

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ В СТРОИТЕЛЬНОЙ СФЕРЕ: ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ И ПОСТАНОВКА ЗАДАЧИ НА СОЗДАНИЕ СИСТЕМЫ ДОССНЕК

Әзілхан Алишер

22azlikanalisher@gmail.com

магистрант 1 курса образовательной программы 7М06105 «Программная инженерия»
Атырауский университет имени Х. Досмухамедова, г. Атырау, Республика Казахстан
Научный руководитель - Ассоциированный профессор, PhD Молдашева Ж. Ж.

Введение. Строительная отрасль является одной из самых документоёмких сфер экономики. На каждом этапе от проектирования до сдачи объекта в эксплуатацию создаются десятки и сотни технических документов: акты скрытых работ, сметы, сертификаты на материалы, исполнительные схемы, чертежи (в том числе DWG) и другие.

Достоверность этих документов напрямую влияет на безопасность объектов, их стоимость и сроки сдачи. В условиях цифровизации строительной отрасли Республики Казахстан и роста объёма документации возникает необходимость автоматизации контроля за их корректностью и соответствием нормативам.

Практика показывает, что значительная часть документации содержит ошибки, противоречия или признаки умышленной фальсификации. Доля проблемных документов достигает 12-25% в зависимости от сегмента строительства. Ручной аудит выявляет не более 40% нарушений, а финансовые потери из-за ошибок и подмен материалов могут составлять до 20-30% сметной стоимости.

Цель статьи это провести обзор существующих программных решений, выявить их недостатки и обосновать необходимость создания новой системы проверки документов, основанной на методах искусственного интеллекта и обработки естественного языка.

2. Проблема поддельных и некорректных технических документов

2.1 Типология нарушений

На основе анализа реальной строительной документации выделены три основные категории проблем:

1. Полная подделка - документ создан без реального выполнения работ или поставки материалов (например, фиктивный акт освидетельствования скрытых работ).

2. Частичная фальсификация - внесение ложных данных в подлинный бланк (завышение объёма земляных работ, подмена марки бетона).

3. Некорректность - формально подлинный документ, но с ошибками против ГОСТ/СНиП (неверные коэффициенты в смете, отсутствие обязательных реквизитов).

2.2 Масштаб и последствия

Анализ отраслевых отчетов и статистических данных за 2023-2025 гг. позволяет количественно оценить масштаб проблемы и ее влияние на строительную отрасль. Качество проектной документации не только не улучшается, но и демонстрирует негативную динамику. В 2025 году количество отрицательных заключений государственной экспертизы на проектно-сметную документацию выросло на **50%** по сравнению с предыдущим годом и достигло **4279** случаев. Доля отрицательных заключений увеличилась с 3,8% до **5,7%** от общего числа выданных. Особенно тревожным выглядит рост отказов по объектам строительства на **55,2%**. При этом важно отметить, что лишь **0,3%** выявленных нарушений относятся к критичным, способным привести к авариям или человеческим жертвам. Остальные **99,7%** нарушений это системные ошибки, некорректное оформление и несоответствие документации требованиям, которые тем не менее требуют исправления.

Еще одна проблема поддельных документов особенно остро стоит в сфере государственных закупок. По данным Федеральной антимонопольной службы (ФАС),

более **60%** компаний, уличенных в предоставлении недостоверных сведений, относятся к строительной отрасли. Наиболее распространенное нарушение это подделка договоров, подтверждающих опыт и квалификацию участников тендеров.

Общая сумма госзакупок, проведенных по фальшивым документам, в 2023 году составила **10 млрд рублей**, а за 2024 год только за первое полугодие уже **6,5 млрд рублей**.

2.3 Примеры реальных инцидентов

Международная практика изобилует примерами, когда некачественная документация приводила к многомиллионным убыткам:

1. **Фальсификация в инфраструктурных проектах (Филиппины, 2025 год):** Аудиторы выявили «призрачные проекты» и несоответствие фактических работ документации на сумму более **306 млн песо**. В ряде случаев выяснилось, что подрядчики получали оплату за строительство объектов, которые существовали еще до заключения контракта.

2. **Неточности в документации коммерческой недвижимости (США):** Около **70%** коммерческих зданий эксплуатируются с устаревшей или недостоверной документацией. Это приводит к перерасходу бюджета на реновацию в размере **15-30%** и задержкам на **4-12 недель**.

3. **Низкое качество данных о материалах (международные данные):** Исследование, охватившее более миллиона документированных поставок, показало, что **95%** записей о доставке материалов содержат серьезные проблемы с данными (отсутствие веса, неверные локации, некорректные идентификаторы). Только **34%** записей достаточно точны для расчета углеродного следа.

2.4 Последствия для безопасности и репутации

Выявленные дефекты после сдачи объекта увеличивают расходы на переделку в **3-5 раз** по сравнению с исправлением на этапе строительства. Более того, случаи использования несертифицированных материалов и несоответствия проектным решениям напрямую влияют на безопасность эксплуатации объектов.

Ручной аудит, который сегодня является основным инструментом контроля, выявляет не более **40%** нарушений, что подтверждает необходимость внедрения автоматизированных систем проверки документации. Согласно исследованию Ассоциации сертифицированных экспертов по борьбе с мошенничеством (ACFE) 2025 года, в строительных компаниях функции по расследованию мошенничества в 30% случаев подчиняются непосредственно генеральному директору или высшему руководству, в 20% совету директоров, в 20% руководителю внутреннего аудита и в 20% руководителю риск-менеджмента [9]. Это свидетельствует о высоком уровне внимания к проблеме мошенничества в строительной отрасли на уровне топ-менеджмента.

3. Анализ существующих решений

3.1 Bit&Beam

Тип: система управления документами с элементами ИИ. Возможности: хранение, поиск, OCR, извлечение сущностей. Ограничения: не проверяет логические противоречия, не работает с чертежами, не проверяет ГОСТ/СНиП.

3.2 PlanRadar

Тип: система контроля качества на стройплощадке. Возможности: фиксация дефектов, фотофиксация, формирование актов. Ограничения: не анализирует сметы и сертификаты, не работает с документами напрямую.

3.3 Glodon

Тип: BIM-экспертиза. Возможности: проверка BIM-моделей на соответствие нормам. Ограничения: требует BIM-модель, не работает с PDF, Word, DWG; ориентирована на китайские стандарты.

3.4 ARCEAK

Тип: научный прототип. Возможности: перевод текста норм в машинный код, использование LLM. Ограничения: экспериментальный уровень, не промышленное использование.

3.5 Сравнительный анализ

В таблице 1 представлено сравнение существующих решений по ключевым критериям.

Характеристика	Bit&Beam	PlanRadar	Glodon	ARCEAK
Проверка PDF / Word	+	-	-	-
Проверка DWG (чертежи)	-	-	-	-
Логические противоречия	-	-	-	-
Орфография / грамматика	-	-	-	-
Проверка ГОСТ / СНиП	-	-	+ (только BIM)	-
Работа без BIM	+	+	-	+

Таблица 1 - Сравнение существующих решений

Простой вывод: на рынке нет системы, которая бы комплексно проверяла все типы документов. Это обосновывает необходимость разработки нового проекта. Я назвал его Doccheck.

4. AI и NLP для проверки документов

Традиционные методы (регулярные выражения, шаблоны, словари) не способны справиться с многообразием формулировок, контекстов и скрытых противоречий в строительных документах. Именно поэтому необходимо привлечение методов искусственного интеллекта (AI) и обработки естественного языка (NLP).

Основные задачи, решаемые с помощью AI и NLP:

- определение типа документа (классификация текста с помощью BERT или LLM);
- выделение сущностей (NER - поиск материалов, дат, подписей);
- поиск логических противоречий (сравнение эмбедингов, Siamese сети);
- проверка соответствия нормам (сравнение числовых параметров);
- анализ чертежей (OCR + послойный анализ).

5. Архитектура системы DocCheck

DocCheck - модульная интеллектуальная система мультимодальной проверки документации. Система состоит из 3 модулей. Первый основных правил названных контрольными точками (КТ), второй проверка данных ии, третьи сравнение двух данных.

5.1 Модуль распознавания и классификации кт 0 и 1

Когда алгоритму попадает документ или документы, программа должен понять его содержание и что за документ таким образом используется две правил кт. КТ0 выполняет задачи как определение форматов файла как PDF, Docx, Excel, DWG и их существование. Конвертирует файлы в читаемый формат. И конечно проверяет, нет ли повреждённых

страниц или слоёв и можно ли извлечь текст с данными еще делает предварительную чистку чтобы не засорять проверку.

КТ1 определяет тип документа: проектная документация, чертёж, ведомость или смешанный тип. Это важно для того, чтобы последующие проверки выполнялись с учётом специфики каждого вида документа.

Таким образом данные полученные от первый двух кт распределяется другим кт и ии для проверки. Всегда первый должен работать кт0 и потом кт1. Если один кт не будет выполнен другие алгоритм не будет работать. Это обязательный условия!

5.2 Остальные правил контрольных точек (КТ-2 - КТ-6)

КТ2 проверяет структуру документа: наличие заголовков, содержания, разделов и других обязательных элементов. Такая проверка позволяет убедиться, что документ оформлен корректно и пригоден для анализа.

КТ3 отвечает за реквизиты и авторские данные: система проверяет наличие информации об авторе, проверяющем и утверждающем документ. Это обеспечивает контроль ответственности за подготовку и проверку документации.

КТ4 проверяет соответствие цифровых данных: объёмы, масштабы, числовые значения, а также имена людей. Такая проверка позволяет выявлять расхождения между различными частями документа или между несколькими документами.

КТ5 отвечает за орфографию: система проверяет правильность написания слов, грамматику и соответствие технической терминологии. Это снижает риск ошибок, которые могут повлиять на точность выполнения работ.

КТ6 проверяет согласованность названия объекта внутри документа: система убеждается, что название объекта не меняется на протяжении всего документа, что особенно важно для комплексных проектов с большим количеством файлов.

Таким образом, последовательное применение КТ2-КТ6 позволяет DocCheck проводить **полную модульную проверку документов**, начиная от базовой доступности и формата файла до точности числовых данных, структуры, реквизитов и правильности орфографии. Каждая контрольная точка выполняет конкретную функцию, обеспечивая высокую точность и надежность проверки строительной документации.

5.3 Модуль ии

ИИ можно самостоятельно обучить или использовать рабочие ии. Для работы я использую второй вариант с api ключом от Gemini потому что он бесплатный!. Он тоже выполняет задачу кт. Извлеченный данные передается с специализированном промтом и его результат возвращается алгоритму.

5.4 Модуль сравнение

Модуль сравнение как говориться названий проверяет данный из двух источников. Именно от алгоритма и от ии. После того как программа будет убежден что данные одинаковые, результат отправляется. Результата можно скачать и увидет.

6. Преимущества системы DocCheck

В таблице 2 представлено сравнение DocCheck с существующими решениями.

Преимущество	DocCheck	Bit&Beam	PlanRadar	Glodon
Комплексность проверки	+	-	-	-
Логические противоречия	+	-	-	-
Орфография / грамматика	+	-	-	-
Анализ DWG-чертежей	+	-	-	-

Преимущество	DocCheck	Bit&Beam	PlanRadar	Glodon
Проверка ГОСТ / СНиП	(в планах)	-	-	+ (только BIM)
Не требует BIM	+	+	+	-
Поддержка русского / казахского языков	+	частично	частично	-

Таблица 2 - Преимущества DocCheck перед аналогами

7. Заключение. В результате проведенного исследования получены следующие результаты:

1. Подтверждена актуальность проблемы: доля некорректных и поддельных документов в строительстве достигает 25%, а ручная проверка выявляет не более 40% нарушений.

2. Обзор существующих решений (Bit&Beam, PlanRadar, Glodon, ARCEAK) выявил отсутствие системы, способной комплексно проверять все типы документов.

3. Обоснована необходимость применения AI и NLP для решения задач классификации, выделения сущностей и поиска логических противоречий.

4. Представлена модульная архитектура DocCheck с пятью контрольными точками (КТ-2 - КТ-6).

5. Показаны ключевые преимущества: комплексность, мультимодальная проверка (PDF, Word, DWG), поддержка русского и казахского языков, работа без BIM.

Дальнейшие работы включают завершение КТ-4 (анализ DWG-чертежей), реализацию КТ-5 (проверка по ГОСТ/СНиП), проведение экспериментов на 400+ документах и публикацию статей по отдельным модулям.

Список использованной литературы:

1. Vaswani A., et al. Attention is all you need. - NeurIPS, 2017.
2. Devlin J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. - NAACL, 2019.
3. Bit&Beam. GitHub. - URL: <https://github.com/BitAndBeam>
4. PlanRadar. Official website. - URL: <https://www.planradar.com>
5. Glodon. Official website. - URL: <https://www.glodon.com>
6. ARCEAK: A framework for automatic compliance checking. - arXiv preprint.
7. Агентство новостей «Строительный бизнес». Количество отрицательных заключений органов экспертизы за год выросло на 50%. - URL: <https://ancb.ru/news/read/20631>
8. Газета.Ru. ФАС выявила рост числа фальшивых договоров в строительстве. - URL: <https://www.gazeta.ru/business/news/2024/07/01/23359759.shtml>
9. CFMA. The Importance of Fraud Prevention in the Construction Industry. - URL: <https://beta.cfma.org/articles/the-importance-of-fraud-prevention-in-the-construction-industry>
10. ACFE. In-House Fraud Investigation Teams: 2025 Benchmarking Report. - URL: <https://www.acfe.com/-/media/files/acfe/pdfs/in-house-fraud-investigation-teams-benchmarking-report-2025.pdf>

11. Visayan Daily Star. NBI starts probe on P306M San Carlos 'ghost projects'. - URL: <https://visayandailystar.com/nbi-starts-probe-on-p306m-san-carlos-ghost-projects>
12. QFlow. The State of Data Quality in Construction. - Construction Management Magazine. - URL: <https://constructionmanagement.co.uk/poor-data-hinders-constructions-sustainability-drive>
13. СНиП РК. Строительные нормы и правила Республики Казахстан.