой рутины). Формулируя иначе: по-настоящему продуктивная работа становится сферой занятий все меньшей и все более элитной части человечества. В наступающей Сингулярности мы увидим, как наконец осуществляются предсказания истинной технологической безработицы.

Другой симптом приближения к Сингулярности: сами по себе идеи станут распространяться все быстрее, и даже самые радикальные из них быстро будут становиться трюизмами. Когда я писал свои первые книги, очень просто было предложить идею, которой для встраивания в культурное сознание понадобятся десятки лет. Сейчас время внедрения идеи – где-то полтора года. (Конечно, может быть, дело в том, что я старею и теряю воображение, но я вижу тот же эффект и у других.) Сингулярность – как пробой звукового барьера: она тем ближе, чем ближе подбираемся мы к критической скорости.

А что можно сказать о наступлении самой Сингулярности? Что можно сказать о ее фактическом явлении? Поскольку она включает в себя взрыв интеллекта, то произойдет она, вероятно, быстрее, чем любая предыдущая техническая революция. Событие, которое вызовет лавину, почти наверняка будет неожиданным – возможно, даже для участвующих в процессе исследователей. («Но ведь все предыдущие модели были дико тормозными, мы лишь слегка подкрутили пару параметров...») Если к тому времени достаточно распространятся сети (став вездесущими встроенными системами), наблюдателю может показаться, что все созданные людьми предметы внезапно проснулись.

А что случится в течение пары месяцев (или пары дней) после этого? В моем распоряжении есть только аналогия, на которую я могу указать: возникновение человечества. Мы окажемся в постчеловеческой эпохе. И при всем моем безудержном технологическом оптимизме иногда я думаю, что мне спокойней было бы наблюдать эти переходные события с расстояния в тысячу лет... а не в двадцать.

Можно ли избежать Сингулярности?

Ну, может быть, ее вообще не будет. Иногда я пытаюсь представить себе симптомы, свидетельствующие, что Сингулярности не суждено возникнуть. Это широко признаваемые аргументы Пенроуза [18] и Серла [21] об отсутствии практического смысла в существовании машинного разума. В августе 1992 года корпорация Thinking Machines Corporation провела семинар по вопросу «Как мы будем строить машину, которая думает» (How We Will Build a Machine that Thinks). Как можно догадаться по названию семинара, участники не слишком поддерживали аргументы против машинного интеллекта. Общим мнением было то, что могут существовать разумы на небиологической основе и что для существования разумов основную роль играют алгоритмы. Однако шли серьезные споры насчет чисто аппаратной мощности, представленной в органических мозгах. Меньшинство считало, что крупнейшие компьютеры 1992 года отстают по мощности от человеческого мозга на три порядка. Большинство же участников соглашались с оценкой Моравеца [16], что до достижения аппаратного паритета нам остается где-то от десяти до сорока лет. Однако было еще одно меньшинство, указывавшее на [6], [20] и предполагавшее, что вычислительная мощность отдельных нейронов может быть существенно выше, чем это принято считать. Если так, то аппаратное обеспечение наших современных компьютеров может даже на *десять* порядков отставать от аппаратуры, которую мы носим под собственным черепом. В этом случае (или, кстати, если критика Пенроуза и Серла обоснована) Сингулярности нам просто никогда не видать. Более того, в начале нулевых мы увидим, что кривые роста производительности аппаратуры начинают постепенно снижаться – это будет вызвано нашей неспособностью к автоматизации сложного проектирования, необходимого для поддержания крутизны этих кривых. В конце концов у нас в руках окажется *очень* мощное аппаратное обеспечение, но усилить его мощность уже не получится. Коммерческая скорость обработки сигналов будет впечатляющей, придающей аналоговый вид даже цифровым операциям, но ничего нигде не «проснется» и не случится взрывного роста интеллекта, который и является сущностью Сингулярности. Это, вероятно, будет рассматриваться как золотой век... но будет также и концом прогресса. Очень похоже на будущее, предсказанное Гюнтером Стентом. Действительно, на странице 137 работы [24] Стент ясно говорит, что достаточным условием для опровержения его построений является развитие трансчеловеческого интеллекта.

Но если технологическая Сингулярность *может* случиться, то она случится. Даже если все правительства мира осознают эту «угрозу» и будут от нее в смертельном ужасе, продвижение к этой цели будет продолжаться. В художественной литературе есть рассказы о законах, принятых для запрета строительства «машины в виде ума человека» [12]. Но конкурентное преимущество (экономическое, военное, даже эстетическое) любого прогресса в автоматизации – настолько императивный фактор, что любые законы или обычаи, запрещающие подобные вещи, гарантируют лишь одно: это преимущество получит кто-то другой.

Эрик Дрекслер [7] изложил красочное видение того, насколько далеко могут завести технические улучшения. Он согласен, что сверхчеловеческие умы появятся в ближайшем будущем – и что такие сущности представляют собой угрозу теперешнему статус кво человечества. Но Дрекслер утверждает, что мы можем физически ограничить подобные сверхчеловеческие устройства или же ввести ограничивающие их правила, чтобы результаты их работы можно было изучать и безопасно применять. Получается все та же ультраинтеллектуальная машина И. Дж. Гуда, только с некоторыми предохранителями. Я возражаю, что такое ограничение внутренне противоречиво. В случае физического ограничения: представьте себе, что вы заперты в собственном доме и доступ данных извне идет только через ваших хозяев. Если эти хозяева думают медленнее вас, скажем, в миллион раз, то практически наверняка

за несколько лет (вашего времени) вы подадите им «полезный совет», который совершенно случайно даст вам свободу. (Эту быстромыслящую форму суперинтеллекта я называю «слабо сверхчеловеческой».) Такое «слабо сверхчеловеческое» существо, вероятно, вырвется на свободу за несколько недель внешнего времени. «Сильно сверхчеловеческий» суперинтеллект будет намного больше, чем просто человеческим разумом, только с быстрыми часами. Трудно сказать, на что будет этот «сильно сверхчеловеческий» суперинтеллект похож, но различие должно быть более глубоким. Представим себе собачий разум, работающий на очень высокой скорости. За тысячу лет додумается эта псина до какого-нибудь из человеческих открытий? (Ну, если собачий разум разумно перепаять и только *потом* запустить на высокой скорости, может, мы что-то иное и увидим...) В основном рассуждения о суперинтеллекте строятся, мне кажется, на модели слабо сверхчеловеческого. Я думаю, что более правильные представления о мире пост-Сингулярности могут быть выведены из размышлений о природе «сильно сверхчеловеческого». К этому я еще вернусь ниже.

Другой подход к Дрекслеровым ограничениям – встроить в создаваемый сверхчеловеческий разум *правила* (законы Азимова). Я считаю, что правила работы, достаточно строгие для гарантии безопасности, заодно снизят возможности устройства по сравнению с его свободными версиями (а потому людская конкуренция будет склонять к разработке этих более опасных моделей). Все же мечта Азимова удивительна: представить себе полного энтузиазма раба, который в тысячу раз превосходит тебя по способностям в любой области. Представить себе создание, которое может удовлетворить любое твоё безопасное (что бы это ни значило) желание, и при этом у него 99,9 % времени свободно для иной деятельности. Это была бы новая вселенная, недоступная нашему пониманию, но населенная благосклонными богами (хотя у меня наверняка возникло бы желание стать одним из них).

Если Сингулярность невозможно ни предотвратить, ни ограничить, то насколько неприятной может оказаться постчеловеческая эпоха? Ну... весьма. Одна из возможностей – физическое истребление человеческого рода. (Или, как это высказал о нанотехнологии Эрик Дрекслер: если учесть, на что эта технология способна, правительства наверняка придут к выводу, что граждане им больше не нужны.) Но физическое уничтожение – это еще, возможно, не самое страшное. Опять же аналогия: вспомним, как мы в различных смыслах связаны с животными. Некоторые грубые физические воздействия маловероятны, и все же... в постчеловеческом мире будет еще полно ниш, где потребуются автоматика, эквивалентная человеку: встроенные системы в автономных устройствах, осознающие себя роботы в низшей функциональности более крупных разумных образований. (Сильно сверхчеловеческий интеллект будет, вероятно, неким Сообществом Разума [15] с весьма разумными компонентами). Некоторые из этих человеческих эквивалентов могут быть использованы всего лишь для цифровой обработки сигналов, они будут более похожи на китов, нежели на людей. Другие могут быть вполне человекообразными, но односторонними, *заикленными* на одной теме в такой степени, что в наше время они оказались бы в психбольнице. Хотя может быть, что эти создания не будут людьми из плоти и крови, они будут в новой среде ближайшим аналогом того, что мы называем сейчас человеком. (И. Дж. Гуду было что сказать по этому поводу, хотя сейчас совет этот может иметь лишь теоретическое значение. Гуд в [11] предложил «золотое метаправило», которое можно перефразировать так: «Обращаясь с низшими так, как ты хотел бы, чтобы с тобой обращались высшие». Это чудесная парадоксальная идея (и большинство моих друзей в нее не верят), поскольку очень трудно сформулировать, что мы на этом выигрываем. Но если бы мы могли следовать этому правилу, то это что-то говорило бы о вероятности такого доброго отношения со стороны нашей вселенной.)

Я выше утверждал, что мы не можем предотвратить Сингулярность, что ее пришествие есть неизбежное следствие природной человеческой конкуренции и неотъемлемых возможностей техники. И все же инициаторы ее – мы. Даже самая огромная лавина возникает из-за

маленького камешка. У нас есть свобода определения начальных условий, приводящих к тому, что события будут не столь враждебны, как в ином случае. Конечно (поскольку мы запускаем лавину), можно сомневаться в том, каким именно должен быть этот ведущий толчок.

Другие пути к Сингулярности: усиление интеллекта

Когда говорят о создании сверхчеловеческого разумного существа, обычно имеют в виду проект ИИ. Но, как я заметил в начале этой статьи, есть и другие пути к сверхчеловеческому. Компьютерные сети и взаимодействие человека с компьютером кажутся более обыденными, чем ИИ, но именно они могут повести к Сингулярности. Я называю этот контрастирующий подход Усилением Интеллекта (УИ). Это процесс, происходящий весьма естественно, и даже его суть в большинстве случаев остается незамеченной разработчиками. Но каждый раз, когда улучшается наша способность получать информацию и делиться ею с другими, мы в некотором смысле усиливаем наш естественный интеллект. Даже сейчас группа людей с докторской степенью и хорошими компьютерами (пусть и не подключенными к сети!) могла бы, вероятно, заработать максимальное число баллов в любом существующем письменном тесте на интеллект.

И весьма вероятно, что УИ – куда более легкая дорога к достижению сверхчеловеческого, нежели ИИ. С людьми самая трудная проблема развития уже решена. Строить из готового материала (из нас самих) должно быть проще, чем сперва сообразить, что же мы собой представляем, а потом строить машины, которые будут представлять собой то же самое. И для такого подхода есть менее гипотетический прецедент. Кэрнс-Смит [5] предполагал, что биологическая жизнь могла начаться как надстройка над еще более примитивной жизнью, основанной на росте кристаллов. Линн Маргулис [14] приводит убедительные соображения в поддержку точки зрения, что мутуализм является великой движущей силой эволюции.

Заметим, что я не предлагаю игнорировать исследования в области ИИ или уменьшить ассигнования на них. Я предлагаю признать, что в изучении сетей и взаимодействия машины с человеком есть нечто столь же глубокое (и потенциально непредсказуемое), как и в исследованиях ИИ. В этом свете можно рассматривать проекты, которые не имеют столь прикладного характера, как обычные работы проектирования интерфейса и сетей, но продвигают нас к Сингулярности по пути УИ.

Вот некоторые возможные проекты, приобретающие особую важность с точки зрения УИ:

- Автоматизация процессов с помощью команды человек/компьютер: для проблем, которые обычно решаются чисто машинными методами (вроде поиска экстремума), разрабатываются программы и интерфейсы, использующие преимущества человеческой интуиции и доступной компьютерной мощности. Учитывая всю причудливость многомерной проблемы поиска экстремума (и тонкие алгоритмы, выработанные для ее решения), можно ожидать весьма интересных средств отображения и контроля, предоставляемых человеческой части команды.

- Создание симбиоза человек/компьютер в искусстве. Сочетание графических возможностей современных машин и эстетической чувствительности человека. Конечно, для разработки компьютерного помощника художнику как инструмента, экономящего труд, требуется огромный объем исследований. Я предлагаю явно поставить себе цель большего слияния умений человека и компьютера, открыто признать возможным такое сотрудничество. В этом направлении чудесную работу сделал Карл Симс [22].

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.