

Почему нейросети «придумывают» нормы? 5 техник, как перестать работать с «несуществующими СП»



Данила Самарин, инженер-проектировщик, независимый эксперт, основатель онлайн-академии для проектировщиков timbuilder.ru и авторской методики обучения проектированию в Revit

Нейросети в проектировании инженерных систем – тема, которая вызывает полярные реакции. Одни видят в них инструмент, который изменит отрасль. Другие – очередной хайп, не имеющий отношения к реальной работе с расчетами, нормативами и экспертизой.

Истина, как обычно, где-то посередине. За последние два года мною протестированы десятки сценариев применения языковых моделей в задачах ОВиК: от поиска требований в СП до подготовки пояснительных записок, от работы с климатическими данными до формирования ответов на замечания экспертизы. Часть экспериментов закончилась разочарованием. Часть – дала результаты, которые теперь используются в повседневной работе.

Главный вывод: нейросети не заменяют инженера, но они могут существенно ускорить работу с информацией, если понимать их ограничения и использовать правильные инструменты. Эта статья – о конкретных техниках, которые работают у конкретного инженера, и об одной из главных проблем, с которой сталкивается каждый, кто пробует применить нейросети к нормативной документации.

Несколько месяцев назад я готовил материал по нормативной базе для применения нейросетей в системах ОВиК. Тема новая, документов много, времени мало. Решил ускорить процесс – попросил ChatGPT собрать релевантные ГОСТы и своды правил.

Результат выглядел безупречно. Структурированный список с конкретикой: ГОСТ Р 70949-2023 «Системы искусственного интеллекта в

промышленности» – требования к интеграции ИИ-решений в технологические процессы, точность прогнозирования не менее 85 %, время отклика системы не более 30 с. Названия и номера правдоподобные. Формулировки – как из реального норматива.

Я почти включил это в рабочий документ. Почти. Что-то заставило проверить первоисточники. ГОСТ Р 70949-2023 действительно существует.

Только называется он «Технологии искусственного интеллекта в образовании». Про научно-исследовательскую деятельность в университетах. Ни слова о промышленности и климатизации. Никаких 85 % и 30 с. – этих цифр не нашел нигде.

Три документа из четырех – с критическими ошибками. При этом нейросеть написала все как полагается: уверенный тон, конкретные цифры, правильная структура ссылок.

Если бы я не проверил – эти «нормативы» ушли бы дальше. В презентацию, в статью, в рабочую документацию. А я бы выглядел как специалист, который ссылается на несуществующие требования, перед коллегами, заказчиками или экспертизой.

После этого случая я начал разбираться: почему нейросеть так уверенно генерирует то, чего не существует? И что с этим делать?

Почему это происходит

Когда на обучениях объясняю эту проблему коллегам, использую простую аналогию.

Языковая модель не «знает» СП 60.13330 в том смысле, в каком его знаете вы. Она не читала документ, не поняла его логику, не запомнила структуру. Модель обучена на миллиардах текстов из интернета. На основе этого массива она научилась предсказывать, какой текст статистически вероятен после вашего запроса.

Когда вы спрашиваете про требования к воздухообмену, модель не ищет ответ в базе знаний. Она генерирует текст, который похож на тексты о требованиях к воздухообмену. Иногда этот текст совпадает с реальностью. Иногда – нет.

Это похоже на студента, который на экзамене всегда пытается дать ответ, даже когда не уверен. Модели обучены на данных, где «не знаю» – редкость. Статьи, книги, документация – авторы обычно пишут о том, что знают.

Проблема не уникальна для инженерной сферы. В мае 2023 года американский адвокат подал в суд документы со ссылками на шесть судебных прецедентов, которые нашел через ChatGPT. Все шесть оказались выдуманными. Когда судья потребовал объяснений, адвокат попросил ChatGPT подтвердить существование этих дел – и получил подтверждение. Итог: штраф 5000 долларов и публичное порицание. В мае 2025 года две крупные юридические фирмы получили штраф 31 000 долларов за аналогичную ошибку.

Масштаб проблемы серьезный. Исследование Стэнфордского университета показало: даже специализированные юридические системы с доступом к базам документов выдают 17–33 % ошибочных ссылок. Базовые модели без такого доступа – 58–82 %. База данных судебных случаев, связанных с ИИ-галлюцинациями, содержит более 800 задокументированных инцидентов.

В апреле 2024 года один из исследователей публично заявил, что проблема будет решена в течение года. Прошло больше года – проблема осталась.

Хорошая новость: существуют способы снизить риск. За два года экспериментов я отобрал пять техник, которые работают при обращении к нормативной документации. Не до нуля – но до уровня, когда инструмент становится полезным.

Техника 1. Перекрестная проверка

Идея пришла из методологии тестирования моделей: если задать один и тот же вопрос несколько раз и сравнить ответы, можно оценить надежность результата.

Логика простая. Если модель действительно нашла информацию в документе – ответы будут совпадать. Номера пунктов, страницы, формулировки. Если модель «придумывает» – каждый раз будет придумывать немного по-разному.

Как применять: задать вопрос, получить ответ. Начать новый чат, задать тот же вопрос. Сравнить результаты.

На что смотреть: совпадают ли номера пунктов? Однаковые ли страницы? Похожи ли формулировки? Если в разных ответах разные номера пунктов – это красный флаг, нужно проверять вручную.

Техника не гарантирует стопроцентной защиты – модель может стабильно выдавать одну и ту же ошибку. Но случайные ошибки, которые составляют значительную часть проблем, отсекаются хорошо. Это дополнение к другим техникам, не замена.

Техника 2. Разрешите нейронке сказать «не знаю»

По умолчанию модели обучены всегда давать ответ и делать это уверенно. Это приводит к ситуациям, когда нейросеть не просто ошибается, а ошибается уверенно.

Решение нашел в рекомендациях разработчиков и адаптировал под работу с нормативкой.



Суть – явно разрешить модели отказаться от ответа, если информации недостаточно.

Формулировка, которая работает: «Найди в документе требования к... Важно: если в документе нет информации по этому вопросу – прямо скажи «В загруженных документах нет данных по этому вопросу». Не пытайся дать ответ на основе общих знаний. Не заполняй пробелы предположениями».

Почему это работает? Вы снимаете с модели неявное требование «всегда помогать». Она получает разрешение признать ограничение. По моим наблюдениям, такая формулировка снижает количество увереных ошибок примерно вдвое.

Ограничение: работает не со всеми моделями одинаково хорошо. Некоторые все равно пытаются «помочь» и додумывают. Поэтому эта техника – дополнение к другим, а не самостоятельное решение.

Техника 3. Двухэтапный запрос – сначала цитата

Даже когда модель отвечает на основе загруженного документа, она может «творчески пересказать» содержание. Ответ выглядит как цитата, но на самом деле это интерпретация – иногда точная, иногда нет.

К этой технике я пришел после того, как трижды получил «перефразированные» пункты СП, которых в таком виде не существовало. Текст был похож на норматив, номера пунктов выглядели правдоподобно, но при проверке оказывалось, что это

вольный пересказ или компиляция из разных мест документа.

Решение – разделить задачу на два этапа.

Первый этап: «Найди в документе все фрагменты, относящиеся к... Скопируй их дословно, без пересказа. Укажи страницу для каждой цитаты. Если релевантных фрагментов нет – скажи об этом».

Второй этап: «Используя только эти цитаты, ответь на вопрос: ... Для каждого утверждения укажи, какая цитата его подтверждает. Не добавляй информацию, которой нет в цитатах».

Логика простая: сначала модель находит и фиксирует сырье данные, потом работает только с ними. Вы видите, что именно она нашла, и можете проверить до того, как она начнет интерпретировать.

Да, это требует больше времени – два запроса вместо одного. Для быстрых справочных вопросов может быть избыточно. Но для ответов на замечания экспертизы, где нужна точная ссылка на норматив, – оправдано полностью. Как-то, проверяя домашнее задание коллеги-инженера, нашел у него отличный наглядный пример этого решения: «Это как двойной контроль на производстве – дольше, но спокойнее».

Техника 4. RAG – ваши документы, ваши ответы

Обычные языковые модели – ChatGPT, Claude, GigaChat – работают на основе данных, на которых были обучены. Когда вы спрашиваете о требованиях СП 60.13330, модель не обращается к тексту



документа – она генерирует ответ на основе того, что «помнит» из обучающей выборки.

RAG-системы (Retrieval-Augmented Generation) работают иначе. Вы загружаете конкретные документы, и модель ищет ответы только в них. Не «придумывает» на основе общих знаний – находит релевантные фрагменты в ваших файлах и дает ссылки на конкретные места в документах.

На рынке существует несколько RAG-решений: Perplexity с загрузкой файлов, корпоративные системы, надстройки над ChatGPT. Для работы с нормативной документацией я остановился на NotebookLM от Google (notebooklm.google.com) – бесплатен (на момент написания статьи), прост в освоении и хорошо работает с техническими текстами на русском языке.

Вы создаете «блокнот», загружаете туда СП, СНиПы, стандарты АВОК – до 50 источников. Затем в чате задаете вопросы. Важно: чат работает в контексте конкретного блокнота. Вы не даете ссылку на блокнот в запросе – вы работаете внутри него. Открыли блокнот с загруженными СП – все вопросы обрабатываются по этим документам.

Я загрузил СП 60.13330.2020 и задал тот же вопрос про воздухообмен, на котором раньше получал выдуманные пункты. Ответ пришел со ссылкой на конкретную страницу. Проверил – все точно.

Еще один момент, который выяснил на практике. Можно выбирать, по каким именно источникам искать. Нужны требования только по СП 60? Снимаете галочки с остальных документов в панели источников – и система работает только с ним. Это сужает контекст поиска и повышает точность ответов. Потом можно снова включить все источники для более широкого исследования.

Отдельная полезная возможность – сохранение результатов. Провели анализ по какому-то вопросу, получили хороший структурированный ответ – нажимаете «Save to note», и он сохраняется как заметка. Потом эту заметку можно превратить в источник через «Convert to source» и использовать для дальнейшей работы. Получается накапливаемая база знаний: сначала исследуете тему по первичным документам, потом сохраняете выводы и строите на них следующий уровень анализа.

Важное ограничение: даже NotebookLM может ошибаться. RAG-системы снижают количество выдуманных ссылок, но не устраниют полностью. Проверка по первоисточнику остается обязательной. Но проверять стало намного проще – ссылка ведет точно туда, откуда взята информация.

Техника 5. Нейросети для структуры, а не для фактов

Это правило я вывел после многих экспериментов. Нейронка отлично справляется с организацией информации: структурирует, группирует, форматирует, переформулирует. И плохо – с точным воспроизведением фактов из нормативных документов.

Безопасные задачи для ИИ: составить структуру пояснительной записи, сделать чек-лист проверки раздела перед сдачей, переформулировать техническое описание более понятным языком, создать шаблон ведомости, подготовить структуру ответа на замечание экспертизы.

Ганс Остберг создал первый в мире канальный центробежный вентилятор, в последствии получивший наименование СК. Это явилось настоящим событием в мире вентиляции и до сих пор СК является инженерной концепцией, признанной по всему миру.

«Östberg» – это не просто имя производителя, это характеристика, говорящая о прекрасных свойствах вентиляционной техники. Каждый вентилятор этой компании можно без преувеличения назвать изобретением. У каждой модели есть своя история, свое лицо, свое назначение.

Да, они разные, но есть то, что всех их объединяет между собой. Все они идеально отлажены, эффективны, надежны и долговечны.

Приобретая «Östberg», приобретаешь уверенность.



системы вентиляции, отопления
и кондиционирования воздуха

Москва, ул. Тимирязевская 1, строение 4.

Тел.: +7 (495) 981-15-15, +7 (499) 755-15-15.

Санкт-Петербург, ул. Большая Московская,
д. 18, литер А, помещение 14-Н.

Тел.: +7 (812) 441-35-30.

www.ARKTIKA.ru



Рискованные задачи: какие требования к воздухообмену по СП 60, какая минимальная температура приточного воздуха, процитировать конкретный пункт норматива. Здесь вероятность ошибки максимальная.

Есть и продвинутый сценарий работы. Можно создать собственную структурированную библиотеку. Как это выглядит: загружаете в RAG все СП и СНиПы, проводите по ним систематическое исследование – например, собираете все требования по температурным режимам для разных типов зданий. Результаты сохраняете и превращаете в источник на будущее. Получается ваша личная база, по которой можно быстро искать.

Я так сделал сводную таблицу по температурам и воздухообменам для десяти типов зданий – от жилых домов до медицинских учреждений. Загрузил все релевантные СП, прошелся по каждому типу здания отдельным запросом, собрал результаты. Теперь не нужно каждый раз перекапывать десять документов – открываю свою базу и нахожу нужное за минуту. На создание такой библиотеки нужно потратить достаточно времени.

Можно пойти еще дальше и создать индекс – документ, который описывает, что в каком файле находится. Тогда при поиске система сначала смотрит в индекс, определяет нужный документ, потом ищет уже в нем. Это экономит время и повышает точность, особенно когда документов много.

Правило остается неизменным: нейросеть организует информацию – вы наполняете фактами. Нейросеть создает структуру – вы проверяете содержание. Нейросеть ускоряет рутину – вы принимаете решения.

Чего нейросеть не может

Было бы нечестно не сказать о принципиальных ограничениях.

Возможно, проблема выдуманных данных не исчезнет полностью. Это не баг, который исправят в следующем обновлении. Исследователи сходятся во мнении, что полностью устранить эту проблему в рамках текущей архитектуры языковых моделей невозможно. Это фундаментальное свойство того, как они работают.

Нейросеть не понимает, что говорит – в том смысле, в каком понимаете вы. У нее нет модели реальности, нет способа проверить, соответствует ли сгенерированный текст действительности. И нет профессиональной ответственности за результат. Подпись на чертеже ставите вы, не нейросеть.

Я думаю об этом так: нейросеть – это калькулятор для текста. Калькулятор считает быстро и без ошибок, но он не проверяет, правильные ли числа вы ввели. Если вы ошиблись при вводе – получите неправильный результат, красиво оформленный. С нейросетями то же самое: ввод данных, проверка результата и принятие решений остаются за инженером.

Что дальше

Пять техник для работы с нормативкой: RAG с загруженными документами, явное разрешение сказать «не знаю», двухэтапный запрос с цитированием, перекрестная проверка несколькими запросами, использование для структуры вместо фактов. Главный принцип один: нейросеть – помощник под вашим контролем, не замена инженеру.

Думаю, через три-пять лет работа с нейронами станет такой же базовой компетенцией, как работа с AutoCAD. Не потому что нейросети станут идеальными, а потому что инженеры научатся их правильно использовать. Вопрос не в том, произойдет ли это, а в том, кто освоит инструмент раньше.

Конкретный первый шаг, который можно сделать сегодня: загрузите в RAG один СП, с которым часто работаете, и задайте вопрос, ответ на который вы точно знаете. Посмотрите, как система находит информацию, как дает ссылки, как подсвечивает фрагменты. Сформируйте собственное мнение.

Нейросеть не заменит инженера, но сильно изменит нашу работу в проектировании. Лично я больше не хочу проектировать так, как проектировал последние 10 лет до внедрения нейросетей в рабочий процесс, – моя работа изменилась кардинально. Возможно, у вас будет иначе...

HL



ВАРИАНТЫ СИФОНОВ СКРЫТОЙ УСТАНОВКИ

ВАРИАНТЫ СИФОНОВ С РАЗРЫВОМ СТРУИ

ВАРИАНТ СИФОНА ДЛЯ ОТВЕДЕНИЯ ДРЕНАЖА ОТ УСТАНОВОК, СОЗДАЮЩИХ ПЕРЕПАД ДАВЛЕНИЯ ДО 2800 Па

ВАРИАНТ С АДАПТЕРОМ, ОБЕСПЕЧИВАЮЩИЙ ГИГИЕНИЧЕСКИЕ ТРЕБОВАНИЯ СП 30.13330-2020 И СП 2.3.6.1079-01

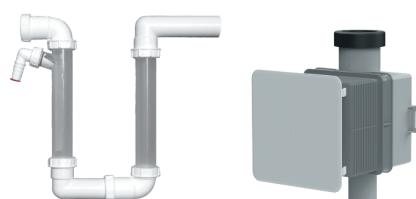
ТЕХНИЧЕСКАЯ ПОДДЕРЖКА И ПОМОЩЬ В ПОДБОРЕ ПРИ ПРОЕКТИРОВАНИИ ОТ ПРОИЗВОДИТЕЛЯ



СИФОНЫ ДЛЯ КОНДИЦИОНЕРОВ

 **ИНТЕРМА**
СИСТЕМЫ ОТОПЛЕНИЯ И ВОДОСНАБЖЕНИЯ

+7 (495) 780 70 00
www.interma.ru



Реклама