

Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»

На правах рукописи

Шадрина Олеся Владимировна

**СТРУКТУРНО-СЕМАНТИЧЕСКИЕ ОСОБЕННОСТИ ТЕРМИНОЛОГИИ
ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»
(КОРПУСНОЕ ИССЛЕДОВАНИЕ)**

Специальность 5.9.8. – Теоретическая, прикладная
и сравнительно-сопоставительная лингвистика

ДИССЕРТАЦИЯ

на соискание ученой степени кандидата филологических наук

Научный руководитель:
доктор филологических наук,
доцент И. М. Петрова

Москва 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1. ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ ТЕРМИНОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»	15
1.1. Терминоведение как актуальная область современной лингвистики	15
1.2. Терминология как объект лингвистического анализа	18
1.3. Определение понятия «термин»	23
1.4. Способы терминообразования	28
1.5. Эволюция терминологии в области искусственного интеллекта: основные этапы формирования	34
1.6. Гармонизация терминологии ИИ в системе международной и национальной стандартизации	38
ВЫВОДЫ ПО ГЛАВЕ 1	43
ГЛАВА 2. МЕТОДОЛОГИЯ ПРОВЕДЕНИЯ КОРПУСНОГО ИССЛЕДОВАНИЯ ТЕРМИНОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»	45
2.1. Инструментарий и методы корпусного анализа терминологии	45
2.1.1. Определение понятий «языковой корпус» и «корпусный менеджер»	45
2.1.2. Обоснование выбора программного комплекса AntConc как инструментария корпусного анализа и обзор его функциональных модулей	48
2.1.3. Параметризация терминологического аппарата: алгоритмы и метрики статистической оценки данных	52
2.2. Дискурсивное пространство научной статьи как среда функционирования и верификации терминосистем	58
2.2.1. Функциональный статус научной статьи как первичного источника данных в прикладной лингвистике	58
2.2.2. Жанровая детерминация корпуса исследования: критерии и принципы формирования эмпирической базы на материале научных статей	63
2.3. Алгоритм и процедурные этапы реализации корпусного анализа в программной среде AntConc	68

2.4. Формирование онлайн-гlossария терминов в области ИИ как отражение структуры тематического поля	81
ВЫВОДЫ ПО ГЛАВЕ 2	86
ГЛАВА 3. СТРУКТУРНО-СЕМАНТИЧЕСКИЕ ОСОБЕННОСТИ ТЕРМИНОСИСТЕМЫ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»	88
3.1. Морфологические и морфолого-синтаксические модели образования однословных терминов искусственного интеллекта	89
3.2. Синтаксическая деривация терминов искусственного интеллекта: двух- и многокомпонентные модели	101
3.3. Семантические особенности образования терминов ИИ	113
3.3.1. Заимствование, привлечение и транстерминологизация	114
3.3.2. Метонимия и метафоризация	120
3.3.3. Термины-эпонимы	139
3.3.4. Синонимия и полисемия	151
ВЫВОДЫ ПО ГЛАВЕ 3	168
ЗАКЛЮЧЕНИЕ	171
СПИСОК ЛИТЕРАТУРЫ	177
ПРИЛОЖЕНИЕ № 1	199
ПРИЛОЖЕНИЕ № 2	200
ПРИЛОЖЕНИЕ № 3	201

ВВЕДЕНИЕ

Представленное исследование посвящено изучению структурно-семантических особенностей терминологии предметной области «искусственный интеллект» (далее – ИИ) в русле приоритетных направлений современной лингвистики – терминоведения и корпусной лингвистики.

Роль термина как базовой единицы номинации и концептуализации в формировании терминосистемы отрасли знания является устойчивым предметом лингвистических исследований [Cabré 1999; Буянова 2012; Лейчик 2022; Шелов 2024 и др.]. В пределах собственной терминосистемы термины обозначают объекты, процессы и свойства конкретного домена, кодируя релевантные признаки и отношения между понятиями, тем самым обеспечивая точность, воспроизводимость и межсубъектную сопоставимость научной коммуникации. Особую значимость приобретает сопоставительный анализ терминосистем родственных и неродственных языков, позволяющий оценивать рациональность и адекватность деривации, жизнеспособность и адаптивность терминологических единиц, а также выявлять механизмы межъязыковой гармонизации и локализации [Серкова 2020; Чжао 2022].

Актуальность исследования обусловлена сложившейся научной парадигмой, в рамках которой структурно-семантические свойства терминологии предметной области «искусственный интеллект» исследуются преимущественно на материале английского языка. Являясь языком-донором и *lingua franca* международного научного сообщества, английский язык во многом определяет направления формирования терминосистемы ИИ [Smith 2024; Dabbara 2025]. Кроме того, активно развиваются исследования, посвященные анализу особенностей перевода терминологии ИИ на русский, китайский, испанский и другие языки [Gómez-Pérez, Machado 2023; Кузнецова, Фуканчик 2023; Баланчуков 2024], а также изучению семантических процессов привлечения терминов и транстерминологизации [Мусаева 2022; Термины и понятия 2024]. В условиях нарастающих процессов глобализации и цифровизации науки изучение механизмов формирования, адаптации и стандартизации терминологии ключевой технологической области приобретает особое теоретическое и прикладное значение.

Вместе с тем термины ИИ ранее не рассматривались комплексно, поэтому целостная картина терминологии этой предметной сферы, учитывающая все многообразие явления, еще формируется. Кроме того, отсутствуют фундаментальные работы, посвященные структурно-семантическим особенностям терминосистемы ИИ, что создает определенные трудности в раскрытии ее функционального и деривационного потенциала [Фомина 2022; Гришин 2023].

В этом контексте в современном терминоведении особую значимость приобретает изучение потенциала цифровых методик, применяемых при анализе письменного академического дискурса. Настоящее исследование проведено с опорой на цифровой инструмент AntConc (версия 4.3.1), позволяющий оптимизировать процесс формирования и обработки эмпирической базы исследования [Сулейманова 2020; Anthony 2023]. Использование подобных технологий не только повышает точность и эффективность исследования, но и способствует более глубокому пониманию структуры и динамики терминологических систем. В связи с этим актуальность настоящей работы значительно возрастает благодаря демонстрации синергии классических лингвистических методов и передовых цифровых технологий, что отражает современные тенденции цифровизации в науке и обеспечивает более глубокий и точный анализ терминологии ИИ.

Объектом настоящего исследования является терминология предметной области «искусственный интеллект» в научном дискурсе, репрезентированная в корпусе высокорейтинговых научных публикаций и соотнесенная с действующими международными стандартами ISO/IEC.

Предметом исследования выступают структурно-семантические особенности терминологических единиц предметной области «искусственный интеллект», проявляющиеся в их морфологической, синтаксической и морфолого-синтаксической организации, а также в процессах семантической деривации (эпонимизация, метонимизация, метафоризация, формирование синонимии и полисемии), выявляемые средствами корпусной лингвистики на материале английского языка.

Цель исследования – выявить и систематизировать структурно-семантические особенности терминологии предметной области «искусственный интеллект» в научном дискурсе на основе разработанной методики корпусного анализа.

Достижение поставленной цели потребовало решения следующих **исследовательских задач:**

- установить ключевые интралингвистические (морфологические, синтаксические, семантические) и экстралингвистические (технологические, институциональные, социокоммуникативные) факторы развития терминологии ИИ в научном дискурсе;
- изучить существующие подходы к анализу терминологии ИИ;
- разработать и апробировать корпусно-ориентированную методику исследования структурно-семантических особенностей терминологии ИИ;
- провести анализ структурно-семантических особенностей терминосистемы ИИ на материале английского языка;
- выявить и систематизировать деривационные механизмы и модели терминообразования в исследуемой области;
- обобщить результаты и представить системное описание структурно-семантической специфики терминосистемы ИИ на материале английского языка.

Теоретическая значимость исследования обусловлена разработкой и обоснованием комплексных подходов к изучению структурно-семантических механизмов формирования и функционирования терминологии предметной области «искусственный интеллект», отражающей динамику обогащения новой терминосистемы под влиянием стремительного развития цифровых технологий. Особое внимание уделено взаимодействию внутренних лингвистических факторов, включая деривационные механизмы и семантическое терминообразование, а также внешних экстралингвистических условий (технологических инноваций, международного статуса английского как *lingua franca*, норм публикации и стандартизации). Полученные результаты расширяют теоретические границы терминологии, лексикологии и корпусной лингвистики, одновременно формируя методологическую базу для будущих исследований в области автоматизации процессов извлечения и унификации терминов ИИ.

Практическая ценность работы заключается в возможности использования ее результатов при проведении исследований в области терминологии ИИ, в курсах лекций и семинаров по корпусной и компьютерной лингвистике, терминоведению и терминографии, в спецкурсах по когнитивистике и прикладной лингвистике, в

написании учебных пособий, подготовке курсовых, проектных и выпускных квалификационных работ, а также исследований по смежной тематике. Онлайн-гlossарий, составленный по результатам исследования, может найти применение в переводческой деятельности, при составлении научной и технической документации, написании научных статей. Кроме того, в работе предлагается описание алгоритма использования цифровых инструментов, что может быть полезно для оптимизации процесса сбора материала при создании эмпирической базы в сходных исследованиях на материале письменного академического дискурса.

Материал исследования опирается на две взаимодополняющие эмпирические компоненты: 1) авторский корпус, включающий 897 рецензируемых англоязычных статей из журналов «Белого списка», отобранных по критериям тематической релевантности и жанровой однородности; 2) терминологическая выборка из 1648 единиц, извлеченная с помощью корпусного менеджера AntConc (версия 4.3.1) и представленная в Приложении № 1 в виде 1648 словарных статей.

Источником верификации полученных результатов выступают:

- толковые и этимологические словари английского языка: Longman Dictionary of Contemporary English URL, Merriam-Webster Dictionary URL, Macmillan Dictionary URL, Cambridge Dictionary URL, Cambridge Dictionary URL, Online Etymology Dictionary URL, Oxford English Dictionary URL, The Concise Oxford Dictionary of English Etymology (2003);

- терминологические словари и справочники: The International Dictionary of Artificial Intelligence (1999), Толковый словарь по искусственному интеллекту (1992), Глоссариум по искусственному интеллекту: 2500 терминов (2023; 2024), Искусственный интеллект: электронный терминологический словарь (2023), Англо-русский толковый словарь по искусственному интеллекту и робототехнике (2022), Machine Learning Glossary URL;

- международные стандарты по искусственному интеллекту, разработанные совместной технической комиссией ISO/IEC JTC 1/SC 42 «Искусственный интеллект» Международной организации по стандартизации (ISO) и Международной электротехнической комиссии (IEC) [International Organization 2025].

Методологическую основу исследования составляют положения, представленные в работах отечественных и зарубежных лингвистов, связанные со следующими аспектами:

- определением особенностей термина (слова или словосочетания для наименования специальных понятий) [Винокур 1939; Wüster 1979; Головин, Кобрин 1987; Sager 1990; Cabré 1999; Авербух 2004; Лейчик 2022];

- применением методов систематизации термина как языка науки [Даниленко 1977; Felber 1984; Гринев-Гриневиц 2008; Буянова 2012; Лейчик 2022];

- присвоением наименования (термина) понятию на основе когнициии в ее лингвистических проявлениях [Lakoff 1987; Cabré 1999; Temmerman 2000];

- анализом научного дискурса и принципов отбора материала исследования на основе жанровых характеристик (исследовательская статья как ключевой жанр институционального научного дискурса; матрица IMRaD) [Карасик 2000; 2004; Чернявская 2001; 2010; 2016];

- анализом метонимии как системного механизма семантической деривации в научно-техническом дискурсе, включая типовые внутримоделные схемы и их операционализацию в корпусе через коллокационные профили и секционнo-жанровые распределения IMRaD [Minsky 1975; Fillmore 1976; Peirsman, Geeraerts 2006; L'Homme 2020];

- ролью метафоры в формировании новых значений термина [Lakoff, Johnson 1980; Stefanowitsch, Gries 2006; Barsalou 2008; The Cambridge Handbook 2008];

- эпонимическими номинациями специальных понятий [Лейчик 2011; Новинская 2013; Какзанова 2019; Кузнецова, Петрулевич 2021];

- введением цифровых технологий в корпусные исследования языка [Suleimanova, Petrova 2020; Suleimanova 2020; Петрова 2022];

- особенностями терминологической лексикографии [Dubichynskyi, Dąbrowska 2021; Полубиченко, Алхастова 2024; Fuertes-Olivera 2025].

Применялся системный подход к интерпретации языкового материала; учитывались лингвофилософские концепции о взаимокорреляции и неразрывной связи мышления, сознания, языка и речи в процессе познавательной деятельности человека.

Настоящее диссертационное исследование посвящено описанию и объяснению структурных и семантических характеристик терминологии предметной области

«искусственный интеллект» и соответствует паспорту научной специальности 5.9.8. «Теоретическая, прикладная и сравнительно-сопоставительная лингвистика», включая следующие направления исследований: П.2. – «Направления современного языкознания и используемые в них методы описания языков. Терминологический аппарат лингвистики. Лингвистические модели. Метаязык современной лингвистики»; П.4. – «Исследование языка, мышления и познания методами когнитивной лингвистики»; П.12. – «Исследование уровневой структуры языка... Теоретические проблемы морфологии; синтаксиса; грамматических категорий; текста и дискурса»; П.23. – «Корпусная лингвистика»; П.24. – «Лингвистические аспекты искусственного интеллекта»; П.28. – «Лингвистическое терминоведение».

Методика исследования определялась целью исследования и поставленными задачами и включала несколько последовательно реализуемых этапов. На первом этапе был сформирован авторский корпус из 897 рецензируемых англоязычных научных статей и материалов ведущих конференций по искусственному интеллекту (отбор источников осуществлялся с опорой на «Белый список» научных журналов), объемом 13 695 508 токенов, и загружен в корпусный менеджер AntConc (версия 4.3.1). Автоматизированный частотный анализ выделил около 153 000 словоформ; в рабочий перечень включались единицы с нормализованной частотностью ≥ 10 PMW, более редкие отсекались. Далее выполнена ручная фильтрация «шумовых» единиц – нерелевантных для терминографического анализа токенов (математико-технические маркеры и фрагменты кода; оценочная и общеупотребительная лексика, не имеющая отраслевой семантики; общезыковые глаголы; имена собственные без терминологической функции и др.). В результате отбора получены 250 тематически релевантных токенов – кандидатов в термины.

На втором этапе осуществлялся углубленный статистический анализ с использованием показателей Mutual Information (MI) – меры ассоциативной силы слов относительно их независимого распределения, а также Log-Likelihood (LL) – критерия статистической значимости отклонения наблюдаемых совместных частот от модели независимого совместного распределения. Это позволило выделить 2300 релевантных слов и устойчивых словосочетаний (коллокаций). Полученная выборка прошла верификацию по авторитетным глоссариям и международным стандартам ISO/IEC, в результате чего в окончательный перечень вошли 1648 терминологических единиц.

На третьем этапе проводился лингвистический анализ отобранных терминов, включая выделение структурно-морфологических и семантических маркеров, специфичных для научно-технического дискурса ИИ. Дополнительно применялись метод лингвистического моделирования для идентификации ключевых признаков терминов; ономазиологический подход для анализа формально-структурных моделей и оценки продуктивности словообразовательных механизмов (морфологические, синтаксические, морфолого-синтаксические); семасиологический анализ для исследования семантических процессов (метонимизация, метафоризация, эпонимизация, полисемия, синонимия); а также статистический анализ на всех этапах исследования для обеспечения точности, воспроизводимости и семантической достоверности результатов.

Научная новизна данного исследования состоит в том, что впервые на основе специально созданного репрезентативного корпуса научных статей с применением инструментария AntConc был проведен комплексный частотно-коллокационный анализ терминологии предметной области «искусственный интеллект», позволивший автоматизировать ее инвентаризацию и детально описать морфемно-словообразовательную организацию (аффиксальные, сложные, сложносокращенные и композитные модели) с оценкой деривационной продуктивности. На этой основе были системно исследованы семантические механизмы адаптации заимствованных и транстерминологизированных единиц, а также выявлены и типологизированы явления полисемии и синонимии с выделением критериев разграничения значений и управления эквивалентностью. Практическим результатом стала разработка и верификация двух взаимосвязанных классификаций терминов ИИ – деривационной и семантической, которые формируют научно-методическую основу для создания корпусно-ориентированного глоссария и последующей системной работы по унификации и стандартизации отраслевой терминологии.

Положения, выносимые на защиту:

1. Современное терминоведение представляет собой комплексную междисциплинарную область, методологический аппарат которой для исследования динамичных терминосистем, подобно терминологии искусственного интеллекта, интегрирует классические лингвистические подходы с корпусно-вычислительными

методами, социокогнитивной и функционально-коммуникативной парадигмами, а также нормами международной стандартизации (ISO 704, ISO 1087).

2. Терминосистема предметной области «искусственный интеллект», функционирующая в научном дискурсе, соотнесена с понятийной структурой соответствующей области знания и поддается многоуровневому описанию по функционально-дискурсивным, семантическим, формально-структурным, мотивационным и нормативным параметрам; данная классификация обеспечивает операциональную основу для корпусного анализа и терминографического моделирования (на материале английского языка).

3. Стандартизация выполняет роль ключевого механизма концептуализации и гармонизации терминологического аппарата ИИ. Основываясь на методологических стандартах терминологической работы (ISO 704, ISO 1087) и их национальных версиях, такие международные нормы, как ISO/IEC 22989, 23053 и 23894, обеспечивают однозначность, воспроизводимость и междисциплинарную интероперабельность терминов, формируя нормативный каркас для профессиональной коммуникации и регуляторики.

4. Многоэтапная процедура выделения терминологических «кандидатов», включающая автоматизированное извлечение единиц из корпуса, последующую фильтрацию «шумовых» токенов и верификацию терминологической релевантности по конкордансам (KWIC), частотным показателям и статистическим мерам сочетаемости (Mutual Information, Log-Likelihood), формирует воспроизводимую эмпирическую базу для дальнейшего структурно-семантического анализа и терминографического описания терминологии предметной области «искусственный интеллект».

5. Формирование и пополнение терминологии ИИ осуществляется преимущественно через интегрированную модель терминообразования, сочетающую семантические, морфологические, синтаксические и морфолого-синтаксические механизмы. Данная модель отражает преобладание многословных номинаций (A+N, N+N, A+N+N) и высокую продуктивность морфологической деривации, системно кодирующей типы концептов (процесс/результат/инструмент/сущность), что вместе с привлечением общеупотребительной лексики и транстерминологизацией создает

условия для оперативного и нормативно управляемого закрепления новых наименований.

6. Терминосистема ИИ демонстрирует управляемые проявления синонимии и полисемии, которые диагностируются корпусными методами (KWIC, распределительные профили, MI/LL) и регламентируются через стандартизованные дефиниции и установление предпочтительных форм. Методика корпусного исследования, основанная на построении репрезентативного специализированного корпуса и статистической триангуляции (частотность, Mutual Information, Log-Likelihood) в связке с качественным анализом, позволяет объективно идентифицировать и разграничивать ядро, периферию и инновационный слой терминов, а также подтверждать их релевантность и устойчивость.

Структура научно-исследовательской работы. Диссертационное исследование состоит из введения, трех глав, заключения, списка литературы, включающего 198 источников (в том числе 94 наименования на иностранном языке), 3 приложений. Общий объем работы составляет 201 страницу печатного текста.

Во **Введении** представлено обоснование актуальности работы, определяются объект и предмет исследования, обозначаются его цель и задачи, теоретическая и практическая значимость, получают описание методы и эмпирический материал, излагаются основные положения, выносимые на защиту.

В **Главе 1** «Теоретико-методологические основы исследования терминологии предметной области “искусственный интеллект”» формируются концептуально-методологические основания работы; определяется место терминоведения в современной лингвистике, разграничиваются понятия «терминология», «терминосистема» и уточняется рабочая дефиниция термина; систематизируются механизмы терминообразования (морфологические, синтаксические, семантические). Представляется диахронический обзор эволюции терминологии ИИ – от кибернетики до трансформерных архитектур; устанавливаются нормативно-стандартные ориентиры гармонизации (ISO 704/1087, ISO/IEC 22989 и др.; ГОСТ; EU AI Act), задающие метаязык описания и принципы последующего корпусного анализа и лексикографической репрезентации.

В **Главе 2** «Методология проведения корпусного исследования терминологии предметной области “искусственный интеллект”» рассматриваются методологические

основы корпусного анализа терминологии ИИ. Формулируются критерии отбора научных текстов и дается обзор ключевых инструментов платформы AntConc. Описываются принципы применения статистических метрик (Frequency, Mutual Information, Log-Likelihood) для выявления терминологических единиц. Детализируются этапы работы с корпусом – от подготовки и предобработки данных до извлечения и интерпретации терминов, и предлагается методика разработки онлайн-гlossария ИИ на базе открытой платформы WordPress.

В **Главе 3** «Структурно-семантические особенности терминосистемы предметной области “искусственный интеллект”» раскрываются основные параметры структурно-семантического профиля терминологии ИИ. Характеризуются морфологические и морфолого-синтаксические модели образования однословных терминов, такие как аффиксация (*binning* (биннинг), *pretraining* (предобучение)), конверсия (to mark up → *markup* (разметка)); словосложение (*keypoints* (ключевые точки)) и др.; описывается синтаксическая деривация двух- и многокомпонентных именных групп с учетом правоголовности, ограниченной вариативности и критериев устойчивости: A+N (*artificial agent* (искусственный агент)), N+N (*data science* (наука о данных)), V-ed+N (*packed data* (упакованные данные)) и др. Анализируются семантические способы номинации: заимствование (из греч. *autonomy* (автономность)); привлечение общеупотребительной лексики (*prompt* (подсказка) → инструкция/вход для LLM); транстерминологизация (из лингвистики в NLP: *parsing* (синтаксический разбор)), а также механизмы адаптации лексических единиц в терминосистеме ИИ: метонимизация (*DALL-E* → система и изображение/выход), метафоризация (*black box* (черный ящик)), эпонимизация (*Turing test* (тест Тьюринга)), сужение/расширение значения (*agent* (агент)). Особое внимание уделяется выявлению полисемии (например, *model*: 1) математическая модель, 2) алгоритмическая модель, 3) предобученная нейросеть) и синонимии (например, *strong AI* (сильный ИИ) / *full AI* (полный ИИ)), что позволяет проследить многослойность значений и конкуренцию обозначений в терминологии ИИ.

В **Заключении** подводятся итоги исследования; полученные результаты систематизируются и интерпретируются в контексте структурно-семантических характеристик терминосистемы ИИ на материале английского языка; формулируются

теоретические и прикладные выводы и намечаются перспективные направления дальнейшей работы.

Приложения содержат материал иллюстративного характера: QR-код для доступа к онлайн-гlossарию терминов ИИ; QR-код для доступа к перечню лексикографических источников (включая онлайн-платформы и репозитории), а также к списку нормативно-правовых актов и стандартов; QR-код, обеспечивающий доступ к материалам, дополняющим основные положения исследования, включая примеры метонимических переносов; классификацию популяционных алгоритмов по метафорическим источникам и когнитивным эффектам; реестр терминов-эпонимов; таксономию синонимии в терминологии ИИ; описание полисемии ключевых терминов ИИ (семантические ядра, значения, источники и корпусные примеры).

Апробация работы. Основные положения и результаты диссертационного исследования были представлены в виде докладов в рамках следующих научных конференций: Международная научно-практическая конференция «Развитие современной науки и технологий в условиях трансформационных процессов» (18 октября 2024 г., Москва); Международная научная конференция «V Казанский международный лингвистический саммит – 2024» (29-31 октября 2024 г., Казань); Международная научно-практическая конференция «Искусственный интеллект и духовная культура» (30 октября 2024 г., Москва); VI международная конференция «Human – AI Synergy for Effective Language Teaching of Science Communication in a Multidisciplinary, Multicultural Context» (21-23 ноября 2024 г., Долгопрудный); Международная молодежная конференция «В многомерном пространстве современной лингвистики: искусственный интеллект vs традиция» (10 декабря 2024 г., Москва); III Всероссийская конференция «Педагогический дискурс в современной научной парадигме и образовательной практике» (3-5 марта 2025 г., Москва); Научно-практическая конференция «Научный старт – 2025» (с элементами научной школы (3 апреля 2025 г., Москва); Всероссийская конференция «Взаимодействие мыслительных и языковых структур» (18-19 сентября 2025 г., Тамбов).

Основные результаты диссертационного исследования отражены в девяти публикациях общим объемом 5,67 п.л., из них четыре работы (4,15 п.л.) опубликованы в изданиях, включенных в Перечень Высшей аттестационной комиссии Министерства высшего образования и науки Российской Федерации.

ГЛАВА 1. ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ ТЕРМИНОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

1.1. Терминоведение как актуальная область современной лингвистики

Терминоведение – сравнительно молодая научная дисциплина, оформлявшаяся в течение XX века. Само название было предложено и введено в научный обиход В. П. Петушковым (1967) и впервые зафиксировано в печати в тезисах Б. Н. Головина [Лейчик 1969; Петушков 1972]. Классическим основателем направления считается австрийский лингвист О. Вюстер, который в начале 1930-х гг. опубликовал программную работу «Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik» («Международное нормирование речи в технике, в особенности в электротехнике»), положившую начало систематическим исследованиям в этой области [Wüster 1931]. Сформированная им Венская школа закрепила ономазиологический принцип и связку «концепт – термин – определение», настаивая на первичности понятийной модели, согласно которой сначала формулируется содержание понятия, затем подбирается оптимальное обозначение и фиксируется нормативная дефиниция [Wüster 1979; Felber 1984]. В рамках этого подхода первичным считалось содержание понятия, а главной целью – практическая стандартизация и унификация технической лексики для минимизации многозначности [Wüster 1979]. Опираясь на принципы гипотезы, описания и предписания, О. Вюстер задал вектору дисциплины прикладной характер, сфокусированный на нормативном регулировании профессиональной коммуникации [Wüster 1979].

В 1960–1970-е годы представители австро-германской школы терминоведения (К. Галински, Г. Будин, Э. Озер и др.) заложили институциональную и методическую основу дисциплины. По инициативе ЮНЕСКО и Австрийского института стандартизации был учрежден международный центр INFOTERM (1971); впоследствии развернуты сетевые проекты TermNet (регистрация в 1988 г.), кодифицированы полный цикл Terminologearbeit и нормативные рамки (DIN 2330/2331; ISO 704/1087), а также внедрены многоязычные терминологические базы – тем самым терминологическая работа была превращена в воспроизводимый международный стандарт [Felber 1984; DIN 2331:2019-12; DIN 2330:2022-07; ISO 1087:2019; ISO 704:2022].

Параллельно с развитием общей лингвистики, формализовавшей описание именных групп и словообразования [Biber, Gray 2016; Plag 2018], в терминоведении наметился переход от нормативных методик [Sager 1990; Arntz, Picht 1991] к когнитивному и социокогнитивному подходам, связавшим термин с процессами категоризации знаний и фреймовым моделированием [Cabré 1999; Temmerman 2000; A Cognitive Linguistics 2012]. К 1980–1990-м годам дисциплина расширилась за счет сопоставительных исследований [Rondeau 1984; Dubuc 1992], управления эквивалентностью [Handbook of Terminology 1997] и появления корпусно-вычислительного направления *automatic term extraction* (ATE). Использование статистических методов (C-value, LL/MI) и лингвистических шаблонов обеспечило эмпирическую базу для автоматизированного извлечения и верификации терминов на больших массивах данных [Kageura 2002; Bowker, Pearson 2002; Evert 2009].

Отечественная традиция терминоведения, восходящая к трудам М. В. Ломоносова и отраслевым словарям XIX века, оформилась в научную школу благодаря работам Д. С. Лотте, С. А. Чаплыгина, А. Ф. Лесохина, А. А. Реформатского и др. Основоположники обосновали требования к термину – однозначность, краткость, отсутствие экспрессии и системность, противопоставив его общелитературному слову по критериям дефинируемости и нормативной закреплённости [Реформатский 1967; Лотте 1961; 1982]. Дискуссии 1950-х годов позволили разграничить статус термина и обиходной лексики, внедрив в практику ономаσιологический и семасиологический методы описания, которые легли в основу советской школы стандартизации и терминографии [Лотте 1961].

Во второй половине XX века в России сформировались три основные школы терминоведения: Московская (системный и нормативно-терминографический подход) [Лотте 1961; Гринев-Гриневиц 2008; Лейчик 2022], Санкт-Петербургская (историко-лингвистическое измерение и динамика терминов) [Кутина 1970; Герд 1981] и Нижегородская (функционально-коммуникативная перспектива) [Немченко 1985; Головин, Кобрин 1987]. К началу XXI века терминоведение утвердилось как комплексная дисциплина, но его место в структуре наук остается предметом дискуссий. Ученые определяют его либо как отрасль языкознания [Татаринов 2006], либо как комплексную или самостоятельную область знаний на пересечении

лингвистики, логики и предметных наук [Гринев-Гриневиц 2008; Буянова, Новоселецкая 2021; Лейчик 2022].

Кроме собственно терминов терминоведение занимается исследованием специальных лексических единиц, таких как *номены* (наименование единичного понятия), *профессионализмы* (слово, ограниченное какой-либо профессией), *профессиональные арготизмы* (экспрессивные слова, обслуживающие интересы узких социальных групп), *предтермины* (специальные лексемы, используемые в качестве терминов для называния новых сформировавшихся понятий, но не отвечающие основным требованиям, предъявляемым к термину), *квазитормины* (используются для означивания несформировавшихся понятий, с размытыми категориальными признаками, не отвечают основным требованиям, предъявляемым к терминам, не имеют своего места в терминосистеме и не зафиксированы в специальных словарях), *терминоиды* (наименования формирующихся понятий), *прототермины* (наименования специальных представлений) [Berger 2021]. Рассмотрению данных лексических единиц посвящены многочисленные труды как отечественных, так и зарубежных исследователей [Даниленко 1977; Handbook of Terminology 1997; Cabré 1999; Temmerman 2000; Суперанская 2012; Berger 2021]. Однако в данной работе мы разделяем позицию В. П. Даниленко о приоритете изучения терминов и номенов как фундаментальных элементов терминологической системы [Даниленко 1977] и потому ограничиваем анализ только этими двумя категориями.

Современный этап (2010–2020-е гг.) отмечен глубокой интеграцией корпусно-вычислительных методов (АТЕ) [Бручес, Батура 2021; Большакова, Семак 2025] и опорой на крупные текстовые массивы, такие как НКРЯ [Савчук 2024]. Важнейшими трендами остаются гармонизация отечественных стандартов с международными нормами ISO 1087/704 [Шелов, Четверикова 2020], совершенствование отраслевых терминобаз [Максимова, Шпальченко 2023] и создание интеллектуальной цифровой инфраструктуры, примером которой служит электронный словарь «Искусственный интеллект» [Искусственный интеллект 2023].

Указанные тенденции в сфере терминографической и нормативной практики переводят работу с терминологией в режим воспроизводимых процедур; это позволяет выделить восемь функций терминоведения: описательную (фиксация терминов), систематизирующую (упорядочивание), нормативную (стандартизация),

теоретическую (изучение природы знака), прикладную (перевод и базы данных), когнитивную (формирование знаний), коммуникативную (точность передачи информации) и прогностическую (анализ векторов развития, актуальный для сферы ИИ). Эволюция дисциплины от локальных практик к самостоятельной области знаний обеспечила интеграцию историко-описательных подходов с корпусной аналитикой, что создает методологическую базу для верифицируемого описания структурно-семантических характеристик терминологии ИИ и правил ее нормативного употребления в профессиональном дискурсе.

Таким образом, современное терминоведение интегрирует историко-описательные и сопоставительные подходы с корпусной и вычислительной аналитикой на основе стандартизированных терминографических процедур, что обеспечивает верифицируемость, сопоставимость и обобщаемость результатов. В рамках данной диссертации указанная методологическая рамка используется для строго проверяемого описания способов терминообразования и структурно-семантических характеристик терминологии ИИ, а также для выработки правил ее согласованного и нормативно управляемого употребления в профессиональной коммуникации.

1.2. Терминология как объект лингвистического анализа

Терминология трактуется как системно организованный массив наименований, соотнесенный с понятийной структурой области и закреплённый в нормативных стандартах [Wüster 1979; ISO 1087:2019; ISO 704:2022], при этом современные подходы смещают акцент с нормативного идеала [Sager 1990; Arntz, Picht 1991] на динамику функционирования терминов в дискурсе и когнитивные механизмы категоризации [Cabré 1999; Temmerman 2022]. Онтологическая и культурно-ориентированная перспективы связывают терминологический массив с моделированием понятийных связей и задачами межъязыковой локализации [Deresker 2002; Diki-Kidiri 2022], что превращает терминологию (включая сферу ИИ) в сеть устойчивых иерархических наименований. В отечественной традиции системность терминологии рассматривается как отражение исторически развивающейся системы понятий [Лотте 1961; Реформатский 1967; Ахманова 2004], где, по мнению В. М. Лейчика, стабилизация массива происходит через процессы селекции номинаций

и междисциплинарной интеграции, подготавливая базу для последующей стандартизации [Лейчик 2022].

Представления о терминологии как совокупности специальных единиц расходятся. Часть исследователей подчеркивает ее особую природу и видит в ней искусственно формируемую подсистему со своими внутренними законами, отличными от общелитературных норм [Капанадзе 1965; Реформатский 1967; Лотте 1982; Лейчик 2022]. Так, А. А. Реформатский метко называл терминологию «службой двух хозяев: системы лексики и системы научных понятий», фиксируя двойственность ее статуса – между языком и знанием [Реформатский 1967: 122]. Другая линия исходит из того, что терминология не изолирована от общих законов языка и естественно входит в его лексическую систему [Головин, Кобрин 1987; Авербух 2004]. Наконец, в рамках субстанциональной трактовки терминология мыслится как система специальных знаков с требованиями однозначности и жесткой формы [Звегинцев 1962; Толикина 1970]. Однако данная позиция неоднократно подвергалась критике за свой нормативно-идеальный характер, поскольку она игнорирует реальные процессы вариативности терминов, динамику их развития и проницаемость границ между специальной и общеупотребительной лексикой [Козловская 2023].

В своем исследовании мы исходим из структурно-системной (второй) позиции и рассматриваем терминологию как автономную, но «проницаемую» подсистему лексики со своими правилами организации и нормирования, взаимодействующую с общим языком. Такой выбор обеспечивает структурный «уклон» анализа и дает практические критерии для работы с вариантностью, эквивалентностью и стандартизацией, избегая при этом чрезмерной жесткости «знаковой» концепции и редуцирования терминологии к обычной лексике.

Современные исследования подтверждают, что системность терминологии является прямым отражением упорядоченности научных понятий предметной области [Головин, Кобрин 1987]. При этом динамичность знания препятствует завершенности системы, выдвигая на первый план деятельностный подход, где термины репрезентируют результаты человеческой практики. В этой связи принципиальной представляется методологическая установка Г. П. Щедровицкого, который призывал рассматривать языковой знак «не в отнесении к объектам, а в отношении к деятельности, элементом которой он является и благодаря которой он получает смысл

и значение» [Щедровицкий 1995: 540–544]. Таким образом, системная связность терминологии является прямым следствием и отражением системной организации порождающей ее познавательной и практической деятельности.

Практическая реализация терминологической работы сегодня опирается на принципы междисциплинарности, интегративности, когнитивной обусловленности и антропоцентризма [Sager 1990; Лейчик 2022]. Это особенно проявляется в сфере ИИ, где базовые термины, такие как *attention mechanism* (механизм внимания) или *reinforcement learning* (обучение с подкреплением), требуют синтеза понятийных рамок математики, информатики и лингвистики для их точной концептуализации.

Интеграция в философском дискурсе трактуется в качестве процесса или действия, направленного на формирование целостности, объединения разнородных элементов в единое целое. Согласно «Логическому словарю-справочнику» Н. И. Кондакова, под интеграцией (от лат. ‘integer’ – целый) понимается «объединение в целое, в единство каких-либо элементов... в теории систем – состояние взаимосвязи отдельных компонентов системы и процесс, обуславливающий такое состояние» [Кондаков 1975: 203].

Применительно к терминологии интегративность может быть определена как системное согласование разнородных терминологических массивов в рамках единой понятийной структуры. Ее основные проявления включают нормирование и гармонизацию терминов (ISO 704/1087); обеспечение многоязычной эквивалентности; применение онтологий и графов знаний, позволяющих формализовать связи типа «род–вид–часть–функция», что имеет ключевое значение для развития терминографических баз данных и иных ИТ-продуктов [Wüster 1979; Handbook of Terminology 1997; Лейчик 2022]. Реализация интегративного подхода способствует минимизации терминологической вариативности, обусловленной межотраслевыми традициями именования, и обеспечивает воспроизводимость результатов терминологической работы.

Когнитивная обусловленность употребления термина в рамках домена означает, что функционирование термина зависит от способов концептуализации предметной области – от того, как структурируется знание (концепты, категории, фреймовые представления) и какие интерпретационные схемы задают профессиональную коммуникацию. В этом смысле термин соотносится не только с формой и значением,

но и с моделью знания, принятой в конкретном домене. Данный подход, развиваемый в когнитивном терминоведении, трактует термин не как статичную «метку», а как инструмент категоризации и структурирования знания, вербально кодирующий концептуальные связи и поддерживающий формирование новых понятий [Гринева-Гриневич 2008; Манерко 2024]. Для полноценного понимания термина необходимо знание не только его значения, но и всей предметной области – его концептуального «ареала».

Взаимодополняющие семасиологический и ономазиологический подходы в терминоведении опираются на когнитивную обусловленность, подчеркивающую мотивированность термина через прототипное и фреймовое моделирование [Fillmore 1976; Rosch 1978; Cabré 1999; Temmerman 2000]. Это обеспечивает выбор прозрачных форм и дефиниций, фиксирующих место понятия в системе [Суперанская 2012; Лейчик 2022], в то время как принцип *non multa, sed multum* (максимум содержательности при минимуме средств) способствует созданию максимально информативных номинаций.

Антропоцентризм в философии понимается как мироотношение, при котором человек выступает центром мироздания и развития [Философский энциклопедический словарь 2006]. В языкознании эта парадигма сохраняется: язык рассматривается как феномен, неразрывно связанный с его носителем. В. фон Гумбольдт одним из первых обосновал антропоцентричность языка, связав ее с направленностью высказывания на адресата [Гумбольдт 2000], а Э. Бенвенист показал, что человек конституируется как субъект «в языке и благодаря языку» [Бенвенист 1974: 293]. Человек при этом выступает не пассивным референтом, а активным субъектом, продуцирующим языковой материал на основе освоения концептов, идей и явлений.

Антропоцентрический подход устанавливает коммуникативную соразмерность термина, требуя, чтобы наименование было точным, стилистически нейтральным, когнитивно доступным для адресата и функционально пригодным в профессиональном дискурсе, включая сферу перевода и локализации. В русле коммуникативно-когнитивных исследований подчеркивается ориентация термина на практику пользователя и культурный контекст – от эргономики языковой формы до учета локальных традиций именования [Бенвенист 1974; Cabré 1999; Гумбольдт 2000; Diki-Kidiri 2022].

Наиболее наглядно антропоцентризм терминологии проявляется в использовании образных и ориентационных моделей номинации, опирающихся на базовый телесный и пространственно-временной опыт человека [Lakoff, Johnson 1980]. В терминологии ИИ широко представлены ориентационные схемы «верх–низ» и «увеличение–уменьшение», ср. *top-down parsing* (верхнеуровневый синтаксический разбор) vs *bottom-up parsing* (нижнеуровневый синтаксический разбор), *upsampling* (повышение дискретизации) vs *downsampling* (понижение дискретизации), а также динамические и пространственные модели, ср. *gradient descent* (градиентный спуск) vs *gradient ascent* (градиентный подъем), *feature space* (пространство признаков), *latent space* (скрытое пространство). Наряду с этим существенную роль играют темпоральные и режимные характеристики, формирующие терминологическую оппозицию способов обработки и обучения, ср. *online learning* (онлайн-обучение), *real-time processing* (обработка в реальном времени).

В прикладном плане это реализуется через проектирование наименований «сверху вниз» от междисциплинарных моделей (интеграция), выбор мотивированных единиц с фреймовой спецификацией (когнитивная обусловленность) и управление вариативностью с учетом адресата (антропоцентризм). Закрепление данных результатов в стандартах обеспечивает воспроизводимость процедур и точность терминологий в быстро меняющихся областях – прежде всего в ИИ, – где термин функционирует не только как имя, но и как инструмент структурирования знания и профессиональной коммуникации [Wüster 1979; Лейчик 2022; ISO 704:2022].

Методология исследования терминологии объединяет лексикологический, семантический и когнитивный анализ с методами корпусной лингвистики и концептуального моделирования, реализуя синтез ономаσιологического и семасиологического подходов. При этом ключевым остается разграничение «терминологии» как стихийно сложившейся совокупности лексем [Berger 2021] и «терминосистемы» как результата сознательной кодификации и стандартизации [Головин, Кобрин 1987; Лейчик 2022]. Несмотря на существование позиции об их фактическом отождествлении в силу системности самого научного знания [Головин, Кобрин 1987; Татаринов 2006], К. Я. Авербух предлагает рассматривать терминосистему как терминологию с эксплицитно выраженными системными свойствами [Авербух 2004]. Таким образом, комплексный инструментарий позволяет описать термин

одновременно как языковой знак и как упорядоченный элемент понятийной структуры предметной области.

Современные исследования подтверждают пересечение терминологии и терминосистемы, где первая выступает как номинативный фонд, а вторая – как динамичная совокупность с коммуникативной функцией [Шарафутдинова 2016]. В данной работе под терминологией понимается широкий массив специализированных наименований (включая номены и периферийные единицы), а под терминосистемой – его структурная проекция в плоскости понятийных связей и иерархий. Опираясь на структурный подход (А. А. Реформатский и др.), мы рассматриваем эти категории как две аналитические проекции одного множества, что позволяет описывать вариативность (синонимию и полисемию) не как дефект, а как системный фактор в условиях естественной динамики таких областей, как искусственный интеллект.

1.3. Определение понятия «термин»

Определение понятия «термин» (от лат. *terminus* – предел, граница) остается дискуссионным из-за его междисциплинарной многозначности. В то время как логики и философы акцентируют внимание на его дефинитивности и понятийности, лингвисты разграничивают «имя понятия» и его толкование [Лейчик 2022]. В философской традиции термин рассматривается как субъект или предикат суждения, самоограничивающий мысль в рамках акта высказывания [Флоренский 1990], и выступает «мостом между субъектом и объектом» [Лосев 2009]. При этом в обыденной и профессиональной речи номинация часто задается индивидуальной перцепцией и закрепляется в узусе при условии ее функциональной пригодности [Зиньковская, Сахно 2021]. Таким образом, трактовка термина эволюционировала от этимологической «предельности» до сложной когнитивной единицы, чей объем содержания варьируется в зависимости от контекста исследования.

В западной традиции базовые положения терминоведения, сформулированные О. Вюстером, основываются на жесткой корреляции «один термин – один концепт» и первичности системного упорядочения понятий [Wüster 1931; 1979]. Х. К. Сагер дополняет эту прагматическую линию, утверждая, что лексикон языка для специальных целей (LSP) стремится к полному соответствию конвенционально выделенным концептам при строгой референции [Sager 1990]. Современные теоретики, такие как

Л. Депенкер, расширяют это понимание, трактуя «специализацию» как приобретение особого смысла по мере возрастания техничности домена [Derecker 2002]. К. Кагеура идет дальше, включая в объем термина любые знаковые формы (формулы, символы, номенклатуру), репрезентирующие специальные понятия [Kageura 2002]. Таким образом, современное западное терминоведение эволюционирует от нормативного идеала однозначности к более гибкому пониманию термина как специализированной знаковой единицы.

В когнитивном терминоведении термин понимается как информационная структура, аккумулирующая профессиональные знания и отражающая концептуальную картину мира конкретной области [Thoiron, Béjoint 2010; A Cognitive Linguistics 2012; L'Homme 2020]. Несмотря на нормативные требования к однозначности и прикладной связи с объектом [Thoiron, Béjoint 2010], представители социокогнитивного и коммуникативного направлений указывают на естественную вариативность, полисемию и размытость границ понятий в реальных терминосистемах [Cabré 1999; Temmerman 2000]. В отечественной традиции также подчеркивается специализированный статус термина как слова, стремящегося к точному выражению системных понятий и требующего научной дефиниции [Реформатский 1967; Суперанская 2012]. Таким образом, современные подходы ставят под вопрос жесткое противопоставление «слова» и «термина», предлагая описывать единицы специального знания через их дискурсивные функции при сохранении их связи с профессиональной коммуникацией [Temmerman 2000; L'Homme 2020].

В. М. Лейчик, опираясь на когнитивный подход, обосновывает статус термина как единицы языка для специальных целей, которая одновременно фиксирует понятие и выступает инструментом познания наравне с приборами и схемами [Лейчик 2022]. В лингвистике доминирует понимание термина как слова или устойчивого словосочетания, соотнесенного с профессиональным понятием и служащего для концентрации и передачи информации [Головин, Кобрин 1987; Лемов 2000]. Несмотря на дискуссию о том, является ли термин особым словом или лишь функциональной ролью лексемы [Винокур 1939], большинство исследователей придерживается субстанциональной трактовки его природы. Релевантной для анализа остается корреляция «термин – терминология – язык науки», предложенная Л. Ю. Буяновой, где терминология рассматривается как системная совокупность единиц, связанных на

понятийном и лексико-грамматическом уровнях [Буянова 2012]. Это позволяет объединить лингвистическую и институционально-дискурсивную перспективы исследования.

Термин совмещает в себе лингвистическое и понятийное начала, обеспечивая материальную фиксацию объекта в виде лексической единицы [Морозова 2004], при этом связь «имя – понятие» всегда детерминирована конкретной сферой профессиональной деятельности [Суперанская 2012]. Р. Г. Пиотровский, рассматривая проблему семантической очерченности, указывает, что требование однозначности не исключает экспрессивного потенциала и внутренней формы термина, которые повышают его функциональную адекватность в дискурсе [Пиотровский 1952]. Данная позиция выступает связующим звеном между дефинитивным подходом Г. О. Винокура и А. А. Реформатского [Винокур 1939; Реформатский 1967] и линией Б. Н. Головина, В. П. Даниленко и А. С. Герда, рассматривающих термин как полноправную лексическую единицу в составе общего языка [Даниленко 1977; Герд 1981; Головин, Кобрин 1987]. В итоге термин предстает как динамичная единица, которая фиксирует строгое научное понятие, но сохраняет лексико-дискурсивные свойства в живой профессиональной коммуникации.

Разнообразие трактовок понятия «термин» обусловлено спецификой научных школ и диапазоном его функций – от логико-понятийной до когнитивно-прагматической [Суперанская 2012]. В настоящем исследовании под термином понимается единица языка для специальных целей, соотнесенная с конкретным концептом и обладающая дефинированным содержанием, точностью референции, стилистической нейтральностью и нормативной кодифицируемостью. Термин выступает элементом онтологии предметной области, значение которого раскрывается через позицию в иерархической сети понятий (родо-видовых, меронимических и ассоциативных связей) [Wüster 1979; Головин, Кобрин 1987; Суперанская 2012; Лейчик 2022; ISO 704:2022]. В качестве иллюстрации системности в работе представлен фреймово-ориентированный профиль термина *retrieval-augmented generation (RAG)* (дополненная поиском генерация), фиксирующий его связи в архитектуре ИИ (см. Таблицу 1).

Таблица 1 – Реляционный профиль термина *retrieval-augmented generation, RAG*
(дополненная поиском генерация)

Тип связи	Узел / роль	Примеры терминов ИИ
Род → Вид	Иерархия класса и подтипов	<i>knowledge-grounded generation</i> (генерация с опорой на знания) → <i>RAG</i> (дополненная поиском генерация) → <i>hybrid RAG</i> (гибридный RAG), <i>query-expansion RAG</i> (RAG с расширением запроса)
Целое → Часть (меронимия)	Подсистемы/компоненты RAG-пайплайна (что входит в состав системы)	<i>document store / knowledge base</i> (хранилище документов/знаний); <i>chunker / splitter</i> (разделитель); <i>embedding model</i> (модель встраивания) и др.
Ассоциативные связи	Связанные процедуры/цели	<i>top-k retrieval</i> (извлечение top-k), <i>grounding</i> (привязка к источникам), <i>citation / source attribution</i> (цитирование/указание источников)
Вариантность / синонимия	Близкие обозначения	<i>knowledge grounding</i> (привязка к знаниям), <i>open-book QA</i> (вопрос–ответ «с открытой книгой»), <i>augmented generation</i> (расширенная генерация)
Противопоставления	Альтернативные режимы/подходы	<i>closed-book generation</i> (генерация «с закрытой книгой», без внешнего извлечения), <i>parametric knowledge vs non-parametric knowledge</i> (параметрические vs непараметрические знания)

Системность термина обеспечивает его интерпретируемость через типы связей ISO 704 (род–вид, часть–целое, ассоциации), что позволяет настраивать корпусные правила извлечения (NP) и ранжировать кандидатов по показателям *termhood* и *unithood* («терминность» и «устойчивость единицы»).

Классическая традиция рассматривает однозначность как нормативный идеал «один концепт – одно обозначение» для исключения синонимии и омонимии [Wüster 1979; Лотте 1961; Реформатский 1967; ISO 704:2022], однако современные социокогнитивные подходы признают неизбежность «контролируемой вариативности» и междисциплинарной полисемии (например, *kernel* (ядро) в операционных системах и *kernel* (ядро) в статистике и машинном обучении) [Cabré 1999; Temmerman 2000; A Cognitive Linguistics 2012]. В настоящем исследовании, основанном на структурном описании, мы стремимся к нормативной однозначности в домене ИИ, управляя возникающей многозначностью через дефиниции, доменные метки и выбор предпочтительных форм на основе корпусной проверки.

Точность термина обеспечивается классической дефиницией через ближайший род и видовое отличие (*genus + differentia*), что минимизирует контекстную

вариативность и риск рассогласования понятий [Wüster 1979; ISO 704:2022]. Однако в новых междисциплинарных сферах, таких как ИИ, часто встречается понятийная размытость, порождающая «терминоиды» и «квазитермины» [Даниленко 1977; Гринев-Гриневиц 2008; Суперанская 2012], семантическая неопределенность которых в данном исследовании устраняется с помощью статистических метрик корпусного анализа. Граница между специальной и общеупотребительной лексикой остается проницаемой из-за наличия «консубстанциональных терминов», возникающих в результате заимствования или детерминологизации [Шелов 2014]. Для корректной интерпретации таких единиц необходим диахронный мониторинг, уточнение онтологических связей и проверка их регистрового распределения в профессиональном дискурсе.

Существенными признаками термина являются его принадлежность к специальной сфере и стилистическая нейтральность, что отличает его от эмоционально окрашенных жаргонизмов и профессионализмов [Даниленко 1977; Суперанская 2012]. Термин должен обладать воспроизводимостью и устойчивостью формы, сохраняя стабильный графический и морфосинтаксический шаблон для обеспечения точности коммуникации [Handbook of Terminology 1997]. В дисциплинарном употреблении выделяют общепринятые стандартизированные единицы (*supervised learning* (обучение с учителем)) и авторские номинации (*StyleGAN*), которые могут со временем пройти кодификацию и войти в основной терминофонд [Temmerman 2000; Лейчик 2022]. Таким образом, закреплённость термина в словарях и учебниках подтверждает его статус инвариантного наименования, способного эволюционировать от локальной авторской идеи до универсального стандарта отрасли.

Принцип языковой экономии побуждает термин передавать максимум содержания минимальными средствами, что ведет к широкому использованию аббревиатур и сокращению описательных конструкций [Лотте 1961; Даниленко 1977; Sager 1990; Суперанская 2012]. Оптимальная длина многокомпонентного термина определяется однозначностью обозначения субпонятий, при этом эмпирически зафиксированная протяженность профессиональных номинаций может варьироваться от 2 до 17 элементов [Головин, Кобрин 1987]. Совокупность признаков – системность, дефинированность, однозначность, нейтральность и устойчивость формы – образует критериальный порог терминологичности, необходимый для отграничения

специальной лексики от общеупотребительной. В рамках данного исследования эта интегральная модель служит базой для анализа эмпирического материала в области искусственного интеллекта, где отсутствие корреляции со строгой системой понятий исключает статус термина для лексической единицы.

1.4. Способы терминообразования

Терминообразование представляет собой двухэтапный процесс: первичное присвоение названия новому понятию на основе его функций и последующая трансформация термина, отражающая эволюцию знания [Татаринов 1996]. Согласно Л. А. Капанадзе, этот процесс завершается наделением слова строгой дефиницией, что превращает его в инструмент фиксации безграничного человеческого познания [Капанадзе 1965]. Интенсивность пополнения терминофонда напрямую зависит от исторических научных прорывов – от открытий И. Ньютона и Р. Бойля в Новое время до идентификации электрона Дж. Дж. Томсоном и создания теории относительности А. Эйнштейном на рубеже XIX–XX веков. Современный этап, начавшийся с научно-технической революции 1950-х годов, обеспечил взрывной рост терминологии в таких стратегических областях, как информатика, нанотехнологии и биотехнологии.

И. Пригожин (1917–2003) предложил постнеклассическую картину мира, отличающуюся особенной неопределенностью, вариативностью, нелинейностью и синергией. Знание, по его мнению, более не существует обособленно, но находится в постоянном взаимном контакте [Пригожин, Стенгерс 1986]. Следовательно, сегодня более продуктивным представляется рассматривать не отдельно взятую дисциплину (ее терминологию, терминологический корпус, терминосистему и т. д.), а «пучки» взаимопроникающих направлений. В технологическом кластере уместно говорить об искусственном интеллекте как наддисциплинарном домене, включающем машинное обучение (а внутри него – глубокое обучение) наряду со смежными областями (например, робототехника, биоинформатика). На уровне профессиональной коммуникации эти подсферы образуют частично автономные терминологические корпуса, однако их лексика и модели терминообразования существенно пересекаются. Специалисты конкретных направлений выступают как создатели и нормировщики терминов, когда того требует упрощение коммуникации, а триггером появления и

закрепления новых наименований служат новые проблемы и задачи, активирующие исследовательскую деятельность.

Терминообразование всегда детерминировано и нацелено на производство моносемичного названия, сжато и емко отражающего признаки именуемого явления. Наименование (номинация) всецело зависит от восприятия именуемого явления человеком. Если восприятие ложно, неточно, ошибочно, то и выбор термина не может быть релевантным. Кроме того, понятия со временем видоизменяются, эволюционируют, и термин перестает отражать основные функции, признаки, утрачивая значимость. Д. С. Лотте, например, говорит о правильно ориентирующих, нейтральных и ложно ориентирующих терминах, хотя даже последние не стоит отвергать, поскольку они несут дефиницию (хотя и ложную) и вплетены (пусть и временно) в одну из терминосистем [Лотте 1961].

В. Г. Гак и Л. Гильбер упоминают три витка жизни термина: рождение, зрелость и умирание [Guilbert 1975; Гак 1998]. На этапе рождения термин существует на ментальном уровне в сознании создателя продукта, требующего номинации, или появляется в результате обсуждения продукта. Термин может зарождаться как в диалогическом взаимодействии, так и при монологическом изложении концепции. В трактовке Б. Н. Головина номинации на данной стадии относятся к сфере речи, а не к системе языка [Головин, Кобрин 1987]. Следующий период – зрелость – связан с терминологическим гнездом, понимаемым как «группа терминов, объединенная на основании общности опорного терминологического элемента, который входит в качестве основной терминологической единицы в каждый термин гнезда и функционирует или как самостоятельный термин (терминологические гнезда с вершиной), или не функционирует как самостоятельный термин (терминологические гнезда без вершины)» [цит. по: Бессонова 1988: 102]. Терминология этого периода зачастую переходит в терминосистему, приобретая свойства устойчивости и целостности. Далее, с течением времени, изменением концептов, возникновением новых теорий, термин умирает, то есть вытесняется более современным или исчезает вовсе вместе с отрицанием, отрицанием понятия. Это происходит в случае, если та или иная теория становится неприменимой и неактуальной [Лейчик 2010: 351–358].

Традиционная теория терминологии ставила в основу принцип системности и однозначности термина, предполагая планомерное формирование терминологии за

счет внутренних ресурсов языка, и прежде всего неоклассических формантов, обеспечивающих родо-видовые связи и «прозрачность» формы [Wüster 1979]. В русле этой парадигмы В. М. Лейчик отмечает, что подавляющее большинство специальных единиц действительно образуется через словообразовательную деривацию – преимущественно суффиксацию и префиксацию. Морфологические способы при этом обеспечивают «прозрачность» внутренней формы термина и системную связь с однокоренными единицами («процесс ↔ результат ↔ носитель/инструмент») [Лейчик 2022]. Показательный пример из терминологии ИИ – *tokenize* → *tokenization* → *tokenizer*, где суффикс *-ation* оформляет название процесса (процедура сегментации текста на токены), *-er* называет средство/агент (модуль/алгоритм, выполняющий сегментацию), а префиксы *de-*, *re-*, *pre-* системно задают значения обратного, повторного и предшествующего действия (*de-tokenization*, *re-tokenization*, *pre-tokenization*).

Общепринятые классификации выделяют следующие основные формально-лингвистические механизмы терминообразования: словообразовательная деривация (аффиксация, словосложение, конверсия), синтаксическое образование терминологических сочетаний (сочетания слов), семантический перенос значений, а также аббревиация и заимствование [Суперанская 2012; Лейчик 2022]. Отдельно отметим эпонимизацию – ономастический механизм номинации по имени ученого/школы, топониму или бренду (антропо-, топо-, брендоэпонимы): *Turing test* (тест Тьюринга), *Bellman equation* (уравнение Беллмана), *Markov decision process* (марковский процесс принятия решений). Функционально эпонимы реализуют метонимический перенос и требуют учета капитализации и деривационных вариантов (*Bayesian*, *Hebbian*, *Hessian*).

Несмотря на функциональную эквивалентность терминов-слов и полилексемных единиц [Циткина 1988], корпусные данные подтверждают доминирование многословных номинаций, составляющих до 50–80% терминофонда [Gries 2024]. Согласно классификации Р. Ф. Прониной, такие сочетания (шаблоны A+N, N+N, NofN) варьируются от чисто терминологических (*logistic regression* (логистическая регрессия)) до образованных из общеупотребительных слов (*bag of words* (мешок слов)), получающих специальный статус только в единстве [Пронина 1986]. Таким образом, морфосинтаксическая структура термина напрямую отражает

его позицию в понятийной сети, обеспечивая точность и интерпретируемость профессиональной коммуникации.

Типология В. П. Даниленко классифицирует специальные единицы по форме (термины-слова, термины-словосочетания, символослова) и частеречной принадлежности, охватывая как непроемные корни (*mesh* (сетка)), так и буквенно-цифровые обозначения (*GPT-4*) [Даниленко 1977]. В зарубежной традиции Х. К. Сагер дополняет этот анализ разграничением первичного терминопобразования, неразрывно связанного с концептуализацией нового знания, и вторичного, которое представляет собой «сконструированный» перенос уже известных понятий в иные языки или домены через заимствование и калькирование [Sager 1990]. Данные принципы закреплены в стандарте ISO 704, который увязывает механизмы номинации (аффиксацию, конверсию, аббревиацию) с онтологической структурой предметной области и требованиями к дефинициям [ISO 704:2022] и позволяет описывать термины ИИ и как статические языковые знаки, и как динамические продукты межъязыковой и междисциплинарной гармонизации.

Таким образом, способы терминопобразования следует рассматривать не только в формально-структурном, но и в когнитивно-прагматическом ключе: с учетом фреймовых знаний о типичных ситуациях, ассоциативных и метафорических моделей, а также жанровых и аудиторных установок. В фреймовой терминологии форма термина коррелирует с его функциональной ролью в соответствующем фрейме; морфологические маркеры системно кодируют тип концепта (процесс, результат, агент/инструмент, свойство), облегчая его интеграцию в онтологию домена [A Cognitive Linguistics 2012; Faber 2022]. Так, продуктивные английские суффиксы и композиционные схемы распределяются по ролям: *X-tion/-ing* – процесс (*recognition* (распознавание), *training* (обучение)), *X-er/-or* – агент/инструмент (*parser* (парсер), *controller* (управляющий модуль)), *X-ment/-ness/-ity* – состояние/свойство (*alignment* (согласование модели), *randomness* (случайность), *adaptability* (адаптивность)), а модели A+N (*deep learning* (глубокое обучение)), N+N (*expert system* (экспертная система)) эксплицируют атрибутивные и меронимические связи внутри фрейма. Такая «формально-семантическая маркировка» повышает прозрачность, облегчает дефинирование и терминографическую интеграцию.

Различие между первичным и вторичным терминообразованием в ИИ проявляется в создании уникальных номинаций для новых феноменов (*in-context learning* (обучение в контексте), *chain-of-thought prompting* (подсказки «цепочки рассуждений»)) или семантической специализации заимствованных лексем (*attention* (внимание)), которые меняют широкое значение на узкий технический слот. С. В. Гринев-Гриневиц синтезировал эти подходы в унифицированную классификационную модель, объединившую достижения Венской школы, функционально-коммуникативной традиции и социокогнитивного направления [Wüster 1979; Sager 1990; Cabré 1999; Temmerman 2000]. Данная модель, опирающаяся на отечественную лексикологическую школу, успешно апробирована в различных междисциплинарных доменах – от медицины до ИКТ [Калиновская 2017; Бурдина 2023]. Таким образом, системная классификация механизмов номинации позволяет эффективно описывать как возникновение принципиально новых терминов ИИ, так и процессы адаптации общенаучного фонда.

Предлагаемая классификация интегрирует в единую систему четыре взаимодополняющие группы способов терминообразования: семантические – терминологизация общеязыковых единиц, специализация/расширение значения, метафора/метонимия, межсистемное заимствование (сюда же отнесены иноязычные заимствования как сопровождаемые смысловыми сдвигами); морфологические – суффиксация, префиксация, префиксально-суффиксальные модели, конверсия, усечение, фонетико-морфологическая адаптация; синтаксические – формирование NP-конструкций (A+N, N+N, V-ed+N, NofN); морфолого-синтаксические – словосложение, эллипсис, аббревиация. Заимствования рассматриваются одновременно как внешний источник формы и как механизм семантического переосмысления [Гринев-Гриневиц 2008]. Классификационная схема с примерами терминов ИИ приведена в Таблице 2.

Таблица 2 – Интегрированная классификация способов терминообразования
(по С. В. Гриневу-Гриневичу)

Группа	Механизм	Примеры терминов ИИ
Семантические	Терминологизация общеупотребительного слова	<i>prompt</i> (подсказка) → <i>prompt</i> (инструкция для вызова отклика LLM); <i>token</i> (жетон) → <i>token</i> (минимальная единица текста)
	Расширение значения	<i>model</i> (модель) макет/образец → алгоритмическая модель
	Специализация значения	<i>attention</i> (механизм внимания) внимание → механизм внимания
	Метафоризация	<i>pruning</i> (отсечение): обрезка (садоводство) → модель разреженности
	Метонимический перенос	<i>BLEU</i> (алгоритм BLEU): метрика → численное значение (<i>optimize BLEU</i>)
	Межсистемное заимствование лексем	<i>entropy</i> (энтропия), <i>perplexity</i> (недоумение) теория информации/статистика → ИИ
	Заимствование иноязычных лексем/элементов	<i>ensemble</i> (ансамбль) из франц., <i>stochastic</i> (стохастический) из древнегреч.; морфемы неоклассического происхождения (<i>auto-</i> , <i>hyper-</i> , <i>meta-</i> , <i>pseudo-</i> , <i>multi-</i> , <i>semi-</i>)
Морфологические	Суффиксация	<i>regulariz-ation</i> (регуляризация); <i>random-ness</i> (случайность)
	Префиксация	<i>pre-training</i> (предобучение), <i>re-ranking</i> (повторное ранжирование)
	Префиксально-суффиксальное образование	<i>re-parameter-ization</i> (перепараметризация), <i>de-token-ization</i> (детокенизация)
	Конверсия	<i>biometric</i> , adj. → <i>biometrics</i> (биометрия), n., <i>record</i> , v. → <i>record</i> (запись), n.
	Фонетико-морфологическое образование	<i>Poissonian noise</i> (шум Пуассона), <i>gaussianization</i> (приведение распределения к гауссовому виду)
	Усечение	<i>config</i> ← <i>configuration</i> (конфигурация), <i>hyperparams</i> ← <i>hyperparameters</i> (гиперпараметры)
Синтаксические	Двухкомпонентные термины-словосочетания	<i>deep learning</i> (глубокое обучение), <i>supervised learning</i> (контролируемое обучение)
	Многокомпонентные термины-словосочетания	<i>digital social innovation</i> (цифровые социальные инновации)
Морфолого-синтаксические	Словосложение	<i>dataset</i> (набор данных), <i>timestep</i> (шаг времени)
	Эллипсис	<i>benchmark program</i> → <i>benchmark</i> (бенчмарк)
	Аббревиация (инициализмы/акронимы)	<i>AI</i> (ИИ), <i>ANN</i> (ИНС)

Обобщение изложенного показывает, что терминообразование – многоуровневый и исторически детерминированный процесс, проходящий две взаимосвязанные стадии (первичное присвоение имени и последующую деривационно-семантическую эволюцию), где решающими выступают строгая дефиниция и лексико-семантическая структура языка. При этом в постнеклассической науке термины формируются внутри «пучков» дисциплин и запускаются новыми проблемными ситуациями. Стремление к моносемии сосуществует с жизненным циклом термина (рождение – зрелость – угасание) и допустимостью «ложно ориентирующих» наименований, временно обеспечивающих коммуникацию. Морфологическая деривация (аффиксация, конверсия, словосложение) придает формам онтологическую «разметку» ролей (процесс/результат/агент/свойство) и повышает прозрачность терминосистем, тогда как синтаксические и морфолого-синтаксические модели (A+N, N+N, эллипсис, аббревиация) объясняют преобладание многословных единиц. Семантические механизмы – терминологизация общеупотребительной лексики, расширение и специализация значения, метафоризация, метонимический перенос, а также заимствования – описывают вторичное конструирование и гармонизацию терминов между сообществами и языками. Интегрированная классификация, объединяющая формальные и когнитивно-прагматические параметры, тем самым задает операциональную схему описания терминологии ИИ.

1.5. Эволюция терминологии в области искусственного интеллекта: основные этапы формирования

Становление терминологии искусственного интеллекта (ИИ) началось в середине XX века, опираясь на математические и нейробиологические модели автоматизации когнитивных функций [McCulloch, Pitts 1943]. Д. Хебб заложил основу теории обучения через укрепление нейронных связей [Hebb 2002], а А. Тьюринг сформулировал поведенческий критерий оценки машинного разума через «имитационную игру», известную как *тест Тьюринга* (Turing test) [Turing 1950]. Окончательное оформление ИИ как научного направления произошло в 1940-х годах в рамках кибернетической традиции Н. Винера, объединившей принципы управления в живых организмах и машинах [Wiener 1961]. На этом этапе был сформирован

базовый терминологический фонд дисциплины, включивший такие универсальные концепты, как *интеллект* (intelligence), *обучение* (learning), *алгоритм* (algorithm), *информация* (information), *обратная связь* (feedback), *нейрон* (neuron).

Институционализация дисциплины закрепила благодаря Дж. Маккарти, внедрившему термин *искусственный интеллект* (Artificial Intelligence) на Дартмутском семинаре [McCarthy 1955]. Ранние инженерные разработки середины XX века способствовали формированию базового понятийного аппарата, включая лексику обучения и подкрепления, а также закрепили представление о вычислительных моделях адаптивного поведения. В это же время в научный обиход вошел термин *машинное обучение* (machine learning), получивший демонстративное воплощение в самообучающейся программе для игры в шашки [Samuel 1959]. Линия развития от механических устройств и кибернетических концепций к цифровым системам обусловила переход от описания изолированных механизмов к моделям усиления интеллектуальной деятельности (например, *человек-в-контуре* (human-in-the-loop), *поддержка принятия решений* (decision support), *когнитивное усиление* (cognitive augmentation)).

В 1960–1970-х годах доминировал символический ИИ, что закрепило в языке единицы *логика* (logic), *эвристика* (heuristic), *представление знаний* (knowledge representation), а создание языка обработки списков *Лисп* (Lisp) Дж. Маккарти ввело технические термины *S-выражение* (S-expression), *интерпретатор* (interpreter), *списочная обработка* (list processing), *сборка мусора* (garbage collection) [Russell, Norvig 2021]. Период «зимы ИИ» (AI winter), вызванный дефицитом ресурсов, сместил фокус на *оценку реализуемости* (feasibility) и *пределов производительности* (performance bounds) систем [National Research Council 1966].

Новый импульс 1980-х годов связан с экспертными системами (DENDRAL, MYCIN, XCON), которые внедрились в оборот понятия *база знаний* (knowledge base), *прямой вывод* (forward chaining), *инженерия знаний* (knowledge engineering), *коэффициент уверенности* (certainty factor) [Баканач 2023; 2024]. Таким образом, терминология этого этапа отразила переход от универсальных алгоритмов поиска (*поиск в глубину* (depth-first search, DFS) и *поиск в ширину* (breadth-first search, BFS)) к прикладным системам, основанным на правилах и специфических знаниях о предметной области.

В 1990-е годы в профессиональном дискурсе утвердился термин *интеллектуальные агенты* (intelligent agents), обозначающий системы, способные автономно выполнять задачи, поставленные человеком [Russell, Norvig 2021]. Показательным событием стало противостояние 1997 года, завершившееся победой компьютера Deep Blue над Г. Каспаровым. Этот эпизод способствовал закреплению в специальном узусе единиц *высокопроизводительный поиск* (high-performance search), *оценочная функция* (evaluation function), *параллельная архитектура* (parallel architecture) [Campbell 2002].

Новое тысячелетие ознаменовалось ускорением развития машинного обучения. Рост *объемов данных и вычислительных ресурсов* (big data; GPU computing), а также прогресс архитектур *глубокого обучения* (deep learning) обеспечили резкий качественный скачок в *компьютерном зрении* (computer vision) и *обработке/понимании естественного языка* (natural language processing/understanding) [Баканач 2023; 2024]. В задачах компьютерного зрения прорыв был связан с внедрением *сверточных нейронных сетей* (convolutional neural networks) на масштабных размеченных наборах (например, *ImageNet*), тогда как для последовательностного моделирования закрепились *рекуррентные нейронные сети* (recurrent neural networks) и архитектура *долгая краткосрочная память* (long short-term memory, LSTM) [Баканач 2023; 2024]. В области *обучения с подкреплением* (reinforcement learning) переход к *глубокому обучению с подкреплением* (deep reinforcement learning) позволил достичь «человеко-сопоставимого» контроля в играх Atari. Параллельно развивались генеративные модели – *генеративно-сопоставительные сети* (generative adversarial networks), а также практики *трансферного обучения* (transfer learning) и *автоматизированного машинного обучения* (automated machine learning / AutoML). На прикладном уровне распространение получили *интернет вещей* (Internet of Things) и *виртуальные/голосовые ассистенты* (virtual assistants).

Статья «Attention Is All You Need» зафиксировала в профессиональном лексиконе термины *трансформер* (transformer), *самовнимание* (self-attention), *многоголовое внимание* (multi-head attention), *позиционное кодирование* (positional encoding) и др. [Vaswani 2017]. Позднее выделились типологии по архитектуре: *только кодировщик* (encoder-only, BERT), *только декодер* (decoder-only, GPT), *кодировщик–декодер* (encoder–decoder, T-5), коррелирующие с задачами понимания, генерации и

преобразования текста. Практики адаптации и согласования моделей кодифицировали понятия *предобучение* (pre-training), *контролируемая доводка/дообучение* (supervised fine-tuning, SFT), *настройка на инструкции* (instruction-tuning), *согласование модели* (alignment) и др.

К новым исследовательским направлениям современного этапа относятся *объяснимый искусственный интеллект* (explainable AI, XAI), а также *федеративное обучение* (federated learning) и сопряженные с ним терминологические кластеры *сохранение приватности* (privacy-preserving), *на устройстве* (on-device), *усреднение моделей* (model averaging) [McMahan 2016]. С 2021 года в научном дискурсе используется зонтичный термин *базовые модели* (foundation models), обозначающий модели, обученные на масштабных данных и адаптируемые к широкому спектру задач.

Одновременно усиливается интеграция ИИ в повседневные и институциональные практики, что актуализирует проблематику этики, безопасности и регулирования и способствует оформлению соответствующих направлений: *ответственный ИИ* (responsible AI), *этичное применение ИИ* (ethical use of AI), *универсальный/сильный ИИ* (artificial general intelligence), *ИИ для обнаружения лекарств* (AI for drug discovery), *ИИ в здравоохранении* (AI-assisted healthcare), *нейроморфные вычисления* (neuromorphic computing).

В 2020-е годы регулятивная политика ЕС оказывает системное влияние на терминологию ИИ. Регламент Artificial Intelligence Act (Regulation (EU) 2024/1689) закрепляет риск-ориентированную таксономию: *недопустимый риск* (unacceptable risk), *высокий риск* (high-risk), *ограниченный риск* (limited risk) – и вводит нормативный аппарат требований и процедур, включая термины *обязательства* (obligations), *прозрачность* (transparency), *оценка соответствия* (conformity assessment) и др. [Regulation (EU) 2024].

Таким образом, эволюция терминологии ИИ представляет собой непрерывный, многоуровневый процесс, соотносящийся с фазами развития самой дисциплины: от кибернетики и символического ИИ к статистическим методам и глубокому обучению, далее – к интеграции в социотехнические экосистемы и регулятивные рамки. Терминологический инвентарь при этом демонстрирует нарастающую интернационализацию (греко-латинские форманты), рост доли многословных единиц и продуктивность серийных деривационных моделей, а также активные семантические

механизмы (метафоризация, эпонимизация и др.). Историческая траектория – от ранних кибернетических и вычислительных моделей к институционализации термина Artificial Intelligence в рамках Дартмутского проекта – показывает, что понятийные новации быстро отражаются в терминологии и формируют ее системные связи (родовые, меронимические, коллокационные). В результате складывается комплексная терминосистема, где устойчивость и стандартизация сосуществуют с динамикой и вариативностью.

1.6. Гармонизация терминологии ИИ в системе международной и национальной стандартизации

Стратегическая значимость развития технологий искусственного интеллекта общепризнанна на национальном и наднациональном уровнях; указанная оценка институционально формализована в соответствующих стратегиях и нормативно-правовых актах. Так, в России цели и задачи в сфере ИИ закреплены в «Стратегии научно-технологического развития Российской Федерации» [О Стратегии НТР 2016], «Стратегии развития информационного общества на 2017–2030 годы» [О стратегии развития 2017], и Указе Президента РФ № 490 от 14 октября 2019 г. «О развитии искусственного интеллекта в Российской Федерации», на основании которого была принята «Национальная стратегия развития искусственного интеллекта на период до 2030 года» [О развитии 2019].

Принятый в Европейском союзе общеобязательный Регламент (EU) 2024/1689 (AI Act) устанавливает гармонизированные, риск-ориентированные правила разработки и применения систем ИИ; документ опубликован 12 июля 2024 г. в “Official Journal” и вступает в действие поэтапно [Regulation (EU) 2024]. Во Франции национальная стратегия “AI for Humanity” и последующие инициативы в рамках “France 2030” закрепляют приоритетные сектора (здравоохранение, транспорт, экология, оборона/безопасность) и меры по развитию инфраструктуры данных и кадрового потенциала [Donner 2018; France 2030]. В Германии действует «Стратегия ИИ федерального правительства» 2018 года с обновлением 2020 года, фокусирующаяся на исследованиях, трансфере и ответственности применения ИИ [Die Bundesregierung 2018; Die Bundesregierung 2020]. В США регулятивную рамку формируют указ президента о безопасной, надежной и ответственной разработке и

использовании ИИ (30.10.2023) и добровольная, но референтная модель управления рисками ИИ NIST (AI RMF 1.0), закрепляющая принципы «доверенного ИИ» и практики оценивания рисков [Artificial Intelligence RMF 2023; Safe, Secure 2023]. В Китае стратегические ориентиры заданы «Планом развития ИИ нового поколения» и последующими нормативами, в частности, «Временными мерами по управлению сервисами генеративного ИИ», которые устанавливают требования к публичным сервисам на основе генеративных моделей [State Council Notice 2017; Interim Measures 2023].

Переход от стратегических и правовых установок к практической реализации политики в области ИИ обеспечивается стандартизацией как механизмом формализации требований к системам ИИ и их интеграции в различные сферы жизнедеятельности [Куприков, Башкирова 2022]. В отличие от рекомендательных практик, стандартизация предполагает нормативную фиксацию терминов и их дефиниций, обеспечивающую однозначность соотнесения термина и определения [Казарина, Казарина 2008].

До 2010-х годов национальные и международные стандарты в области ИИ отсутствовали, а терминология формировалась преимущественно в научной и отраслевой литературе. При этом ключевые концепты (*искусственный интеллект, экспертная система, нейронные сети*) не имели строгих определений. Начало системной стандартизации связано с созданием в 2017 году технического подкомитета ISO/IEC JTC 1/SC 42 в рамках Международной организации по стандартизации (ISO) совместно с Международной электротехнической комиссией (IEC) для разработки международных стандартов. Базовым нормативным источником является ISO/IEC 22989:2022 (Artificial intelligence – Concepts and terminology), фиксирующий понятийно-терминологический каркас ИИ на английском языке и предназначенный для разработки последующих стандартов и междисциплинарной коммуникации [ISO/IEC 22989:2022]. Комплементарным документом выступает ISO/IEC 23053:2022 (Framework for AI Systems Using Machine Learning), обеспечивающий согласованное описание архитектуры, компонентов и функций ML-систем [ISO/IEC 23053:2022]. В части управления рисками ключевым ориентиром служит ISO/IEC 23894:2023 (AI – Guidance on risk management), задающий процедуры интеграции риск-менеджмента в жизненный цикл ИИ-систем [ISO/IEC 23894:2023]. Методологическую основу

терминологической работы формируют ISO 704:2022 и ISO 1087:2019, на которые опираются при построении дефиниций и установлении правил образования терминов [ISO 1087:2019; ISO 704:2022]. В данной рамке стандартизация понимается как нормативная фиксация терминоединиц и их дефиниций, обеспечивающая воспроизводимость междисциплинарной коммуникации [Куприков, Башкирова 2022].

По состоянию на 2025 год подкомитетом SC 42 опубликовано 39 международных стандартов, охватывающих ключевые аспекты функционирования систем ИИ – от управления и терминологии до жизненного цикла и оценки рисков. Кроме того, готовится проект поправки ISO/IEC 22989:2022/DAmD 1 «Generative AI», призванный инкорпорировать терминологию, связанную с генеративным ИИ, и согласовать ее с актуальными направлениями работ подкомитета. Это отражает тенденцию к ускоренному пересмотру понятий в связи с технологическими изменениями. Особенностью стандартизации в области ИИ является сокращенный жизненный цикл терминов и необходимость оперативной унификации между практическим употреблением и нормативной фиксацией во избежание отставания от индустрии.

Координацию разработки и внедрения нормативной базы осуществляет подкомитет SC 42, выступающий основной площадкой для программ ИИ в ISO/IEC и взаимодействующий с региональными структурами (в т. ч. по линии CEN/CENELEC в Европейском союзе), что облегчает международную гармонизацию и унификацию понятий при адаптации к национальным правовым режимам. Под гармонизацией в данном контексте понимается обеспечение сопоставимости научно-технической терминологии на национальном и международном уровнях, что предполагает согласование понятийного аппарата, устранение дублирования и противоречий, а также выработку единых подходов к его интерпретации [Гринев-Гриневиц 2008; Суперанская 2012]. Данный подход позволяет интегрировать международные стандарты в национальные нормативные системы без потери смысловой точности терминов и обеспечивает единое профессиональное пространство для специалистов разных стран.

В Европейском союзе стандарты ISO/IEC принимаются комитетом CEN-CENELEC/JTC 21 как EN ISO/IEC для целей презумпции соответствия AI Act; показательный пример – терминологический стандарт EN ISO/IEC 22989 (Concepts and

terminology) и сопутствующие рамочные документы для ML-систем (EN ISO/IEC 23053) и менеджмент-систем ИИ (EN ISO/IEC 42001) [ISO/IEC 22989:2022; ISO/IEC 23053:2022; ISO/IEC 42001:2023]. На национальном уровне в Европе стандарты ISO/IEC внедряются без изменений, обеспечивая единообразие дефиниций: DIN EN ISO/IEC 22989 (Германия), BS EN ISO/IEC 22989 (Великобритания), NF EN ISO/IEC 22989 (Франция), – что поддерживает единообразие дефиниций и понятийного аппарата в научно-технической документации и регуляторике [DIN EN ISO/IEC 22989; BS EN ISO/IEC 22989:2023; NF EN ISO/IEC 22989:2023]. В США зеркальный комитет INCITS публикует идентичные версии стандарта ISO/IEC 22989:2022, позволяя использовать тот же терминологический каркас в практиках управления рисками и соответствия [INCITS 2023]. В Китае действует национальный терминологический стандарт GB/T 41867-2022 для ИИ; он служит национальной основой для унификации определений и согласуется с международной повесткой при двуязычных проектах и переводах [GB/T 41867-2022].

В России национальные стандарты на русском языке разрабатываются Росстандартом (Федеральное агентство по техническому регулированию и метрологии); ключевым субъектом выступает технический комитет по стандартизации «Искусственный интеллект» (ТК 164), учрежденный в 2019 году приказом Росстандарта (№ 1732 от 25.07.2019) и функционирующий как зеркальная структура подкомитета SC 42. К достижениям ТК 164 относятся утверждение базовых национальных стандартов: ГОСТ Р 59276–2020 Искусственный интеллект. Доверие к системам искусственного интеллекта. Концепции и терминология [ГОСТ Р 59276–2020], ГОСТ Р 59277–2020 Искусственный интеллект. Классификация систем искусственного интеллекта [ГОСТ Р 59277–2020], ГОСТ Р 59278–2020 Искусственный интеллект. Применение в интерактивных руководствах [ГОСТ Р 59278–2020], а также участие в ключевых международных мероприятиях (ЦИПР, AI Global Dimension) и интеграция в работу ISO/IEC JTC 1/SC 42. За синхронизацию отечественных требований к качеству и управлению данными с разрабатываемыми в ISO/IEC сериями отвечает подкомитет ПК 02 «Данные».

Совокупность реализуемых мер формирует целостную и согласованную терминологическую и методическую основу в области искусственного интеллекта, в рамках которой международные стандарты задают концептуальные и процедурные

ориентиры, а национальные документы обеспечивают их адаптацию и закрепление в нормативно-правовой и профессиональной практике. В этом контексте стандартизация выступает не только средством унификации, но и механизмом концептуализации: фиксирует ключевые дефиниции, поддерживает межъязыковую эквивалентность, служит базой для онтологий и контролируемых словарей и влияет на юридическую определенность, этические принципы и процедуры оценки соответствия.

Для нашего корпусного исследования терминологии ИИ стандартизированные документы являются опорным источником, позволяющим выделить ядро понятийного аппарата, проверить устойчивость дефиниций в различных контекстах и выявить расхождения между нормативной фиксацией и реальным употреблением в научной коммуникации.

ВЫВОДЫ ПО ГЛАВЕ 1

Анализ литературы по проблеме диссертационной работы послужил основой для теоретико-методологического обоснования данного исследования, отраженного в следующих выводах.

1. Терминоведение трактуется как междисциплинарная область, объединяющая лингвистику, логику и предметно-научные подходы; эмпирическое описание опирается на корпусно-вычислительные процедуры (АТЕ, коллокационный анализ и др.) и согласуется с международными стандартами терминологической работы (ISO 704; ISO 1087), что гарантирует воспроизводимость и сопоставимость результатов.

2. Под терминологией понимается широкий набор специализированных наименований (кодифицированные термины и номены), описываемых в единстве языковой формы и понятийного содержания; терминосистема – структурная проекция того же массива в плоскости внутренних отношений (родо-видовых, меронимических, ассоциативных), отражающих организацию системы понятий. В интересах экспозиционной экономии оба термина используются как функциональные эквиваленты там, где различие не влияет на выводы; при необходимости оно оговаривается специально. Принята структурно-системная позиция: терминология – автономная, но пронизываемая подсистема лексики, взаимодействующая с общим языком и обладающая собственной организацией.

3. Термин определяется как единица языка для специальных целей (LSP) с набором дифференциальных признаков: соотнесенность с понятием в системе домена, дефинированность, нормативная установка на однозначность, стилистическая нейтральность, формальная устойчивость, кодифицируемость. Эти признаки используются как операциональные критерии отбора и описания единиц в корпусе; наблюдаемая вариативность рассматривается как эмпирический факт и регулируется через дефиниции и контекстную проверку.

4. Способы терминообразования в терминологии ИИ систематизированы в интегрированной классификации, охватывающей семантические, морфологические, синтаксические и морфолого-синтаксические механизмы; данная схема задает маршрут анализа появления, адаптации и специализации терминов и обеспечивает сопоставимость продуктивности моделей.

5. Предложена многоуровневая система классификационных оснований, релевантных корпусной аннотации: функционально-дискурсивный, содержательно-онтологический, формально-структурный и частеречный, семантический, регистро-нормативный и частотно-дисперсионный уровни. Их интеграция образует согласованную схему описания, необходимую для онтологического моделирования и лексикографической фиксации.

6. Эволюция терминологии ИИ реконструируется как нелинейный процесс, коррелирующий со сменой научно-технологических парадигм: от кибернетики и символического ИИ через экспертные системы и машинное обучение к глубокому обучению и крупномасштабным языковым моделям. Зафиксированы тенденции к интернационализации, рост доли именных многословных единиц и высокая роль семантических механизмов (метафоризация, эпонимизация).

7. Стандартизация выступает ключевым механизмом концептуализации и гармонизации понятийного аппарата: международные и национальные нормы (в т. ч. ISO/IEC 22989; ISO/IEC 23053; ISO/IEC 23894) формируют регулятивный каркас, обеспечивающий однозначность, воспроизводимость и сопоставимость употребления в научной, инженерной и регуляторной коммуникации.

ГЛАВА 2. МЕТОДОЛОГИЯ ПРОВЕДЕНИЯ КОРПУСНОГО ИССЛЕДОВАНИЯ ТЕРМИНОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

Актуальность изучения терминологии предметной области «искусственный интеллект» обуславливает применение корпусных методов, обеспечивающих анализ терминов в их естественных контекстах функционирования. В данной главе изложены методологические основания исследования, включающие построение репрезентативного корпуса научных текстов, посвященных ИИ, извлечение кандидатов в термины с помощью инструмента AntConc и их последующую обработку с использованием статистических метрик в сочетании с качественной лингвистической верификацией.

Под кандидатами в термины понимаются однословные и многословные лексические единицы, автоматически выделенные из корпуса по совокупности распределительных и статистических признаков (нормализованная частотность, показатели ключевости (keyness), меры ассоциативной силы (Mutual Information, Log-Likelihood), устойчивость сочетаемости в KWIC) и подлежащие экспертному отсеву с исключением служебных и общенаучных маркеров. Отдельное внимание уделяется интеграции полученных результатов в авторский монолингвальный онлайн-гlossарий, отражающий тематическое поле современной терминологии ИИ.

При этом корпусные метрики и программные инструменты рассматриваются в работе как средство лингвистической верификации, а не как самостоятельный объект описания. Частотные и коллокационные показатели используются для операционализации критериев терминологичности и устойчивости, выделения терминологического ядра и обоснованного отбора единиц, подлежащих дальнейшему структурно-семантическому анализу.

2.1. Инструментарий и методы корпусного анализа терминологии

2.1.1. Определение понятий «языковой корпус» и «корпусный менеджер»

Корпусная лингвистика рассматривается в современной науке о языке как один из наиболее продуктивных эмпирических подходов к изучению языка и дискурса,

поскольку опирается на анализ больших массивов текстовых данных и тем самым позволяет переходить от интуитивных обобщений к проверяемым наблюдениям. По оценке А. П. Кононенко и Л. А. Недосека, исследователи все активнее используют корпусные технологии как инструментарий решения конкретных научных задач [Кононенко, Недосека 2023]. Практика применения корпусных методов охватывает широкий спектр направлений – от лингвистического анализа дискурсивных практик и задач преподавания/изучения языка до междисциплинарных гуманитарных исследований, что демонстрирует масштаб и вариативность методов корпусной обработки и интерпретации данных [O’Keeffe, McCarthy 2010; Давидюк 2024; Moreno-Ortiz 2024].

В теоретико-методическом плане корпус обычно понимается как структурированная коллекция текстов, потенциально сопровождаемая разметкой (морфологической, синтаксической, семантической), то есть как специально организованный массив данных, обеспечивающий систематическое наблюдение языковых единиц в их естественном употреблении [O’Keeffe, McCarthy 2010]. В ранней корпусной традиции, связанной с проектированием Brown Corpus, корпус трактовался как совокупность текстов, репрезентативных для языка, диалекта или подмножества естественного языка и предназначенных для лингвистического анализа [Захаров 2005]. В более поздней формулировке Дж. Синклера акцентируется принцип селекции материала ради исследования вариативности языка, то есть корпус мыслится как коллекция текстов естественного языка, целенаправленно выбранных для изучения разнообразия употребления [Sinclair 1991]. В свою очередь Н. В. Козлова подчеркивает, что корпус включает оцифрованные как письменные, так и устные высказывания, хранящиеся на специальных платформах и доступные в электронном виде [Козлова 2013]. В рамках современного определения Л. Селивана корпус описывается как собрание реальных образцов использования языка в виде текстов, специально предназначенное для его изучения [Selivan 2023]; данная установка принципиальна и для прикладных задач, поскольку в образовательной и исследовательской практике корпусные данные дополняют и корректируют традиционные ресурсы (словарные и учебные описания), а также используются при проектировании учебных материалов и заданий, основанных на наблюдении за реальным языком [Vyatkina 2024; Pérez-Paredes, Boulton 2025].

Операционально наиболее емкая дефиниция, задающая рамку для методического проектирования корпуса, представлена у В. П. Захарова: «...под названием лингвистический, или языковой корпус текстов понимается большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [Захаров 2005: 3]. Существенно, что в этой трактовке корпус включает как массив текстов, так и систему управления текстовыми и лингвистическими данными – корпусный менеджер, задача которого состоит в поиске данных по корпусу, получении статистической информации и предоставлении результатов пользователю в удобном формате [Там же: 3].

Данный функционал успешно реализован в ряде широко используемых платформ – Sketch Engine, Open Corpus Workbench (BNCweb), ANNIS, Корп (Språkbanken), а также в Национальном корпусе русского языка. Эти системы, являясь готовыми исследовательскими инфраструктурами, обеспечивают продвинутый лингвистический поиск, автоматическую статистику (включая коллокации), работу с подкорпусами и воспроизводимый экспорт результатов. Так, Written BNC2014 (через CQPweb) спроектирован как сопоставимый корпус к BNC1994 и применяется для диахронических и жанровых сопоставлений при фиксированном дизайне и метаданных [Brezina 2021]. Инфраструктура Корп служит основой для построения диахронических эмбедингов по газетным массивам и картирования семантических сдвигов; среда ANNIS демонстрирует работу с многоуровневыми аннотациями и сложными AQL-запросами на материале жестового языка. Версия 2.0 НКРЯ (2020–2023) предлагает обновленные корпуса с нейросеточной аннотацией и модернизированными интерфейсными решениями, а в экосистеме Sketch Engine доступны веб-корпусы для мониторинга трендов (Timestamped JSI, TenTen с обновлениями 2021–2022), что позволяет оперативно отслеживать терминологические новации.

Поскольку сопоставимый по тематике и параметрам корпус текстов по искусственному интеллекту в готовом виде не выявлен, использование преднастроенных платформ затруднительно. В связи с этим корпус создается специально под задачи исследования и развертывается в корпусном менеджере, поддерживающем загрузку и индексацию авторского текстового массива.

Таким, образом, в работе под «корпусом» понимается специально сформированная под исследовательскую задачу электронная коллекция текстов, структурированная и документированная метаданными и, при необходимости, лингвистической разметкой, обеспечивающая воспроизводимый доступ к контекстам и количественным показателям употребления. Под «корпусным менеджером» – программная среда для загрузки и нормализации авторского массива, его индексации и управления метаданными/версиями, выполнения формализованных запросов с учетом лемматизации и синтаксической структуры, получения стандартных статистик (конкордансы, частотные и ключевые списки, коллокации, дисперсия) и экспорта результатов вместе с параметрами поиска. Такое операциональное разграничение задает рамку последующих методических решений и описания проектирования специализированного корпуса для анализа структурно-семантических особенностей терминологии ИИ.

2.1.2. Обоснование выбора программного комплекса AntConc как инструментария корпусного анализа и обзор его функциональных модулей

История инструментов для работы с электронными корпусами текстов (корпусных менеджеров и конкордансеров) тесно связана с развитием корпусной лингвистики и вычислительной техники. По мере роста объемов цифровых данных и усложнения исследовательских задач менялись принципы хранения, доступа и анализа текстов, а также требования к интерфейсу и воспроизводимости процедур. В ранний период становления корпусных исследований ключевую роль играли специализированные программные решения, ориентированные на базовые операции – построение конкордансов и подсчет частотности, что обеспечивало эмпирическую основу для описания языковых закономерностей (например, COCOA, CLOC) [Захаров 2005; O’Keeffe, McCarthy 2010]. Характерной особенностью таких инструментов являлась функциональная ограниченность и зависимость от конкретной вычислительной среды, что затрудняло интерактивную работу с данными и переносимость процедур между корпусами.

Распространение персональных компьютеров и развитие пользовательских интерфейсов изменили практику корпусного анализа. Корпусные инструменты стали интерактивными, а их функционал расширился за счет модулей поиска,

статистической обработки, выявления ключевых слов, анализа коллокаций и тематических связей (например, TACT, WordSmith Tools, MonoConc). В результате корпусный анализ стал в большей степени опираться не только на операции «поиска и подсчета», но и на комплексные процедуры описания контекстов употребления, сравнения подкорпусов и формирования интерпретационных гипотез. Данное обстоятельство соответствует современному пониманию корпусной методологии как сочетания количественного и качественного анализа [O’Keeffe, McCarthy 2010; Давидюк 2024].

Следующий этап связан с развитием веб-технологий и распределенных вычислений, которые актуализировали удаленный доступ к корпусам, совместную работу и стандартизированные интерфейсы анализа. Веб-платформы и облачные решения сделали корпусные данные более доступными, а инструменты – более ориентированными на масштабируемость и интеграцию с различными типами лингвистической разметки и аналитическими модулями (например, BNCweb, Sketch Engine, Voyant Tools). Это повлекло за собой диверсификацию спектра задач – от быстрых разведочных процедур (KWIC, тренды, ключевые слова) до более сложных моделей извлечения паттернов и коллокационных профилей, при сохранении требований к прозрачности исследовательских шагов и воспроизводимости результатов [Петрова 2022; Давидюк 2024].

Параллельно развивались свободно распространяемые и открытые решения, ориентированные на работу с пользовательскими данными и исследовательскую гибкость (например, AntConc, Corpus Workbench/CWB, LancsBox), а также специализированные среды для аннотирования мультимодальных данных (например, ELAN), что расширило корпусную методологию за пределы чисто текстовых массивов. В целом эволюция корпусных инструментов может быть описана как переход от узкоспециализированных решений для базовых операций к многофункциональным интерактивным средам и веб-платформам, обеспечивающим масштабируемый анализ и поддержку разнообразных процедур количественного описания. Этот процесс способствовал институционализации корпусных методов в гуманитарных и социальных исследованиях и закрепил корпусные инструменты как обязательный компонент исследовательской инфраструктуры [Suleimanova 2020; Suleimanova, Petrova 2020; Петрова 2022].

Критический анализ функционального потенциала и методологических ограничений корпусных менеджеров послужил основанием для выбора AntConc в качестве основного инструмента данного исследования. К числу его ключевых преимуществ относятся доступность для пользователей с различным уровнем подготовки, интуитивно понятный интерфейс, широкий спектр функций для текстового анализа (включая конкорданс, коллокации, частотный анализ, кластеризацию, создание частотных словарей и выделение ключевых слов), поддержка различных текстовых форматов, возможность гибкой настройки параметров анализа и кросс-платформенная совместимость [Suleimanova 2020; Anthony 2024]. Помимо основных перечисленных преимуществ AntConc позволяет исследовать новейшие языковые явления на самостоятельно собранном и подготовленном для анализа материале.

Программный пакет AntConc предоставляет доступ к десяти функциональным модулям, интегрированным в графический интерфейс приложения [Anthony 2024]. Данные инструменты становятся доступными для использования через специализированную панель управления после завершения процедуры инсталляции программного обеспечения на локальный компьютер. В Таблице 3 приведены основные функции платформы AntConc.

Таблица 3 – Основные функции корпусного менеджера AntConc

Функция	Описание	Применение
KWIC (Key Word In Context)	Показывает ключевое слово в контексте (слева и справа)	Анализ употребления слова в разных контекстах, изучение коллокаций
Plot	Визуализирует распределение слова/фразы по тексту в виде штрих-кода	Определение частоты и местоположения слова в документе
File View	Отображает полный текст файла с подсветкой выбранных слов	Поиск конкретных примеров употребления в оригинальном тексте
Cluster	Отбирает группы слов с заданным количеством элементов слева и справа от целевого слова	Выявление устойчивых словосочетаний, клише, шаблонов
N-Gram	Анализирует фиксированные последовательности из N слов (например, биграммы = 2 слова, триграммы = 3 слова и т.д.)	Поиск часто встречающихся фраз (биграммы, триграммы и т. д.)

Функция	Описание	Применение
Collocate	Выявляет слова, которые часто встречаются рядом с заданным словом	Изучение коллокаций, семантических связей между словами
Word List	Отображает список всех слов в корпусе с их частотностью	Исследование лексического состава и частотных характеристик текста
Keyword List	Сравнивает частотность слов в корпусе с эталонным корпусом	Выявление ключевых слов, характерных для данного текста
Wordcloud	Визуализирует частотность слов в виде облака тегов	Быстрое представление самых употребительных слов
ChatAI	Предоставляет доступ к большим языковым моделям (LLM) посредством создания промптов	Использование ИИ для ответов на общие лингвистические вопросы, поиска закономерностей, углубленного анализа

В рамках настоящего исследования используется комплекс инструментов AntConc, отобранных по принципу функциональной релевантности лингвистическим задачам описания терминологической системы ИИ; поэтому обзор модулей ограничивается теми возможностями, которые непосредственно обеспечивают получение и интерпретацию результатов. Во-первых, генерация частотных списков (Word List) позволяет выделить терминологическое ядро области и наметить его периферию на основе распределительных характеристик, что служит первичным этапом отбора терминологических кандидатов. Во-вторых, конкордансный анализ (KWIC) обеспечивает проверку терминологического статуса единиц через их употребление в естественных контекстах и позволяет уточнять контекстную семантику, в том числе при разграничении значений и вариантов употребления. В-третьих, анализ коллокаций (Collocate) используется для выявления статистически значимых сочетаний, что дает возможность фиксировать устойчивые многокомпонентные термины и типовые синтаксические модели номинации. Наконец, функция выделения ключевых слов (Keyword) позволяет обнаружить лексические доминанты дискурса и оценить их вклад в тематическое и концептуальное структурирование научных текстов по ИИ. Перечисленные процедуры обеспечивают системное и всестороннее исследование терминологической системы искусственного интеллекта.

Таким образом, выбор AntConc обусловлен достаточным функциональным охватом, доступностью и методологической прозрачностью, что позволяет выстроить

воспроизводимый протокол корпусного исследования терминологии ИИ. Инструмент поддерживает базовые процедуры идентификации и систематизации единиц и обеспечивает удобный экспорт результатов в универсальные аналитические среды, тем самым повышая надежность и сопоставимость данных.

2.1.3. Параметризация терминологического аппарата: алгоритмы и метрики статистической оценки данных

Ключевой особенностью корпусных лингвистических исследований выступает применение разнообразных статистических методов для анализа языковых единиц в рамках определенного языкового корпуса [Захаров 2005; Gries 2024], позволяющих получить надежные и верифицируемые данные при условии использования одного и того же материала исследования и компьютерной программы. В отличие от методов традиционной лингвистики, обращающейся преимущественно к интроспективным и интерпретативным методам при проведении лингвистического анализа, применение статистики помогает избежать субъективности результатов [Теоретические и прикладные 2016].

К числу наиболее распространенных статистических мер, которыми располагают современные корпусные менеджеры, относятся: абсолютная и нормализованная частотность (Frequency, Normalized Frequency), коэффициент взаимной информации (Mutual Information, MI), куб взаимной информации (MI^3), T-критерий (T-score), Z-критерий (Z-score), метод максимального правдоподобия (Log-Likelihood), коэффициент Дайса (Dice), логарифмическое соотношение (Log Ratio) и др. [Anthony 2023].

Вместе с тем использование методов математической статистики в лингвистике сопряжено с рядом методологических сложностей, что обуславливает скептическое отношение некоторых исследователей к представлению языковых явлений в формате числовых данных [Gries 2024]. Во-первых, эффективность статистического анализа зависит от корректного выбора метрик, соответствующих конкретным исследовательским задачам. Во-вторых, количественные данные, полученные из корпусов, требуют содержательной лингвистической интерпретации. В-третьих, существующий инструментарий нуждается в непрерывном дальнейшем

совершенствовании для повышения точности описания языковых процессов и проверки научных гипотез.

Одним из основных преимуществ программы AntConc является высокая степень интеграции статистических мер непосредственно в пользовательский интерфейс. Это позволяет исследователю производить количественный и сопоставительный анализ без необходимости экспортировать данные в сторонние приложения. Все вычисления осуществляются автоматически, а результаты отображаются в виде упорядоченных таблиц в таких инструментах, как Collocate, Cluster, N-Gram, Keyword, Word List. Интерфейс предлагает быстрый выбор нужной метрики из выпадающего списка, что упрощает смену критерия сортировки и адаптацию аналитической стратегии под конкретные цели исследования. Например, при анализе корпуса научных публикаций по машинному обучению использование меры MI выявляет устойчивые терминологические пары (*training set, neural network*), тогда как T-score и Log-Likelihood акцентируют более частотные сочетания с широкой сочетаемостью. В Таблице 4 представлена классификация основных статистических инструментов, доступных в AntConc, с указанием их функционального назначения.

Таблица 4 – Основные статистические меры корпусного менеджера AntConc

Категория	Название меры	Назначение
Частотные показатели	Абсолютная частотность	Количество вхождений единицы в корпус
	Нормализованная частотность	Количество вхождений на 1 млн токенов/словоупотреблений; позволяет сравнивать разные корпуса
Коллокационная связанность	Mutual Information (MI)	Сила ассоциации между словами (насколько тесно связаны)
	MI ³ (MI cubed)	Уточненная версия MI, более чувствительная к сильным связям
	T-score	Учитывает не только силу ассоциации, но и надежность частотных данных
	Log-Likelihood (LL)	Статистическая значимость связи между словами
Ключевость (keyness)	Dice, Log Dice	Баланс частот и совместных вхождений
	LL (G ²), биномиальный критерий	Значимость различий относительных частот между основным и эталонным корпусами
Ранжирование кластеров	LL, MI	Оценка устойчивости словосочетаний

Остановимся на приоритетных для данного исследования статистических мерах более подробно.

В корпусной лингвистике частотность является одной из базовых количественных характеристик, позволяющих оценить распространенность языковых единиц (слов, словосочетаний, грамматических конструкций) в текстах [Плунгян 2024; Gries 2024]. Действительно, в рамках когнитивно-лингвистических исследований установлена значимая корреляция между частотностью лексических единиц и степенью их закрепления (entrenchment) в когнитивных структурах [Hrǝica 2024].

Различают два основных типа частотности:

- абсолютную частотность (общее количество употреблений исследуемой языковой единицы в корпусе). Например, если слово *algorithm* (алгоритм) встречается в корпусе объемом 1 млн слов 1500 раз, его абсолютная частотность равна 1500;

- нормализованную частотность (показывает количество употреблений языковой единицы на определенный объем текста (обычно на 1 млн или 10 тыс. слов), что позволяет сравнивать данные из корпусов разного размера, а также выявлять статистически значимые различия в распределении терминов между подкорпусами []).

Нормализованная частотность рассчитывается по формуле:

$$\text{Нормализованная частотность} = \left(\frac{\text{Абсолютная частотность}}{\text{Объём корпуса в словах}} \right) \times N,$$

где N – базовый масштаб (например, 1 000 000).

Например, если лексема *method* встречается 50 раз в тексте объемом 250 тыс. слов, ее нормализованная частотность на 1 млн слов составит:

$$\left(\frac{50}{250\,000} \right) \times 1\,000\,000 = 200$$

Для интерпретации значений нормализованной частотности в англоязычных корпусах были предложены ориентировочные пороги, позволяющие отнести лексему к той или иной зоне частотности. В настоящем исследовании частотные показатели рассчитываются по данным Corpus of Contemporary American English (COCA), тогда как интерпретация частотности опирается на шкалу частотных диапазонов (frequency bands), принятую в Oxford English Dictionary [OED 2025]:

- высокочастотная лексика: более 1000 вхождений на миллион слов (pmw);
- среднечастотная: от 100 до 1000 pmw;

- низкочастотная: от 10 до 100 pmw;
- редкая лексика: менее 10 pmw.

Частотность терминов в специализированных корпусах может существенно расходиться с общеязыковыми распределениями; отсюда вытекает требование осторожной интерпретации. При этом использование нормализованных показателей остается базовым инструментом сопоставительного анализа ключевых единиц – в процедурах анализа ключевости (keyness analysis), идентификации коллокаций и оценке терминологической представленности [Anthony 2023].

Применение нормализованной частотности является необходимым этапом количественного анализа лексики при исследовании терминологических единиц в рамках одного корпуса. Несмотря на то что нормализация частотности изначально была разработана для сопоставления разных корпусов или подкорпусов [Gries 2024], она также играет важную роль в обеспечении интерпретируемости данных в пределах одного массива текстов. В случае анализа терминологии в англоязычном корпусе по тематике искусственного интеллекта нормализованные значения позволяют исключить влияние объема текстов на результаты и объективно оценить распространенность термина относительно общего объема данных.

Значение нормализации возрастает при выполнении коллокационного анализа, поскольку частотность лексических сочетаний зависит от плотности контекста, длины документов и распределения терминов в корпусе. Как подчеркивает С. Грис, нормализованная частотность (на 10 тыс./1 млн слов) обеспечивает сопоставимость данных корпусов разного объема и уместна даже для внутрикорпусных сопоставлений, однако сама по себе она недостаточна. В структурно неоднородных корпусах ее необходимо дополнять мерами распределенности (dispersion) и/или работать со скорректированными частотами, чтобы избежать искажений из-за «кластеризации» употреблений [Gries 2024]. Кроме того, нормализация частотных данных является стандартной практикой в корпусной лингвистике и входит в базовые процедуры количественной обработки лексем в инструменте AntConc, что подтверждает ее методологическую состоятельность.

Вместе с тем интерпретация показателей частотности требует обязательного учета лингвистического контекста, так как высокая частотность не всегда коррелирует с семантической или функциональной значимостью единицы. Следовательно,

частотный анализ должен дополняться качественными методами исследования для получения валидных лингвистических выводов.

В корпусном анализе терминологической лексики количественные показатели выступают мощным инструментом для выявления устойчивых словосочетаний и терминологически значимых коллокаций. В терминологической парадигме корпусной лингвистики важнейшим является понятие коллокационной базы – множества лексем, встречающихся в пределах заданного контекстного окна вокруг исследуемого, или «ядерного» слова [Теоретические и прикладные 2016]. Каждая единица этой базы подлежит многомерной статистической интерпретации: рассчитываются вероятностные параметры совместной встречаемости, степень синтагматической связанности и устойчивость сочетания.

Наиболее востребованы две ассоциативные меры – Log-Likelihood (LL) и Mutual Information (MI), реализованные в ряде лингвистических инструментов, включая AntConc. Они отражают разные аспекты лексической связности и дополняют друг друга при анализе терминологических структур. Log-Likelihood (LL) оценивает статистическую значимость отклонения наблюдаемой частоты коллокации от ожидаемой при случайном распределении, что позволяет выявлять устойчивые словосочетания в тематически однородных корпусах [Anthony 2023]. Мера особенно чувствительна к высокочастотным лексемам и эффективна для обнаружения терминологических конструкций в специализированных текстах. Например, коллокация *training data* характеризуется высокой частотностью в подкорпусе машинного обучения и значением $LL > 500$, что указывает на статистически значимую связь компонентов и устойчивый терминологический статус сочетания.

Однако при множественных сравнениях (например, при анализе всех возможных сочетаний слов в корпусе) возрастает риск ошибок первого рода – фиксации случайных статистически значимых связей. Для его минимизации применяется коррекция Бонферрони (Bonferroni correction), предполагающая снижение уровня статистической значимости (p-value) путем деления на число независимых проверок, что повышает надежность вывода [Bonferroni 1936]. В AntConc (версия 4.3.1) предусмотрено включение/отключение данной коррекции и выбор уровня p-value для отсека нерелевантных коллокаций: $p < 0.05$, $p < 0.01$, $p < 0.001$, $p < 0.0001$.

Использование коррекции Бонферрони в сочетании со строгим уровнем значимости ($p < 0.001$ или $p < 0.0001$) особенно актуально при работе с объемными корпусами, включающими тысячи потенциальных сочетаний: оно повышает статистическую достоверность и снижает риск ложноположительных выводов. Это принципиально для терминологических исследований, где требуется эмпирически подтвердить устойчивые семантические связи между единицами. Выбор указанных параметров соответствует рекомендациям по контролю множественных сравнений в корпусной лингвистике и статистике [Benjamini, Hochberg 1995] и поддерживает принципы надежности и воспроизводимости анализа.

Статистическая мера взаимной информации (Mutual Information, MI) необходима для оценки степени ассоциативной связи между словами; она отражает, насколько вероятно совместное появление слов по сравнению с ожидаемым при независимом распределении. MI особенно эффективна при выявлении редких, но устойчивых сочетаний, которым она, как правило, присваивает высокие значения. Например, коллокация *vanishing gradient* (исчезающий градиент), несмотря на относительно низкую нормализованную частотность, получает высокое значение $MI > 5$, что указывает на сильную синтагматическую связанность компонентов и специализированный характер сочетания. Встречаемость этого термина ограничена корпусами по глубокому обучению, что подчеркивает его терминологическую специфику. Интерпретация результатов на основе MI опирается на фундаментальные работы в области вычислительной лингвистики и статистического анализа текстов [Church, Hanks 1990; Evert 2009]. Значения MI выше 3 принято считать индикатором сильной лексической ассоциации, близкие к нулю – свидетельством независимости, а отрицательные значения ($MI < 0$) могут указывать на взаимное избегание слов в корпусе.

С учетом методологических ограничений каждой метрики в исследованиях терминологической сочетаемости рекомендуется комбинировать MI с другими статистическими показателями, в частности с логарифмическим отношением правдоподобия (LL). Такая комбинация позволяет учитывать как редкие, так и высокочастотные коллокации, повышая достоверность выявления семантически значимых связей. MI обеспечивает чувствительность к лексемам с низкой частотностью (*fuzzy logic* (нечеткая логика), *zero-shot learning* (обучение с нулевым

сэмплом), *adversarial attack* (сопязательная атака)), тогда как LL – к часто встречающимся единицам, типичным для специализированных подкорпусов (например, *machine learning* (машинное обучение), *training model* (обучающая модель), *optimization algorithm* (алгоритм оптимизации)). Комбинированное использование этих мер обеспечивает более комплексное и надежное описание коллокационной структуры терминов, что особенно важно для анализа профессионального языка и построения терминологических моделей [Evert 2009; Gries 2024].

Таким образом, статистические метрики корпусной лингвистики обеспечивают выявление терминологически релевантных единиц, моделирование их коллокационной структуры и оценку распределенности на воспроизводимой основе. Совместное применение частотности, MI и LL создает эффект триангуляции, обеспечивая комплексное описание терминов и базу для последующих когнитивных и структурно-онтологических интерпретаций.

2.2. Дискурсивное пространство научной статьи как среда функционирования и верификации терминосистем

2.2.1. Функциональный статус научной статьи как первичного источника данных в прикладной лингвистике

В настоящем исследовании корпус научных статей выступает не пассивным источником данных, а методологическим каркасом терминологического исследования. Выбор данного материала обусловлен прежде всего тем, что жанр научной статьи выступает институциональным пространством порождения, определения, аргументации и закрепления новой терминологии, а также уточнения значений уже существующих обозначений [Temmerman 2000; Hyland 2004].

Нормативно-стандартизационная традиция, закрепленная в современных стандартах терминологической работы, описывает связку «объект → понятие → определение → обозначение» и предъявляет требования к дефинициям и терминам, трактуя стандарты как механизм кодификации [ISO 1087:2019; ISO 704:2022]. Следовательно, для выявления терминологических инноваций и реконструкции понятийных отношений методологически оправданно обращаться к жанрам, где концентрируются эксплицитные определения и таксономические связи, прежде всего

к исследовательской статье; стандарты и словари в этой перспективе образуют вторичный кодифицирующий слой, а научно-популярные тексты – диффузионный слой, ориентированный на распространение и упрощение знания.

При этом техническая документация (спецификации, API-документация, руководства разработчика или white papers и др.) нередко фиксирует первичные описания функциональных возможностей продуктов и технологий [ISO/IEC/IEEE 26514:2011]. Однако коммуникативные цели и дискурсивные характеристики технической документации существенно отличаются от таковых у научной статьи. Основная функция документации носит инструктивный и описательный характер, она ориентирована на практикующих разработчиков и инженеров и ставит целью описание способов использования технологии, а не обоснование ее научной новизны или концептуальных оснований [ISO/IEC/IEEE 26514:2011]. В терминологическом плане в документации термины чаще употребляются как данность, реже подвергаясь детальному определению и концептуальному размежеванию, в то время как научная статья представляет собой жанр, где новые термины активно номинируются, концептуализируются и встраиваются в систему дисциплинарного знания через аргументацию и сопоставление.

Существенным является и различие в режимах институциональной валидации. Рецензируемая научная статья проходит внешнюю экспертную оценку и тем самым обеспечивает согласование терминологического употребления с нормами академического сообщества и дисциплинарными стандартами аргументации [COPE 2003]. Техническая документация, напротив, в большинстве случаев создается внутри организаций и ориентирована на задачи внедрения, поэтому, как правило, не предполагает сопоставимой процедуры независимой научной экспертизы и междисциплинарной верификации понятийного аппарата. В рамках настоящего исследования документация рассматривается как смежный, комплементарный материал, полезный для уточнения имплементационных значений, параметров и эксплуатационных ограничений, однако не как замена научной статье, поскольку целью является описание терминосистемы на этапе концептуального становления и первичной институционализации в научном дискурсе.

Дискурсивно-коммуникативный (М. Кабре) и социокогнитивный (Р. Теммерман) подходы осуществляют принципиальный сдвиг парадигмы: от статичных «идеальных»

понятийных систем – к анализу функционирования термина в реальных текстах. Термин трактуется как динамическая единица дискурса, семантика и форма которой детерминированы коммуникативными задачами и когнитивными моделями конкретной области знания. Такое переосмысление не только обосновывает методологию корпусного извлечения терминов из специализированных публикаций, но и императивно требует учета их вариативности и полисемии как неотъемлемых свойств термина в дискурсе [Cabré 1999; Temmerman 2000].

В корпусно-ориентированной терминографии уже разработаны стандартизированные процедуры автоматического извлечения терминов (АТЕ) из научных статей, включая методы оценки терминокандидатства (termhood) и цельности единицы (unithood), статистические метрики и анализ специализированной лексики [Kageura, Umino 1996]. При этом методологическим узлом выступает концепт «knowledge-rich contexts» (KRC), позволяющий извлекать определения и отношения между понятиями из текста [Meyer 2001]. Для исследовательских статей характерна регулярная распространенность KRC по секциям и наличие устойчивых лексико-синтаксических маркеров (например, *is defined as*; *X is a type of Y*; *such as*; *consists of*; *also known as* и др.), что делает возможными формализованный поиск и алгоритмическую детекцию.

Семантически ориентированные модели – Frame-Based Terminology (П. Фабер) и лексико-семантическая модель М.-К. Л’Ом – связывают понятийное моделирование с корпусной наблюдаемостью и используют KRC как эмпирические опоры для реконструкции фреймов и валентностных профилей термина [L’Homme 2020; Faber 2022]. Современные обзоры по АТЕ и работы по извлечению структурированных знаний с участием больших языковых моделей (LLM) дополнительно подтверждают статус статьи как ключевого эмпирического источника для масштабного терминологического анализа [Polak, Morgan 2024].

В отечественном терминоведении тезис о приоритете материала сферы функционирования имеет давнюю и четкую аргументацию. В. П. Даниленко одной из первых системно сформулировала проблему двух сфер бытования термина: сферы фиксации – лексикографические и нормативные жанры (специальные словари, терминологические стандарты, перечни рекомендуемых единиц), где термин как бы помещается в «идеальные условия» и очищается от неоднозначности и вариативности;

и сферы функционирования – научная речь и специальная литература в широком смысле, где термины реально рождаются, определяются и аргументируются [Даниленко 1977: 38–39].

Дальнейшее развитие эта мысль получила в трудах В. М. Лейчика, который на основе дихотомии «фиксация/функционирование» разделил обе области на несколько подтипов (разные виды словарей и банков данных – в фиксации; научные, деловые, публицистические тексты и т. д. – в функционировании), показал первичность сферы функционирования и сформулировал проблематику «термин и текст», в рамках которой были заложены основы терминоведческой теории текста. Разработка данной теории позволила выделить терминопорождающие, терминоиспользующие и терминофиксирующие тексты, что напрямую объясняет институциональный статус научной статьи как первичной среды порождения и уточнения терминов, тогда как словари и стандарты выполняют функцию вторичной нормализации [Лейчик 2022]. Современные обобщения Л. М. Алексеевой о «вековом пути» российского терминоведения и С. В. Гринева-Гриневича, анализирующего эволюцию дефиниций термина, также подтверждают, что институциональное признание термина первично достигается в корпусе научных публикаций, тогда как его фиксация в словарях и стандартах является закономерным, но производным результатом [Алексеева, Мишланова 2021; Гринева-Гриневич 2022].

Практическое подтверждение связки «функционирование → фиксация» выражается в наблюдаемом корпусно-прикладном сдвиге. Он проявляется прежде всего в автоматизации извлечения терминов из массивов статей [Бручес, Батура 2021; Большакова, Семак 2025]. Параллельно развиваются отраслевые кейсы и неографические исследования, демонстрирующие процессы формирования и вариативности терминологии в новых доменах (цифровое общество, цифровые медиа). Эти исследования расширяют понимание сферы функционирования термина за пределы традиционно академических жанров [Бузальская 2021]. Наконец, на стороне фиксации появляются современные проекты лексикографической и стандартизированной кодификации (онлайн-глоссарии лабораторий, дискуссии о принципах фиксации формирующейся терминологии) [Иконникова 2025].

Применительно к терминологии ИИ исследователями введен и описан новый объект – «научный дискурс с ключевым термином “искусственный интеллект”»

[Термины и понятия 2024]. Показано, что именно тексты исследовательских статей и родственных жанров служат первичной средой появления, аргументации и предварительной нормализации терминов. Выделяются типовые коммуникативные зоны (в т. ч. дефиниционные и аргументативные), что дает эвристики для корпусного поиска соответствующих контекстов; при этом формализованный набор меток специально не задается. Это прямо поддерживает выбор исследовательской статьи как первичной зоны профессионального употребления и, следовательно, как базы для терминоизвлечения.

Корпусно-ориентированную линию развивает А. С. Мусаева. На материале русскоязычного массива текстов, посвященных ИИ, исследователь подробно описывает механизмы терминообразования (в т. ч. заимствование, транстерминологизацию, словообразовательные модели), что дает готовые эвристики для автоматического выделения терминов-кандидатов и их морфологического нормирования [Мусаева 2022].

Методологически значима и работа А. А. Клементьевой, демонстрирующая режимы функционирования термина «искусственный интеллект» в научном и публицистическом дискурсах; тем самым уточняется граница между «первичной» сферой (научные статьи) и смежными типами текстов, в которые термины распространяются [Клементьева 2022]. Для мониторинга динамики в более широком русскоязычном дискурсивном поле показателен корпусный анализ Ю. И. Бедненко по материалам НКРЯ: опираясь на тематическую кластеризацию и жанровое профилирование упоминаний «искусственного интеллекта», исследование сопоставляет дисциплинарный (научный) слой с медийными и учебно-научными регистрами, что позволяет описать траектории распространения и семантических сдвигов термина [Бедненко 2025].

Таким образом, корпус научных статей обеспечивает репрезентативность данных на всех этапах – от отбора терминов до анализа их функционирования в дискурсе; корпусные процедуры гарантируют воспроизводимость результатов, а практические кейсы (в т. ч. по ИИ) подтверждают, что именно статьи служат зоной первичной нормализации и оптимальным материалом для автоматизированного извлечения терминов и описания их контекстных профилей.

2.2.2. Жанровая детерминация корпуса исследования: критерии и принципы формирования эмпирической базы на материале научных статей

В рамках институционального научного дискурса, ориентированного на производство и верификацию знания, исследовательская статья выступает базовым жанром коммуникации науки. Она встроена в систему социальных ролей (автор – рецензент – редактор – читатель-эксперт) и нормативных процедур (рецензирование, редакционная экспертиза, типовая структура изложения), благодаря чему именно в этом жанре реализуются ключевые механизмы терминогенеза: номинация, дефинирование, таксономизация, аргументация и легитимация новых терминов. С опорой на подходы В. И. Карасика и В. Е. Чернявской к пониманию дискурса научная статья рассматривается нами как жанровый узел межтекстового научного пространства, где взаимодействие множества тематически связанных текстов делает процессы появления и закрепления терминологии наблюдаемыми и воспроизводимыми [Карасик 2000; Чернявская 2001]. Научному дискурсу присущи фиксированные цели (создание и проверка знания), ролевая структура и высокая степень нормирования (например, формат IMRaD), что обуславливает повышенную терминологическую плотность и наличие устойчивых дефиниционно-таксономических контекстов [Карасик 2004]. В таких условиях именно публикации в жанре исследовательской статьи обеспечивают оптимальные возможности для выявления и анализа новых терминов в корпусном исследовании.

Согласно типологии, разработанной В. Е. Чернявской и опирающейся на доминирующую текстовую функцию, научные тексты можно разделить на: 1) академические (научно-теоретические); 2) научно-информационные; 3) научно-критические; 4) научно-популярные; 5) научно-учебные [Чернявская 2010: 38–39]. Особый статус отводится академическим текстам, функция которых заключается в фиксации и трансляции принципиально нового знания, полученного в результате индивидуального или коллективного творчества. Данные тексты обладают первичностью по отношению к вторичным научным жанрам (обзорам, рецензиям, учебным пособиям), представляющим интерпретацию или систематизацию уже существующих данных.

К числу академических текстов, ориентированных на трансляцию нового оригинального знания, относится научная статья. Структура научной статьи

воспроизводит логику исследовательского процесса и включает следующие этапы: постановку проблемы, формулировку гипотезы, аргументацию и выводы, что обеспечивает соответствие эпистемологическим стандартам научного познания [Чернявская 2016: 63].

Научная статья содержит формализованное представление результатов исследования, направленного на решение конкретной научной или прикладной задачи. Являясь законченным научным произведением, опубликованным в рецензируемых периодических (научные журналы) или неперiodических изданиях (сборники научных трудов, материалы конференций и т. д.), научная статья включает элементы нового знания, такие как новые экспериментальные данные, новые объяснения существующих явлений или новые подходы и методы в теоретических и экспериментальных исследованиях.

В зависимости от целевой установки выделяются следующие типологические категории научных статей: теоретическая статья (систематизированное изложение теоретико-методологических основ исследуемой проблемы, критический анализ существующих концепций), обзорная статья (комплексный анализ текущего состояния научного направления, выявление трендов, противоречий и перспектив развития), эмпирическая статья (детализированное описание этапа исследования или прикладных результатов, включая методологию, данные и их интерпретацию), исследовательская статья (публикация завершеного исследования с полным циклом – от постановки проблемы до верификации результатов) [Сулейманова 2020].

Жанровая формализация и стандартизованная структура изложения IMRaD (Introduction, Methods, Results, and Discussion) способствуют обнаружению и описанию терминов в устойчивых коммуникативных позициях. Так, раздел “Introduction” фиксирует постановку проблемы и терминологические рамки. Раздел “Methods” позволяет зафиксировать употребление технической лексики: алгоритмы, фреймворки – *PyTorch*, *TensorFlow*; датасеты; параметры обучения, что облегчает точную идентификацию терминов и их связей с методологиями. Благодаря наличию гиперссылок на исходные коды, размещенные на открытых платформах (GitHub, Hugging Face и др.), возможно уточнение значений терминов через их практическую реализацию – например, отслеживание того, как концепт *attention mechanism* (механизм внимания) реализуется в программных скриптах.

Разделы “Results” и “Discussion” показывают функциональные роли терминов, их синонимию, вариантность и распределение по тематическим подсферам, а также содержат социотехническую терминологию, связанную с этическими, правовыми и управленческими аспектами внедрения ИИ. Здесь широко употребляются такие термины, как *algorithmic bias* (алгоритмическое смещение), *AI governance* (управление ИИ), *fairness* (справедливость), *transparency* (транспарентность/объяснимость), отражающие современный дискурс на стыке технологий и гуманитарных наук. Кроме того, в описаниях практических кейсов часто встречаются практико-ориентированные единицы – например, *edge AI* (периферийный ИИ), *real-time inference* (вывод в реальном времени), *deployment pipeline* (пайплайн развертывания). Охват включает фундаментальные (архитектуры моделей) и прикладные исследования (медицинская диагностика и др.), что обеспечивает разнообразие терминов и междисциплинарные связи (например, *bio-inspired neural networks* (биоинспирированные нейронные сети), *clinical NLP* (клиническая обработка естественного языка). Дополнительную ценность создают метаданные публикаций, позволяющие стратифицировать материал по направлениям внутри ИИ и тем самым обеспечивать достаточный охват и сравнимость данных.

В рамках оценки эффективности научной деятельности значимыми являются не столько количественные показатели публикационной активности, сколько качество публикаций, соотносимое с нормами научной коммуникации и экспертизы [Сычев, Загребельный 2013]. В условиях настоящего исследования индикатором институциональной авторитетности источников выступают издания, входящие в «Белый список» и отнесенные к 1-му уровню, поскольку такая классификация предполагает повышенные требования к редакционной политике, рецензированию, научной добросовестности и устойчивости издательских практик. Публикация в журналах первой категории «Белого списка» рассматривается как маркер включенности исследования в верифицируемое профессиональным сообществом научное общение и как подтверждение того, что представленные результаты прошли процедуры внешней экспертной оценки и соответствуют стандартам академического качества.

С практической точки зрения научная статья удовлетворяет ключевым требованиям к источнику терминологического материала: 1) предоставляет

аутентичный и актуальный язык профессионального сообщества; 2) опирается на рецензирование и редакционную экспертизу; 3) встраивается в сеть цитируемости и интертекстуальных зависимостей, что делает прослеживаемыми заимствование, стандартизацию и конкуренцию вариантов; 4) обладает формализованной структурой, обеспечивающей воспроизводимое извлечение единиц и их контекстов для корпусного анализа.

Таким образом, научная статья из изданий, включенных в «Белый список» и отнесенных к первой категории, выступает не только жанровым маркером академического дискурса, но и одним из ключевых каналов терминологической актуализации и институционализации в области ИИ. Стандартизованная композиция (включая IMRaD) и институциональные механизмы экспертной проверки обеспечивают надежность и сопоставимость терминологического материала. Это обуславливает практическую установку анализировать терминологию ИИ не по отдельным текстам, а по корпусу рецензируемых публикаций, где ролевые ожидания и жанровая организация делают процессы номинации, уточнения значения и закрепления терминов наблюдаемыми и воспроизводимыми. В этой системе координат стандарты и словари относятся к сфере фиксации как вторичной кодификации, научно-популярные тексты – к зоне распространения и адаптации знания, тогда как научные статьи образуют первичную сферу функционирования, оптимальную для корпусного терминоанализа.

Формирование эмпирической базы исследования осуществлялось как воспроизводимый протокол отбора и включения публикаций, ориентированный на контролируемую репрезентативность выборки и минимизацию методологических смещений. На первом этапе были заданы рамки наблюдения: англоязычные рецензируемые исследовательские статьи по тематике искусственного интеллекта и смежных направлений, обеспечивающие сопоставимость жанра и регистра. Хронологические границы корпуса охватывают период 2020–2024 годов, что позволяет фиксировать актуальные тенденции развития области и соответствующие лексико-дискурсивные особенности современного научного письма.

На втором этапе был сформирован пул изданий по институциональному критерию: в качестве базового реестра использовался «Белый список» научных изданий; в корпус включались только журналы и серии, отнесенные к первой

категории (1-й группе) «Белого списка». Такое ограничение задает проверяемый критерий отбора источников и обеспечивает единообразие требований к редакционной политике и процедурам экспертной оценки.

Далее применялись тематические и жанровые фильтры. Тематическая релевантность устанавливалась на основе метаданных (ключевые слова, рубрики, предметные области) и подтверждалась полнотекстовыми индикаторами (описание моделей, методов и данных; наличие устойчивого терминологического ядра домена ИИ); публикации с периферийными упоминаниями ИИ не включались в выборку. Жанровая релевантность обеспечивалась ориентацией на исследовательские статьи (research article); материалы, не представляющие полного цикла исследования (краткие тезисы без полнотекста, редакционные сообщения, новости, нерцензируемые тексты), отсеивались на этапе отбора. Таким образом, выборка определялась не формальным присутствием словосочетания Artificial Intelligence в названии журнала, а сочетанием институционального критерия («Белый список», первая категория) и предметно-жанровых признаков публикации.

На следующем шаге фиксировались правила включения/исключения и применялась регламентированная отбраковка по заранее заданным основаниям: дублирование записей, нерелевантность тематике, несоответствие жанру, а также отсутствие полнотекста в пределах допустимых каналов доступа. Наличие открытого доступа не использовалось как самостоятельный критерий включения, чтобы предотвратить смещение выборки в сторону публикаций, доступных без подписки (open access). Полнотекстовые версии обеспечивались комбинированием каналов: открытый доступ издателя, институциональные ресурсы и подписки, авторские версии в репозиториях и препринт-сервисы, а также легальные копии, предоставленные авторами по запросу. При невозможности получить полнотекстовую версию публикация исключалась с фиксацией причины, что повышало прозрачность и проверяемость протокола.

Для обеспечения репрезентативности корпус дополнительно стратифицировался по направлениям внутри ИИ (машинное обучение, обучение с подкреплением, компьютерное зрение, обработка естественного языка, системы поддержки принятия решений, предиктивная аналитика, аппаратные решения, применение ИИ в здравоохранении, а также этика и правовое регулирование) и по типу

исследования (фундаментальные работы об архитектурах и методах; прикладные кейсы). Это снижало риск перекоса в пользу одной подсферы или одного исследовательского тренда и обеспечивало сопоставимость тематических срезов.

На завершающем этапе выполнялась техническая нормализация корпуса: унификация форматов, удаление сервисных повторов, проверка корректности метаданных, что необходимо для воспроизводимых процедур корпусного поиска и извлечения терминов. В результате корпус представляет собой институционально и жанрово однородную эмпирическую базу, сформированную на основе изданий 1-го уровня «Белого списка» и предметно-жанровых критериев, с документированной отбраковкой и прозрачной стратегией доступа к полнотекстам; это обеспечивает валидность корпусных наблюдений и воспроизводимость получаемых результатов.

Примененный подход позволил сформировать корпус, отражающий актуальный язык профессионального сообщества и опирающийся на контент, прошедший внешнюю экспертную оценку. Совокупность публикаций из разных источников снижает влияние индивидуально-авторских особенностей и обеспечивает достаточную полноту материала для прослеживания процессов терминотворчества и первичной нормализации терминологии в области ИИ.

2.3. Алгоритм и процедурные этапы реализации корпусного анализа в программной среде AntConc

В данном разделе подробно описана подготовка корпуса научных статей к анализу в AntConc и ключевые процедуры корпусного исследования: извлечение ключевых слов, построение конкордансов, изучение коллокаций и анализ частотных характеристик терминов. Особое внимание уделено применению возможностей AntConc для словообразовательного и семантического анализа терминологического материала, что позволяет выявить структурные и смысловые особенности терминов в контексте научного дискурса.

Этап 1. Сбор эмпирического материала. Эмпирическая база исследования сформирована в цифровом формате и включает 897 рецензируемых англоязычных исследовательских статей за 2020–2024 годы со структурой IMRaD. Источники отбирались вручную из научных журналов и серий, включенных в актуальную версию Единого государственного перечня научных изданий («Белый список»); в корпус

включались публикации изданий первой категории (1-й уровень) (*Artificial Intelligence Review, Journal of Artificial Intelligence Research, Artificial Intelligence in Medicine, IEEE Transactions on Neural Networks and Learning Systems, AI & Society*). Принадлежность изданий к «Белому списку», а также их библиографические и классификационные сведения (название, издатель, предметная рубрика и др.) уточнялись по данным ресурса JournalRank РЦНИ, поддерживающего обновляемый реестр и предоставляющего экспорт метаданных.

Этап 2. Подготовка корпуса к анализу. На данном этапе осуществляется предварительная обработка текстов, включающая очистку научных статей от нерелевантных элементов (метаданных, библиографий, сносок, формульных обозначений). Тексты сохраняются в формате .txt с кодировкой UTF-8 для корректного отображения символов. Каждый файл корпуса соответствует одной статье, что позволяет анализировать распределение терминологии на уровне отдельного документа. Несмотря на поддержку разных форматов в AntConc (версия 4.3.1), формат .txt является предпочтительным для лингвистических исследований, поскольку не содержит скрытой разметки, гиперссылок и иных структур, способных исказить токенизацию и извлечение коллокаций [Anthony 2023]. Прозрачность этого формата повышает воспроизводимость и вычислительную эффективность обработки корпуса.

Этап 3. Импортирование корпуса статей в AntConc и настройка параметров анализа. Процесс включает три основных шага:

1) Загрузка подготовленных текстов в программу AntConc посредством инструмента Corpus Manager для дальнейшего использования в качестве основного корпуса (AI Target Corpus):

Corpus Manager → Raw Files → AddFile(s)/Add Folder → Create.

Корпус сформирован с учетом критериев репрезентативности: достаточного объема (для стабилизации относительных частот), сплошного отбора релевантных контекстов, итеративной верификации данных и соблюдения «закона текстового блока» [Палийчук 2022].

2) Импортирование справочного корпуса (AI Reference Corpus) на основе терминов ИИ из «Глоссария по искусственному интеллекту» [Баканач 2023; 2024]. Корпус включает 2500 терминологических единиц, отобранных по критериям релевантности и охватывающих ключевые концепты и технологии ИИ (machine learning, neural network, reinforcement learning и др.). Визуализация структуры корпусов представлена на Рисунке 1.

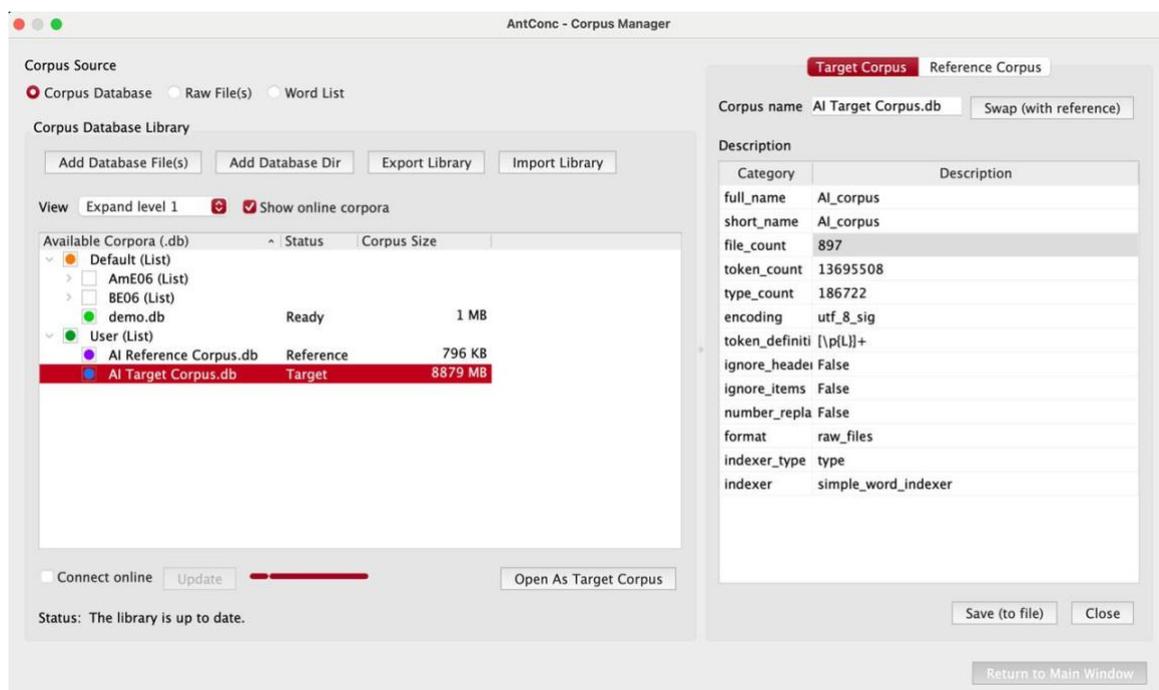


Рисунок 1 – Структура AI Target Corpus (13 695 508 токенов) и AI Reference Corpus (8500 токенов) с основными параметрами корпусной обработки

3) Настройка глобальных параметров анализа в AntConc. Версия 4.3.1, позволяющая адаптировать обработку англоязычных текстов под задачи исследования. Ключевые настройки, использованные для терминологического анализа корпуса по ИИ, приведены в Таблице 5.

Таблица 5 – Настройки AntConc для анализа терминосистемы ИИ

Категория	Параметр / Настройка	Назначение / Примечание
Версия и формат данных	AntConc 4.3.1, .txt, UTF-8	Простая структура, без скрытой разметки; корректная токенизация
Экспорт данных	.csv, .txt, .xml	Для последующей обработки и визуализации результатов
Нормализация частот	на 1 000 000 слов/токенов (PMW)	Обеспечивает сопоставимость частот терминов в текстах разного объема
Точность вычислений	3–4 цифры (для нормализованной частотности); 4–5 цифр (для LL, MI)	Гарантирует точность при статистическом анализе коллокаций

Категория	Параметр / Настройка	Назначение / Примечание
Регистр	учет регистра (вкл.); приведение к нижнему регистру (по необходимости)	Для различения аббревиатур (<i>AI, ML</i>) и унификации лексем (например, <i>Model / model</i>)
Токенизация	пробел, табуляция, знаки препинания	Разделение текста на лексемы для частотного анализа
Фильтрация лексики	стоп-листы (репозиторий GitHub); регулярные выражения (regex)	Очистка от служебной, структурной и «шумовой» лексики
Дополнительные стоп-слова	chapter, figure, page, section, work, represented и др.	Исключение типичных для научного дискурса единиц, но нерелевантных для терминосистемы ИИ

Важным условием получения валидных статистических показателей является исключение функциональной лексики, что реализуется подключением списка стоп-слов в AntConc (Global Settings → Tool Filters → Hide words in file). В исследовании используется набор stopwords-iso [Stopwords ISO 2023], дополненный узкоспециальными «шумовыми» маркерами научного дискурса (chapter, figure, table, page, section, journal и др.), что снижает искажения при расчете частот и коллокаций. В целом настройки AntConc 4.3.1 (нормализация частот, точная токенизация, фильтрация нерелевантной лексики и учет регистра) обеспечивают воспроизводимый анализ англоязычного корпуса по ИИ и позволяют выявлять значимые терминологические единицы и их распределение в научном дискурсе.

Этап 4. Частотный и ключевой анализ терминологической лексики. Корпус объемом 13 695 508 токенов подвергается двухуровневому анализу:

1) Частотный анализ выполняется модулем Word List, формирующим список словоформ с абсолютной и/или нормализованной частотностью; в итоговый перечень включаются только единицы с нормализованной частотностью ≥ 10 PMW (экспорт в .txt и .csv). Омнимичные формы и варианты написания уточняются посредством анализа конкордансов (KWIC), что позволяет правильно лемматизировать термины. Сведение к леммам выполняется через шаблоны с подстановочным символом * (например, *transformer** → *transformer/transformation*).

2) Анализ ключевых слов (Keyword List) основан на сопоставлении AI Target Corpus и AI Reference Corpus; отбираются лексемы с наивысшими показателями ключевости (keyness) – статистической метрикой, отражающей степень отклонения фактической частоты слова в основном корпусе от ожидаемой по справочному корпусу [Anthony 2023].

3) В результате частотного анализа и анализа ключевости (keyness analysis) выявлено 250 наиболее частотных (диапазон частот 100–3000) и тематически релевантных однословных лексических единиц (преимущественно существительные и ряд прилагательных), составляющих основу терминологического поля научного дискурса ИИ (см. Таблицу 6). Данная выборка послужила базой для последующего коллокационного анализа и поисковым перечнем для выявления вторичных терминологических реализаций в корпусе (производных, композитов, аббревиатур/акронимов).

Таблица 6 – Высокочастотные однословные лексемы научного дискурса ИИ

Лексема	Абсолютная частотность	Норм. частотность (на 1М слов)	Лексема	Абсолютная частотность	Норм. частотность (на 1М слов)
<i>learning</i>	42882	3131	<i>neural</i>	14621	1067
<i>data</i>	40392	2949	<i>system</i>	13821	1008
<i>model</i>	40288	2941	<i>network</i>	13404	978
<i>set</i>	22086	1760	<i>approach</i>	12764	931
<i>algorithm</i>	19455	1420	<i>problem</i>	12705	927
<i>method</i>	16821	1228	<i>language</i>	12576	918
<i>training</i>	16352	1193	<i>analysis</i>	12553	916
<i>performance</i>	15459	1128	<i>image</i>	11969	873
<i>number</i>	15156	1106	<i>function</i>	11755	858

Этап 5. Коллокационный анализ (модуль Collocate) для идентификации устойчивых терминологических словосочетаний. Список из 250 наиболее частотных однословных лексем рассматривается как предварительное терминологическое ядро, репрезентирующее ключевые концепты области и служащее основой для последующей смысловой детализации и иерархизации понятий через производные и уточняющие термины [Кондратюкова 2012]. С учетом того, что в научно-технической терминологии доминируют термины-словосочетания [Лотте 1961; Даниленко 1977; Гринев-Гриневиц 2008], приоритетной задачей становится выявление устойчивых коллокаций, уточняющих параметры, функции и объекты исследования.

Коллокационный анализ выполняется в AntConc (модуль Collocate) с фильтрацией и ранжированием результатов по частотности (Frequency), охвату (Range) и статистическим показателям связанности (LL, MI), подробно рассмотренным ранее; дополнительно учитывалась нормализованная частотность (на 1 млн токенов). В качестве базового уровня статистической значимости принят порог $p < 0.001$, что, согласно таблице критических значений распределения χ^2 при 1 степени свободы, соответствует пороговому значению метрики Log-Likelihood (LL) > 10.83 [Agresti 2002]. Установление данного порога позволяет отсеивать случайные совпадения и сохранять лишь статистически значимые связи между лексемами – это особенно важно при идентификации устойчивых терминологических коллокаций.

На первоначальном этапе статистического анализа приоритет отдается метрике LL, обладающей высокой чувствительностью к релевантным ассоциациям (в том числе низкочастотным) и позволяющей с достаточной степенью достоверности установить наличие неслучайной встречаемости лексем. Например, коллокат *network* демонстрирует высокий уровень связности с лексемами *neural* (LL = 33 932), *adversarial* (LL = 1411) и *convolutional* (LL = 978), что указывает на их терминологическую релевантность. Напротив, сочетания *network* с лексемами *connection* (LL = 9.3), *problem* (LL = 7.6) и *device* (LL = 4.8) не достигают установленного порогового значения, поэтому они исключаются из дальнейшего анализа как нерепрезентативные.

На втором этапе анализа применяется коэффициент взаимной информации MI, позволяющий выявлять устойчивые синтагматические связи и терминологические сочетания даже при низкой абсолютной частоте; значения $MI > 3$ обычно трактуются как индикатор устойчивой коллокации и потенциального специализированного термина. Одновременно MI помогает исключать частотные, но слабосвязанные сочетания, не обладающие терминологическим потенциалом. Например, при высокой встречаемости *AI utilization* низкое значение MI указывает на отсутствие прочной лексико-семантической связи, что позволяет не включать его в перечень релевантных единиц.

Таким образом, совместное использование метрик LL и MI обеспечивает статистически достоверный и содержательно релевантный отбор терминов из корпуса научных текстов по ИИ.

Результаты коллокационного анализа для лексемы *network* демонстрируют спектр устойчивых терминологических сочетаний, репрезентирующих ключевые направления ИИ (например, *neural network* (нейронная сеть), *adversarial network* (сопоставительная сеть), *convolutional network* (сверточная нейронная сеть)); наиболее релевантные коллокаты, ранжированные по LL и MI, представлены в Таблице 7.

Таблица 7 – Пример анализа коллокаций с лексемой *network*

Collocate	Frequency	MI	LL	Интерпретация
<i>neural</i>	4111	7.1	33932	термин (высокие MI и LL)
<i>model</i>	209	1.4	147	часто встречается, но не термин (низкий MI)
<i>diffusive</i>	8	8,6	84	узкий термин (высокий MI, средний LL)

Анализ данных показывает, что *neural network* (нейронная сеть) выступает ключевым термином, тогда как *neural model* (нейронная модель), несмотря на высокую частотность, не демонстрирует достаточной терминологической устойчивости и потому рассматривается как преимущественно контекстная коллокация. Напротив, *diffusive network* (диффузионная сеть) характеризуется высокой концептуальной значимостью при ограниченной распространенности, что позволяет отнести его к узкоспециализированным терминам.

Для обеспечения точного и воспроизводимого извлечения коллокаций ядерных терминов в корпусе научных статей по ИИ были заданы специальные параметры модуля Collocate (AntConc 4.3.1), представленные в Таблице 8.

Таблица 8 – Технические параметры модуля Collocate (AntConc 4.3.1)) при работе с AI Target Corpus

Параметр	Опции	Рекомендации для ИИ-терминов
Search Query (поисковый запрос)	Word (слово) Case (регистр) Regex (шаблон)	Word – для общего анализа; Case – для учета регистра при исследовании аббревиатур; Regex – для поиска сложных шаблонов и анализа коллокаций с частицами/предлогами
Window Span (размер контекстного окна)	диапазон слов вокруг искомого термина (L – слева, R – справа)	Максимально 5 слов слева (L) и справа (R) от искомого слова для охвата синтаксически и семантически значимых связей в пределах предложения
Min Frequency (порог частотности)	минимальное количество совместных вхождений коллоката с термином (от 1 до ∞)	5 – для статистической надежности
Min Range (минимальное расстояние)	минимальное расстояние между термином и коллокатом (от 1 до ∞)	2 – для исключения предлогов/местоимений; 1 – для анализа ближайших слов (например, прилагательных)
Sort by (сортировка)	сортировка по показателям Type, Type End, Frequency (L/R; L; R), Range, Likelihood (LL), Effect (MI)	отображение коллокатов относительно ключевого слова с указанием их позиций (левый/правый контекст); частотности и соответствующих статистических значений LL и MI (в порядке убывания значений)

Оценка устойчивости лексико-семантических связей в выборке высокочастотных единиц позволила выделить структурно значимые сочетания и сформировать перечень из ≈2300 лексем и устойчивых словосочетаний, включая ядерные термины с наибольшей сочетаемостью (например, *learning* (обучение), *method* (метод), *algorithm* (алгоритм) и др.). Отобранные единицы прошли верификацию в соответствии с международными стандартами терминологической деятельности (ISO/IEC 22989:2022; ISO/IEC TR 24028:2020; ISO 18374:2025 и др.) [International Organization 2025] и с опорой на авторитетные специализированные глоссарии и словари по искусственному интеллекту [Толковый словарь 1992; Raynor 1999; Искусственный интеллект 2023; Баканач 2023; 2024; Machine Learning Glossary 2025]. В результате сформирован AI Verified Corpus из 1648 терминологических единиц для последующего словообразовательного и семантического анализа, сохраненный в формате .xml для обеспечения

воспроизводимости, прозрачности отбора и интеграции с цифровыми инструментами обработки.

Этап 6. Словообразовательный и семантический анализ терминологических единиц. На данном этапе исследовательский фокус смещается на расширение эмпирической (статистической) базы за счет комплексного словообразовательного и семантического анализа отобранных терминологических единиц. Несмотря на преобладание ручной обработки, основанной на лингвистической интерпретации и классификации, инструментарий AntConc позволяет частично автоматизировать ключевые операции: поиск по заданным шаблонам (регулярным выражениям), извлечение производных форм от ядерных терминов и первичную оценку семантической сочетаемости.

Словообразовательный анализ охватывает морфемный разбор, описание продуктивных способов и моделей терминообразования, а также учет заимствованных элементов на основе корпусных списков и этимологических справочников (Online Etymology Dictionary, Oxford English Dictionary и др.). В AntConc используется модуль Cluster, формирующий n-граммы (2–4 слова) вокруг ядерного элемента (например, *learning* (обучение) → *machine / deep / continual learning* (машинное / глубокое / непрерывное обучение)), что позволяет проследить расширение базового понятия в более сложные терминологические конструкции и оценить продуктивность словосложения в различных контекстах. Дополнительно задействуется инструмент Word List в сочетании с Regex (регулярными выражениями) для извлечения деривационных моделей, композитов (в т.ч. дефисных) и акронимов; применяемые шаблоны приведены в Таблице 9.

Таблица 9 – Регулярные выражения для выявления словообразовательных моделей в модуле Word List AntConc

Задача анализа	Описание функции	Пример Regex-шаблона	Пример извлечения
Дериваты от корня/основы	Любые формы с вхождением корня/основы	\b\w*<ROOT>\w*\b	<i>training</i> <i>retraining</i>
Суффиксальные формы	Слова с заданным суффиксом	\b\w+ization\b \b\w+er\b \b\w+tion\b	<i>anonymization</i> <i>encoder</i> <i>classification</i>
Префиксальные формы	Слова с заданной приставкой	\bmeta\w+ \bhyper\w+	<i>metadata</i> <i>hyperparameter</i>
Дефисные композиты (KWIC)	Термины с дефисным написанием	\b\w+-\w+\b	<i>self-driving</i> <i>neuro-fuzzy</i>
Многокомпонентные цепочки (≤3 дефиса) (KWIC)	Композиты из 2–4 компонентов	\b(?:\w+-){1,3}\w+\b	<i>state-of-the-art</i> <i>end-to-end</i>
Акронимы	Выбор последовательностей из двух и более заглавных букв	\b[A-Z]{2,}\b	<i>AI, NLP, GPU</i>
CamelCase-композиции	Слова, где две и более «части» начинаются с заглавной буквы	\b(?:[A-Z][a-z+]){2,}\b	<i>FuseMed</i>

* CamelCase-композиции (от англ. “camel case” – «верблюжий регистр») – это сложные слова, в которых несколько лексических компонентов объединены в одно «слово» без пробелов или дефисов, причем каждый компонент (кроме, возможно, первого) начинается с заглавной буквы. Такой способ записи часто используется в программировании и в технических терминах для компактного обозначения многословных понятий.

Параметр Sort by Range позволяет оценивать устойчивость и распространенность моделей по числу документов (например, регулярность суффиксов -ization, -er и структурные шаблоны A+N, N+N, N+N+N), отсеивать локализованный «шум» и проследивать экспансию ядерных терминов (*boosting* → *gradient boosting* → *light gradient boosting*).

Семантический анализ ориентирован на уточнение значений, выявление эпонимических, метафорических и метонимических номинаций, установление синонимии, тематических иерархий и контекстной полисемии. Методологически он базируется на корпусном контекстном анализе (KWIC) и экспертной интерпретации, где контекст понимается как совокупность лексико-синтаксических, прагматических и тематических параметров [Firth 1957], а в когнитивной перспективе – как источник активации фреймов и сценариев [Fillmore 1976; Temmerman 2000].

Интерпретационной основой выступает категориально-понятийная классификация. Каждая терминединица (а при полисемии – каждый смысл) соотносится с типом обозначаемой сущности – метод/алгоритм (*beam search* (поиск по лучу), *backpropagation* (обратное распространение ошибки)), модель/архитектура (*transformer* (трансформер), *variational autoencoder* (вариационный автокодировщик)), задача/проблема (*text classification* (классификация текстов); *object detection* (обнаружение объектов), компонент/модуль (*encoder* (кодировщик); *attention layer* (слой внимания)), метрика/функция потерь (*F1 score* (оценка F1); *BLEU* (метрика BLEU); *IoU* (пересечение по объединению)), параметр/гиперпараметр (*learning rate* (скорость обучения); *dropout* (дропаут)), ресурс/датасет/бенчмарк (*ImageNet*; *GLUE*), инструмент/библиотека (*PyTorch*; *scikit-learn*) и др. Предложенная классификация не подменяет собой контекстуальный анализ, а служит его концептуальным каркасом, что позволяет строго разводить полисемичные употребления (например, различать *mechanism* как понятие метода и *layer* как понятие архитектуры), проверять синонимию (только в пределах одной концептуальной категории) и выстраивать тематически обусловленные иерархии.

Эмпирическое исследование проводится посредством комплексного анализа корпуса: изучаются конкордансы (KWIC), коллокации (*Collocate*) и частотные списки непрерывных словосочетаний (N-Gram) по всему корпусу. Такой подход позволяет выявлять регулярные паттерны употребления, дефиниционные конструкции и семантические связи между терминами, опираясь на количественные показатели.

Контекстный анализ терминологии ИИ проводится в три взаимосвязанных этапа, первый из которых посвящен эпонимизации, метафоризации и метонимии. *AntConc* используется как фильтр кандидатов; модуль *Word List* с *Regex* выделяет единицы с именами собственными и формантами, характерными для названий теорем и тестов (например, `\b[A-Z][a-z]+(test|theorem|model)\b`: *Turing test*, *Bayes theorem*, *Markov model*), после чего кандидаты проверяются по KWIC и N-Gram для анализа контекстов и исключения нерелевантных вхождений. Метафорические термины выявляются через *Collocate* и N-Gram по устойчивым семантически маркированным сочетаниям (например, *bag of words* (мешок слов)), при этом устойчивость и терминологизация подтверждаются частотностью и распределенностью по корпусу с помощью *Sort by Range*. Метонимия диагностируется по смене синтаксико-

семантических рамок одного термина (X) и уточняется шаблонным поиском типовых конструкций (*optimize X*, *X score*, *achieves N on X* и др.), а также проверкой устойчивости соупотреблений через KWIC, Collocate и N-Gram.

На втором шаге разграничиваются полисемичные употребления и формируются синонимические ряды. Термин квалифицируется как полисемичный при наличии ≥ 2 устойчивых коллокационно-синтаксических кластеров с минимальным пересечением топ-коллокаций и различающимися онтологическими или глагольно-валентными рамками; статистическая поддержка устанавливается в окне L5–R5 ($MI \geq 3.0$; $LL \geq 10.83$; $p < 0.001$). Например, *mechanism* может реализовываться как «понятие метода» (в рамках глагольного управления типа *implement/optimize mechanism*) и как «элемент архитектуры» (в сочетаниях типа *attention mechanism*), что фиксируется различием коллокационного окружения. Синонимия выявляется по явным маркерам и распределительному сходству: аппозитивным/дефиниционным конструкциям («X, also known as Y», «X (Y)», «X is called Y»), совпадению терминологических «голов» (model, algorithm, system, network, function, rate) и близости управляющих глаголов (train, evaluate, implement, optimize и др.). Например, связка *dropout* (дропаут) / *dropout rate* (уровень дропаута) в корпусе функционирует как аппозитивная подстановка (X (Y)). Пара признается синонимической при наличии ≥ 2 независимых маркеров, ≥ 5 KWIC-примеров на каждый термин и ≥ 5 общих ключевых коллокаций; конструкции типа «Unlike X, Y...» исключаются.

На завершающем этапе терминология распределяется по тематическим подкорпусам (Machine Learning, NLP, Computer Vision и др.), где с помощью Keyword List и Collocates выделяются репрезентативные термины и устойчивые сочетания, а их функции уточняются по KWIC. Так, в подкорпусе «Робототехника» термин *robot* чаще всего сочетается со словами *navigation* и *motion*, тогда как в подкорпусе «Машинное обучение» *robot* не играет ключевой роли, уступая место терминам *training*, *data*, *algorithm*.

Таким образом, корпусный семантический анализ терминологии ИИ обеспечивает многоаспектное описание значений терминов, выявление метафорических и метонимических переносов, установление синонимических связей и создание тематически структурированной картины терминологической системы. Полученные данные легли в основу разработки глоссария с уточненными

дефинициями, предпочтительными терминами и структурированными синонимическими рядами.

Этап 7. Извлечение определений терминов и примеров их употребления (инструмент KWIC). На заключительном этапе из корпуса извлекаются дефиниции терминов и репрезентативные контексты их употребления для статей онлайн-гlossария *aiterms.ru*. Для этого используется инструмент KWIC, отображающий ключевое слово в право- и левосторонних контекстах и позволяющий отбирать примеры употребления, фиксировать функции термина (определение, описание, применение), выявлять дефиниционные маркеры (*is defined/known as, refers to, can be described as* и др.), а также отмечать вариативность дефиниций и реализаций.

Пример дефиниционного контекста: *landmark detection* (обнаружение ключевых точек) – «*One of the tasks of computer vision is **landmark detection**. It can be described as an object recognition task with localization...*» (перевод: «Одной из задач компьютерного зрения является обнаружение ключевых точек. Ее можно описать как задачу распознавания объектов с локализацией: алгоритму необходимо не только обнаружить объект на изображении, но и оценить положение этого объекта») (Anna Kuznetsova 2021), где термин вводится и поясняется через метаязыковую конструкцию *can be described as*, уточняя его место в понятийной иерархии.

Пример недефиниционного употребления: «*Recent advances in **artificial intelligence** have significantly improved...*» (перевод: «Последние достижения в области искусственного интеллекта значительно повысили эффективность диагностических инструментов в медицинской визуализации») (Emily Zhang 2022), где *artificial intelligence* (искусственный интеллект) выступает тематическим компонентом высказывания и иллюстрирует естественное функционирование термина в научном дискурсе.

Таким образом, использование KWIC на данном этапе позволяет обеспечить достоверность, репрезентативность и практическую значимость лексикографического материала, адаптированного к задачам научного терминологического описания в цифровой среде.

Разработанный методологический аппарат, основанный на применении корпусно-ориентированного подхода с использованием инструментария AntConc, позволяет провести многоаспектное исследование терминосистемы искусственного

интеллекта в научном дискурсе. Такой подход согласуется с современными требованиями терминоведческого исследования: опираться на репрезентативные данные корпусов и применять количественные метрики, не пренебрегая качественным лингвистическим анализом.

2.4. Формирование онлайн-гlossария терминов в области ИИ как отражение структуры тематического поля

В современной лексикографии под гlossарием понимается «толковый словарь устарелых, малоупотребительных, специальных и т. п. слов, составляемый обычно по какому-нибудь тексту, преимущественно старинному» [Крысин 2023: 87]. На практике гlossарием сегодня именуется любое «собрание гloss, непонятных слов или выражений с толкованием (толковый гlossарий) или переводом на другой язык (переводной гlossарий)» [Зарицкая 2013: 173]. Как правило, тематическое наполнение гlossария связано с узкоспециализированной терминологией в определенной отрасли.

Разработка специализированных гlossариев по искусственному интеллекту стала важным аспектом работы в данной области ввиду стремительного появления новых терминов, понятий и технологий, требующих систематизации и упрощения [Толковый словарь 1992; Raynor 1999; Англо-русский толковый 2022; Баканач 2023; 2024]. В последние годы наибольшую актуальность приобрело создание онлайн-гlossариев по ИИ (например, онлайн-гlossарии компаний Google, IBM, Arpen, Deep AI), поскольку они обладают рядом неоспоримых преимуществ, присущих большинству онлайн-словарей, таких как доступность, сравнительно быстрая обновляемость, легкий поиск нужных слов, интерактивность, включающая примеры использования терминов в контексте и возможность предоставления комментариев от пользователей, мультимедийные элементы (аудио-, видеоматериалы и изображения), многоязычность, кросс-платформенность и др. [Fuertes-Olivera 2025].

В русле современных представлений об электронно-цифровом словаре как специфическом лексикографическом объекте, отличающемся динамичностью и интерфейсной обусловленностью описания, проектирование онлайн-гlossария по искусственному интеллекту предполагает акцент на дифференцирующих свойствах цифровых ресурсов [Полубиченко, Алхастова 2024]: 1) гибкость – инструментарий интерфейсных и поисковых средств, обеспечивающих простую формулировку

запросов, удобство взаимодействия и оперативное получение релевантной информации; 2) динамичность – возможность оперативного редактирования, обновления и расширения записей как составителем, так и при модерации пользовательских предложений; 3) масштабируемость – практически неограниченный объем лексических единиц и сопроводительных материалов (комментарии, ссылки, примечания, перекрестные ссылки и др.); 4) многоязычность – поддержка параллельных терминологических представлений и межъязыковой сопоставимости.

Разработка онлайн-гlossария терминологии искусственного интеллекта базируется на интеграции классических лексикографических принципов с современными цифровыми решениями. Предлагаемая методология, структурированная в виде последовательных этапов, обеспечивает системность терминологического анализа и реализацию функционального потенциала электронных ресурсов.

1. Отбор, анализ и систематизация специальной лексики. На первом этапе формируется исходный массив терминов для лексикографического описания. Отбор терминологических единиц на английском языке осуществляется методом корпусного исследования с использованием платформы AntConc. Полученный лексический материал систематизируется в реляционной базе данных (таблица Excel), где каждому термину присваивается тематическая категория (Machine Learning, NLP, Computer Vision и др.). На завершающем шаге этапа проводится комплексный лексикографический анализ, предусматривающий дефинирование, выявление синонимических и полисемических отношений, а также описание особенностей употребления.

2. Проектирование гlossария терминологии ИИ на основе лексикографических принципов. Составление онлайн-гlossария по ИИ осуществлялось в соответствии с системой лексикографических критериев, ключевыми из которых являются авторская установка, макрокомпозиция и микрокомпозиция [Гринев-Гриневиц 2008; Dubichynskyi, Dąbrowska 2021]. Кратко охарактеризуем проект в рамках данной системы.

Гlossарий имеет следующие элементы авторской установки: охватывает пласт специальной лексики, тематически связанный со сферой функционирования искусственного интеллекта; включает максимальное число специальных лексем

терминосистемы ИИ (> 3000 ЛЕ на английском и русском языках), исключает жаргонизмы, профессионализмы; ориентирован на широкий круг читателей; носит информационно-справочный характер.

На уровне макрокомпозиции в глоссарии используется принцип сплошной организации словарных статей: термины располагаются в алфавитном порядке, также могут быть упорядочены по тематическим разделам (см. пункт «Техническая реализация»), есть возможность осуществления поиска по термину. Глоссарий содержит вступительное слово и основной информационный материал (основной корпус). Многозначные слова представляются через нумерованные значения в одной статье (например, *bias* (предвзятость) – 1) смещение в данных; 2) систематическая ошибка алгоритма; 3) этическая предвзятость ИИ-системы).

Микрокомпозиционная организация словарной статьи реализует следующие принципы:

- 1) термин и его аббревиатура (при наличии);
- 2) грамматическая информация (часть речи);
- 3) тематическая атрибуция (machine learning, NLP, computer vision);
- 4) определение (краткое и точное объяснение термина);
- 5) примеры использования термина в контексте;
- 6) синонимы и/или связанные термины (если применимо);
- 7) аудиопроизношение термина на английском языке.

Каждая словарная статья дополнена ссылками на источники цитирования (гиперссылки на статьи, стандарты, словари) или на платформы генеративных моделей, использующихся при создании изображений и аудиопроизношения (см. Рисунок 2).

<p>text-to-speech noun</p> <p>abb. TTS</p> <p>1. <i>the method of converting ASCII text to a speech waveform</i></p> <p style="text-align: right;">"text-to-speech ." A Dictionary of Computing. Encyclopedia.com. 14 Sep. 2024 <https://www.encyclopedia.com>.</p> <p>2. <i>TTS specifically refers to a type of speech synthesis that converts written text into spoken words. TTS systems analyze the text and produce an audio output that mimics natural speech. TTS is one of the most common applications of speech synthesis.</i></p> <p><u>Examples of <i>text-to-speech</i> in a sentence:</u></p> <ul style="list-style-type: none"> In this work, we propose Prosody-TTS, improving prosody with masked autoencoder and conditional diffusion model for expressive <i>text-to-speech</i>. (Huang et al. 2023) TTS is another assistive technology that can improve communication accessibility for deaf and mute individuals. (Zaineldin et al. 2024) 	<p>NLP</p>
---	------------

Рисунок 2 – Пример словарной статьи: *text-to-speech* (синтез речи)

3. Техническая реализация онлайн-гlossария. Веб-приложение для просмотра и поиска терминов разработано на базе открытой CMS WordPress [WordPress 2025], что обеспечило быстрое развертывание интерфейса навигации и поиска при минимальных инфраструктурных затратах (хостинг и домен – <http://aiterms.ru>), а также за счет развитой экосистемы шаблонов и плагинов, доступной документации и поддержки сообщества при невысоких требованиях к компетенциям разработчика. Терминологическая база и система внутренней навигации (перекрестные ссылки) реализованы с использованием модулей управления гlossарием; ключевым решением стал плагин WP Glossary, обеспечивающий создание карточек терминов, алфавитный и тематический указатели и автоматизированную навигацию по взаимосвязанным единицам. Дополнительно применено тегирование, позволяющее группировать термины по темам и поддерживать многомерную классификацию без перестройки структуры гlossария; используются теги для фильтрации по областям ИИ: AI Fundamentals, Machine Learning (ML), Deep Learning, Natural Language Processing (NLP), Computer Vision (CV), Robotics, Neural Networks, Expert Systems, AI Ethics. Например, термин *text-to-speech* маркируется тегом NLP (см. Рисунок 2).

4. Наполнение гlossария. Данный этап представляет собой систематическую интеграцию терминологических единиц, их дефиниций и контекстных примеров,

направленную на формирование структурированного и репрезентативного ресурса, и включает ввод данных (ручное добавление терминов или загрузка данных из таблиц / внешних баз данных) и загрузку мультимедийного материала (плагин Audio Attachment).

5. Тестирование и доработка глоссария. Работы реализуются на двух уровнях: функциональный контроль и пользовательская оценка. Функциональный контроль включает проверку поисковых механизмов, внутренней/внешней ссылочной структуры, отображения мультимедийных компонентов и эргономики интерфейса (удобство, согласованность, доступность). Пользовательская оценка предполагает сбор обратной связи целевой аудитории (опросы, комментарии) и анализ предложений для совершенствования структуры и содержания.

6. Поддержка и обновление лексикографического ресурса. Непрерывный этап жизненного цикла, обеспечивающий актуальность, достоверность и соответствие современным языковым реалиям, включает регламентированное пополнение словника, ревизию и уточнение дефиниций с учетом изменений употребления, корректировку объема значения, добавление контекстов и примеров, контроль качества (вычитка, устранение орфографических, грамматических и терминологических ошибок), документирование и датирование правок (журнал версий), а также внедрение структурных и интерфейсных улучшений.

Таким образом, методика разработки онлайн-глоссария по ИИ включает создание репрезентативного корпуса текстов, анализ терминологических единиц (статистический, семантический, словообразовательный), разработку структурированного глоссария, техническую реализацию онлайн-платформы, наполнение глоссария, а также тестирование, доработку, поддержку и обновление.

ВЫВОДЫ ПО ГЛАВЕ 2

В Главе 2 сформирован методологический каркас корпусного исследования терминологии искусственного интеллекта, ориентированный на эмпирическую проверяемость, контролируруемую репрезентативность и воспроизводимость процедур на всех этапах работы с данными.

Во-первых, уточнены операциональные определения ключевых понятий «языковой корпус» и «корпусный менеджер», а также задана исследовательская логика, в рамках которой корпус трактуется как специально сконструированная и документированная коллекция текстов, обеспечивающая систематический доступ к контекстам функционирования единиц и к количественным параметрам их употребления в естественном узусе. Тем самым корпус выступает не вспомогательным приложением, а методической основой проектирования дальнейших процедур терминоизвлечения, оценки устойчивости и лингвистической верификации. Корпусный менеджер, в свою очередь, трактуется как специализированная программная среда для управления текстовым массивом и проведения многоступенчатого лингвистического анализа. Разграничение этих понятий закрепляет единообразие терминологического аппарата исследования и фиксирует исходную установку, согласно которой корпусно-ориентированный подход выступает не отдельным техническим приемом, а базисом построения и проверки описания терминосистемы на материале профильных публикаций.

Во-вторых, обоснован выбор AntConc как основной инструментальной среды и показано, что его функциональные модули обеспечивают полный цикл процедур корпусного терминоведческого анализа. Word List и Keyword List используются для построения частотного профиля и выявления терминологического ядра; KWIC – для контекстной проверки терминологичности и отбора доказательных примеров употребления; Collocates, Cluster и N-Gram – для фиксации устойчивых многокомпонентных единиц и типовых структурных паттернов. Такая конфигурация процедур задает единый воспроизводимый протокол – от первичного извлечения кандидатов до установления их коллокационной связанности и подготовки количественных данных для дальнейшей обработки. В результате AntConc используется не как описательный ресурс, а как средство операционализации

критериев termhood и unithood («терминность» и «устойчивость единицы»), что усиливает методологическую прозрачность и проверяемость исследования.

В-третьих, описана и стандартизирована процедура построения корпуса и контроля качества данных. Сформирован цифровой массив из 897 англоязычных рецензируемых исследовательских статей за 2020–2024 годы, отобранных по институциональному критерию (издания 1-го уровня «Белого списка») и предметно-жанровым фильтрам. Общий объем корпуса превысил 13,6 млн словоупотреблений (токенов), что гарантирует достаточную статистическую надежность при вычислении частот и мер связанности. Процедура включает документированные правила включения и исключения, контроль дубликатов, а также получение полнотекстовых версий через легальные каналы доступа с фиксацией причин исключения при невозможности доступа. Дополнительная стратификация по направлениям ИИ и типам исследований снижает риск тематического перекоса и повышает сопоставимость внутрикорпусных срезов.

В-четвертых, представлен воспроизводимый цикл корпусного анализа в AntConc, который включает сбор и предобработку файлов, очистку и унификацию форматов с кодировкой UTF-8, настройку фильтров, проведение частотного, ключевого и коллокационного анализа, последующую лингвистическую верификацию и переход к словообразовательному и семантическому описанию терминов. Особое внимание уделено методам фильтрации «шумовых» элементов – от применения стоп-листов до коррекции Бонферрони, что гарантирует высокую точность и валидность получаемых лингвистических данных. Отбор единиц реализован как многоступенчатая модель, где количественные показатели выполняют функцию первичного фильтра, а окончательное решение принимается на основе статистических метрик, контекстной интерпретации и сопоставления с нормативными и лексикографическими источниками.

Наконец, показана прикладная результативность предложенной методики. Результаты корпусных процедур интегрируются в авторский онлайн-гlossарий, где корпусные данные обеспечивают доказательность дефиниций, репрезентативность иллюстраций и тематическую систематизацию материала, а также поддерживают обновляемость ресурса по мере развития терминосистемы ИИ.

ГЛАВА 3. СТРУКТУРНО-СЕМАНТИЧЕСКИЕ ОСОБЕННОСТИ ТЕРМИНОСИСТЕМЫ ПРЕДМЕТНОЙ ОБЛАСТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

В данной главе представлено корпусно-ориентированное описание структурно-семантических особенностей терминологии ИИ. В центре анализа находятся структурные модели терминообразования: аффиксация, словосложение (в т. ч. с неоклассическими формами), конверсия, эллипсис и редуктивные стратегии (сложносокращенные слова, телескопии, аббревиация); семантические механизмы пополнения: заимствование, привлечение, транстерминологизация; механизмы концептуализации и адаптации: метафоризация, метонимизация, эпонимизация; полисемия и синонимия как параметры вариативности значения. Описание опирается на формально ориентированную классификацию и результаты корпусного подсчета.

Отправной точкой структурно-семантического анализа выступает формально ориентированная классификация, предусматривающая разделение терминологических единиц на термины-слова и термины-словосочетания [Даниленко 1977; Гринев-Гриневич 2008; Лейчик 2022]. В исследуемом корпусе зафиксировано следующее распределение в рамках указанной дихотомии: термины-слова – 450 ТЕ (27%; например, *agent* (агент), *classifier* (классификатор)); термины-словосочетания – 1198 ТЕ (73%; например, *reinforcement learning* (обучение с подкреплением), *stochastic semantic analysis* (стохастический семантический анализ). Процентное соотношение указанных типов в терминосистеме ИИ представлено на Рисунке 3.

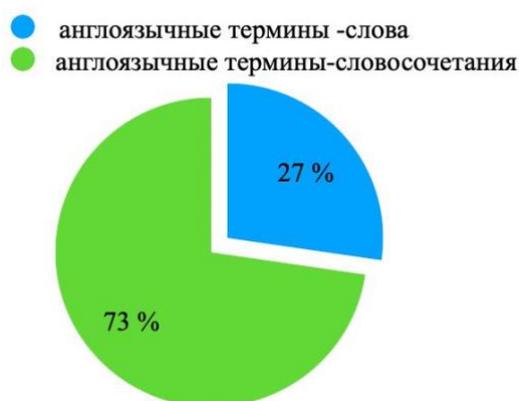


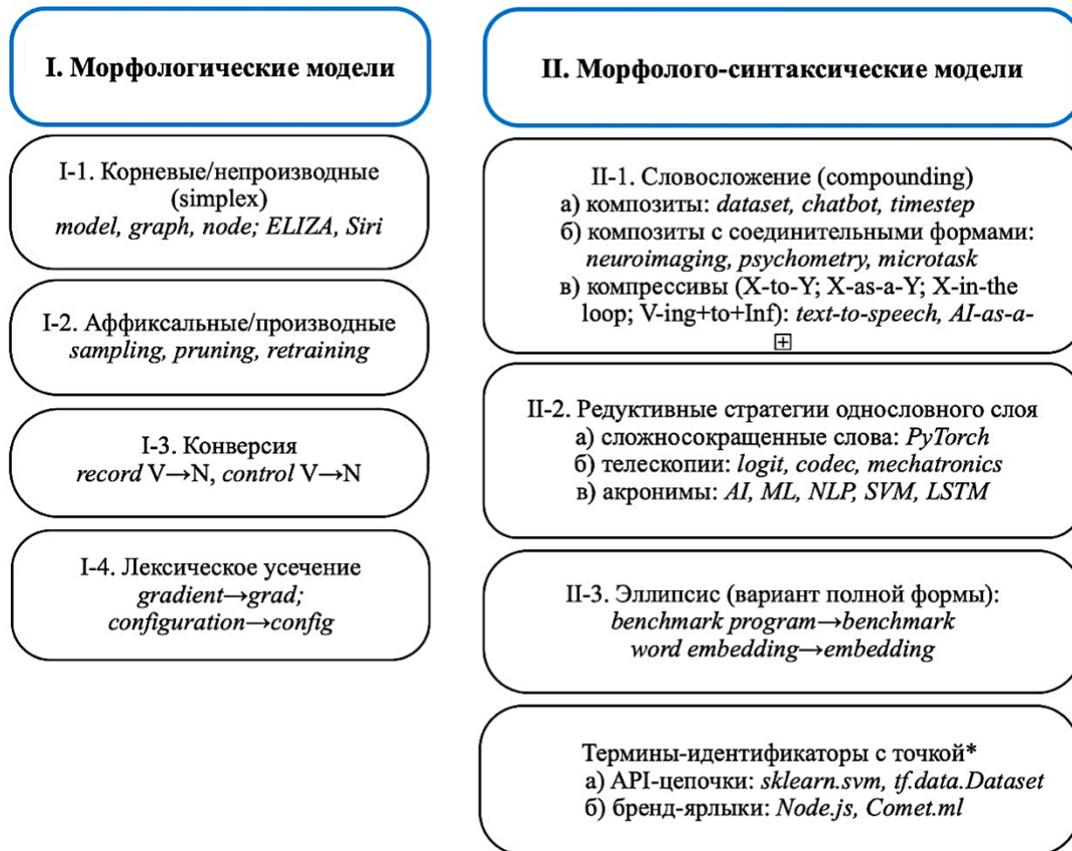
Рисунок 3 – Соотношение терминов-слов и терминов-словосочетаний
в терминосистеме ИИ

Учитывая установленное соотношение, дальнейшее изложение организовано по возрастанию синтагматической сложности – от однословного слоя к многословному. Методически это соответствует корпусному подходу, в рамках которого сначала формируется опорный список однословных терминов, затем на его основе извлекаются многословные единицы через коллокационный поиск с использованием статистических метрик (MI, LL) и структурных шаблонов (A+N, N+N, N+prep+N и др.). Для однословных единиц применяются морфемный разбор, частотные показатели и деривационно-редуктивный анализ (аффиксация, словосложение; редуктивные стратегии – сложносокращенные слова, акронимы, телескопии); для многословных – частотно-коллокационный анализ с последующей синтаксической и контекстной (KWIC) верификацией. На заключительном шаге результаты используются для функциональной стратификации терминов на стабильное «ядро» и инновационную «периферию», при этом многословность не тождественна принадлежности к последней.

3.1. Морфологические и морфолого-синтаксические модели образования однословных терминов ИИ

При доминировании многословных комбинаций именно однословный слой задает норму называния и служит опорой для извлечения и интерпретации сложных единиц. Принимая во внимание положения работ В. П. Даниленко и С. В. Гринева-Гриневича, а также других исследователей, разработавших структурные классификации терминов [Даниленко 1977; Cabré 1999; Гринева-Гриневич 2008; Суперанская 2012; Лейчик 2022], выделим основные классы терминов-слов: корневые или производные (*corpus* (корпус), *metric* (метрика)); аффиксальные или производные (*underfitting* (недообучение)); сложные (*database* (база данных), *keypoints* (ключевые точки)); аббревиатуры (*AI* (ИИ), *bioBERT*, *theranostics* (тераностика)).

Для операционализации дальнейшего анализа сведем эти типы в единую схему, фиксирующую классы и подгруппы морфологических и морфолого-синтаксических моделей терминообразования (см. Рисунок 4).



*Термины-идентификаторы с точкой – тип записи, не способ терминообразования

Рисунок 4 – Методы образования однословных терминов ИИ:
классы и подгруппы

Количественное распределение терминов-слов представлено на Рисунке 5. При подсчете не учитывались случаи эллипсиса и лексического усечения, рассматриваемые в работе как варианты полной терминологической формы.

● непроизводные ● аффиксальные ● сложные
● термины-идентификаторы ● сокращения ● конверсия

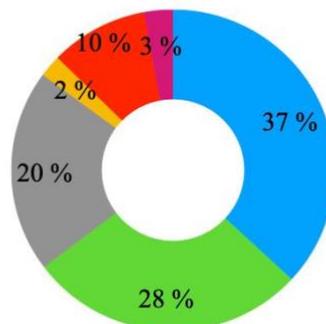


Рисунок 5 – Количественное распределение терминов-слов по типам морфологического и морфолого-синтаксического терминообразования (N = 450; 27% корпуса)

Корпусный анализ выявил существенный массив высокочастотной **непроизводной лексики** – 166 единиц (37% однословного слоя). Показательной характеристикой данного пласта является его преимущественно заимствованный характер (139 ТЕ), отражающий глубинную историческую интерференцию и процесс систематической адаптации научной лексики, сопровождавший формирование дискурса ИИ. Этимологически крупнейшая доля единиц восходит к латинскому языку (*algorithm* (алгоритм), *parameter* (параметр), *corpus* (корпус), *data* (данные) и др.), греческому – (*autonomy* (автономность), *semantics* (семантика), *homophone* (омофон)), а также немецкому – (*entropy* (энтропия), *batch* (пакет), *token* (токен)). Зафиксированы отдельные примеры из итальянского (*replica* (копия)) и чешского (*robot* (робот)). Ряд единиц демонстрирует сложные пути заимствования: латинско-французские (*action* (действие), *computation* (вычисление), *verification* (верификация)) и латинско-греческие (*praxis* (практика)); ср. *algorithm* (алгоритм), восходящий через средневековую латынь к арабскому источнику. Одновременно в однословном слое учитываются непроизводные номены – собственные наименования индивидуально идентифицируемых артефактов и ресурсов предметной области (модели, системы, платформы, датасеты, библиотеки, бенчмарки), не требующие обязательной системной кодификации: *ELIZA*, *Siri*, *Ada*, *Kaggle* и др.

Непроизводные термины формируют в дискурсе ИИ консервативное, интернациональное «ядро» понятийного аппарата. Этимологические связи с классическими языками науки (латинским и древнегреческим), а также с европейскими языками-донорами позволяют рассматривать их как элемент «интернационального ядра» научной лексики [Wüster 1979; Cabré 1999; Лейчик 2022]. Этот слой фиксирует историческую преемственность и обеспечивает интеграцию ИИ в широкий междисциплинарный и наднациональный контекст, где интернационализмы выполняют функцию унификации понятийного аппарата и облегчают межъязыковую коммуникацию.

Аффиксальные (производные) термины составляют 125 единиц (28% однословного слоя: суффиксальные – 94, префиксальные – 30, префиксально-суффиксальные – 1). Продуктивные **суффиксальные модели** и их распределение приведены в Таблице 10.

Таблица 10 – Продуктивные суффиксальные модели в терминосистеме ИИ

Суффикс	Значение суффикса	Кол-во ТЕ	Примеры
-ing	обозначение действий, процессов или операций	32	<i>sketching</i> (скетчинг), <i>scoring</i> (оценивание)
-ity	абстрактное качество, состояние, характеристика	13	<i>stationarity</i> (стационарность), <i>sparsity</i> (разреженность)
-ation/ -ization	процессы, методы или результаты процессов	13	<i>visualization</i> (визуализация), <i>anonymization</i> (анонимизация)
-or/-er	агент/ инструмент	12	<i>quantifier</i> (квантификатор), <i>transformer</i> (трансформер), <i>layer</i> (слой)
-ence/ -ance	абстрактное состояние, способность или характеристика	3	<i>resilience</i> (устойчивость), <i>convergence</i> (сходимость), <i>performance</i> (производительность)

Численное преобладание суффиксально образованных терминов в составе производной лексики согласуется с деривационной типологией английского языка и указывает на ведущую функцию суффиксации в терминообразовании. Для терминосистемы ИИ это означает, что систематизация понятийного аппарата преимущественно достигается посредством суффиксальной экспликации признаков и категориальных значений, что поддерживает устойчивость и унификацию терминологии.

Данные Таблицы 10 показывают высокую продуктивность суффиксов -ing, -ity, -er/-or, -ation/-ization, что отражает ключевые тенденции номинации в терминологии ИИ. Суффикс *-ing* служит для обозначения процессов и процедур (*sampling* (выборка), *pruning* (отсечение), *debugging* (отладка)), фиксируя динамический характер исследуемых явлений. Суффикс *-ity* выражает абстрактные понятия и качества, часто отражающие свойства систем, моделей или алгоритмов (*flexibility* (гибкость), *complexity* (сложность), *variability* (вариативность)). Суффиксы *-or/-er* указывают на агента или инструмент действия (*encoder* (энкодер), *parser* (парсер), *optimizer* (оптимизатор)), закрепляя в терминологии роль специализированных модулей и компонентов. Суффиксы *-ation/-ization* обозначают процессы и результаты действий (*normalization* (нормализация), *optimization* (оптимизация)), обеспечивая высокий уровень системности и регулярности в наименовании операций.

Таким образом, суффиксация в терминосистеме искусственного интеллекта выполняет две ключевые функции. С одной стороны, она способствует номинализации процессов и состояний, превращая динамические явления в устойчивые термины; с другой – обеспечивает категориальную дифференциацию понятийного аппарата, позволяя разграничивать действия, их результаты, характеристики и участников. Высокая частотность указанных словообразовательных элементов подтверждает их деривационную продуктивность в профессиональном дискурсе ИИ, а также иллюстрирует общее стремление к стандартизации и прозрачности лексических структур, что облегчает восприятие и интерпретацию новых терминов специалистами в междисциплинарном контексте.

Префиксальная деривация в исследуемом корпусе менее продуктивна, чем суффиксальная, что соответствует общим тенденциям английского словообразования, включая специализированную лексику [Plag 2018]. Корпусный репертуар префиксальных формантов, участвующих в образовании однословных терминов ИИ, представлен в Таблице 11.

Таблица 11 – Префиксальные модели в терминосистеме ИИ

Префикс	Значение префикса	Кол-во ТЕ	Примеры
re-	повтор и итеративность	3	<i>retraining</i> (дообучение), <i>recall</i> (отзыв), <i>recombination</i> (рекомбинация)
self-	автономия и самонастройка	3	<i>self-attention</i> (механизм самовнимания), <i>self-training</i> (самообучение)
hyper-	градуация степени	3	<i>hyperplane</i> (гиперплоскость), <i>hyperheuristic</i> (гиперэвристика)
post-	после-/постэтап	2	<i>posthuman</i> (постчеловек), <i>postprocessing</i> (постобработка)
over-	превышение/пере-	2	<i>overfitting</i> (переобучение), <i>oversampling</i> (передискретизация)
co-	совместность/координация	2	<i>co-training</i> (совместное обучение), <i>co-dadaption</i> (коадаптация)
super-	над-/усиление	2	<i>superalignment</i> (сверхсогласование), <i>superintelligence</i> (сверхинтеллект)
de-	снятие/реверсия	2	<i>decomposition</i> (декомпозиция / разложение), <i>denoising</i> (шумоподавление)

Префикс	Значение префикса	Кол-во ТЕ	Примеры
meta-	мета-уровень/ вторичный порядок	2	<i>metadata</i> (метаданные), <i>meta-learning</i> (метаобучение)
under-	недостаточность/ дефицит	2	<i>underfitting</i> (недообучение), <i>undersampling</i> (недостаточное сэмплирование)

Префиксация в терминосистеме ИИ выполняет преимущественно семантически дифференцирующую функцию. Наиболее активны модели, связанные с повторяемостью, автономностью и усложнением (*re-*, *self-*, *hyper-*), тогда как менее продуктивные префиксы уточняют вторичные признаки.

В корпусе отмечен единичный случай **префиксально-суффиксальной деривации** – *trans-human-ism* (трансгуманизм), где префикс маркирует направленность, а суффикс – категориальную принадлежность.

Конверсия (нулевая деривация, безаффиксное словообразование) представлена 13 ТЕ, в которых переход из одной части речи в другую происходит без изменения основы. Типичные примеры: *record* V → *record* N (запись), *control* V → *control* N (контроль), *mark up* V → *markup* N (разметка), *prompt* V → *prompt* N (промт). В широком деривационном толковании сюда же отнесен случай деадъективной субстантивации *biometric* A → *biometrics* N (биометрия), где формальный маркер *-s* выступает как средство лексикализации «полевого» существительного.

Лексическое усечение (*clipping*, *truncation*) трактуется как сокращение исходного слова или словосочетания до одного сегмента (обычно начальной части или основы). В отличие от аббревиации (инициальной/звуковой), усечение не образует новую единицу из инициалов компонентов, а реализует графемно-морфемное сокращение без изменения лексической природы термина. В современной морфологии усечение описывается как самостоятельный механизм вне «регулярных» словообразовательных правил, но устойчиво продуктивный в техно-научном регистре [Mattiello 2013].

В данной работе усеченные формы не выделяются как отдельная модель в сводной типологии (рассматриваются как варианты по отношению к полным формам), однако их частотность зафиксирована. В корпусе обнаружено 14 случаев, например, *network* → *net* (сеть); *graphic formula* → *graph* (графическая формула) и др.

Усечения обеспечивают компрессию при сохранении семантического ядра и высокой распознаваемости в профессиональном контексте. Такие формы чаще характерны для устной и полупрофессиональной письменной коммуникации; вместе с тем в научных статьях они регулярно используются при описании экспериментальных настроек и технических процедур (например, «update params», «apply norm», «calc grads»), а также в коде, формулах и схемах, где *conv* функционально предпочтительнее *convolutional layer* (сверточный слой). Количественно уступая более продуктивным моделям, усечения остаются функционально значимыми как средство экономии формы и маркер профессионального регистра.

Морфолого-синтаксическое терминообразование включает способы формирования однословных терминов из словосочетаний через синтаксико-морфологические преобразования. К ним относятся эллипсис (с переносом значения на опорный компонент), словосложение (в т. ч. компрессивы и термины-символослова) и редуцированные модели. Эти пограничные механизмы противопоставляются аффиксации и характеризуются устойчивой продуктивностью в научно-техническом регистре [Cabré 1999; Kageura 2002;].

Эллипсис представляет собой опущение компонента терминологического словосочетания, при котором оставшееся слово репрезентирует весь исходный комплекс значений. В корпусе зафиксировано 9 ТЕ, например: *benchmark program* (бенчмарк-программа) → *benchmark, data sampling* (семплирование/выборка данных) → *sampling*. Как и усечения, эллиптические формы трактуются как вариант полной номинации; в количественном учете они рассматриваются как термины-словосочетания и, соответственно, исключаются из статистики терминов-слов.

В терминосистеме ИИ одним из наиболее продуктивных способов словообразования выступает **словосложение** (compounding), традиционно рассматриваемое как второй по значимости механизм в английском языке. В корпусе зафиксировано 92 сложных однословных термина ($\approx 20\%$ выборки однословных ТЕ). Типология компаундинга наиболее вариативна и включает классические N+N-композиции, образования со связными (неоклассическими) формами, дефисные компрессивы (X-to-Y, X-as-a-Y, X-in-the-loop, V-ing+to+Inf) и символично-буквенные композиции.

Подавляющее большинство сложных терминов представляют собой специальные лексические единицы, состоящие из двух корневых морфем – 34 термина ($\approx 37\%$ от числа сложных единиц). Примерами могут служить термины *database* (база данных), *keypoints* (ключевые точки), *timestep* (шаг времени), *benchmark* (бенчмарк) и др. Эти образования отличаются высокой степенью семантической прозрачности и закрепляют фундаментальные понятия предметной области.

Следующую подгруппу составляют термины с соединительными (неоклассическими) формами (combining forms): *uni-*, *back-*, *neuro-*, *geo-*, *bio-*, *micro-*, *radio-*, *cross-*, *mini-*, *-gram*, *auto-*, *-bot*, *-athon*, *neo-* и др. Всего зафиксирована 21 единица. Эти образования не относятся к деривации в узком смысле, так как образуются посредством сочетания двух и более основ, каждая из которых вносит семантический компонент: *neuroimaging* (нейровизуализация), *biometrics* (биометрия), *microtask* (микрзадание), *backpropagation* (обратное распространение) и др. Такая трактовка поддерживается как англоязычной морфологической традицией, так и русской терминоведческой школой. В англоязычной литературе neoclassical combining forms рассматриваются как «связанные корни» в составе сложных слов, а не как аффиксы [Plag 2018]; в русской традиции сочетания типа нейро-/био- + основа описываются как сложение с формантами, то есть вне «узкой» деривации [Гринева-Гриневич 2008; Лейчик 2022].

Встречаются также термины, полученные сочетанием двух соединительных форм, каждая из которых обладает высокой степенью автономии: *phonology* (фоносистема), *psychometry* (психометрия), *bigram* (биграмма) и др. Сюда же относится термин *N-gram*, где буквенный символ *N* указывает на переменную величину, а формант *-gram* выполняет функцию словообразовательной основы, вместе они формируют неделимую лексическую единицу.

Внутри категории сложных терминов выделяются устойчивые дефисные компрессивы технологического регистра. Модель **X-to-Y** компактно кодирует преобразование (*text-to-speech* (синтез речи), *speech-to-text* (распознавание речи)) и активно порождает неологизмы (*prompt-to-3D* (генерация 3D по текстовому промпту); *code-to-text* (преобразование кода в текст)). Модель **X-as-a-Y** обслуживает сервисные решения (*API-as-a-service* (API как сервис); *AI-as-a-service* (ИИ как сервис)), а **X-in-the-**

loop фиксирует включенность агента в контур принятия решений (*human-in-the-loop* (человек в контуре); *AI-in-the-loop* (ИИ в контуре)).

Рассматриваемые модели являются компрессивами – фразовыми/формульными композитами, сворачивающими развернутую парафразу (mapping of X to Y; provision of X as a Y; inclusion of X in the loop) в дефисную цепочку с «замороженными» связками -to- / -as-a- / -in-the- и формирующими лексикализованные единицы с головным компонентом (service/platform/loop), функционирующие как одно слово (например, *AI-as-a-service offering* (предложение формата «ИИ как сервис»), *human-in-the-loop pipeline* (конвейер с человеком в контуре)). Аббревиатуры в левой позиции (*AI, API, GPU, CPU*) выступают как основы-модификаторы, устойчиво комбинируются, сохраняют дефисную орфографию, допускают дальнейшую деривацию (*AI-as-a-service-based* (на основе сервисной модели ИИ)) и блокируют внутренние грамматические вариации (связка as/a не варьируется; множественное число маркируется на головном имени). По результатам операционных тестов словосложения (лексикализация, эндоцентричность, единая акцентная группа, недоступность внутренней флексии, атрибутивное распределение) конструкции X-to-Y, X-as-a-Y и X-in-the-loop квалифицируются как композиты, а не свободные словосочетания, что позволяет единообразно включать их в слой словосложения [Plag 2018].

Отдельно выделяется компрессив **V-ing+to+Infinitive** (*learning-to-rank, LTR* (обучение ранжированию) и *learning-to-learn, L2L* (метаобучение)), где герундий кодирует процесс, а инфинитив – целевую ориентацию («PROCESS aimed at achieving GOAL»). Эта модель обеспечивает семантическую компрессию и экспликацию метаконцептов; в интерпретации Х. К. Сажера и К. Кагеуры такие конструкции соотносятся с грамматикализацией, мотивированной частотностью и потребностью в стандартизации научно-технического дискурса [Sager 1990; Kageura 2002].

Наконец, в корпусе зафиксирована ограниченная группа сложных **терминов-символослов** (10 ТЕ), включающих буквенные и числовые обозначения: *Q-function* (функция-Q), *F1-score* (оценка F1), *k-means* (метод k-средних) и др. Несмотря на наличие символьного компонента, данные единицы сохраняют свойства композитов, поскольку каждый элемент вносит семантический вклад в значение термина [Даниленко 1977; Гринев-Гриневиц 2008].

Таким образом, словосложение обеспечивает базовую номинацию и компактное кодирование процедур, конфигураций и сервисных моделей, повышая семантическую плотность терминологии ИИ за счет неоклассических формантов и символично-буквенных компонентов при сохранении распределительной целостности лексемы.

Сокращения (аббревиация) – эффективный способ сжатия развернутых терминологических словосочетаний (в среднем в 5–10 раз), распространение которого усиливается по мере удлинения терминов и роста доли трех- и более компонентных наименований. В настоящем исследовании редуцированный блок включает три модели: сложносокращенные слова, телескопии и инициальную/звуковую аббревиацию; их объединяют сокращение протяженности исходной номинации, лексикализация результата и вторичный характер новообразования.

Ключевым для нашего анализа является различие диахронической и синхронической аббревиации, основанное на уровне языковой интеграции и определяющее правила терминографического учета. **Диахроническая аббревиация** – это лексикализованная единица, функционирующая как обычное слово (нижний регистр, словарная фиксация, деривационная продуктивность, участие в композитообразовании); такие элементы учитываются как однословные термины и входят в типологическую статистику (например, *Lisp*, *Pytorch*, *Matplotlib*). **Синхроническая аббревиация** – дискурсивный, «здесь-и-сейчас» вариант полной номинации (например, *AI*, *ML*, *NLP*, *LSTM*; *ономастические акронимы BERT, LaMDA, BART*), который вводится через раскрытие полной формы и далее используется ради ее экономии. В корпусной методике такие сокращения нормализуются к полной форме, помечаются как варианты (*variant-of* = полная номинация), их частоты агрегируются к базовой единице [Kageura 2002; ISO 1087:2019; ISO 704:2022].

Диагностика лексикализации опирается на классические критерии: понижение регистра (но не всегда) / орфографическая интеграция, словарная фиксация, деривационная активность [Plag 2018]. Такой подход, поддерживаемый и русской терминоведческой традицией в описании вариативности и «словоидов» [Даниленко 1977; Гринев-Гриневиц 2008; Лейчик 2022], обеспечивает теоретическую корректность и сопоставимость количественных результатов.

В основе образования **сложносокращенных слов** лежит усечение исходных компонентов до «стержней» их линейной конкатенации без наложения, при этом

границы между сегментами остаются прозрачными, а развертывание источников – детерминированным. В корпусе фиксируются типичные образования: *PyTorch* ← *Py(thon)* + *Torch*; *datalog* ← *data* + *logic*; *AlexNet* ← *Alex* + *net(work)* (эпоним + усеченная основа) и др.; всего 33 ТЕ.

У параллельной редуктивной практики – **клиппинга многословных терминов** – сохраняется исходная модификаторная структура, но компоненты сокращаются до минимально различительных опор и сливаются в единый токен. В терминосистеме ИИ устойчиво фиксируются следующие формы: *recsys* (рекомендательная система) (← *recommendation system*), *logreg* (логистическая регрессия) (← *logistic regression*) и *linreg* (линейная регрессия) (← *linear regression*) и др. Подобные формы, часто используемые в коде и профессиональном сленге, при наличии лексикализации рассматриваются как переходные между синтаксическим и морфологическим уровнями и засчитываются в однословный слой.

Современные сложносокращенные термины часто включают несколько сокращенных элементов и соединяют полнозначные слова, усеченные основы, акронимы и словообразовательные форманты. Так, *Matplotlib* восходит к *MATLAB* + *plotting* + *library*, *Scikit-learn* – к *scientific* + *toolkit* + *learn* и далее сокращается до *sklearn*; к этому же типу относятся *mobile-BERT*, *KL-ONE*, *bioBERT*, *roBERTA* и др.

В корпусной классификации такие образования, как *mobileBERT*, *distilBERT* и *bioBERT*, трактуются как гибридные композиты с правой головностью (голова = *BERT*). В отличие от них, *roBERTa* и *deBERTa* являются не словосложением, а акронимной деривацией – переосмысленными номенами внутри «семейства BERT», что обуславливает их учет среди акронимных производных.

Телескопия (*blending*) определяется как создание новой лексемы путем наложения и взаимного проникновения усеченных фрагментов компонентов, где стык недетерминирован и часто сопровождается фонографической спайкой. В корпусе представлены характерные примеры телескопии (всего 7 ТЕ): *logit* (логит) ← *logarithmic* + *unit*, *flaxformer* (трансформер на Flax) ← *Flax* + *transformer* и др.

Статус телескопии в терминоведческой литературе дискуссионен, что отражает разнообразие методологических подходов – от формального [Heller, Macris 1968], акцентирующего структурные особенности, до формально-семантического, учитывающего как план выражения, так и план содержания [Лейчик 2022]. Это

порождает различные варианты классификации, включая отнесение телескопии к аббревиатурному словообразованию [Гринев-Гриневиц 2008]. В настоящем исследовании телескопия операционально выделена как самостоятельная редуктивная модель на основе признака наложения на стыке компонентов и учитывается в однословном слое отдельной подрубрикой.

Пределом редукции является **акронимия** – инициальная или звуковая аббревиация, например, *KAI* (корейский ИИ), *ANFIS* (адаптивная нейро-нечеткая система вывода), а также цифро-буквенные комплексы вроде *YOLOv8* и *GPT-4*. Всего в корпусе зафиксировано 170 случаев акронимии, что указывает на высокую степень формальной компрессии однословного слоя при сохранении терминологической однозначности. В рамках концептоцентричной нормализации инициальные/звуковые акронимы трактуются как синхронические варианты полной номинации и не образуют самостоятельных типов; их частоты агрегируются к базовой терминоединице. Исключение составляют диахронически лексикализованные акронимы (например, *Lisp* (язык программирования Лисп), учитываемые как автономные однословные термины.

Особый интерес представляют аббревиатуры, омонимичные полнозначным словам (*MARS*, *ALVINN*, *ADAM*). Эти единицы обладают двойной мотивацией, а их положительные качества – произносимость и запоминаемость – при отсутствии опасности смешения с общеупотребительной лексикой ведут к их широкому распространению в терминосистемах разных языков.

В отдельную группу терминов-слов (10 ТЕ) выделены **термины-идентификаторы**: 1) **namespace-цепочки** программных экосистем, где точка служит иерархическим разделителем «платформа/пакет/модуль/класс» и кодирует адресацию внутри API: *sklearn.svm*, *tf.data.Dataset*, *tf.Example*, *tf.layers*, *torch.nn*, *jax.numpy* (левый компонент часто – аббревиатура экосистемы: *tf* = *TensorFlow*; *nn* = *neural networks*; *svm* = *support vector machine*); 2) **брендонимы с доменным маркером** (точка имитирует доменную зону/технологический ярлык): *Node.js*, *Comet.ml*, *Wit.ai*, *React.js*, где точка не задает модульную иерархию, а выступает технологическим ярлыком (*.js*, *.ml*, *.ai*) и входит в онимную категорию: *Node.js*, *Comet.ml*, *Wit.ai*. Эти единицы включаются в сводную статистику как «термины-идентификаторы» с нормализацией орфографии и, при необходимости, агрегацией к базовой концепции/записи API, без распределения по моделям аббревиатурного словообразования.

Суммарно аббревиация, сложносокращенные слова, телескопии, а также термины-идентификаторы образуют согласованный набор редукативных стратегий, который обеспечивает высокую степень формальной экономии при сохранении терминологической точности. Доля английских терминов ИИ (12% от общего объема терминов-слов), созданных с помощью этих морфолого-синтаксических механизмов, подтверждает их продуктивность и объясняет устойчивое распространение в корпусе.

Таким образом, однословные термины ИИ формируются главным образом за счет суффиксации и словосложения, включая композиты с неоклассическими формантами и формульные модели, которые обеспечивают компактную номинацию процессов, объектов и конфигураций. Редукативные стратегии (сложносокращенные формы, телескопии, аббревиации) усиливают тенденцию к экономии выражения при сохранении семантической прозрачности и профессиональной распознаваемости. Методологическое разграничение лексикализованных единиц и синхронных вариантов полной номинации (включая эллипсис, усечения и акронимы) обеспечивает корректный учет и сопоставимость количественных результатов без двойного счета.

3.2. Синтаксическая деривация терминов искусственного интеллекта: двух- и многокомпонентные модели

Корпусный анализ терминосистемы ИИ демонстрирует абсолютное преобладание терминов-словосочетаний (1198 ТЕ; 73%) над однословными терминами (450 ТЕ; 27%), что подтверждает синтагматически-композиционный тип терминообразования в данной области. Это обусловлено необходимостью точного отражения функциональных и иерархических характеристик понятий высокотехнологичных доменов [Bowker, Pearson 2002].

Синтаксическое терминообразование, опирающееся на лексико-синтаксические сочетания из двух и более слов [ISO 1087:2019], является одним из ведущих способов пополнения терминологии. По оценке С. В. Гринева-Гриневица, таким путем образуется до 60–95% терминологического фонда европейских языков [Гринева-Гриневиц 2008], что подтверждается корпусными исследованиями специализированных доменов [Adjali 2022].

Аналитический строй английского языка способствует легкости образования многокомпонентных наименований через юкстапозицию (постановку рядом), что

ведет к «синтаксической компрессии» – концентрации смысла в атрибутивных именных группах с правоголовой структурой [Biber, Gray 2016]. Например, цепочка, образованная на основе высокочастотного ядерного термина *model* (модель), *predictive model* (предиктивная модель) → *ensemble predictive model* (ансамблевая предиктивная модель) → *explainable ensemble predictive model* (объяснимая ансамблевая предиктивная модель), демонстрирует наращивание дополнительных признаков (предсказательность → ансамблевость → интерпретируемость). Подобные многоступенчатые сочетания характеризуются высокой номинативной емкостью и внутренней связностью; сложное терминологическое словосочетание функционирует как свернутое определение, отвечая установке научного подязыка на лаконичность и точность [Bowker, Pearson 2002; L'Homme 2020; Лейчик 2022].

Наряду с этим устойчиво функционируют идиоматичные терминологические словосочетания (термины-фразеологизмы), формируемые механизмами концептуальной метафоры и метонимии [Lakoff, Johnson 1980; Temmerman 2000; Суперанская 2012]. Показателен метафорический термин *curse of dimensionality* (проклятие размерности), значение которого не выводится композиционно через сумму значений компонентов, а идиоматически именуется несколько эффектов роста размерности (взрыв объема данных/поиска, разреженность выборок, деградация метрик). Единица воспроизводится в фиксированной форме, однозначно соотносена с понятием и функционирует как термин.

Структурный анализ терминов-словосочетаний, выполненный методами корпусной лингвистики, выявил устойчивые закономерности их компонентной организации. Ключевой процедурой было выделение коллокационных кластеров вокруг высокочастотных ядерных терминов (*model*, *network*, *learning*, *system*) в AntConc (модуль Collocates, окно ±5, ранжирование по ассоциативным мерам), что позволило систематизировать типовые шаблоны A+N, N+N, V-ing+N, V-ed+N и др. «Профиль» конструкции описывался через длину номинации (число лексем) с разграничением двукомпонентных и многокомпонентных единиц; кандидаты верифицировались по конкордансам (KWIC). Процентное распределение по длинам показано на Рисунке 6.



Рисунок 6 – Компонентная протяженность терминов-словосочетаний: процентное распределение ((N = 1198; 73% корпуса)

Базовой и наиболее частотной моделью терминологических словосочетаний в терминосистеме ИИ является двухкомпонентная атрибутивная конструкция (72% от общего числа терминологических словосочетаний), в которой правый (головной) компонент служит видовой или родовой основой, а левый препозитивный модификатор систематически вносит дифференцирующий признак [Суперанская 2012: 99]. Так, в терминах-словосочетаниях, образованных на основе ядерного термина *logic* (логика), головной компонент употребляется в одном и том же значении – «the science or study of careful reasoning using formal method» (наука о формах, методах и законах интеллектуальной познавательной деятельности, формализуемых с помощью логического языка) [Longman Dictionary 2025]. Второй же уточняет значение, указывая на дополнительный план-содержание: *belief logic* (логика веры), *common-sense logic* (логика здравого смысла), *constructive logic* (конструктивная логика). Эти конструкции совмещают номинативную и частично дефинитивную функции, компактно эксплицируя существенные признаки концепта.

Типология атрибутивной двухкомпонентной модели определяется морфологической природой и синтаксической ролью левого компонента. Далее рассмотрим ее ключевые реализации, чтобы операционально развести шаблоны N+N, A+N, N+V-ing, A+V-ing, V-ed+N, V-ing+N, V-ed+V-ing, N+abb / abb+N и сопоставить их корпусные доли; предлоговые группы N+prep+N и редкие притяжательные формы N's+N учитываются как малочисленные варианты.

N+N (noun+noun) – 353 ТЕ (≈40%). Головной правый компонент – таксономический «якорь» в системе терминов, отражающий принадлежность к определенной тематической группе; левый компонент несет видовую (специализирующую) характеристику, уточняющую значение термина. Такая структура обладает высокой степенью мотивированности и прозрачности, а сам процесс терминообразования предполагает регулярное присоединение новых определяющих компонентов к устойчивым родовым основаниям. Например, в термине *error backpropagation* (обратное распространение ошибки) первый компонент *error* (ошибка) указывает на характер передаваемой информации, тогда как второй компонент *backpropagation* (обратное распространение) является головным, определяющим родовую принадлежность термина к области методов обучения нейронных сетей. В данном случае *backpropagation* выступает как устойчивая основа, к которой может быть присоединено множество атрибутивных уточнений: *batch backpropagation* (пакетное/батчевое обратное распространение ошибки); *gradient backpropagation* (обратное распространение градиента) и др. В качестве препозитивного определения может выступать не только имя нарицательное, но и имя собственное, образуя так называемые термины-эпонимы (*Ebert test* (тест Эберта)). Встречаются также конструкции с двойным антропонимом в препозиции: *Damerau–Levenshtein distance* (расстояние Дамерау-Левенштейна).

A+N (adjective+noun) – 263 ТЕ (≈30%). Левый компонент (прилагательное) кодирует дифференцирующий признак, правая голова (существительное) фиксирует таксономический класс (*model, system, network, algorithm*): *artificial intelligence* (искусственный интеллект), *neural network* (нейронная сеть). В данной модели преобладают относительные прилагательные с формантами -al, -ic, -ive, -ary/-ory (*neural, probabilistic, generative, supervisory*). Семантика признака распределяется по четырем подтипам: доменные (*linguistic, visual*), структурные (*convolutional, recurrent*),

функциональные (*generative, robust*), методологические (*probabilistic, bayesian*). Модель конкурирует с N+N (*language model vs linguistic model*); выбор зависит от наличия устоявшегося относительного прилагательного и степени институционализации головы. Формально модель A+N компактна, серийна и малоградуируема, поэтому служит «рабочим» шаблоном для быстрых расширений (*supervised learning* (обучение с учителем) → *semi-supervised learning* (полуобучение с учителем) → *self-supervised learning* (самообучение с учителем) → *unsupervised learning* (обучение без учителя)).

N+V-ing (gerund) (noun + gerund head) – 58 ТЕ (≈7%). Левый именной компонент параметризует процесс (механизм, ресурс, объект, критерий), правая голова-герундий номинирует процесс/метод. Модель компактно репрезентирует динамические процедуры и образует серийные гнезда типа *X learning / X mining / X modeling: reinforcement learning* (обучение с подкреплением), *representation learning* (обучение представлениям), *curriculum learning* (поэтапное обучение).

A+V-ing (adjective + gerund head) – 45 ТЕ (≈5%). Шаблон представляет собой атрибутивное словосочетание, где первый компонент (прилагательное) выполняет функцию определяющего элемента, а второй (герундий) – роль головного компонента, обозначающего таксономическую категорию процесса, метода. Герундий, как правило, именуется центральный процесс, тогда как прилагательное уточняет его свойства, режимы или параметры. Подобные структуры высокопродуктивны в научно-технической терминологии и обладают высокой степенью семантической мотивированности. Например, в термине *incremental learning* (пошаговое обучение, инкрементное обучение) прилагательное *incremental* указывает на способ осуществления процесса – поэтапный, нарастающий, а герундий *learning* обозначает сам процесс обучения.

В оппозиции N+V-ing ↔ A+V-ing первая модель предпочтительна при необходимости субстантивной спецификации источника, объекта, критерия (*representation/metric/curriculum learning*), вторая – при акценте на режимном или качественном параметре процесса (*incremental/active/deep learning*).

Отдельный подкласс составляют смешанные модели **N+abb** и **abb+N** (47 ТЕ; ≈5,5%), где *abb* – это аббревиатура (чаще всего технического или методологического характера), а *N* – существительное. Такой формат обеспечивает компактные, серийно

воспроизводимые наименования методов, архитектур и метрик ИИ. В конструкции *abb+N* аббревиатура выступает препозитивным детерминатором, а головой является существительное (*CNN model* (модель сверточной нейронной сети)); в *N+abb* головным компонентом служит аббревиатура, слева – доменная спецификация (*game AI* (игровой ИИ)). Эти схемы активно используются в профессиональной коммуникации благодаря высокой частотности акронимов, маркирующих алгоритмы, модели и подходы.

В двухкомпонентном слое фиксируется ряд моделей с глагольной морфологией в атрибутивной позиции, которые кодируют аспект действия, способ или результат, обеспечивая синтаксическую компрессию:

- **V-ed+N** (24 ТЕ; ≈2,8%) – препозитивное причастие прошедшего времени маркирует результативный, завершённый признак (*trained model* (обученная модель), *compressed representation* (сжатое представление));

- **V-ing (gerund)+N** (17 ТЕ; ≈2%) – препозитивный герундий обозначает процесс/процедуру и задает целевое назначение головного компонента (*learning algorithm* (алгоритм обучения), *training set* (обучающая выборка));

- **V-ed+V-ing** (7 ТЕ; ≈0,8%) – препозитивное причастие прошедшего времени конкретизирует способ или режим действия, выраженного герундием (*supervised learning* (контролируемое обучение), *regularized training* (регуляризованное обучение));

- **V-ing (participle)+N** (5 ТЕ; ≈0,6%) – препозитивное причастие настоящего времени указывает на активное свойство, динамическое состояние объекта (*learning system* (обучающаяся система), *streaming data* (поточковые данные)).

Незначительная доля терминов ИИ (2,6% двухкомпонентных терминов) образована по модели **N+prep+N** (именная группа с постпозиционной предложной детерминацией), например, *theory of computation* (теория вычислений), *area under the curve* (площадь под кривой). Ограниченная представленность модели обусловлена общим дрейфом научно-технического дискурса к лексико-синтаксической компрессии. Препозитивные беспредложные конструкции (*graph theory* (теория графов)) короче, легче унифицируются, проще встраиваются в длинные именные цепочки и лучше поддаются последующей акронимизации. Тем не менее **N+prep+N** сохраняет функциональную нишу там, где требуется явная экспликация ролевых отношений и устранение двусмысленности (*of* маркирует таксономическую или

комплементарную связь, *under* – пространственно-метрический параметр, *without* – операционный режим). В таких контекстах постпозиционный формат обеспечивает наибольшую интерпретационную прозрачность при умеренной потере компактности.

Конструкция **N's+N (noun's+noun)** встречается крайне редко ($\approx 0,5\%$ двухкомпонентных единиц) и характеризуется низкой словообразовательной продуктивностью. Притяжательный падеж ('s) обычно маркирует принадлежность метода, инструмента или концепции определенному автору/системе, при этом одушевленность компонента терминологически нерелевантна. Например, в термине *experimenter's bias* (предвзятость/систематическая ошибка экспериментатора) модификатор *experimenter's* указывает на источник/носителя искажения, головной компонент *bias* называет сам феномен (предвзятость, систематическая ошибка). Нередко такая связь может быть равноценно выражена моделями N+N (*experimenter bias*) или N+of+N (*bias of the experimenter*), что делает использование 's в терминологии факультативным и стилистически маркированным.

К непродуктивным моделям относятся также редкие форматы **A+A (adjective+adjective)** и **abb+Ving (abbreviation+gerund)**. Первая модель представлена в названии IBM *Deep Blue* (техническая характеристика + корпоративная отсылка), вторая – в единицах типа *AI planning* (автоматическое планирование) и *GPU computing* (GPU-вычисления). Эти модели периферийны и чаще связаны с брендонимами и уникальными наименованиями.

Анализ структурных типов английских двухкомпонентных терминов-словосочетаний в области ИИ показывает доминирование моделей атрибутивного уточнения базового термина, где определяющим компонентом выступают прилагательные, существительные (в т. ч. отглагольные), а также причастия в препозиции. Атрибут обычно обозначает функцию, свойства, структуру, способ действия или форму, тогда как головной компонент задает родовую категорию и включает единицу в соответствующую терминологическую парадигму.

В ходе исследования также установлено, что высокочастотные однословные термины (*model* (модель), *network* (сеть), *learning* (обучение), *inference* (инференс/вывод) и др.) функционируют двояко: как правая голова (head) многословного наименования и как левый препозитивный классификатор (modifier) в конструкциях N+N. В первом случае они задают родо-видовую категорию (*predictive*

model (предиктивная модель)), во втором – специфицируют и таксономизируют другое ядро (*model selection* (выбор модели)). Такая смена позиции повышает деривационную продуктивность лево-ветвящихся структур (*attention* → *attention head* (голова механизма внимания) / *attention map* (карта внимания)).

В рассматриваемом корпусе трехкомпонентные терминологические словосочетания составляют 21% (255 ТЕ) от общего числа терминов-словосочетаний; наиболее продуктивные модели приведены в Таблице 12.

Таблица 12 – Продуктивные модели трехкомпонентных терминов

Модель	Продуктивность	Примеры
A+N+N	74 ТЕ	<i>active learning strategy</i> (стратегия активного обучения)
N+N+N	53 ТЕ	<i>machine learning model</i> (модель машинного обучения)
A+A+N	48 ТЕ	<i>subordinate artificial intelligence</i> (подчиненный искусственный интеллект)
V-ed+N+N	15 ТЕ	<i>weighted sum model</i> (модель взвешенной суммы)
V-ed+A+N	10 ТЕ	<i>embodied cognitive science</i> (воплощенная когнитивная наука)
N+A+N	7 ТЕ	<i>consumer artificial intelligence</i> (потребительский искусственный интеллект)
Abb+N+N	6 ТЕ	<i>Adam optimization algorithm</i> (алгоритм оптимизации Adam)
N of A+N	6 ТЕ	<i>Internet of medical things, IoMT</i> (интернет медицинских вещей)
Part I+N+N	4 ТЕ	<i>vanishing gradient problem</i> (проблема исчезающего градиента)
A+V-ed+N	3 ТЕ	<i>generative pre-trained transformer, GPT</i> (генеративный предварительно обученный трансформер)

Всего выявлено 34 формальные модели образования трехкомпонентных терминов, 24 из которых охватывают 31 терминоединицу, что составляет ≈2% от общего корпуса. В силу низкой частотности эти модели не достигают принятого порога продуктивности и потому остаются за рамками детального анализа.

Как правило, трехкомпонентные терминологические словосочетания формируются на базе двухкомпонентных конструкций, обладающих более тесными структурно-семантическими связями. Двухкомпонентные единицы функционируют как самостоятельные термины и служат опорным ядром для дальнейшего наращивания – либо посредством добавления головы, либо за счет добавления еще одного

модификатора при сохранении правоголовной структуры: *error propagation model* (модель распространения ошибки) ← *error propagation* (распространение ошибки), *propagation model* (модель распространения).

В зависимости от положения исходного двучленного сочетания в трехчленной конструкции выделяются три базовые модели и одна смешанная: 1) $A > (B > C)$ – централизованная модель, при которой к уже существующему двухкомпонентному термину добавляется препозитивный уточняющий элемент: *deep learning method* (метод глубокого обучения) ← *learning method* (метод обучения) + *deep* (глубокий); 2) $(A, B) > C$ – параллельная модель, при которой два однородных атрибутивных компонента подчиняются одному головному элементу: *visual semantic embedding* (визуально-семантическое встраивание) ← *visual embedding* (визуальное встраивание), *semantic embedding* (семантическое встраивание); 3) $(A > B) > C$ – децентрализованная модель, при которой двухкомпонентное сочетание выполняет функцию сложного определяющего элемента: *natural language processing* (обработка естественного языка) ← *natural language* (естественный язык) + *processing* (обработка); 4) $\{A > (B)\} > C$, – смешанный вариант из двух пересекающихся сочетаний, в котором промежуточный компонент – головной в одном и определяющий в другом: *deep feature extraction* (глубокое извлечение признаков) ← *deep feature* (глубокий признак/признаковый уровень) и *feature extraction* (извлечение признаков).

Во всех вышеуказанных случаях третий компонент является головным и обозначает родовую категорию понятия, тогда как первые два компонента уточняют и конкретизируют его значение. Такая структурная организация демонстрирует высокую степень мотивированности и формализованности терминологических словосочетаний, что делает возможной их автоматическую идентификацию и классификацию с применением корпусных технологий и алгоритмов обработки естественного языка.

Многокомпонентные терминологические словосочетания с числом элементов более трех образуют в терминосистеме ИИ узкую ($\approx 7\%$) группу, характеризующуюся высокой синтаксической сложностью и максимальной степенью спецификации. В корпусе зафиксированы 63 четырехкомпонентных и 11 пятикомпонентных терминов; шести-, семи- и восьмикомпонентные конструкции единичны (по одной ТЕ на модель). Примеры включают следующие образования: *four-layer convolutional neural network* (четырёхслойная сверточная нейронная сеть) – четырехкомпонентный термин;

partially observable Markov decision process (частично наблюдаемый марковский процесс принятия решений) – пятикомпонентный термин; *recommender systems and intelligent decision support systems* (рекомендательные системы и интеллектуальные системы поддержки принятия решений) – шестикомпонентный термин; *extensive way of developing the hardware and software complex for artificial intelligence* (экстенсивный путь развития аппаратного и программного комплекса для искусственного интеллекта) – восьмикомпонентный термин.

Многочленные номинативные группы, как правило, вырастают из устойчивых двучленных баз (N+N, A+N) через последовательное наращивание слева (*four-layer > convolutional > neural > network*) либо координацию однородных модификаторов (*object > detection & instance > segmentation > model*). При этом сохраняется правоголовость: таксономический элемент расположен справа (*model, system, network, process, algorithm*), слева наслаиваются уточнения (прилагательные, причастия, числительные, лексикализованные блоки N+N, формульные модификаторы *human-in-the-loop, end-to-end* и др.), возможны предлоговые цепочки (N+prep+N) и координация без утраты семантической цельности. Практически такие сочетания часто выступают источниками акронимии (*partially observable Markov decision process* (частично наблюдаемый марковский процесс принятий решений) → *POMDP*) и обеспечивают точную номинацию сложных понятий в научной коммуникации. В современной терминосистеме ИИ синтаксическое и морфологическое терминообразование функционируют как единый цикл уплотнения и развертывания знаний (синтаксис → морфология → синтаксис). Исходной точкой обычно выступают правоголовые именные группы (A/N+N, A+N+N и др.), быстро создающие локально мотивированные номинации, например, *language model* (языковая модель) → *large language model* (большая языковая модель). Далее включаются механизмы морфологической компрессии: словосложение (*chat + bot* → *chatbot* (чатбот)), эллипсис (*error backpropagation algorithm* (алгоритм обратного распространения ошибки) → *error backpropagation* → *backprop*), сокращения (*large language model* (большая языковая модель) → *LLM*), что подтверждает тезис С. В. Гринева-Гриневица о компрессивной природе этих процессов [Гринева-Гриневиц 2008]. Полученные краткие формы быстро стандартизируются и становятся новыми «строительными блоками» синтаксиса (*LLM* → *LLM pre-training* (предобучение больших языковых моделей)).

Такой механизм номинации репрезентативен для значительной части терминологии ИИ и демонстрирует продуктивное взаимодействие синтаксической и морфологической деривации, обеспечивая несколько эффектов. Во-первых, он реализует когнитивную экономию: развернутые дефиниции и многокомпонентные номинации сворачиваются в компактные ярлыки (*convolutional neural network* (сверточная нейронная сеть) → *ConvNet*), снижая издержки научной коммуникации при высокой информационной плотности исследований [Cabré 1999; Temmerman 2000]. Во-вторых, поддерживается онтологическая связность, поскольку сокращенные формы сохраняют родо-видовую архитектуру исходных терминологических групп и позволяют без потерь переключаться между уровнями абстракции (например, *model* → *predictive model* → *ensemble predictive model* → *explainable ensemble models (EEM)*). В-третьих, обеспечивается технологическая масштабируемость, поскольку стандартизированные акронимы и бленды упрощают именование архитектур и ресурсов (*YOLOv8*, *roBERTa*) и облегчают последующее порождение дериватов (*roBERTa-large*, *GPT-4-style prompting*).

Для лексикографической и корпусной практики данный континуум задает операциональную схему: извлекать многословные термины по шаблонам, соотносить их с морфологическими вариантами и агрегировать частоты по «семействам» (например, *convolutional neural network* ~ *ConvNet* ~ *CNN*; *reinforcement learning policy gradient* ~ *policy gradient* ~ *PG*) с учетом орфографической вариативности.

Результаты структурного анализа позволили зафиксировать общий профиль терминосистемы ИИ в двух верхних классах – термины-слова и термины-словосочетания – и предложить авторскую классификацию, показанную на Рисунке 7. Классификация построена как взаимно исключающая (во избежание двойного счета) и отражает три уровня организации: 1) морфологический блок для терминов-слов, где производные (*simplex*) трактуются как нулевой деривационный тип (не как способ образования), а производные образуются посредством аффиксации, конверсии, усечения; 2) морфолого-синтаксические способы (словосложение, эллипсис, сокращения); 3) синтаксические модели многословных терминов (двух- и многокомпонентные конструкции). Такое разбиение обеспечивает сопоставимость количественных подсчетов и операционализирует различие между формой записи и типом словообразовательной модели, позволяя проследить переходы между классами

(от развернутых сочетаний к компрессионным вариантам и обратно в составе новых словосочетаний).

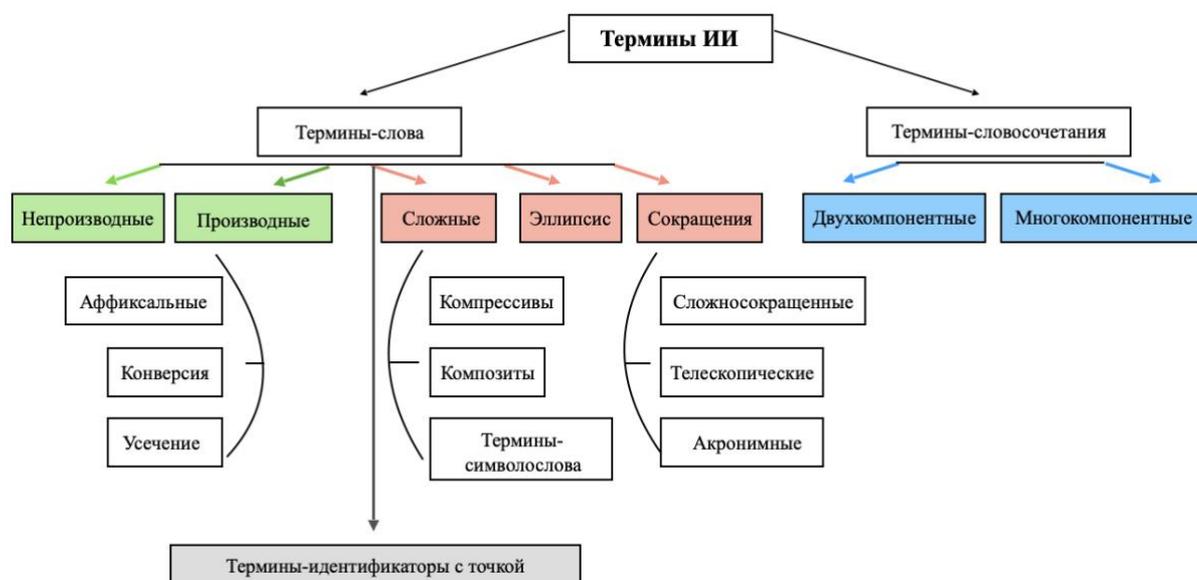


Рисунок 7 – Структурный профиль терминосистемы ИИ

Таким образом, синтаксическое терминообразование не противопоставляется морфологическому, а функционирует с ним как единый континуум «развертывание → компрессия → повторное развертывание». Двух- и многокомпонентные конструкции выступают системообразующими единицами терминологических полей, поскольку явно маркируют родо-видовые отношения и обеспечивают информативность при относительной компактности; их устойчивость поддерживается воспроизводимостью, семантической прозрачностью и коммуникативной экономией (в том числе через эллипсис и аббревиации в прозрачном контексте). Это согласуется с положением о преимущественном формировании терминов из ресурсов естественного языка посредством словообразовательных и синтаксических механизмов [Лейчик 2022] и соответствует трактовке терминологии как динамической, когнитивно мотивированной системы в отечественной и зарубежной традиции [Cabré 1999; Temmerman 2000; Суперанская 2012; Лейчик 2022].

Соотнесение морфосинтаксических результатов с метриками MI/LL и данными KWIC (см. главу 2) показывает, что стратификация терминосистемы определяется не длиной термина, а его операциональными признаками (частотность, дисперсия, MI/LL, KWIC-стабильность). К слою высокой стабильности относятся не только однословные

непроизводные единицы (например, *algorithm, data, model*), но и ряд производных терминов, ставших центральными концептами домена. Показательна лексема *learning*, функционирующая как опорный центр конфигураций типа *machine/deep/reinforcement learning* и демонстрирующая широкую дистрибуцию по поддоменам и предсказуемое контекстное окружение (KWIC). Среди многословных единиц устойчивый слой формируют прежде всего классические конструкции A+N и N+N, включая модели с герундиальной головой (*artificial intelligence* (искусственный интеллект), *neural network* (нейронная сеть), *deep learning* (глубокое обучение)), которые характеризуются высокой частотностью, семантической цельностью и коллокационной устойчивостью и служат конструктивными опорами дискурса. Зона текущих инноваций пополняется преимущественно более специализированными производными и композитными однословными единицами (*outpainting, hypernetwork*), для которых характерны узкая дистрибуция и менее устойчивые коллокационные профили. К периферии также относятся вариативные многословные модели, включая предлоговые расширения (*reinforcement learning from human feedback*), формульные шаблоны (*X-to-Y, X-as-a-Y, in-the-loop*) и компактные обозначения новейших концептов (*GPT, GAN, RAG, XAI*), демонстрирующие повышенную ключевость в анализируемом временном срезе. Таким образом, многословность и производность не тождественны периферийности.

3.3. Семантические особенности образования терминов ИИ

В данном параграфе семантическое терминообразование рассматривается как совокупность смысловых процессов, благодаря которым лексемы получают специализированные значения в терминосистеме ИИ. Анализ охватывает заимствование иноязычных лексем и терминоэлементов, терминологизацию общеупотребительных слов, транстерминологизацию из смежных дисциплин, а также сопутствующие механизмы адаптации терминов. На основе корпусных методов вводится операционализированная типология семантических путей пополнения терминологического инвентаря ИИ с оценкой их продуктивности и стабильности; уточняются когнитивные функции метафоры и метонимии в научном дискурсе.

3.3.1. Заимствование, привлечение и транстерминологизация

Семантическое терминообразование понимается как формирование новых терминов за счет переосмысления уже имеющихся лексем без изменения их формы; ключевыми механизмами выступают терминологизация общеупотребительных слов, расширение/сужение значения, метафорический и метонимический перенос, а также межсистемное заимствование терминов [Гринев-Гриневиц 2008; Лейчик 2022; Мусаева 2022]. Терминологическая традиция по-разному обозначает и разграничивает эти процессы. В работах С. В. Гринева-Гриневица «заимствование» трактуется как переход из общего языка в специальный, а «привлечение» – как перенос термина между терминологиями [Гринев 1993]; также используются термины «перенос» [Лотте 1961: 99], «вторичная терминологизация» [Даниленко 1977: 21], «трансфер» [Мальцева 2021: 150] и «транстерминологизация» [Мусаева 2022].

Показательно, что энциклопедический словарь В. А. Татарина «Общее терминоведение» фактически не разграничивает привлечение и транстерминологизацию, трактуя их как один и тот же процесс – перенос термина из одной терминосистемы в другую [Татарин 2006]. В данном словаре «привлеченные термины» определены как «термины из других отраслей, используемые в данной терминосистеме» [Татарин 2006: 146], а «транстерминологизация» – как «процесс перехода термина из одной отрасли в другую» [Там же: 309].

Учитывая данную вариативность, в настоящем исследовании принимается операциональная дифференциация по А. С. Мусаевой – 1) «заимствование» как межъязыковой переход (иноязычные термины, терминологические элементы и др.), 2) «привлечение» как перенос единиц из общего языка в терминологию, 3) «транстерминологизация» как горизонтальный переход терминов между специальными областями [Мусаева 2022]. Такой каркас позволяет последовательно описывать, как в дискурсе ИИ общеупотребительные и междисциплинарные единицы получают специализированные значения и закрепляются в качестве терминов.

Заимствование как универсальный механизм остается неотъемлемой частью эволюции терминологии, но проявляется типологически разнородно. В рамках контактной лингвистики выделяют: прямые заимствования (включая экзотизмы без перевода); калькирование – структурные и семантические кальки (дословный или моделирующий перевод); семантические заимствования (расширение/сдвиг значения

единицы языка-реципиента под влиянием языка-источника); частичные заимствования формантов (аффиксоидов/ комбинируемых основ), благодаря которым образуются гибридные термины [Loanwords 2009]. Степени интеграции при этом варьируются от поверхностной адаптации (орфографической/фонетической) до глубокой грамматической ассимиляции и деривационной продуктивности в языке-реципиенте (способности порождать регулярные словообразовательные гнезда) [Durkin 2014]. В классических обзорах именно рост деривационной активности служит признаком «вхождения» в систему языка-реципиента.

Для терминосистемы ИИ это различие иллюстрируется следующими примерами. Прямые исторические заимствования, такие как *algorithm* (алгоритм) (← средневек.-лат. *algorismus* от имени аль-Хорезми) и *robot* (робот) (чеш. *robota*, популяризовано К. Чапеком), демонстрируют полную интеграцию и продуктивные деривационные гнезда: *algorithmic, algorithmically; robotics, roboticist, cobot* [OED 2025; Online Etymology 2025]. Калькирование представлено терминами *principal components* (главные компоненты) ← нем. *Hauptkomponenten* в статистике и *steepest descent* (метод наискорейшего спуска) ← фр. *méthode de la plus forte descente* в оптимизации и машинном обучении [Andrei 2022]. Лексемы *eigenvalue* (собственное значение) и *eigenvector* (собственный вектор) являются частичными кальками с немецкого (← нем. *eigenwert/eigenvektor* с заимствованием форманта *eigen-* (собственный)). В англоязычных словарях фиксируется немецкое происхождение префиксоидной основы *eigen-* [Merriam-Webster 2025]. Частичные заимствования формантов и гибридизация проявляются в продуктивных неоклассических моделях научного английского (комбинируемые основы греко-латинского происхождения: *neuro-, cyber-, bio-, auto-, -morphic, -centric, -logy, -metry* и др., свободно сочетающиеся с английскими основами (*neuro-symbolic* (нейросимволический), *data-driven* (основанный на данных), *context-aware* (контекстно-осведомленный), *bio-inspired* (биоинспирированный))).

Греко-латинские форманты в современном англоязычном научно-техническом дискурсе функционируют как органичный, давно натурализованный компонент морфологической системы. Они свободно комбинируются с английскими основами и между собой, обеспечивая интернационально прозрачную семантику. В сфере ИИ этот «строительный набор» особенно активен при номинации новых понятий и архитектур,

поэтому такие элементы функционируют как нормативная часть морфосинтаксиса научного английского [Durkin 2014].

В ряде случаев заимствования остаются периферийными и культурно или дисциплинарно маркированными, выступая как интерпретационные метки и «сигналы происхождения». Так, *Ansatz* (анзац) – германизм, фиксируемый лексикографическими источниками как термин преимущественно научного употребления; при включении в англоязычные тексты он нередко сохраняет исходную графику, допускает конкуренцию вариантов множественного числа (*Ansätze / Ansatzes*) и демонстрирует ограниченную морфологическую интеграцию [OED 2025]. *Gestalt* (гештальт) описывается словарями как термин, закрепленный в психологической традиции и связанный с целостным восприятием/организацией образа [Online Etymology 2025]; в корпусном материале он реализуется в устойчивых словосочетаниях, отражающих соответствующий понятийный контекст (например, *Gestalt principles, Gestalt grouping*). *Umwelt* (умвельт) лексикографически определяется как ‘окружающий/перцептивный мир’ (в т. ч. в биолого-философской традиции) и сохраняет признаки заимствования при ограниченной деривационной продуктивности [OED 2025]; в корпусе данный термин употребляется для обозначения воспринимаемой среды агента (например, *agent’s Umwelt*). Латинизмы *a priori* (априори, ‘до опыта’), *a posteriori* (апостериори, ‘после опыта’), *ad hoc* (ад хок, ‘специально для данного случая’) функционируют главным образом как дискурсивные маркеры и, как правило, не формируют разветвленных терминологических семей [Online Etymology 2025]. Наконец, *Go* и *Shogi* закреплены в словарях как культурно специфические названия игр [OED 2025]; в академическом дискурсе они могут выступать в качестве доменов/задач для тестирования игровых ИИ, оставаясь при этом периферийными по отношению к терминологическому ядру (наблюдение по корпусу).

Перечисленные единицы показательны как диагностические признаки контакта (источника идей и каналов междисциплинарного влияния), однако комплекс корпусных индикаторов – ограниченная частота, слабая ключевость, узкий круг устойчивых коллокаций, отсутствие деривационных рядов – показывает, что они не определяют структуру тезауруса ИИ и сосуществуют с доминирующими внутриязыковыми механизмами номинации [Durkin 2014].

Привлечение в данном исследовании рассматривается как процесс терминологизации единиц общелитературного языка с сохранением формы слова, но с приобретением нового специализированного значения в профессиональном дискурсе, что отражает динамику терминологической лексики и ее системное варьирование в научном языке [Мусаева 2022]. В отечественной традиции его разграничивают с межъязыковым заимствованием и транстерминологизацией; в первом случае источник – иной язык, во втором – смежная терминосистема; при привлечении же задействуется внутриязыковой ресурс [Даниленко 1977; Гринев 1993; Термины и понятия 2024]. Наиболее показательными примерами привлеченных терминов ИИ считаем *agent* (действующее лицо) → автономная программная сущность/агент в ИИ-среде; *prompt* (подсказка/стимул) → инструкция/вход для LLM (ср. *prompt engineering*, *prompt tuning*); *hallucination* (галлюцинация) → генеративная ошибка/несоответствие действительности в выходах LLM; *guardrail* (ограждение) → политики и технические ограничения поведения модели. Во всех примерах семантический сдвиг опирается на дисциплинарную операционализацию (задаваемую дефинициями, метриками и типовыми конструкциями), а в ряде случаев сопровождается дополнительным метафорическим переосмыслением, что соответствует описанным механизмам семантического терминообразования в терминоведении [Термины и понятия 2024].

Транстерминологизация понимается как «горизонтальный» перенос терминологической единицы из одной терминосистемы в другую с последующей семантической адаптацией в новом профессиональном контексте; в отличие от межъязыкового заимствования источник здесь – иная предметная область того же языка, а не другой язык [Гринев-Гриневиц 2008; Мусаева 2022]. Для терминологии ИИ типичны траектории из математики и статистики (*kernel* (ядро) → *kernel trick* (ядерный трюк), *kernel method* (ядерный метод)), физики (*annealing* (отжиг) → *simulated annealing* (симулированный отжиг)), когнитивных наук и лингвистики (*attention* (внимание) → *attention mechanism* (механизм внимания), *self-attention* (самовнимание)), биологии/эволюционной теории (*population* (популяция)), а также экономики, теории игр и управления (*payoff* (выигрыш)).

В ходе формирования терминосистемы ИИ процессы привлечения и транстерминологизации находятся в состоянии диалектического единства, формируя сложную многоуровневую иерархию терминологических пластов, где трансфер

смыслов происходит нелинейно. Показателен термин *memory* (память), который демонстрирует гибридный путь развития. Из общелитературного языка он сначала получает научную фиксацию в когнитивных науках (биологическая функция хранения и извлечения информации), а затем транстерминологизируется в ИИ, где переосмысливается как формальный механизм поддержания и обновления состояний в архитектурах нейронных сетей (например, *memory cell* (ячейка памяти) в LSTM), сохраняя общеязыковое значение в роли семиотического «якоря». Аналогичен путь термина *neuron* (нейрон), который из обозначения нервной клетки трансформируется в абстрактный вычислительный элемент (искусственный нейрон) с атрибутами *activation* (активация), *weight* (вес/«синапс»), закрепляясь в номинациях типа *neural network* (нейронная сеть), *convolutional neural network* (сверточная нейронная сеть); сходные сдвиги наблюдаются у слов *reward* (вознаграждение → скалярный сигнал в RL) и *policy* (политика → правило выбора действия).

Кроме того, наблюдается миграция терминов между пластами. Единица может начать свой путь как «чистая» терминологизация, затем, будучи формализована в ИИ, транстерминологизироваться в смежные области, и наоборот. Так, термин *reinforcement learning* (обучение с подкреплением), зародившись в психологии обучения, в ИИ получил строгую математическую формулировку (марковские процессы принятия решений, функция ценности), а затем вновь был адаптирован когнитивной нейронаукой для описания дофаминергических механизмов оценки вознаграждения.

Возможны и гибридные номинации, в которых компоненты разного происхождения сходятся в одном термине. Лексема *convolutional neural network* (сверточная нейронная сеть) объединяет *convolutional* (транстерминологизация из математики сигнальной обработки), *neural* (транстерминологизация из нейробиологии через кибернетику) и *network* (терминологизация общеязыкового «сеть», переосмысленного как формальный граф/архитектура).

В рамках динамики терминологической эволюции искусственного интеллекта наблюдается феномен семантической дивергенции, при которой термины, попав в дисциплинарный контекст, продолжают развиваться, приобретая специализированные значения, дистанцированные от исходной семантики. Характерным примером служит термин *attention* (внимание), изначально представляющий собой прямую

терминологизацию общелитературной лексемы, обозначающей когнитивный акт концентрации на объекте. В терминосистеме ИИ данный термин кристаллизовался в качестве ключевого архитектурного элемента – *attention mechanism* (механизм внимания), составившего основу архитектуры трансформеров. В дальнейшем он породил автономную терминологическую подсистему, включающую такие производные понятия, как *self-attention* (самовнимание) или *multi-head attention* (многоголовое внимание), семантика которых целиком определяется операциональным контекстом машинного обучения и не может быть выведена из бытового значения исходной единицы.

Принимая во внимание, что процессы привлечения и транстерминологизации не являются изолированными и формируют динамичную сеть, в которой общелитературный язык поставляет интуитивно мотивированные лексемы, смежные дисциплины – готовые формализованные концепции, а терминосистема ИИ выступает интегратором, перерабатывающим и адаптирующим эти ресурсы в новые специализированные понятия, в настоящем исследовании не вводится жесткого разграничения единиц по источнику (общий язык vs. смежные науки) и не проводится их отдельный количественный учет; аналитический фокус смещен на стратегии семантической адаптации. К ключевым механизмам относятся: 1) расширение значения – добавление специализированного, но по объему более широкого применения без вытеснения общезыкового смысла (например, *agent* (действующее лицо) → *agent* (автономная программная сущность/агент в ИИ среде); *prompt* (подсказка) → *prompt* (инструкция модели)) [L’Homme 2020; ISO 704:2022]; 2) специализация (сужение) – закрепление узкого операционального содержания (например, *attention* (внимание) → *self-attention* (механизм взвешивания контекстов), *reward* (вознаграждение) → *reward* (скалярная обратная связь в RL)) [Cabré 1999; Temmerman 2000]; 3) метафоризация – перенос по сходству образа/структуры (например, *hallucination* (галлюцинация) → *hallucination* (генеративная ошибка LLM), *guardrail* (ограждение) → *guardrail* (политики/механизмы безопасности)) [Lakoff, Johnson 1980; A Cognitive Linguistics... 2012]; 4) метонимический перенос – замещение по смежности (например, «процедура → результат»: *loss* (функция потерь) → *loss* (ее численное значение); *minimize loss* = уменьшать величину потерь; *define the loss* = задавать функцию потерь) [Barcelona 2000; L’Homme 2020]; 5) эпонимизация –

номинация от имени ученого / автора / топонима / мифонима / брендони́ма: *Bellman equation* (уравнение Беллмана); *Markov chain* (марковская цепь); *Chernoff bound* (оценка Чернова) и др. [Новинская 2013; Какзанова 2019].

Таким образом, привлечение, заимствование и транстерминологизация образуют единую динамику семантического терминообразования в ИИ, где расширение и сужение выступают базовыми механизмами переосмысления формы без ее замены. Именно их взаимодействие системно порождает полисемию (наслаивание регистров и субдоменов), конфигурации синонимии (дубликаты/варианты, условные синонимы), а также запускает метафоризацию и метонимизацию как регулярные стратегии адаптации смысла; эпонимизация добавляет идентификационный ресурс и ускоряет нормирование. Эти процессы эмпирически наблюдаемы и поддаются операционализации в корпусе через ключевость, ассоциативные меры, энтропию профилей, дисперсию по субкорпусам и дефиниционные/числовые паттерны.

3.3.2. Метонимия и метафоризация

В настоящем исследовании категория фрейма служит методологической основой интерпретации таких лингвистических явлений, как метонимия и метафора. Под фреймом понимается когнитивная структура, представляющая собой организованное знание о типичной ситуации, задающая ожидания относительно ее участников, свойств и отношений [Fillmore 1976]. Так, М. Мински еще в 1974 году предложил рассматривать фреймы как структуры данных, хранящие стереотипные ситуации, где верхний уровень состоит из фиксированных элементов, а нижние уровни содержат слоты для заполнения конкретной информацией [Minsky 1975]. Позднее Ч. Филлмор перенес идею фреймов в лингвистику и разработал фреймовую семантику, согласно которой значение лексической единицы интерпретируется через активируемый ею фрейм – прототипическую сцену / сценарий, релевантный данному слову [Fillmore 1976].

Концепция фрейма получила широкое развитие в трудах как зарубежных, так и отечественных исследователей. В частности, И. Гофман рассматривал фрейм как схему интерпретации событий, позволяющую индивиду идентифицировать, воспринимать, обозначать и классифицировать бесчисленное множество происходящих вокруг него событий. При этом автор выделял первичные и преобразованные фреймы, главное

отличие между которыми заключается в том, что первичный фрейм представляет собой естественный способ понимания ситуации, тогда как преобразованный фрейм – это ситуация, которая была каким-то образом изменена, притворна или является копией другой. Следует отметить, что фрейм, по И. Гофману, это не когнитивная, а интерактивная структура, организующая социальный опыт индивида и позволяющая ему участвовать в сложных, но упорядоченных взаимодействиях [Гофман 2004].

Согласно В. В. Красных, фреймы характеризуются иерархической (атомарной) организацией, преимущественно пирамидальной или древовидной формы. Они состоят из вершины – имени темы стереотипной ситуации (например, ПОКУПКА) и нижних уровней – слотов (ПРОДАВЕЦ, ПОКУПАТЕЛЬ, ТОВАР, ДЕНЬГИ, МЕСТО) и атомов (ХЛЕБ, САХАР, МОЛОКО, КНИГИ), заполненных вариативными составляющими [Красных 2004: 27].

Н. Н. Болдырев определяет фрейм как фундаментальную единицу организации знаний в сознании человека, как «пакет информации, знание о стереотипной, часто повторяющейся ситуации, хранящееся в долговременной памяти в виде целостной структуры, сложного концепта, и используемое для категоризации нового знания и передачи накопленного ментального опыта» [Болдырев 2021: 54]. Следовательно, фрейм выполняет функцию категоризации нового знания и передачи накопленного ментального опыта, упрощения процесса восприятия, так как для анализа каждой новой ситуации человек использует уже имеющиеся данные, и предугадывания развития ситуации, закрепленной в том или ином фрейме. При этом исследователь четко разграничивает фрейм и другие когнитивные структуры – сценарий и концепт. По сравнению с концептом фрейм является более крупной, комплексной и структурированной единицей знания, а по сравнению со сценарием он более статичен, включает знания об участниках и атрибутах ситуации [Болдырев 2000].

Введение в категориальный аппарат науки о языке фреймовой модели вместе с родственными понятиями домена, идеализированной когнитивной модели ICM, схемы и сценария быстро показало высокую объяснительную силу для лингвистики [Fillmore 1976; Lakoff 1987]. Фреймовая теория стала основой фреймово-семантических исследований Ч. Филлмора, стимулировала прикладные лексикографические проекты, включая Berkeley FrameNet, начатый в 1998 году для английского языка [Baker 1998], и сформировала когнитивно-ориентированное направление в терминоведении [Cabr e

1999; Temmerman 2000; A Cognitive Linguistics 2012; L'Homme 2020]. В этой парадигме фреймы применяются для моделирования специализированных понятийных систем в русле Frame-Based Terminology у П. Фабер [A Cognitive Linguistics 2012] и для лексикографического описания терминов на основе фреймовой семантики в работах М.-К. Л'Ом [L'Homme 2020]. Такой подход позволяет учитывать многомерную природу терминов и извлекать релевантную семантическую информацию из корпусов.

Одним из явлений, которое эффективно объясняется через фреймы, является **метонимия** – перенос значения по смежности. В когнитивной лингвистике метонимия рассматривается не просто как риторический прием, а как когнитивный механизм, основанный на внутримоментных ассоциациях. Иными словами, метонимический сдвиг – это отображение типа «носитель → цель» внутри единой концептуальной области знания [Barcelona 2000]. В рамках фреймового подхода метонимию целесообразно описывать как внутримоментный доступ к одной сущности через другую в пределах одного и того же фрейма или ИСМ [Lakoff 1987]. Такое понимание соотносится с классическим структурным принципом смежности (по Р. Якобсону) и противопоставляется метафоре, которая основана на междоментном переносе и сопоставлении разных областей знаний [Jakobson 1960; Lakoff, Johnson 1980]. Метонимия же сохраняет общий домен и лишь перераспределяет фокус внимания внутри него [Gibbs 1999]. Ключевую роль при этом играет салиентность – степень выдающегося, когнитивно заметного статуса того или иного элемента ситуации. Именно салиентные (наиболее релевантные в контексте) слоты фрейма могут выступать «точками доступа», через которые вызывается в целом вся фреймовая структура [Lakoff, Johnson 1980; Barcelona 2000].

В результате метонимические связи образуют системную сеть взаимоотношений внутри одного фрейма. Такая сеть имеет радиальную организацию, согласно которой значения «разрастаются» от центрального, прототипического узла к периферийным элементам по принципу «прототип и вариации». В центре располагаются наиболее типичные участники ситуации, а на периферии – менее заметные детали; при этом любой элемент фрейма потенциально может стать актуальным, если контекст делает его значимым [Lakoff 1987; Peirsman, Geeraerts 2006]. Метонимический сдвиг запускается в тот момент, когда некоторый компонент фрейма приобретает статус наиболее важного для коммуникативной цели – и тогда этот компонент начинает

обозначать собой всю ситуацию в целом (при сохранении исходного концептуального фрейма). Типичные когнитивные отношения, лежащие в основе метонимии, включают связи «часть – целое», «причина – следствие», «средство – действие», «контейнер – содержимое» и др. Например, название части может употребляться для обозначения целого (и наоборот); средство или инструмент – для обозначения действия, которое с его помощью совершается; вместилище – для содержимого и т.п.

Важно, что чаще пусковыми точками метонимии служат именно прототипные, центральные элементы фрейма (они привычнее и значимее для носителей языка), тогда как периферийные элементы привлекаются реже. Такой прототипно-радиальный принцип организации метонимических переносов был подробно описан в работах по когнитивной лингвистике [Peirsman, Geeraerts 2006]. Существенный вклад в осмысление метонимии как внутримоделного феномена внесли исследования в русле когнитивной семантики и прагматики [Barcelona 2000], где метонимия трактуется как основанный на смежности механизм в пределах единой когнитивной модели, порождающий регулярные выводы и сдвиги значения. В этой перспективе фреймовый подход позволяет описывать метонимию как систему внутримоделных перераспределений салиентности, отражающих организацию знания, коммуникативные приоритеты и механизмы категоризации в языке.

В научном дискурсе требования точности и одновременной экономности выражения приводят к тому, что термины нередко употребляются метонимически – когда вместо полного описательного названия явления называется лишь его наиболее значимый компонент. Особенно это характерно для быстроразвивающихся областей, таких как искусственный интеллект (ИИ), где набор общепринятых ключевых понятий (моделей, задач, метрик, данных и пр.) выступает когнитивным каркасом коммуникации.

Для выявления метонимических употреблений терминов в терминосистеме ИИ применялись методы корпусной лингвистики с опорой на фреймовый подход. Автоматизированный поиск по ключевым словам в сочетании с конкордансным просмотром в режиме KWIC в программе AntConc и расчетом ассоциативных метрик Mutual Information и Log-Likelihood позволил сформировать список кандидатов на метонимию и задать дальнейшую верификацию. Поисковые запросы строились на базе типовых конструкций, где ожидается внутримоделный перенос значения, например,

improves X by N или *achieves N on X*, где X обозначает датасет или бенчмарк и тем самым репрезентирует не объект данных, а достигнутый на нем показатель, а также *optimize X*, где X как название метрики интерпретируется как ее числовое значение либо процедура оптимизации. Инвентарь лексико-синтаксических шаблонов дополнялся перечнем сопряженных терминов, в который вошли датасеты *ImageNet*, *MNIST*, *COCO*, метрики *BLEU*, *F1-score*, *accuracy*, *loss*, а также модели *BERT*, *GPT*, *ResNet*. Для каждого шаблона выполнялась выборка контекстов в корпусе с последующей проверкой метонимического статуса, при этом конкордансные окна предоставляли ближайшее окружение ключевой единицы X и позволяли установить актуализируемый смысл. Так, в примере *Our model achieves 89.7 on ImageNet* X трактуется как показатель точности на наборе ImageNet, а в примере *We optimize BLEU* наименование метрики выступает обозначением целевого показателя качества перевода, а не методики подсчета.

Количественная оценка тенденций осуществлялась коллокационным анализом с вычислением показателей связности, в том числе числа вхождений конструкций типа *achieves* рядом с названиями датасетов и *optimize* рядом с названиями метрик, а также дополнительно рассматривалось распределение выявленных метонимий по секциям IMRaD с сопоставлением частот в разделах статей “Abstract”, “Introduction”, “Methods”, “Results” и “Discussion”. Наконец, результаты соотносились с терминоведческими критериями терминологизации, то есть закрепления выражений в специализированном значении, при этом рост доли устойчивых метонимических употреблений и стабилизация формул рассматривались как индикаторы формирования новых сокращенных наименований: так, если сочетание вида *X score* повсеместно интерпретируется как результат на X, то X получает самостоятельную терминологическую интерпретацию, а соответствующие случаи подлежат отражению в глоссариях и словарных статьях.

После идентификации конкретных случаев метонимических употреблений терминов была проведена их классификация по типам фреймовых сдвигов. В качестве общей когнитивной опоры принят фрейм научного эксперимента по обучению и оценке модели. В его упрощенной конфигурации выделяются слоты Model (модель/алгоритм), Dataset/Benchmark (данные или бенчмарк-задача), Metric (метрика качества или критерий), Result (результат/показатель качества), Task (задача или цель

эксперимента), Output (выходные данные/продукт работы модели) и др. На этой основе установлено семь регулярных типов внутримоделных переносов в терминологии ИИ, каждый из которых соотнесен с парой фреймовых слотов (например, часть–целое, контейнер–содержимое, инструмент–действие, агент–продукт и др.):

1) benchmark/датасет → оценка/результат (*our model improves **GLUE** by 2.1 points* = GLUE-score; *achieves 89.7 top-1 on **ImageNet*** = тестовая точность на ImageNet; *new SOTA on **AI benchmark*** = «балл бенчмарка»); здесь реализуется связь типа «контейнер → содержимое», поскольку имя набора данных (контейнер) репрезентирует содержащиеся в нем показатели;

2) метрика → ее численное значение/процедура (*we optimize **BLEU*** = «процедура/значение BLEU», *lower **cross-entropy*** = «меньшее значение потери»);

3) тип модели → конкретный экземпляр/вариант (*we fine-tune **BERT** for NER* = «BERT-based модель»; *evaluate **RoBERTa** on MNL1* = «конкретная конфигурация»);

4) система → ее продукт/выход (*detecting bias in **ChatGPT*** = «в выходах ChatGPT»; *filter **DALL-E*** = «сгенерированные изображения»);

5) задача/режим → стандартный результат/метрика (*better **ASR** on LibriSpeech* = «более низкий WER/CER»; *state-of-the-art **NER*** = «F1 на датасете»);

6) метод/процедура → артефакт/параметр (*apply **LoRA*** = «LoRA-адаптеры/веса»; *aggressive **pruning*** = «прореженная сеть»);

7) организация/бренд → система/результаты (*evaluate **OpenAI** on...* = «модель/API OpenAI» – встречается в текстах неформально).

В каждом случае метонимия актуализирует наиболее релевантный для коммуникативной цели салиентный слот. В сообщении о новизне фокус смещается на результат эксперимента, который задается через упоминание бенчмарка или метрики как стандартной меры успеха. Имя набора данных, например, *ImageNet*, выступает точкой доступа к информации о достигнутой точности и активирует соревновательный подфрейм; выражения вида *achieves 89.7 on **ImageNet*** интерпретируются как достигнутая точность 89,7% по принятому протоколу ImageNet, даже если числовой показатель не назван явно. В этом случае действует перенос типа «контейнер → содержимое», поскольку название теста репрезентирует представленные в отчете результаты. Такая компрессия поддерживает высокую информационную плотность и

опирается на общие для сообщества знания о задачах, протоколах и шкалах, поэтому интерпретация не страдает и для профессиональной аудитории остается однозначной.

Модель «метрика → ее численное значение или процедура» трактуется как перенос по схеме «инструмент и результат». Имя метрики не называет объект измерения, а служит синонимом получаемого показателя либо самой процедуры подсчета. В корпусе это чтение поддерживают глаголы оптимизации и изменения величины, например, *optimize*, *maximize*, *increase*, *decrease*, *lower*, *reduce*, а также соседство с числами и единицами измерения, как в *we optimize BLEU*, *lower cross-entropy*. Буквальные употребления (без метонимического переноса) встречаются в определениях и теоретических описаниях, где метрика вводится или выводится формально. Практический критерий различения опирается на ближайший контекст; при наличии возле наименования метрики количественного индикатора или глагола изменения величины с высокой вероятностью активируется метонимия, и метрика интерпретируется как числовой результат либо как процедура его получения.

В наименованиях моделей (*BERT*, *GPT*) метонимия проявляется как перенос «тип → экземпляр». Архитектура BERT представляет семейство моделей, однако формула *fine-tune BERT for our task* отсылает к конкретной конфигурации, дообученной под задачу. Возникает семантическое сужение, при котором родовое имя употребляется для единичного объекта. Такой перенос опирается на прототипную организацию «модельного пространства», в рамках которой тип занимает центральную позицию, а индивидуальные реализации рассматриваются как вариативные воплощения, достаточные для обсуждения результатов. Онтологически это соответствует отношению «класс – член класса» и, при соответствующем прочтении, со связью «целое – часть», если архитектура понимается как целое, а ее конкретная настройка – как часть.

Перенос «система → продукт/выход» устойчиво проявляется в дискурсе генеративных моделей. Фраза *bias in ChatGPT* читается как смещение в сгенерированных ChatGPT ответах, то есть имя системы метонимически обозначает ее выход. Такой механизм согласуется со схемами «автор – произведение» (например, когда говорят «читаю Толстого», имея в виду книги Л. Н. Толстого) и «орудие – результат» и регулярно поддерживается контекстными сигналами (*detect*, *filter*, *audit*, *analyze*) в окружении названий систем. Показательны конструкции, где имя системы

прямо номинирует результат ее работы, например, такие как *audit ChatGPT outputs*, *filter DALL·E generations*, *analyze Google results for X*, что снимает двусмысленность и фиксирует перенос. Корректное прочтение опирается на разделяемые в сообществе фреймовые знания о работе модели и на выделение слота *Output*, поэтому для специалиста такое толкование не требует дополнительных пояснений.

Далее фиксируется модель «задача или режим → стандартный результат или метрика». Название задачи репрезентирует ее типовую шкалу оценки, поэтому *better ASR on LibriSpeech* подразумевает более низкий *WER* или *CER*, а *state-of-the-art NER* – высокий *F1* на профильном датасете. Диагностические сигналы в KWIC-окнах включают прилагательные *better*, *stronger*, *state-of-the-art*, глаголы сравнения или улучшения, а также соседство с именами датасетов. Буквальное чтение возникает при описании модулей или пайплайнов (*ASR pipeline*, *NER module*).

Метонимический сдвиг «метод или процедура → артефакт, параметр или полученное состояние» проявляется там, где наименование техники репрезентирует производимый ею объект либо установленную настройку. В описаниях экспериментов формулы *apply LoRA* обычно предполагает добавление LoRA-адаптеров и новых весов, *aggressive pruning* соотносится с получением прореженной сети и заданным уровнем разреженности. Корпусные индикаторы включают глаголы материализующего действия *apply*, *add*, *use*, *prune* и ближайшие имена артефактов *adapters*, *weights*, *masks*, *tokens* либо числовые маркеры гиперпараметров. В случаях, когда метод определяется, классифицируется или выводится теоретически, как в *LoRA is a low-rank adaptation method*, метонимия не активируется и чтение остается буквальным.

Наконец, в текстах встречается перенос «организация или бренд → система или ее результаты», характерный для оценочных и сравнительных описаний. Имя компании метонимически обозначает модель или API и их выходы, поэтому *evaluate OpenAI on ...* интерпретируется как оценка конкретной системы OpenAI и ее результатов. В корпусе сдвиг проявляется в сравнительных и бенчмарк-контекстах, где отсутствует явное название модели, встречаются глаголы *evaluate*, *compare*, *benchmark* и рядом стоят задачи или датасеты. Уточнения о релизах, финансировании и персоналиях читаются буквально. Для формального стиля рекомендуется указывать точное имя системы и версию (например, *GPT-4o*, *Gemini 1.5*, *Llama-3*).

Одновременно выявлены пограничные случаи, требующие контекстуальной интерпретации. Термины *entropy* (энтропия), *temperature* (температура), *momentum* (момент/импульс), *diffusion* (диффузия) чаще всего употребляются строго в математико-физическом значении, но при определенном окружении могут приобретать метафорический оттенок. Лексемы *memory* (память), *attention* (механизм внимания), *agent* (агент), *autonomy* (автономность) в методологических разделах уже терминологизированы, однако метонимия проявляется, когда они обозначают результат или конкретный экземпляр (например, *this attention improves BLEU* – «этот механизм внимания улучшает BLEU»). Аналогично, *pipeline* (конвейер/поток обработки), *filter* (фильтр), *window* (окно), *gate* (вентиль/затвор) являются конвенциональными метафорами, но демонстрируют метонимические сдвиги, когда форма начинает обозначать процесс или результат (*through the pipeline in 2 ms* = «все стадии обработки за 2 мс»). Развернутые иллюстрации, регулярные шаблоны и дополнительные примеры метонимических сдвигов приведены в Приложении № 3.1.

Корпусный анализ подтверждает, что указанные метонимические модели не единичны, а регулярно воспроизводятся в научном ИИ-дискурсе. В среднем в каждой статье корпуса обнаружено от 3 до 5 случаев подобных переносов. Наиболее частотными были шаблоны, связанные с бенчмарками и метриками. Например, конструкции *achieve N on X* присутствовали приблизительно в 40% статей, а *improve X by Y* (где X – показатель на бенчмарке) – в 25%. Упоминания оптимизации метрик (*optimize [metric]* или *reduce [loss]* и т. п.) встретились примерно в 30% текстов. Использование названий моделей вместо конкретных экземпляров (особенно для популярных моделей типа BERT, GPT, ResNet) наблюдалось реже (около 15% статей), что может быть связано с тем, что в методологических разделах обычно все же указывается конкретная конфигурация модели. Метонимия типа «система → ее выход» (как в случае ChatGPT) пока относительно новая и встречалась в ≈10% документов, преимущественно в последние годы, что коррелирует с появлением крупных генеративных моделей в 2022–2023 годах.

Распределение метонимических употреблений по разделам IMRaD демонстрирует устойчивую функционально-жанровую асимметрию. В разделах статей “Abstract” и “Introduction” их доля максимальна: требования к краткости и эксплицитации вклада стимулируют формулы типа *dataset/metric → result* и *model →*

result (например, *the model achieves SOTA on ImageNet*), обеспечивающие высокую информативную плотность. В разделах “Results” и “Discussion” метонимия также используется, но чаще сопровождается контекстуализацией (указанием условий эксперимента, вариативности метрик), что смещает фокус к интерпретации и сопоставлению результатов. В разделе “Methods” преобладает «буквальный» регистр перечисления компонентов (модели, датасеты, метрики) и процедур, где метонимические сокращения практически не мотивированы. Тем самым наибольшая концентрация метонимий наблюдается в тех частях статьи, которые институционально ориентированы на презентацию новизны и ключевого результата.

Помимо секционной структуры ряд иных жанрово-коммуникативных условий также благоприятствует использованию метонимических сдвигов. Во-первых, устойчивость и «общезначимость» определенных слотов фрейма (например, *dataset, metric, model, output*) делают их вероятными кандидатами на роль такой «точки доступа» к целому эксперименту. Во-вторых, высокая степень конвенционализации домена – наличие общепринятых бенчмарков, метрик, процедур – в сочетании с целевой установкой дискурса (оперативно обозначить научную новизну, сравнить результаты, вывести ключевой показатель в тезис) – усиливает тенденцию к метонимическим сокращениям. В-третьих, немаловажны редакционно-институциональные факторы: жесткие ограничения по объему, требования рецензентов и стилистические гайды поощряют использование экономичных формулировок. В-четвертых, свойство английского языка легко образовывать сжатые номинальные конструкции (композиции типа $N+N$, акронимы) технически облегчает использование подобных сокращений в научных текстах. Наконец, существенным фактором остается риск двусмысленности и профиль аудитории. Чем выше уровень общих знаний сообщества о бенчмарках и метриках, тем безопаснее использование метонимических переносов; в междисциплинарных и прикладных публикациях метонимия проявляется более сдержанно.

Представленный анализ показывает, что фреймовая перспектива дает целостное объяснение метонимии в терминосистеме ИИ и описывает ее как регулярный механизм внутридоменного профилирования. Выявленные переносы укладываются в общую когнитивную модель научно-технического дискурса и подтверждают тезис о метонимии как об экономии выражения и средстве выделения релевантных признаков

внутри одного фрейма [Rosch 1978]. Упоминание салиентного элемента – бенчмарка, метрики, архитектуры или системного выхода – активирует у читателя заранее разделяемые знания о задаче, протоколе и шкале оценки, поэтому сжатые формулы остаются однозначными для профессиональной аудитории.

С точки зрения теории фреймов наблюдается конкуренция слотов за экспликацию. Выбор актуализируемого элемента не случаен и определяется как когнитивными факторами – прототипностью и частотностью в сообществе, так и прагматическими условиями – жанром, коммуникативной задачей и редакционными ограничениями [Fillmore 1976]. В исследовательских текстах чаще выдвигаются Dataset Benchmark и Metric, когда требуется быстро показать превосходство на общепринятом тесте, тогда как Model Architecture профилируется при акценте на новизне подхода. Такой баланс поддерживает высокую информативную плотность без потери интерпретируемости и согласуется с принципом релевантности в научной коммуникации [Sperber, Wilson 1986].

Практическая значимость проведенного анализа метонимии в терминосистеме ИИ имеет два взаимодополняющих измерения. Во-первых, терминографическая фиксация регулярных переносов улучшает нормализацию и перевод терминов. Например, словарная статья для термина *ImageNet* в контексте ИИ должна отражать не только *ImageNet* – база данных изображений, но и переносное употребление: *ImageNet* – бенчмарк-тест, результат на данном тесте (например, *achieve X on ImageNet* – добиться точности X). Во-вторых, при автоматическом извлечении знаний из текстов (например, для построения онтологий или баз знаний) понимание метонимии обеспечивает корректную типизацию отношений и предотвращает онтологические ошибки. Например, вместо формально возможного, но неверного *model – achieves – ImageNet* система должна выводить *model – has accuracy – 89.7 on ImageNet*. Последовательная аннотация корпусов с пометой метонимических сдвигов снижает неоднозначность терминов и повышает точность семантического анализа, позволяет правильно классифицировать отношения.

Вышесказанное также может свидетельствовать о формировании своеобразного профессионального сленга внутри научного сообщества ИИ. Метонимия, будучи по природе экономией выражения, служит маркером принадлежности. Иными словами, тем, «кто в теме», понятны сокращения и переносы, тогда как постороннему читателю

они могут показаться неочевидными. Перспективным представляется сопоставительный анализ подобных моделей в смежных доменах – биомедицина, финтех и другие области, что позволит оценить степень универсальности механизма и специфику его терминологической реализации.

В целом фреймовый анализ демонстрирует высокую объяснительную силу и объединяет лексическую семантику, когнитивную теорию и практику корпусной терминографии. Метонимия проявляется не как факультативный стилистический прием, а как устойчивый способ категоризации и передачи знаний в научном дискурсе ИИ.

Метафора, являясь одним из наиболее сложных и коннотативно насыщенных языковых феноменов, на протяжении десятилетий остается в центре междисциплинарного внимания, объединяя перспективы лингвистики, философии, литературоведения и шире – гуманитарного знания в целом [The Cambridge Handbook 2008]. Исторически метафора рассматривается как один из базовых когнитивных механизмов, обеспечивающих концептуальное соотнесение конкретного и абстрактного и участвующих в формировании символических систем и смыслообразования.

В современной науке метафора рассматривается не просто как лингвистический прием, а как когнитивный инструмент, формирующий восприятие, мышление и культурное понимание. В работе «Метафоры, которыми мы живем» Дж. Лакофф и М. Джонсон совершили революцию в теории метафоры, утверждая, что метафоры носят не только лингвистический, но и концептуальный характер. Они показали, что метафоры формируют наше восприятие, мышление и действия, структурируя абстрактные понятия в терминах конкретного опыта: «Суть метафоры заключается в понимании и переживании одного вида вещей или опыта в терминах другого» [Lakoff, Johnson 1980: 5]. Таким образом, метафора трактуется как когнитивная операция, направленная на трансформацию говорящим языковой репрезентации действительности и ее вербального выражения.

Концептуальная теория метафоры Дж. Лакоффа и М. Джонсона продолжает эволюционировать, порождая разнообразные исследовательские направления. Современные подходы не отменяют, а развивают исходную теорию, образуя многоуровневую исследовательскую парадигму. Нейрокогнитивные и телесно-

обусловленные модели [Feldman 2006; Barsalou 2008] раскрывают физиологические основы метафорического проецирования. Теория первичных метафор [Grady 1997] вводит уровень элементарных соответствий, концепция смешения ментальных пространств [Fauconnier, Turner 2002] описывает динамическую сборку значений, в то время как исследования культурной вариативности [Kövecses 2005] ограничивают претензии теории на универсальность. В итоге формируется интегративная картина, где метафора понимается как многоуровневый процесс с нейронной и телесной опорой, базовыми схемами, динамическим конструированием смысла и культурно-прагматической модуляцией. Такая модель продуктивно соотносится с корпусным анализом и аннотацией метафор в научном дискурсе ИИ.

Кроме того, в научной литературе накоплено значительное число исследований, посвященных феномену метафорического мышления [The Cambridge Handbook 2008; Guo 2023]. Как отмечает К. Казо, тема метафорического познания стала центральной в континентальной философии, начиная с И. Канта. Философы рассуждали о том, каким образом разум конструирует реальность на основе многообразных чувственных впечатлений. Наши когнитивные способности формируют зрительный образ мира посредством визуальных стимулов, поступающих через органы зрения, и слуховой – через слуховые раздражители. К. Казо ссылается на Ницше, который утверждает, что истина «происходит от фундаментально метафорической природы формирования понятий, представляющей собой ряд творческих скачков: от нервного раздражения – к зрительному образу (первая метафора), от образа – к звуку как обозначающему (вторая метафора)» [Cazeaux 2007: 104].

Цифровая эпоха изменила не только способы коммуникации, но и способы концептуализации реальности – и метафора по-прежнему остается ключевым когнитивным инструментом, соединяющим абстрактную сферу технологий с человеческим опытом. Такие термины, как *cloud computing* (облачные вычисления), *virality* (виральность) и *streaming* (стриминг), наглядно показывают, как цифровые метафоры делают сложные – и зачастую невидимые – технологические процессы осязаемыми. В фреймовой перспективе [Fillmore 1976] это достигается за счет междоменной проекции, при которой каждый термин активирует донорский фрейм со слотно-ролевой структурой и переносит ее на технический объект. Так, *cloud computing* вызывает фрейм ПОГОДЫ/ОБЛАКА (удаленное «вместилище», масштабируемая

«емкость», повсеместный доступ), *virality* – фрейм ЗАБОЛЕВАНИЯ/ЭПИДЕМИИ (контент как «возбудитель», пользователь как «носитель/хозяин», репликация через шеринг, нарастание охватов), *streaming* – фрейм ТЕЧЕНИЯ ЖИДКОСТИ (источник, канал, пропускная способность, скорость/задержка, «утечки» как потери пакетов).

В стремительно эволюционирующем лингвистическом ландшафте терминологическая система искусственного интеллекта формируется на стыке специализированного знания и повседневного когнитивного опыта, активно опираясь на метафоризацию как один из ключевых механизмов концептуального переноса. Метафоры не только упрощают восприятие сложных технических понятий, но и формируют устойчивые когнитивные фреймы, закрепляющиеся в научном дискурсе.

Терминосистема ИИ, подобно другим научным дисциплинам, развивается за счет лексики общелитературного языка и терминов, заимствованных из смежных областей, таких как математика, психология, биология, физика и социология. В процессе интеграции в новую концептуальную среду данные единицы нередко подвергаются переосмыслению – как по линии структурной аналогии (например, архитектура, модель, слой), так и в рамках функциональной или процессуальной схожести. На этой основе выделяются два устойчивых сценария метафоризации:

1) первичная метафоризация в терминосистеме-доноре, когда термин уже содержит образную нагрузку до переноса (например, *tree* (дерево) в теории графов → *decision tree* (дерево решений) в ИИ без изменения метафорического основания);

2) вторичная (интерпретативная) метафоризация в ИИ-контексте, когда общезыковая или междисциплинарная единица получает новое когнитивное наполнение внутри ИИ-дискурса (например, *agent* (агент) как автономная вычислительная сущность, *blackboard system* (система «классная доска») как архитектура совместного пространства обмена данными).

Методологический аппарат исследования на этапе анализа метафорической номинации включал выявление и отбор терминов, претерпевших метафорическое переосмысление, с последующим многоаспектным анализом: дефиниционным, концептуальным и контрастивным. Ключевую роль в интерпретации терминов играл контрастивный анализ, позволивший сопоставить метафоризованные единицы, функционирующие в научном дискурсе, с их прототипами в терминосистемах-донорах. Это дало возможность установить не только семантические соответствия, но

и зафиксировать трансформации значений, обусловленные когнитивной переработкой и адаптацией к ИИ-контексту. Тем самым была реконструирована структура когнитивных моделей, определяющих лингвистическое оформление новых технических концептов.

Для эмпирической верификации использовался корпусный инструментарий. Модули KWIC и Collocate в среде AntConc обеспечили анализ типичных контекстов употребления и идентификацию устойчивых словосочетаний, содержащих метафорические смещения. Статистические показатели LL и MI обеспечили объективную идентификацию значимых коллокаций, которые интерпретируются как маркеры скрытых фреймовых структур, формирующих основу профессионального метаязыка.

Наибольшая плотность метафорических номинаций (более 85 ТЕ) зафиксирована в сегменте оптимизационного поиска наилучшего решения или модели для конкретной задачи (см. Приложение № 3.2). В данной области метафора функционирует в двух взаимодополняющих режимах: ономаσιологическом – как средство наименования методов и алгоритмов (например, *genetic algorithm* (генетический алгоритм), *swarming* (роевое поведение)); и семасиологическом – как инструмент описания процедур и динамики поиска (например, *crossover* (скрещивание), *mutation* (мутация), *pheromone trail* (феромонный след)). Такая двойственная роль метафоры указывает на ее глубинную встраиваемость в когнитивную архитектуру профессионального мышления и на ее участие в процессах терминологической стабилизации.

Современным подходом к решению многомерных задач непрерывной оптимизации в области искусственного интеллекта является использование популяционных методов. К ним относятся эволюционные алгоритмы (включая генетические и метод дифференциальной эволюции), алгоритмы, вдохновленные живой и неживой природой (в частности метод роя частиц и алгоритм бактериального поиска), а также методы, основанные на моделировании поведения человеческого общества. Общей чертой этих стратегий является использование модели популяции фиксированного размера, где каждая особь (или агент) представляет потенциальное решение задачи [Beheshti 2013]. Поиск глобального экстремума осуществляется за счет взаимодействия особей в процессе коллективного поиска. Конкретные алгоритмы

реализуют различные механизмы взаимодействия, имитирующие поведенческие и функциональные аспекты своих биологических или физических прототипов [Beheshti 2013].

Анализ терминов-метафор, описывающих данные алгоритмы, показал, что ключевые концепты данного класса методов являются результатом транстерминологизации из биологических дисциплин, где они обозначают процессы развития и адаптации живых организмов. Основной механизм переноса – метафоризация по функциональному, процессуальному и структурному сходству между биологическими и искусственными системами [Shadrina 2025]. Так, термин *population* (популяция) в биологии означает группу организмов одного вида, проживающих на определенной территории и способных к размножению; в популяционных алгоритмах *population* – это набор потенциальных решений задачи, где каждое решение называется *individual* (особь), а процесс их адаптации моделируется как *evolution* (эволюция), включающая стадии *mutation* (мутация), *crossover* (скрещивание) и *selection* (отбор) [Goldberg 1989]. Процедуры отбора решений в ИИ аналогичны биологическому отбору. Например, *roulette wheel selection* (метод рулетки) моделирует вероятностный отбор, где шанс особи быть выбранной пропорционален ее приспособленности; для целенаправленного сохранения лучших решений применяются такие методы, как *tournament selection* (турнирный отбор) и *elitism* (элитизм) [Goldberg 1989].

Термин *individual* (особь) в этом контексте имеет обобщающее значение и варьирует в зависимости от типа алгоритма: *chromosome* (хромосома) – в генетических алгоритмах, *particle* (частица) – в физически инспирированных моделях, *agent* (агент) – в роевых алгоритмах. Такое варьирование подтверждает степень метафорической трансформации исходных понятий в терминологической системе ИИ и их адаптацию к когнитивной модели оптимизационного поиска [Shadrina 2025].

Основываясь на теории концептуальной метафоры, мы разделили отобранные термины-метафоры на четыре категории и выделили сферы-источники метафорической экспансии: «Эволюция», «Живая природа», «Неживая природа», «Социум».

Эволюционные методы представляют собой одно из старейших направлений в области метаэвристик, основанное на дарвиновских принципах биологической

эволюции и реализующее популяционный поиск через операции вариативности и отбора [Goldberg 1989]. В рамках данного направления закрепились такие алгоритмы, как *genetic algorithm*, *GA* (генетический алгоритм), *genetic programming*, *GP* (генетическое программирование) и *evolution strategies*, *ES* (эволюционные стратегии), а также модификации, ориентированные на задачи непрерывной оптимизации, прежде всего *differential evolution*, *DE* (дифференциальная эволюция) [Goldberg 1989; Beheshti 2013]. Все представленные эволюционные алгоритмы объединяют общие принципы, такие как использование популяции, функции приспособленности, операторов эволюции и итеративного процесса. Они вдохновлены природными процессами и направлены на решение задач оптимизации. Несмотря на различия в реализации, эти алгоритмы имеют общую цель – поиск оптимальных решений через эволюцию и адаптацию.

Концептуальные метафоры, заимствованные из сферы живой природы (зоологии, экологии, микробиологии), играют фундаментальную роль в описании и разработке биоинспирированных алгоритмов искусственного интеллекта. Их становление началось в 1990-х годах с формирования концепции *swarm intelligence* (роевой интеллект), вдохновленной наблюдениями за децентрализованными самоорганизующимися системами, такими как колонии насекомых, стаи птиц или косяки рыб [Beheshti 2013].

Дальнейшее развитие данного класса методов привело к расширению метафорического репертуара, включая номинации, непосредственно использующие зоонимы – термины-метафоры, основанные на названиях животных [Vitanova-Ringaseva 2024]. Классическим примером является *ant colony optimization*, *ACO* (муравьиный алгоритм), использующий модель феромонных следов для решения комбинаторных задач. Широкий спектр других природных явлений также был формализован в вычислительные методы. Например, *firefly algorithm*, *FA* (алгоритм светлячков) и *glowworm swarm optimization algorithm* (алгоритм оптимизации роя светлячков) моделируют притяжение на основе биолюминесценции; *bat algorithm* (алгоритм летучей мыши) реализует правила эхолокации; *grey wolf optimizer*, *GWO* (оптимизация серого волка) формализует иерархическую охотничью стратегию.

Среди значимых методов данного направления следует выделить *fish school search*, *FSS* (поиск косяком рыб), где вес каждой особи играет роль памяти и позволяет

избежать необходимости отслеживания глобально лучшего решения, а также алгоритм *artificial bee colony, ABC* (искусственный пчелиный рой). Объединяющей чертой этих методов является моделирование взаимодействия агентов через правила притяжения, отталкивания и кооперации, что математически описывается с помощью вероятностных переходных моделей и стохастических динамических систем [Beheshti 2013].

Самостоятельное и методологически значимое направление в области метаэвристик сформировали алгоритмы, черпающие вдохновение в законах физики и химии. В отличие от биологически инспирированных подходов, эти методы опираются на абстрактные принципы неживой природы, устанавливая аналогии между оптимизацией и динамикой физических систем; при этом важной остается общая популяционная логика поиска и использование стохастического обновления кандидатов [Beheshti 2013; Shadrina 2025]. К числу типичных номинаций данного класса относятся, например, *simulated annealing, SA* (имитация отжига), *gravitational search algorithm, GSA* (гравитационный поисковый алгоритм), *harmony search, HS* (алгоритм гармонического поиска). Особое место занимают модели с космологической метафорикой, такие как *Big Bang–Big Crunch, BB-BC* (Большой взрыв–Большое сжатие) [Shadrina 2025].

Наконец, последнюю крупную подгруппу составляют алгоритмы, метафорически ориентированные на модели социального поведения; их терминологический репертуар привлекает лексику социологии, психологии и экономики для репрезентации взаимодействия, кооперации и конкуренции, что позволяет описывать оптимизацию как коллективное принятие решений и обмен знаниями внутри популяции [Shadrina 2025]. Показательными примерами здесь являются *brain storm optimization, BSO* (оптимизация мозгового штурма) и *cultural algorithm, CA* (культурный алгоритм), концептуализирующие поиск как группогенерацию решений и социально опосредованную передачу знаний соответственно [Shadrina 2025].

Обобщая результаты, отметим, что четырехчастная классификация метафорических источников – «Эволюция», «Живая природа», «Неживая природа», «Социум» – адекватно отражает ономазиологические и семасиологические траектории становления терминологии ИИ и ее жанровую стабилизацию. Корпусные данные

показывают, что метафора в ИИ не является стилистическим орнаментом: она выступает когнитивно-операциональным механизмом, обеспечивающим перенос структур и процедур из донорских доменов в вычислительные контексты, где закрепляется через устойчивые дефиниции, коллокационные профили и продуктивные модели терминообразования. Более того, анализ показал, что метафорический перенос в терминологии ИИ может распространяться не только на отдельные термины, но и на целые концептуальные фреймы, заимствованные из смежных научных областей. Этот вывод согласуется с теорией фреймов М. Мински, согласно которой сложные знания структурированы в единицы – фреймы, содержащие «слоты» для определенных параметров и сценариев их взаимодействия [Minsky 1975]. Наглядным примером может служить алгоритм искусственной иммунной системы, который эмулирует функциональность биологической иммунной системы через сеть взаимосвязанных метафорических соответствий:

antigens → входная задача (слот «угроза» в иммунном фрейме);

antibody → потенциальное решение (слот «защитный механизм»);

killer-cells → агенты оптимизации (слот «устранение угрозы»).

Глубокий перенос фрейма демонстрирует, как метафора не только заменяет термины, но и активирует целостную когнитивную схему. Это в свою очередь позволяет исследователям экономить ресурсы, используя готовые ментальные структуры из биологии, и предсказывать поведение системы по аналогии с исходным фреймом (например, предсказывать «адаптацию» алгоритма по аналогии с иммунитетом). Кроме того, это позволяет расширить модель за счет добавления новых слотов (например, ячеек памяти для хранения удачных решений). Подобные заимствования иногда приводят к семантическим сдвигам. Например, термин *affinity* (первоначально биологическое «родство, близость») в контексте ИИ приобретает значение «мера соответствия решения». Это наблюдение служит подтверждением гипотезы М. Мински о гибкости фреймов как динамических структур [Minsky 1975].

Проведенный анализ подтверждает, что биологические, физические и социальные метафорические источники в терминологии ИИ выполняют когнитивную, а не декоративную функцию. В рамках теории концептуальной метафоры [Lakoff, Johnson, 1980] и фреймового подхода [Minsky 1975] они организуют понимание, переводя абстрактные алгоритмы (эволюционные, роевые, физические, социальные) в

интуитивные ментальные модели через проекцию знакомых доменов опыта; структурируют терминосистему, формируя семантически связанные кластеры вокруг общих концептуальных ядер; обеспечивают эвристическую и мнемоническую ценность, снижая когнитивные издержки интерпретации. Корпусная верификация (KWIC-конкордансы и меры ассоциативной силы LL/MI) фиксирует траектории стабилизации таких проекций и позволяет оперативно диагностировать семантический дрейф. Тем самым метафора выступает ключевым механизмом концептуализации, превращающим сложные ИИ-конструкции в операционализируемые схемы и повышающим междисциплинарную прозрачность дискурса.

Таким образом, фреймовый подход обеспечивает единообразную интерпретацию метафоры и метонимии как системных механизмов концептуализации в терминологии искусственного интеллекта. Метонимия реализует внутридоменное профилирование, согласно которому салиентный слот (бенчмарк, метрика, модель, выход) выступает «точкой доступа» к целому экспериментальному сценарию. Метафора, напротив, обеспечивает междоменное отображение, актуализируя структурно насыщенные области-источники (эволюционные процессы, природные явления, физические законы, социальные взаимодействия) для концептуализации новых технических реалий. При этом корпусные методы (инструменты AntConc и статистические метрики ассоциативной силы LL, MI) операционализируют и количественно подтверждают регулярность соответствующих лексико-семантических шаблонов. Практическая значимость данного подхода проявляется в таких прикладных направлениях, как терминографическая фиксация вторичных значений, онтологическое моделирование и автоматизированное извлечение знаний, – везде, где требуется точная типизация отношений и устранение семантической неоднозначности.

3.3.3. Термины-эпонимы

Введение терминов-эпонимов в научный дискурс остается дискуссионным вопросом, несмотря на значительное количество современных исследований эпонимии и процессов апеллятивизации имен собственных [Кузнецова, Петрулевич 2021; Солнцева 2022]. Ученые демонстрируют отсутствие консенсуса относительно места и ценности подобных единиц в терминосистемах. Так, Е. А. Лобач акцентирует их существенные недостатки, включая громоздкость, низкую продуктивность, неполное

отражение понятия и нарушение принципов систематизации [Лобач 1986: 91]. Д. С. Лотте указывает на их амбивалентность: с одной стороны, нейтральность, с другой – частое отсутствие прозрачной связи с номинируемым объектом или понятием [Лотте 1961: 27]. В противовес этой критике А. В. Суперанская видит в именах собственных неисчерпаемый резерв для номинации, особенно востребованный в условиях бурного развития науки и техники [Суперанская 2023]. Эта полемика и противоречивость оценок обуславливают необходимость специального рассмотрения терминов-эпонимов в рамках данного исследования.

Как отмечает Д. В. Нерубленко, высокая частотность эпонимов в научно-техническом дискурсе обусловлена рядом факторов. Во-первых, эпонимические термины выполняют мемориальную функцию, закрепляя в научной традиции имена исследователей, впервые открывших или описавших определенный феномен, орган, заболевание, вещество, технологию и т. п. Во-вторых, отсутствие синонимов у эпонимов способствует устранению семантической двусмысленности и повышает точность профессиональной коммуникации. В-третьих, эпонимы, как правило, представляют собой компактные номинации, соответствующие закону экономии языковых средств, в отличие от громоздких описательных конструкций [Нерубленко 2022]. Так, термин *Turing machine* (машина Тьюринга) является кратким и общепринятым обозначением, заменяющим развернутое описание «a mathematical model of computation that defines an abstract machine that manipulates symbols on a strip of tape according to a table of rules» («абстрактная вычислительная модель, которая оперирует символами на потенциально бесконечной ленте в соответствии с конечным набором правил»), что существенно облегчает его употребление в научной и учебной практике [Баканач 2024].

В настоящем исследовании эпоним трактуется в соответствии с определением, предложенным Е. М. Какзановой, согласно которому эпонимом считается «термин, содержащий в своем составе имя собственное (антропоним, топоним или мифоним), а также имя нарицательное, обозначающее научное понятие» [Какзанова 2019]. Кроме того, в работе учитывается позиция Н. В. Новинской, подчеркивающей, что «именно современному языку науки и техники присуща тенденция к активной эпонимизации» [Новинская 2005: 147]. Такой подход позволяет рассматривать эпонимы как значимый и продуктивный способ терминообразования в терминологии ИИ.

Далее эпоним операционализируется как лексема с проприальным компонентом (антропоним/топоним/мифоним/культуроним) и терминологической «головой» (theorem, test, model, network, process, chain, filter, operator, equation, architecture, criterion, algorithm, entropy, equilibrium и др.). Выделение кандидатов в термины-эпонимы выполнялось в AntConc (Word List) с регулярными выражениями, учитывающими дефисы, притяжательные формы ('s/s), многочастные фамилии и прилагательные с окончаниями -ian/-ean (Bayesian, Gaussian и т. п.). Контекстная верификация проводилась в KWIC и N-Gram (окно L5–R5) на заранее очищенном от служебных зон корпусе; терминологичность подтверждалась по коллокационным профилям в Collocates при порогах MI $\geq 3,0$ и LL $\geq 10,83$ ($p < 0,001$) с контролем дисперсии (Sort by Range). Для исключения искажений выполнялись нормализация графики (Bayes'/Bayes's; регистр; дефис/эндаш) и разведение омонимов (антропоним ↔ акроним: *Akaike* ↔ *AIC*; топоним ↔ продукт: *Loihi*; культуроним ↔ маркетинговое имя: *Prometheus*, *Hermes*). Каждая единица атрибутировалась по типу производящей основы и по основной семантической функции (алгоритм/модель; вычислительная архитектура; концептуальная/теоретическая рамка; математический инструмент/оператор; бренд-номинация). Результаты сопоставлялись с авторитетными источниками (стандарты ISO/IEC, словари и глоссарии по ИИ, профильные учебники и обзоры).

В проанализированном корпусе терминов ИИ выявлено 140 терминов-эпонимов ($\approx 8,5\%$ от общего объема терминологических единиц; см. Приложение № 3.3). При количественном доминировании иных лексических слоев эпонимы демонстрируют высокую терминологическую значимость: отражают механизмы наименования, быстро реагируют на научные открытия и институциональные практики, обеспечивают узнаваемость концептов и онтологическую дифференциацию. Открытость к пополнению и прочные связи с историей идей и их авторами делают данный слой приоритетным объектом исследования в диссертации, посвященной структурно-семантическим особенностям терминологии ИИ.

Важным аспектом изучения эпонимов в ИИ является реконструкция их ранней генеалогии и установление первичных номинаций. При этом ключевые фигуры – Т. Байес, А. А. Марков и К. Шеннон – предшествуют формализации ИИ как дисциплины, однако их имена вошли в ИИ-дискурс значительно позже, главным

образом через вероятностное моделирование и теорию информации. Напротив, одним из первых и последовательно используемых эпонимов, непосредственно задавшим концептуальный горизонт области, представляется обозначение *Turing machine* (машина Тьюринга), введенное А. Тьюрингом в 1936 году в статье «On Computable Numbers, with an Application to the Entscheidungsproblem» [Turing 1950]. Эта абстрактная вычислительная модель заложила основы теории вычислимости и оказала прямое влияние на ранние дискуссии об ИИ задолго до того, как термин *artificial intelligence* (искусственный интеллект) был предложен Дж. Маккарти [McCarthy 1955]. Тем самым *Turing machine* выступает не только терминологически стабильным ранним эпонимом, но и концептуальным краеугольным камнем, вокруг которого позднее формировались ключевые парадигмы ИИ.

Другие значимые термины-эпонимы поступательно входили в лексику ИИ через междисциплинарные переносы (статистика, теория информации, нейронауки, обработка изображений), а их канонизация закреплялась в учебниках, обзорных работах и стандартах (например, *Metropolis–Hastings algorithm* (алгоритм Метрополиса–Гастингса) [Russell, Norvig 2021]).

Ряд эпонимов в ИИ демонстрируют косвенное, метафорически опосредованное соотнесение с референтом, обусловленное междисциплинарными аналогиями. Так, *Boltzmann machine* (машина Больцмана) и *restricted Boltzmann machine* (ограниченная машина Больцмана) названы в честь Л. Больцмана, однако эта связь не авторская, а аналогическая – с опорой на аппарат статистической механики, прежде всего параметр *Boltzmann distribution* (распределение Больцмана), положенный в основу энергетических вероятностных моделей в нейросетях. В подобных случаях уместно говорить об аллюзивной эпонимии, когда имя собственное выступает когнитивным ориентиром и метафорой принципа, а не указанием на прямое авторство или первичное конструирование структуры [Дронова 2006].

Наряду с «классическими» примерами некоторые математические термины-эпонимы – *Chernoff bound* (граница Чернова), *Hausdorff dimension* (размерность Хаусдорфа), *Lyapunov function* (функция Ляпунова) – иллюстрируют междисциплинарный перенос фундаментальных понятий в ИИ. Возникнув вне области ИИ, эти наименования стали ключевыми для анализа устойчивости, обобщающей способности и сходимости в машинном обучении и теории управления. Их адаптация

нередко протекает без прямой апелляции к первоисточникам, а при функциональной переинтерпретации семантическая прозрачность снижается и связь с именем-референтом приобретает скорее ассоциативно мотивированный характер. Рассмотренные случаи свидетельствуют, что эпонимизация в ИИ – это не просто акт увековечения памяти, а динамический семиотический процесс, формируемый концептуальными аналогиями, историческим наслоением и конвергенцией дисциплин.

В ходе рассмотрения терминов-эпонимов ИИ было установлено, что основной для их образования являются следующие группы имен собственных: антропонимы, мифонимы и имена вымышленных персонажей, а также топонимы. Антропонимы (имена или фамилии либо реально существовавших людей, либо вымышленных персонажей) имеют самую широкую представленность ($\approx 79\%$ исследуемых терминов-эпонимов). Значительная часть терминологии, связанной с искусственным интеллектом, демонстрирует прямую атрибуцию фундаментальным теоретикам, чьи имена прочно встроены в концептуально-лексическую архитектуру области. При этом, как отмечает Н. В. Новинская, эпоним, образованный от имени либо фамилии реального человека, сам по себе не несет содержательной информации, а лишь отражает признак принадлежности. Однако в научном и научно-техническом дискурсе антропонимический эпоним занимает особое место, так как в термине-эпониме и результат открытия, достижения, изобретения, и его автор слиты воедино, являя таким образом историю развития науки [Новинская 2013: 110].

Показателен кластер эпонимов, названных в честь Т. Байеса (1701–1761) – английского статистика и пресвитерианского священника, чье наследие легло в основу широкого семейства вероятностных моделей, центральных для современного машинного обучения. В проанализированном корпусе «байесовская» линия формирует наиболее разветвленное лексическое семейство – 16 терминологических единиц: *Bayes' Theorem* (теорема Байеса), *Bayesian framework* (байесовская парадигма), *Bayesian network* (байесовская сеть), *Bayesian optimization* (байесовская оптимизация) и др. Эти наименования широко представлены во всех тематических подкорпусах, особенно в машинном обучении и обработке естественного языка, и устойчиво сочетаются с лексемами *probability* (вероятность), *uncertainty* (неопределенность), *inference* (вывод), *prior* (априорное), что отражает их центральную роль в вероятностных моделях рассуждения.

Имя А. Тьюринга (1912–1954) закреплено в лексиконе ИИ через несколько высокочастотных терминов, прежде всего *Turing machine* (машина Тьюринга), *neural Turing machine* (нейронная машина Тьюринга) и *Turing test* (тест Тьюринга), доминирующих в теоретическом дискурсе и литературе по когнитивному моделированию и выступающих концептуальными опорами для обсуждения природы вычислений, архитектуры интеллектуальных систем и границ машинного познания. В научных текстах эти термины образуют устойчивые коллокации с лексемами *intelligence* (интеллект), *reasoning* (рассуждение), *symbolic processing* (символическая обработка), *human cognition* (человеческое познание), что указывает на их стабильную релевантность в дебатах об основаниях ИИ и моделировании человеческого мышления.

Среди других фигур, чьи имена прочно закрепились в лексиконе ИИ, – А. А. Марков, Дж. Буль, Д. Хебб и К. Шеннон. Каждому соответствуют ключевые теоретические концепты с высокой сочетаемостью в ИИ-дискурсе, особенно в областях вероятностного моделирования, нейронных вычислений и теории информации (*Boolean algebra/logic* (булева алгебра/логика), *Hebbian learning/theory* (хеббовское обучение/теория), *Shannon's entropy* (энтропия Шеннона)). Эпонимы, восходящие к А. А. Маркову: *Markov chains* (цепи Маркова), *Markov model* (модель Маркова), *Markov process* (процесс Маркова), *Markov decision process, MDP* (процесс принятия решений Маркова) – составляют основу моделирования стохастических систем, в частности, в обучении с подкреплением и задачах поэтапного принятия решений.

На фоне преимущественно «канонизированных» мужских эпонимов особо выделяется язык программирования *Ada* – идеологически значимый мемориальный антропонимический эпоним, названный в честь Ады Лавлейс. Он не сводится только к технической номинации, поскольку данный термин-эпоним одновременно маркирует признание женского вклада и кодирует современные дискурсы инклюзивности в STEM-дисциплинах. Институциональная поддержка (разработка языка программирования *Ada* в начале 1980-х годов под эгидой Минобороны США) закрепила название, сделав его образцовым примером исторически мотивированного именования и лексикализации личного имени в устойчивую терминологическую единицу вычислительной техники и ИИ.

Наряду с мемориальными эпонимами (антропонимами ученых и теоретиков) в терминологии ИИ устойчиво присутствует пласт наименований, происходящих от мифологических образов, вымышленных персонажей и объектов массовой культуры ($\approx 17\%$). Эти термины служат не только техническими ярлыками, но и семиотическими «якорями», вводящими в профессиональный дискурс расширенные социокультурные смыслы и повествовательные рамки.

Отдельную подгруппу образуют мифонимы – имена собственные, восходящие к античным мифам и легендам. Имена героев и богов древнегреческого и древнеримского пантеона активно интегрировались в европейские языки посредством перехода в имена нарицательные (ср.: рус. *Геркулес*, англ. *Hercules*, фр. *Hercule*, нем. *der Hercules* – 1) герой мифов; 2) созвездие; 3) силач) [Егорова 2014: 155]; при этом, как отмечает Р. З. Мурясов, такие наименования часто сохраняют черты своих мифологических прототипов [Мурясов 2015: 952]. В частности, *Theano* – библиотека численного вычисления на Python – этимологически восходит к женскому имени Феано, которое носили несколько персонажей древнегреческой мифологии: жрица Афины в Трое; одна из дочерей царя Даная, убившая мужа в первую брачную ночь; жена царя Италии Метапонта, одна из дочерей царя Левктра, проклятию которых приписывалось катастрофическое поражение спартанской армии от рук фиванцев в битве при Левктрах, а также жена выдающегося математика античности – Пифагора. Функционально библиотека *Theano* стала одним из ранних инструментов для символического задания вычислительных графов, автоматического дифференцирования и ускорения на графических процессорах (GPU); в 2010-е годы она широко использовалась в исследованиях по глубокому обучению и заметно повлияла на последующие фреймворки.

К мифонимам примыкает явление омонимической аббревиатурной эпонимии, когда технические аббревиатуры совпадают с именами мифологических или культурно значимых персонажей и тем самым добавляют символический слой к терминологическому знаку. Так, *SETH* (сильная экспоненциальная временная гипотеза) отсылает к древнеегипетскому богу Сету; *ADAM* (глубокая адаптивная автономная машина, МФТИ) активизирует библейский образ «первого» человека, метафорически маркируя «первый/основополагающий» искусственный разум. Подобные акронимы выступают как мифологические семиотические ориентиры,

усиливающие нарративное обрамление систем и алгоритмов, оставаясь при этом полноценными терминоединицами. Одновременно они выполняют сугубо функциональную роль – обеспечивают терминологическую экономию и высокую информационную плотность выражения [Лейчик 2011].

Существенную роль в формировании терминов-эпонимов играют персонажи литературы и массовой культуры. Так, система ранней обработки естественного языка *ELIZA* (Элиза) названа в честь Элизы Дулиттл из «Пигмалиона» Б. Шоу; эта отсылка концептуально подчеркивает идею имитации «преобразующего» диалога с пользователем. Связанный термин *ELIZA effect* (эффект Элизы) описывает склонность пользователей антропоморфизировать ИИ-системы и широко обсуждается в HCI (человеко-компьютерное взаимодействие) и когнитивной науке. Аналогично, наименования *PARRY*, *ALICE*, *Dante* и *Rosie* апеллируют к культурным архетипам и литературным персонажам, формируя ожидаемые сценарии взаимодействия: *PARRY* – конфронтационное, «оппонирующее» общение, *ALICE* – разговорная машина с правил-ориентированным «поведением», *Dante* – навигация по морально нагруженным и «пограничным» тематикам, *Rosie* – домашняя автоматизация и сервисные функции. Тем самым такие эпонимы выступают метафорическими рамками, которые не только маркируют технологии, но и задают интерпретацию их функций, влияя на ожидания пользователей и исследовательские дискурсы о том, «как» и «что» делают машины.

Особый случай – *Python* (язык программирования Python), один из наиболее распространенных языков программирования в ИИ. Вопреки ожиданию «научной» или мифологической мотивации, название происходит от британской комик-труппы *Monty Python*. Автор языка Г. ван Россум выбрал его за краткость, уникальность и ироничность – качества, вдохновленные абсурдистским юмором труппы. При всей «легковесности» источника номинализации Python стал фундаментом экосистемы ИИ, вокруг которого сформировался устойчивый стек инструментов и практик – от научных вычислений и анализа данных до прототипирования и промышленного развертывания моделей; фактически он выступает *lingua franca* современного машинного обучения.

Рассмотренные примеры демонстрируют многоуровневую этимологию терминов ИИ. Имена, заимствованные из мифов, литературы и поп-культуры, повышают запоминаемость и идентичность, одновременно встраивая в технический лексикон символические, иронические или ценностные смыслы. Тем самым эпонимы

подчеркивают творческую и культурно укорененную природу именования в области искусственного интеллекта.

Топонимы составляют $\approx 4\%$ от числа исследуемых терминов-эпонимов; их основная функция в терминах и терминологических словосочетаниях – сохранение исходного именования через отсылку к географическому месту происхождения объекта, даже когда мотивирующая связь уже неочевидна. Пространственная маркированность топонима обеспечивает его связь с локусом и культурно-исторической перспективой и тем самым поддерживает «память» о происхождении обозначаемого объекта [Кузнецова, Петрулевич 2021]. Например, *Monte Carlo methods* (методы Монте-Карло) представляют собой класс стохастических процедур для оценки исходов при неопределенности [Баканач 2024]. Название навеяно городом Монте-Карло (Монако), известным казино и играми случая; базовый принцип метода – генерация случайных чисел – метафорически соотнесен с рулеткой. Несмотря на метафорическое происхождение, топоним закрепился в техническом словаре как устойчивый маркер случайности, вероятности и вероятностного вывода.

В то же время, как справедливо замечает А. Брунов, переход топонима в сферу специализированной терминологии может сопровождаться сдвигом значения, что приводит к омонимической вариативности [Брунов 2011]. Так, нейроморфный исследовательский чип *Loihi* использует асинхронные спайковые нейронные сети (SNN) для адаптивных, самоизменяющихся, событийно-управляемых параллельных вычислений. Его название восходит к подводному вулкану Лоихи у побережья Гавайев. По заявлению Intel Labs, топоним выбран для подчеркивания «мозгоподобной» обработки, однако связь между географическим именем и вторичным референтом (чипом) неочевидна и требует интерпретации и контекстуализации.

Таким образом, попадая в терминологическую систему ИИ, топонимы склонны утрачивать локативно-идентифицирующую функцию и приобретать новую роль номинативных и дифференцирующих маркеров. Такие ресемантизированные топонимы помогают различать технологические концепты и артефакты, одновременно тонко сохраняя культурный или метафорический резонанс.

Помимо установления типов производящих основ для терминов-эпонимов, каждый термин оценивался по основной семантической функции. В результате выделено пять групп:

- 1) алгоритм / модель (42%) – *Kalman filter* (фильтр Калмана), *Viterbi algorithm* (алгоритм Витерби), *Dijkstra's algorithm* (алгоритм Дейкстры);
- 2) вычислительная архитектура (21%) – *Turing machine* (машина Тьюринга), *von Neumann architecture* (архитектура фон Неймана);
- 3) концептуальная / теоретическая рамка (14%) – *Turing test* (тест Тьюринга), *Nash equilibrium* (равновесие Нэша);
- 4) математический инструмент / оператор (11%) – *Koopman operator* (оператор Купмана), *Bellman function* (функция Беллмана);
- 5) название бренда (12%) – *Prometheus* («Прометей»), *Nous-Hermes* («Нус-Гермес») и др. – имена, формирующие образ системы и задающие ожидания аудитории.

Первые четыре группы формируют вычислительно-методологическое ядро терминосистемы; они охватывают способы конструирования и анализа (методы/модели), организацию вычислений (архитектуры), теоретико-концептуальные основания и математический инструментарий оценки и доказательства. Пятая группа относится к номинативно-репрезентативному слою, кодифицируя способы именования и публичного позиционирования технологических объектов.

В современной терминологии ИИ заметно растет доля эпонимов корпоративного и брендового происхождения, что отражает конвергенцию технологических инноваций, коммерческих стратегий и языковой практики. Нередко выбираются антропоморфные или мифологически маркированные имена, размывающие границу между функциональным обозначением и маркетинговой репрезентацией. Показательны: голосовые помощники *Alexa* (Amazon; антропоним, возможная аллюзия на Александрийскую библиотеку), *Siri* (Apple; скандинавский антропоним, унаследованный от одноименного стартапа), *Watson* (IBM; отсылка к руководителю компании Т. Дж. Ватсону и доктору Ватсону у Конан Дойла), чип *Loihi* (Intel; топоним), платформа *DALL·E* (портманто от С. Дали и WALL·E.), проект *Prometheus*, линейка моделей *Nous-Hermes*, система *NeoGanesh*, платформа *Wipro HOLMES* (акроним с очевидной аллюзией на Шерлока Холмса). Такие наименования выступают не только идентификаторами, но и риторическими инструментами, формирующими образ технологии и ожидания пользователей.

Параллельно формируется пласт брендоподобных наименований в Python-экосистеме инструментов, ставших *de facto* стандартом в ИИ: *NumPy* (*Numerical Python*), *SciPy* (*Scientific Python*), *PyTorch* (Meta AI; наследует *Torch*), *Seaborn* (название вдохновлено персонажем Сэмом Норманом Сиборном из сериала “The West Wing”; в коде часто сокращается как *sns*). Префикс/суффикс *Py* выступает как экосистемный маркер и механизм «со-брендинга», укрепляя идентичность сообщества вокруг языка Python. Формально это не «чистые» эпонимы, а бренд-связанные неологизмы, консолидирующие разработчиков и отражающие модульный, кооперативный *ethos open source*.

Еще одним значимым наблюдением стало наличие у терминов-эпонимов синонимов без компонента – имени собственного, что свидетельствует о сложных процессах терминологической эволюции в научном дискурсе. Данное явление отражает три ключевых аспекта. Во-первых, замена термина-эпонима на дескриптивный термин-дублет (например, *Rumelhart algorithm* (алгоритм Румельхарта) → *backpropagation* (обратное распространение)) демонстрирует приоритет функциональной ясности над историческим признанием. В ситуации конкуренции вариантов научное сообщество, как правило, закрепляет обозначение, которое прямо указывает на принцип работы метода и облегчает его интерпретацию вне узкой исследовательской традиции. После периода параллельного использования обоих вариантов предпочтение обычно смещается в сторону функционального названия, поскольку оно поддерживает междисциплинарную коммуникацию и снижает порог усвоения ключевых концептов при обучении.

Во-вторых, ряд *Bayesian Network* (байесовская сеть) – *Belief Network* (сеть убеждений) – *Probabilistic Network* (вероятностная сеть) демонстрирует баланс между научной традицией и прагматикой. Имя Т. Байеса сохраняется как дань историческому вкладу, тогда как нейтральные синонимы подчеркивают математическую природу модели, делая термин интуитивно понятным вне контекста авторства.

В-третьих, преобладание акронимной формы *SAT* / *SAT problem* (задача выполнимости) над развернутыми вариантами – *Boolean satisfiability problem* (задача выполнимости булевых формул) и *Propositional satisfiability problem* (задача выполнимости высказываний) – фиксирует сдвиг к унифицированной, инструментально ориентированной номинации. В инженерной информатике и

индустрии *SAT* выступает междисциплинарным стандартом, формы *Boolean/Propositional satisfiability* сохраняются преимущественно в логике и формальных методах. Будучи краеугольной задачей теории сложности (NP-полная по теореме Кука–Левина), акроним *SAT* функционирует как самодостаточная категория, семантика которого не нуждается в маркировке источником («Boolean», «propositional»); наблюдаются деэпонимизация и терминологическая экономия за счет акронимизации.

Наблюдаемая вариативность эпонимических наименований в ИИ не опровергает тезис Д. В. Нерубленко об отсутствии «внутриэпонимической» синонимии [Нерубленко 2022], а указывает на более сложную траекторию терминологической эволюции. Речь идет не о конкуренции нескольких эпонимов за один и тот же денотат, а о смене типа номинации под воздействием дисциплинарной динамики и прагматики научной коммуникации. Такой сдвиг согласуется с наблюдением, что на ранней стадии признания концепции сильнее выражена мемориальная функция имени, тогда как по мере канонизации метода усиливаются требования операциональной прозрачности, скорости и интернациональности коммуникации, стимулирующие переход к деперсонифицированным дескриптивным наименованиям [Переверзева 2010].

Графическая неоднородность терминов-эпонимов, обусловленная орфографической нормой английского языка, требует отдельного рассмотрения. Вследствие обязательной капитализации имен собственных эпонимический компонент обычно сохраняет написание с заглавной буквы как в исходной форме (*Siri, Markov grammar*), так и в адъективных дериватах (*Bayesian network*). Однако процессы лексикализации приводят к утрате проприальной семантики и переходу к строчному написанию у части единиц (*algorithm, einsum*), что может свидетельствовать о закреплении термина в ядре языковой системы как общепринятой семантически автономной единицы, о завершении этапа деэпонимизации, когда связь с антропонимом (ал-Хорезми) или топонимом (Лоихи) перестает актуализироваться в языковом сознании, и о функциональном перерождении из уникального обозначения в типовой лингвистический знак с классовой референцией. Дополнительную сложность создает орфографическая вариативность на стыке специализированного и общеупотребительного регистров (*Python / python, PyTorch / pytorch*), задающая контекстно-зависимый выбор нормы в научно-технических текстах.

Таким образом, термины-эпонимы в терминологии ИИ – структурно значимый, системообразующий слой английской лингвокультуры, ценность которого выходит за пределы мемориальной функции. При сохраняющейся номинативной продуктивности и динамике пополнения эпонимы обеспечивают экономию выражения, устранение неоднозначности и высокую узнаваемость концептов, несмотря на ограничения (полилексемность, сниженная мотивационная прозрачность) и наблюдаемые тенденции к деэпонимизации и акронимизации. Учитывая неоднородность связи имени с понятием (прямая, аллюзивная, эвристическая), оптимальна «двухконтурная» стратегия нормализации: сохранять канонические эпонимы там, где они повышают точность и коммуникативную эффективность, одновременно поддерживая дескриптивные эквиваленты для междисциплинарной и учебной коммуникации. Такая политика облегчает терминографию и перевод, улучшает онтологическое моделирование и автоматическую нормализацию в естественной обработке языка.

3.3.4. Синонимия и полисемия

В терминообразовании ИИ расширение и сужение выступают взаимодополняющими механизмами, опосредующими привлечение и транстерминологизацию; их совместное действие структурно объясняет рост полисемии и конфигурации синонимии [Sager 1990; Cabré 1999]. Расширение увеличивает объем референции, насаждая новые контекстные профили без устранения старых и тем самым продуцируя межрегистровую и межподдоменную полисемию, а также сближение с соседними номинациями (условная/аспектная синонимия). Яркими примерами служат термины *model* (модель), *agent* (агент) и *prompt* (промпт), значения которых варьируют в зависимости от субдомена ИИ (машинное обучение, компьютерное зрение, обработка естественного языка). Сужение, напротив, фиксирует вход в терминологическую систему, усиливает интенционал термина и снижает кросс-доменную полисемию. Например, термин *token* (токен), обладающий широким значением в компьютерных науках, в контексте NLP сужается до «единицы субсловной сегментации». Несмотря на снижение полисемии, сужение поддерживает локальную вариативность в форме дублетов и терминологических вариантов [Kageura 2002].

В рамках предложенной модели метафоризация, эпонимизация и метонимизация рассматриваются как вторичные семантические механизмы, которые не подменяют базовые процессы семантического расширения и сужения, а настраивают их траекторию и тем самым уточняют динамику полисемии и синонимических отношений в терминосистеме. Применительно к терминологии ИИ это позволяет объяснить, каким образом перенос уже освоенных когнитивных схем в новую предметную область инициирует появление новых смыслов и приводит к росту полисемии и аспектной синонимии. Показателен термин *attention* (внимание), заимствованный из когнитивной психологии и психолингвистики и получивший в ИИ-дискурсе дополнительную семантическую нагрузку. В современных публикациях он демонстрирует регулярную полисемию: 1) общий принцип фокусирования на релевантной информации, 2) конкретный компонент модели (*attention layer / attention module*), а также 3) результат его работы – веса или распределение (*attention weights / attention distribution / attention map*), что соответствует метонимическому сдвигу типа «механизм ↔ продукт». Параллельно формируется поле аспектной синонимии и вариативных номинаций: в ряде контекстов *attention mechanism* функционирует в качестве близкого по смыслу обозначения наряду с *alignment mechanism / (soft) alignment* (особенно в традиции seq2seq-моделей), тогда как выражения *attention weights*, *attention distribution* и *attention map* выступают как частично взаимозаменяемые наименования одного и того же объекта с разными фокусами интерпретации.

Эпонимизация представляет собой специализированный способ терминообразования, преимущественно связанный с сужением интенционала. Термины, образованные по этому принципу (например, *Dijkstra's algorithm* (алгоритм Дейкстры)), фиксируются в узком, специализированном значении, что снижает их полисемию, но одновременно продуцирует локальную вариативность и дублиеты (например, *Turing test* (тест Тьюринга) vs. *The Imitation Game* (имитационная игра)), выступая маркером терминологической вариативности [Temmerman 2000].

Метонимия как механизм расширения позволяет использовать часть для обозначения целого или смежный признак, что иллюстрируется семантической эволюцией термина *prompt* (промпт), который метонимически расширился от обозначения «ввода данных» до «задачи или инструкции для ИИ-системы». С другой

стороны, метонимия способствует семантическому сужению, закрепляя термин в узком профессиональном контексте и формируя синонимические ряды через когнитивные ассоциации [Barcelona 2000].

Таким образом, интеграция метафоризации, эпонимизации и метонимии в предлагаемую концепцию позволяет моделировать их как взаимосвязанные семантические механизмы, обслуживающие процессы расширения и сужения. Их совместное действие обеспечивает структурную и когнитивную сложность терминосистемы ИИ, детерминируя динамику синонимии и полисемии.

Неразрешенная по сей день проблема определения синонимии, уходящая корнями в истоки филологии, естественным образом проявляется и в терминологии, вызывая не меньшую полемику, чем синонимия общеупотребительной лексики. Характерной чертой современной лингвистической ситуации стал «взрыв» терминологической синонимии, непосредственно связанный с бурным прогрессом в научных дисциплинах [Cabré 1999: 72].

В научной литературе представлена критическая оценка терминологической синонимии как деструктивного феномена [Лотте 1961; Толикина 1970]. Сторонники этой позиции утверждают, что синонимия нарушает структурную целостность и затрудняет функционирование терминосистемы; противоречит принципу однозначности термина (отсутствию синонимии), вводя избыточные элементы; усложняет освоение терминологии и детерминирует неоднозначность в передаче и восприятии информации.

Д. С. Лотте определял идеальный термин через краткость, однозначность и отсутствие синонимии, но отмечал распространенность синонимии в реальной терминологии (наличие нескольких терминов для выражения одного понятия) [Лотте 1961]. Анализируя генезис синонимии (в т. ч. через заимствования на ранних этапах формирования терминологии), Д. С. Лотте предлагает классификацию терминов-синонимов (абсолютные / относительные, полноправные / неполноправные) [Лотте 1982]. Под абсолютными синонимами понимаются лексемы с тождественным содержанием, тогда как относительные имеют частичное совпадение значений. Несмотря на детальное рассмотрение, исследователь оценивает синонимию преимущественно как негативное явление: «Существование для какого-либо понятия двух или более терминов, если один из них в момент введения (в частности, путем

заимствования) даже являлся полным синонимом другого, влечет собою почти всегда сперва “ограничение” сферы применения, затем “ограниченность использования в качестве элемента” для построения составного термина (сложного слова, словосочетания) и, наконец, “сужение” его значения, “расширение” или “смещение”» [Лотте 1982: 25–26].

Е. Н. Толикина рассматривает синонимию как явление, противоречащее фундаментальному требованию к термину – отсутствию синонимических соответствий. Исследователь полагает, что это приводит к нарушению структурной целостности терминосистемы, затрудняет ее функционирование и может быть причиной неоднозначного восприятия представляемой термином информации [Толикина 1970].

С. В. Гринев-Гриневиц углубляется в анализ негативных последствий синонимии для научной коммуникации и дает следующее объяснение: «Наличие нескольких синонимичных терминов вызывает у пользующихся ими специалистов стремление <...> находить между ними разницу (часто несущественную), что приводит к искажению их содержания <...> использование разных терминов вызывает неуверенность в том, что пользующиеся говорят об одном и том же понятии. Это приводит к затруднению взаимопонимания и вызывает бесчисленные споры о терминах. Поэтому в работах, посвященных нормализации терминологии, традиционно выдвигается требование отсутствия синонимов <...>» [Гринев-Гриневиц 2008: 103]. И. С. Квитко указывает на противоречие между естественным развитием науки, которое порождает синонимию, и потребностью научного стиля в точности, что создает конфликт и стремление избавиться от синонимии в терминосистемах [Квитко 1976: 60].

Ряд лингвистов рассматривают синонимию терминов в положительном ключе как важный инструмент для семантического обогащения и повышения коммуникативной эффективности терминологических систем. В. П. Даниленко признает наличие синонимии в терминологии и рассматривает ее как естественное и полезное явление для развития терминосистемы: « <...> синонимы не могут быть правилом, но, с другой стороны, наличие множества причин и предпосылок для появления и сосуществования синонимичных наименований одного понятия делает

синонимы слишком частым исключением, чтобы так категорически их запрещать» [Даниленко 1977: 176].

Причину нерешенности проблем синонимии в терминологии видят прежде всего в смешении понятий «синоним», «вариант», «дублет», «дублетность» и их производных [Щербина 2000]. Терминологическая синонимия и дублетность часто отождествляются, поскольку в обоих случаях одно понятие обозначается разными терминами [Даниленко 1977]. Однако ряд авторов принципиально разграничивают эти понятия. Сторонники строгого подхода полагают, что дублеты – это абсолютно эквивалентные термины, не имеющие стилистических различий, тогда как синонимия с присущими ей семантическими и стилистическими нюансами характерна для общего языка; в терминосистемах, по их мнению, допустима лишь дублетность [Толикина 1970; Квитко 1976]. В то же время С. В. Гринев-Гриневиц, Б. Н. Головин, Р. Ю. Кобрин видят в дублетности частный случай синонимии – абсолютную эквивалентность, обусловленную спецификой научного понятия [Головин, Кобрин 1987; Гринев-Гриневиц 2008]. Эту позицию разделяет С. Д. Шелов (ограничивает дублетность заимствованно-исконными парами) [Шелов 2014].

Критики категоричного отрицания синонимии указывают, что отсутствие стилистических оттенков само по себе не доказывает дублетность и не исключает синонимических отношений [Гречко 1987]. Допустимость стилистически окрашенных терминов-синонимов (книжные, разговорные, профессионализмы) поддерживается В. П. Даниленко [Даниленко 1977].

Дискуссионным остается вопрос о разграничении вариантности и синонимии. В ряде работ эти термины фактически употребляются как взаимозаменяемые. Так, С. В. Гринев-Гриневиц утверждает: «...лексемы, различающиеся даже одной буквой, являются разными самостоятельными терминами и, поскольку они служат для называния одного понятия, полностью соответствуют признакам абсолютных синонимов» [Гринев 1993: 109–110]. Вместе с тем существует подход, согласно которому вариантность понимается как сугубо формальная множественность обозначений одного и того же термина (орфографические, морфологические, словообразовательные, графические, аббревиатурные и др.), тогда как синонимия – как конкуренция различных терминов с близкими, но не тождественными значениями [Даниленко 1977; Татаринев 2006].

В настоящей работе, вслед за С. В. Гриневым-Гриневичем, мы принимаем нормативный подход и рассматриваем любые лексемы, называющие одно и то же понятие, но отличающиеся формально (вплоть до одной графемы, способа дефисного написания, варианта транслитерации/аббревиатуры), как абсолютные синонимы внутри синонимического поля данного понятия. Термин «условные синонимы» резервируется для единиц с близкими, но не тождественными значениями [Гринева-Гриневич 2008]. При описании терминосистемы искусственного интеллекта мы учитываем молодость и междисциплинарность области, в которой синонимия – естественный эффект эволюции и межотраслевых контактов. В этой связи классификация С. В. Гринева-Гриневича – с разграничением абсолютных и условных синонимов и включением вариантов в поле синонимии – представляется наиболее адекватной практике современной терминологии ИИ. Такой выбор позволяет фиксировать реальную вариативность, не подменяя ее жестким нормативным предписанием. Одновременно регистрируя дубликаты и варианты, мы предлагаем предпочтительные формы и правила их употребления, что существенно для быстро развивающейся терминосистемы, а также для задач стандартизации, переводческой согласованности и автоматизированной обработки.

На практике синонимия в терминосистеме ИИ формируется как результат взаимодействия трех процессов: 1) параллельной номинации в разных научных школах и исследовательских подразделениях индустрии (например, *few-shot learning* \approx *small-sample learning* / *low-shot learning* – вариативность между сообществами компьютерного зрения и современного машинного обучения); 2) конкуренции межъязыковых и внутриязыковых стратегий наименования – заимствования, калькирования, транстерминологизации и морфологического варьирования (например, *regularization* \approx *shrinkage/penalization* – перенос статистической терминологии в машинное обучение); 3) маркетингово-мотивированной ребрендизации сходных технических решений (например, *foundation model* \approx *large pre-trained model* / *LLM*). На корпусном уровне это проявляется в сосуществовании нескольких лексем с высоким перекрытием коллокационных профилей вокруг одной и той же онтологической позиции (ср. «объяснимость»: *explainability*, *interpretability*, *transparency*, *comprehensibility*), в асимметрии их распределения по позициям внутри научной статьи (заголовков, аннотация, ключевые слова, методы, обсуждение), по поддисциплинам и

площадкам публикации (NLP, CV, HCI; журналы/конференции), а также в различиях деривационных семейств и акронимизации (*interpretable, post-hoc interpretability vs explainable, XAI*). Эти наблюдения плохо согласуются с постулатом «нулевой синонимии» без потери эмпирической точности и, следовательно, требуют явного учета при корпусном анализе терминосистем [Temmerman 2000; L’Homme 2020].

Для выявления синонимических рядов в корпусе научных текстов по ИИ применялся контекстный анализ конкордансов (KWIC) с акцентом на синтаксические паттерны, маркирующие синонимические отношения (апозитивные и дефиниционные конструкции («X, also known as Y»; «X, also called Y»; «X (Y)»; «X is called Y»), координационные конструкции («X and Y», «X or Y»); статистический анализ коллокаций в окне L5–R5 с применением мер Mutual Information ($MI \geq 3.0$) и Log-Likelihood ($LL \geq 10,83$; $p < 0,001$) для проверки распределительной близости членов ряда (совпадение правых голов в сочетаниях N+N (например, model, algorithm, system, network, function, rate) и левых глаголов управления (например, train, evaluate, implement, design, optimize, apply, fine-tune, compute, calculate, define)); а также шаблонный анализ (Cluster, N-Gram) с извлечением устойчивых лексико-синтаксических рамок, специфичных для научного дискурса ИИ (операциональные: applying X to Y; сравнительно-оценочные: «X outperforms Y»; классификационные: «X belongs to the family of Y»). Для минимизации ложноположительных случаев учитывались контрастивные и перечислительные конструкции («Unlike X, Y ...»; списки когипонимов в «X and Y»); такие употребления не рассматривались как маркеры синонимии при отсутствии апозитивного или дефиниционного сигнала.

Синонимический ряд формировался при соблюдении двух условий: функциональная эквивалентность в контекстах (перекрытие типичных правых голов и левых глаголов управления) и наличие не менее двух независимых явных индикаторов синонимии в KWIC из разных источников. Устойчивым считался ряд, в котором каждый член представлен не менее чем 5 репрезентативными контекстами KWIC и которому соответствует не менее 5 общих ключевых коллокаций (по LL) при минимальном расхождении распределительных профилей. Проверка корректности и согласование результатов осуществлялись путем сопоставления полученных рядов с авторитетными глоссариями и стандартизирующими ресурсами.

По результатам кластеризации терминов по критерию функциональной эквивалентности было выделено и систематизировано 257 синонимических рядов (см. Приложение № 3.4) и выполнена их систематизация. В основу систематизации положено различие двух базовых типов: 1) **абсолютные синонимы** – обозначения одного и того же понятия, совпадающие по денотату и распределению функций и различающиеся лишь формально (*pre-training* (предобучение) \approx *pretraining*; *support vector machine* (машина опорных векторов) \approx *SVM*); 2) **условные синонимы** – обозначения с частичным семантическим перекрытием, функционирующие как взаимозаменяемые только в определенных жанровых, дисциплинарных или коммуникативных контекстах (например, *creative computing* (творческие вычисления) \approx *creative computation*). Абсолютные синонимы в свою очередь подразделяются на **варианты** – формальные модификации исходного термина (*tokenization* (токенизация) \approx *tokenisation*) и **дублиеты** – эквивалентные термины с разной морфологической структурой (например, *first-order logic* (логика первого порядка) / *first-order predicate calculus* (исчисление предикатов первого порядка)).

Вариантность, обусловленная широкой акронимизацией научного английского в ИИ, составляет крупнейший класс (200 рядов в нашей выборке). Мы фиксируем полные и краткие варианты (*winter of artificial intelligence* \leftrightarrow *AI winter*); среди кратких выделяем: графические (*human in the loop* \leftrightarrow *human-in-the-loop*); усеченные (*network* \rightarrow *net*), эллиптические (*artificial neural network* \rightarrow *neural network*), сложносокращенные (*capsule neural network* \rightarrow *CapsNet*), акронимные (*neural network* \rightarrow *NN*), гибридные – смешение полной и краткой форм (*artificial intelligence system* \rightarrow *AI system*), а также комбинированные: аббревиация+эллипсис (*adaptive gradient algorithm* \rightarrow *AdaGrad*); акронимия+сокращение (*artificial intelligence for IT operations* \rightarrow *AIOps*); алфанумерический тип (*sequence-to-sequence (model)* \rightarrow *seq2seq*). Такая стратификация позволяет системно описывать формальную множественность обозначений при сохранении единого понятийного содержания. Полный вариант, как правило, имеет несколько кратких: *artificial neural network (ANN)* – *neural network (NN)* – *neural net*.

В ходе исследования выявлено 29 групп абсолютных синонимов-дублиетов: *relation extraction* (извлечение отношений) / *relationship extraction* (извлечение отношений) – одно понятие в NLP; *Boolean satisfiability problem* (задача выполнимости

булевых формул) / *propositional satisfiability problem*, *SAT* (задача выполнимости высказываний) – два обозначения одной и той же задачи SAT и др.

Типология дублетов по происхождению включает три класса: одновременные (диахронические), ареальные (географически/сообщественно маркированные) и разноязычные. Корпус фиксирует единичные, но показательные случаи переименования, отражающие смену парадигмы, уточнение смысла или избегание нежелательных коннотаций. Так, употреблявшийся А. Тьюрингом в 1940–1950-е годы термин *machine intelligence* (машинный интеллект) акцентировал аналогию с человеческим интеллектom; с 1956 года его вытеснил дублет *artificial intelligence* (искусственный интеллект), сместив фокус на «искусственную» природу систем и сохранив предшественника главным образом в исторических контекстах. Ареальные дублеты также редки. Например, термин *brain-machine interface* (интерфейс мозг–машина) относительно чаще встречается в биомедицинских публикациях США, тогда как общепринятым обозначением является *brain-computer interface* (интерфейс мозг–компьютер). Разноязычные дублеты в пределах корпуса встречаются эпизодически и обычно связаны с переводами стандартов, двуязычными источниками или цитированием национальных инициатив. Например, *federated learning* (англ.) / *apprentissage fédéré* (фр.) / *aprendizaje federado* (исп.) – встречаются в англоязычных текстах при ссылках на локальные регуляторные или образовательные инициативы; как дублет фиксируются, если переводная форма используется в основном тексте, а не только в глоссарии.

Небольшая представленность одновременных и ареальных дублетов в современной терминосистеме ИИ объясняется спецификой формирования области как глобально унифицированной и технологически ориентированной дисциплины. Во-первых, стремительный рост ИИ в последние десятилетия уменьшает период устойчивой конкуренции альтернативных наименований, тем самым сужая диахроническую вариативность. Во-вторых, доминирование англоязычного академического и инженерного сообщества приводит к первичному появлению терминов на английском (преимущественно в американской традиции), минимизируя вероятность устойчивых ареальных расхождений. Наконец, международные стандарты, метрики и цифровые платформы публикаций ускоряют унификацию и

глобальное распространение предпочтительных форм, снижая потребность в параллельных локальных наименованиях.

Условные синонимы в терминосистеме ИИ – это пары или кластеры терминов с близким, но не тождественным значением. Различия обычно возникают из-за разного объема понятия (род–вид), онтологического статуса («функция» vs «архитектура»), научной традиции или рабочего контекста применения. Кроме того, часть условной синонимии имеет метонимическую природу, проявляющуюся в переносе обозначения между смежными онтологическими ролями (*механизм* ↔ *инстанция/слой*: *self-attention* ↔ *self-attention layer*; *метрика* ↔ *целевая функция*: *Earth Mover's Distance* ↔ *Wasserstein loss*; *свойство/функция* ↔ *архитектура*: *autoassociative memory* ↔ *autoassociation network*).

По данным корпуса зафиксировано 28 рядов условной синонимии. В таких парах взаимозаменяемость обычно допустима в операциональных разделах (“Methods” и его подразделы, “Results”, “Discussion”), тогда как в дефинициях и таксономиях термины следует разводить. Например, *personal data* (персональные данные) и *personally identifiable information, PII* (персонально идентифицируемая информация) расходятся по правовому объему: «*Under the GDPR, device identifiers and IP addresses were treated as **personal data**...*» («в соответствии с GDPR идентификаторы устройств и IP-адреса рассматриваются как персональные данные...») и «*...the public release excludes **PII** such as names...*» («...в открытой публикации исключается ПИ, например имена...»). В ЕС-контексте *personal data* охватывает более широкий набор идентификаторов (в т. ч. IP-адреса, cookie-ID), тогда как в США *PII* чаще соотносится с явными идентификаторами личности (имя, e-mail, ID). В операциональном описании допустима подстановка: «*The released dataset contains no **personal data** / **PII**, and all identifying fields were removed during de-identification*» («Публикуемый набор данных не содержит персональных данных / ПИ; все идентифицирующие поля удалены на этапе деидентификации»).

В эволюционных алгоритмах отношения «род–вид» иллюстрирует пара *recombination (рекомбинация)* ↔ *crossover (скрещивание)*: «*The GA employs **recombination**...; specifically, uniform **crossover** is used...*» («генетический алгоритм использует рекомбинацию; конкретно применяется равномерный кроссовер...»). В методическом контексте взаимозаменяемость также возможна: «*We maintained population diversity by enabling **recombination (crossover)** with a probability of 0.7.*» («Мы

поддерживали разнообразие популяции, включая рекомбинацию (кроссовер) с вероятностью 0,7»).

Таким образом, условная синонимия в терминологии ИИ представляет собой не случайную вариативность, а системно обусловленное явление, отражающее когнитивные механизмы категоризации и функциональную организацию научного дискурса.

Некоторые пары занимают промежуточную позицию между дублетами и условными синонимами. Так, *dialogue system* (диалоговая система) / *conversational agent* (разговорный/диалоговый агент) часто используются как абсолютные синонимы, но «agent» может имплицировать автономность/воплощенность, что делает пару контекстно-условной; *machine listening* (машинное восприятие звука) / *computer audition* (компьютерный слух) во многих корпусах ведут себя как дублеты, но *computer audition* нередко шире (включает синтез/ретривал), поэтому пара условная по объему.

Также для терминосистемы ИИ характерно наличие формально сходных, но семантически различающихся терминов; лексемы имеют близкое звучание или общую основу, однако обозначают разные понятия или разные аспекты одного процесса. В терминоведении такие случаи описываются как паратермины или смежные термины, формирующие семантические поля с высокой степенью пересечения, но не являющиеся синонимами [Temmerman 2000; Татаринев 2006; Гринев-Гриневиц 2008; L'Homme 2020]. Показателен контраст терминов *prompt engineering* (промт-инжиниринг) vs *prompt design* (промт-дизайн). *Prompt engineering* – более широкое и прикладное понятие, включающее в себя разработку, тестирование и оптимизацию запросов (промтов) с целью улучшения результатов генеративной модели. *Prompt design* – более узкое и концептуальное понятие, акцентированное на структуре, формулировке и лингвистических особенностях самих запросов. Аналогично, пара *prompt template* (шаблон промта) vs *prompt pattern* (паттерн промта) иллюстрирует смежность без тождества. *Prompt template* – конкретная реализуемая заготовка с параметрами/слотами для подстановки (операциональный артефакт в задании), тогда как *prompt pattern* – обобщенная схема/прием решения (концептуальный образец, например, *few-shot pattern*, *chain-of-thought pattern*), из которого могут выводиться разные шаблоны. Их семантические поля пересекаются (обе единицы обслуживают одну задачу – управляемую генерацию), но онтологический статус различается

(«конкретная реализация» vs «обобщенная конструкция»), что не позволяет считать их синонимами.

В анализируемых примерах поверхностное формальное сходство часто оказывается обманчивым, что требует учета лингвистического и семантического контекста каждой единицы. Надежная обработка должна сочетать корпусную верификацию (коллокационные профили, распределительные характеристики; метрики LL/MI), использование контекстуальных и семантических моделей и экспертную оценку для разведения паратерминов и подлинных синонимов. Такой подход минимизирует концептуальное смешение и повышает точность автоматизированных процедур извлечения и обработки терминов.

Таким образом, терминосистема ИИ характеризуется динамичной эволюцией и контекстно-зависимой синонимией. Абсолютные синонимы (в т. ч. варианты и дублиеты) отражают процессы стандартизации и унификации; условные – обслуживают разнообразные коммуникативные задачи; ложные дублиеты требуют строгой дифференциации. Синонимия в ИИ – не дефект, а показатель многослойности концептов; ее управляемость обеспечивается нормированием и корпусной диагностикой коллокаций и контекстов в научных статьях.

Полисемия в терминологии трактуется как наличие у одной номинативной единицы нескольких взаимосвязанных значений в пределах одной предметной области, обычно возникающих в результате видоизменения и развития первоначального значения этой единицы [Ахманова 2004]. В классической терминологии полисемия рассматривается как нежелательное отклонение от идеала «один термин – одно понятие». Еще Д. С. Лотте – основоположник отечественного терминоведения – считал многозначность серьезным недостатком терминосистемы и предлагал различать случаи, когда один термин называется многозначным при более чем двух значениях: 1) значения термина не имеют между собой семантической связи либо связаны лишь отдаленно; 2) термин объединяет относительно близкие по смыслу понятия [Лотте 1961].

Существует и иная точка зрения, согласно которой полисемия существует, но в пределах одной терминосистемы, поскольку принцип однозначности практически недостижим в терминологии в целом [Кутина 1970]. В. М. Лейчик объясняет многозначность использованием терминов в близких отраслях знания; причиной

данной многозначности является транстерминологизация терминов. Кроме того, рассматривается категориальная многозначность, при которой в результате метонимического переноса одним термином можно обозначать и процесс, и результат [Лейчик 2022]. По мнению Е. Н. Толикиной, только категориальная многозначность относится к разряду истинной полисемии и имеет право на существование в терминосистемах [Толикина 1970]. Особым видом полисемии называют авторскую, которую понимают как «обозначение уже существующим в науке термином новых объектов и понятий в зависимости от авторской исследовательской позиции, принадлежности к научной школе» [Головин, Кобрин 1987: 51].

Представители социокогнитивного подхода подчеркивают, что явления полисемии и синонимии в терминологии – это естественные этапы развития терминосистемы; их не следует рассматривать исключительно негативно, напротив, их изучение и описание способствует достижению большей точности в профессиональной коммуникации [Cabré 1999; Temmerman 2000]. Современные терминоведы также отмечают, что многозначность терминов – интегральная черта специализированной коммуникации, требующая не запрета, а управления и учета при создании терминологических ресурсов [Kageura 2002; L'Homme 2020]. Более того, на практике полисемия нередко устраняется самим контекстом. В специализированных текстах различные смыслы термина обычно проявляются через разные коллокации или уточняющие словосочетания, благодаря чему неоднозначность минимизируется в рамках узкопрофильного дискурса.

В данной работе мы исходим из допустимости полисемии в пределах одной терминосистемы (ИИ) и трактуем ее как функционально оправданный ресурс – механизм семантической экономии и адаптивности. Несмотря на утрату термином точности, зависимость от контекста, нарушение базового принципа унивокальности в профессиональной коммуникации, полисемия довольно распространена в языке для специальных целей и привлекает внимание исследователей, рассматривающих ее как нормативную характеристику зрелых терминосистем, требующую точной контекстной разметки и прозрачных правил корпусного извлечения [Temmerman 2000; L'Homme 2020; Ноу 2025].

Для анализа полисемии в корпусе научных статей по ИИ применялся комплексный подход с использованием корпусного менеджера AntConc. Методика

включала: контекстный анализ конкордансов (KWIC) с фокусировкой на синтаксических маркерах смены значения (конструкции «in this context», «meaning that», «refers to»); статистическую оценку коллокационных профилей в окне L5–R5 по метрикам Mutual Information ($MI \geq 3,0$) и Log-Likelihood ($LL \geq 10,83$; $p < 0,001$) для выявления значимых различий в лексическом окружении; шаблонный анализ (Cluster, N-Gram) с извлечением устойчивых рамок употребления (правые головы: X system/model/agent; левые глаголы управления: deploy/use vs acts/decides).

Полисемичные термины описывались через семантическое ядро – минимальный инвариант, разделяемый их употреблениями; операционально ядро устанавливалось по стабильному пересечению наборов ключевых коллокаций и повторяющихся метаязыковых/дефиниционных маркеров в KWIC. Для индуктивного выделения вариантов интерпретации контексты (KWIC) группировались в кластеры на основе общих коллокационно-синтаксических паттернов; устойчивым считался кластер при наличии не менее 10 репрезентативных вхождений и не менее 5 уникальных ключевых коллокаций. Термин квалифицировался как полисемичный при наличии ≥ 2 таких кластеров с минимальным пересечением топ-коллокаций и различающимися онтологическими/вербальными рамками при сохранении единого семантического ядра. Валидация результатов осуществлялась двухконтурным сопоставлением: 1) с международными стандартами ISO/IEC по ИИ [International Organization 2025], 2) с рецензируемой научной литературой (монографии, обзорные и методологические статьи); спорные случаи разрешались либо в пользу норматива ISO/IEC, либо при подтверждении не менее чем двумя независимыми публикациями.

С применением описанной методики было установлено, что в терминологии ИИ полисемия выражена заметно слабее, чем синонимия. Корпусная выборка фиксирует небольшой, но устойчивый слой многозначных единиц – 21 лексему (1% наблюдаемых терминов), что свидетельствует о тенденции терминосистемы к смысловой экономии при высокой функциональной вариативности употребления. Полный список выявленных полисемичных терминов представлен в Приложении № 3.5.

Главные источники многозначности – привлечение общеязыковых слов (например, *model*: «архитектура/семейство» vs «экземпляр обученной системы»; *agent*: «программный агент/участник среды»), где на базовое значение наслаиваются специализированные чтения; транстерминологизация из смежных областей (например,

attention, reward, entropy, gradient), когда сохраняется «донорское» ядро, но возникает операциональная интерпретация (*gradient* как «вектор производных» vs «метод оптимизации»); метонимические сдвиги типа процесс → результат, функция → значение, модуль → архитектура: *loss* «функция» → *loss* «числовое значение»; *decoder* «компонент модели» → «самостоятельный модуль/семейство моделей». Такая конфигурация подтверждает нормативный характер умеренной полисемии в зрелых терминосистемах при условии контекстной разметки и прозрачных правил корпусного извлечения.

Показательный случай контролируемой полисемии – термин *artificial intelligence* (искусственный интеллект). Он используется для обозначения ряда концептуально взаимосвязанных смыслов – от области знания до конкретной технологической реализации. В частности, в нормативных и академических источниках фиксируются следующие значения:

- раздел компьютерной науки: «branch of computer science devoted to developing data processing systems» [ISO/IEC/IEEE 24765:2017];
- научная дисциплина: «discipline concerned with the building of computer systems» [ISO/IEC 39794-16:2021];
- технология, система или программное обеспечение: «capability of an engineered system to acquire, process and apply knowledge and skills» [ISO/IEC TR 24029-1:2021];
- способность машины к интеллектуальному поведению: «capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning» [ISO/IEC 2382:2015];
- индивидуальный агент или сущность: «an instance of this type of software; a (notional) entity exhibiting such intelligence» [Oxford English Dictionary 2025].

Эволюция терминологического значения сопровождается метонимическими переходами: от абстрактного научного направления к технологии (*абстракция* → *приложение*); от процесса разработки к результату – системе или программе (*деятельность* → *продукт*); от системы – к агенту, действующему в среде (*структура* → *функция* / *субъект*). В корпусе соответствующие интерпретации разводятся по коллокационным профилям: *AI research/field/discipline* ↔ *AI system/technology/software* ↔ *AI agent*.

Эти значения формируют общее семантическое ядро, связанное с идеей искусственного интеллекта как способности технических систем к имитации, воспроизведению и развитию форм интеллектуального поведения. Несмотря на различие в уровнях абстракции и сфере применения, все значения термина *AI* (ИИ) демонстрируют семантическую связь и развиваются на основе общего концепта «искусственного разума».

Анализ многозначных единиц терминосистемы ИИ позволил установить, что при сравнительно низкой общей доле полисемии в терминологии ИИ устойчиво проявляется системная многозначность. Каждое значение опирается на компактное семантическое ядро и расходится по типовым траекториям: процесс → ресурс/фаза/результат (*training, inference*); класс/принцип → экземпляр/модуль (*model, ReLU*); структура → визуализация (*graph*); параметр → статистическое/социальное смещение (*bias*). Ключевые механизмы – метонимизация, а также сужение и расширение референтной области; в профессиональном дискурсе полисемия функционирует не как «размывание», а как управляемое распределение интерпретаций. Корпусная проверка на англоязычном материале подтверждает регулярность указанных траекторий.

Таким образом, в терминосистеме ИИ синонимия количественно преобладает как следствие междисциплинарности и быстрого обновления практик именования, тогда как полисемия выражена слабее и функционирует как управляемый ресурс семантической экономии. В нормативном описании мы различаем абсолютные эквиваленты (включая варианты и дублиеты) и условные синонимы (типовые оси расхождения: род–вид, механизм–инстанция, метрика–функция потерь, область охвата). Полисемия главным образом возникает при привлечении общеязыковых единиц, транстерминологизации и метонимических сдвигах (процесс → результат/ресурс, функция → показатель и др.); интерпретации разводятся контекстом и коллокационными профилями. Корпусные методы (KWIC, распределительные характеристики, LL/MI) фиксируют жанрово-поддисциплинарное распределение, стадии стабилизации и семантический дрейф. Это позволяет нормировать синонимию (предпочтительные формы с полями ссылок) и инвентаризировать значения с онтологической разметкой и правилами разрешения неоднозначности, что критично

для стандартизации, переводческой согласованности и автоматизированной обработки.

Переходя от частных траекторий к системному уровню описания, представим суммарную классификацию семантических способов терминообразования (см. Рисунок 8).



Примечание. ----- возможная внутримоделная синонимия и полисемия терминов ИИ

Рисунок 8 – Семантическая классификация способов терминообразования в терминосистеме ИИ

Схема консолидирует основные пути пополнения терминосистемы ИИ – заимствование, привлечение общеязыковых единиц и транстерминологизацию, а также механизмы их семантической адаптации (метафоризацию, метонимизацию, расширение и сужение значения, эпонимизацию), демонстрируя потенциальные, но не обязательные выходы в синонимию и/или полисемию; во многих случаях закрепляется одна однозначная, поддающаяся стандартизации номинация.

ВЫВОДЫ ПО ГЛАВЕ 3

Результаты Главы 3 демонстрируют, что терминосистема предметной области «Искусственный интеллект» формируется как многоуровневая, динамически балансирующая структура. Количественно преобладают термины-словосочетания – около 73% единиц (1198 ТЕ), однако стратификация не сводится к длине обозначения. Ядро формируют устойчивые интернационализмы и часть стабилизированных многословных терминов, тогда как на периферии находятся производные и сложные единицы вместе с вариативными многословными конструкциями. Такая конфигурация одновременно поддерживает преемственность понятийного аппарата и ускоряет терминологический прирост в зонах спецификации и технологических новаций, что выражается в росте сложных обозначений и компрессивных формул.

Морфологический срез подтверждает ведущую роль суффиксации как механизма категоризации и номинализации процессов. Продуктивны, в частности, суффиксы *-ing* (процедуры и операции), *-ity* (абстрактные свойства), *-ation/-ization* (процессы/результаты), *-er/-or* (агенты/инструменты). Такая деривация не только упорядочивает признаки по регулярным формантам, но и повышает унификационный потенциал терминосистемы, позволяя минимизировать вариативность при расширении онтологического охвата. Наряду с суффиксацией задействованы словосложение и редукативные стратегии – от неоклассических композитов до формульных шаблонов типа *X-to-Y*, *X-as-a-Y*, *X-in-the-loop*; аббревиатуры и телескопии обеспечивают экономию формы без утраты семантической прозрачности, причем синхронные акронимы трактуются как варианты полной формы, а диахронно лексикализованные – как элементы однословного слоя. Совместное действие этих механизмов поддерживает «тонкую настройку» между экономией знака и предсказуемостью интерпретации.

Синтаксическая деривация выступает главным источником терминологического прироста. Среди терминов-словосочетаний преобладает двухкомпонентная модель [модификатор + головная основа] (около 72% всех терминов-словосочетаний), системно кодирующая дифференциальные признаки через левый препозитивный модификатор при стабильном значении головной основы (*logic*, *network*, *model* и др.). Многоступенчатые конструкции функционируют как «свернутые определители», одновременно повышая номинативную емкость и поддерживая внутреннюю связность.

Продуктивные трехкомпонентные паттерны обычно опираются на устойчивые двучленные базы; в верхнем сегменте встречаются и идиоматические единицы (*curse of dimensionality* (проклятие размерности), *black box* (черный ящик)), которые демонстрируют терминологизированную метафоричность и фиксированность формы. Тем самым сегмент словосочетательной номинации функционирует в «двойном режиме»: 1) обеспечивает прозрачную интерпретацию за счет регулярных композитов и предсказуемой композиционности, 2) закрепляет сложные явления посредством идиоматизированных конструкций, формируя краткие и легко распознаваемые наименования.

Существенным итогом анализа структуры терминов ИИ стало выявление континуума «развертывание ↔ компрессия», определяющего циклическую динамику терминообразования в данной предметной области. Первичная номинация, как правило, оформляется синтаксически – в виде атрибутивного словосочетания с правым головным компонентом, обеспечивающего точную спецификацию понятия (напр., *large language model* (большая языковая модель)). По мере закрепления концепта активируются компрессивные механизмы (словосложение, усечение, акронимизация), порождающие компактные эквиваленты и варианты (ср. *large language model* (большая языковая модель) → *LLM* (большая языковая модель)). Далее сокращенные формы вновь включаются в синтаксические модели как модификаторы и опорные «строительные блоки», формируя повторяющийся цикл «синтаксис → морфология → синтаксис» (напр., *LLM pre-training* (предобучение большой языковой модели)). Данная динамика поддерживает одновременно когнитивную экономию за счет сокращения длины наименований и онтологическую связность, поскольку сокращенные ярлыки сохраняют родо-видовые отношения и наследуют семантическое ядро исходной номинации, позволяя масштабировать терминологические серии без потери интерпретируемости.

Механизмы семантической адаптации в ИИ образуют согласованную систему, в которой метафоризация и метонимизация действуют как когнитивные «ускорители» концептуализации (активируя донорские фреймы и типовые сдвиги типа «процесс → результат», «механизм → инстанция/модуль»), эпонимизация выполняет роль социокогнитивного «якоря» (закрепляя наименования за авторитетными источниками и повышая распознаваемость/нормируемость), а расширение и сужение значения

(генерализация/специализация) калибруют объем референции и выстраивают родовые проекции без утраты семантического ядра. На корпусном уровне все четыре механизма проявляются как устойчивые коллокационные паттерны и жанрово-поддисциплинарные профили.

Лексико-семантическая вариативность проявляется в количественном преобладании синонимии (включая варианты и дублиеты), тогда как полисемия выражена слабее и выступает управляемым ресурсом семантической экономии. Синонимия распределяется между абсолютными синонимами (часто различающимися формально/морфологически) и «условными» парами, где сближение обусловлено отношениями род–вид, механизм–инстанция, метрика–функция потерь или жанрово-дисциплинарными режимами употребления. Полисемия локализована в ограниченном слое единиц ($\approx 1\%$), преимущественно у общеязыковых слов, на которые наслаиваются специализированные значения (*model, agent, prompt*), а также у терминов, где регулярно возникает метонимия «процесс → результат/величина». Эти результаты подтверждают, что для ИИ-дискурса характерна стратегия «тонкого разведения» значений и предпочтение нормирования синонимических рядов с одновременной инвентаризацией значений многозначных единиц.

Методологически корпусные процедуры (KWIC, распределительные профили, LL/MI) продемонстрировали высокую дискриминативную силу, фиксируя стадии стабилизации и направления семантического дрейфа. Полученные результаты напрямую конвертируются в практики терминографической нормализации. В прикладном плане на основании полученных результатов можно дать ряд рекомендаций: профилировать однословные единицы по операциональным критериям (частотность, степень кодификации, устойчивость коллокаций) без введения дихотомий; для многословных наименований – фиксировать композиционные паттерны и устойчивые коллокации; применять регистрово-жанровую детализацию к метонимическим употреблениям; трактовать синхронные акронимы и эллиптические формы как варианты полной номинации для целей учета и поиска; осуществлять выбор предпочтительных наименований в синонимических гнездах с прозрачными перекрестными ссылками; использовать формульные шаблоны как контролируемые «терминоконструкторы».

ЗАКЛЮЧЕНИЕ

В настоящем исследовании предпринята попытка комплексного анализа структурно-семантических особенностей терминологии предметной области «искусственный интеллект». Рассматриваемый пласт лексики трактуется как совокупность специальных наименований, формирующаяся под влиянием стремительного технологического прогресса и интенсивной междисциплинарной коммуникации. Актуальность работы определялась высокой скоростью терминологического обновления, разнообразием номинативных моделей, неоднородностью нормотворческих практик, а также необходимостью создания воспроизводимой методики корпусного описания терминов для задач терминографии, стандартизации и профессионального перевода.

Объектом исследования выступила терминология предметной области «искусственный интеллект» в научном дискурсе, репрезентированная в корпусе высокорейтинговых научных статей и соотнесенная с действующими международными стандартами ISO/IEC. **Предмет** исследования составили структурно-семантические особенности терминологических единиц предметной области в целом «искусственный интеллект», проявляющиеся в их морфологической, синтаксической и морфолого-синтаксической организации, а также в процессах семантической деривации (эпонимизация, метонимизация, метафоризация, формирование синонимии и полисемии), выявляемые средствами корпусной лингвистики на материале английского языка.

Для определения структурно-семантических особенностей терминосистемы предметной области «искусственный интеллект» и установления ее внутренних закономерностей была разработана и апробирована воспроизводимая методика корпусного анализа, обеспечившая систематизированную обработку репрезентативного эмпирического материала, отражающего когнитивные, деривационные и дискурсивные процессы, лежащие в основе формирования терминологии ИИ. Уникальность предложенного подхода заключается в интеграции инструментов корпусного анализа на всех стадиях исследования – от первичного отбора материала до интерпретации семантических сдвигов и диагностики контекстной вариативности.

Первичная эмпирическая база исследования формировалась как авторский корпус из 897 англоязычных рецензируемых исследовательских статей за 2020–2024 годы, отобранных из журналов первой категории («Белый список», 1-й уровень). Корпус включал полнотекстовые публикации по тематике искусственного интеллекта, в которых систематически фиксируются описания моделей, методов и данных, что обеспечивает релевантность материала задачам терминологического анализа и сопоставимость жанрово-регистровых характеристик. Инструментарием для получения вторичного экспериментального материала стал корпусный менеджер AntConc, выбор которого был обоснован в результате сравнения с аналогичными платформами для проведения корпусного анализа. В ходе двухуровневой статистической обработки, включающей фильтрацию по частотному порогу (≥ 10 случаев на миллион токенов) и анализ с применением статистических метрик (Mutual Information, Log-Likelihood), были выделены релевантные частотные и ключевые единицы, зафиксированы устойчивые словосочетания (коллокации). По итогам терминографической верификации составлен репрезентативный реестр из 1648 терминологических единиц, ставший эмпирической основой для комплексного лингвистического анализа.

В работе показано, что сочетание корпусно-ориентированных процедур (AntConc: Word List, KWIC, Collocate) с количественно-статистическими метриками (нормализованные частоты, Mutual Information, Log-Likelihood), лингвистическим описанием и дефиниционно-компонентным анализом, верифицированным по ISO 704/1087 и актуальным отраслевым стандартам (ISO/IEC 22989, 23053, 23894), формирует надежный методологический каркас для изучения высокотехнологичных доменов, включая терминосистему ИИ. Его дополнение ономаσιологическим и семасиологическим анализом, дистрибутивно-коллокационным профилированием, а также элементами онтологизации (классы, отношения, ограничения) обеспечивает согласование лексической формы с понятийной структурой предметной области. Экспертная валидация и процедурная триангуляция укрепляют воспроизводимость методики и результатов корпусного анализа и обеспечивают межисследовательскую сопоставимость выводов.

Процедурная схема корпусного анализа с использованием AntConc, включающая статистическую триангуляцию (нормализованная частотность, Mutual

Information, Log-Likelihood), качественный KWIC-анализ и поиск по регулярным выражениям (Regex), обеспечивает объективный, верифицируемый и воспроизводимый отбор и разграничение терминологических единиц. Подход надежно дифференцирует ядро и периферию по критериям релевантности и устойчивости.

Проведенный анализ формально-структурных характеристик выявленных терминологических единиц показал, что в исследованной выборке термины-словосочетания составляют 73%, однословные – 27%. Ядро терминологической системы формируется преимущественно однословными наименованиями, включая устойчивые интернационализмы (например, *algorithm* (алгоритм), *data* (данные)), и устоявшимися многословными наименованиями (например, *neural network* (нейросеть), *language model* (языковая модель)), тогда как к периферии относятся инновационные дериваты и сложные образования (например, *large language model*, *LLM* (большая языковая модель), *reinforcement learning* (обучение с подкреплением), *human-in-the-loop* (человек в цикле)), мотивированные ядерными компонентами и обеспечивающие текущий терминологический прирост. Усложнение понятийного аппарата технологии коррелирует с ростом структурной сложности термина; в корпусном материале данная закономерность выявляется посредством специально разработанных регулярных шаблонов.

Установлено, что формирование и пополнение терминологии ИИ определяется взаимодействием семантических, морфологических, синтаксических и морфолого-синтаксических механизмов. В корпусе научных текстов доминируют многословные терминологические модели A+N, N+N, A+N+N, которые обеспечивают высокую степень спецификации: *neural network* (нейронная сеть), *model compression* (сжатие модели), *reinforcement learning agent* (агент обучения с подкреплением), тогда как продуктивная деривация системно кодирует типы концептов – процесс, результат, инструмент, сущность, свойство: *pre-training* (предварительное обучение), *embedding* (встраивание), *tokenizer* (токенизатор), *fairness* (справедливость). Привлечение общеязыковой лексики, в том числе с частичной семантической переинтерпретацией, ускоряет номинацию новых понятий: *agent* (агент), *prompt* (промпт); транстерминологизация из смежных и базовых дисциплин (математика, информатика, статистика) обеспечивает нормативно управляемое закрепление наименований: *kernel*

(ядро), *entropy* (энтропия). В совокупности эти механизмы поддерживают гибкость и при этом стандартизируемость терминологии ИИ.

Вариативность терминов (в форме синонимии и частичной полисемии) эффективно управляется за счет корпусной диагностики (KWIC, коллокационные профили, MI/LL) и применения стандартизованных предпочтительных наименований, что повышает уровень однозначности и согласованности в профессиональной коммуникации.

Таким образом, теоретико-методологическая база исследования опирается на обоснование междисциплинарного статуса современного терминоведения, требующего интеграции классических лингвистических подходов с когнитивно-дискурсивной парадигмой и методами корпусной лингвистики. Разработанная теоретическая рамка и ее процедурная реализация в специализированном протоколе позволили осуществить переход от фрагментарного описания к комплексному многоуровневому анализу терминологии, объединяющему функционально-коммуникативные и вычислительные аспекты лингвистического моделирования.

Системность и понятийная соотнесенность терминологического аппарата области «искусственный интеллект» подтверждаются разработкой операциональной классификации, которая легла в основу типологизации единиц по семантическим, формально-структурным и мотивационным параметрам. Особая роль в процессе концептуализации и гармонизации данной терминосистемы отводится механизмам стандартизации: использование положений международных и отраслевых стандартов (ISO 704, 1087, 22989) в качестве внешних критериев верификации позволило обосновать принципы согласования дефиниций и фиксации предпочтительных номинаций в рамках терминографически релевантной модели.

Эмпирическая достоверность и воспроизводимость результатов обеспечены алгоритмизацией процесса выделения терминологических кандидатов, включающей автоматизированное извлечение, фильтрацию «шумовых» токенов и статистическую оценку сочетаемости методами Mutual Information и Log-Likelihood. Анализ конкордансов и коллокационных профилей позволил сформировать верифицированный реестр из 1648 единиц, послуживший объективной базой для исследования динамических процессов в профессиональном дискурсе и доказавший

эффективность применения корпусных методов для выявления терминологического ядра предметной области.

Интегрированная модель терминообразования в сфере искусственного интеллекта характеризуется доминированием многословных номинаций, отражающих синтаксически-композиционный способ спецификации понятий, при сохранении высокой продуктивности морфологической деривации и транстерминологизации. Исследование показало, что возникающие в системе проявления синонимии и полисемии носят управляемый характер и поддаются диагностике через дистрибутивный анализ. Сочетание различных каналов пополнения терминосистемы обеспечивает ее адаптивность к запросам профессиональной коммуникации, создавая условия для последующей нормативной кодификации и упорядочивания терминологической вариативности.

Предложенная в диссертации методика корпусного анализа терминологии обладает трансферабельностью и может быть использована как модель для исследования иных терминосистем быстроразвивающихся областей знания. Полученные результаты уточняют взаимосвязь когнитивных механизмов и структурно-семантических процедур терминообразования, а также позволяют более последовательно описывать процессы концептуализации и категоризации в профессиональном дискурсе. Данный подход стимулирует изучение когнитивных процедур через призму языковых форм, а также коммуникативных и референциальных аспектов терминологии.

Исследование также вносит вклад в типологию терминологических систем, так как полученные результаты актуальны для семантической и деривационной типологии, расширяют представление о роли морфологических, синтаксических и семантических механизмов в формировании терминологии; также можно говорить о ценности результатов исследования для корпусной лингвистики и лексикографической практики, учитывая разработанную методику корпусного анализа и созданный онлайн-гlossарий.

Кроме того, апробированная в работе методика применения корпусного менеджера AntConc расширила спектр исследовательских операций и усилила объективность результатов за счет стандартизируемых процедур и параметризуемых настроек анализа. Использованный формат корпусного

исследования, сочетающий автоматизированное извлечение терминов, статистическую верификацию (частотность, ключевость, ассоциативные меры) и последующее лингвистическое описание, обеспечивает мониторинг терминосистемы как в синхроническом срезе, так и при сопоставлении временных периодов. Одновременно анализ практики работы с AntConc выявляет баланс преимуществ и ограничений инструмента. С одной стороны, он позволяет получать результаты, согласованные с заранее заданными формальными критериями и воспроизводимыми параметрами поиска, с другой – не обеспечивает автоматического различения тонких семантических и прагматических оттенков, вследствие чего обязательным этапом остается ручная контекстная верификация и интерпретация.

Созданный в рамках исследования онлайн-гlossарий может использоваться в переводческой деятельности, при написании научной и технической документации, а также в образовательных целях. Разработанные принципы нормирования терминологии ИИ, учитывающие вариативность, синонимию и полисемию, способствуют унификации профессиональной коммуникации в данной области.

Перспективы дальнейшей работы видятся в диахроническом расширении корпуса для количественного моделирования эволюции терминов; межжанровом сопоставлении (патенты, техническая документация, учебные тексты); межъязыковом анализе для картирования путей заимствования; интеграции распределительных семантических моделей для автоматического различения значений и кластеризации синонимических гнезд; онтологическом связывании терминосистемы ИИ с формальными таксономиями смежных областей (статистика, НСИ и др.).

Таким образом, терминология предметной области «искусственный интеллект» предстает как сложная, динамически развивающаяся и при этом структурно упорядоченная система, отражающая концептуальную организацию соответствующего научно-технологического знания. Ее становление определяется взаимодействием внутриязыковых механизмов номинации и деривации с экстралингвистическими факторами технологического, институционального и социокультурного характера. Проведенное исследование показывает, что корпусно-ориентированный подход в сочетании с многоуровневым лингвистическим анализом позволяет инвентаризировать и описать терминологический массив, а также выявить закономерности его организации.

СПИСОК ЛИТЕРАТУРЫ

1. **Авербух, К. Я.** Общая теория термина / К. Я. Авербух. – Иваново : Ивановский государственный университет ; Центр социальной поддержки женщин и семьи, 2004. – 251 с. – ISBN 5-89729-059-8. – Текст : непосредственный.

2. **Алексеева, Л. М.** Вековой путь российского терминоведения / Л. М. Алексеева, С. Л. Мишланова. – Текст : непосредственный // Научный диалог. – 2021. – № 9. – С. 9–34. – DOI <https://doi.org/10.24224/2227-1295-2021-9-9-34>.

3. **Баланчуков, Д. А.** Введение графических заимствований терминологии из сферы искусственного интеллекта в классификацию неисконной лексики китайского языка / Д. А. Баланчуков, И. В. Чекулай. – Текст : непосредственный // Вопросы журналистики, педагогики, языкознания. – 2024. – № 43(4). – С. 486–495. – DOI <https://doi.org/10.52575/2712-7451-2024-43-4-486-495>.

4. **Бедненко, Ю. И.** Репрезентация концепта «искусственный интеллект» в Национальном корпусе русского языка / Ю. И. Бедненко. – Текст : непосредственный // Евразийский гуманитарный журнал. – 2025. – № 1. – С. 30–41.

5. **Бенвенист, Э.** Общая лингвистика / Э. Бенвенист ; пер. с франц. – Москва : Прогресс, 1974. – 446 с. – Текст : непосредственный.

6. **Бессонова, Е. В.** Методы лингвистического анализа терминологии / Е. В. Бессонова. – Текст : непосредственный // Термины в научной и учебной литературе : межвузовский сборник / под ред. В. Н. Немченко. – Горький : Горьковский государственный университет, 1988. – С. 98–105.

7. **Болдырев, Н. Н.** Когнитивная семантика. Введение в когнитивную лингвистику : Курс лекций / Н. Н. Болдырев. – 5-издание, переработанное и дополненное. – Тамбов : Тамбовский государственный университет имени Г.Р. Державина, 2021. – 236 с. – ISBN 978-5-00078-451-8. – Текст : непосредственный.

8. **Болдырев, Н. Н.** Фреймовая семантика как метод когнитивного анализа языковых единиц / Н. Н. Болдырев. – Текст : непосредственный // Проблемы современной филологии : межвузовский сборник научных трудов / отв. ред. Е. В. Алтабаева. – Мичуринск : Мичуринский государственный педагогический институт, 2000. – Вып. 1. – С. 36–45.

9. **Большакова, Е. И.** Методы и средства извлечения терминов из текстов для терминологических задач / Е. И. Большакова, В. В. Семак. – Текст : непосредственный // Программные продукты и системы. – 2025. – Т. 38, № 1. – С. 5–16. – DOI 10.15827/0236-235X.149.005-016.

10. **Брунов, А. В.** Топонимы в их отношении к терминам и детерминологическим единицам / А. В. Брунов. – Текст : непосредственный // Вестник МГГУ им. М. А. Шолохова. Серия: Филологические науки. – 2011. – № 2. – С. 61–68.

11. **Бручес, Е. П.** Метод автоматического извлечения терминов из научных статей на основе слабо контролируемого обучения / Е. П. Бручес, Т. В. Батура. – Текст : непосредственный // Вестник НГУ. Серия: Информационные технологии. – 2021. – Т. 19, № 2. – С. 5–16. – DOI 10.25205/1818-7900-2021-19-2-5-16.

12. **Бузальская, Е. В.** Научная статья: тенденции изменения модели / Е. В. Бузальская. – Текст : непосредственный // Жанры речи. – 2021. – № 2(30). – С. 90–100. – DOI 10.18500/2311-0740-2021-2-30-90-100.

13. **Бурдина, О. Б.** Социокогнитивный контекст развития фармакопейной терминологии / О. Б. Бурдина. – Текст : непосредственный. // Когнитивные исследования языка. – 2023. – № 4(55). – С. 189-192.

14. **Буянова, Л. Ю.** Термин как единица логоса / Л. Ю. Буянова. – Москва : Наука, 2012. – 218 с. – ISBN 978-5-9765-1133-018. – Текст : непосредственный.

15. **Буянова, Л. Ю.** Термин как когнитема: эволюционно-прагматический аспект / Л. Ю. Буянова, Д. И. Новоселецкая. – Текст : непосредственный // Гуманитарные исследования. – 2021. – № 4(80). – С. 6-10. – DOI 10.21672/1818-4936-2021-80-4-006-010.

16. **Винокур, Г. О.** О некоторых явлениях словообразования в русской технической терминологии / Г. О. Винокур. – Текст : непосредственный // Труды Московского института истории, философии и литературы. Филологический факультет. Т. V : сборник статей по языковедению / под ред. М. В. Сергиевского, Д. Н. Ушакова, Р. О. Шор. – Москва : Московский институт истории, философии и литературы, 1939. – С. 3–54.

17. **Гак, В. Г.** Языковые преобразования / В. Г. Гак. – Москва : Языки русской культуры, 1998. – 763 с. – ISBN 5-7859-0063-7. – Текст : непосредственный.

18. **Герд, А. С.** Формирование терминологической структуры русского биологического текста / А. С. Герд. – Ленинград : Ленинградский государственный университет, 1981. – 112 с. – Текст : непосредственный.

19. **Головин, Б. Н.** Лингвистические основы учения о терминах / Б. Н. Головин, Р. Ю. Кобрин. – Москва : Высшая школа, 1987. – 105 с. – Текст : непосредственный.

20. **Гофман, И.** Анализ фреймов: эссе об организации повседневного опыта / И. Гофман. – Москва : Институт социологии РАН, 2004. – 752 с. – ISBN 5-93947-011-4. – Текст : непосредственный.

21. **Гречко, В. А.** Лексическая синонимика современного русского литературного языка / В. А. Гречко. – Саратов : Саратовский государственный университет, 1987. – 153 с. – Текст : непосредственный.

22. **Гринеv, С. В.** Введение в терминоведение / С. В. Гринеv. – Москва : Московский Лицей, 1993. – 309 с. – ISBN 5-88564-052-1. – Текст : непосредственный.

23. **Гринеv-Гринеvич, С. В.** Еще раз к вопросу об определении термина / С. В. Гринеv-Гринеvич, Э. А. Сорокина, М. А. Молчанова. – Текст : непосредственный // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. – 2022. – № 3. – С. 710–729. – DOI <https://doi.org/10.22363/2313-2299-2022-13-3-710-729>.

24. **Гринеv-Гринеvич, С. В.** Терминоведение : учебное пособие для студентов высших учебных заведений / С. В. Гринеv-Гринеvич. – Москва : Академия, 2008. – 304 с. – ISBN 978-5-7695-4951-9. – Текст : непосредственный.

25. **Гришин, М. М.** Особенности английской терминологии искусственного интеллекта / М. М. Гришин, Д. В. Крутяков, Н. В. Струнина – Текст : непосредственный // Язык как основа современного межкультурного взаимодействия : материалы X международной научно-практической конференции (Пенза, 30–31 октября 2023 г.) / отв. ред. Д. Н. Жаткин, И. В. Куликова. – Пенза : Пензенский государственный технологический университет, 2023. – С. 65–72.

26. **Гумбольдт, В. фон.** Избранные труды по языкознанию / В. фон Гумбольдт ; пер. с нем. ; ред. и предисл. Г. В. Рамишвили. – 2-е изд. – Москва : Прогресс, 2000. – 396 с. – ISBN 5-01-004661-X. – Текст : непосредственный.

27. **Давидюк, Т. И.** Корпуса и корпусные исследования языков Российской Федерации / Т. И. Давидюк, А. А. Кибрик, Д. Д. Мордашова. – Текст :

непосредственный // Вестник Российской академии наук. – 2024. – Т. 94, № 9. – С. 804–813.

28. **Даниленко, В. П.** Русская терминология: опыт лингвистического описания / В. П. Даниленко. – Москва : Наука, 1977. – 246 с. – Текст : непосредственный.

29. **Дронова, Е. М.** Семантические характеристики аллюзивных имен собственных / Е. М. Дронова. – Текст : непосредственный // Язык, коммуникация и социальная среда. – 2006. – Вып. 4. – С. 184–189.

30. **Егорова, О. А.** Специфика мифологической основы волшебного вымысла в сказках Британских островов / О. А. Егорова. – Текст : непосредственный // Вестник МГУ. Серия 19: Лингвистика и межкультурная коммуникация. – 2014. – № 1. – С. 87–97.

31. **Зарицкая, Л. А.** Лексикографические критерии составления глоссария англоязычных терминов для будущего дизайнера / Л. А. Зарицкая. – Текст : непосредственный // Вестник Оренбургского государственного университета. – 2013. – № 11(160). – С. 174–180.

32. **Захаров, В. П.** Корпусная лингвистика : учебно-методическое пособие / В. П. Захаров. – Санкт-Петербург : Санкт-Петербургский государственный университет, 2005. – 48 с. – Текст : непосредственный.

33. **Звегинцев, В. А.** Очерки по общему языкознанию / В. А. Звегинцев. – Москва : Московский государственный университет, 1962. – 384 с. – Текст : непосредственный.

34. **Зиньковская, А. В.** Психолингвистические предпосылки формирования неологизмов (на примере единиц периода пандемии коронавирусной инфекции COVID-19) / А. В. Зиньковская, А. А. Сахно. – Текст : непосредственный // Ученые записки Новгородского государственного университета. – 2021. – № 5(38). – С. 553–556.

35. **Иконникова, В. А.** Перспективы лексикографической фиксации формирующихся терминологий (на примере нейролингвистического онлайн-глоссария) / В. А. Иконникова, Е. В. Глушко, А. С. Галюченко. – Текст : непосредственный // Вопросы лексикографии. – 2025. – № 36. – С. 45–66.

36. **Казарина, С. Г.** Стандартизационные процессы в терминологии / С. Г. Казарина, Н. Б. Казарина. – Текст : непосредственный // Физическая культура, спорт – наука и практика. – 2008. – № 4. – С. 48–50.

37. **Какзанова, Е. М.** Представление терминов-эпонимов в юридических словарях / Е. М. Какзанова. – Текст : непосредственный // Вестник Курганского государственного университета. – 2019. – № 2(53). – С. 63–67.

38. **Калиновская, В. В.** Лингвокультурологический подход как способ понимания иноязычной правовой системы / В. В. Калиновская. – Текст : непосредственный // Вестник Университета имени О. Е. Кутафина. – 2017. – № 11(39). – С. 67–71. – DOI <https://doi.org/10.17803/2311-5998.2017.39.11.067-071>.

39. **Капанадзе, А.** Взаимодействие терминологической и общеупотребительной лексики / А. Капанадзе. – Текст : непосредственный // Развитие лексики современного русского языка / под ред. Е. А. Земской, Д. Н. Шмелева. – Москва : Наука, 1965. – С. 23–36.

40. **Карасик, В. И.** О типах дискурса / В. И. Карасик. – Текст : непосредственный // Языковая личность : институциональный и персональный дискурс : сборник научных трудов / под ред. В. И. Карасика. – Волгоград : Перемена, 2000. – С. 5–20.

41. **Карасик, В. И.** Языковой круг: личность, концепты, дискурс / В. И. Карасик. – Москва : Гнозис, 2004. – 390 с. – ISBN 5-7333-0143-0. – Текст : непосредственный.

42. **Квитко, И. С.** Термин в научном документе / И. С. Квитко. – Львов : Вища школа, 1976. – 125 с. – Текст : непосредственный.

43. **Клементьева, А. А.** К вопросу о функционировании термина «искусственный интеллект» в современном научном и публицистическом дискурсе / А. А. Клементьева. – Текст : непосредственный // Мир русского слова. – 2022. – № 4. – С. 14–23. – DOI [10.24412/1811-1629-2022-4-14-23](https://doi.org/10.24412/1811-1629-2022-4-14-23).

44. **Козлова, Н. В.** Лингвистические корпуса: определение основных понятий и типология / Н. В. Козлова. – Текст : непосредственный. // Вестник НГУ. Лингвистика и межкультурная коммуникация. – 2013. – № 1. – С. 79–88.

45. **Козловская, Н. В.** Транстерминологизация в сфере искусственного интеллекта: к постановке вопроса о субтерминологии / Н. В. Козловская, Ю. В. Сложеникина. – Текст : непосредственный // Art Logos. – 2023. – № 3(24). – С. 98–118.

46. **Кондратюкова, Л. К.** Заимствования и интернационализмы в терминологии англоязычной компьютерной техники / Л. К. Кондратюкова. – Текст : непосредственный // Вестник терминологии. – 2012. – № 5. – С. 45–60.

47. **Кононенко, А. П.** Лингвистический потенциал компьютерных технологий в современной филологии / А. П. Кононенко, Л. А. Недосека. – Текст : непосредственный // Гуманитарные и социальные науки. – 2023. – Т. 97, № 2. – С. 50–54.

48. **Красных, В. В.** Анализ дискурса в свете концепции фрейм-структур сознания / В. В. Красных. – Текст : непосредственный // Культурные слои во фразеологизмах и в дискурсивных практиках : сборник статей / отв. ред. В. Н. Телия. – Москва : Языки славянской культуры, 2004. – С. 243–250.

49. **Кузнецова, А. В.** Семиозис топонимики в пространстве городской идентичности / А. В. Кузнецова, И. А. Петрулевич. – Текст : непосредственный // Гуманитарные и социальные науки. – 2021. – Т. 89, № 6. – С. 72–78. – DOI 10.18522/2070-1403-2021-89-6-72-78.

50. **Кузнецова, Д. А.** Особенности перевода терминологии в области искусственного интеллекта / Д. А. Кузнецова, Т. С. Фуканчик. – Текст : электронный // Гагаринские чтения – 2023 : сборник тезисов докладов XLIX Международной молодежной научной конференции (Москва, 11–14 апреля 2023 г.). – Москва : Перо, 2023. – С. 808–810. – URL: <https://gagarin.mai.ru/files/2023/abstracts2023.pdf> (дата обращения: 14.07.2025).

51. **Куприков, Н. М.** Вопросы стандартизации в сфере искусственного интеллекта / Н. М. Куприков, Е. А. Башкирова. – Текст : непосредственный // Компетентность. – 2022. – № 3. – С. 14–18.

52. **Кутина, Л. Л.** Языковые процессы, возникающие при становлении научных терминологических систем / Л. Л. Кутина. – Текст : непосредственный // Лингвистические проблемы научно-технической терминологии : материалы совещания, проведенного АН СССР в Ленинграде (30 мая – 2 июня 1967 г.) / отв. ред. С. Г. Бархударов. – Москва : Наука, 1970. – С. 82–94.

53. **Лейчик, В. М.** Динамика термина (три возраста термина) / В. М. Лейчик. – Текст : непосредственный // Non multum, sed multa = Немного о многом. У когнитивных истоков современной терминологии : сборник научных трудов в честь В. Ф. Новодрановой / редкол.: Е. С. Кубрякова, В. З. Демьянков, А. Г. Кочкарева. – Москва : Авторская академия, 2010. – С. 351–359.

54. **Лейчик, В. М.** Место терминологии в системе современных наук (к постановке вопроса) / В. М. Лейчик. – Текст : непосредственный // Научно-техническая информация. – 1969. – № 8. – С. 5–8.

55. **Лейчик, В. М.** Обсуждение проблем эпонимии в современной науке / В. М. Лейчик. – Текст : непосредственный // Язык и право: актуальные проблемы взаимодействия : сборник статей / отв. ред. В. Ю. Меликян. – Ростов-на-Дону : Донское книжное издательство, 2011. – С. 134–142.

56. **Лейчик, В. М.** Терминоведение: предмет, методы, структура / В. М. Лейчик. – 5-е изд. – Москва : ЛЕНАНД, 2022. – 248 с. – ISBN 978-5-9710-9188-2. – Текст : непосредственный.

57. **Лемов, А. В.** Система, структура и функционирование научного термина (на материале русской лингвистической терминологии) / А. В. Лемов. – Саранск : Мордовский государственный университет, 2000. – 192 с. – ISBN 5-7103-0498-0. – Текст : непосредственный.

58. **Лобач, Е. А.** Имя собственное в немецкой математической терминологии / Е. А. Лобач. – Текст : непосредственный // Подготовка и использование научно-технических словарей в системе информационного обеспечения : тезисы докладов Всесоюзной конференции (Москва, 14–16 октября 1986 г.). – Москва : Русский язык, 1986. – С. 90–92.

59. **Лосев, А. Ф.** Философия имени / А. Ф. Лосев. – Москва : Академический проект, 2009. – 300 с. – ISBN 978-5-8291-1044-4. – Текст : непосредственный.

60. **Лотте, Д. С.** Вопросы заимствования и упорядочения иноязычных терминов и терминоположений / Д. С. Лотте. – Москва : Наука, 1982. – 149 с. – Текст : непосредственный.

61. **Лотте, Д. С.** Основы построения научно-технической терминологии: Вопросы теории и методики / Д. С. Лотте. – Москва : АН СССР, 1961. – 160 с. – Текст : непосредственный.

62. **Максимова, И. В.** Лексикографическая типология терминологических словарей: когнитивный вектор развития (на материале терминологических словарей русского, английского и французского языков предметной области «военная авиация») / И. В. Максимова, Э. П. Шпальченко. – Текст : непосредственный // Филологические науки. Вопросы теории и практики. – 2023. – № 2. – С. 615–623.

63. **Мальцева, Е. И.** Семантические процессы при трансфере *verba dicendi* в язык искусства / Е. И. Мальцева. – Текст : непосредственный // Вестник Московского государственного лингвистического университета. Серия: Гуманитарные науки. – 2021. – № 9(851). – С. 148–161. – DOI 10.52070/2542-2197_2021_9_851_148.

64. **Манерко, Л. А.** Основополагающие методы когнитивно-дискурсивного терминоведения / Л. А. Манерко. – Текст : непосредственный // Преподаватель XXI век. – 2024. – № 2, ч. 2. – С. 390–405. – DOI 10.31862/2073-9613-2024-2-390-405.

65. **Морозова, Л. А.** Терминознание: основы и методы / Л. А. Морозова. – Москва : Прометей, 2004. – 143 с. – ISBN 5-7042-1407-X. – Текст : непосредственный.

66. **Мурясов, Р. З.** Мифонимы в системе языка / Р. З. Мурясов. – Текст : непосредственный // Вестник Башкирского университета. – 2015. – Т. 20, № 3. – С. 952–956.

67. **Мусаева, А. С.** Терминообразование в сфере искусственного интеллекта / А. С. Мусаева. – Текст : непосредственный // Верхневолжский филологический вестник. – 2022. – № 2(29). – С. 166–173.

68. **Нерубленко, Д. В.** Антропонимы как маркеры вторичной номинации в научно-техническом дискурсе / Д. В. Нерубленко. – Текст : непосредственный // Актуальные проблемы филологии : материалы международной научно-практической конференции (Екатеринбург, 28 апреля 2022 г.) / под ред. И. А. Семухиной. – Екатеринбург : Уральский государственный педагогический университет, 2022. – С. 383–389.

69. **Новинская, Н. В.** Классификация эпонимических названий по семантическому признаку: семантические поля / Н. В. Новинская. – Текст : непосредственный // Вестник Астраханского государственного технического университета. – 2005. – № 5. – С. 147–154.

70. **Новинская, Н. В.** Термины-эпонимы в языке науки / Н. В. Новинская. – Текст : непосредственный // Русистика. – 2013. – № 4. – С. 34–38.

71. **Палийчук, Д. А.** Корпусные технологии в лингвистических исследованиях / Д. А. Палийчук. – Текст : электронный // Гуманитарные исследования. История и филология. – 2022. – № 6. – С. 34–41. – URL: <https://cyberleninka.ru/article/n/korpusnyetehtnologii-v-lingvisticheskikh-issledovaniyah-1> (дата обращения: 12.02.2026).

72. **Переверзева, Н. А.** Метафорическая деперсонификация в современном немецком языке / Н. А. Переверзева – Текст : непосредственный // Вестник

Челябинского государственного университета. Филология. Искусствоведение. – 2010. – № 11(192), вып. 42. – С. 100–104.

73. **Петрова, И. М.** Современные цифровые технологии в лингвистических исследованиях : учебное пособие для обучающихся по направлению «Лингвистика» / И. М. Петрова, А. М. Иванова, В. В. Никитина. – Москва : Языки народов мира, 2022. – 259 с. – ISBN 978-5-6048046-8-1. – Текст : непосредственный.

74. **Петушков, В. П.** Лингвистика и терминоведение / В. П. Петушков. – Текст : непосредственный // Терминология и норма : О языке терминологических стандартов : сборник статей / отв. ред. В. П. Даниленко. – Москва : Наука, 1972. – С. 102–116.

75. **Пиотровский, Р. Г.** К вопросу об изучении термина / Р. Г. Пиотровский. – Текст : непосредственный // Ученые записки Ленинградского государственного университета. – 1952. – № 161. – С. 21–36.

76. **Плунгян, В. А.** Корпусная лингвистика на современном этапе / В. А. Плунгян. – Текст : непосредственный // Вестник Российской академии наук. – 2024. – Т. 94, № 9. – С. 787–794. – DOI 10.31857/S0869587324090018.

77. **Полубиченко, Л. В.** Англоязычные паремиологические словари в электронно-цифровую эпоху: проблемы и перспективы / Л. В. Полубиченко, З. Р. Алхастова. – Текст : непосредственный // Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация. – 2024. – № 3. – С. 97–112. – DOI 10.55959/MSU-2074-1588-19-27-3-7.

78. **Пригожин, И.** Порядок из хаоса. Новый диалог человека с природой / И. Пригожин, И. Стенгерс. – Москва : Прогресс, 1986. – 432 с. – Текст : непосредственный.

79. **Пронина, Р. Ф.** Перевод английской научно-технической литературы : учебное пособие для вузов / Р. Ф. Пронина. – 3-е изд., испр. и доп. – Москва : Высшая школа, 1986. – 174 с. – Текст : непосредственный.

80. **Реформатский, А. А.** Термин как член лексической системы / А. А. Реформатский. – Текст : непосредственный. // Проблемы структурной лингвистики : сборник статей / отв. ред. С. К. Шаумян. – Москва : Наука, 1967. – С. 103–125.

81. **Савчук, С. О.** Национальный корпус русского языка 2.0: новые возможности и перспективы развития / С. О. Савчук, Т. А. Архангельский, А. А. Бонч-Осмоловская,

О. В. Дони́на, Ю. Н. Кузнецова, О. Н. Ляшевская, Б. В. Орехов, М. В. Подрядчикова. – Текст : непосредственный // Вопросы языкознания. – 2024. – № 2. – С. 7–34. – DOI 10.31857/0373-658X.2024.2.7-34.

82. **Серкова, Н. И.** О сходстве и различиях двух терминосистем (сопоставительный анализ) / Н. И. Серкова. – Текст : непосредственный // Язык и культура : вопросы современной филологии и методики обучения языкам в вузе : материалы VII Всероссийской научно-практической конференции с международным участием (Хабаровск, 21 мая 2020 г.) / отв. ред. И. Ф. Уманец. – Хабаровск : Тихоокеанский государственный университет, 2020. – С. 266–270.

83. **Солнцева, А. В.** Эпонимы в спортивном дискурсе: структурно-семантические характеристики / А. В. Солнцева. – Текст : непосредственный // Романские языки: прагматика и дискурс (способы выражения и содержание) : материалы Международной научной конференции (Москва, 28–29 июня 2022 г.) / отв. ред. И. В. Скуратов. – Москва : Московский государственный областной университет, 2022. – С. 194–196.

84. **Сулейманова, О. А.** Принципы и методы лингвистических исследований / О. А. Сулейманова, М. А. Фомина, И. В. Тивьяева. – 2-е изд., перераб. и доп. – Москва : Языки Народов Мира, 2020. – 352 с. – ISBN 978-5-89191-091-1. – Текст : непосредственный.

85. **Суперанская, А. В.** Общая теория имени собственного / А. В. Суперанская. – 4-е изд., стереотип. – Москва : URSS, 2023. – 366 с. – ISBN 978-5-397-06922-9. – Текст : непосредственный.

86. **Суперанская, А. В.** Общая терминология: вопросы теории / А. В. Суперанская, Н. В. Подольская, Н. В. Васильева ; отв. ред. Т. Л. Канделаки. – 6-е изд. – Москва : URSS, 2012. – 243 с. – ISBN 978-5-397-02414-3. – Текст : непосредственный.

87. **Сычев, М. Ф.** К вопросу о соответствии современного научного журнала международным стандартам качества / М. Ф. Сычев, А. В. Загребельный. – Текст : непосредственный // Экономические и социальные перемены : факты, тенденции, прогноз. – 2013. – № 6(30). – С. 201–206.

88. **Татаринов, В. А.** Теория терминоведения : в 3 т. – Т. 1 : Теория термина: история и современное состояние / В. А. Татаринов. – Москва : Московский лицей, 1996. – 311 с. – ISBN 5-88564-125-0. – Текст : непосредственный.

89. **Теоретические и прикладные аспекты корпусных исследований** : сборник научных трудов / отв. ред. Л. А. Кочетова. – Волгоград : Волгоградский государственный университет, 2016. – 84 с. – ISBN 978-5-9669-1633-6. – Текст : непосредственный.

90. **Термины и понятия искусственного интеллекта в лингвистическом освещении** / А. С. Мусаева, Ю. В. Сложеникина, Л. М. Гареева [и др.]. – Москва : Спутник+, 2024. – 193 с. – ISBN 978-5-9973-6887-6. – Текст : непосредственный.

91. **Толикина, Е. Н.** Некоторые лингвистические проблемы изучения термина / Е. Н. Толикина. – Текст : непосредственный // Лингвистические проблемы научно-технической терминологии : материалы совещания, проведенного АН СССР в Ленинграде (30 мая – 2 июня 1967 г.) / отв. ред. С. Г. Бархударов. – Москва : Наука, 1970. – С. 53–67.

92. **Флоренский, П. А.** У водоразделов мысли / П. А. Флоренский. – Москва : Правда, 1990. – 448 с. – Текст : непосредственный.

93. **Фомина, И. Н.** Англоязычная терминология искусственного интеллекта: состав и пути пополнения / И. Н. Фомина. – Текст : непосредственный // Язык и культура в эпоху интеграции научного знания и профессионализации образования. – 2022. – № 3-1. – С. 230–235.

94. **Циткина, Ф. А.** Терминология и перевод (к основам сопоставительного терминоведения) / Ф. А. Циткина. – Львов : Вища школа, 1988. – 156 с. – ISBN 5-11-000558. – Текст : непосредственный.

95. **Чернявская, В. Е.** Дискурс как объект лингвистических исследований / В. Е. Чернявская. – Текст : непосредственный // Текст и дискурс. Проблемы экономического дискурса : сборник научных статей / отв. ред. В. Е. Чернявская. – Санкт-Петербург : Санкт-Петербургский государственный университет экономики и финансов, 2001. – С. 11–22.

96. **Чернявская, В. Е.** Интерпретация научного текста / В. Е. Чернявская. – 5-е изд. – Москва : URSS ; ЛИБРОКОМ, 2010. – 128 с. – ISBN 978-5-397-00914-0. – Текст : непосредственный.

97. **Чернявская, В. Е.** Научно-исследовательская статья как вербализация нового научного результата (на материале лингвистики) / В. Е. Чернявская – Текст : непосредственный // Жанры речи. – 2016. – № 1(3). – С. 56–64.

98. **Чжао, Ц.** Сопоставительный анализ терминов подгруппы «Бионанотехнология» в терминосистеме биомедицинской инженерии в русском, английском и китайском языках / Ц. Чжао. – Текст : непосредственный // Вестник Чувашского государственного педагогического университета им. И. Я. Яковлева. – 2022. – № 2(115). – С. 139–146. – DOI 10.37972/chgpru.2022.115.2.020.

99. **Шарафутдинова, Н. С.** О понятиях «Терминология», «Терминосистема» и «Терминополе» / Н. С. Шарафутдинова. – Текст : непосредственный // Филологические науки. Вопросы теории и практики. – 2016. – № 6-3(60). – С. 168–171.

100. **Шелов, С. Д.** Вопросы толкования специальной лексики в толковых словарях (из истории теории и практики отечественной лексикографии) / С. Д. Шелов. – Текст : непосредственный // Труды Института русского языка им. В. В. Виноградова. – 2014. – № 1. – С. 434–453.

101. **Шелов, С. Д.** Некоторые новые результаты науки о терминах / С. Д. Шелов. – Текст : непосредственный // Слово и термин : сборник статей. – Москва : Институт русского языка им. В. В. Виноградова РАН, 2024. – С. 14–23.

102. **Шелов, С. Д.** Универсальный терминологический словарь: новое в русской терминографии / С. Д. Шелов, Т. Д. Четверикова. – Текст : непосредственный // Язык. Культура. Перевод: научные парадигмы и практические аспекты : материалы международной научно-практической конференции / под ред. В. А. Иконниковой, Н. Д. Паршиной. – Москва : МГИМО(У) МИД России, 2020. – С. 19–25.

103. **Щедровицкий, Г. П.** Избранные труды / Г. П. Щедровицкий. – Москва : Школа культурной политики, 1995. – 800 с. – ISBN 5-88969-001-9. – Текст : непосредственный.

104. **Щербина, С. И.** К вопросу о терминологической синонимии / С. И. Щербина. – Текст : непосредственный // Русский язык : история, диалекты, современность : сборник научных трудов / отв. ред. Л. Ф. Копосов. – Москва : Московский педагогический университет, 2000. – Вып. II. – С. 105–109.

105. **A Cognitive Linguistics View of Terminology and Specialized Language** / P. Faber (ed.). – Berlin : De Gruyter Mouton, 2012. – 307 p. – ISBN 9783110275568. – Текст : непосредственный.

106. **Adjali, O.** Building Comparable Corpora for Assessing Multi-Word Term Alignment / O. Adjali, E. Morin, P. Zweigenbaum – Текст : электронный // LREC 2022 –

13th Language Resources and Evaluation Conference (Marseille, Jun 2022). – Marseille, 2022. – P. 3103–3112. – URL: <https://hal.science/hal-03803881v1/document> (дата обращения: 17.09.2025).

107. **Agresti, A.** *Categorical Data Analysis* / A. Agresti. – 2nd ed. – Wiley, 2002. – 712 p. – ISBN 0-471-36093-7. – DOI <https://doi.org/10.1002/0471249688.scard>. – Текст : непосредственный.

108. **Andrei, N.** *Steepest Descent Methods* / N. Andrei. – Текст : электронный // *Modern Numerical Nonlinear Optimization*. – Springer, 2022. – P. 81–107. – URL: <https://www.springerprofessional.de/en/steepest-descent-methods/23615412> (дата обращения: 17.08.2025). – DOI 10.1007/978-3-031-08720-2_3.

109. **Anthony, L.** *AntConc (Version 4.3.1) : Computer software* / L. Anthony. – Текст : электронный. – Tokyo : Waseda University, 2024. – URL: <https://www.laurenceanthony.net/software/AntConc> (дата обращения: 11.08.2025).

110. **Anthony, L.** *Common Statistics Used in Corpus Linguistics* / L. Anthony. – Текст : электронный. – 2023. – URL: https://laurenceanthony.net/resources/statistics/common_statistics_used_in_corpus_linguistics.pdf (дата обращения: 11.08.2025).

111. **Arntz, R.** *Einführung in die Terminologearbeit* / R. Arntz, H. Picht. – 2. Aufl. – Hildesheim : Olms, 1991. – 344 S. – (Studien zu Sprache und Technik). – ISBN 3-487-07235-1. – Текст : непосредственный.

112. **Baker, C. F.** *The Berkeley FrameNet Project* / C. F. Baker, C. J. Fillmore, J. B. Lowe. – Текст : непосредственный // *Proceedings of the 36th Annual Meeting of ACL and 17th International Conference on Computational Linguistics (COLING-ACL 1998)*. – Montreal, Canada, 1998. – P. 86–90. – DOI 10.3115/980845.980860.

113. **Barcelona, A.** *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective* / A. Barcelona. – Berlin ; New York : Mouton de Gruyter, 2000. – 356 p. – ISBN 978-3-1-10-17556-1. – Текст : непосредственный. – DOI 10.1515/9783110894677.

114. **Barsalou, L. W.** *Grounded Cognition* / L. W. Barsalou. – Текст : непосредственный // *Annual Review of Psychology*. – 2008. – Vol. 59. – P. 617–645.

115. **Beheshti, Z.** *A Review of Population-based Meta-Heuristic Algorithms* / Z. Beheshti, S. M. H. Shamsuddin. – Текст : электронный // *International Journal of Advances in Soft Computing and its Applications*. – 2013. – Vol. 5, № 1. – P. 1–35. – URL: https://i-csrs.org/Volumes/ijasca/IJASCA07_REVISIED.pdf (дата обращения: 14.08.2025).

116. **Benjamini, Y.** Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing / Y. Benjamini, Y. Hochberg. – Текст : непосредственный // Journal of the Royal Statistical Society : Series B (Methodological). – 1995. – Vol. 57, № 1. – P. 289–300.

117. **Berger, O.** Синтаксические термины в русском и чешском языках: сопоставительный аспект (на материале выбранных терминов) / O. Berger. – Brno : Masaryk University Press, 2021. – 141 с. – ISBN 978-8-021-09807-7. – Текст : непосредственный.

118. **Biber, D.** Grammatical Complexity in Academic English: Linguistic Change in Writing / D. Biber, B. Gray. – Cambridge : Cambridge University Press, 2016. – 278 p. – ISBN 978-0-5-119-2077-6. – Текст : непосредственный. – DOI 10.1017/CBO9780511920776.

119. **Bonferroni, C. E.** Teoria statistica delle classi e calcolo delle probabilità / C. E. Bonferroni. – Текст : непосредственный // Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. – 1936. – Vol. 8. – P. 3–62. – DOI 10.4135/9781412961288.n455.

120. **Bowker, L.** Working with Specialized Language: A Practical Guide to Using Corpora / L. Bowker, J. Pearson. – London : Routledge, 2002. – 256 p. – ISBN 978-0-2-034-6925-5 – Текст : непосредственный. – DOI 10.4324/9780203469255.

121. **Brezina, V.** The Written British National Corpus 2014 – design and comparability / V. Brezina, A. Hawtin, T. McEnery. – Текст : непосредственный // Text & Talk. – 2021. – Vol. 41, № 5–6. – P. 595–615. – DOI 10.1515/text-2020-0052.

122. **Cabré, M. T.** Terminology: Theory, Methods and Applications / M. T. Cabré. – Amsterdam : John Benjamins Publishing, 1999. – 250 p. – ISBN 978-9-0-272-1634-2. – Текст : непосредственный. – DOI <https://doi.org/10.1075/tlrp.1?locatt=mode:legacy>.

123. **Campbell, M.** Deep Blue / M. Campbell, A. J. Hoane, Jr., F.-h. Hsu. – Текст : непосредственный // Artificial Intelligence. – 2002. – Vol. 134, № 1–2. – P. 57–83. – DOI [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1).

124. **Cazeaux, C.** Metaphor and Continental Philosophy: From Kant to Derrida / C. Cazeaux. – Taylor & Francis ; Routledge, 2007. – 240 p. – ISBN 978-0-2-033-2683-1. – Текст : непосредственный. – DOI 10.4324/9780203326831.

125. **Church, K. W.** Word Association Norms, Mutual Information, and Lexicography / K. W. Church, P. Hanks. – Текст : непосредственный // Computational Linguistics. – 1990. – Vol. 16, № 1. – P. 22–29.

126. **Committee on Publication Ethics (COPE).** Committee on Publication Ethics (COPE). Guidelines on good publication practice. – Текст : электронный // The COPE Report 2003 / ed. C. White. – London : BMJ Books, 2003. – P. 69–71. – URL: <https://www.ntnu.edu/documents/140096/1275971792/COPEguidelines.pdf/> (дата обращения: 25.12.2025).

127. **Dabbara Rajesh, S. N. P.** Lingua Franca 2.0: The Enduring Power of English Amidst Artificial Intelligence / S. N. P. Dabbara Rajesh. – Текст : непосредственный // International Journal of Language, Literature and Culture. – 2025. – Vol. 5, № 4. – P. 61–64. – DOI 10.22161/ijllc.5.4.9.

128. **Depecker, L.** Entre signe et concept. Éléments de terminologie générale / L. Depecker. – Paris : Presses Sorbonne Nouvelle, 2002. – 200 p. – ISBN 978-2-87854-234-9. – Текст : непосредственный.

129. **Diki-Kidiri, M.** Cultural Terminology: an Introduction to Theory and Method / M. Diki-Kidiri. – Текст : непосредственный // Theoretical Perspectives on Terminology : Explaining Terms, Concepts and Specialized Knowledge / P. Faber, M.-C. L’Homme (eds.). – Amsterdam : John Benjamins, 2022. – P. 197–216. – DOI 10.1075/tlrp.23.09dik.

130. **Dubichynskyi, V.** О лексикологии и лексикографии в Варшавском университете (XV Международная конференция «Современные проблемы лексикографии LEX-2021»). Варшавский университет. Кафедра русистики. Варшава, 20–21.05.2021 г.) / V. Dubichynskyi, M. Dąbrowska. – Текст : непосредственный // Studia Rossica Gedanensia. – 2021. – № 8. – С. 313–318. – DOI 10.26881/srg.2021.8.25.

131. **Dubuc, R.** Manuel pratique de terminologie / R. Dubuc. – 3e éd. – Québec : Brossard ; Linguatex, 1992. – 144 p. – ISBN 9782920324145. – Текст : непосредственный.

132. **Durkin, P.** Borrowed Words : A History of Loanwords in English / P. Durkin. – Oxford University Press, 2014. – 491 p. – ISBN 9780199574995. – Текст : непосредственный.

133. **Evert, S.** Corpora and Collocations / S. Evert. – Текст : непосредственный // Corpus Linguistics : An International Handbook / eds. A. Lüdeling, M. Kytö. – Berlin ; New York : De Gruyter Mouton, 2009. – P. 1212–1248. – DOI 10.1515/9783110213881.2.1212.

134. **Faber, P.** Frame-Based Terminology / P. Faber. – Текст : непосредственный // Theoretical Perspectives on Terminology : Explaining Terms, Concepts and Specialized Knowledge; P. Faber; M.-C. L’Homme (eds.). – Amsterdam ; Philadelphia : John Benjamins, 2022. – P. 353–376. – DOI 10.1075/tlrp.23.16fab.

135. **Fauconnier, G.** The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities / G. Fauconnier, M. Turner. – Basic Books, 2002. – 464 p. – ISBN 9780465087860. – Текст : непосредственный.

136. **Felber, H.** Terminology Manual / H. Felber. – Paris : UNESCO, Infoterm, 1984. – 426 p. – URL: <https://archive.org/details/terminology-manual-by-h.-felber> (дата обращения: 17.08.2025). – Текст : электронный.

137. **Feldman, J.** From Molecule to Metaphor: A Neural Theory of Language / J. Feldman. – MIT Press, 2006. – 384 p. – ISBN 0-262-06253-4. – Текст : непосредственный.

138. **Fillmore, C. J.** Frame Semantics and the Nature of Language / C. J. Fillmore. – Текст : непосредственный // Annals of the New York Academy of Sciences. – 1976. – Vol. 280, № 1. – P. 20–32. – DOI 10.1111/j.1749-6632.1976.tb25467.x.

139. **Firth, J. R.** Papers in Linguistics 1934–1951 / J. R. Firth. – London : Oxford University Press, 1957. – 233 p. – ISBN 9780197135488. – Текст : непосредственный.

140. **Fuertes-Olivera, P. A.** A Guide to Practical Online Lexicography / P. A. Fuertes-Olivera. – 1st ed. – Routledge, 2025. – 334 p. – ISBN 9781003426998 – Текст : непосредственный. – DOI 10.4324/9781003426998.

141. **Gibbs, R. W., Jr.** Speaking and Thinking with Metonymy / R. W. Gibbs, Jr. // Metonymy in Language and Thought; / K.-U. Panther, G. Radden (eds.). – Amsterdam ; Philadelphia : John Benjamins, 1999. – P. 61–76. – Текст : непосредственный.

142. **Goldberg, D. E.** Genetic Algorithms in Search, Optimization, and Machine Learning / D. E. Goldberg. – Reading, MA : Addison-Wesley, 1989. – 412 p. – ISBN 9780201157673. – Текст : непосредственный.

143. **Gómez-Pérez, A.** Inteligencia artificial y lengua española / A. Gómez-Pérez, S. Muñoz Machado. – Madrid : E.T.S. de Ingenieros Informáticos (UPM), 2023. – 146 p. – Текст : непосредственный.

144. **Grady, J. E.** Foundations of Meaning: Primary Metaphors and Primary Scenes : PhD thesis // Grady Joseph Edward. – UC Berkeley, 1997. – 299 p. – Текст : непосредственный.

145. **Gries, S. Th.** Corpus Linguistics: Quantitative Methods / S. Th. Gries // The Encyclopedia of Applied Linguistics / C. A. Chapelle (ed.). – John Wiley & Sons, 2024. – Текст : непосредственный. – DOI 10.1002/9781405198431.wbeal20003.

146. **Guilbert, L.** La créativité lexicale / L. Guilbert. – Paris : Larousse, 1975. – 285 p. – ISBN 9782030703403. – Текст : непосредственный.

147. **Guo, Z.** Conceptual Metaphor and Cognition: From the Perspective of the Philosophy of Language / Z. Guo. – Текст : непосредственный // Journal of Innovation and Development. – 2023. – Vol. 3, № 1. – P. 73–75. – DOI 10.54097/jid.v3i1.8424.

148. **Handbook of Terminology Management. Vol. 1: Basic Aspects of Terminology Management** / S. E. Wright, G. Budin (eds.). – Amsterdam ; Philadelphia : John Benjamins, 1997. – 370 p. – ISBN 9789027221544. – Текст : непосредственный.

149. **Hebb, D.** The Organization of Behavior / D. Hebb. – New York : Psychology Press, 2002. – 279 p. – ISBN 9781135631901. – Текст : непосредственный.

150. **Heller, L. G.** A Typology of Shortening Devices / L. G. Heller, J. Macris. – Текст : непосредственный // American Speech. – 1968. – Vol. 43, № 3. – P. 201–208.

151. **Hou, Y.** Pruning Translation of Logical and Accidental Polysemy in Traditional Chinese Medicine Terminology / Y. Hou. – Текст : непосредственный // Terminology. – 2025. – Vol. 31, № 2. – P. 311–334. – DOI 10.1075/term.24007.hou.

152. **Hržica, G.** The Role of Entrenchment and Schematisation in the Acquisition of Rich Verbal Morphology / G. Hržica, S. Košutar, T. B. Botica, P. Milin. – Текст : непосредственный // Cognitive Linguistics. – 2024. – DOI 10.1515/cog-2023-0022.

153. **Hyland, K.** Disciplinary Discourses: Social Interactions in Academic Writing / K. Hyland. – University of Michigan Press, 2004. – 211 p. – ISBN 9780582419049. – Текст : непосредственный.

154. **Jakobson, R.** Closing Statement: Linguistics and Poetics / R. Jakobson // Style in Language / T. A. Sebeok (ed.). – Cambridge : MIT Press, 1960. – P. 350–377. – Текст : непосредственный.

155. **Kageura, K.** Methods of Automatic Term Recognition: A Review / K. Kageura, B. Umino. – Текст : непосредственный // Terminology. – 1996. – Vol. 3, № 2. – P. 259–289.

156. **Kageura, K.** The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth / K. Kageura. – Amsterdam ; Philadelphia : John Benjamins, 2002. – 322 p. – ISBN 9789027223289. – Текст : непосредственный. – DOI 10.1075/tlrp.5.

157. **Kövecses, Z.** Metaphor in Culture: Universality and Variation / Z. Kövecses. – Cambridge University Press, 2005. – 314 p. – DOI 10.1017/CBO9780511614408. – ISBN 0-521-84447-9. – Текст : непосредственный.

158. **L’Homme, M.-C.** Lexical Semantics for Terminology: An Introduction / M.-C. L’Homme. – John Benjamins, 2020. – 263 p. – DOI 10.1075/tlrp.20. – ISBN 9789027204677. – Текст : непосредственный.

159. **Lakoff, G.** Metaphors We Live By / G. Lakoff, M. Johnson. – Chicago : University of Chicago Press, 1980. – 192 p. – ISBN 9780226470993. – Текст : непосредственный.

160. **Lakoff, G.** Women, Fire, and Dangerous Things: What Categories Reveal about the Mind / G. Lakoff. – Chicago : The University of Chicago Press, 1987. – DOI 10.7208/chicago/9780226471013.001.0001. – ISBN 978-0-226-46804-4. – Текст : непосредственный.

161. **Loanwords in the World’s Languages** : A Comparative Handbook / M. Haspelmath, U. Tadmor (eds.). – Berlin : De Gruyter Mouton, 2009. – 1081 p. – ISBN 9783110218435. – Текст : непосредственный. – DOI 10.1515/9783110218442.

162. **Mattiello, E.** Extra-Grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena / E. Mattiello. – Berlin ; Boston : De Gruyter Mouton, 2013. – 349 p. – ISBN 9783110295399. – Текст : непосредственный. – DOI 10.1515/9783110295399.

163. **McCarthy, J.** A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence / J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon. – 1955. – 13 p. – URL: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (дата обращения: 17.09.2025). – Текст : электронный.

164. **McCulloch, W.** A Logical Calculus of the Ideas Immanent in Nervous Activity / W. McCulloch, W. Pitts. – Текст : непосредственный // *Bulletin of Mathematical Biophysics*. – 1943. – № 5. – P. 115–133.

165. **McMahan, H. B.** Communication-Efficient Learning of Deep Networks from Decentralized Data / H. B. McMahan [et al.]. – Текст : непосредственный // *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Journal of Machine Learning Research*. – 2017. – Vol. 54. – P. 1273–1282.

166. **Meyer, I.** Extracting Knowledge-Rich Contexts for Terminography: A Conceptual and Methodological Framework / I. Meyer. – Текст : непосредственный // *Recent Advances in Computational Terminology* / D. Bourigault, C. Jacquemin, M.-C. L’Homme (eds.). – Amsterdam ; Philadelphia : John Benjamins, 2001. – P. 279–302.

167. **Minsky, M.** A Framework for Representing Knowledge / M. Minsky. – Текст : непосредственный // *The Psychology of Computer Vision* / P. Winston (ed.). – New York : McGraw-Hill, 1975. – P. 211–277.

168. **Moreno-Ortiz, A.** Keywords / A. Moreno-Ortiz. – Текст : непосредственный // *Making Sense of Large Social Media Corpora*. – Cham : Palgrave Macmillan, 2024. – P. 85–116. – DOI 10.1007/978-3-031-52719-7_4.

169. **National Research Council. Language and Machines: Computers in Translation and Linguistics (ALPAC Report).** – Washington, D.C. : National Academies Press, 1966. – 124 p. – Текст : непосредственный.

170. **O’Keeffe, A.** *Routledge Handbook of Corpus Linguistics* / A. O’Keeffe, M. McCarthy. – London : Routledge, 2010. – 682 p. – ISBN 9780415464895. – Текст : непосредственный. – DOI 10.4324/9780203856949.

171. **Peirsman, Y.** Metonymy as a Prototypical Category / Y. Peirsman, D. Geeraerts. – Текст : непосредственный // *Cognitive Linguistics*. – 2006. – Vol. 17, №. 3. – P. 269–316. – DOI 10.1515/COG.2006.007.

172. **Pérez-Paredes, P.** *Data-Driven Learning in and out of the Language Classroom* / P. Pérez-Paredes, A. Boulton. – Cambridge : Cambridge University Press, 2025. – 350 p. – ISBN 9781009511384. – Текст : непосредственный. – DOI <https://doi.org/10.1017/9781009511384>.

173. **Plag, I.** *Word-Formation in English* / I. Plag. – 2nd ed. – Cambridge : Cambridge University Press, 2018. – 245 p. – ISBN 9780521525633. – Текст : непосредственный.

174. **Polak, M. P.** Extracting Accurate Materials Data from Research Papers with Conversational Language Models and Prompt Engineering. – Текст : непосредственный / M. P. Polak, D. Morgan // *Nature Communications*. – 2024. – Vol. 15. – Article 1569. – DOI 10.1038/s41467-024-45914-8.

175. **Rondeau, G.** Introduction a la terminologie / G. Rondeau. – 2-e éd. – Québec : Gaëtan Morin, 1984. – 227 p. – ISBN 9782891051378. – Текст : непосредственный.

176. **Rosch, E.** Principles of Categorization / E. Rosch. – Текст : непосредственный // *Cognition and Categorization* / E. Rosch, B. B. Lloyd (eds.). – Hillsdale : Lawrence Erlbaum, 1978. – P. 27–48.

177. **Russell, S.** Artificial Intelligence: A Modern Approach / S. Russell, P. Norvig. – 4th ed. – Harlow : Pearson, 2021. – 1136 p. – ISBN 978-1292401133. – Текст : непосредственный.

178. **Sager, J. C.** A Practical Course in Terminology Processing / J. C. Sager. – Amsterdam ; Philadelphia : John Benjamins Publishing Company, 1990. – 253 p. – ISBN 9789027274342. – Текст : непосредственный. – DOI <https://doi.org/10.1075/z.44>.

179. **Samuel, A. L.** Some Studies in Machine Learning Using the Game of Checkers / A. L. Samuel. – Текст : непосредственный // *IBM Journal of Research and Development*. – 1959. – Vol. 3, №. 3. – P. 211–229.

180. **Selivan, L.** Corpus Linguistics and Vocabulary Teaching / L. Selivan. – Текст : непосредственный // *Demystifying Corpus Linguistics for English Language Teaching* / K. Harrington, P. Ronan (eds.). – Cham : Springer, 2023. – P. 139–161. – DOI 10.1007/978-3-031-11220-1_8.

181. **Shadrina, O.** Thinking out of the Box: Population-Based AI Terms through the Lens of Metaphor / O. Shadrina, O. Marunovich, D. Kuznetsov. – Текст : непосредственный // *The Journal of Teaching English for Specific and Academic Purposes*. – 2025. – Vol. 13, No 3. – P. 515–527. – DOI 10.22190/JTESAP250820039S.

182. **Sinclair, J.** Corpus, Concordance, Collocation / J. Sinclair. – Oxford : Oxford University Press, 1991. – 179 p. – ISBN 9780194371445. – (Describing English Language). – Текст : непосредственный.

183. **Smith, J.** Standard Language Ideology in AI-Generated Language / J. Smith, A. Wexler, L. Chen, R. Doe. – 2024. – URL: <https://arxiv.org/abs/2406.08726> (дата обращения: 11.08.2025). – Текст : электронный. – DOI 10.48550/arXiv.2406.08726.

184. **Sperber, D.** *Relevance: Communication and Cognition* / D. Sperber, D. Wilson. – Oxford : Blackwell, 1986. – 279 p. – ISBN 9780674754751. – Текст : непосредственный.

185. **Stefanowitsch, A.** *Corpus-Based Approaches to Metaphor and Metonymy* / A. Stefanowitsch, S. Th. Gries. – Текст : непосредственный // *Corpus-Based Approaches to Metaphor and Metonymy* / eds. A. Stefanowitsch, S. Th. Gries. – Berlin ; New York : De Gruyter Mouton, 2006. – P. 1–16. – (Trends in Linguistics. Studies and Monographs. Vol. 171).

186. **Suleimanova, O. A.** *The Learning and Educational Potential of Digital Tools in Humanities and Social Science* / O. A. Suleimanova, I. A. Guseinova, A. A. Vodyanitskaya. – DOI 10.17770/sie2020vol4.4851. – Текст : непосредственный // *Society. Integration. Education : Proceedings of the International Scientific Conference, Rēzekne (22–23 May 2020)*. – Rēzekne : Rēzeknes Tehnoloģiju akadēmija, 2020. – P. 657–669.

187. **Suleimanova, O. A.** *Using Big Data Experiments in Cognitive and Linguo-Cultural Research in English and Russian* / O. A. Suleimanova, I. M. Petrova. – Текст : непосредственный // *Journal of Siberian Federal University. Humanities and Social Sciences*. – 2020. – Vol. 13, № 3. – P. 385–393. – DOI 10.17516/1997-1370-0561.

188. **Temmerman, R.** *Towards New Ways of Terminology Description: The Sociocognitive Approach* / R. Temmerman. – Amsterdam ; Philadelphia : John Benjamins, 2000. – 258 p. – Текст : непосредственный. – DOI 10.1075/tlrp.3. – ISBN 9789027223265.

189. **Temmerman, R.** *Units of Understanding in Sociocognitive Terminology Studies* / R. Temmerman // *Theoretical Perspectives on Terminology : Explaining Terms, Concepts and Specialized Knowledge*; P. Faber, M.-C. L’Homme (eds.). – Amsterdam : John Benjamins, 2022. – P. 331–352. – (Terminology and Lexicography Research and Practice; Vol. 23). – ISBN 9789027211064. – Текст : непосредственный. – DOI 10.1075/tlrp.23.15tem.

190. **The Cambridge Handbook of Metaphor and Thought** / ed. by R. W. Gibbs, Jr. – Cambridge University Press, 2008. – 566 p. – ISBN 9780511816802. – Текст : непосредственный. – DOI 10.1017/CBO9780511816802.

191. **Thoiron, P.** *La terminologie, une question de termes* / P. Thoiron, H. Béjoint. – Текст : непосредственный // *Meta : journal des traducteurs*. – 2010. – Vol. 55, № 1. – P. 105–118. – DOI 10.7202/039605ar.

192. **Turing, A. M.** Computing Machinery and Intelligence / A. M. Turing. – Текст : непосредственный // Mind. – 1950. – Vol. 59, №. 236. – P. 433–460.

193. **Vaswani, A.** Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar [et al.]. – Текст : непосредственный // Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017) (Long Beach, 4–9 December 2017). – 2017. – P. 5998–6008.

194. **Vitanova-Ringaceva, A.** Zoonym Phraseological Units as Hate Speech in Youth Vocabulary / A. Vitanova-Ringaceva, D. Marolova, D. Kuzmanovska, B. Ivanova. – Текст : непосредственный // Education and New Developments; M. Carmo (ed.). – World Institute for Advanced Research and Science, 2024. – Vol. 1. – P. 185–189.

195. **Vyatkina, N.** Corpus-Based Applications in Language Teaching and Research: The Case of Data-Driven Learning of German / N. Vyatkina. – New York : Routledge, 2024. – 264 p. – ISBN 9781032122182. – Текст : непосредственный.

196. **Wiener, N.** Cybernetics: Or Control and Communication in the Animal and the Machine / N. Wiener. – 2nd revised ed. – Cambridge : MIT Press, 1961. – 212 p. – ISBN 9780262730099. – Текст : непосредственный.

197. **Wüster, E.** Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie / E. Wüster. – Wien : Springer, 1979. – 272 S. – ISBN 978-0387815428. – Текст : непосредственный.

198. **Wüster, E.** Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik / E. Wüster. – Berlin : VDI-Verlag, 1931. – 431 S. – Текст : непосредственный.

ПРИЛОЖЕНИЕ № 1

QR-код для доступа к онлайн-гlossарию терминов ИИ ([www. aiterms.ru](http://www.aiterms.ru))



ПРИЛОЖЕНИЕ № 2

QR-код для доступа к перечню лексикографических источников
(включая онлайн-платформы и репозитории), а также к списку
нормативно-правовых актов и стандартов



ПРИЛОЖЕНИЕ № 3

QR-код для доступа к дополнительным материалам, иллюстрирующим и расширяющим результаты исследования (корпусные примеры, классификации, таксономии)

