ВЕРНОР ВИНДЖ

СИНГУЛЯРНОСТЬ

ИЗДАТЕЛЬСТВО АСТ
МОСКВА

УДК 004.8
ББК 32.813
В48

Серия «Эксклюзивная классика»

Vernor Vinge

THE COMING TECHNOLOGICAL SINGULARITY
WHAT IF THE SINGULARITY DOES NOT HAPPEN?
THE COOKIE MONSTER

Перевод с английского *М. Левина, В. Гришечкина*

Серийное оформление *А. Фереза, Е. Ферез*

Компьютерный дизайн *А. Чаругиной*

Печатается с разрешения литературных агентств The Lotts Agency и Andrew Nurnberg.

Виндж, Вернор.

В48 Сингулярность [сборник] / Вернор Виндж ; [пер. с англ. М. Левина, В. Гришечкина]. — Москва : Издательство АСТ, 2022. — 224 с. — (Эксклюзивная классика).

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.

УДК 004.8
ББК 32.813

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 удержать 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 натыкались 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 умозрительная 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 аппарату-