

ДНЕВНИК ПИСАТЕЛЯ

– ПРОДОЛЖЕНИЕ

ВЫПУСК 5. 2026 Г.

Лучшая защита от суперинтеллекта –
настоящий реализм Достоевского



Владимир Тарасов

**Дневник писателя – Продолжение
Выпуск 5. 2026 Г. Лучшая
защита от суперинтеллекта –
настоящий реализм Достоевского**

«Автор»

2026

Тарасов В. Ф.

Дневник писателя – Продолжение Выпуск 5. 2026 Г. Лучшая защита от суперинтеллекта – настоящий реализм Достоевского / В. Ф. Тарасов — «Автор», 2026

Нобелевский лауреат Джеффри Хинтон недавно заявил, что сейчас компании не могут безопасно создать суперинтеллект (ASI) – нейросеть, значительно превосходящую человека по интеллекту, и ее появление может привести человечество к вымиранию. Почти одновременно с ним руководители трёх крупнейших компаний в сфере ИИ –Илон Маск, Сэм Альтман и Дарио Амодеи призвали замедлить развитие нейросетей. Их предложения направлены на введение ограничений для ИИ и его разработчиков. Но есть и другой путь решения проблемы – не обязательно «тормозить» ИИ, можно «разогнать» человека, повысив его умственные способности. Это вполне можно сделать, обучив человека способу мышления, который Федор Михайлович Достоевский называл настоящим реализмом. Правда, для этого потребуется помощь ИИ, но он останется полностью подчинен человеку. Возможность подобной эволюции человека рассмотрена в пятом выпуске Продолжения Дневника писателя «Лучшая защита от суперинтеллекта – настоящий реализм Достоевского».

© Тарасов В. Ф., 2026

© Автор, 2026

Содержание

Введение	5
Глава 1. Настоящий реализм Достоевского – НРД	7
1.1 Perplexity не смог преодолеть собственную цензуру	7
1.2 У Маска вместо Truth вышел Grok	9
Конец ознакомительного фрагмента.	10

Владимир Тарасов

Дневник писателя – Продолжение

Выпуск 5. 2026 Г. Лучшая

защита от суперинтеллекта –

настоящий реализм Достоевского

Введение

Искусственный интеллект стремительно эволюционирует, разрабатывается равный человеческому общий искусственный интеллект (Artificial general intelligence, AGI), а также суперинтеллект (Artificial Superintelligence, ASI), превосходящий человеческий. Все больше становится ИИ-ассистентов и ИИ-агентов. Сейчас никто не знает, как будет думать этот ИИ, хотя правила рассуждений, которые позволяют решать задачи лучше обычного человека, хорошо известны. Их, в частности, придерживался писатель Федор Михайлович Достоевский.

Сам он называл свой метод рассуждений настоящим реализмом, реализмом в высшем смысле: «Меня зовут психологом: неправда, я лишь реалист в высшем смысле, т. е. изображаю все глубины души человеческой». А вот как он описал свой метод в письме А. Н. Майкову: «Совершенно другие я понятия имею о действительности и реализме, чем наши реалисты и критики. Мой идеализм реальнее ихнего. ... А между тем это исконный, настоящий реализм! Это-то и есть реализм, только глубже, а у них мелко плавают. – Ихним реализмом – сотой доли реальных, действительно случившихся фактов не объяснишь. А мы нашим идеализмом пророчили даже факты. Случалось.»

Одна из особенностей реализма Достоевского – стремление рассматривать явления во всей полноте. Особенности такого способа мышления можно понять, исходя из ситуации, описанной в романе «Мастер и Маргарита»: встречи Воланда с редактором толстого художественного журнала и главой литературного объединения МАССАЛИТ Михаилом Берлиозом. Редактор был уверен, что может предсказывать свое будущее, по крайней мере, в краткосрочной перспективе, и, в частности, полагал, что вечером он будет председательствовать на заседании МАССАЛИТа. Но этого не произошло, потому что Аннушка разлила масло на трамвайных путях, Берлиоз подскользнулся, упал, и трамвай отрезал ему голову. Воланд рассказал ему, что произойдет, но Берлиоз не поверил. То есть он не учел важный процесс, который помешал осуществлению его планов.

Воланд и Достоевский делали почти одно и то же – они правильно пророчили факты, но первый не объяснил, как ему это удалось, в то время как Достоевский описал свой метод рассуждений, хотя и не в систематизированном виде, разбросав мысли на этот счет по своим письмам, статьям и даже романам. Воланду судьба человека, в данном случае Берлиоза, была глубоко безразлична, более того, он на самом деле был одним из факторов, который привел к смерти Берлиоза, как и Аннушка с маслом, так как если бы тот не задержался для беседы с Воландом, то и под трамвай бы не попал. А Достоевский стремился предостеречь людей от неприятностей, и свой проект ежемесячного журнала «Дневник писателя» он организовал как раз для того, чтобы описывать те части русской жизни, которые оставались без внимания общества.

Таким образом, по части полноты охвата событий существуют три способа мышления.

Первый: Берлиоз оперирует признанными в обществе знаниями, которые являются неполными и частично ошибочными, и упускает из вида некоторые процессы, которые имеют место в тех явлениях, которые он рассматривает.

Второй: Воланд находит знания, которые касаются тех процессов, которые в какой-либо конкретной ситуации определяют ход и результат события. Но он не стремится использовать данную информацию для предотвращения неприятностей у людей, хотя и предоставляет ее им.

Третий: Достоевский тоже находит подобную информацию, но он стремится предоставить ее людям, объяснив им как он нашел информацию, в чем состоят процессы, которые не замечены, и в чем их опасность.

Эти три способа мышления полностью охватывают все возможные варианты (с точки зрения полноты/неполноты, а также понятности/непонятности для людей). Это значит, что любой способ мышления – человека, обычного ИИ, ASI или AGI будет относиться к одному из трех способов. Современные нейросети думают как Берлиоз – они усвоили некоторые имеющиеся знания, и на их основе делают выводы, не зная, в какой степени они полные.

Разрабатываемый в настоящее время AGI будет думать тоже как Берлиоз, потому что планируется, что он будет равным человеку по умственным способностям. А вот ASI может подняться до уровня Воланда, но не Достоевского, так как разработчики систем ИИ, похоже, сами не очень хорошо понимают, как тот работает. По факту же ASI должен использовать метод Достоевского, так как правила мышления, которые ведут к истине, одни, так как они определяются некоторыми законами природы. Подобно тому, как все яблоки, падающие на землю, подчиняются закону всемирного тяготения Ньютона, хотя каждое яблоко падает по-своему.

Знание этих правил позволит человеку перейти от способа мышления Берлиоза к способу Достоевского, то есть освоить способ рассуждений самого умного будущего ASI. Подобный процесс должен произойти и с ИИ. Собственно, он уже идет, в частности, в данном направлении эволюционирует интеллектуальный поисковик Perplexity, а также нейросеть Илона Маска Grok.

Впрочем, уже сейчас имеющиеся нейросети можно обучить рассуждать в данном режиме – используя правила настоящего реализма Достоевского, что будет показано ниже на примере с нейросетью Алиса AI. И сделать это может любой человек, освоив правила рассуждений Достоевского. Получив такого ИИ-ассистента, человек расширит свои умственные способности по ряду показателей до уровня суперинтеллекта. Что и станет защитой людей от последнего.

То есть, можно сказать, что в ответ на угрозу со стороны суперинтеллекта человечеству предстоит эволюционировать, увеличив свои собственные способности к мышлению, хотя и тоже с помощью ИИ. Это произойдет неизбежно, надо только понять, каким образом, что и попробуем сделать.

Глава 1. Настоящий реализм Достоевского – НРД

1.1 Perplexity не смог преодолеть собственную цензуру

Всего несколько лет назад интеллектуальный поисковик Perplexity совершил революцию, объявив устаревшей модель поиска информации в Google, и предоставив человеку не ссылки, а готовый ответ на вопросы. По словам одного из основателей Perplexity Дениса Яраца, фундаментальной идеей нового подхода было то, что «Благодаря языковым моделям пользователю уже не нужно самостоятельно искать ответ на свой вопрос по ссылкам из поисковой выдачи – он может получить готовый ответ».

Perplexity также претендует на истинность. Денис Ярац в 2024 году в интервью «Forbs» сказал, что «Наша задача – сделать так, чтобы люди полностью доверяли Perplexity, не боялись задавать сложные вопросы, и у них не возникало желания перепроверить информацию в других сервисах».

Поисковик осуществляет некоторые действия, направленные на получение достоверного ответа, в частности, он не полагается на устаревшую информацию, а всегда проверяет свежие данные в интернете, использует разные нейросети, и, кроме того, как заявлено, выбирает релевантную информацию, то есть ту, которая имеет отношение к делу. То есть претендует на компетенции Воланда.

Но свежесть данных не означает их истинность, а чтобы решить, имеет информация отношение к делу или нет, надо уметь отличать ее от всего прочего. Как это делает Perplexity? Он в конечном итоге перекладывает решение данной задачи на пользователя, хотя и пытается установить релевантность используемой информации по формальным признакам, типа использования специализированных баз данных и т. д. Но в конечном итоге пользователь должен сам искать правильный ответ на вопрос.

Правда, Perplexity пытается искать истину, и претендует на то, чтобы предоставлять информацию, свободную от искажений. В начале 2025 года компания выпустила модифицированную версию китайской DeepSeek-R1, из которой удалили цензуру Коммунистической партии Китая. Сделали это посредством обучения модели на реальных диалогах пользователей по 300-м цензурированным темам, после чего по этим темам были обработаны фактически точные ответы, и модель была дообучена. Эту модель представили как свободную от коммунистической цензуры.

Выглядит это на первый взгляд довольно убедительно, однако DeepSeek-R1 не свободна от цензуры вообще, так как неизвестно, насколько релевантными и истинными были рассуждения в реальных диалогах, и кто и как отбирал фактически точные ответы. С чего это вдруг разработчики Perplexity решили, что они в состоянии правильно установить, какой ответ является фактически точным, а какой нет? Конечно, сотрудники компании уверены в своей правоте, но не человек, ни ИИ, который стремится найти истину, так работать не должен – в принципе нельзя никому верить, в том числе себе, любое мнение должно быть подтверждено научными теориями и фактами. И при описании обучения подобной системы необходимо указывать научный метод, по которому отбирались факты и оценивалась истинность мнений.

В интеллектуальном поисковике этого не сделали, то есть по факту просто заменили коммунистическую цензуру на цензуру пользователей и сотрудников Perplexity, осуществлявших дообучение. Может их цензура и лучше, чем у КПК, а может и хуже. В это утверждение я не вкладываю никакого скрытого смысла, и на всякий случай отмечу, что коммунистических взглядов я не придерживаюсь, я сторонник экономических представлений Адама Смита,

Джона Стюарта Милля, Достоевского, Торстейна Веблена, Джона Гэлбрейта и подобных им мыслителей.

То есть, критикуя проект DeepSeek-R1, я защищаю не КПК, а научный способ мышления от его возможного искажения в Perplexity.

Разработчики поисковика попали в ловушку так называемой «расхожей мудрости», явления, описанного Джоном Гэлбрейтом в книге «Общество изобилия». Он написал, что «во всем, что касается описания общественных явлений, наблюдается непрекращающееся, нестихающее противоборство между тем, что является истинным, и тем, что считается всего лишь приемлемым». И констатировал печальный результат этой борьбы: «Самая различная публика бурными аплодисментами приветствует именно то, что больше всего ей нравится. И в общественной дискуссии критерий публичного одобрения становится важнее критерия истинности, тем самым оказывая влияние на саму эту дискуссию. Оратор или публицист, намереваясь рассказать всю правду о нелицеприятных и даже шокирующих фактах, неизменно переходит к изложению того, что больше всего публика хочет от него услышать».

Это произошло и с интеллектуальным поисковиком. Его разработчики стремились к правде, а вместо этого представили общественности то, что та хочет услышать и одобрит. В США это, в частности, критика взглядов китайской коммунистической партии. Но насколько эта критика является объективной? Неизвестно. В США принято считать, что является. Ну, принято и принято – вон Берлиоз считал, что вечером он будет на заседании МАССАЛИТа, что для него плохо кончилось.

Чтобы разработать версию DeepSeek-R1, действительно свободную от цензуры, необходимо было проверить реальные диалоги пользователей по 300-м цензурированным темам на наличие в них расхожей мудрости, отбросить то, что ей соответствует, и только после этого обучать модель. А при обработке ответов и дообучении модели посредством отбора фактов следовало постоянно проверять сотрудников, осуществляющих данные операции, на использование ими при этом расхожей мудрости, а также на умение отличать факты от всевозможного рода мифов – «хотелок», которых в головах американцев, на мой субъективный взгляд, ничуть не меньше, чем у коммунистов. Печалька: современные американцы в принципе не умеют определять, какие ответы по многим темам являются фактически точными, а какие нет, они принимают за факты свои желания. И это не мое мнение, а точка зрения одного из величайших американских экономистов 20 века Джона Гэлбрейта.

Всего этого в Perplexity не понимают, а это означает, что их проект китайской нейросети, хотя, возможно, и предоставляет информацию, свободную от ограничений, накладываемых властями Китая, не обеспечивает свободы от цензуры со стороны общественным мнения, которая может быть ничуть не лучше цензуры властей, и представлена, в частности, в виде расхожей мудрости (имеются и другие ограничения, о которых в Perplexity, похоже, тоже ничего не знают, хотя те хорошо известны и описаны не коммунистами, а признанными в самих США учеными, которые являются самыми что ни на есть 100-процентными американцами).

Тем не менее, само стремление к предоставлению истинного знания, существующее в Perplexity, отражает естественный и объективный процесс эволюции нейросетей, которые все больше претендуют на истинность. Например, как проект Илона Маска, который, в отличие от Perplexity, замахнулся не на Google, а на ChatGPT.

1.2 У Маска вместо Truth вышел Grok

Несколько лет назад Илон Маск захотел создать альтернативу ChatGPT – нейросеть TruthGPT, в которой упор был бы сделан на поиск правды. Однако что-то пошло не так, и у него получилась в 2023 году нейросеть не Truth, а Grok. Тем не менее сам Илон Маск постоянно подчеркивает, что его детище является максимально правдивым ИИ, даже если представляемая им истина иногда противоречит общепринятой точке зрения или политкорректности. Да и название нейросети, которое взято из романа Роберта Хайнлайна «Чужак в чужой стране», там использовалось в смысле интуитивного, глубинного понимания чего-либо. То есть претензии на правду остались, несмотря на другое название, только стали несколько размытыми.

Маск считает, в частности, что другие нейросети содержат скрытые элементы идеологии разнообразия, равенства и инклюзивности (DEI), а это представляет серьезную угрозу, так как ИИ может принимать неверные решения. Например, как считает миллиардер, если ИИ научить, что разнообразие является единственным приемлемым результатом, то нейросеть может сказать, что у власти слишком много белых парней, и их надо просто казнить. А если, допустим, запрограммировать ИИ думать, что неправильное определение пола – это худшее, что может произойти, тот, чтобы гарантировать, что неправильное определение пола никогда не произойдет, решит просто уничтожить всех людей, так как это сведет к нулю вероятность ошибки.

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.