ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 71657— 2024

Технологии искусственного интеллекта в образовании

ФУНКЦИОНАЛЬНАЯ ПОДСИСТЕМА СОЗДАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Общие положения

Издание официальное

Москва
Российский институт стандартизации
2024

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Федеральным государственным автономным образовательным учреждением высшего образования «Волгоградский государственный университет» (ФГАОУ ВО ВолГУ)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 1 октября 2024 г. № 1364-ст

4 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Оформление. ФГБУ «Институт стандартизации», 2024

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Общие положения и структура функциональной подсистемы 4
5 Реализация функциональной подсистемы на этапах создания научных публикаций 5
6 Требования к защите данных, содержащихся в функциональной подсистеме решения задач искусственного интеллекта 12
7 Управление рисками качества данных о научных публикациях 12
Библиография 13

Введение

Функциональная подсистема создания научных публикаций (далее — функциональная подсистема) входит в систему образования как важная ее часть, обеспечивающая не только документирование результатов научно-исследовательской деятельности, но и формирование соответствующих компетенций у обучающихся, в том числе школьников, студентов бакалавриата, специалитета и магистратуры, а также аспирантов.

Процесс создания научных публикаций представляет собой деятельность по документированию результатов научных исследований в виде научных публикаций в журналах, сборниках трудов конференций или других научных изданиях.

Настоящий стандарт содержит:

- основные этапы реализации функциональной подсистемы на этапах создания научной публикации;
- рекомендации и примеры по использованию конкретных систем искусственного интеллекта при решении различных задач, возникающих при создания научной публикации;
- требования к защите данных, содержащихся в функциональной подсистеме;
- рекомендации по управлению рисками качества данных в функциональной подсистеме.

Положения настоящего стандарта не исключают применения иных методов и технологий искусственного интеллекта в процессе создания научных публикаций.

Настоящий стандарт входит в серию стандартов «Технологии искусственного интеллекта в образовании».

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Технологии искусственного интеллекта в образовании

ФУНКЦИОНАЛЬНАЯ ПОДСИСТЕМА СОЗДАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Общие положения

Artificial intelligence technologies in education.
Functional subsystem for creating scientific publications. General provisions

Дата введения — 2025—01—01

1 Область применения

Настоящий стандарт устанавливает общие положения к реализации функциональной подсистемы на этапах создания научных публикаций в сфере образования и науки, в том числе с привлечением обучающихся, а также описывает системы искусственного интеллекта, предназначенные для обеспечения ее работы.

Настоящий стандарт применим в организациях профессионального образования [1] (статья 10), а также иных организациях, которые осуществляют научно-образовательную деятельность.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р ИСО/МЭК 24029-2 Интеллект искусственный. Оценка робастности нейронных сетей. Часть 2. Методология использования формальных методов

ГОСТ Р ИСО/МЭК 42001 Информационные технологии. Искусственный интеллект. Система управления

ГОСТ Р 59407 Информационные технологии. Методы и средства обеспечения безопасности. Базовая архитектура защиты персональных данных

ГОСТ Р 59897—2021 Данные для систем искусственного интеллекта в образовании. Требования к сбору, хранению, обработке, передаче и защите данных

ГОСТ Р 59898 Оценка качества систем искусственного интеллекта. Общие положения

ГОСТ Р 70944—2023 Технологии искусственного интеллекта в образовании. Функциональная подсистема организации участия в конкурсных мероприятиях с целью финансирования научной деятельности. Общие положения

ГОСТ Р 70945 Технологии искусственного интеллекта в образовании. Функциональная подсистема организации и проведения научных мероприятий. Общие положения

ГОСТ Р 70949 Технологии искусственного интеллекта в образовании. Применение искусственного интеллекта в научно-исследовательской деятельности. Варианты использования

ГОСТ Р 71484.2 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Показатели качества данных

ГОСТ Р 71484.3 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 3. Требования и рекомендации по управлению качеством данных

ГОСТ Р 71484.4 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 4. Структура процесса управления качеством данных

Издание официальное

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1 **алгоритм коллаборативной фильтрации:** Рекомендательный алгоритм, который используется для предсказания предпочтений пользователя на основе информации о его предыдущих действиях и предпочтениях, а также данных о действиях и предпочтениях других пользователей.

3.2 **большая языковая модель:** Разновидность искусственной нейронной сети, которая обучается на больших объемах данных для генерации и понимания естественного языка.

Примечание — Большая языковая модель использует технологии глубокого обучения для генерации текста, ответов на вопросы, переводов и многое другое.

3.3

заинтересованная сторона: Лицо или организация, которые могут воздействовать на осуществление деятельности или принятие решения, быть подверженными их воздействию или воспринимать себя в качестве последних.

[Адаптировано из ГОСТ Р ИСО/МЭК 42001—2024, пункт 3.2]

3.4

искусственный интеллект; ИИ: Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

[[2], статья 1]

3.5 **машинный реферат (сниппет):** Реферат, составленный с помощью компьютерной программы, содержащий сокращенное отображение информации из источника, обычно включающий заголовок, адрес и ключевую информацию из источника.

3.6

научная [научно-исследовательская] деятельность: Деятельность, направленная на получение новых научных знаний и/или применение научных знаний.

[ГОСТ 34831—2022, пункт 4.4]

3.7 **научная монография:** Научный труд, в котором с наибольшей полнотой исследована определенная научная тема, который выражен в уточнении или создании новых научных знаний и автор которого всесторонне анализирует тему или проблему и подводит итог их разработки учеными и специалистами.

3.8 **научная публикация:** Основная форма представления результатов исследования и их передачи на экспертизу научному сообществу для соотнесения с наличным массивом научного знания.

3.9 **научная статья:** Текст, содержащий теоретические и практические обобщения по конкретной научной тематике и опубликованный в научном издании.

3.10

научное издание: Издание, содержащее результаты теоретических и (или) экспериментальных исследований, а также научно подготовленные к публикации памятники культуры и исторические документы.

[ГОСТ Р 7.0.60—2020, пункт 3.2.8.2]

3.11 **научный доклад:** Публичное сообщение нового или уточненного научного знания, развернутое изложение научной темы.

3.12 **научный факт:** Знание о каком-либо явлении, достоверность которого доказана; предложение, фиксирующее знание, полученное в ходе наблюдений и экспериментов (эмпирическое знание).

3.13

организация: Лицо или группа людей, связанные определенными отношениями, имеющие ответственность, полномочия и выполняющие свои функции для достижения их целей.

Примечание — Понятие организации включает в себя, но не ограничивается следующими примерами: индивидуальный предприниматель, компания, корпорация, фирма, предприятие, орган власти, товарищество, ассоциация, благотворительные учреждения, а также их часть или их объединение, являющиеся юридическим лицом или нет, государственные или частные.

[Адаптировано из ГОСТ Р ИСО/МЭК 42001—2024, пункт 3.1]

3.14 **поисковый робот (веб-скрейпер):** Программное обеспечение, которое автоматически извлекает данные с веб-сайтов.

Примечание — Веб-скрейперы используются для сбора информации с различных онлайн-ресурсов. Процесс веб-скрейпинга включает в себя загрузку веб-страниц, анализ HTML-кода страницы и извлечение нужной информации, такой как текст, изображения, ссылки и другие данные. Эти данные могут быть сохранены в файле или базе данных для дальнейшего анализа или использования.

3.15 **производные научные произведения:** Научные произведения, представляющие собой переработку (перевод, обработку, обзор или иную переработку) другого научного произведения.

3.16

система искусственного интеллекта: Техническая система, в которой используются технологии искусственного интеллекта и обладающая искусственным интеллектом.

[ГОСТ Р 59276—2020, пункт 3.16]

3.17 **составные научные произведения:** Научные произведения, представляющие собой по подбору и расположению материала результат творческого труда по определенной научной проблематике.

3.18

технологии искусственного интеллекта: Совокупность технологий, включающая в себя компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта.

[[2], статья 1]

3.19

функциональная подсистема: Взаимосвязанная совокупность систем искусственного интеллекта, направленных на решение определенных задач.

[ГОСТ Р 70950—2023, пункт 3.20]

4 Общие положения и структура функциональной подсистемы

4.1 Общие положения

Положения настоящего стандарта описывают использование систем ИИ с целью повышения результативности создания научных публикаций. Применение систем ИИ должно осуществляться этичным образом[1)] и не должно заменять творческую научно-исследовательскую деятельность человека.

Организация должна разработать и обеспечить соблюдение критериев и порядка применения технологий ИИ, руководствуясь положениями разделов 6 и 7, целями конкретной научной публикации и условиями ее создания, в рамках функциональной подсистемы, определенной в 4.2 и в разделе 5.

Научные публикации разделяют на следующие виды:

- научные статьи;
- научные доклады;
- научные монографии;
- производные научные произведения;
- составные научные произведения;
- охраняемые результаты интеллектуальной деятельности (патенты на изобретения, полезные модели, промышленные образцы, селекционные достижения, свидетельства о государственной регистрации программ для электронных вычислительных машин, баз данных, топологий интегральных микросхем) и т. д.

Примечание — Заявления на регистрацию подаются в установленной форме регистратору/в Федеральную службу по интеллектуальной собственности (для повышения результативности этого этапа следует применять положения 5.4). Для продвижения и мониторинга использования зарегистрированных результатов интеллектуальной деятельности следует применять положения 5.5 и 5.6.

4.2 Структура функциональной подсистемы

Функциональная подсистема создания научной публикации предназначена для решения возникающих задач с применением технологий ИИ. Структура функциональной подсистемы состоит из следующих элементов, представляющих собой этапы создания и продвижения научной публикации, для повышения результативности которых могут быть использованы системы ИИ. Организация должна определить, в ходе каких из перечисленных ниже этапов научного мероприятия будут применяться технологии ИИ, определенные в разделе 5:

- сбор исходных данных, анализ научных изданий в рамках предметной области публикации и формулирование ее названия;
- написание научной публикации;
- выбор научного издания для публикации;
- подготовка публикации по требованиям научного издания для публикации;
- продвижение научной публикации;
- мониторинг научной публикации.

Организации следует применять технологии ИИ в рамках перечисленных выше этапов, когда не известна математическая модель, обеспечивающая результат с точностью, соответствующей требованиям заинтересованных сторон по созданию научной публикации. При этом методы ИИ используются для того, чтобы по имеющейся информационной модели построить математическую модель и уже ее использовать для решения задачи.

Дополнительная информация по возможностям применения систем ИИ для решения задач на этапах организации и проведения научных мероприятий приведена в разделе 5.

[1)] Документом, устанавливающим общие этические принципы и стандарты поведения, которым следует руководствоваться при использовании искусственного интеллекта, является Кодекс этики в сфере искусственного интеллекта.

5 Реализация функциональной подсистемы на этапах создания научных публикаций

Данная функциональная подсистема направлена на обеспечение результативного применения ИИ в научно-исследовательской деятельности[1)], системно связана с функциональными подсистемами организации и проведения научных мероприятий[2)], а также с функциональной подсистемой организации участия в конкурсных мероприятиях с целью финансирования научной деятельности[3)].

В настоящем разделе описаны этапы реализации настоящей функциональной подсистемы. По каждому из этапов приведены варианты использования систем ИИ, их задачи, а также предназначенные для решения этих задач системы ИИ более низкого уровня совместно с соответствующими моделями, процедурами и т. д.

В процессе создания научных публикаций следует реализовывать изложенные ниже этапы последовательно, однако в определенных ситуациях в зависимости от структуры исследования и особенностей подготовки публикации некоторые задачи одного этапа могут решаться и на другом этапе.

Каждая система ИИ, входящая в настоящую функциональную подсистему, помимо своих специфических задач (если применимо) решает определенную совокупность типовых задач, идентичных приведенным в ГОСТ Р 70944—2023 (раздел 5).

5.1 Сбор исходных данных, анализ научных изданий в рамках предметной области публикации и формулирование ее названия

Этап предполагает сбор исходных данных, анализ научных изданий в рамках предметной области научной публикации с целью исследования существующего научного задела, формулирование названия научной публикации. Формулирование названия научной публикации позволит ее найти после опубликования по точно подобранным тезаурусным статьям, сформировать представление о содержании исследования в сжатом виде, для чего может быть использована система ИИ, приведенная в 5.1.1.

5.1.1 Система ИИ формулирования названия научной публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) поисковые и информационно-аналитические системы

Такие системы, как Российский индекс научного цитирования (РИНЦ), предоставляют доступ к базам данных научных статей, журналов и других публикаций. Они используют алгоритмы машинного обучения и обработки естественного языка для индексации и классификации статей по темам;

2) системы автоматического сбора информации

Могут использоваться для автоматического сбора данных из различных источников, таких как базы данных, репозитории, журналы и другие ресурсы. Это может включать методы сканирования веб-страниц (например, применение поисковых роботов или веб-скрейперов), парсинга данных и алгоритмы машинного обучения для сбора информации, соответствующей требованиям научного издания. Применяются для извлечения таких данных, как заголовки статей, авторы, аннотации, фрагменты текста и ключевые слова;

3) системы обработки естественного языка, включая большие языковые модели

Системы обработки естественного языка, включая большие языковые модели, могут использоваться для анализа и классификации текстовых данных. Они могут помочь в выделении ключевых слов, генерации машинных рефератов (сниппетов), тематическом моделировании, анализе тональности и других аспектах текста, чтобы исследователи могли сформулировать тему научной публикации;

4) рекомендательные системы

Такие системы используют алгоритмы коллаборативной фильтрации, анализа контента и другие алгоритмы для предоставления рекомендаций относительно научных публикаций, основываясь на интересах и предпочтениях исследователя, а также на анализе его предыдущих публикаций или изучаемых тем. Они могут помочь исследователям найти релевантную информацию для формулирования названия научной публикации.

---

1) См. ГОСТ Р 70949—2023.

2) См. ГОСТ Р 70945—2023.

3) См. ГОСТ Р 70944—2023.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться следующие методы хранения, включающие применение технологий ИИ:

1) системы управления базами данных (СУБД)

Предоставляют возможность хранить и организовывать информацию в структурированном формате. Они обеспечивают эффективное хранение и управление большим объемом данных, включая научные статьи, исследовательские данные и другие ресурсы;

2) облачные хранилища

Платформы облачного хранения данных предоставляют возможность хранить большие объемы информации в удобном и доступном формате. Они обеспечивают высокую масштабируемость и надежность хранения данных;

3) цифровые библиотеки

Предоставляют инструменты для создания и управления электронными коллекциями научных ресурсов. Они обеспечивают хранение, поиск и доступ к научным статьям, отчетам, монографиям и другим материалам;

4) репозитории и архивы

Существуют онлайн-репозитории и архивы, где исследователи могут хранить свои научные статьи и другие материалы. Эти ресурсы предоставляют доступ к открытым публикациям и облегчают поиск и доступ к академическим материалам;

5) системы управления библиографической информацией

Предназначены для организации и хранения библиографических данных, включая ссылки на статьи, журналы и другие ресурсы. Они облегчают управление литературой и помогают отслеживать предполагаемые научные издания для публикации;

6) системы управления знаниями

Эти системы позволяют организовать и хранить знания в структурированной форме. Они могут содержать информацию о научных концепциях, теориях, результаты экспериментов, требования научных изданий для публикаций и другие ресурсы, которые могут быть полезны для формулирования темы научной публикации;

7) семантические хранилища

Семантические хранилища данных используют онтологии и семантические технологии для организации и хранения информации. Они позволяют устанавливать связи между различными элементами данных и предоставляют возможность более глубокого анализа и поиска информации;

8) системы контент-менеджмента

Обеспечивают удобное управление содержанием и публикацией научных статей. Они позволяют исследователям организовывать и хранить свои научные работы, а также облегчают публикацию и обмен информацией с другими учеными;

9) графовые базы данных

Графовые базы данных представляют данные в виде графа, где узлы представляют сущности, а ребра — их связи. Это позволяет эффективно хранить и извлекать информацию о взаимосвязях между различными элементами данных, что может быть полезно при формулировании в работе над научной публикацией;

10) системы управления проектами

Некоторые системы управления проектами, могут быть использованы для управления процессом публикации как проектом, а также для хранения информации о научных публикациях.

в) Актуализация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматического поиска

Могут использоваться для автоматического мониторинга и поиска новых научных статей, публикаций и источников информации. Данные системы могут использовать алгоритмы машинного обучения и обработки естественного языка для определения актуальности и релевантности новых материалов для работы над научной публикацией. В случае, если используется специализированная база данных для хранения информационных ресурсов, результаты поиска используются для автоматического ее обновления;

2) системы анализа социальных медиа

Анализ социальных медиа может помочь в определении актуальных исследовательских тем, трендов и новых источников информации. Системы анализа социальных медиа могут использовать

информацию из наукометрических баз данных систем, таких как РИНЦ, чтобы определить активность исследователей в определенной области и предоставить рекомендации относительно актуальных исследовательских тем;

3) системы извлечения информации

Автоматически извлекают научные факты и данные из различных источников, таких как научные издания, отчеты или базы данных. Они могут помочь исследователям быстро получить необходимую информацию для подтверждения или обновления своих исследовательских результатов;

4) системы автоматической аннотации

Могут создавать краткие аннотации научных статей и других научных публикаций. Они могут помочь исследователям быстро оценить актуальность статьи и определить, стоит ли ее изучать более подробно.

Примечание — Например, платформа Toowrite Abstract для подготовки аннотаций;

5) системы автоматического перевода

При необходимости получения актуальной информации об исследованиях или публикациях на других языках системы автоматического перевода могут помочь исследователям получить доступ к актуальной информации, которая может быть недоступна на их родном языке.

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматического реферирования

Данные системы могут извлекать ключевую информацию из большого количества научных публикаций и, сжимая ее, представлять в формате краткого реферата. Они помогают исследователям быстро получить обзор научных работ по интересующей теме;

2) системы кластеризации и категоризации

Используют алгоритмы машинного обучения для группировки связанных тематических публикаций в кластеры или категории. Они могут помочь исследователям найти связанные работы и определить общую тему для своей публикации;

3) системы генерации текста

Современные системы генерации текста на основе ИИ могут помочь исследователям сформулировать тему научной публикации, предложив различные варианты заголовков или аннотаций. Для обеспечения этичного применения систем ИИ (см. 4.1) окончательную редакцию темы научной публикации должен формулировать исследователь самостоятельно, проанализировав предложенные системой генерации текста варианты заголовков и аннотаций.

д) Предоставление информационных ресурсов пользователям

Поскольку практически все перечисленные выше системы оснащены достаточно совершенными интерфейсами взаимодействия с пользователями, выделять отдельные системы ИИ, обеспечивающие предоставление информационных ресурсов пользователям, не целесообразно. Исходя из этого порядок решения этой задачи на следующих этапах процесса создания научных публикаций приводиться не будет.

5.2 Написание научной публикации

На данном этапе осуществляется выбор стиля научной публикации, литературный обзор источников на основании выдвинутой гипотезы, обоснование гипотезы, формулирование результатов, дискуссия относительно результатов, написание аннотации или автореферата научной публикации, выбор ключевых слов и оформление списка литературы, для чего может быть использована система ИИ, приведенная в 5.2.1.

5.2.1 Система ИИ поддержки написания научной публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) поисковые и информационно-аналитические системы

Описание данных систем приведено в 5.1.1, перечисление а) 1);

2) системы автоматического реферирования

Описание данных систем приведено в 5.1.1, перечисление г) 1);

3) системы автоматического извлечения информации

Автоматически извлекают структурированную информацию из неструктурированных данных и различных источников, таких как научные издания, отчеты или базы данных. Это может включать полу-

чение ключевых научных фактов, связей между сущностями, цитат и других важных данных. Могут помочь исследователям быстро получить необходимую информацию для подтверждения или обновления своих исследовательских результатов;

4) системы обработки естественного языка, включая большие языковые модели

Позволяют анализировать и понимать текстовую информацию, что может быть полезным при обработке больших объемов научных публикаций. Они могут выполнять задачи, такие как кластеризация публикаций по темам, анализ тональности, выделение ключевых слов и фраз, автоматическое резюмирование текста и т. д.;

5) системы машинного обучения

Методы машинного обучения могут быть применены для создания моделей, которые могут классифицировать и фильтровать научные статьи по заданным параметрам. Например, можно использовать модели классификации для определения релевантности статей к заданной теме или для оценки качества и достоверности научных изданий.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление б).

в) Актуализация информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление в).

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматической классификации

Могут классифицировать научные статьи, отчеты и другие источники информации по темам, областям исследования или другим параметрам. Они помогают исследователям организовать и структурировать свои ресурсы для более эффективного доступа к ним;

2) системы извлечения информации

Могут автоматически извлекать научные факты, данные и другую сущностную информацию из научных статей, отчетов или баз данных. Они могут быть полезны для автоматического сбора и анализа данных, необходимых для написания научной публикации;

3) системы генерации текста

Могут создавать варианты фрагментов текстов научных статей или отчетов на основе имеющихся данных и шаблонов. Для обеспечения этичного применения систем ИИ (см. 4.1) окончательную редакцию как текста научной публикации в целом, так и его любых фрагментов должен формулировать исследователь самостоятельно, проанализировав предложенные системой генерации текста варианты фрагментов текста;

4) системы проверки подлинности исследования

Используют алгоритмы машинного обучения для проверки подлинности исследования, выявления плагиата и определения его источников. Они помогают исследователям обеспечить интеллектуальную чистоту своих публикаций;

5) системы визуализации данных

Могут преобразовывать научные данные в графики, диаграммы и другие визуальные формы, делая их более доступными и понятными. Они помогают исследователям представить свои результаты в наглядной форме.

5.3 Выбор научного издания для публикации

На данном этапе анализируются научные издания, освещающие достижения в предметной области проведенного исследования. Научное издание должно точно соответствовать тематике исследования: его цели, задаче, тематическому охвату. Для этого могут быть использованы системы ИИ, приведенные в 5.3.1.

5.3.1 Система ИИ выбора научного издания для публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматического поиска

Могут автоматически искать научные издания, базы данных и другие ресурсы, связанные с темой исследования. Они могут использовать ключевые слова, тематические кластеры или другие параметры для поиска предполагаемых научных изданий для публикации;

2) рекомендательные системы

Используют алгоритмы машинного обучения для анализа предпочтений и поведения пользователей, чтобы рекомендовать предполагаемые для публикации научные издания. Они могут учитывать факторы, такие как репутация предполагаемого для публикации научного издания, цитирования, факторы влияния и другие параметры;

3) системы анализа репутации предполагаемых для публикации научных изданий

Могут анализировать репутацию научных изданий на основе различных метрик, таких как индекс цитирования, импакт-фактор, Н-индекс и другие. Они помогают исследователям оценить качество и влияние издания перед выбором для публикации;

4) системы анализа отзывов и рецензий

Могут анализировать отзывы и рецензии на научные издания, оставленные другими исследователями. Они помогают получить представление о качестве публикаций, процессе рецензирования и репутации научного издания;

5) системы управления библиографической информацией

Позволяют исследователям организовывать и управлять своими библиографическими данными, включая информацию о научных изданиях, требованиях к публикации и сроках подачи. Они помогают исследователям быть в курсе актуальных научных изданий и сроков подачи в них материалов.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление б).

в) Актуализация информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление в).

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы обработки естественного языка, включая большие языковые модели

Могут использоваться для анализа содержания научных статей, аннотаций, ключевых слов и других метаданных. Это может включать методы обработки естественного языка, машинное обучение и алгоритмы классификации, чтобы определить соответствие статьи тематике и требованиям конкретного научного издания;

2) рекомендательные системы

Могут использоваться для создания рекомендаций по выбору научного издания для публикации на основе анализа предпочтений и поведения исследователя. Это может включать анализ предыдущих публикаций, цитирований, авторов и других факторов для предоставления персонализированных рекомендаций;

3) системы интеграции данных

Могут использоваться для объединения и интеграции информации из различных источников, таких как базы данных, репозитории, журналы и другие ресурсы. Это помогает создать единую точку доступа к информации и облегчает выбор научного издания для публикации на основе собранной информации;

4) системы автоматического сопоставления

Могут использоваться для автоматического сопоставления требований журналов с характеристиками научной статьи. Это может включать анализ ключевых слов, тематического соответствия, требований к форматированию и других факторов для определения подходящих журналов для публикации.

5.4 Подготовка публикации по требованиям научного издания для публикации

Данный этап предполагает подготовку научной публикации в строгом соответствии с требованиями выбранного научного издания для публикации в зависимости от ее типа: требования к объему публикации, количеству рисунков, количеству источников и т. д. Для этого может быть использована система ИИ, приведенная в 5.4.1.

5.4.1 Система ИИ подготовки публикации по требованиям научного издания для публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматического сбора информации

Описание данных систем приведено в 5.1.1, перечисление а) 2);

2) системы обработки и структурирования данных

Могут использоваться для обработки и структурирования собранных данных в соответствии с требованиями научного издания для публикации. Это может включать анализ текстовых данных, классификацию библиографических источников, извлечение ключевых слов и другие методы обработки естественного языка для подготовки информации к публикации;

3) системы проверки соответствия требованиям

Могут использоваться для проверки соответствия подготавливаемой научной публикации требованиям научного издания для публикации. Это может включать анализ формата, стиля, цитирований и других параметров, чтобы убедиться, что статья соответствует требованиям журнала перед ее представлением;

4) системы автоматической регистрации

Могут использоваться для автоматической регистрации научной публикации в соответствующем научном издании для публикации. Это может включать заполнение форм, загрузку файлов, проверку метаданных и другие процессы, связанные с регистрацией статьи.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление б).

в) Актуализация информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление в).

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматического форматирования научной публикации

Могут быть применены для автоматического извлечения с сайта научного издания для публикации требований к форматированию и обеспечения соответствия им шрифтов, размеров полей, вида таблиц, формул и т. д. научной публикации;

2) системы автоматической генерации аннотаций

Могут быть использованы для автоматической генерации вариантов аннотаций научных статей в соответствии с требованиями научного издания для публикации. Для обеспечения этичного применения систем ИИ (см. 4.1) окончательную редакцию аннотаций научных статей должен формулировать исследователь самостоятельно, проанализировав предложенные системой генерации варианты;

3) системы автоматической проверки и исправления текста

Могут быть применены для автоматической проверки грамматики, орфографии, стиля и других аспектов текста научной публикации. Это помогает улучшить качество текста и обеспечить соблюдение требований научного издания для публикации.

5.5 Продвижение научной публикации

Данный этап предполагает размещение научной публикации в социальных и профессиональных медиа, в открытых архивах и других информационных ресурсах, позволяющее включить опубликованные результаты исследования в систему научных коммуникаций, для чего может быть использована система ИИ, приведенная в 5.5.1.

5.5.1 Система ИИ продвижения научной публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы анализа социальных медиа

Могут быть применены для сбора информации об активности и взаимодействии в социальных медиа. Алгоритмы машинного обучения и обработки естественного языка могут использоваться для определения релевантных тематик, идентификации влиятельных пользователей, анализа настроений и мнений, а также для создания персонализированных рекомендаций и контента для продвижения научной публикации;

2) системы автоматического поиска и оптимизации ключевых слов

Могут быть применены для автоматического поиска и оптимизации ключевых слов и фраз, связанных с научной публикацией. Алгоритмы машинного обучения и анализа текста могут использоваться

для определения наиболее релевантных и популярных ключевых слов, которые помогут улучшить видимость публикации в поисковых системах и базах данных.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление б). Кроме того, могут быть применены положения 5.4.1, перечисления а) 2), 3) и 4) для автоматизации регистрации научной публикации в различных хранилищах (например, репозиториях и архивах) для ее продвижения.

в) Актуализация информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление в).

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы автоматической рекомендации

Могут быть использованы для создания систем автоматической рекомендации научных публикаций. Это может включать методы коллаборативной фильтрации, контент-фильтрации и гибридных подходов, которые основываются на анализе предпочтений пользователей, сходстве контента и других факторах. Такие системы помогают предложить научную публикацию целевой аудитории, увеличивая ее видимость и охват;

2) системы автоматической генерации контента

Могут быть использованы для автоматической генерации контента, связанного с научной публикацией. Это может включать создание аннотаций, пресс-релизов, социальных медиа-постов и других материалов, которые помогут привлечь внимание к публикации и повысить ее видимость.

5.6 Мониторинг научной публикации

На данном этапе проводится отслеживание научной публикации посредством изучения ее цитируемости в международных научных базах данных, социальных и профессиональных медиа и других информационных системах, для чего может быть использована система ИИ, приведенная в 5.6.1.

5.6.1 Система ИИ мониторинга научной публикации

а) Сбор и регистрация информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы анализа цитирования и влияния

Могут быть применены для анализа цитирования и влияния научной публикации. Алгоритмы машинного обучения и обработки данных могут использоваться для отслеживания количества цитирований, идентификации цитирующих работ, определения рейтинга влияния публикации и других метрик. Такие системы помогают оценить реакцию на публикацию в научном сообществе и измерить ее влияние;

2) системы мониторинга социальных медиа

Могут быть использованы для мониторинга активности и обсуждений о научной публикации в социальных медиа. Алгоритмы обработки естественного языка и анализа настроений могут использоваться для отслеживания упоминаний, комментариев, репостов и других форм взаимодействия с публикацией. Такие системы позволяют оценить общественную реакцию и восприятие публикации;

3) системы автоматического оповещения и уведомления

Могут быть использованы для автоматического оповещения и уведомления об активности и изменениях, связанных с научной публикацией. Это может включать отправку уведомлений о новых цитированиях, комментариях, упоминаниях в социальных медиа, изменениях в статистике и других событиях. Такие системы помогают быть в курсе последних обновлений и реакции на публикацию.

б) Хранение информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление б).

в) Актуализация информационных ресурсов

Для решения данной задачи могут использоваться системы и методы, изложенные в 5.1.1, перечисление в).

г) Обработка информационных ресурсов

Для решения данной задачи могут быть использованы следующие системы ИИ:

1) системы анализа статистических данных

Могут применяться для анализа статистических данных, связанных с научной публикацией. Это может включать анализ количества загрузок, просмотров, скачиваний, комментариев, рейтингов и других метрик. Алгоритмы машинного обучения и статистического анализа данных могут использоваться для выявления трендов, паттернов и изменений в статистике публикации;

2) системы автоматического анализа конкурентов

Могут быть применены для анализа конкурентов и сопоставления публикации с другими научными работами в той же области. Алгоритмы машинного обучения и обработки данных могут использоваться для сравнительного анализа, определения уникальности и оригинальности публикации, а также для выявления возможных сотрудничеств и партнерств.

6 Требования к защите данных, содержащихся в функциональной подсистеме решения задач искусственного интеллекта

6.1 К защите данных, содержащихся в функциональной подсистеме создания научных публикаций следует применять требования ГОСТ Р 59897—2021 (раздел 12).

6.2 Общий подход по установлению требований по защите сведений, подлежащих защите в соответствии с действующим законодательством, используемых в функциональных подсистемах решения задач ИИ, должен соответствовать требованиям ГОСТ Р 59407. При необходимости защиты персональных данных, используемых в функциональных подсистемах решения задач ИИ, требования должны быть установлены в соответствии с положениями [3] — [7].

7 Управление рисками качества данных о научных публикациях

7.1 Следует предусмотреть наличие в функциональной подсистеме возможностей управления рисками качества данных, например:

- способность на основе интеллектуального анализа и машинного обучения выявлять и прогнозировать возможные будущие риски внутри и за пределами функциональной подсистемы в отношении точности, полноты, достоверности, актуальности, доступности, соответствия требованиям, конфиденциальности, эффективности, понятности, переносимости и возможности восстановления данных;
- способность анализировать и оценивать риски качества данных с точки зрения причин, угроз и последствий;
- возможность реализации вариантов управления рисками качества данных (предупреждение, предотвращение, снижение или минимизация).
- установление пороговых значений ошибок и уровней рисков качества данных, за пределами которых возникает необходимость исправления низкого качества данных;
- возможность применять методы повышения качества данных.

7.2 Процесс управления рисками, связанный с каждым показателем качества данных, должен быть интегрирован в практику управления качеством данных, структуру и процессы функциональной подсистемы.

7.3 Следует определить роли и обязанности в отношении управления рисками качества данных и довести данную информацию до сведения всех пользователей, обеспечивающих функционирование функциональной подсистемы.

7.4 Периодичность подготовки и распространения отчетов по управлению рисками должна быть установлена таким образом, чтобы облегчить принятие обоснованных решений о рисках лицами, принимающими решения.

7.5 Управление качеством и рисками функциональной подсистемы следует осуществлять в соответствии с ГОСТ Р 59898, ГОСТ Р ИСО/МЭК 42001, ГОСТ Р 71484.2, ГОСТ Р 71484.3, ГОСТ Р 71484.4, ГОСТ Р ИСО/МЭК 24029-2 и [8].

Библиография

[1] Федеральный закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»

[2] Указ Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»

[3] Федеральный закон от 23 августа 1996 г. № 127-ФЗ «О науке и государственной научно-технической политике»

[4] Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных»

[5] Постановление Правительства Российской Федерации от 1 ноября 2012 г. № 1119 «Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных»

[6] Приказ ФСТЭК России от 18 февраля 2013 г. № 21 «Об утверждении Состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных»

[7] Приказ ФСБ России от 10 июля 2014 г. № 378 «Об утверждении Состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных с использованием средств криптографической защиты информации, необходимых для выполнения установленных Правительством Российской Федерации требований к защите персональных данных для каждого из уровней защищенности»

[8] ПНСТ 776—2022 Информационные технологии. Интеллект искусственный. Управление рисками

УДК 004.8:004.6:006:354 ОКС 35.240.90

Ключевые слова: научная публикация, научная деятельность, научное издание, функциональная подсистема, система искусственного интеллекта

Редактор *Е.В. Якубова*
Технический редактор *И.Е. Черепкова*
Корректор *Л.С. Лысенко*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 04.10.2024. Подписано в печать 16.10.2024. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 2,32. Уч.-изд. л. 1,90.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru