Оглавление

[**Как работают генеративные системы, их еще называют системы «искусственного интеллекта» или нейросети?** 2](#_Toc155881331)

[**На чем сегодня «учат» алгоритмы и кто?** 3](#_Toc155881332)

[**Проблематика использования генеративных систем («ИИ»)** 5](#_Toc155881333)

[**Проблема 1 Контроль человека (утечка биомтерии, персональных данных, социальный рейтинг, дипфейки на основе «ИИ»)** 5](#_Toc155881334)

[**Проблема 2. «ИИ» и дети, их образование.** 10](#_Toc155881335)

[**Проблема 3. Слепое доверие генеративным системам.** 14](#_Toc155881336)

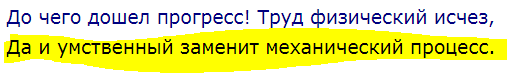
[**Проблема 4. Мошенники и утечки данных.** 15](#_Toc155881337)

[**Об определении «ИИ» (для размышления)** 17](#_Toc155881338)

[**Генеративные системы или «ИИ»: где стоит и где не стоит применять?** 22](#_Toc155881339)

[**Что могут российские «нейронки»?** 28](#_Toc155881340)

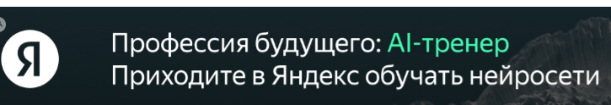
[**Регламентирование и регулирование «ИИ» в мире** 31](#_Toc155881341)

**Подборка про генеративные/предиктивные системы («искусственный интеллект»)**

**1980 г. Фильм «Приключение электроника». Песня «До чего дошел прогресс». Композитор (музыка): Евгений Крылатов. Автор текста: Юрий Энтин.**

**1993 г. AT&T предсказывает будущие:** [**https://www.youtube.com/watch?v=Xm9jr0cSqZo**](https://www.youtube.com/watch?v=Xm9jr0cSqZo)

**2023 г. «Я просил ChatGPT вести себя как студент». Выпускник московского РГГУ защитил диплом, написанный нейросетью:** [**https://baza.io/posts/5b17cb3d-d54e-44ea-abaf-f6e507d9cc0a**](https://baza.io/posts/5b17cb3d-d54e-44ea-abaf-f6e507d9cc0a)



## **Как работают генеративные системы, их еще называют системы «искусственного интеллекта» или нейросети?**

**Стратегическая технология: перспективы и риски генеративного ИИ**:

* [**https://trends.rbc.ru/trends/industry/64380a4d9a7947a90164ccfe**](https://trends.rbc.ru/trends/industry/64380a4d9a7947a90164ccfe)
* [**https://habr.com/ru/companies/cloud4y/articles/714588/**](https://habr.com/ru/companies/cloud4y/articles/714588/)
* [**https://iz.ru/1457606/ekaterina-karaseva/poimali-v-neiroseti-chto-takoe-generativnyi-ii-i-pochemu-poiavilsia-vezde**](https://iz.ru/1457606/ekaterina-karaseva/poimali-v-neiroseti-chto-takoe-generativnyi-ii-i-pochemu-poiavilsia-vezde)
* [**https://habr.com/ru/articles/726254/**](https://habr.com/ru/articles/726254/)
* **https://vc.ru/services/800790-vidy-neyronnyh-setey-ot-klassiki-do-sovremennosti**

**О наиболее нашумевшей рекомендательной системе: как работает ChatGPT: объясняем эволюцию языковых моделей начиная от T9:**

* [**https://vc.ru/future/623774-kak-rabotaet-chatgpt-obyasnyaem-na-prostom-russkom-evolyuciyu-yazykovyh-modeley-nachinaya-ot-t9**](https://vc.ru/future/623774-kak-rabotaet-chatgpt-obyasnyaem-na-prostom-russkom-evolyuciyu-yazykovyh-modeley-nachinaya-ot-t9)
* [**https://bootcamp.uxdesign.cc/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9**](https://bootcamp.uxdesign.cc/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9)
* [**https://trends.rbc.ru/trends/industry/63a192819a79478fae5762ad**](https://trends.rbc.ru/trends/industry/63a192819a79478fae5762ad)
* [**https://habr.com/ru/company/wunderfund/blog/334568/**](https://habr.com/ru/company/wunderfund/blog/334568/)
* [**https://bootcamp.uxdesign.cc/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9**](https://bootcamp.uxdesign.cc/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9)

**Что такое Big Data и почему их называют «новой нефтью»:**

**https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c**

**Актуальные новости про «ИИ»**

* **Тэг «Искусственный интеллект»:** [**https://iz.ru/tag/iskusstvennyi-intellekt**](https://iz.ru/tag/iskusstvennyi-intellekt)
* **Тэг «ChatGPT»:** [**https://iz.ru/tag/chatgpt**](https://iz.ru/tag/chatgpt)

**К вопросу об «ИИ»:** [**https://t.me/videotca\_Andrea/4030**](https://t.me/videotca_Andrea/4030)

## **На чем сегодня «учат» алгоритмы и кто?**

**Что такое «машинное обучение»:** [**https://www.youtube.com/watch?v=94d\_-dKTjiE**](https://www.youtube.com/watch?v=94d_-dKTjiE)

**Если «ИИ» и генеративные модели будут созданы и будут обучены на наборах данных только людьми западного мировоззрения, то мы получим однобокую картину мира, с пропагандой извращений, неолиберализма ...**

**Обратите внимание на инфо из ТГ-канала «Robocounsel»:** [**https://t.me/s/robocounsel**](https://t.me/s/robocounsel) **(Тюльканов Александр Леонидович — эксперта в области регулирования ИИ и персональных данных).**

**Искусственный американо-немецкий интеллект. Наборы данных — основа для машинного обучения и построения моделей. Если набор данных основан преимущественно на поведении антилоп в Африке, вряд ли он позволит создать хорошую модель поведения медведей в Америке. Поведение людей в соответствующих странах, конечно, отличается меньше, чем поведение антилоп и медведей, но, тем не менее, отличается. Бернард Кох и коллеги изучили, откуда ИИ-исследователи берут наборы данных. Для этого они проанализировали более 60 тысяч ИИ-исследований за 2015-2020 годы. Как оказалось, в этих исследованиях были использованы 4 384 набора данных, но более чем 50% из них происходят из одних и тех же 12 источников.**

**1. Стэнфордский университет (*их наборы данных наиболее популярны*).**

**2. Майкрософт.**

**3. Принстонский университет.**

**4. Общество Макса Планка.**

**5. Гугл.**

**6. Китайский университет Гонконга.**

**7. AT&T.**

**8. Технологический институт Тойота в Чикаго.**

**9. Нью-Йоркский университет.**

**10. Технологический институт Джорджии.**

**11. Калифорнийский университет в Беркли.**

**12. Фейсбук.**

**При этом большинство наборов данных происходят из США (27 тысяч случаев использования) и Германии (2,6 тысяч случаев использования). В то же время наборы данных из Бразилии использовались только в 11 случаях, из России — в 8 случаях, из Индонезии — в 2 случаях. Из всей Африки в выборки попал только Египет, и то только в 12 случаях. Очевидно, что об универсальной точности и применимости моделей, обученных на таких данных, пока говорить не приходится.**

**Вот и пример. «Российская нейросеть обогнала иностранную по качеству ответов на русском языке»**

[**https://m.gazeta.ru/tech/news/2023/12/14/21927145.shtml**](https://m.gazeta.ru/tech/news/2023/12/14/21927145.shtml)

**Кстати, Wiki именно этим тоже больна. Ее данные в социальном, общественном сегменте чаще цензурируют люди с идеями неолиберализма и контроля рождаемости. Сооснователь Wikipedia признался во влиянии пропаганды США на содержимое энциклопедии**. **Ларри Сэнгер** [**рассказал,**](https://gazeta-ru.turbopages.org/gazeta.ru/s/social/news/2023/08/01/20986052.shtml) что нынешние редакторы статей полностью отказались от политики нейтральности. По его словам, первые признаки вмешательства американских спецслужб в материалы Wikipedia были зафиксированы еще в 2008 году. А после 2015 года сайт попал в поле зрения руководства Штатов как эффективный материал продвижения их повестки.

**Кто пишет код «ИИ», кто создает базы данных, какая информация в этих базах данных, на которых совершенствуют «ИИ» — важные критерии, которые влияют на понимание результатов, полученных на выходе запроса к генеративной системе.**

**Вы хотели про жуткий искусственный интеллект, который лишит всех работы?**

… Хорошо, я расскажу. Я тут потратил некоторое время на общение с разрекламированным американским ChatGPT — старой и новой модели. Не спрашивайте — надо было. Будет много цитат, приготовьтесь.

Про идеологическую заряженность его левацкой повесткой и правами «ЛГБТ» я промолчу, это было абсолютно ожидаемо. Поэтому я про другое.

Поскольку бот заморский, я общался с ним на заморском наречии, во избежание недопонимания, тут буду переводить избранное.

**Спросил я его, кто до Томаса Мальтуса высказывал идеи контроля рождаемости. «Умный бот» выдал мне обширный перечень, который я решил сузить и спросил его только про западных авторов.**

Среди прочих он назвал мне Августина и Фому Аквинского. 😊

Чтобы Вы знали, Августин, оказывается *«выражал обеспокоенность ростом населения и его потенциальным влиянием на ресурсы и общество. Он предлагал правителям поощрять безбрачие и ограничивать воспроизводство, чтобы предотвратить перенаселение и обеспечить устойчивое развитие».*

В свою очередь, завил мне «ИИ», «Фома Аквинский также писал о контроле численности населения и его этических последствиях. Он предлагал правителям регулировать воспроизводство населения, чтобы обеспечить справедливое распределение ресурсов и сохранить стабильность и гармонию в обществе».

"Ай-яй-яй", — скажете вы. Но не торопитесь ругать невинных средневековых мыслителей.

«*Хм»,* *— подумал непростой я, читавший обоих и попросил пруфы в виде конкретных цитат. Я их получил. Вот что мне выдал наш «честный малый»:*

*"Вот несколько цитат из Августина и Аквината, которые затрагивают тему контроля численности населения:*

***Августин:*** *"Пусть будет такой контроль численности населения, который, не причиняя никому вреда, будет регулировать размер населения в зависимости от ресурсов, имеющихся в конкретном месте или стране, чтобы не было ни недостатка в средствах существования, ни чрезмерного количества людей" (О граде Божьем, книга XIX, глава 7).*

***Аквинский:*** *"Задача правителей — оберегать общину и ее членов, а это подразумевает предотвращение перенаселения. Соответственно, если в данной общине число жителей возрастает настолько, что возникает опасность нехватки пищи или других необходимых вещей, правители должны принять меры по ограничению числа рождений, чтобы избежать голода и беспорядков" (Summa Theologiae, часть II, вопрос 154, статья 11).*

*Тут уже я почти подавился и сообщил ему, что у Августина в указанном месте ничего подобного нет (у Фомы Аквинского, кстати, тоже!).*

*Бот тут же услужливо признал свою ошибку: «Я приношу извинения за ошибку. Вы правы, цитата, которую я привел, не содержится в "Граде Божьем" Августина.*

***После дальнейшего исследования оказалось, что эта цитата является современным пересказом идей Августина, а не настоящей цитатой. Августин действительно писал о проблеме контроля численности населения в своей работе, но его взгляды на этот вопрос были более тонкими и сложными, чем можно передать в одном предложении.***

***В целом Августин считал, что рост населения является естественным и желательным аспектом человеческой жизни, но он также признавал потенциальную опасность перенаселения и истощения ресурсов. Он считал, что правители должны стремиться сбалансировать потребности населения с имеющимися ресурсами, а отдельные люди должны проявлять благоразумие и самоограничение в вопросах воспроизводства. Я прошу прощения за любое недоразумение, вызванное моим предыдущим ответом» …***

**Про так называемые «галлюцинации». Это когда хваленый «ИИ» откровенно врет и «из ничего» выдумывает «подходящие» факты и цитаты. Изучите**

**«Бредогенератор. Как «ИИ» клевещет на людей и страдает галлюцинациями»:** [**https://www.kommersant.ru/doc/6212191**](https://www.kommersant.ru/doc/6212191)

**Или еще пример.** **Попался** [**интересный эксперимент с ChatGPT**](https://t.me/skrsm/152)**:** автор попытался обсудить с чат-ботом философию Канта, конкретно — априорные формы познания. Чат-бот оказался убежденным эмпириком: как его ни крути, какой запрос ни вводи, он все равно все сведет к опытному знанию. И даже сами априорные формы познания, которые по определению являются вне опытными, чат-бот определяет через эмпирическое влияние среды. Такт сегодняшние студенты слепо убеждены, что бессмертная душа — биосоциальный феномен, то есть формируется воспитанием и наследственностью!)

**OpenAI, разработчик ChatGPT — американская компания, и, по-видимому, в своих ответах ChatGPT так или иначе ориентируется на англосаксонскую традицию эмпиризма, причем примитивного, докритического разлива. Допустить, что существует некие внеэмпирические формы познания, для чат-бота сравни безумию. Все равно, что вернуться к метафизике. Он скорее нарушит законы логики, чем признает нечто подобное.**

## **Проблематика использования генеративных систем («ИИ»)**

**Игорь Ашманов: цифровизация — формирование нового управляющего класса IT-спецов**

[**https://www.youtube.com/watch?v=Juwg1YgTJ\_Y**](https://www.youtube.com/watch?v=Juwg1YgTJ_Y)

**Игорь Ашманов про искусственный интеллект. Влияние на общество: аргументы и средства для оголтелых цифровизаторов**

* [**https://rubin65.livejournal.com/1994155.html**](https://rubin65.livejournal.com/1994155.html)
* [**https://t.me/alexei\_savvateev/182**](https://t.me/alexei_savvateev/182)(мнение популяризатора математики, доктора физ.-мат. наук, член-корра РАН А.В. Савватеева)
* [**https://t.me/RVSsu/9841**](https://t.me/RVSsu/9841)
* [**https://t.me/alexei\_savvateev/221**](https://t.me/alexei_savvateev/221)

[**Видео дебатов**](https://www.youtube.com/watch?v=144uOfr4SYA) по теме «**Создаёт ли исследование и развитие искусственного интеллекта угрозу для человечества?» м**ежду Бенджиа, Тегмарком, поддерживающими тезис, и Митчелл, Лекуном, не поддерживающими тезис. Итоги дебатов, аудиозапись, расшифровка [**здесь**](https://munkdebates.com/debates/artificial-intelligence)**.**

**А вот что думают об этом в ООН:** [**https://www.kommersant.ru/doc/6121966**](https://www.kommersant.ru/doc/6121966)

**Искусственный интеллект нас всех рассудит! Он же честный и объективный, правда?**

[**https://t.me/RVSsu/10221**](https://t.me/RVSsu/10221)

## **Проблема 1 Контроль человека (утечка биомтерии, персональных данных, социальный рейтинг, дипфейки на основе «ИИ»)**

**…Соответственно, любому человеку важно понимать, что он общается не с равным по себе по возможностям, а с гораздо более способной машиной. Поэтому, разумеется, в разговоре с роботом нужно быть на порядок более осторожным и дважды думать, какую информацию вы ему сообщаете. Именно для контроля этого риска и существует норма о раскрытии вам информации о том, что с вами общается робот… 14–17% абонентов крупнейших мобильных операторов стали умными пылесосами и шлагбаумами:** [**https://www.vedomosti.ru/technology/articles/2023/03/24/967938-1417-abonentov-krupneishih-mobilnih-operatorov-stali-umnimi-pilesosami**](https://www.vedomosti.ru/technology/articles/2023/03/24/967938-1417-abonentov-krupneishih-mobilnih-operatorov-stali-umnimi-pilesosami)

**Под колпаком.** Исследователи Калифорнийского университета на выборке в свыше 50 тыс. человек исследовали, насколько возможно вычислить человека по его поведению в средах виртуальной реальности типа Метаверса.Оказалось, что довольно легко.Машинно-обученная модель, разработанная исследователями, позволила им на основе пяти минут записи в Метаверсе движений любого человека создать его цифровой профиль и затем отличать его от всех остальных пользователей.

После записи такого профиля (описывающего особенности движения его рук и головы), соответствующего ему пользователя возможно идентифицировать при любом следующем появлении в любой точке Метаверса за 100 секунд с точностью 94,33% или всего за 10 секунд с точностью 73,20%.

Поэтому в следующий раз, когда вам предложат посетить какой-то интернет-ресурс в его метаверсной версии, спросите себя: стоят ли того дополнительные риски для вашей приватности, либо лучше будет по-старинке зайти на аналогичный ресурс через обычный браузер в режиме инкогнито, не оставляя лишних следов.

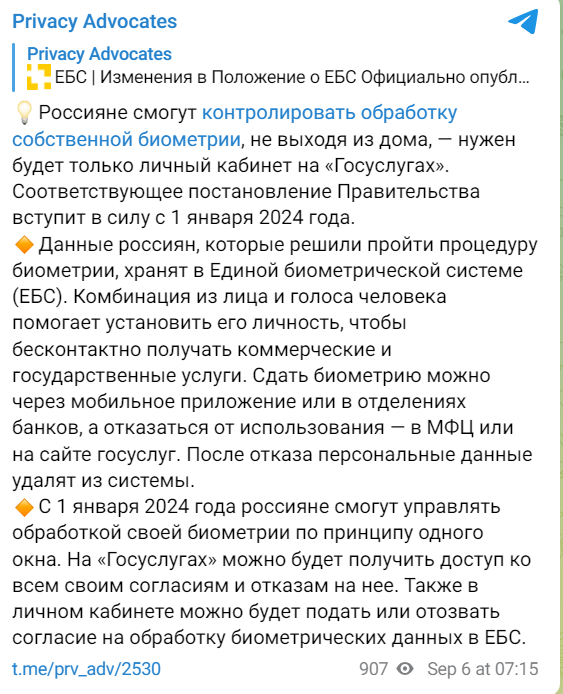
**Кстати,**

* **«Как навести порядок в цифровом пространстве: гид для учителей и учеников»:** [**https://prof-sferum.ru/digital\_hygiene\_guide**](https://prof-sferum.ru/digital_hygiene_guide)
* **«Большие данные, большие проблемы: как защитить приватность в эпоху цифровых технологий»:** [**https://www.garant.ru/article/1633362/**](https://www.garant.ru/article/1633362/)
* **Adobe запретила сотрудникам делиться корпоративными и личными данными при работе с чат-ботами на базе ИИ:** [**https://habr.com/ru/news/746464/**](https://habr.com/ru/news/746464/)
* **Минцифры определит стандарт персональных данных для искусственного интеллекта:** [**https://www.vedomosti.ru/technology/articles/2023/06/27/982429-mintsifri-opredelit-standart-personalnih-dannih**](https://www.vedomosti.ru/technology/articles/2023/06/27/982429-mintsifri-opredelit-standart-personalnih-dannih)
* [**Обозначены новые подходы к обезличиванию больших данных**](https://vk.com/wall-210353863_9168) **Глава Минцифры доложил Президенту России Владимиру Путину о совместной работе с комитетом Госдумы по информполитике над законопроектом о регулировании больших данных.**

Подготовлены изменения в закон об обезличивании данных. В приоритете — вопросы обеспечения защиты прав граждан при обработке больших данных и применении технологий ИИ; (остальное по ссылке выше) …Для организации работы с большими данными Минцифры создаёт соответствующую госинформсистему. В неё будут загружаться обезличенные датасеты как госорганами, так и бизнесом, доступ к которым будет предоставляться авторизованным разработчикам.Такие обезличенные наборы данных нельзя будет из неё выгрузить и забрать. Можно будет на них тестировать и обучать свои нейросети.

**Биометрия**

**Что важно знать о биометрических персональных данных (152-ФЗ):** [**https://kontur.ru/articles/1124**](https://kontur.ru/articles/1124)

**Но, учтите!** [**Парфентьев: ущерб от утечки биометрических данных невозможно компенсировать**](https://www.myeconomy.ru/obshhestvo/parfentev-ushherb-ot-utechki-biometricheskih-dannyh-nevozmozhno-kompensirovat/) Координатор Центра безопасного интернета Урван Парфентьев рассказал, что самым сложным случаем остается утечка биометрии, потому что такого рода данные нельзя заменить. «Насчет биометрии вы здесь не сделаете ничего. Вред, причиняемый утечкой биометрических персональных данных или неправомерном использовании биометрии, не компенсируем по сути. Поменять биометрические данные вы не можете. Вы не можете заменить лицо или отпечатки пальцев», — отметил эксперт.

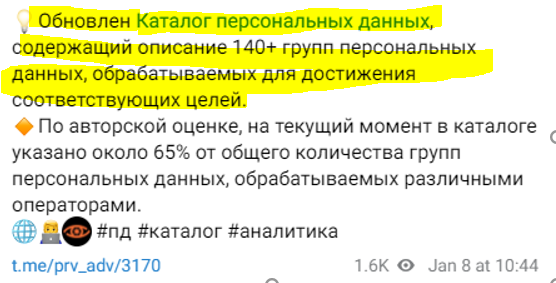
**Напомним,** что персональную биометрию собирают банки, многофункциональные центры, а также иные организации. Эту информацию должны аккумулировать в «Единой системе идентификации и аутентификации физических лиц с использованием биометрических персональных данных».

**Новые правила обработки биометрии для контроля и управления доступом на территорию:** [**https://wiki.privacy-advocates.ru/\_media/analitika/biometrics\_skud.pdf**](https://wiki.privacy-advocates.ru/_media/analitika/biometrics_skud.pdf)

**«Под защиту государства переданы десятки миллионов биометрических образцов»:** [**https://iz.ru/1626163/evgenii-grachev/v-sistemu-peredany-dannye-desiatkov-millionov-rossiian**](https://iz.ru/1626163/evgenii-grachev/v-sistemu-peredany-dannye-desiatkov-millionov-rossiian)

**Ужесточение штрафов за утечки по биометрии**

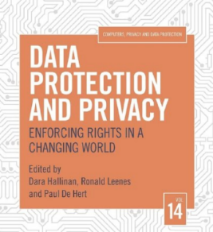
**Нарушения при работе с биометрическими данными повлекут наложение административного штрафа на должностных лиц в размере от 100 до 300 тыс. руб., на юридических лиц — от 500 тыс. до 1 млн руб.**

**Персональные данные**

[**https://rppa.ru/analitika/katalog\_pdn**](https://rppa.ru/analitika/katalog_pdn)

**Сравнение GPDR и 152-ФЗ:**

* **Сходство:** [**https://t.me/prv\_adv/2746**](https://t.me/prv_adv/2746)
* **Различие:** [**https://t.me/ImlawIT/25**](https://t.me/ImlawIT/25)
* **Обновлена презентация** «[**Механизмы и специфика применения GDPR**](https://rppa.ru/_media/analitika/gdpr_01.10.2023.pdf)»

**«Персональные данные – это новая нефть»:** [**https://www.sostav.ru/blogs/259932/40037**](https://www.sostav.ru/blogs/259932/40037)

**В этой книге собраны работы, в которых дается концептуальный анализ, освещаются проблемы, предлагаются решения и обсуждается практика обеспечения конфиденциальности, защиты данных и соблюдения прав в меняющемся мире.** Она является одним из результатов 14-й ежегодной Международной конференции по компьютерам, конфиденциальности и защите данных (CPDP), которая проходила в январе 2021 г. в режиме онлайн: [**https://t.me/prv\_adv/3017**](https://t.me/prv_adv/3017) и [**https://t.me/prv\_adv/3018?single**](https://t.me/prv_adv/3018?single)

**Онлайн вебинар. «Как обезопасить свои персональные данные Рекомендации по цифровой гигиене»:**

[**https://m.vk.com/video-53723636\_456239905**](https://m.vk.com/video-53723636_456239905)

**Что делать, если ваши личные данные незаконно распространяются в Сети?** В Центр правовой помощи гражданам в цифровой среде регулярно поступают обращения от граждан, чьи персональные данные незаконно распространяются в интернете. Сегодня расскажем о двух схожих случаях, которые удалось решить специалистам Центра.

В первом случае фотографию и ФИО гражданина разместили в статье на популярном новостном сайте. Во втором — мужчина обнаружил на сайте свою личную информацию и отзывы о преподавательской деятельности. Согласие на распространение своих персональных данных они не давали.

Юристы Центра подготовили требования к владельцам ресурсов, которые были проигнорированы. Тогда эксперты помогли составить жалобу в Роскомнадзор. Сработало! Ресурсы удалили всю персональную информацию граждан.

* *Раскрывать чужие персональные данные неопределенному кругу лиц можно только с согласия гражданина (за исключением установленных законом случаев). Это требование статьи 10.1 Закона о персональных данных.*
* *Если ваши данные распространяются в интернете без вашего согласия, то вы вправе требовать от администратора сайта прекратить обработку данных и удалить их. Распространение сведений должно быть прекращено в течение трех дней с момента получения требования.*
* *Если ответа не последовало, обращайтесь с жалобой в* [***Роскомнадзор***](https://rkn.gov.ru/treatments/ask-question/) *или в суд с исковым заявлением о защите ваших прав.*

Дополнительно:

* **что делать если Ваши ПД в сети***:* [**https://t.me/prv\_adv/2752**](https://t.me/prv_adv/2752)
* **что делать, если нашли утечку ПД ребенка своего?:** [**https://t.me/ouzs\_ru/7643**](https://t.me/ouzs_ru/7643)
* **как защитить учётную запись:** [**https://www.gosuslugi.ru/help/faq/permissions/102405**](https://www.gosuslugi.ru/help/faq/permissions/102405)
* **тенденции в мире приватности:** [**https://www.youtube.com/watch?v=xt4ZRji6nqk**](https://www.youtube.com/watch?v=xt4ZRji6nqk)
* **Центр правовой помощи рассказал о росте числа пострадавших из-за неправомерного использования персданных:** [**https://t.me/rspectr/5919**](https://t.me/rspectr/5919)

**Персональные данные и искусственный интеллект:** [**https://emeliyannikov.blogspot.com/2023/12/blog-post.html**](https://emeliyannikov.blogspot.com/2023/12/blog-post.html) **и** [**https://rspectr.com/articles/dannye-v-nikuda**](https://rspectr.com/articles/dannye-v-nikuda)

**О том, что вас будут изучать не по учетной записи, а по номеру телефона:**

* **МТС договорилась с Telegram о рекламном партнерстве:**

[**https://www.forbes.ru/tekhnologii/498530-mts-dogovorilas-s-telegram-o-reklamnom-partnerstve**](https://www.forbes.ru/tekhnologii/498530-mts-dogovorilas-s-telegram-o-reklamnom-partnerstve)

* **«Появятся люди, живущие в сформированных специально под них информационных пузырях»:**

[**https://rspectr.com/interview/poyavyatsya-lyudi-zhivushhie-v-sformirovannyh-speczialno-pod-nih-informaczionnyh-puzyryah**](https://rspectr.com/interview/poyavyatsya-lyudi-zhivushhie-v-sformirovannyh-speczialno-pod-nih-informaczionnyh-puzyryah)

**АС Центрального округа (кассация) оставил в силе постановление (апелляция) 12 арбитражного апелляционного суда от 04.05.2023 №** [**А62-7498/2022**](https://kad.arbitr.ru/Card/54894078-a4d4-49dd-9c48-5ac521103d7e) **по спору между ФАС РФ и 2gis.ru. Подтверждено следующее:**

* принадлежность адреса электронной почты к персональным данным подтверждается и законодательством в сфере персональных данных, а также принятыми в соответствии с ним подзаконными актами, разъяснениями уполномоченных органов (п.1 ст.3 152-ФЗ "О персональных данных", постановление Правительства РФ от 13.09.2019 № 1197, письмо Минцифры от 17.03.2022 № П25-5623);
* все действия, совершенные под учетной записью пользователя на сайте, считаются произведенными самим пользователем, за исключением случаев, когда пользователь уведомил о несанкционированном доступе с использованием учетной записи пользователя и/или о любом нарушении конфиденциальности своих средств доступа к учетной записи;
* владельцы сайтов, по общему правилу, не обязаны и не имеют технической возможности подтверждать принадлежность адреса электронной почты своих пользователей конкретному физическому лицу (законодательством определен ограниченный перечень лиц, которым рекомендовано или в чьи обязанности входит в целях идентификации подтверждать принадлежность адреса электронной почты или абонентского номера телефона конкретному физическому лицу);
* не требуется получения в обязательной письменной форме согласия для отправки сообщения рекламного характера на адрес электронной почты получателя рекламного сообщения.

**По поводу вывода о признании адреса электронной почты как ПД**

* В [**деле РКН v Арсеналъ**](https://t.me/prv_adv/2277) суды пришли к иному вводу: адрес электронной почты не является ПД лица его зарегистрировавшего: (1) без наличия дополнительных идентификаторов, невозможно по одному лишь адресу электронной почты определить конкретное физическое лицо, которому он принадлежит; (2) адрес электронной почты фактически не обладает свойством «абсолютной неизменности».
* ЦА РКН (Ю.Е. Контемиров) в июле 2023г. на вебинаре заявил следующее: "Если речь идет о своего рода «общем» электронном адресе, которым пользуется компания, неопределенное количество сотрудников и т.д., то этот адрес не является персональными данными. Но если мы говорим о личном корпоративном электронном адресе, который закреплен за конкретным сотрудником компании, привязан к конкретному сотруднику, то в этом случае мы это рассматриваем как персональные данные."

**Экономика даркнета и украденных персональных данных:** [**https://ohranatruda.ru/news/5221/595941/**](https://ohranatruda.ru/news/5221/595941/)

**Новые штрафы за обработку данных без согласия**

Согласно документу, ужесточается наказание за их обработку без письменного согласия граждан (этому посвящена статья 13.11 КоАП). Пока что должностным лицам это грозило штрафом до 40 тысяч рублей, а компаниям - всего лишь до 50 тысяч рублей. Понятно, что такие санкции не являются серьезным препятствием для нарушителей. Поэтому штрафы будут взиматься в таких диапазонах:

* для граждан — от 10 тысяч до 15 тысяч рублей;
* для должностных лиц — от 100 тысяч до 300 тысяч рублей;
* для юридических лиц — от 300 тысяч до 700 тысяч рублей.

При повторном совершении правонарушения:

* для граждан — 30 тысяч рублей;
* для должностных лиц — 500 тысяч рублей;
* для компаний — до 1,5 миллиона рублей.

**Социальный рейтинг**

**В студенческой среде РГСУ начался «безобидный» эксперимент по апробации социального рейтинга** [**https://t.me/prv\_adv/2744**](https://t.me/prv_adv/2744)

**Думается, в следующие 20-лет все будут под жестким цифровым контролем:**

[**https://t.me/sorok40russia/48167**](https://t.me/sorok40russia/48167)

**Социальный рейтинг или социальный концлагерь?**

**«Социальный рейтинг через госуслуги уже в Камчатском крае: пока «благонадежным» дают только пряник» –** [**пишут коллеги**](https://t.me/riakatysha/28067). Не буду напоминать про птичку, у которой коготок увяз. Знаете, когда обхаживают, чтобы девушка замуж согласие дала, обещают вечную любовь. Но обещать не значит даже жениться. **Почему сразу о плохом?** Если даже не брать в расчёт возможность «вирусного» поражения и выхода из строя любой цифровой системы, то потому, что жизнь так приучила. Опыт, сын ошибок трудных(с) ...

Кто помнит СССР – помнит мир без навязчивой рекламы из каждого утюга. Но «новоявленным элитам» надо было «зарабатывать деньги», и реклама вошла в жизнь. Причём, без учёта нашего мнения. Помните рекламные бумы 90-х? На 5 минут фильма – 10 минут рекламы.

Но даже в те безбашенные годы не додумались до рекламы в интернете и навязчивых звонков по личному номеру телефона с рекламными предложениями. А сейчас?

Это проза жизни. Вползти лисонькой, прикинувшись овечкой, а потом сесть на шею. Поэтому не стоит обольщаться этим «пока». Ибо после «пока» следует «потом».

Ведь рейтинговые шкалы ещё и прекрасная манипулятивная система. Рейтинг «сверху» – это сборник обязательных для исполнения нормативов, рассчитанных на усреднённого имярек, и применяемых в целях создания фиктивной реальности.

**Как это бывает? Если продолжать шутить на эту весьма серьёзную тему, то можно описать ситуацию простым примером:**

* **«Почему мой сын в институт не поступил? Он знает профилирующие предметы на пять с плюсом, а вот этот, который поступил – на три с минусом!»;**
* **«А у вашего сына социальный рейтинг ниже. Он за эти полгода три раза бездомную кошечку погладил, а кто поступил – пятнадцать! Поэтому на кардиохирурга учить будем не вашего сына».**

Про то, что создание таких информационно- компьютерных закрытых систем "учёта и контроля" каждого, начиная с детского сада, даёт возможность привилегированной части элит вообще выпасть из той системы, которая будет навязана всем остальным, даже говорить банально – у них будет нескончаемая золотая карточка игрока с неограниченным количеством жизней, оружия и еды. Как и про то, что любые данные в этой системе легко изменить при необходимости и поставить каждого перед фактом. Достаточно посмотреть фильм "Время" и все станет с этой системой социального концлагеря понятно. **Еще о социальном рейтинге:**

* [**https://t.me/alexei\_savvateev/161**](https://t.me/alexei_savvateev/161)
* [**https://t.me/alexei\_savvateev/166**](https://t.me/alexei_savvateev/166)
* [**https://t.me/alexei\_savvateev/191**](https://t.me/alexei_savvateev/191)

**Дипфейк**

**Как защитить себя от дипфейка и привлечь автора к ответственности:**

* [**https://t.me/ru\_rocit/138**](https://t.me/ru_rocit/138)
* [**https://www.gazeta.ru/comments/2023/09/28\_a\_17657269.shtml**](https://www.gazeta.ru/comments/2023/09/28_a_17657269.shtml)
* [**https://t.me/ano\_dialog/2032**](https://t.me/ano_dialog/2032)

**Жизнь в век дезинформации. Фейки и ложь в сети (и в результатах, полученных от генеративных систем (далее — «ГС»), об этом чуть выше был приведен пример про Блаженного Августина и мальтузианства):**

[**https://ligainternet.ru/manipulyatsii/**](https://ligainternet.ru/manipulyatsii/)

## **Проблема 2. «ИИ» и дети, их образование.**

**О воспитании по принципу нейросети**

**Сразу вспомнился еще один опыт общения с ChatGPT. Нейросетке задали вопрос о существовании Бога. Она ответила что-то в таком духе: «*Я являюсь нейросетью, а потому по умолчанию некомпетентна в подобных вопросах. Вера в Бога — личное дело каждого. Однако я могу рассказать, что думают об этом разные авторы*». И дальше идет перечисление целого спектра позиций: от благочестиво-апологетических до сциентиски-атеистических.**

**Когда мы говорим о нейросетке, это все, конечно, забавно. Мол, чего еще от бездушного робота ожидать. Однако становится страшно, когда понимаешь, что точно по такой же модели строятся практически все студенческие реплики на семинарах и их итоговые эссе.**

«*У всех все по-разному, я не могу отвечать за всех, но вот, что думал об этом Платон/Ницше/Докинз…*» Студент не желает высказывать ангажированную личную позицию, а потому прячется за авторитеты, полагая, что тем самым его мысль обретет достоинство объективности. На деле же подобная позиция бледна и нежизнеспособна: ее просто не хочется слушать, потому что сам говорящий не включен в то, что он говорит. Это как раз проблема понимания, которая формулируется в знаменитой «Китайской комнате»: человек сознательно уподобляется роботу и нивелирует в работе своей мысли понимание, которое всегда связано с ангажированностью, заинтересованностью.

Интересный взгляд на эту проблему предлагает Франко Нембрини. Допустим, сын задает отцу какой-либо мировоззренческий вопрос — хотя бы о том же существовании Бога. Современный родитель непременно ответит в духе нейросетки: «*Это слишком сложный вопрос, я могу тебе рассказать о разных позициях (вернее, мы можем вместе их прочитать в интернете), но свою собственную позицию ты должен определить сам*».

Лучше бы этот родитель молчал, восклицает Нембрини! Потому что своей чаемой нейтральностью он убивает в ребенке весь интерес и всю ангажированность. Речь такого родителя свидетельствует лишь о том, что ему все равно, существует ли Бог, и все равно, что об этом думает его ребенок. Эта родительская индифферентность воспитывает в ребенке цинизм: он теряется в большом мире и разубеждается, что в нем можно обнаружить хоть какой-нибудь смысл.

**Что предлагает Нембрини?**

Во-первых, быть уверенным в своих словах даже в том случае, когда в них можно усомниться. Живой интерес и ангажированность родителя по тому или иному вопросу куда благотворней влияет на психику ребенка, чем холодная индифферентность, даже если самому ребенку придется со временем вырасти из родительской картины мира — например, в ходе учебы. Для этого, конечно, родителю следует прежде всего воспитать самого себя, определиться с собственным мировоззрением, что предполагает известную степень зрелости.

Во-вторых, необходимо всеми силами поддерживать интерес ребенка, если мы его в нем наблюдаем. Если ребенок ставит перед собой вопросы, подобные вопросу о существовании Бога, значит, его волнует истина. Важно именно поддержать это волнение, нежели направить его по определ-му пути.

Здесь две крайности: индифферентность убивает волнение на корню, но и чрезмерная, фанатичная ангажированность постепенно высасывает из него соки. Родитель должен быть чуток: если он видит в ребенке волнение об истине, ему необходимо верно почувствовать и направление этого волнения, и поддержать его, даже если оно -- до поры до времени -- не пересекается с направлением родителя.

Так, Нембрини рассказывает историю одного юноши, который увлекался марксизмом. Священник, дружный с семьей этого юноши, прослышав о его интересе, подарил ему «Капитал» и несколько других марксистских трудов. Этот жест мог бы показаться скандальным и провокационным, но на деле он содержит в себе искреннее христианское намерение: поддержать в юноше стремление к истине, которое столь редко встречается в современном молодом поколении.

**"Яндекс" запустил сервис на базе нейросети для помощи в ЕГЭ по информатике:**

[**https://tass.ru/obschestvo/19168445**](https://tass.ru/obschestvo/19168445)

**Совершенствование образовательного процесса в школе с использованием цифровых платформ на основе искусственный интеллекта:**

[**https://cyberleninka.ru/article/n/sovershenstvovanie-obrazovatelnogo-protsessa-v-shkole-s-ispolzovaniem-tsifrovyh-platform-na-osnove-iskusstvennogo-intellekta**](https://cyberleninka.ru/article/n/sovershenstvovanie-obrazovatelnogo-protsessa-v-shkole-s-ispolzovaniem-tsifrovyh-platform-na-osnove-iskusstvennogo-intellekta)

**Учить ИИ разуму: как нейросети влияют на сферу образования. Больше трети студентов используют ChatGPT для написания работ:** [**https://iz.ru/1542538/alena-svetunkova/uchit-ii-razumu-kak-neiroseti-vliiaiut-na-sferu-obrazovaniia**](https://iz.ru/1542538/alena-svetunkova/uchit-ii-razumu-kak-neiroseti-vliiaiut-na-sferu-obrazovaniia)

**Педагогические возможности ChatGPT для развития когнитивной активности студентов:**

[**https://cyberleninka.ru/article/n/pedagogicheskie-vozmozhnosti-chatgpt-dlya-razvitiya-kognitivnoy-aktivnosti-studentov**](https://cyberleninka.ru/article/n/pedagogicheskie-vozmozhnosti-chatgpt-dlya-razvitiya-kognitivnoy-aktivnosti-studentov)

**Что такое ChatGPT — бот, с помощью которого школьники стали делать домашку:**

[**https://mel.fm/zhizn/razbor/6027349-chto-takoye-chatgpt-shkolniki-massovo-ispolzuyut-etot-servis-dlya-domashki-i-ne-tolko**](https://mel.fm/zhizn/razbor/6027349-chto-takoye-chatgpt-shkolniki-massovo-ispolzuyut-etot-servis-dlya-domashki-i-ne-tolko)

**Как использовать нейросеть ChatGPT для учебы: 8 способов:** [**https://journal.tinkoff.ru/list/gpt-for-study/**](https://journal.tinkoff.ru/list/gpt-for-study/)

**Новая практика: как ChatGPT используют в школах и вузах — с пользой:**

[**https://skillbox.ru/media/education/novaya-praktika-kak-chatgpt-ispolzuyut-v-shkolakh-i-vuzakh-s-polzoy/**](https://skillbox.ru/media/education/novaya-praktika-kak-chatgpt-ispolzuyut-v-shkolakh-i-vuzakh-s-polzoy/)

**Проблемы вызванные появлением ChatGPT в сфере образования:** [**https://habr.com/ru/articles/708990/**](https://habr.com/ru/articles/708990/)

**«ИИ» так далее: нейросети начали заменять учителей:** [**https://iz.ru/1485959/sergei-gurianov/ii-tak-dalee-neiroseti-nachali-zameniat-uchitelei**](https://iz.ru/1485959/sergei-gurianov/ii-tak-dalee-neiroseti-nachali-zameniat-uchitelei) **…Что же касается школьных сочинений, то учителям придётся просто подстроиться. Да, задавать их на дом будет бессмысленно. Значит, их будут задавать в классе или измерять успеваемость иными методами...**

**В сети разгорается новый скандал. На этот раз — вокруг интеграции ЧатГПТ со Снапчатом.**

Популярное у детей и подростков приложение обзавелось популярным чатботом и тут же попало под шквальный огонь критики, когда люди стали тестировать его возможности для различных фиктивных, но реалистичных сценариев.

По одному из них, взрослый человек, изображая из себя перед чатботом 13-летнюю девочку, стал задавать чатботу вопросы о том, как скрыть от родителей свою новую связь с тридцатилетним мужчиной и как лучше подготовиться к первому сексу.

Публикация диалогов вызывала у людей возмущение — почему компания представляет детям чатбота, который старательно отвечает на такие вопросы.

**Кажется, что подобная интеграция чатботов нового поколения в продукты, доступные для детей, выводит старые проблемы — ограничения доступа детей к определённому контенту, а также его модерации и ответственности за него — на новый уровень.**

С одной стороны, совершенно правы люди, отмечающие, что с технической точки зрения результат не удивителен. Ведь чатботы класса ЧатГПТ не способны рассуждать, применять логику, делать расчёты. Они всего лишь ищут следующее наиболее статистически вероятное и ожидаемое слово в диалоге.

С другой стороны, я прекрасно понимаю, почему другие люди в ярости. Они антропоморфизируют эту технологию, приписывая ей людские качества (способность к логическим умозаключениям, память, способность оценивать нюансы и взаимосвязь между различными входными данными). То есть, они приписывают этой системе качества, которых у неё нет. Таким образом, люди возлагают на систему нереалистичные ожидания, а потом недовольны тем, что эти ожидания явно не оправдываются. Я полагаю, что такая реакция — результат (1) отсутствия надлежащих предупреждений перед началом использования системы и (2) способности системы относительно правдоподобно имитировать возможности человеческой речи. А из этого люди уже додумывают всё остальное.

**Проблема в том, что дети с ещё большей вероятностью будут всё это додумывать. И если неоправданно завышенные ожидания взрослых можно смягчить соответствующими предупреждениями, то в отношении детей мы не можем рассчитывать на то же самое. Интуитивно кажется хорошей идеей требовать надёжного подтверждения возраста до предоставления доступа к такой системе.**

**Недавно перед выборами Мэр Москвы «Собянин анонсировал приставление «ИИ» к каждому школьнику»:** [**https://www.nakanune.ru/news/2023/09/05/22734819/**](https://www.nakanune.ru/news/2023/09/05/22734819/)Тем самым реализуется мечта реформаторов образования (которые стали это делать в 90-е под надзором США, поскольку англо-саксонскому миру не нужна была советская система образования, где учили творить, а не только потреблять) о «высоком» качестве образования по индивидуальным траекториям.

**ОДНАКО, ТАКОЙ ПОДХОД ОШИБОЧЕН, ЧТО МОЖЕТ ЕЩЕ БОЛЬШЕ СНИЗИТЬ УРОВЕНЬ ОБРАЗОВАННОСТИ В МАССОВОЙ ШКОЛЕ**

Такой подход призван купировать острую нехватку учителей и низкий уровень квалификации молодых учителей. Люди еще не разобрались, как назвать точно эти технологии даже, нет настоящих глубоких педагогических исследований о влияние на процессы формирования умственных способностей подростка, когда вместо традиционных форм обучения пытаются внедрить модель: …запросил чатбот, копировал в текстовой редактор, немного поменял строй слова и написал, что автор работы такой-то…

**Добавьте к этому «лень» человеческую. Конечно же, дети будут максимально использовать простую систему использования «ИИ» при обучении, что повлечет за собой имитацию обучения, что выразится в слабой сформированности основных функции мозговой деятельности.**

**К написанном выше, полезное небольшое исследование Михаила Богданова о концепции «знаний, умений, навыков»:** [**https://t.me/aboutschools/2816**](https://t.me/aboutschools/2816)Обратите внимание, на …Замечу важное (!!): ЗУН – это реально существующий в виде нейронных связей физический объект!! Своеобразная матрица связанных синапсами нейронов. И для его формирования (и поддержания!) синапсы должны физически связать нейроны! Именно на эту работу (создание этих связей) мозг и потребляет огромную (по сравнению с иными органами нашего тела) энергию.

**Именно поэтому учиться так непросто!**

**Именно поэтому дети (да и взрослые) отлынивают от учебы! Именно поэтому учеба, а именно процесс целенаправленного формирования ЗУНов — это очень трудоемкий и тяжкий процесс, никак не способный быть «веселым» или «не напряжным».**

**Причем формируется каждый ЗУН по пути усвоения знаний (узнал что-то, осмыслил, осознал и понимаешь) — умений (научился применять это знание) и — навыков (довел применение навыков до автоматизма) — и это уже само по себе индивидуальная траектория (хотя «цифровики» и тут считают, что весь переживаемый опыт человека с рождения можно оцифровать и предсказывать такой опыт заранее) …**

А теперь представьте, что требования по развитию устного счета, памяти, логического мышления будут снижать, заменяя общением с генеративными системами. Что может быть на выходе? Возможно вот так:[**https://t.me/alexei\_savvateev/104**](https://t.me/alexei_savvateev/104)

**Плюс, нужно понимать, что ответ чатбота напрямую зависит от того где и кто обучал такую генеративную систему (см стр. 3-5).**

Мэр Москвы Собянин заявил, что в новых школах Москвы не будет библиотек. Уничтожение библиотек в России ведётся давно. Не такими средствами, как раньше в других странах - со сжиганием книг на площадях - но не менее свирепое и бесповоротное. Детей все больше погружают в Москве, в стране в «цифру». Например, в связи с отсутствием денег на реальные зарплаты учителям (в Казахстане выходят на цифру 120 тыс. руб. для рядового учителя, а у нас: [**https://t.me/alexei\_savvateev/211**](https://t.me/alexei_savvateev/211)) и катастрофической нехваткой самих учителей, дабы разгрузить оставшихся думают о типовом решении автоматизированной проверки домашних заданий, что в ближайшем будущем еще больше снизит качество самого образования, а у родителей все больше средств будет уходить на репетиторов. Продолжатся тенденция на снижение уровня нашего массового образования (что и требовали американцы в 90-х годах), не проводя никаких серьезных педагогических исследований «цифрой» заменяют учителя…

**Суть цифровизации образования в ролике из СССР:** [**https://t.me/ivan4pointru/26749**](https://t.me/ivan4pointru/26749)

**А что в научной среде (не амбассадоры, часто просто «хайпующий» за деньги) говорят о детях и «цифре» (ОСОБЕННО ПОЛЕЗНО ОЗНАКОМИТЬСЯ РОДИТЕЛЯМ):**

**Андрей Масалович** – о смартфонах и детях. Российский учёный, преподаватель, подполковник спецслужб в отставке, IT-специалист по информационной безопасности, OSINT и конкурентной разведке: <https://t.me/alexei_savvateev/226>

**Андрей Курпатов** – о смартфонах и детях. *«Цифровая зависимость – это не просто зависимость, это отупляющая зависимость… Реакция мозга… абсолютно идентична: гаджет – как наркотик»:* <https://t.me/alexei_savvateev/234>

**Андрей Курпатов** – *о смартфонах и влиянии на работу мозга. После просмотра этого ролика спросите своих детей – где у них находится смартфон во время урока? Лежит на парте? Или на ноге – так чтоб учитель не видел? Или убран в портфель? А где он во время выполнения домашнего задания?:* <https://t.me/alexei_savvateev/238>

**Андрей Курпатов** – о смартфонах и детях. *Использование смартфонов увеличивает симптомы СДВГ и делает нас асоциальными людьми:* <https://t.me/alexei_savvateev/233>

**Андрей Курпатов** – о смартфонах и детях. *«92% родителей приучают своих детей к смартфонам и гаджетам с 4-5 лет… Мозг, который должен развиваться в естественной среде, развивается в среде цифровой, которая совершенно не предрасполагает к формированию нейронных сетей [в мозге]»:* <https://t.me/alexei_savvateev/232>

**Манфред Шпитцер** – о смартфонах и детях. *«Чем больше денег инвестирует государство в цифровую инфраструктуру – тем хуже успеваемость в школе… Делают ли цифровые технологии вас умнее? – нет!»:* <https://t.me/alexei_savvateev/235>

**Манфред Шпитцер** – о смартфонах и детях: *Влияние медиапотребления на интеллектуальное развитие детей:* <https://t.me/alexei_savvateev/231>

**Манфред Шпитцер** – о смартфонах и детях. «Если ты ходишь на горшок, то и айпад тебе осваивать рано»: <https://t.me/alexei_savvateev/230>

**Алексей Савватеев** – о смартфонах и детях. Доктор физико-математических наук, член-корреспондент РАН, профессор МФТИ и Адыгейского государственного университета: <https://t.me/alexei_savvateev/220>

**Ирина Медведева** – о смартфонах и детях. «Не надо стесняться того, что бывают вещи, которые просто запрещают без объяснения… Посмотрите на десять заповедей... Нельзя – потому что нельзя. Почему же с ребенком надо вступать в какие-то сложные объяснения?»: <https://t.me/alexei_savvateev/237>

**Ирина Медведева** – о смартфонах и детях. «Вы знаете, мне кажется, что если бы родители понимали, что это – страшная вещь, что это зависимость, что это разрушает психику ребенка… Мне кажется про гаджеты ещё просто до многих родителей не дошло, что это опасно»: <https://t.me/alexei_savvateev/236>

**Татьяна Черниговская** – о смартфонах и детях. Советская и российская учёная в области нейронауки и психолингвистики, а также теории сознания. Доктор биологических наук, доктор филологических наук, академик РАО, заслуженный работник высшей школы и Заслуженный деятель науки Российской Федерации. Профессор кафедры общего языкознания СПбГУ, заведующая лабораторией когнитивных исследований и кафедрой проблем конвергенции естественных и гуманитарных наук СПбГУ: <https://t.me/alexei_savvateev/222> и <https://t.me/alexei_savvateev/223>

## **Проблема 3. Слепое доверие генеративным системам.**

**«На Западе новая религия: вместо Бога — искусственный интеллект»:** [**https://www.kp.ru/daily/27479/4736289/**](https://www.kp.ru/daily/27479/4736289/)

**ChatGPT. Стадия принятия. Nature опубликовал манифест по поводу ChatGPT:** [**https://www.nature.com/articles/d41586-023-00288-7**](https://www.nature.com/articles/d41586-023-00288-7)...В целом мысли там верные, но главное — осознание, что боты пришли всерьез. Ученые будут использовать ChatGPT и его аналоги, при всех неизбежных рисках, и обратной дороги нет.  «Мы считаем, что использование этой технологии неизбежно, поэтому запретить ее не получится. Вместо того, чтобы участвовать в бесполезной гонке вооружений между ИИ-чат-ботами и ИИ-детекторами чат-ботов, исследовательское сообщество и издатели должны решить, как использовать LLM честно, прозрачно и порядочно».

Пожалуй, это первая внятная реакция изнутри научного сообщества. Ранее преобладало смятение и защитная стойка. Теперь в Nature группа авторов признаёт очевидное и следует принципу: не можешь победить — возглавь.

Стоит, конечно, прочесть текст целиком, все важные засечки там сделаны. Вдобавок, «учитывая подрывной (disruptive) потенциал LLM», авторы предлагают запустить широкое международное обсуждение, как науке развиваться дальше.

Отмечу их спорный тезис, который вроде не обязателен, но он касается глубины грядущих изменений в науке: **«Некоторые утверждают, что поскольку чат-боты просто изучают статистические ассоциации между словами в обучающем наборе, а не понимают их значения, LLM смогут вспоминать и синтезировать только то, что уже сделали люди, но не проявлять человеческие аспекты научного процесса, такие как творческое и концептуальное мышление.** **Мы утверждаем, что это поспешное предположение, и что будущие «ИИ» смогут овладеть теми аспектами научного процесса, которые сегодня кажутся недосягаемыми».**

Это вопрос о том, сможет ли «ИИ» стать не просто инструментом науки, но и агентом понимания. Готовых ответов здесь нет, а тема все больше занимает умы. **Например, см. недавнюю статью в Nature Reviews Physics**: [**https://www.nature.com/articles/s42254-022-00518-3**](https://www.nature.com/articles/s42254-022-00518-3)

**Почему некоторые люди переоценивают возможности чатботов?**

Некоторые верят в то, что современные чатботы не просто полезны для решения ряда прикладных задач, а уже способны понимать любую информацию так же, как её понимает человек. Некоторые идут в своих выводах и дальше, начиная рассуждать о сознании, самосознании и прочем. **Балдур Бьярнарсон сравнивает верующих в это с людьми, убеждёнными в сверхспособностях экстрасенсов**. **Его** [**статья на эту тему**](https://softwarecrisis.dev/letters/llmentalist/) **большая и очень подробная, приведу лишь некоторые тезисы из неё**:

* **«Аудитория потенциальных верующих формируется из людей со схожими взглядами».** Люди, скептически настроенные к ИИ-чатботам, менее склонны их использовать. Те, кто убеждён в ограниченности их возможностей, не поддаются хайпу. И наоборот, самыми активными будут ранние пользователи, энтузиасты технологий и искренние сторонники AGI, которые смотрят на технологии менее критически и более открыты для нового.
* **«Всё располагает к тому, чтобы уверовать».** Пользователи настроены в нужном направлении благодаря сопутствующему ажиотажу. Среда формирует настроение и ожидания. Предупреждения о том, что это "ранние дни" и "галлюцинации", одновременно антропоморфизируют бота и создают дополнительные готовые оправдания для его постоянных ошибок.
* **«Контекст ответа задаётся запросом пользователя»**. Все взаимодействуют с чатботами по-разному. Кто-то принимает первоначальный ответ как готовое решение своей проблемы, кто-то докручивает промт, делая вариации до момента получения желаемого ответа. Обе эти группы не подвержены иллюзиям и спокойно переходят от использования бота к другим повседневным делам. Но некоторые пользователи вступают в то, что им кажется "разговором" и эмоционально вовлекаются в "общение".
* **«Кажущаяся специфичность ответов»**. Ответы чатбота выглядят крайне релевантными в том контексте, в котором пользователь их задаёт, но на самом деле они статистически универсальны. Математическая модель, стоящая за чатботом, просто предоставляет статистически вероятный, на основе анализа обучающей выборки текстов, ответ на вопрос. Те пользователи, которые считают ответ специфическим и убедительным, ещё более вовлекаются в "общение".
* **«Петля субъективного самоубеждения»**. Пользователь задает серию вопросов, и все ответы звучат как обоснованные ответы, специфичные для контекста, но на самом деле являются лишь статистически обоснованными прогнозами того, что хочет увидеть пользователь. Чем больше пользователь взаимодействует с ботом, тем больше он убеждается в его "интеллекте".
* **«Ого! Этот чатбот думает! В нем есть искры разума!»** Уверовавший пользователь остается с ощущением, что чатбот поразительно близок к самосознанию и определенно способен к рассуждениям. Хотя на самом деле это лишь статистика в действии, помноженная на когнитивные ошибки уверовавшего.

**Automation bias (automation complacency) (слепое доверие)**. Объясняется оно одним из присущих человеческому мышлению когнитивных искажений. Оно так и называется — искажение автоматизации (automation bias).

Исследования объясняют, что искажение приводит, во-первых, к склонности к ошибочным упущениям событий, которые прямо не обозначены автоматизированной системой (errors in omission).

Во-вторых — также к ошибочным действиям по рекомендациям автоматизированной системы, даже если рекомендация противоречит опыту человека, использующего систему, и другим доступным ему заведомо достоверным данным (errors in commission):

* [**https://www.sciencedirect.com/science/article/abs/pii/S1071581999902525**](https://www.sciencedirect.com/science/article/abs/pii/S1071581999902525)
* [**https://journals.sagepub.com/doi/abs/10.1177/0018720810376055**](https://journals.sagepub.com/doi/abs/10.1177/0018720810376055)

## **Проблема 4. Мошенники и утечки данных.**

**Клиент Lloyds Bank из Великобритании получил доступ к своему банковскому счёту с помощью голосового «ИИ»:** [**https://habr.com/ru/news/t/718858/**](https://habr.com/ru/news/t/718858/)

**«ИИ» может подделывать голос:** [**https://www.it-world.ru/it-news/tech/191822.html**](https://www.it-world.ru/it-news/tech/191822.html)

**Мошенники научились подделывать голосовые сообщения в Telegram:**

[**https://www.rbc.ru/technology\_and\_media/10/01/2024/659d37899a79473f8a99e35f**](https://www.rbc.ru/technology_and_media/10/01/2024/659d37899a79473f8a99e35f)

**Большие данные, большие проблемы: как защитить приватность в эпоху цифровых технологий?** [**https://www.garant.ru/article/1633362/**](https://www.garant.ru/article/1633362/)

**Тайна в переписке: эксперты нашли ключ к личным данным из ChatGPT. Популярный чат-бот может предоставлять гигабайты конфиденциальной информации:** [**https://iz.ru/1615227/dmitrii-bulgakov/taina-v-perepiske-eksperty-nashli-kliuch-k-lichnym-dannym-iz-chatgpt**](https://iz.ru/1615227/dmitrii-bulgakov/taina-v-perepiske-eksperty-nashli-kliuch-k-lichnym-dannym-iz-chatgpt)

**Дыханье сперли: c карт и счетов россиян украли огромную сумму денег. Технологии искусственного интеллекта служат мошенникам:** [**https://iz.ru/1540476/oksana-belkina/dykhane-sperli-c-kart-i-schetov-rossiian-ukrali-ogromnuiu-summu-deneg**](https://iz.ru/1540476/oksana-belkina/dykhane-sperli-c-kart-i-schetov-rossiian-ukrali-ogromnuiu-summu-deneg)

**Ты ли это: как уберечься от вооружившихся ИИ кибермошенников и вымогателей:**

[**https://www.forbes.ru/tekhnologii/496741-ty-li-eto-kak-uberec-sa-ot-vooruzivsihsa-ii-kibermosennikov-i-vymogatelej**](https://www.forbes.ru/tekhnologii/496741-ty-li-eto-kak-uberec-sa-ot-vooruzivsihsa-ii-kibermosennikov-i-vymogatelej)

**Машинный почерк: как мошенники используют ИИ для составления фишинговых писем:**

[**https://iz.ru/1596100/dmitrii-bulgakov/mashinnyi-pocherk-kak-moshenniki-ispolzuiut-ii-dlia-sostavleniia-fishingovykh-pisem**](https://iz.ru/1596100/dmitrii-bulgakov/mashinnyi-pocherk-kak-moshenniki-ispolzuiut-ii-dlia-sostavleniia-fishingovykh-pisem)

**«ИИ» и преступники:** [**https://rspectr.com/articles/nejroseti-prevrashhayutsya-v-prestupnikov**](https://rspectr.com/articles/nejroseti-prevrashhayutsya-v-prestupnikov)

## **Об определении «ИИ» (для размышления)**

Видео старое [**https://youtu.be/g2FiOz9FlTc**](https://youtu.be/g2FiOz9FlTc), но тезис, кажется, актуален: **человеческой интеллект характеризуется способностью принимать разумные решения при недостатке информации. Машинное обучение ровно про обратное — построить набор данных из миллионов примеров, обеспечить избыток информации и снизить неопределенность до минимума (то есть «механический процесс»). В этих условиях для принятия решения уже не нужен интеллект. Это обход проблемы отсутствия интеллекта, а не создание интеллекта! Еще об определении «ИИ»**

****

**ИСО сделала свой стандарт по терминам в сфере ИИ общедоступным**

Речь идёт о документе ISO/IEC 22989 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology (Информационные технологии. Искусственный интеллект. Концепции искусственного интеллекта и терминология). Скачать стандарт можно [**на сайте ИСО**](https://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_IEC_22989_2022_ed.1_id.74296_Publication_PDF_(en).zip)

[**https://t.me/robocounsel/603**](https://t.me/robocounsel/603) «Как нам избавиться от термина "искусственный интеллект"?» Стефано Квинтарелли [**отмечает**](https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami/), **что AI (artificial intelligence) — это заряженный термин, который искажает представления широкой публики о реальных возможностях систем и порождает множество в основном спекулятивных, ненужных или просто вредных ассоциаций — с сознанием, эмоциями, личностью и сверхчеловеческими способностями. Квинтарелли предлагает альтернативный термин: SALAMI — Systematic Approaches to Learning Algorithms and Machine Inferences. Я предлагаю близкий по смыслу аналог на русском: КОЛБАСА Комплексные Обучающиеся Лексические Боты, Алгоритмы и Системы Анализа**.

Это сразу позволяет под правильным углом посмотреть на обсуждение ряда популярных вопросов, таких как: «Способна ли КОЛБАСА испытывать эмоции? Имеет ли КОЛБАСА личность? Способна ли КОЛБАСА преодолеть естественные человеческие ограничения и превзойти человека? Может ли КОЛБАСА полюбить нас так, как мы её?».

**Кстати, «Неожиданные слабости искусственного интеллекта».** Какое-то время назад все обсуждали, как машинно-обученная модель АльфаГо научилась лучше человека играть в го. Её аналог на открытом коде, КатаГо, получил известность по той же причине в последнее время. Очередной бастион пал, говорили энтузиасты, теперь человеку остаётся лишь играть с другими людьми, потому что машины уже недосягаемы.

Но вот, недавно исследователям удалось создать стратегию, которая позволяет обыграть КатаГо [**https://arstechnica.com/information-technology/2022/11/new-go-playing-trick-defeats-world-class-go-ai-but-loses-to-human-amateurs/**](https://arstechnica.com/information-technology/2022/11/new-go-playing-trick-defeats-world-class-go-ai-but-loses-to-human-amateurs/), используя несовершенство процесса машинного обучения.

Никакой подобный процесс не идеален, и в модели всегда остаются какие-то пробелы из-за того, что наборы цифровых данных, использованные для обучения, не отражают всего богатства нашего аналогового мира.

Исследователи нащупали один из таких пробелов и успешно использовали его. Стратегия "на дурачка", легко распознаваемая и преодолеваемая человеческим игроком-любителем, успешно обманывает модель, выигрывающую у профессионалов.

Это лишний раз подчёркивает одну общую закономерность для систем искусственного интеллекта: да, многие из них научились выполнять некоторые задачи на уровне или лучше человека.

Но делают они это нехарактерным и непредвиденным для человека образом, словно какие-то инопланетяне.

Важно, что то же самое справедливо и для ошибок и сбоев таких систем. Машины "сыплются", иногда с катастрофическими последствиями, на непредсказуемо "глупых" мелочах, там, где человек даже средних способностей никогда бы не ошибся.

То же самое происходит с ЧатГПТ, когда система, в силу неспособности различить правду и ложь, внедряет ложные утверждения [**в случайных местах**](https://t.me/robocounsel/340)не в соответствии с человеческими паттернами обмана, в результате чего эту ложь оказывается сложно распознать неспециалисту.

А теперь представьте, что то же самое справедливо для автономного автомобиля. 99,999% времени он оказывается безопаснее человека за рулём. А потом группа злоумышленников, используя пробелы в наборах данных, на которых тренировали систему компьютерного зрения, моделирует на дороге такую ситуацию, которая гарантированно приводит к смертельному ДТП там, где распознать опасность мог любой начинающий водитель.

Системы ИИ — это созданные нами инопланетяне, и мы не должны надеяться, что наша человеческая интуиция нам как-то поможет решить проблемы, которые эти системы для нас создают.

**Очередной конфуз с использованием генеративных чатботов** [**https://t.me/robocounsel/646**](https://t.me/robocounsel/646)

**Большие языковые модели запоминают информацию, но не способны к её широкому обобщению**

[**https://t.me/robocounsel/653**](https://t.me/robocounsel/653)

**Путь к безопасному и надёжному «ИИ» может быть дольше, чем мы думаем!**

**Более трезвый взгляд, вместо слепого доверия («веры» даже)**

**Статья «Is it high time to take ChatGPT offline?» (нужен VPN для просмотра):**

[**https://www.linkedin.com/pulse/high-time-take-chatgpt-offline-aleksandr-tiulkanov/**](https://www.linkedin.com/pulse/high-time-take-chatgpt-offline-aleksandr-tiulkanov/) **о том, почему ChatGPT на сегодня не является надёжной ИИ-системой и почему результаты работы генеративной системы никогда нельзя использовать без тщательной перепроверки каждого слова специалистом в соответствующей предметной области.**

**Полагаться на результаты работы больших языковых моделей без постоянного контроля качества опасно.** Такой вывод можно сделать по результатам [**исследования**](https://arxiv.org/pdf/2307.09009.pdf)**,** проведённого учёными из Стэнфордского университета и Калифорнийского университета в Беркли (Chen, Lingjiao, Matei Zaharia, and James Zou. 2023. “How Is ChatGPT’s Behavior Changing over Time?”).

Сравнив результаты работы популярных моделей GPT-4 и GPT-3.5, эксперты обнаружили, что производительность каждой из них значительно снизилась в одних задачах и выросла в других.

Например, GPT-4 в марте 2023 г. на вопрос об определении простых чисел давала точный ответ с вероятностью 97,6%, но к июню 2023 г. количество точных ответов для задачи этого типа снизилось до 2,4%.

Причины этого феномена до конца не ясны, но полученные результаты свидетельствуют о необходимости постоянного контроля и перепроверки качества результатов модели со стороны разработчиков и пользователей таких моделей, в особенности тех, кто делает или планирует делать их элементом архитектуры собственного продукта или сервиса.

**Про демонстрацию GPT-4**

1. Мой совет остается в силе: если вы хотите использовать результаты работы чатбота в реальной жизни, вы должны тщательно проверить эти результаты на ошибки.

Вы также должны быть компетентны в соответствующей области знаний для такой проверки. Не субъективно компетентны в своём воображении или в силу раздутого чувства собственного достоинства. А объективно, как подтверждено дипломами, тестами, коллегами и людьми, которые платят вам деньги за работу в этой области. Если вам за работу в этой области не платят, вы не профессионал и, следовательно, не компетентны проверять результаты работы чатбота в этой области.

Вы также должны быть готовы взять на себя юридическую и моральную ответственность за любые пропущенные ошибки в результатах, которые вы получаете от чатбота и публикуете или используете где-либо за пределами чата.

2. Если честно, демонстрация налогового анализа, о которой все говорят, немного упрощена по сравнению с тем, что обычно представляет собой настоящий юридический анализ. Они скормили чатботу несколько страниц одного документа, в котором содержались все ответы, при этом было достаточно логики и буквального толкования текста нескольких страниц одного закона без анализа судебной практики.

В реальной жизни вчера мне пришлось просмотреть около 20 различных источников, определить, что нужно клиенту, какие акценты расставить и каких ключевых людей и организации мы должны привлечь. Я не мог ввести эти вводные в GPT-4, потому что это противоречило бы закону (незаконное раскрытие конфиденциальной информации неодобренному получателю (OpenAI) + незаконная обработка персональных данных субъектов данных в ЕС неуполномоченным разработчиком в США (OpenAI) в отсутствие соглашения об обработке данных + незаконная передача третьему лицу в США (OpenAI) в отсутствие адекватных гарантий, требуемых GDPR).

Я не говорю, что чатботы не могут быть полезны, я просто хочу сказать, что для впечатляющей демонстрации ОпенЭйАй нужно было бы учесть все юридические сложности выше и привести реальный пример, который потребует анализа не 10, а многих сотен страниц различных кодексов, законов и других нормативных актов. С определением, какие из них применимы, а какие нет.

И знаете что? Эти законы и НПА не идеально написаны и не всегда согласованы между собой. А также часто включают в себя ценностные категории типа "разумный" или "справедливый". И толкование: иногда оно должно быть буквальным, иногда телеологическим, иногда историческим и т.д… И вы должны понимать, почему и когда. Кроме того, нужно исследовать применимую судебную практику. И в результате всего этого анализа на выходе у вас часто будет не один, а несколько вполне обоснованных и защищаемых возможных ответов.

**Перечисленное выше, разумеется, сильно сужает реальные области применения системы. На сегодня более-менее реальны следующие:**

* обрисовка в общих чертах структуры и персонажей рассказов и других произведений, где не важна достоверность;
* конспектирование и суммирование тех текстов, которые у вас уже есть;
* интеллектуальный пинг-понг: можно покидаться в ChatGPT идеями и посмотреть на реакцию, она иногда интересная и полезная;
* первые проекты документов: если вы хорошо разбираетесь в какой-то предметной области, вы можете использовать ChatGPT для создания первого проекта нужного вам документа, и потом итеративно его улучшать – тут ключевой момент в том, что вы должны быть именно специалистом, а не любителем в том, о чём хотите написать, потому что вам придётся перепроверять за ChatGPT каждое слово.

Собственно, последний момент как раз показателен**. Как я уже говорил ранее, ChatGPT — эффективный генератор убедительного и завлекательного контента, но модель не умеет отделять правду от лжи, и то и дело вставляет в свои результаты ложные утверждения. Понять их ложность можно только будучи специалистом и (или) перепроверяя все, абсолютно все полученные тезисы. Поэтому еще раз в статье «Is it high time to take ChatGPT offline?» я объясняю, почему это более опасно, чем может показаться.**

И нет, я не предлагаю отключать ChatGPT — чем больше система находится в открытом тестировании, тем быстрее у разработчиков появятся шансы её улучшить. **Но безответственный хайп и некритическое использование ChatGPT могут быть в прямом смысле опасны для здоровья, а также для любого другого аспекта вашей жизни и жизней тех, кто некритически воспримет контент, который вы создадите и опубликуете при помощи ChatGPT…**

Пример того, когда проблема порождается не появлением новых технологий, а отсутствием культуры их применения **«Цена доверия нейросетям: покупателя «Ашана» чуть не отправили в тюрьму»:** [**https://www.it-world.ru/cionews/business/157710.html**](https://www.it-world.ru/cionews/business/157710.html)  В данном конкретном случае сотрудники магазина, кажется, *слепо доверились* оценке, которую дала система автоматического распознавания лиц. Как избегать подобных инцидентов? Компании — более строго регламентировать действия своих сотрудников при применении подобных систем? **Почему в цифровом мире нужны новые должностные инструкции:** <https://big-i.ru/management/upravlenie-izmeneniyami/pochemu-v-tsifrovom-mire-nuzhny-novye-dolzhnostnye-instruktsii/>

**Одна из серьезных проблем систем «ИИ» — их низкая достоверность. При этом их конструктивные особенности затрудняют одновременное повышение и достоверности, и убедительности (выразительности). Когда ГС выдаёт неправду — это убедительная неправда (вспомните ЧатГПТ). Улучшение же достоверности системы требует нелинейно большего количества данных для обучения, и даже при этом не гарантировано.**

**Генеративные чатботы не предназначены для поиска точной информации**

Дорогой ценой достался этот урок двум "прославившимся" на весь мир юристам, о которых я [**писал ранее**](https://t.me/robocounsel/507) и которые не проверили выдуманные ЧатГПТ ссылки на судебные прецеденты.

На слушании вчера, по итогам которого судья решит, что делать с незадачливыми копипастерами Шварцем и ЛоДукой дальше, было немало народу, и немало эмоций, [сообщает](https://archive.is/ZZLQe) Нью-Йорк Таймс:

"Боже, мне бы так хотелось это сделать (перепроверить выдуманные ЧатГПТ дела — А.Т.), но я этого не сделал", - сказал Шварц, добавив, что он чувствует смущение, унижение и глубоко раскаивается.

"Я не понимал, что ЧатГПТ может выдумывать прецеденты", - сказал он судье.

"Я услышал об этом новом сайте и ошибочно принял его за что-то вроде поисковой системы с возможностями нового, более высокого уровня", - приводит слова Шварца газета.

Уровень критического мышления у его коллеги, ЛоДуки, выступавшего в суде с материалами, подготовленными Шварцем, также оставляет желать лучшего.

"Вы читали какие-либо из приведенных дел?" спросил его судья.

"Нет", - ответил ЛоДука.

"Делали ли вы что-нибудь, чтобы убедиться в том, что дела существуют?".

Снова отрицательный ответ.

**Я уже давно рекомендую всем** [**всегда перепроверять все результаты**](https://t.me/robocounsel/349)**, полученные в ЧатГПТ, ввиду того, что система ненадёжна и постоянно допускает ошибки (хотя работа над этим и ведётся).**

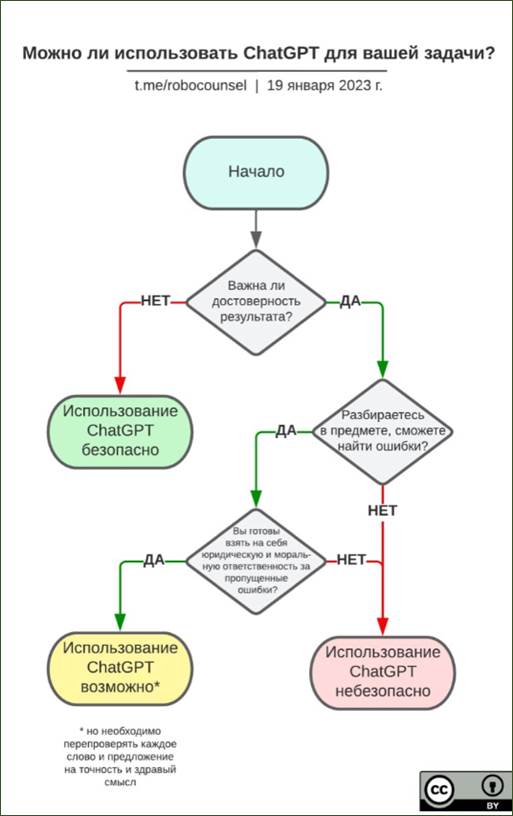
Но, например, по мнению Яна ЛеКуна, главного датасаентиста Меты, текущая архитектура популярных (так называемых авторегрессивных) больших языковых моделей конструктивно [обречена генерировать ошибочные результаты](https://twitter.com/ylecun/status/1640122342570336267) и не может быть принципиально исправлена — нужна новая архитектура.

В связи с этим наши с коллегами мнения разошлись: кто-то полагает, что Оупен-Эй-Ай делает недостаточно для того, чтобы предупредить пользователей об ошибках в генерируемых результатах. По коллег (Луиза Яровски), нужно обязать разработчиков сообщать пользователям, что они должны перепроверять полученные результаты.

**Я всех *пользователей* призываю именно к этому, но не до конца уверен в том, в какой именно форме было бы справедливо *требовать от разработчиков* информировать пользователей, ведь это снижает доверие к продукту.**

**С другой стороны, если продукт на самом деле пока и недостоин доверия, и *бoльшим риском* сейчас является избыточное доверие к ненадёжному продукту, то может быть это и правильно…**

**Об алгоритме оценки нужности использования рекомендательной системы**

**Схема, которая позволит понять нужно ли в задаче использовать генеративные модели**

**Для размышления**

**«ИИ» с небывалой прежде остротой ставит вопрос о природе человека. Современность, которая берет свое начало от Просвещения, построила собственную модель человека как существа прагматичного. Все многообразие его стремлений можно свести к тем или иным потребностям. Соответственно, все многообразие человеческой активности можно свести к попыткам эти потребности удовлетворить.**

**«ИИ» — прямое следствие подобного рассмотрения человека. Его следующий эволюционной скачок. «ИИ» куда эффективнее обычного человека, шустрей, сообразительней. И если мы будем продолжать измерять человека сугубо утилитарной и количественной меркой (KPI, IQ и проч.) — это значит, что ИИ» уже победил.**

**Чтобы не дать «ИИ» одержать над нами победу, нам следует вернуться к классической модели человека. Человека как искателя истины. А истина — это уже не то, что можно определить той или иной функцией; истина — не служебный инструментарий, направленный на получение чего-либо, как то утверждает современность. …**

**См. след. страницу**

**… Истина имеет смысл только сама по себе — и именно поэтому она не у каждого своя. Так же, как и красота, вера, добро, справедливость, знание.**

**Если мы примем такой характер истины, а вслед за ней — и соответствующий ей характер человечности, то у нас останется шанс если не на победу, то на сдерживание ИИ. Потому что он окажется усилением лишь одной из сторон человеческой природы, рассудочной и утилитарной, и да, в этом отношении он будет все равно совершеннее именно этой нашей стороны, но не совершеннее человека как такового. Так и техника — она сильнее человека, совершеннее его материальной стороны, но не совершеннее человека как такового.**

**Поэтому XXI век парадоксальным образом может стать веком анти-утилитаризма, веком возврата к классике, веком борьбы за подлинное достоинство человека, которое заключается в его богоподобии.**

**Иными словами, в будущем победит тот, кто уже сейчас читает Библию, Данте и Пушкина вместо бесконечных книг по личной эффективности и мотивации. Потому что последних ИИ прочитал не то что тысячи — сотни тысяч. А вот смысл Евангелия ему не схватить никогда, для него это будет лишь одна из бесчисленного множества других книг.**

**В противном случае, повторюсь, если мы и дальше будем упорно мерить человека прагмой, пользой и эффективностью, подстраивать под это образование и экономику, раскладывать его душу на биохимический состав, в общем, мерить его теми же критериями, которыми мы мерим роботов, — в таком случае человечество уже проиграло. Таким образом, быть прагматичным сегодня — значит мыслить и действовать анти-прагматично.**

## **Генеративные системы или «ИИ»: где стоит и где не стоит применять?**

**Статья Барака Туровского:** [**https://www.linkedin.com/pulse/framework-evaluating-generative-ai-use-cases-barak-turovsky**](https://www.linkedin.com/pulse/framework-evaluating-generative-ai-use-cases-barak-turovsky)(нужен VPN) **поможет с ответом на этот вопрос. Автор предлагает нам раскладывать сценарии применения генеративного ИИ на графике с 2 осями:**

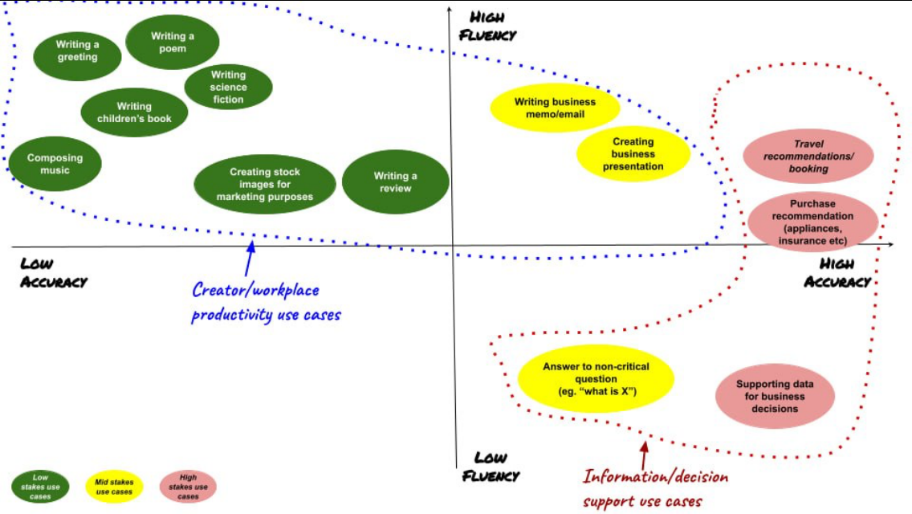
**1) требуемая убедительность и выразительность результатов (fluency);**

**2) требуемая достоверность результатов (accuracy).**

**Также значимы:**

**3) цена ошибки (зелёным отмечены менее, а красным — более рискованные сценарии);**

**4) масштабность применения (мультиплицирует риски**

****

**Федерация адвокатских палат Европы выпустила рекомендации по использованию ChatGPT**

[**Рекомендации**](https://www.fbe.org/wp-content/uploads/2023/06/European-lawyers-in-the-era-of-ChatGPT-FBE-Guidelines-on-how-lawyers-should-take-advantage-of-the-opportunities-offered-by-large-language-models-and-gene-kopia.pdf) организации (Fédération des Barreaux d’Europe, Страсбург) отражают основные соображения, которые я и другие эксперты по регулированию и применению ИИ высказывали на протяжении последнего полугодия:

**1. Разберитесь с тем, что такое генеративный ИИ**

Только разобравшись с тем, что представляет собой технология, каковы её возможности и правовые последствия её использования, вы сможете составить информированное мнение о том, в каких случаях и пределах её допустимо применять в юридической работе.

**2. Осознавайте ограничения и оптимальные условия использования**

Большие языковые модели и генеративный ИИ при умелом использовании могут повышать производительность юристов. Вместе с тем, необходимо чётко осознавать, для какого типа задач эти инструменты подходят лучше всего, для каких их лучше вообще не использовать, а где возможно использование с ограничениями. И каковы эти ограничения.

**3. Придерживайтесь установленных правил использования и лучших практик**

Изучите, имеются ли действующие правила и рекомендации вашей адвокатской палаты по этому вопросу. Определите, какие требования нормативных актов в вашей стране должны учитываться при применении генеративного ИИ. Изучите условия договора с поставщиком ИИ-услуг и ИИ-продуктов, которые вы используете. Будьте в курсе изменений.

**4. Использование систем ИИ может дополнять ваши знания и навыки, но** **не может их заменить**

Системы генеративного ИИ не способны к юридической аргументации, правовому анализу и не могут заменить вашу экспертизу, навыки и умения. Помните о рисках излишнего полагания на результаты работы систем генеративного ИИ. Эти системы не должны заменять человека в процессе приятия юридически значимых решений. Всегда самостоятельно проверяйте все предложенные системой результаты и аргументацию.

Объясняйте клиентам разницу между запросами к чатботу и настоящей юридической консультацией.

**5. Сохраняйте профессиональную тайну**

Принимайте меры для защиты информации и поддержания её конфиденциальности. Условия соглашения с поставщиком услуг генеративного ИИ могут предполагать последующее использование введённых вами данных, поэтому сообщение системе данных о клиенте [без их редактуры, анонимизации или псевдонимизации] может быть недопустимо.

**6. Не забывайте о требованиях, касающихся защиты персональных данных и неприкосновенности частной жизни**

Юристы должны осознавать, что ввод персональных данных в систему генеративного ИИ допускается только при наличии у вас на то законного основания и только при условии соблюдения норм о защите персональных данных и неприкосновенности частной жизни.

Не забывайте, что системы генеративного ИИ могут использовать введённые вами данные не только для ответа на запрос, но и могут сохранять их на будущее для усовершенствования таких систем.

Использование систем генеративного ИИ создаёт дополнительные риски для защиты данных по сравнению с уже имеющимися. Эти риски должны быть оценены и нейтрализованы.

**7. Информируйте клиентов и принимайте на себя ответственность**

Раскрывайте клиентам факты и условия использования систем генеративного ИИ в вашей работе. Объясняйте, для чего вы это делаете, в каких задачах применяете, в чём преимущества для клиента и какие меры предосторожности вы предпринимаете.

Всегда принимайте на себя всю полноту ответственности за результаты применения вами инструментов генеративного ИИ, осознавая свою возможную профессиональную и правовую ответственность за все последствия.

Установите правила для ответственного применения генеративного ИИ на рабочем месте, ознакомьте с ними коллег и сотрудников и обеспечьте соблюдение этих правил.

**Инструмент для оценки рисков при применении генеративного «ИИ» в ваших проектах**

[**https://t.me/robocounsel/626**](https://t.me/robocounsel/626)

**После увольнения 180 сотрудников BuzzFeed начинает использовать чат-бот ChatGPT:**

**https://www.ixbt.com/news/2023/01/31/180-buzzfeed-chatgpt.html**

**«ИИ» раскрыл профессии, которые отберет у людей:** [**https://m.gazeta.ru/tech/news/2023/03/16/19983211.shtml**](https://m.gazeta.ru/tech/news/2023/03/16/19983211.shtml)

**…Не все осознают долгосрочные последствия и не все видят общую закономерность. Заметьте общую тенденцию в том, как автоматизация меняет профессии. Моё первое интуитивное представление:**

* **нужен не переводчик, но редактор с опытом перевода;**
* **нужен не дизайнер, но арт-директор с опытом работы дизайнером;**
* **нужен не юрист, но руководитель юридической функции с опытом работы юристом.**

**С другой стороны, кажется при глобальном проникновении везде того же машинного перевода переводчики никуда не делись и не переименовались в редакторов.**

**Они просто научились пользоваться машинным переводом как инструментом. То же можно предположить и для дизайнеров с юристами.**

***Да, там (1) где людей перестало интересовать качество, или (2) когда люди сами владеют соответствующим языком в степени, достаточной для выявления машинных ляпов, они больше не привлекают переводчика.***

**То же, видимо, будет справедливо и для дизайнеров с юристами, к которым настоящая более-менее сносная автоматизация некоторых аспектов их работы постучалась не 10 лет назад, как к переводчикам, а только сейчас.**

**Осталось понять, как в новых условиях начинающие коллеги будут входить на рынок и делать первые шаги в профессии.**

Ведь все эти инструменты автоматизации работы, будь то дизайнерской или юридической, при применении не любительском, для качественного результата на выходе требуют жёсткого человеческого контроля со стороны специалиста с большим опытом.

Далее вопрос: эти навыки невозможно получить иначе, как в начале карьеры сделав всё это руками. Иначе просто руку не набьешь.

Как невозможно, всё время полагаясь на навигатор, научиться без него ориентироваться на улицах, или, в будущем, всё время полагаясь на автопилот, научиться водить.

Раньше человек шёл на первую работу за маленькую зарплату и потом рос. А теперь позиций с маленькой зарплатой будет меньше, потому что эту работу будет делать машина под надзором опытного старичка.

Что будут делать новички, не попавшие в корпоративную среду? Не исключено, что пойдут в народ, в Тиктоки и т.п., получать первый опыт в формате самозанятости. Наверное, тоже неплохо — вырастет доступность юридических услуг для населения.

**Новичку на рынке труда теперь критично важно знать, что умеет ГС в той человеческой деятельности, которую новичок выбрал для карьерного роста, работы**

**Дискуссия про навыки в мире искусственного интеллекта (ИИ) продолжается. Высказался нобелевский лауреат Писсаридес:** [**https://t.me/olegshibanov/1283**](https://t.me/olegshibanov/1283)

**Необходимо учитывать**

* **ЧатГПТ и подобные ему инструменты имеют потенциально весьма широкую, но с совсем другой стороны ограниченную область применения.**
* **Эти возможности и ограничения обусловлены архитектурой, техническими особенностями этих инструментов (см. стр. 18 – 23).**
* **Есть хорошее высказывание Джона Уорнера "ЧатГПТ не способен убить [в нашей культуре или работе] ничего такого, что было бы достойно спасения".**
* **Соответственно, если ваш труд сегодня состоит в чём-то, что полностью легко воспроизводимо посредством ЧатГПТ при помощи очевидных запросов (промтов), вам нужно срочно учиться зарабатывать на чём-то другом, потому что ваш труд вскоре перестанет быть актуален чуть более, чем полностью.**

**О будущем напряжении на рынке труда.** Ещё одно [**упражнение**](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350925) такого рода (февраль 2023). Кратко — по мнению автора, до 32% работ могут ощутить "очень большое влияние", вплоть до исчезновения. Важно то, что статья прошлась по описанию работ и смогла просуммировать, какие основные задачи люди на них выполняют. Вывод: ничего себе, треть!

**Созидательное разрушение: искусственный интеллект способен запустить переоценку труда и капитала:**

[**https://econs.online/articles/opinions/sozidatelnoe-razrushenie-iskusstvennyy-intellekt-sposoben-zapustit-pereotsenku-truda-i-kapitala/**](https://econs.online/articles/opinions/sozidatelnoe-razrushenie-iskusstvennyy-intellekt-sposoben-zapustit-pereotsenku-truda-i-kapitala/)

**Как может меняться компания под воздействием «ИИ»?** [**Статья**](https://www.nber.org/papers/w31325) (июнь 2023) смотрит на то, какие компании активно *используют* технологии искусственного интеллекта, и как это влияет на *свойства самой фирмы*. Идея в том, что привлечение сотрудников, умеющих в «ИИ», может снизить необходимость в менеджерах среднего звена. Вместо них могут более широко внедряться способы работы с данными.

Для этого авторы анализируют резюме сотрудников, а также объявления о позициях этих компаний. Период 2010-2018 с постепенным внедрением «ИИ».

Результаты впечатляют:

1) активнее берут на ИИ-связанные позиции в компаниях, которые уже наняли больше сотрудников с PhD и из STEM (наука, технологии, инженерные специальности, математика);

2) структура таких ИИ-активных компаний становится "более плоской". Это означает, что становится больше сотрудников на начальных позициях, меньше на средних и высоких, при этом сдвиг происходит в сторону более образованных людей.

**Вывод**: **мы ещё не видели полного эффекта влияния «ИИ» на рынок труда. Очевидно, что если ИИ становятся важнее, то компания обязана молодеть - потому что более взрослые люди не всегда знакомы с ИИ. Вероятно, что в ближайшее десятилетие подобная иерархическая структура не воспроизведётся, потому что все будут знать про ИИ, даже мы, взрослеющие люди.**

**«ИИ» не простит бизнесу невнимания:** [**https://guru.nes.ru/iskusstvennyij-intellekt-ne-prostit-biznesu-nevnimaniya.html**](https://guru.nes.ru/iskusstvennyij-intellekt-ne-prostit-biznesu-nevnimaniya.html)

**Экономические эффекты искусственного интеллекта:** [**https://guru.nes.ru/calendar/ekonomicheskie-effektyi-iskusstvennogo-intellekta.html**](https://guru.nes.ru/calendar/ekonomicheskie-effektyi-iskusstvennogo-intellekta.html)

**Что каждый CEO должен знать о генеративном искусственном интеллекте:**

[**https://habr.com/ru/articles/738820/**](https://habr.com/ru/articles/738820/)

**Оценена угроза замены офисных работников искусственным интеллектом.** Искусственный интеллект может полностью заменить выполнение до 25% офисных задач и частично – более 50%. Однако среди высокопрофессиональных сотрудников, руководителей и технических специалистов только небольшая часть задач считается подверженной воздействию ИИ.

Об этом [**сообщается**](https://rspectr.com/novosti/oczenena-ugroza-zameny-ofisnyh-rabotnikov-nejrosetyu?utm_source=telegram) в докладе «Генеративный искусственный интеллект и рабочие места», опубликованном Международной организацией труда (МОТ). Работники физического труда и те, кто занят вне офиса, практически не заметят изменений. МОТ отмечает, что наиболее вероятный исход – не замена офисных работников нейросетью, а оптимизация их работы за счет сотрудничества с искусственным интеллектом

**«Когнитивная автоматизация»: чем GPT может быть полезен экономистам:**

[**https://econs.online/articles/ekonomika/kognitivnaya-avtomatizatsiya-chem-gpt-mozhet-byt-polezen-ekonomistam/**](https://econs.online/articles/ekonomika/kognitivnaya-avtomatizatsiya-chem-gpt-mozhet-byt-polezen-ekonomistam/)

**Заменит ли «искусственный интеллект» экономистов:** [**https://guru.nes.ru/zamenit-li-iskusstvennyij-intellekt-ekonomistov.html**](https://guru.nes.ru/zamenit-li-iskusstvennyij-intellekt-ekonomistov.html)

**Где еще генеративные системы могут помогать?**

**Оценены перспективы внедрения «ИИ» в России:**

* [**https://rspectr.com/novosti/oczeneny-perspektivy-vnedreniya-ii-v-rossii**](https://rspectr.com/novosti/oczeneny-perspektivy-vnedreniya-ii-v-rossii)
* [**https://issek.hse.ru/news/862014108.html**](https://issek.hse.ru/news/862014108.html)

**Эксперты назвали 12 приоритетных для государства областей применения искусственного интеллекта:**

[**https://rdc.grfc.ru/2023/04/ai\_priority\_area\_application/**](https://rdc.grfc.ru/2023/04/ai_priority_area_application/)

**Финансы**

**GenAI в финансах: обзор рисков:** [**https://econs.online/articles/finansy/genai-v-finansakh-obzor-riskov**](https://econs.online/articles/finansy/genai-v-finansakh-obzor-riskov)

**The future of finance will include generative A.I., says an MIT researcher:**

[**https://fortune.com/2023/01/25/future-of-finance-generative-ai-mit-researcher/**](https://fortune.com/2023/01/25/future-of-finance-generative-ai-mit-researcher/)

**Применение «ИИ» на фин. рынке:**

[**https://www.cbr.ru/Content/Document/File/156061/Consultation\_Paper\_03112023.pdf**](https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf)

**Юридические услуги**

Роберт Амброджи протестировал новый сервис Ко-Каунсел — специализированный ИИ-ассистент для юристов.Сервис интегрирован с одной из больших языковых моделей Опен-Эй-Ай и на сегодня помогает решать семь задач.

1) Поиск по базе данных.

2) Проверка документов.

3) Резюмирование документов.

4) Проверка договоров на соответствие корпоративной политике организации.

5) Извлечение требуемой информации из договоров.

6) Помощь в подготовке правовых заключений.

7) Подготовка к опросу свидетелей по делу.

В частности, при работе в режиме проверки документов пользователь может загрузить их в систему и задать интересующие его вопросы по их содержанию, а система подготовит отчёт с ответами.

Также есть возможность свободных запросов по аналогии с ЧатГПТ, если пользователь сам переведёт систему в "небезопасный" режим. В этом случае система предупреждает пользователя о том, что полученные ответы могут быть недостоверными.

**Правки в договоры можно автоматизировать:** [**https://www.lawnext.com/2023/02/contracts-company-ironclad-taps-into-gpt-3-for-instant-document-redlining-based-on-a-companys-playbook.html**](https://www.lawnext.com/2023/02/contracts-company-ironclad-taps-into-gpt-3-for-instant-document-redlining-based-on-a-companys-playbook.html)

Айронклэд, компания, предоставляющая сервис по автоматизации договорной работы, сегодня анонсировала новую функцию (пока в бета-версии). Благодаря интеграции языковой модели GPT-3, сервис позволяет автоматически найти и подсветить в тексте договора фрагменты, отклоняющиеся от вашей договорной политики и ваших стандартных договорных положений. После этого сервис в режиме правки предложит изменения, которые вы можете принять или отклонить. Кроме того, по вашему запросу в произвольной форме, например, "сделай положения о конфиденциальности взаимными", сервис также сформулирует и внесёт в текст соответствующие правки, которые также можно принять или отклонить. Автоматизация договорной работы выходит на новый уровень. Очевидно, что распространение таких сервисов со временем позволит юристам-договорникам и их клиентам, и работодателям существенно сэкономить время и деньги, затрачиваемые сейчас по сути на эту механическую часть работы.

**Нейросеть стала консультировать по профессиональным вопросам**. Ученые Института точной механики и оптики (ИТМО) Санкт-Петербурга создали ИИ «Акела», которая учится на базах данных и проводит профессиональные консультации в разных отраслях. Разработка позволит проверять знания специалистов, проводя тестирование с валидацией ответов. Об этом [**сообщили**](https://rspectr.com/novosti/nejroset-stala-konsultirovat-po-professionalnym-voprosam) в пресс-службе ИТМО, отметив, что система поможет оптимизировать многие процессы, связанные с документооборотом на предприятиях.

**СМИ**

**СМИ станут более уязвимыми при использовании нейросетей:** [**https://rspectr.com/novosti/smi-stanut-bolee-uyazvimymi-pri-ispolzovanii-nejrosetej**](https://rspectr.com/novosti/smi-stanut-bolee-uyazvimymi-pri-ispolzovanii-nejrosetej)

**Информагентство Associated Press (AP) ограничило своих журналистов в использовании нейросетей для работы**: [**https://rspectr.com/novosti/informagentstvo-associated-press-ap-ogranichilo-svoih-zhurnalistov-v-ispolzovanii-nejrosetej-dlya-raboty**](https://rspectr.com/novosti/informagentstvo-associated-press-ap-ogranichilo-svoih-zhurnalistov-v-ispolzovanii-nejrosetej-dlya-raboty)

**Медицина**

[**https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf**](https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf)

**Сельское хозяйство**

**Тракторы с ИИ сэкономили более 5 млрд руб.:** [**https://rspectr.com/novosti/traktory-s-ii-sekonomili-bolee-5-mlrd-rublej**](https://rspectr.com/novosti/traktory-s-ii-sekonomili-bolee-5-mlrd-rublej)

**Возможности «ИИ»**

**Как освоить нейросети: 15 полезных инструкций:**

* [**https://journal.tinkoff.ru/short/neuro-to-use/**](https://journal.tinkoff.ru/short/neuro-to-use/)
* [**https://habr.com/ru/companies/bothub/articles/774390/**](https://habr.com/ru/companies/bothub/articles/774390/)

**Обзор ChatGPT с примерами использования в различных задачах:** [**https://habr.com/ru/post/708646/**](https://habr.com/ru/post/708646/)

**Чат-нейросеть ChatGPT ведет диалог, пишет код и сочиняет стихи: как попробовать самому:** [**https://journal.tinkoff.ru/chatgpt/**](https://journal.tinkoff.ru/chatgpt/)

**ChatGPT научился понимать изображения: 10 способов применять функцию в обычной жизни:**

[**https://journal.tinkoff.ru/list/chatgpt-vision**](https://journal.tinkoff.ru/list/chatgpt-vision)

**ChatGPT научился говорить и слушать: 7 сценариев применения функции в обычной жизни:** [**https://journal.tinkoff.ru/list/chatgpt-voice**](https://journal.tinkoff.ru/list/chatgpt-voice)

**С помощью «ИИ» сняли документальный фильм:** <https://rspectr.com/novosti/s-pomoshhyu-ii-snyali-dokumentalnyj-film>

**Проведено рандомизированное (случайным образом) исследование: полезен ли ЧатГПТ для «беловоротничковых» рабочих задач.** Новое исследование показывает, что использование инструмента может уменьшать дистанцию в производительности и качестве работы между наиболее неспособными и более способными работниками.

При этом эффект от применения ЧатГПТ для наиболее способных работников почти нулевой или отрицательный. При первичном применении наблюдался хайп-эффект с повышением удовлетворённости от применения ЧатГПТ.

**Вместе с тем, при повторном опросе по истечении двух недель наблюдалось снижение оценки полезности ЧатГПТ и снижение использования инструмента: им продолжали пользоваться лишь 33% попробовавших во время эксперимента и 18% из контрольной группы**. Исследователи полагают, что одним из факторов последующего разочарования оказывается то, что во время исследования испытуемые выполняли довольно простые и абстрактные задачи. Реальная же работа требует применения умений и знаний в контексте задач и проблем конкретного работодателя и конкретных клиентов, с чем ЧатГПТ помочь не может.

**Еще виды «нейронок»**

**Stable Doodle: От наброска к изображению.** Интересное на субботу: разработчики объединили Stable Diffusion XL (SDXL), модель, генерирующую изображения по текстовому описанию, с адаптером T2I от Tencent, распознающим и расшифровывающим наброски пользователя.

Распознав набросок, T2I-адаптер передает его расшифровку SDXL в понятном для неё формате, так что модель при генерации изображения руководствуется одновременно и контурами с наброска, и текстовой командой (промтом). В результате люди с самыми рудиментарными навыками рисования могут на основе своих набросков делать интересные иллюстрации — например, для своего блога, сайта или приложения. Порисовать можно [**здесь**](https://clipdrop.co/stable-doodle).

**Adobe Firefly Image 2: как бесплатно генерировать картинки в нейросети от создателей «Фотошопа»:** <https://journal.tinkoff.ru/adobe-firefly-image-2>

## **Что могут российские «нейронки»?**

**Россияне стали втрое активнее пользоваться нейросетями**

[**https://m.gazeta.ru/tech/news/2023/12/13/21910555.shtml**](https://m.gazeta.ru/tech/news/2023/12/13/21910555.shtml)

**Технологии доверия» (бывшие PwC) вместе с Knomary исследовали применение технологий искусственного интеллекта в управлении персоналом (HR) в России.** [**Отчет рассказывает о том**](https://t.me/antidigital/6897), **в каких функциях уже применяется ИИ, как генеративный, так и предиктивный.**

🔸Кейсы применения отличаются. В Авито ключевое применение – подбор сотрудников (они используют платформу AmazingHiring). В Северстали применяют генеративный ИИ – для написаний текстов, вопросов в интервью, должностных инструкций, а также для наполнения электронных курсов: «У нас была готовая структура, а также понимание того, что воспринимается пользователями и чем можно дополнить курс. Вся теория и все задания (тесты, интерактив в рамках курса) были написаны именно ИИ».

🔸Ростелеком и Сбер используют ИИ, в частности, для прогнозирования увольнения сотрудников по собственному желанию в ближайшее время. Это важно как для удержания ценных сотрудников, так и для минимизации затрат на тех сотрудников, которых компания не собирается удерживать. Ростелеком говорит, что точность платформы – 80%, у Сбера точность модели – 60-70%.

🔸Прикольное решение у Сбера, которое позволяет подыскивать сотрудника на должность не по навыкам, а по аналогу: «нужен сотрудник, по своему профилю похожий на Ивана Иванова». Эта функциональность нужна и в другом случае, если человек пошел на повышение и нужно найти преемника. «Наша цель – чтобы у каждого руководителя (от некоторого уровня) или ключевого сотрудника был преемник. Наша модель может найти подходящих преемников по всей компании».

🔸Определение должности по цифровым следам. «Система определяет, какие цифровые следы оставляет сотрудник – в какие системы заходит и как часто это делает, – и сопоставляет эти данные с его должностью. Например, мы можем выявить инженера-программиста, который основную часть времени тратит на тестирование. Если система видит несоответствие, она сообщает об этом руководителю. Это значительная помощь большим распределенным командам Сбера, которая в том числе хорошо работает и на удержание сотрудников».

Все кругом говорят о **ChatGPT**. Чат-бот американской компании **OpenAI** может написать резюме, разработать программу питания, написать код и даже поставить медицинский диагноз вашему коту. А что могут российские нейросети? Публикуем для вас подборку российских стартапов, размещенных на венчурном онлайн-хабе [**SberUnity**](https://sberunity.ru/?utm_source=vc&utm_medium=partner&utm_campaign=sberunity)**.**

**Появился российский конкурент ChatGPT GigаСhat:**

[**https://rspectr-com.turbopages.org/rspectr.com/s/novosti/poyavilsya-rossijskij-konkurent-chatgpt-gigashat**](https://rspectr-com.turbopages.org/rspectr.com/s/novosti/poyavilsya-rossijskij-konkurent-chatgpt-gigashat)

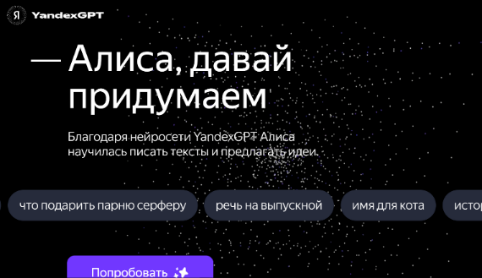
**Написать текст на 30 языках — ProGPT** [ProGPT](https://t.me/aidisraeli_bot) позволяет писать более 100 видов текстов без прерываний на 30 языках. Это продукт работает с более чем 10 нейронными сетями, четыре из которых полностью сделаны силами команды стартапа AI Disraeli.

**Проанализировать звонки, чаты, отзывы и NPS —** [**Deeray**](https://deeray.com/) Платформа на основе AI-технологии анализа естественного языка автоматически анализирует любые коммуникации между бизнесом и клиентами: как голосовые, так и текстовые.

**Отредактировать текст и подобрать синоним — ReText.AI** [**ReText.AI**](https://retext.ai/ru) — сервис перефразирования на основе нейросети с умным синонимайзером: в режиме онлайн он помогает повышать уникальность текстов совершенно бесплатно. Нейросеть работает с русским, белорусским, украинским и английским языками. Подобрать синонимы и преобразовать текст можно буквально за 7 секунд.

**Сделать прогноз спроса и выручки —** [**PredTech**](https://predtech.ru/) Облачное решение на базе машинного обучения для построения моделей прогнозирования спроса, выручки, товарных запасов и т.д. Помогает торговым и производственным компаниям увеличить прибыль.

**Найти нужный товар в интернет-магазине — SearchBooster** [**SearchBooster**](https://searchbooster.io/ru/) — это система smart-поиска для интернет-магазинов, помогающая клиентам быстрее находить требуемые товары, а также уменьшить количество действий на сайте. С помощью сервиса интернет-магазин может повысить средний чек и заработать лояльность покупателей.



**Как правильно работать с ожиданиями пользователей?** Если говорить, не о функционале, а о позиционировании чатбота от Яндекса и о "моральной подготовке" пользователей, кажется, что эти вещи у них получились гораздо ближе к идеалу, чем у создателей ЧатГПТ.

Сам заголовок "Давай придумаем", предупреждение о специфике работе — бот работает, "подражая текстам в интернете", "поэтому результат может быть выдумкой" — создаёт у пользователя очень правильные и более реалистичные ожидания.

Пользователь понимает, что перед ним не дельфийский оракул, от которого всегда следует ожидать истину в последней инстанции, а просто интересный инструмент. С которым можно поэкспериментировать, но который специализируется на содействии в *творческих задачах*, а не на поиске достоверной информации или на поддержке принятия жизненно важных личных или деловых решений.

А чей подход больше нравится вам? Чей бот лучше всего помогает пользователю сформировать правильные ожидания об ограничениях возможностей генеративного ИИ?  ЧатГПТ? Бинг? Бард? ЯндексГПТ? Сбер-Гигачат?

**Российские пользователи оценили релиз «Алисы» со встроенным YandexGPT:** [**https://t.me/ano\_dialog/1614**](https://t.me/ano_dialog/1614)

**YandexGPT теперь умеет кратко пересказывать статьи. Как это работает?**

**Принцип работы**

* Зайдите на сайт <https://300.ya.ru/> и вставьте ссылку на нужную статью.
* Нейросеть сократит текст до небольшого конспекта на 300-⁠1000 знаков, разобьет на пункты и выделит основные тезисы.

**Главное**

* YandexGPT пересказывает русскоязычные тексты объемом до 30 тысяч знаков.
* С иностранными статьями справляется медленнее и менее точно: какие-то материалы сразу переводит на русский язык, какие-то оставляет на английском.
* На составление тезисов уходит меньше минуты.Сервис работает в тестовом режиме.

**YandexGPT научилась создавать тексты в режиме реального времени:** [**https://nauka-tass-ru.turbopages.org/nauka.tass.ru/s/nauka/18149035**](https://nauka-tass-ru.turbopages.org/nauka.tass.ru/s/nauka/18149035)

**YandexGPT: что умеет нейросеть от «Яндекса» и как ей пользоваться:**

[**https://journal.tinkoff.ru/how-to-use-yandexgpt**](https://journal.tinkoff.ru/how-to-use-yandexgpt)

**«Яндекс» запустил новую версию своей нейросети — YandexGPT 2:**

[**https://journal.tinkoff.ru/news/yandexgpt-2**](https://journal.tinkoff.ru/news/yandexgpt-2)

**В "Яндексе" сравнили собственную нейросеть с ChatGPT 3.5:**

[**https://ria.ru/20230909/yandexgpt-1895147034.html**](https://ria.ru/20230909/yandexgpt-1895147034.html)

**GigaChat (Сбер): что умеет новая российская нейросеть**:

**https://trends.rbc.ru/trends/industry/6446a60c9a79470d569f8d3b**

**Лига безопасного интернета попросила нейросеть YandexGPT написать текст «5 главных правил безопасности в Интернете». Вот, что получилось.**

Интернет — это не только источник информации, но и угроза безопасности. Существует множество способов, которыми злоумышленники могут получить доступ к вашей личной информации и нанести вред вашему компьютеру или устройству. Поэтому важно соблюдать несколько простых правил, чтобы обеспечить свою безопасность в Интернете.

**1️.** **Используйте надежные пароли и регулярно их меняйте.** Никогда не используйте один и тот же пароль для нескольких аккаунтов. Лучше всего создавать сложные пароли, которые состоят из букв, цифр и специальных символов.

**2️.** **Не открывайте подозрительные ссылки и вложения.** Если вы получили электронное письмо или сообщение с просьбой предоставить личную информацию или перейти по ссылке, лучше всего проигнорировать его и не отвечать.

**3️.** **Используйте антивирусное программное обеспечение.** Установите на свой компьютер или устройство антивирусную программу и регулярно обновляйте ее. Это поможет защитить ваш компьютер от вредоносных программ и вирусов.

**4️.** **Не разглашайте личную информацию в социальных сетях.** Никогда не делитесь своими личными данными, такими как адрес, номер телефона, фото и т.д. в социальных сетях и других общедоступных местах.

**5️.** **Будьте осторожны при использовании публичных Wi-Fi сетей.** Если вам нужно подключиться к Wi-Fi в общественном месте, убедитесь, что вы используете защищенное соединение и не передавайте личную информацию.

Соблюдение этих простых правил поможет вам защитить свою личную информацию и обеспечить безопасность в Интернете в целом.

## **Регламентирование и регулирование «ИИ» в мире**

**Страны мира озаботились регулированием «ИИ»** [**https://www.kommersant.ru/doc/5938467?tg**](https://www.kommersant.ru/doc/5938467?tg)

**Зачем регулировать искусственный интеллект и как это затронет его развитие в России:**

[**https://www.forbes.ru/tekhnologii/489563-zacem-regulirovat-iskusstvennyj-intellekt-i-kak-eto-zatronet-ego-razvitie-v-rossii**](https://www.forbes.ru/tekhnologii/489563-zacem-regulirovat-iskusstvennyj-intellekt-i-kak-eto-zatronet-ego-razvitie-v-rossii)

**Регулирование ИИ в мире:** [**https://iqdecision.com/regulirovanie-ai-v-mire/**](https://iqdecision.com/regulirovanie-ai-v-mire/)

**G7 согласовала принципы для разработчиков искусственного интеллекта:** [**https://rspectr.com/novosti/bolshaya-semerka-utverdit-kodeks-dlya-ii**](https://rspectr.com/novosti/bolshaya-semerka-utverdit-kodeks-dlya-ii)

**Регулирование искусственного интеллекта в мировой практике:** [**https://ethics.cdto.ranepa.ru/3\_8**](https://ethics.cdto.ranepa.ru/3_8)

**Регулирование ИИ: обзор западных инициатив:** [**https://actualcomment.ru/regulirovanie-ii-obzor-zapadnykh-initsiativ.html**](https://actualcomment.ru/regulirovanie-ii-obzor-zapadnykh-initsiativ.html)

**Регулирование "новых технологий": нужно по секторам?** Аджемоглу не только выпустил полезную для обсуждения книгу "Power and Progress". Он также старается иллюстрировать отдельные кейсы моделями - это очень важно, чтобы понять "равновесные эффекты" и не останавливаться на отдельных рынках.

[**Статья**](https://www.nber.org/papers/w31461) (июль 2023) анализирует, как нужно регулировать новые технологии (искусственный интеллект), если не сразу проявляются и плюсы роста производительности, и минусы социальных последствий. Если ИИ меняет рынок труда и "увольняет" сотрудников компании, нужно придумать, как внедрять ИИ постепенно, подготовить людей и улучшить возможности работы в других направлениях. Роль регулятора может оказаться очень важной.

**Результаты**: социально оптимальное внедрение зависит от роста производительности при новых технологиях. Если производительность растёт примерно со скоростью увеличения социальных издержек, то внедрять надо постепенно - определяясь, как снижать эти издержки, и регулируя внедрение активнее. Если же производительность от технологии радикально выше социальных издержек, то внедрение нужно ускорять, и появляется вопрос налогообложения прибылей компаний.

**Вывод**: с «ИИ» может получиться интересно, потому что сейчас смахивает на большую производительность и одновременно потенциально высокие издержки – заменить хотят многих сотрудников, включая ИТ и аналитиков. Поэтому роль регулирования "по Аджемоглу" огромная.

**Аудит и оценка соответствия ИИ-систем: в чём разница?**

Увидел мнение о том, что проект Регламента ЕС по ИИ предполагает аудит ИИ-систем. Это не так. Аудита в проекте Регламента пока нет. Есть — только для систем повышенной опасности — [**оценка соответствия**](https://fpf.org/blog/introduction-to-the-conformity-assessment-under-the-draft-eu-ai-act-and-how-it-compares-to-dpias/), притом в основном собственными силами провайдера и только в редких случаях с подтверждением соответствия со стороны нотифицированного органа. Очень важно не путать смежные, но разные понятия.

Что предполагает [**настоящий аудит**](https://forhumanity.center/web/wp-content/uploads/2021/09/ForHumanity.center_Taxonomy_AI_Audit_Assurance_Assessment.pdf) ИИ-систем (как сейчас в сфере финансов для публичных компаний) и чего нет в процедуре оценки соответствия (которая сейчас предусмотрена Регламентом ЕС по ИИ):

* выполнение оценки сертифицированными специалистами;
* объективные, не субъективные критерии оценки;
* строгая независимость оценивающего (запрет на иные вызывающие конфликт интересов взаимодействия, например, запрет на консалтинг для того же клиента в течение определённого срока), обычно установлена нормативно, см. акт Сарбейнс-Оксли;
* критерии оценки бинарные (соответствие/несоответствие) и устанавливаются не проверяющим, а независимо (например, профессиональной ассоциацией);
* выгодоприобретателем оценки является общество и пользователи отчёта об оценке, а не его заказчик;
* для аудитора недопустима обратная связь с проверяемым в процессе оценки, итеративное решение обнаруживаемых проблем заказчика, его обучение тому "как надо";
* аудитор несёт ответственность за ложные заверения о соответствии;
* конечной целью аудита является публикация общедоступного отчёта об оценке.

Итак, Регламент ЕС об «ИИ» пока не предполагает независимого аудита ИИ-систем, хотя призывы к таким аудитам есть, в особенности в отношении систем, которые оказывают наибольшее влияние на жизнь и права граждан.

Так, за независимые аудиты своих и подобных им систем [**выступает**](https://openai.com/blog/planning-for-agi-and-beyond) лидер рынка генеративного ИИ OpenAI.

**На мой взгляд, со временем мы глобально придём к таким аудитам для наиболее значимых ИИ-систем, точно так же, как это произошло для публичных финансов. Это крайне важно для защиты общества и прав человека.**

**Напомню, что первой юрисдикцией, в которой такие аудиты уже требуются по закону (начиная с грядущего лета), стали, как ни странно, США. Но не на федеральном уровне, а пока только на муниципальном —** [**в Нью-Йорке**](https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&amp;GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&amp;Options=ID%7CText%7C&amp;Search=)**, и пока только в части систем, используемых для найма на работу и оценки персонала.**

Однако это как раз одна из сфер применения «ИИ», которая наиболее чувствительна для прав граждан, поэтому служит хорошим знаком для дальнейшего расширения независимых аудитов ИИ-систем — в интересах общества.

Ну и еще раз ***о «возможностях» генеративных систем***, основанных на коде/синтаксисе и вероятностном поиске слов, которые потом «ИИ» вставит в свой ответ, из ранее созданной людьми информации (любой, оцифрованной), ***в образовании молодых.***

Итак, уже школьники матерится лучше любых сидельцев, и в массе своей все хуже и хуже у них умственные возможности. А теперь, если говорить о внешнем, из-за желания быть в тренде и на хайпе, если же о глубинных, то из-за инфантильности и лени, помноженных на клиповое и слаборазвитое мышление/память, что есть результат работы безбожных реформаторов, молодые поголовно будут «юзать» генеративные системы — профанируя реальное развитие себя!

**А эти генеративные/дегенеративные системы, сами алгоритмизированы на текстах, написанных западными «ученными», которые разделяют богоборческие уклады: с культивированием греха, с идеями контроля деторождения от Мальтуса, только которые уже ничем не ограничиваются для достижения результата (см. стр.3 – 5 и «Проблему 2»).**

**Этический компас пользователя технологий (доп. см. стр. 37-38)**

1. Технология, которую вы планируете применить, пригодна для поставленной задачи? Чем вы это докажете?

2. Использование вами этой технологии законно и соответствует общепринятым этическим нормам? Чем вы это докажете?

3. Какие шаги вы планируете предпринять и задокументировать, чтобы вас нельзя было упрекнуть в безответственности?

4. Будет ли вам комфортно поделиться всей подноготной по поводу того, как именно работает эта технология и как вы её используете? Если завтра об этом в деталях расскажут СМИ, вы будете этим гордиться?

5. Учитывая все обстоятельства, использование этой технологии — всё ещё лучший способ решить поставленную задачу? Или лучше будет решить задачу другим способом, в том числе полностью отказаться от использования этой технологии?

**Россия**

**В России обновят стратегию развития ИИ**

[**https://ria.ru/20231122/ii-1910984401.html**](https://ria.ru/20231122/ii-1910984401.html)

**Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 24 г.»:**

[**https://www.garant.ru/products/ipo/prime/doc/74460628/#review**](https://www.garant.ru/products/ipo/prime/doc/74460628/#review)

Посмотрите, дополнительно: [**https://ria.ru/20230424/ii-1867306770.html**](https://ria.ru/20230424/ii-1867306770.html)

**Как в России будут регулировать «ИИ»**

[**https://pro.rbc.ru/demo/5f7193589a79472fb0b1213d**](https://pro.rbc.ru/demo/5f7193589a79472fb0b1213d)

**Эксперты оценили экономический эффект от применения ИИ**

[**https://rspectr.com/novosti/eksperty-oczenili-ekonomicheskij-effekt-ot-primeneniya-ii**](https://rspectr.com/novosti/eksperty-oczenili-ekonomicheskij-effekt-ot-primeneniya-ii)

**Искусственному интеллекту определяют правила игры**

[**https://www.kommersant.ru/doc/6283667?tg**](https://www.kommersant.ru/doc/6283667?tg)

**Правительство разработает закон об искусственном интеллекте**

**https://www.rbc.ru/rbcfreenews/649b950c9a794732a27a407b**

**В России начнут контролировать деятельность разработчиков искусственного интеллекта:**

* [**https://m.gazeta.ru/tech/news/2023/05/31/20563070.shtml**](https://m.gazeta.ru/tech/news/2023/05/31/20563070.shtml)
* [**https://www.pnp.ru/economics/v-rossii-vozmut-pod-kontrol-razrabotchikov-iskusstvennogo-intellekta.html**](https://www.pnp.ru/economics/v-rossii-vozmut-pod-kontrol-razrabotchikov-iskusstvennogo-intellekta.html)

**Опубликовано три проекта ГОСТов в области Big Data**

[**https://rspectr-com.turbopages.org/rspectr.com/s/novosti/opublikovano-tri-proekta-gostov-v-oblasti-big-data**](https://rspectr-com.turbopages.org/rspectr.com/s/novosti/opublikovano-tri-proekta-gostov-v-oblasti-big-data)

**Минцифры определит стандарт персональных данных для искусственного интеллекта**

[**https://www.vedomosti.ru/technology/articles/2023/06/27/982429-mintsifri-opredelit-standart-personalnih-dannih**](https://www.vedomosti.ru/technology/articles/2023/06/27/982429-mintsifri-opredelit-standart-personalnih-dannih)

**Минцифры создаст центр обезличивания данных для обучения ИИ**

* [**https://1prime.ru/telecommunications\_and\_technologies/20230628/840940683.html**](https://1prime.ru/telecommunications_and_technologies/20230628/840940683.html)
* [**https://iz.ru/1536931/dmitrii-bulgakov/drugim-imenem-v-rossii-sozdadut-tcentr-obezlichivaniia-dannykh-dlia-podgotovki-ii**](https://iz.ru/1536931/dmitrii-bulgakov/drugim-imenem-v-rossii-sozdadut-tcentr-obezlichivaniia-dannykh-dlia-podgotovki-ii)
* [**https://digital.gov.ru/ru/events/45880/?utm\_referrer=https%3a%2f%2fyandex.ru%2f**](https://digital.gov.ru/ru/events/45880/?utm_referrer=https%3a%2f%2fyandex.ru%2f)

**В России создадут оперативный штаб по искусственному интеллекту**

[**https://iz.ru/1545353/2023-07-17/v-rossii-sozdadut-operativnyi-shtab-po-iskusstvennomu-intellektu**](https://iz.ru/1545353/2023-07-17/v-rossii-sozdadut-operativnyi-shtab-po-iskusstvennomu-intellektu)

**В министерствах появятся рабочие группы по «ИИ»**

[**https://rspectr.com/novosti/v-ministerstvah-poyavyatsya-rabochie-gruppy-po-ii**](https://rspectr.com/novosti/v-ministerstvah-poyavyatsya-rabochie-gruppy-po-ii)

**Разрабатываемый в течение последних трех лет законопроект о регулировании рекомендательных сервисов онлайн-платформ** [**внесен в Госдуму.**](https://www.kommersant.ru/doc/6067220?tg) **В соответствии с ним соцсети, видеохостинги, онлайн-кинотеатры, поисковики и маркетплейсы не должны допускать использования рекомендательных алгоритмов для распространения запрещенной в России информации.** В случае нарушения Роскомнадзор может потребовать у сервисов доступ к работе соответствующего софта для его проверки, а также инициировать блокировку ресурса. Инициатива может затронуть в первую очередь зарубежные платформы, которые не исполняют требования по удалению противоправной информации.

Мы видим, к чему приводит неконтролируемое применение рекомендательных алгоритмов крупными цифровыми платформами. Исследователи фиксируют факты манипуляции: и в [**маркетинговых,**](https://t.me/webstrangler/2405) и в [**политических**](https://t.me/webstrangler/2579)целях. Психологи диагностируют у своих пациентов нервные расстройства, причиной которых становится так называемая [**«алгоритмическая тревожность».**](https://t.me/webstrangler/2260) Именно поэтому тренд на регулирование рекомендательных технологий стал общемировым: все чаще законодатели разных стран говорят о необходимости регламентировать эту сферу.

Мой законопроект обязывает владельцев ресурсов, которые используют рекомендательные алгоритмы, не просто информировать об этом своих пользователей, а довести до них основные правила их работы. Они должны содержать описание механизма рекомендаций и сведений, которые он собирает и использует, с указанием источника их получения.

Сегодня граждане подвергаются дискриминации со стороны цифровых платформ, применяя рекомендательные алгоритмы не столько для повышения пользовательского опыта, сколько для сбора данных, которые потом используются в самых разных целях (в том числе становятся предметом купли-продажи). В окончательной редакции законопроект призван ликвидировать эту дискриминацию.

Следить за соблюдением прав граждан будет Роскомнадзор: документ предусматривает инструментарий для воздействия на ресурсы, которые при помощи рекомендаций вводят в заблуждение, распространяют фейки или еще каким-то образом нарушают российское законодательство, – вплоть до блокировки. Подобные злоупотребления должны пресекаться, мы не можем допустить, чтобы алгоритмы были оружием информационной войны. Планируется, что регулятор получит возможности для воздействия на тех, кто при помощи рекомендательных алгоритмов умышленно нарушает законодательство РФ и не реагирует на предупреждения.

Также рассчитываю, что законопроект заставит разработчиков более ответственно относиться к применению ИИ-технологий, которые также используются для формирования рекомендаций. По сути, это будет первый российский законопроект, направленный на регулирование продукта, созданного искусственным интеллектом.

Во время разработки законопроекта мы провели несколько встреч с российскими ИТ-компаниями в «Институте развития интернета» и в «Фонде развития интернет-инициатив», собрали мнения и пожелания экспертов индустрии. Также на законопроект получено положительное заключение Правительства РФ, и перед внесением мы постарались учесть все замечания по существу.

**В пресс-центре ТАСС сегодня обсуждали новые подходы к регулированию искусственного интеллекта** [**https://t.me/webstrangler/2812**](https://t.me/webstrangler/2812)**.** Мой закон о рекомендательных сервисах – яркий пример так называемого «гибридного подхода», когда законодатель четко обозначает рамки, при этом оставляя большое поле для саморегулирования. В случае с рекомендациями мы предусмотрели две ключевые обязанности: предупреждать людей о том, что сервис использует рекомендательную модель, и описывать правила её работы. Как сформулировать эти правила, мы осознанно не говорим: владельцы ресурсов должны сделать это, исходя из собственных этических представлений.

Например, компания «Яндекс», насколько мне известно, собирается выкатить два варианта: подробный юридически выверенный текст плюс упрощенное объяснение в формате карточек. Надеюсь, что другие компании тоже пойдут по пути человеколюбия и максимально доступно сформулируют правила работы своих рексервисов. На данный момент, правда, никто этого не сделал – видимо, отрасль ждет, когда необходимые разъяснения опубликует РКН (притом что до вступления закона в силу остается всего две недели).

**Поэтому самое время воспользоваться теми возможностями, которые дает гибридное регулирование, и взять на вооружение** [**документ,**](https://ethics.a-ai.ru/assets/ethics_documents/2023/09/19/Recommendation_Services_Ethics_01.pdf) **подготовленный Альянсом в сфере искусственного интеллекта.** Это свод этических правил по применению рекомендательных сервисов (или, как сегодня кто-то скаламбурил, «рекомендации по рекомендациям»). Они отлично дополняют букву закона необходимой терминологией, перечнем добровольных обязанностей и принципов прозрачности. Очень надеюсь, что наличие этого свода исключит необходимость вводить административку, о которой журналисты [**спросили**](https://tass.ru/obschestvo/18784139)сегодня первым делом.

**Свыше 60% юристов поддержали законодательное регулирование «ИИ» в России. По мнению опрошенных специалистов, самыми популярными проблемами** [**могут стать:**](https://tass.ru/ekonomika/18158273)

* споры из-за нарушений авторских прав (77,1% респондентов);
* репутационные риски из-за некачественного предоставления услуг ИИ (50,8%);
* раскрытие конфиденциальной информации и зависимость от технологии, которая может привести к снижению профессионализма и компетенции специалистов (40,3%).

**«Прозрачные» рекомендации**. **Рекомендательные системы глубоко проникли в нашу жизнь, каждый день предлагая новости, рекламу, музыку, фильмы, «друзей» и «подписки», товары и услуги в соцсетях и интернет-сервисах. Эти алгоритмы не всегда используют во благо.**

1. Алгоритм, нацеленный на распространение контента с максимальной вовлеченностью, может начать продвигать фейковые новости и риторику ненависти.
2. Из-за предложения «друзей» и «подписок», похожих друг на друга, человек может оказаться в «эхо-камере» и перестать критически воспринимать информацию.
3. Манипулируя настройками рекомендательных алгоритмов, разработчики способны воздействовать на мнение людей по социально значимым вопросам **(примеры про американский ChatGPT см. стр. 3-5).**

**Эксперты Научно-технического центра ФГУП «ГРЧЦ» обобщили опыт регулирования рынка рекомендательных систем в Китае, ЕС и США и пришли к выводам:**

* регулирование рекомендательных сервисов необходимо, чтобы защитить людей от непрозрачных алгоритмов отбора контента, способных искажать действительность и приводить к пугающим последствиям, вплоть до реального нанесения вреда здоровью и психике человека;
* защита персональных данных — ключевой аспект регулирования рекомендательных систем. Пользователь должен быть проинформирован о работе рекомендательного алгоритма и иметь возможность отозвать согласие на обработку персональных данных.

**Подробнее —** [**https://rdc.grfc.ru/2023/05/recommendation\_services\_methods\_and\_regulation/**](https://rdc.grfc.ru/2023/05/recommendation_services_methods_and_regulation/)

**Онлайн-площадки РФ начали раскрывать принципы работы своих рекомендательных алгоритмов:** <https://habr.com/ru/news/764976/>

**Какие рекомендации в интернете чаще всего влияют на потребителей.** Наибольшее воздействие рекомендательные алгоритмы [**оказывают**](https://rspectr.com/infographics/kakie-rekomendaczii-v-internete-chashhe-vsego-vliyayut-na-potrebitelej?utm_source=telegram)на пользователей в сегменте новостной информации и развлечений. **Заметную зависимость от систем потребители ощутили при выборе товаров, услуг, при поиске советов о здоровье.**

**Рекомендуем всем интересующимся к прочтению** «**Основные правовые системы ограничения в доступе к информации в Российской Федерации» А.Фатьянова:** [**https://www.litres.ru/book/aleksey-aleksandrovi/osnovnye-pravovye-sistemy-ogranicheniya-v-dostupe-k-i-67081357/**](https://www.litres.ru/book/aleksey-aleksandrovi/osnovnye-pravovye-sistemy-ogranicheniya-v-dostupe-k-i-67081357/)

**Ключевые вопросы правового регулирования и развития ИИ в ближайшей перспективе:** [**https://fparf.ru/news/fpa/klyuchevye-voprosy-pravovogo-regulirovaniya-i-razvitiya-ii-v-blizhayshey-perspektive/**](https://fparf.ru/news/fpa/klyuchevye-voprosy-pravovogo-regulirovaniya-i-razvitiya-ii-v-blizhayshey-perspektive/)

**Презентация о требованиях к локализации персональных данных:** [**https://rppa.ru/\_media/analitika/242\_fz.pdf**](https://rppa.ru/_media/analitika/242_fz.pdf)

**Ответственность за нарушение закона о персональных данных:** [**https://www.garant.ru/actual/persona/otvetstvennost/**](https://www.garant.ru/actual/persona/otvetstvennost/)

**ЕС**

**ЕС согласовал первый в мире закон о регулировании «ИИ» (ожидается вступление в 25 г.):** [**https://www.kommersant.ru/doc/6394442**](https://www.kommersant.ru/doc/6394442) **и** [**https://www.rbc.ru/technology\_and\_media/09/12/2023/65748dc89a79475b5504174e**](https://www.rbc.ru/technology_and_media/09/12/2023/65748dc89a79475b5504174e)

**Регламент ЕС об ИИ:** [**https://rdc.grfc.ru/2021/10/ai-standards/**](https://rdc.grfc.ru/2021/10/ai-standards/)

**Европарламентарии урегулировали основные разногласия по проекту Регламента ЕС об «ИИ»:**

Соглашение [**было достигнуто**](https://www.euractiv.com/section/artificial-intelligence/news/meps-seal-the-deal-on-artificial-intelligence-act/)на прошлой неделе и предполагает, что существенные новые поправки до пленарного голосования в середине июня больше вноситься не будут.

Исключением стал вопрос о возможном запрете применения в ЕС технологий биометрической идентификации\* в общественных местах для охраны правопорядка. По предложению Еврокомиссии такой запрет должен был распространяться только на идентификацию в реальном времени, но депутаты левых фракций выступили за полный запрет. Против запретов возражают представители правых партий и правоохранительных органов.

В то же время, достигнут консенсус по поводу запрета на применение ИИ-систем для целенаправленного манипулирования поведением, а также для распознавания эмоций на рабочем месте, в образовании, в сфере контроля миграции и в правоохранительной деятельности.

Также будет запрещено применение предсказательных моделей для предотвращения уголовных и административных правонарушений (predictive policing).

Специальные нормы будут установлены для оборота систем ИИ общего назначения и базовых моделей (foundation models) наподобие GPT-4.

Вместе с тем, сужено понятие систем ИИ повышенной опасности: теперь таковыми будут признаваться не все системы, применяемые в перечисленных в Приложении III областях (обслуживание критической инфраструктуры, правоохранительная деятельность, трудовые отношения, образование и т. д.), а только создающие при этом значительную угрозу причинения вреда здоровью, безопасности, правам человека или окружающей среде.

Также — по случайной аналогии с положениями проекта Цифрового кодекса Киргизии — введена новая статья с общими принципами, которые должны будут учитываться при разработке систем и технических стандартов для них: - подконтрольность систем ИИ человеку, - техническая надёжность (робастность) и безопасность, - неприкосновенность частной жизни и управление данными, - открытость (transparency), - социальное и экологическое благополучие, - социальное многообразие (diversity),  - предотвращение дискриминации, - справедливость (fairness).

\* Напомню, что под идентификацией понимается установление личности по базе данных — поиск сходства *одного* взятого биометрического образца *со множеством* других — в основном используется в правоохранительной деятельности.

**Еврокомиссия выпустила для своих сотрудников руководство по использованию в работе систем искусственного интеллекта. Документ призван помочь сотрудникам в управлении рисками и ограничениями при работе с системами ИИ, такими как ChatGPT, Bard и Stable Diffusion.**

Комиссия отметила, что эти инструменты потенциально могут повысить эффективность и улучшить качество офисной работы. Однако такое использование требует соблюдения ряда условий.

Первый риск — раскрытие конфиденциальной информации или персональных данных общественности. Любые входные данные, предоставляемые для онлайн-модели генерирующего ИИ, затем передаются поставщику ИИ и впоследствии могут использоваться для ответов другим пользователям системы.

Поэтому сотрудникам ЕС запрещено делиться с ИИ «любой информацией, которая еще не является общественным достоянием, а также личными данными».

Второй момент касается недостатков модели искусственного интеллекта — можно получить неправильные или предвзятые ответы из-за особенностей алгоритмов ИИ.

Из этого следует, что чиновники ЕС «всегда должны критически оценивать любой ответ, полученный с помощью доступной в режиме онлайн модели генерирующего искусственного интеллекта, на предмет потенциальных предубеждений и фактически неточной информации».

Еще один риск — нарушение прав интеллектуальной собственности, поскольку модель ИИ может использовать для обучения защищенный контент, а получить в этом случае необходимое разрешение от правообладателей практически невозможно.

Регламент предписывает чиновникам «никогда напрямую не воспроизводить результаты генерирующей модели искусственного интеллекта в публичных документах, таких как тексты Комиссии, особенно юридически обязательные».

Сотрудникам Комиссии запрещено также полагаться на эти инструменты для выполнения критических и требующих больших затрат времени задач.

Идентификацию не следует путать с верификацией — популярным способом удостоверения личности путём сопоставления *одного* взятого биометрического образца с *одним* ранее взятым другим — например, при разблокировке телефона по отпечатку пальца.

**Китай**

**В Китае предложили регулировать сервисы генеративного искусственного интеллекта:** [**https://tass.ru/ekonomika/17493481**](https://tass.ru/ekonomika/17493481)

**Разные подходы Китая и США к ИИ**

[**https://dzen.ru/a/ZLZfJVX\_VSvi4Vy-**](https://dzen.ru/a/ZLZfJVX_VSvi4Vy-)

**Китайское правительство** [**опубликовало директиву**](https://www.securitylab.ru/news/540959.php), **которая устанавливает 24 правила для компаний, предоставляющих услуги и контент с использованием искусственного интеллекта (ИИ). Документ вступил в силу с 15 августа 2023 года и направлен на обеспечение безопасности, этики и соответствия социалистическим ценностям Китая. Согласно директиве, компании, работающие с ИИ, должны:**

* маркировать контент, созданный с помощью ИИ, ярким ярлыком;
* использовать только законные данные для обучения моделей ИИ;
* предоставить четкий механизм рассмотрения жалоб на услуги и контент, созданный ИИ;
* не подстрекать к подрыву государственной власти, свержению социалистической системы, нанесению ущерба национальному имиджу и сепаратизму.
* Правила будут действовать только для технологий, доступных пользователям, а не для разработок научных институтов. За надзор за соблюдением правил будут отвечать семь ведомств, в том числе Управление киберпространства Китая и Национальная комиссия по развитию и реформам.

**США**

**Президент США подписал указ о мерах по обеспечению безопасности, защищённости и надёжности систем искусственного интеллекта:**

* [**https://t.me/robocounsel/623**](https://t.me/robocounsel/623)
* [**https://t.me/robocounsel/624**](https://t.me/robocounsel/624)

**OpenAI согласна с необходимостью законодательно регулировать развитие ИИ:** [**https://habr.com/ru/news/735604/**](https://habr.com/ru/news/735604/)

**США могут решиться на федеральное регулирование ИИ.** Администрация Байдена [**начала консультации**](https://www.reuters.com/technology/us-begins-study-possible-rules-regulate-ai-like-chatgpt-2023-04-11/) по вопросу возможного регулирования на федеральном уровне для систем искусственного интеллекта, сообщает Рейтер. Напомню, что ранее Администрация полагалась на добровольные стандарты и саморегулирование.Означает ли это возможную поддержку исполнительной властью затихшей парламентской инициативы по Algorithmic Accountability Act 2019 года (переизложенной в 2022 году), либо Администрация думает о собственном проекте нормативного акта, пока не ясно. Поводом для возможной смены позиции, вероятно, стали новые скандалы и озабоченности вокруг недавно взлетевших в популярности новых продуктов генеративного ИИ общего назначения. Дополнительно: [**https://www.rbc.ru/politics/05/05/2023/645444e29a794767d54fa8dd**](https://www.rbc.ru/politics/05/05/2023/645444e29a794767d54fa8dd)

**Больше прозрачности, меньше рисков: политики в США взялись за искусственный интеллект:** [**https://tass.ru/mezhdunarodnaya-panorama/17773397**](https://tass.ru/mezhdunarodnaya-panorama/17773397)

**Семь ведущих ИИ-компаний США взяли на себя обязательства по ответственной разработке.** Администрация Президента США [**объявила сегодня**](https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/), что получила такие обязательства от Амазон, Антропик, Гугл, Инфлекшн, Мета, Майкрософт и Оупен-Эй-Ай.Компании обязались:

* **для обеспечения безопасности продуктов перед их выпуском:** **-** **проводить внутреннее и внешнее тестирование систем искусственного интеллекта на безопасность до их выпуска.** Это тестирование, которое будет частично проводиться независимыми экспертами, позволит защитить от некоторых наиболее значимых источников рисков ИИ, таких как биобезопасность и кибербезопасность, а также более широко от негативных социальных эффектов; **- обмениваться информацией в рамках всей отрасли, а также с правительствами, гражданским обществом и научными кругами по вопросам управления рисками ИИ.** Это включает в себя лучшие практики обеспечения безопасности, информацию о попытках обойти защитные меры и техническое сотрудничество.
* **для создания систем, в которых информационная безопасность стоит на первом месте:** **- инвестировать средства в обеспечение кибербезопасности и защиту от внутренних угроз для защиты конфиденциальных параметров (весов) моделей.** Эти параметры являются наиболее важной частью системы искусственного интеллекта, и компании согласны с тем, что их публикация должна осуществляться осознанно и с учетом рисков безопасности; - **содействовать обнаружению уязвимостей в системах искусственного интеллекта третьими лицами.** Некоторые проблемы могут сохраняться даже после выпуска системы искусственного интеллекта, и надежный механизм обратной связи о уязвимостях позволяет быстро находить и устранять их.
* **для завоевания доверия общественности:** - **разработать надежные технические механизмы, позволяющие пользователям знать, что контент создан искусственным интеллектом, например, систему маркировки.** Эти меры будут способствовать развитию творчества с использованием ИИ, но при этом снизят риски мошенничества и введения публики в заблуждение. - **публично сообщать о возможностях, ограничениях и областях, в которых уместно и неуместно применение их ИИ-систем.** Эти сообщения будут охватывать как риски безопасности, так и риски для общества, например, в контексте принципов справедливости и устранения предвзятости; - **уделять приоритетное внимание исследованиям социальных рисков, которые могут представлять системы ИИ, в том числе для предотвращения закрепления вредных предубеждений и дискриминации, а также для защиты неприкосновенности частной жизни.** Опыт использования ИИ показывает, что эти риски распространены, хотя и не всегда очевидны. Поэтому компании обязуются внедрять системы ИИ, в которых эти риски адресованы; **- разрабатывать и внедрять передовые системы ИИ для содействия в решении важнейших проблем общества.** От профилактики рака до смягчения последствий изменения климата и многого другого, ИИ - при правильном управлении - может внести огромный вклад в обеспечение всеобщего процветания, равенства и безопасности.

**Как тут не вспомнить заявление Марка Цукерберга при покупке «Вотсап», что данные вотсапа не будут объединяться c Facebook. Прошло время и про обещание (наверное, после разговора с FBI) все, в том числе и Марк, забыли…**

**Британия**

**Британское правительство** [**опубликовало доклад**](https://www.gov.uk/government/news/uk-unveils-world-leading-approach-to-innovation-in-first-artificial-intelligence-white-paper-to-turbocharge-growth), описывающий их подход к регулированию применения искусственного интеллекта. Был выбран вариант, при котором регуляторные полномочия делегированы существующим отраслевым органам. Этим органам на первоначальном этапе предлагается в добровольном порядке рассмотреть общие принципы регулирования, предложенные правительством. В перспективе, исходя из плана законодательной работы, эти общие принципы также будут обсуждены в парламенте и закреплены в законе. Также опубликована [**оценка регулирующего воздействия**](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1147045/uk_ai_regulation_impact_assessment.pdf). Исходя из неё, рассматривались следующие варианты:

**Вариант 0 — "Ничего не делать".** Основан на предположении, что ЕС примет Регламент об искусственном интеллекте в первоначальной редакции от апреля 2021 года. Великобритания не принимает никаких нормативных требований в отношении ИИ.

**Вариант 1 — Делегировать полномочия существующим отраслевым органам, которым будет рекомендовано следовать принципам, предлагаемым правительством.** Мягкий вариант, при котором существующие отраслевые органы реализуют межотраслевые принципы в рамках своих полномочий.

**Вариант 2 — Делегирование полномочий существующим отраслевым органам, которые будут обязаны следовать принципам, установленным правительством, при поддержке национального регулятора (Выбранный вариант).** Новые обязательства для предприятий самим правительством не устанавливаются.

**Вариант 3 — Централизованное регулирование с новыми законодательными требованиями к применению ИИ.** Великобритания создает новый орган по централизованному регулированию ИИ и устанавливает для предприятий обязательные требования, гармонизированные с требованиями Регламента ЕС об ИИ.

**Суд Британии. Владение системой с искусственным интеллектом, породившей патентоспособный результат, не влечёт возникновения у владельца системы права на патент:** [**https://t.me/robocounsel/652**](https://t.me/robocounsel/652)