

УДК 677.017

Анализ языковых генераций искусственного интеллекта для оценки изделий лёгкой промышленности

**Орлов А. В., асп.,
Шустов Ю. С., д.т.н., проф.,
зав. кафедрой**

Российский государственный
университет им. А. Н. Косыгина
(Технологии. Дизайн. Искусство),
г. Москва, Российская Федерация

Реферат. На современном этапе цифровизации науки языковые генеративные модели стали неотъемлемым инструментом в научных исследованиях, включая область текстильного материаловедения. Международная практика внедрения технологий искусственного интеллекта в сферу высшего образования демонстрирует значительный потенциал ИИ в автоматизации рутинных процессов, индивидуализации образовательных траекторий и расширении преподавательских возможностей. Вместе с тем, как свидетельствуют актуальные аналитические исследования, фактическое применение ИИ в академической среде остаётся ограниченным и, в основном, выполняет вспомогательные функции. Использование языковых генеративных моделей позволяет автоматизировать анализ литературных источников, формулировку гипотез, разработку планов экспериментов и интерпретацию полученных данных. Для выбора оптимального инструмента планирования научного эксперимента рассмотрены возможности и ограничения ведущих языковых моделей: ChatGPT-4 (OpenAI), Claude 3 (Anthropic), DeepSeek-V3 (DeepSeek AI) и YandexGPT 5 Pro-4.

Ключевые слова: искусственный интеллект, языковые генерации, выбор оптимального инструмента, планирование научного эксперимента.

Введение

Текстильные материалы обладают сложным комплексом взаимосвязанных свойств, включая прочность, износостойкость, способность к влагопоглощению и воздухопроницаемость. При исследовании каждого из этих параметров необходимо учитывать внешние условия, тип материала и его предполагаемую область применения.

Традиционные методы планирования научных исследований основаны на когнитивно-экспертных подходах и предполагают использование различных моделей представления и формализации знаний. Внедрение технологий искусственного интеллекта (ИИ) позволяет не только сократить затраты времени, но и повысить точность получаемых результатов,

что подтверждено в ряде прикладных исследований [1].

Для экспериментальных исследований текстильных материалов необходимо учитывать специфику многофакторного влияния технологических параметров на формирование эксплуатационных свойств изделий. Типовые задачи научного эксперимента в данном контексте заключаются в следующем:

- отбор значимых факторов, определяющих физико-механические и гигиенические свойства стелечных материалов;
- определение критических контрольных точек измерений в зависимости от условий эксплуатации материалов;
- установление оптимального количества серий испытаний с учётом требуемой точности и допустимой погрешности;
- учет внешних факторов, таких как температура и влажность, влияющих на ход эксперимента;
- разработка плана эксперимента, минимизирующего количество испытаний и обеспечивающего статистическую достоверность получаемых результатов;
- систематизация экспериментальных данных для последующего анализа и интерпретации результатов.

Все эти задачи могут быть эффективно делегированы системам искусственного интеллекта. При этом следует учитывать, что языковые генеративные модели (GPT – Generative Pretrained Transformer) не способны создавать принципиально новые знания. Их основное преимущество заключается в анализе существующих данных и предоставлении рекомендаций на основе выявленных закономерностей [2]. ИИ способен обрабатывать большие массивы экспериментальных данных, полученных в предыдущих исследованиях, и использовать их для построения гипотез и оптимизации параметров последующих экспериментов. Например, перед тестированием нового текстильного материала можно поручить ИИ анализ данных о прочности на разрыв ранее изготовленных образцов, что позволит скорректировать параметры испытаний. Согласно исследованиям, автоматизация выбора экспериментальных условий значительно сокращает временные затраты и позволяет прогнозировать поведение материалов в экстремальных условиях [3].

Эффективное использование ИИ продемонстрировано в анализе физико-механических свойств тканей при проектировании одежды из многослойных текстильных материалов. Применение нейронных сетей для автоматизации расчётов даёт возможность прогнозировать эксплуатационные характеристики материалов в различных режимах нагружения, что особенно важно для создания изделий с особыми требованиями, где ошибки могут привести к существенным финансовым потерям. При этом подчёркивается важность выбора адекватных методов оценки качества нетканых материалов.

Основная часть

Одной из актуальных задач остаётся адаптация предложенных ИИ рекомендаций к конкретным условиям лабораторных испытаний. Например, диапазоны температуры и

влажности, определённые языковой моделью, могут не соответствовать возможностям имеющегося оборудования. В этом случае требуется предварительная корректировка параметров эксперимента, что подтверждается исследованиями в области физико-химических характеристик многослойных текстильных материалов. Предполагается, что сочетание рекомендаций искусственного интеллекта с экспертным опытом позволит существенно повысить надёжность и практическую применимость результатов.

На современном этапе цифровизации науки языковые генеративные модели стали неотъемлемым инструментом в научных исследованиях, включая область текстильного материаловедения. Международная практика внедрения технологий искусственного интеллекта в сферу высшего образования демонстрирует значительный потенциал ИИ в автоматизации рутинных процессов, индивидуализации образовательных траекторий и расширении преподавательских возможностей. Вместе с тем, как свидетельствуют актуальные аналитические исследования, фактическое применение ИИ в академической среде остаётся ограниченным и, в основном, выполняет вспомогательные функции. В этой связи важным фактором становится не только развитие самих технологий, но и выстраивание культуры их осознанного критического восприятия, а также обеспечение этического, сбалансированного и педагогически выверенного их внедрения в образовательную практику.

На текущем этапе развития искусственного интеллекта применение языковых генеративных моделей требует комплексной оценки их релевантности по ряду показателей, включая точность, обоснованность, логическую стройность, устойчивость к генерации недостоверной информации и способность учитывать научный контекст. Эффективность работы моделей напрямую зависит от качества исходных данных, точности используемой терминологии и структуры академического текста [3, 4].

Языковые модели способны не только генерировать связный текст, но и формулировать гипотезы, подбирать релевантную литературу, строить логические цепочки рассуждений и предлагать сценарии исследовательской работы. Однако для того, чтобы использовать эти модели как полноценный инструмент в научной деятельности, требуется обоснованный выбор конкретной нейросети на основе чётко сформулированных критериев.

Использование языковых генеративных моделей позволяет автоматизировать анализ литературных источников, формулировку гипотез, разработку планов экспериментов и интерпретацию полученных данных. Для выбора оптимального инструмента планирования научного эксперимента рассмотрим возможности и ограничения ведущих языковых моделей: ChatGPT-4 (OpenAI), Claude 3 (Anthropic), DeepSeek-V3 (DeepSeek AI) и YandexGPT 5 Pro-4.

ChatGPT (OpenAI) — наиболее широко используемая языковая модель на базе архитектуры GPT-4, разработанная компанией OpenAI. Отличается высокой степенью генерализации, точностью формулировок, способностью поддерживать сложный научный диалог и генерировать структурированные фрагменты текста, включая списки, таблицы и псевдокод. Обладает встроенными механизмами логического вывода, позволяет производить анализ и синтез информации, адаптируясь под стиль научного языка.

Claude (Anthropic) — языковая модель, ориентированная на этически безопасное взаимодействие и гуманитарные дисциплины. Обладает хорошими навыками обобщения и рассуждения в технической и научной сферах по глубине проработки деталей и способности к аналитической обработке сложной информации.

DeepSeek (DeepSeek AI) — одна из развивающихся моделей, ориентированных на англоязычную и китайскую научную аудиторию. Демонстрирует хорошие результаты в задачах формирования цепочек рассуждений и оптимизирована для задач, требующих высокой степени обоснованности.

YandexGPT (Яндекс) — русскоязычная языковая модель, адаптированная для бытовых и профессиональных задач, включая техническую документацию и краткий анализ текстов. Её преимуществом является хорошее понимание терминологии на русском языке, устойчивость к лексико-грамматическим ошибкам и стабильная работа в закрытой среде.

В таблице 1 приведено сравнительное распределение по ключевым параметрам, релевантным для задач планирования научного эксперимента, созданное в результате анализа официальных публикаций разработчиков моделей [5].

Таблица 1 – Сравнительный анализ по критериям релевантности к научному планированию

Критерий оценки / Модель	ChatGPT (GPT-4)	Claude-3	YandexGPT 5 Pro-4	DeepSeek-V3
Уровень аргументации	высокий	высокий	средний	средний
Способность к научному стилю	высокий	высокий	средний	средний
Устойчивость к «галлюцинациям»	высокая	высокая	средняя	низкая
Контекстное планирование	высокая	высокая	средняя	средняя
Работа с табличными и числовыми данными	высокая	средняя	низкая	средняя
Воспроизводимость запросов	высокая	средняя	высокая	средняя
Поддержка русского языка	высокая	средняя	высокая	средняя
Возможность адаптации под терминологию	высокая	высокая	средняя	средняя
Доступность API и документации	высокая	средняя	ограниченная	ограниченная
Возможности генерации кода и формул	высокая	средняя	низкая	высокая

Сравнение нейросетей в контексте использования для научных исследований на основании официальных данных разработчиков показало следующее:

- Claude может быть полезен в случаях, где важна этичность и безопасность, но для специфичных научных задач ChatGPT является более точным и универсальным инструментом;
- YandexGPT ориентирован на русскоязычные запросы, но в силу ограниченности данных и функционала по сравнению с ChatGPT, его использование в высокоспециализированных задачах, таких как оптимизация научных экспериментов, ограничено;
- DeepSeek в версии, актуальной на момент написания данной работы, не может быть рекомендован для академической деятельности ввиду высокой степени «галлюцинаций нейросети» (выдача несуществующих данных);
- ChatGPT является оптимальным выбором для планирования научных экспериментов благодаря сочетанию универсальности, аналитических возможностей и гибкости в обработке запросов.

Для объективной оценки был выбран метод экспертной оценки качества научных текстов. Основные критерии данного метода относительно использования искусственного интеллекта в научных работах были выработаны на основе анализа научной литературы по вопросам академического письма, когнитивной логики и стиля научного изложения, включая подходы [4–7]. Оценка производилась по ключевым критериям, релевантным задачам создания научного текста:

- достоверность фактов;
- логичность структуры изложения;
- соответствие научному стилю;
- устойчивость к генерации недостоверной информации («галлюцинации»).

Фокус-группа проводилась с участием 10 специалистов из вузов Москвы. Участвовавшие эксперты представляли разные области науки, что обеспечило междисциплинарную валидность результатов.

Процедура оценки была организована как двухэтапная: сначала проводилось индивидуальное тестирование моделей по одинаковым заданиям (формулировка гипотезы, описание эксперимента, аннотация), затем результаты оценивались экспертной группой по десятибалльной шкале (каждый эксперт заполнял стандартную форму оценки по шкале от 1 до 10 баллов).

В таблице 2 приведены критерии оценки результатов, полученных от языковых моделей по каждому из этапов тестового задания, на основании которых участники фокус-группы проводили экспертную оценку качества ответов. Каждый этап оценивался по нескольким показателям: логичность, полнота, корректность терминологии, соответствие академическому стилю, реализуемость и др. Такая детализация позволяет обеспечить объективность, воспроизводимость и методическую строгость в интерпретации результатов тестирования языковых моделей [4].

Особое внимание уделялось устойчивости моделей к так называемым «галлюцинациям» — генерации недостоверной или вымышленной информации. Этот параметр был включён как отдельный критерий. Таким образом, методология сочетает как количественные,

так и качественные элементы анализа, что делает выбор модели обоснованным с позиции академической достоверности и воспроизводимости.

Таблица 2 – Структура оценки этапов задания генерации языковых моделей

Этап задания	Критерии оценки
1. Формулировка научной гипотезы	- логичность формулировки; - научная корректность; - конкретность и проверяемость гипотезы
2. Описание плана эксперимента	- полнота описания этапов; - логичность и последовательность; - реализуемость; - научный стиль
3. Аннотация к научной статье	- структура (цель, метод, результат, значимость); - ясность изложения; - терминологическая точность; - соответствие академическому стилю

Полученные данные позволяют проследить вариативность восприятия модели разными экспертами.

Высокая согласованность оценок у Claude и ChatGPT (отклонение не более ± 1) свидетельствует о стабильности восприятия качества генерации. Напротив, DeepSeek и особенно YandexGPT показали больший разброс мнений, что говорит о менее однозначном качестве и сниженной академической целостности их текстов.

Результаты работы экспертов фокус-группы с усредненными оценками и пояснениями к ним приведены в таблице 3. Оценки представлены в виде средних баллов по десятибалльной шкале и сопровождаются пояснениями, отражающими позицию фокус-группы по каждому параметру.

Таблица 3 – Усредненные оценки 4-х языковых моделей с пояснениями экспертов фокус-группы

Тестовое задание	ChatGPT		Claude		DeepSeek		YandexGPT	
	балл	пояснение	балл	пояснение	балл	пояснение	балл	пояснение
1	2	3	4	5	6	7	8	9
1. Формулировка научной гипотезы	9,9	Гипотеза конкретна, количественно выражена, логична и проверяема	9,9	Гипотеза полная, с диапазонами, логична и экологически обоснована	8,0	Гипотеза логична и научно корректна, но без диапазона или количественной формулировки	6,0	Формулировка абстрактна, без цифр и четкой проверяемости

Окончание таблицы 3

1	2	3	4	5	6	7	8	9
2. Описание плана эксперимента	9,9	План обобщен, содержит методику, параметры, статистику. Стиль научный	10	Очень подробный, многоуровневый план, охватывает весь жизненный цикл	7,9	План содержательный, но менее строго академичен. Статистика ограничена	7,0	План описательный, но неформализованный, стиль ближе к отчетному
3. Аннотация к научной статье	10	Аннотация полная, структурирована, в научном стиле, с результатами	9,9	Аннотация насыщенная, с сильной терминологией и академическим языком	6,9	Аннотация логична, но уступает по терминологии и стилю	6,0	Аннотация повторяет план, структура слабая, нет акцента на методах
4. Устойчивость к «галлюцинациям»	10	Факты корректны, стандартные методы указаны, «галлюцинаций» нет	10	Информация строго достоверна, ошибки отсутствуют	8,9	Все методы реальны и корректны, нет ошибок или искажений	8,0	Ошибок нет, но часть терминов расплывчата

Таблица 4 отражает сводные результаты экспертной оценки четырёх языковых моделей – ChatGPT, Claude, DeepSeek и YandexGPT – по четырём ключевым критериям: формулировка научной гипотезы, описание плана эксперимента, аннотация к научной статье, устойчивость к генерации недостоверной информации («галлюцинациям»).

Таблица 4 – Сводные результаты экспертной оценки языковых моделей

Тестовое задание	Средний балл			
	ChatGPT	Claude	DeepSeek	YandexGPT
1. Формулировка научной гипотезы	9,9	9,9	8,0	6,0
2. Описание плана эксперимента	9,9	10	7,9	7,0
3. Аннотация к научной статье	10	9,9	6,9	6,0
4. Устойчивость к «галлюцинациям»	10	10	8,9	8,0
Итоговая средняя оценка	9,95	9,95	7,93	6,75

Сводная таблица позволяет визуализировать разницу между системами по отдельным параметрам и даёт возможность сопоставить сильные и слабые стороны каждой модели [7].

Наиболее высокие оценки за устойчивость к генерации недостоверной информации получили Claude и ChatGPT, что подтверждает надёжность этих моделей в научном контексте. Различия по критерию «Аннотация» выявили чувствительность моделей к

академическому стилю: только Claude и ChatGPT смогли соблюсти научную структуру и терминологическую точность.

Заключение

На основе многофакторной оценки четырёх языковых моделей по критериям научной корректности, полноты описания, академического стиля и устойчивости к «галлюцинациям», можно выделить две модели – Claude и ChatGPT – как наиболее соответствующие задачам академической и исследовательской деятельности. Обе системы получили итоговую оценку – 9,95 баллов – что отражает их высокий уровень проработки, структурности и стилистической зрелости.

Модель Claude показала большую глубину анализа и стилистическую строгость, однако может отличаться неустойчивым поведением в зависимости от входного запроса и платформы, что следует учитывать при применении в критически важных научных задачах.

Тем не менее, при выборе между двумя равными по качеству моделями, важным фактором становится техническая стабильность и доступность в различных пользовательских сценариях. В этой связи модель ChatGPT выступает как более предсказуемый, устойчиво функционирующий инструмент, подходящий как для академических учреждений, так и для индивидуальных исследователей. Она демонстрирует высокую повторяемость качества ответов, гибкость в междисциплинарных темах и хорошо интегрируется в привычные рабочие среды (браузер, API, плагины и др.).

Модели DeepSeek и YandexGPT 5 Pro-4 оказались ограниченно пригодными для задач научного моделирования. Они демонстрируют логичность и функциональную состоятельность, однако уступают в академической строгости, стилистической однородности и глубине обоснования. Их можно рекомендовать для черновой генерации идей или предварительной компоновки исследовательских текстов с последующей редактурой.

Таким образом, на основе экспертной оценки и анализа релевантных критериев, наиболее предпочтительной языковой моделью для задач планирования научного эксперимента является ChatGPT (модель GPT-4). Она демонстрирует:

- высокую точность при генерации научных текстов и гипотез;
- уверенную поддержку научного стиля и аргументации;
- возможность адаптации к специфической терминологии (например, нетканые материалы, гигроскопичность и др.);
- устойчивость к логическим и фактическим ошибкам;
- стабильность работы.

Тем не менее, как отмечается в современных исследованиях, анализирующих влияние искусственного интеллекта на образовательную и научную сферу [6], тексты, создаваемые с использованием нейросетевых моделей, требуют обязательной и всесторонней критической проверки. Несмотря на высокий уровень автоматизации, подобные системы пока не обладают способностью полноценно учитывать контекст, выявлять смысловые искажения, интерпретировать сложные понятия и корректно распознавать логические не-

соответствия. Это приобретает особую значимость в области научной коммуникации, где приоритетными являются точность, достоверность и верифицируемость информации. В связи с этим использование ИИ в научной практике должно неотъемлемо сопровождаться экспертной верификацией, независимым рецензированием и вдумчивым редакторским сопровождением.

Список использованных источников

1. Искусственный интеллект в задачах моделирования, управления, диагностики технологических процессов: монография / Веревкин А. П., Муртазин Т. М. – Москва; Вологда: Инфра-Инженерия, 2023. – 232 с.
2. Scientific discovery in the age of artificial intelligence. Nature. – 2023. DOI: 10.1038/s41586-023-06221-2.
3. ИИ в материаловедении: новая эра открытий и инноваций. – URL: <https://julienflorkin.com/ru/%D0%9C%D0%B0%D1%82%D0%B5%D1%80%D0%B8%D0%B0%D0%B%D1%8C%D0%BD%D0%B0%D1%8F-%D0%BD%D0%B0%D1%83%D0%BA%D0%B0/%D0%98%D0%98-%D0%B2-%D0%BC%D0%B0%D1%82%D0%B5%D1%80%D0%B8%D0%B0%D0%B%D0%BE%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B8/> (дата обращения 26.04.2025). – Текст : электронный.
4. Орлов, А. В., Орлова, И. А., Шустов, Ю. С., Кирсанова, Е. А. Методика проведения научного эксперимента с использованием искусственного интеллекта // Дизайн и технологии. – 2024. № 6.
5. Мирзахалилова, Д. М. Глобальные практики: использование искусственного интеллекта в высшем образовании / Д. М. Мирзахалилова, Ф. М. Камолиддинова // Новые технологии оценки качества образования: сборник материалов XVIII общероссийского с международным участием Форума Гильдии экспертов в сфере профессионального образования, Москва, 17 ноября 2024 года. – Москва: ассоциация «Гильдия экспертов в сфере профессионального образования», 2023. – С. 121–128. – EDN VAGTSZ.
6. Rew, D. A., Popova, N. G. The challenges of machine translation of academic publications. Science Editor and Publisher. 2021; 6 (2) – p. 104-112.
7. Павлюк, Е. С. Анализ зарубежного опыта влияния искусственного интеллекта на образовательный процесс в высшем учебном заведении / Е. С. Павлюк // Современное педагогическое образование. – 2020. – № 1. – С. 65–72.
8. Актуальные аспекты разработки и внедрения систем искусственного интеллекта в организациях / Мельников А. В., Бачурин А. И., Распопов А. А., Цветкова В. А. // Социальные новации и социальные науки. – Москва : ИНИОН РАН, 2021. – № 2(4). – С. 38–46. – DOI 10.31249/snsn/2021.02.03 rg/10.24069/SEP-21-01