

О применении искусственного интеллекта и больших языковых моделей в области почвенно-экологических исследований (обзор)

© 2026. Н. А. Боков¹, аспирант, м. н. с.,

И. Г. Широких^{1, 2}, д. б. н., зав. лабораторией, в. н. с.,

¹Федеральный аграрный научный центр Северо-Востока им. Н. В. Рудницкого,
610007, Россия, г. Киров, ул. Ленина, д. 166а,

²Институт биологии Коми научного центра Уральского отделения
Российской академии наук,
167982, Россия, г. Сыктывкар, ул. Коммунистическая, д. 28,
e-mail: biotekhnologiya@fanc-sv.ru

В последние годы в научных исследованиях активно развиваются методы, основанные на использовании искусственного интеллекта (ИИ) и больших языковых моделей (БЯМ), которые способны обрабатывать и интерпретировать значительные объёмы текстовой и структурированной (таблицы, числовые и категориальные данные, данные датчиков и т. п.) информации. Экология почвы как область, характеризующаяся быстрым накоплением разнородных данных, включая результаты метагеномных анализов почвенных микробиомов, может потенциально извлечь существенную пользу от применения подобных инструментов. В настоящем обзоре рассмотрены основные направления использования ИИ в исследованиях в области почвенной экологии и микробиологии, включая анализ научной литературы, аннотирование и систематизацию микробиологических, агрохимических и агрофизических данных, а также создание моделей для прогнозирования на основе метагеномных данных почвенных микробиомов показателей плодородия и продуктивности сельскохозяйственных угодий. Обсуждаются ключевые преимущества данных подходов, их ограничения и возможные методологические риски. Несмотря на необходимость дальнейшего усовершенствования и оптимизации имеющихся к настоящему времени технологий ИИ и БЯМ, выполненный анализ литературы свидетельствует об их значительном потенциале в обеспечении дальнейшего развития почвенно-экологических исследований и достижении практических результатов в аграрном производстве.

Ключевые слова: искусственный интеллект, нейронные сети, машинное обучение, большие языковые модели, экология микроорганизмов, почвенный микробиом, метагеномика, аграрное производство.

On the application of artificial intelligence and large language models in soil and environmental research: a review

© 2026. N. A. Bokov¹ ORCID: 0000-0002-1000-1192²

I. G. Shirokikh^{1, 2} ORCID: 0000-0002-3319-2729²

¹Federal Agricultural Research Center of North-East named N. V. Rudnitsky,
166a, Lenina St., Kirov, Russia, 610007

²Institute of Biology of Komi Science Centre of the Ural Branch
of the Russian Academy of Sciences,
28, Kommunisticheskaya St., Syktyvkar, Russia, 167982,
e-mail: biotekhnologiya@fanc-sv.ru

In recent years, scientific research has seen the rapid development of methods based on artificial intelligence (AI) and large language models (LLMs), which are capable of processing and interpreting significant volumes of text and structured information, including tables, numerical and categorical data, and sensor data and other heterogeneous information, using specialized and multimodal AI systems. Soil ecology, as a field characterized by the rapid accumulation of heterogeneous data, including the results of metagenomic analyses of soil microbiomes, can potentially benefit significantly from the application of such tools. The aim of this review is to identify, analyze, and summarize current approaches to the use of AI and LLM technologies in soil and environmental research. The review examines the main areas of AI application, including scientific literature search and analysis, semantic search for relevant publications, citation mapping, text summarization, and the use of generative AI assistants. Particular attention is paid to the application of AI methods in the analysis of soil microbiomes and to the potential use of metagenomic data for predicting soil fertility and agricultural productivity. The challenges associated with adapting

AI methods originally developed for biomedical research to the specific characteristics of soil and environmental datasets are also considered. The key advantages of these approaches, their limitations, and potential methodological risks, including those related to data quality and the reliability of AI-generated information, are discussed. Despite the need for further improvement and optimization of currently available AI and LLM technologies, the literature review demonstrates their significant potential for advancing soil and environmental research and supporting practical applications in agriculture.

Keywords: artificial intelligence, neural networks, machine learning, large language models, microbial ecology, soil microbiome, metagenomics, agricultural production.

Почва представляет собой верхний тончайший слой (в среднем 1–2 м) земной коры, сформированный в результате почвообразовательного процесса и обладающий плодородием. В отличие от почвообразующих горных пород, почва характеризуется наличием органического вещества, активной биоты, развитой системой горизонтов и плодородием – способностью обеспечивать рост и развитие растений. Представление о почве как особом природном образовании, формирующемся в течение длительного периода времени при взаимодействии компонентов и факторов живой и неживой природы, было сформулировано в конце 19 века основоположником русской школы почвоведения В. В. Докучаевым [1, 2].

Почвенные ресурсы играют ключевую роль в развитии сельского хозяйства, функционировании экосистем и обеспечении продовольственной безопасности государства. Вместе с тем, почва – истощаемый и плохо возобновляемый ресурс. Образование новых слоёв почвы происходит зачастую медленнее, чем её деградация под действием природных и антропогенных факторов, что приводит к нарушению круговоротов воды и биогенных элементов, сокращению биологического разнообразия, ставит под угрозу сельское хозяйство [3, 4]. Наиболее значимый и заметный вклад в ухудшение ситуации вносит растениеводство. Большая часть территории России непригодна для возделывания сельскохозяйственных культур, поэтому регионы с благоприятными климатическими условиями испытывают большую нагрузку на почвенный покров, что ведёт к истощению и загрязнению почв и водных ресурсов. Внесение удобрений, использование ядохимикатов и нарушение целостности почвенных горизонтов в результате механической обработки, использование монокультур и ограниченные севообороты являются основными причинами негативных экологических последствий сельскохозяйственной практики: истощения почвенных ресурсов и деградации окружающей среды [5, 6], утраты биоразнообразия – главного показателя биологического здоровья почвы [7]. Устойчивые методы ведения сельского

хозяйства, сохранение и восстановление почв, а также меры по ограничению использования минеральных удобрений играют ключевую роль в поддержании здоровья почвы и продуктивности растениеводства [8, 9].

Фундаментальную роль в биогеохимических циклах, круговоротах биогенных элементов и устойчивости агроэкосистем в целом играют микроорганизмы [10–12]. Почвенная микробиология становится в последние десятилетия одной из ключевых областей биологических и экологических исследовательских проектов. Микробная биомасса и микробное разнообразие – это ключевые показатели здоровья почвы. Для их определения используются как традиционные методы, основанные на культивировании грибов и бактерий на селективных средах, так и самые передовые мультиомиксные подходы, значительно расширяющие разнообразие выявляемых таксонов, а также позволяющие прогнозировать функциональную структуру и метаболический профиль почвенного микробиома [13].

Одним из наиболее распространённых методов изучения таксономической структуры микробных сообществ является метабаркодирование, основанный на анализе выделяемой из почвы общей ДНК. Проведение анализа включает амплификацию с использованием полимеразной цепной реакции (ПЦР) специфических участков, таких как 16S рРНК для бактерий и ITS (Internal Transcribed Spacer – внутренний транскрибируемый спейсер) и/или 18S рРНК для грибов, за которой следует высокопроизводительное секвенирование, например, с использованием методов секвенирования нового поколения (Next Generation Sequencing – NGS) [14]. В исследованиях с помощью гено-молекулярных методов на показатели почвенного микробиома существенно влияют метод отбора проб, размер выборки, способы экстракции ДНК и методы биоинформатического анализа [15]. В числе недостатков данного подхода отмечают потенциальные искажения, возникающие во время экстракции и амплификации ДНК, трудности в идентификации живых/мёртвых микробных клеток и близкородственных

видов или подвидов, проблемы, связанные с неполнотой имеющихся в настоящее время баз геномных данных [16, 17]. Кроме того, присутствие ингибиторов ПЦР (гуминовые и фульвокислоты, тяжёлые металлы и полисахариды) в образцах почв может осложнить процесс амплификации и привести к неполной представленности некоторых таксонов [18]. Для преодоления названных ограничений разрабатываются новые методические подходы и совершенствуются существующие технологии. Например, для повышения таксономического разрешения рекомендуется комбинировать несколько генетических маркеров. А платформы PacBio (Pacific Bioscience) и Oxford Nanopore, разработанные для анализа протяжённых участков генома, позволяют получать данные о нуклеотидных последовательностях РНК и ДНК с более высокой точностью по сравнению с традиционными методами NGS, а также более качественно собирать геномы и подробнее характеризовать сложные микробные сообщества [18]. Выявление присутствующих в почве родов и/или видов микроорганизмов может раскрыть ценные сведения об окружающей среде, поскольку определённые микроорганизмы могут служить «видами-индикаторами» или «биомаркерами», под которыми понимают таксономические группы со значительными различиями в численности в разных микробиомах [16].

Наряду с техническим совершенствованием научного инструментария, огромное значение в современной науке имеет способность человека систематизировать отдельные явления, факты, процессы, потому что без этого разделения на различные категории их было бы очень сложно глубоко анализировать, полноценно изучать, находить им применение и использовать на благо человечества. Вместе с тем, быстрый рост объёма экспериментальных данных и научных публикаций создаёт серьёзные трудности для их систематизации, интерпретации и интеграции в единую концептуальную картину. Так, развитие высокопроизводительных методов секвенирования привело к экспоненциальному увеличению объёма данных, характеризующих таксономический состав, функциональные гены и метаболический потенциал почвенных микробных сообществ [19]. Анализ подобных данных требует применения сложных биоинформатических инструментов и междисциплинарных подходов, объединяющих знания в области микробиологии, экологии, статистики

и информатики. Кроме того, исследователи сталкиваются с проблемой информационной перегрузки, связанной не только с объёмом первичных данных, но и с быстрым ростом числа публикаций, зачастую содержащих фрагментарные или слабо сопоставимые результаты [20].

В условиях возрастающего количества научной информации и увеличения её сложности для восприятия всё большую актуальность приобретает в научно-исследовательской деятельности использование методов искусственного интеллекта (ИИ). Особый интерес представляют большие языковые модели (БЯМ или Large Language Models – LLM), способные обрабатывать и интерпретировать значительные объёмы текстовой информации, выявлять скрытые закономерности и формировать связные текстовые представления знаний. В отличие от традиционных алгоритмов машинного обучения, ориентированных на решение узкоспециализированных задач, БЯМ демонстрируют универсальность и могут использоваться для более широкого спектра аналитических и интерпретационных задач.

Библиометрический анализ, проведённый в базе данных Web of Science за период 2000–2024 гг., показал, что различным аспектам использования методов ИИ в научных исследованиях были посвящены 6883 статьи [21]. Однако модели для оценки состояния здоровья почвы и прогнозирования продуктивности и устойчивости агроэкосистем, разработанные с использованием методов ИИ на основе физико-химических и микробиологических параметров почвенного покрова, в литературе встречаются достаточно редко.

Цель настоящего обзора – поиск, анализ и обобщение информации по использованию технологий ИИ и БЯМ при проведении почвенно-экологических исследований.

Объекты и методы исследования

Сбор информации по теме обзора включал электронный поиск научных публикаций в базах данных PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), Scopus (<http://www.scopus.com/>), Академия Google (<https://scholar.google.ru>) и в Научной электронной библиотеке eLIBRARY.RU (<http://www.elibrary.ru>) за последнее десятилетие. Поисковые запросы включали следующие ключевые слова и словосочетания: «искусственный интеллект», «большие языковые модели», «почвенное микробное сообщество», «почвенный мета-

геном» «прогнозирование здоровья почвы», «мониторинг», «метабаркодинг», «здоровье почвы», «микробные индикаторы», «растениеводство», «агрономия». Отобранная литература была рассмотрена, проанализирована, распределена по категориям и представлена в соответствующих разделах обзора.

Интеграция ИИ в анализ научной информации. Традиционные методы работы с научной литературой, основанные на ручном анализе источников, требуют значительных временных затрат и не всегда позволяют охватить весь массив доступной информации [22]. В условиях быстрого роста числа научных публикаций исследователи сталкиваются с проблемой эффективного поиска, отбора и обобщения релевантных публикаций. Интеграция ИИ в рабочий процесс способствует увеличению скорости и масштабируемости скрининга литературы, позволяя быстро обрабатывать и анализировать большие объёмы информации. Использование ИИ не только облегчает процесс поиска нужной литературы, но и позволяет выявлять скрытые закономерности и взаимосвязи между различными работами, генерировать новые идеи и идентифицировать ранее неочевидные связи между результатами исследований. Эта способность выявлять скрытые закономерности особенно ценна при проведении научных исследований [23]. Современные БЯМ могут систематизировать информацию по темам, выделяя пробелы в знаниях (белые пятна) и помогая пользователям быстрее воспринимать сложный контент. При работе с литературой современные ИИ-агенты используют БЯМ, но исследования, посвящённые непосредственно ИИ-помощникам, находятся пока на начальной стадии. Нерешённой, например, остаётся проблема галлюцинаций (ложных или сфабрикованных результатов), полученных в результате «рассуждений» ИИ [23]. Галлюцинации БЯМ возникают из-за создания правдоподобно звучащих, но неверных или сфабрикованных фактов. Галлюцинации могут включать вымышленные цитаты, выдуманные факты или вводящую в заблуждение информацию. Таким образом, чрезмерная зависимость пользователя от резюме ИИ, без проверки и перепроверки, может приводить к появлению в научных работах не существующей или заведомо неверной информации. Пользователи склонны слепо доверять контенту, сгенерированному ИИ, в случаях, когда резюме кажутся авторитетными (резюме или обзор, написанный на основе реально существующей статьи), даже

если в действительности это не так [24]. Кроме того, возможности ИИ ограничены данными, которые были использованы в процессе предварительного обучения модели [25]. Предвзятость или пробелы в знаниях, особенно когда модели не имеют доступа к платным или специализированным источникам научной литературы, могут исказить рассуждения или непреднамеренно упустить важную информацию, ограничивая точность и полноту анализа научной литературы [23].

Существуют также проблемы с воспроизводимостью и прозрачностью, которые являются критичными требованиями к научному результату: ИИ-агенты действуют в режиме «чёрного ящика», каждый раз генерируя немного разные ответы на один и тот же запрос. Это делает воспроизводимость и проверяемость результата, к сожалению, не всегда возможными.

Существуют также этические проблемы, связанные с плагиатом. Текст, созданный с помощью ИИ, может непреднамеренно содержать заимствования или создавать новый синтетический текст, что ставит под вопрос его авторство. Также существует опасение, что чрезмерная автоматизация может подрывать человеческие навыки аналитической деятельности, такие как критическое мышление и творческий анализ, который всегда был неотъемлемой частью научной работы [26].

Инструменты для отбора и анализа литературы. Такие инструменты, как Abstrackr, ASReview, EPPI-Reviewer и Rayyan предназначены для автоматического поиска статей по заголовкам, аннотациям, ключевым словам. Abstrackr и ASReview – это системы на основе машинного обучения, разработанные для помощи в написании литературных обзоров, способные к полуавтоматическому отбору литературы по заданной теме [27]. Платформы Rayyan и EPPI-Reviewer ориентированы на управление рабочим процессом, отбор и метаанализ литературы, но требуют обучения пользователей по работе с данными программами и предлагают ограниченные функции ИИ [27]. Выбор инструмента зависит от конкретных задач, предпочтений пользователя и типа информации, которую необходимо найти.

Генеративные ИИ-помощники. Относимые к этой группе ChatGPT, DeepSeek, Gemini, YandexGPT, GigaChat и другие могут реферировать научные статьи, интерпретировать результаты, формулировать новые концепции, предлагать темы для проведения дальнейших исследований, поддерживать

принятие решений, трактовать исследовательские гипотезы [28]. Пример использования таких ИИ-помощников – это генерация идей и составление резюме. Сгенерированный ими текст следует рассматривать исключительно как предварительный, требующий редактуры и серьёзной перепроверки. Любые вызывающие сомнение, чрезмерно обобщённые, слабо обоснованные или неточные утверждения, сгенерированные ИИ, необходимо рассматривать критически или отбрасывать.

Инструменты семантического поиска и поиска релевантных научных публикаций реализуют подход, основанный не на совпадении ключевых слов, а на понимании смысла и контекста запроса. К этой группе инструментов относятся Consensus, Elicit, Semantic Scholar и Perplexity, которые помогают в семантическом поиске литературы, предлагая пользователю более релевантные результаты по сравнению с традиционными методами поиска, основанными на ключевых словах [29]. Инструменты семантического поиска, управляемые ИИ, могут выявлять концептуально связанные работы, которые не имеют одинаковых ключевых слов, автоматически

проверять большие объёмы аннотаций на релевантность и рекомендовать валидные или тематически связанные статьи посредством анализа сети цитирования. Эти возможности особенно ценны, если литература по изучаемому предмету отличается фрагментарностью, а также в междисциплинарных исследованиях, где релевантная литература часто разбросана по нескольким областям знаний.

Инструменты для составления карт цитирования Connected Papers, Litmaps, Research Rabbit и Scite изучают контекст цитирования, анализируют связи цитирования, выявляют фундаментальные и практически ориентированные исследования, визуализируют научные сети цитирований (рис. 4).

Инструменты для составления карт цитирования и визуализации данных могут быть полезными для понимания состояния изученности того или иного вопроса, написания обзоров, но характеризуются отсутствием глубоких аналитических функций и имеют ограничения при работе со статистическими данными [29, 31].

Инструменты для составления резюме и анализа текста. К этой группе относятся

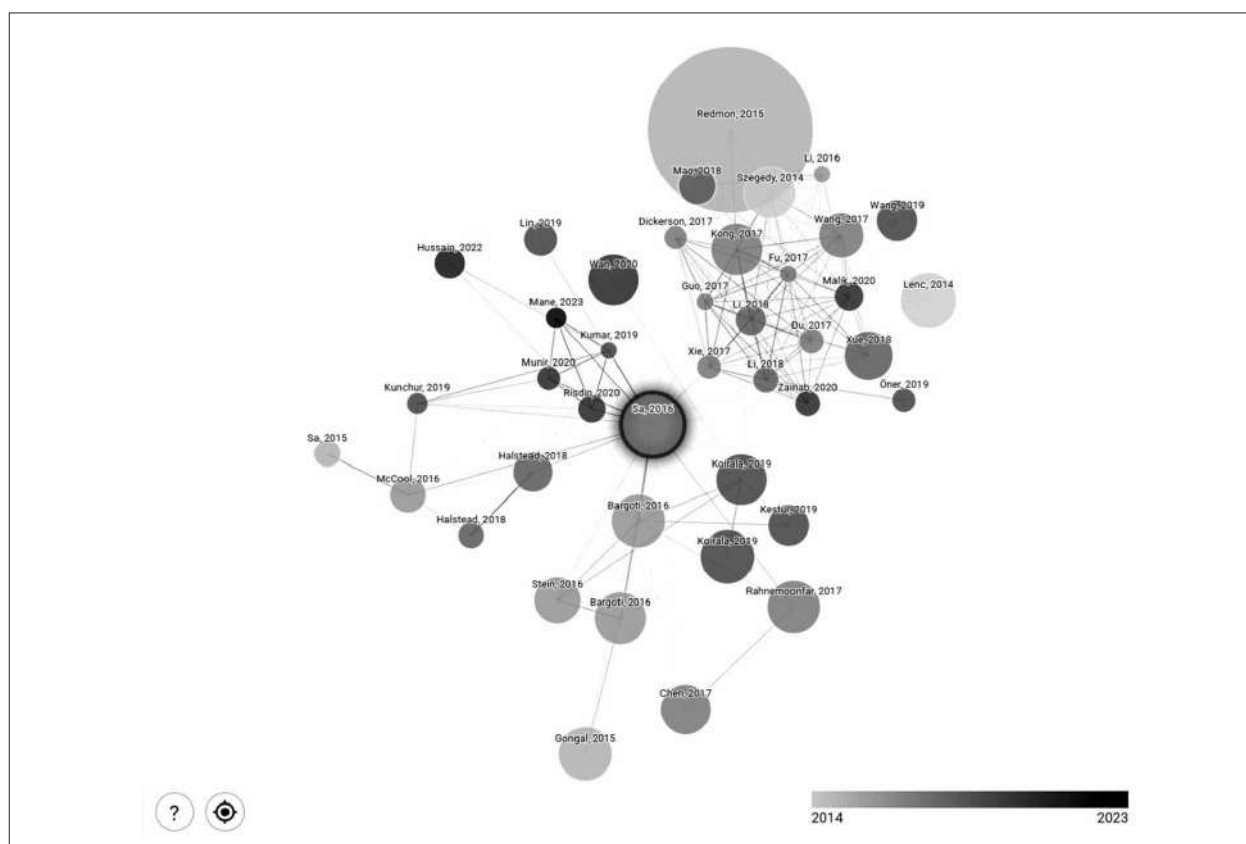


Рис. 1. Карта цитирования статьи (выделена в центре рисунка), созданная в программе Connected Papers. Цит. по [30]

Fig. 1. Citation map of the article (highlighted in the center of the figure), created in the Connected Papers program. Cited from [30]

NotebookLM, Scholarcy, SciSpace, SciSummary и Paper Digest. Они позволяют выделять ключевые моменты, технически упрощать содержание текста, составлять резюме [29].

Инструменты ИИ могут использоваться на различных этапах процесса обзора литературы для повышения эффективности, точности и полноты отображения информации. Одно из основных применений – полностью или частично автоматизированный поиск литературы, при котором базы данных, управляемые ИИ, и алгоритмы обработки естественного языка могут извлекать релевантные источники более эффективно, чем традиционный поиск по ключевым словам [32]. Расширенный поиск организован в форме иерархически выстроенной системы фильтров и логических операций по атрибутам элементов.

Ещё одна ценная функция – извлечение метаданных, библиографических сведений, оформление списков библиографических источников, что снижает нагрузку на пользователя. Системы ИИ могут извлекать из статей в структурированном формате методологическую информацию и сортировать информационные записи по названию, автору, типу источника, дате, издательству, языку и другим характеристикам. Для обеспечения контроля за этим процессом в работе должны быть контрольные точки проверки с обязательным участием человека.

Инструменты на базе ИИ способны выявлять при обзоре литературы закономерности, классифицировать литературу или оценивать её релевантность. Исследователи при этом должны критически проверять полученные результаты, чтобы избежать обусловленной алгоритмом предвзятости или неверной интерпретации исходных данных. В конечном итоге, наиболее эффективно рассматривать ИИ как вспомогательный инструмент, дополняющий человека, а не заменяющий его творческое участие.

Технологии ИИ в анализе почвенных микробиомов

Состав почвенного микробиома отражает экологическое состояние почвы, её здоровье, поскольку присутствующие в почве микроорганизмы принимают непосредственное участие в биогеохимических процессах. В связи с этим микробные сообщества обладают некоторым прогностическим потенциалом, позволяя судить о том, как эти процессы будут протекать в будущем [33]. Микробиомы могут служить

моделью для прогнозирования экосистемных процессов, поскольку они непосредственно в них участвуют.

Способствовать повышению эффективности управления почвенным плодородием и в целом экологической устойчивости сельскохозяйственного производства может использование универсального индекса качества почвы (Soil Quality Index, SQI). Этот индекс должен быть достаточно гибким, чтобы учитывать региональные особенности, но при этом позволять проводить сравнительные оценки. Достижения в области ИИ открыли новые возможности по созданию прогностических моделей в этой сфере. Так, в Бразилии, наряду с данными по продуктивности растений, физическими и химическими параметрами почвы, в расчёт SQI были включены почвенные метагеномные данные, а также использованы алгоритмы машинного обучения [34]. Это значительно улучшило прогностическую ценность модели, по сравнению с ранее существовавшими. Возможность практического применения данного индекса сдерживают ограниченность имеющегося у авторов массива данных (микробиологических, молекулярных, физических, химических за период 3–5 лет), и необходимость высоких вычислительных мощностей для обработки и интерпретации результатов.

Параметры почвенного микробиома также использовали в другом исследовании, направленном на разработку прогностической модели, включающей 12 ключевых показателей здоровья почвы [35]. В работе применяли алгоритмы машинного обучения, в частности, метод «случайного леса» (Random forest – RF) [36] и метод опорных векторов (Support Vector Machine – SVM) [37]. В качестве исходных авторы использовали метагеномные данные по бактериальным сообществам из 949 образцов почв, отобранных в пределах различных сельскохозяйственных угодий Северной Америки. Выделено несколько таксонов, включая *Pyrinomonadaceae*, *Nitrososphaeraceae* и *Candidatus Udeaobacter*, наличие которых значимо связано со здоровьем почвы. Авторы отмечают, что биологические показатели здоровья модель прогнозировала с наибольшей точностью, в то время как с химическими и физическими показателями возникли некоторые затруднения. Дальнейшее повышение точности прогнозирования может быть достигнуто посредством увеличения регионально-специфичных наборов данных. Такой вывод свидетельствует о наличии у рассмотренных

моделей фундаментальных ограничений: невозможность прямого переноса прогнозов с одного на другие типы почв и высокая ресурсоёмкость, обусловленная потребностью в огромных массивах информации для оптимизации работы системы.

Подобная работа была проделана в Новой Зеландии для прогнозирования физико-химических характеристик почвы с использованием метода RF [38]. В исследовании определяли состав бактериальных сообществ с помощью метагеномного анализа образцов почв, отобранных в период с 2013 по 2018 гг. в десяти различных регионах страны, с площади, охватывающей приблизительно 196 тыс. км². В общей сложности проанализировано более чем 3000 образцов почв, отобранных с 606 участков с различными типами землепользования. Предложенная модель выявила устойчивую корреляцию между составом бактериальных сообществ и типом землепользования. Выявленная связь обусловлена чувствительностью бактерий к изменениям физико-химических свойств почвы. Наибольший вклад в формирование почвенного прокариотного сообщества вносят такие показатели как кислотность (pH) и соотношение углерода и азота (C:N). Установлено, что различные типы землепользования способствуют формированию бактериальных сообществ с характерным составом. Это даёт возможность определять по составу микробиома тип землепользования с точностью до 85 %. Полученные результаты свидетельствуют о потенциале метагеномных инструментов в оценке качества почвы и прогнозировании её состояния.

Машинное обучение с использованием RF применялось также в штате Иллинойс (США) при исследовании взаимодействия растений с почвенными микроорганизмами [39]. Авторы продемонстрировали преимущества полногеномного исследования ассоциаций (Metagenome-wide association studies – MWAS) и машинного обучения в прогнозировании урожайности сельскохозяйственных культур на основе состава почвенного микробиома. Установлено, что различия в урожайности сельскохозяйственных культур на шести обследованных полях в значительной степени определялись вариативностью таксонов в почвенном микробиоме. С помощью алгоритмов машинного обучения выявлена связь между большей представленностью таксонов *Bradyrhizobium* и *Gamma-proteobacteria* в почве участков с высокой урожайностью, а таксонов *Actinobacteria*, *Ascomycota*, *Planctomycetales*

и стрептофитовых водорослей – в почве участков с более низкой урожайностью. Предложенная модель способна успешно предсказывать продуктивность сельскохозяйственных культур на основе состава микробиома на уровне порядка с максимальной точностью 0,79.

Рассмотренные примеры демонстрируют ряд общих закономерностей, в частности – зависимость организации почвенных микробиомов от факторов окружающей среды. Состав бактериальных сообществ тесно связан с физико-химическими характеристиками почвы, среди которых наиболее значимыми являются кислотность (pH), соотношение углерода и азота, а также содержание органического вещества. Изменение характера землепользования сопровождается изменением этих параметров, влекущих закономерную перестройку структуры микробных сообществ, и приводит к формированию специфических микробиомов, характерных для различных типов экосистем и агроценозов. Отдельные таксоны микроорганизмов могут выступать индикаторами продуктивности сельскохозяйственных культур и состояния почвы. Использование методов машинного обучения позволило выявить эти зависимости и построить прогностические модели, основанные на анализе большого числа одновременно действующих факторов, что при применении традиционных статистических подходов было бы затруднительным или вовсе невозможным.

Известны другие работы, в которых методы ИИ были использованы для изучения характера взаимных связей между характеристиками почвенных микроорганизмов и физико-химическими свойствами почв [40], для прогнозирования биомассы почвенных микроорганизмов [41], активности почвенных ферментов [42], оценки микробиологических характеристик почвы и структуры бактериального сообщества [43]. Анализ характеристик микробиома в этих работах опирается на широкий спектр вычислительных методов – от традиционных статистических [44] до более сложных алгоритмов на основе ИИ.

Совокупность полученных результатов свидетельствует о том, что методы ИИ позволяют не только автоматизировать обработку метагеномных данных, но и выявлять ранее трудно обнаруживаемые связи между составом почвенных микробных сообществ, физико-химическими свойствами почвы, характером землепользования и продуктивностью агроэкосистем.

Проблема адаптации заимствованных из биомедицины методов ИИ к почвенно-экологическим исследованиям

Необходимо отметить, что методы искусственного интеллекта изначально получили широкое распространение в биомедицинских исследованиях – при анализе роли кишечной микробиоты в поддержании здоровья и развитии заболеваний человека и животных. Впоследствии данный методологический подход был распространён на задачи медицинской диагностики, где алгоритмы машинного обучения (Machine Learning – ML) [45] и глубокого обучения (Deep Learning – DL) [46] применяются для выявления ассоциаций между составом микробиоты и различными патологическими состояниями [47–49]. Адаптация заимствованных из области медицины и здравоохранения методов машинного обучения к нуждам сельского хозяйства сопряжена с многочисленными проблемами, включая количество и качество данных, зачастую разреженность точек отбора проб, гетерогенность почвы, отсутствие однозначных биомаркеров. Ввиду влияния множества неконтролируемых факторов – погоды, топографии и т. д. образцы почв гораздо более гетерогенны, чем биомедицинские. Кроме того, наборы данных по почвенным микробиомам уступают в объёме и глубине секвенирования наборам данных по микробиому человека. Ограниченность почвенных данных зачастую диктует необходимость более сложной биоинформатической обработки в сравнении с данными по микробиому человека [45].

Если исследования микробиома человека сосредоточены в медицине на прогнозировании заболеваний, то в сельском хозяйстве основной целью является идентификация биомаркеров, связанных с плодородием почвы и/или урожайностью сельскохозяйственных культур. Тем не менее, примеры биомедицинского применения методов ИИ могут служить ориентиром для изучения почвенных микробиомов. Алгоритмы машинного обучения могут внести значительный вклад в дальнейшее развитие знаний о здоровье почвы, анализируя большие данные о микробных сообществах и делая прогнозы относительно различных экологических показателей [47].

Процедура анализа данных почвенных микробиомов с использованием методов ИИ сходна с обычной биоинформатической обработкой данных. Важным этапом анализа

микробиома является предварительная подготовка данных, обеспечивающая адекватное формирование входных массивов для точного моделирования методами ML и DL (рис. 2).

Нормализация данных микробиомов необходима для корректировки различий в глубине секвенирования между образцами. Для решения проблемы асимметрии распределений могут быть использованы логарифмические или иные виды преобразований. Для снижения размерности используются такие методы, как анализ главных компонент (PCA) [50], стохастическое вложение соседей с t-распределением (t-distributed Stochastic Neighbor Embedding – t-SNE) [51] и равномерная аппроксимация и проекция многообразия (UMAP – Uniform Manifold Approximation and Projection) [52]. Эти подходы значительно ускоряют обучение моделей [53].

Следующим шагом после предварительной обработки данных является выбор и обучение подходящей модели ML или DL, в зависимости от рабочей гипотезы исследования и характера имеющихся данных. При работе с микробиомами наиболее часто применяются методы: RF [36], градиентного бустинга [54] и метод SVM [37], а в задачах глубокого обучения – нейронные сети, такие как свёрточные нейронные сети (Convolutional Neural Network – CNN) [55], рекуррентные нейронные сети (Recurrent Neural Network – RNN) [56] и автокодировщики [57]. Модели глубокого обучения, в частности CNN и RNN, используются для анализа сложных наборов данных микробиомов и прогнозирования функциональных свойств на основе таксономического состава микробиома. Эти модели специально предназначены для обработки огромных объёмов данных, генерируемых технологиями высокопроизводительного секвенирования геномов [58]. Новые нейронные сети, основанные на физических принципах (Physics-informed neural networks – PINN), перспективны в моделировании роста микробных популяций (в том числе почвенных) с учётом факторов окружающей среды, предоставляя мощный инструмент для изучения динамики развития микробиомов [59]. Применяя этот алгоритм, можно получить всестороннее представление о том, как микроорганизмы взаимодействуют друг с другом, и как эти взаимодействия сказываются на окружающей среде, будь то система агроценоза или человеческий организм [58].

Особенности различных подходов к машинному обучению для анализа данных ми-

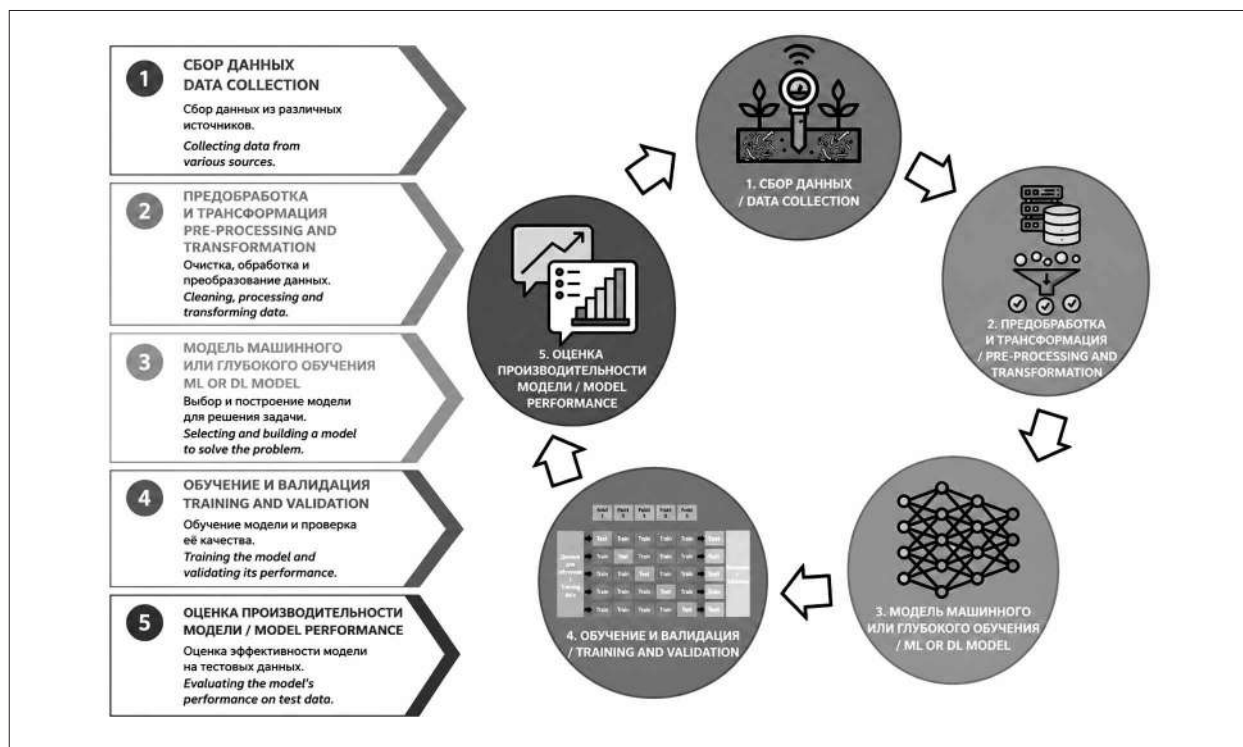


Рис. 2. Основные этапы разработки моделей машинного обучения или глубокого обучения. Цит. по [48] с изменениями

Fig. 2. Main stages of machine learning or deep learning model development. Cited from [48] with modifications

кробиома рассмотрены различными авторами [58–61]. Отмечаются достижения в методах DL и их преимущества перед классическими методами ML.

Следующим шагом в создании прогностических моделей на основе ИИ становятся модели, в которых наряду с физическими, химическими и биологическими характеристиками почвы, интегрированы также агрометеорологические данные. Это обеспечит возможность сопоставлять изменения в составе микробного сообщества с более широким набором различных параметров окружающей среды в целом. Прогресс в данной области уже есть, но для улучшения прогностической точности современных моделей всё ещё требуется слишком большой набор разнородных данных. Кроме того, для хранения всей собранной информации важно использовать централизованную базу данных, фиксировать контекстные метаданные о методах отбора, локациях и условиях отбора проб, что обеспечит основу для дальнейшей интерпретации результатов.

Используя методы метагеномики, исследователи могут выявлять и идентифицировать полезные микроорганизмы (азотфиксирующие бактерии, микоризные

грибы и другие симбиотические микробы) в здоровых почвах, которые улучшают плодородие почвы и стимулируют рост растений. На основе собранных данных и понимания микробной экологии могут быть описаны микроорганизмы-биоиндикаторы, отвечающие за плодородие почв. Если индикаторные виды в почве найдены не были, алгоритмы могут рекомендовать конкретные микробные инокулянты или агрономические практики с учётом уникальных особенностей почвы и выращиваемых культур. Включение этих элементов в систему поддержки принятия решений (СППР) позволит сельхозпроизводителям принимать более обоснованные решения, способствующие устойчивому ведению сельского хозяйства. Обратная связь, поступающая от фермеров, позволит судить об эффективности рекомендованных методов и наблюдаемых изменениях в состоянии почвы, а со временем – точнее настраивать алгоритмы, что критически важно для постоянного совершенствования СППР [62]. Реализация этого процесса потребует, по-видимому, значительного времени, однако результаты будут иметь весомое значение, как для конкретных сельхозпроизводителей, так и для мирового аграрного сектора в целом.

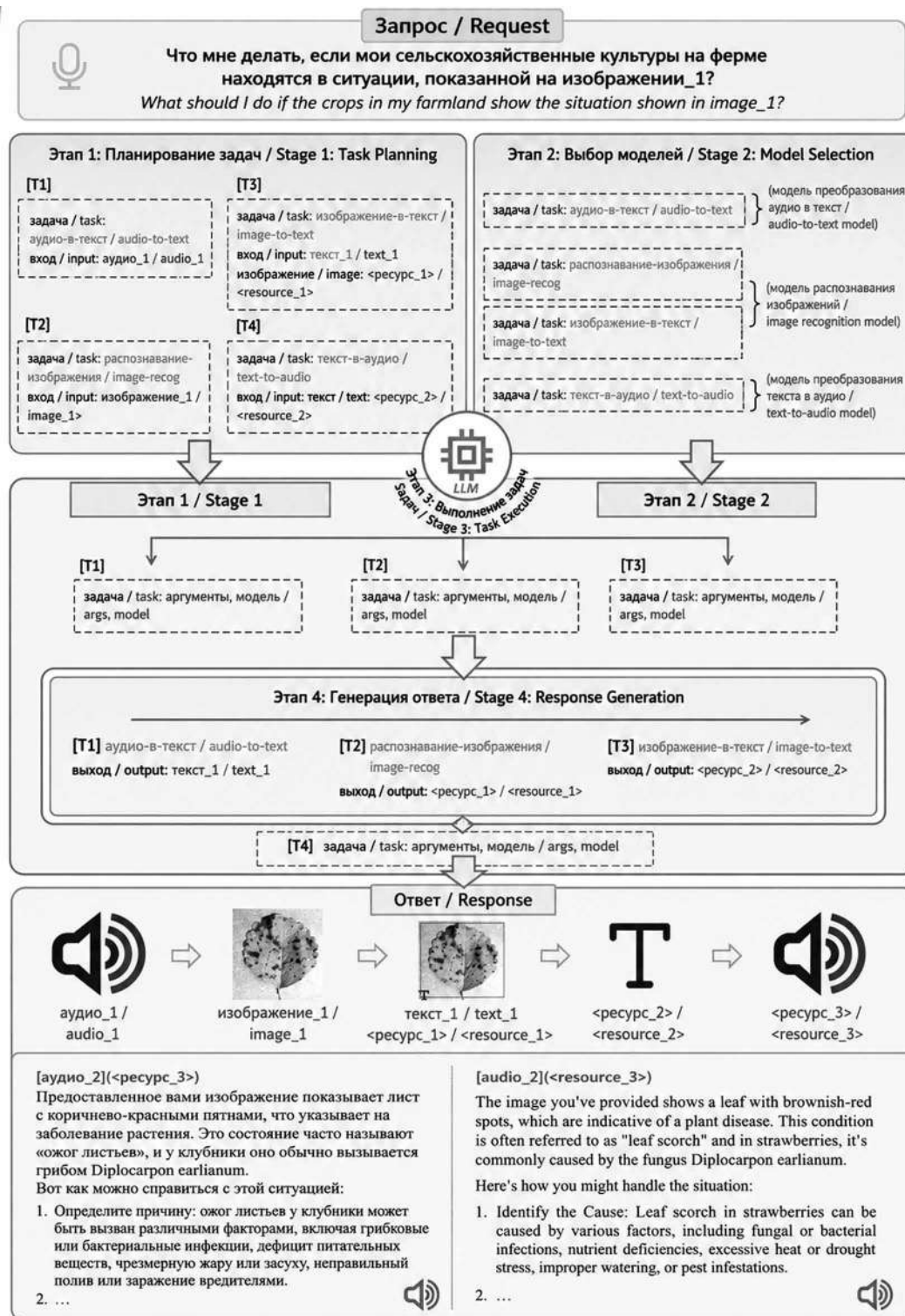


Рис. 3. Схема работы HuggingGPT при запросе по поражению листьев растений пятнами.
 Цит. по [70] с изменениями
 Fig. 3. Flow chart of HuggingGPT for query on leaf spot lesions.
 Cited from [70] with modifications

Перспективы использования технологий ИИ в сельском хозяйстве

Современное сельское хозяйство переходит к «умному» агропроизводству, основанному на роботизации, ИИ, интернете вещей (Internet of Things – IoT) и больших данных для оптимизации урожайности сельскохозяйственных культур при минимизации затрат на воду, удобрения и пестициды [63]. Такие устройства, как датчики влажности почвы, метеостанции и автоматизированные системы орошения, позволяют фермерам отслеживать и контролировать сельскохозяйственные процессы в режиме реального времени. Подключённые к сети интернет датчики обеспечивают непрерывный сбор данных об условиях окружающей среды и состоянии урожая, что позволяет осуществлять точное управление агрономическими процессами и своевременно вносить необходимые корректировки.

Анализ больших данных для выявления скрытых закономерностей, корреляций, тенденций позволит фермерам принимать более обоснованные решения [64]. Наборы данных, собранных с датчиков, расположенных непосредственно на возделываемом участке (датчики влажности, температуры, pH) или дистанционно от обрабатываемой площади (с дронов или спутников), могут с помощью новых вычислительных технологий помочь фермерам оптимизировать распределение ресурсов, прогнозировать урожайность, управлять рисками [65, 66].

Благодаря интеграции новых технологий, фермеры могут выйти за рамки традиционных методов и принимать решения, основываясь на данных, приходящих в режиме реального времени [65]. Сегодня процесс принятия решений облегчается СППР, которые постепенно развивались, включая более сложные инструменты и методы [67]. Благодаря внедрению алгоритмов машинного обучения и прогностических моделей, СППР действительно могут анализировать сложные наборы данных, предлагая новые идеи и практические рекомендации [68]. Предложен ИИ-агент, построенный на основе БЯМ, под названием HuggingGPT. Данная система использует БЯМ как ядро принятия решений в области сельского хозяйства, решая задачи в режиме реального времени, например, отвечая на вопросы фермеров [69]. Пример работы HuggingGPT по формированию ответа на запрос по поражению листьев растений пятнистостью представлен на рисунке 3.

Для ответа на поставленный вопрос система использовала модели: Amazon transcribe, Whisper [71] для преобразования голоса в текст, модель vit-gpt2 для распознавания входящего изображения и генерации текста на его основе, преобразование текстового ответа в аудио производилось с помощью модели FastSpeech [72], в качестве контролера использовали ChatGPT. Большая языковая модель выступает в роли «мозга» агента, выполняя важные функции. При получении запроса на выполнение задачи БЯМ сначала делит всю задачу на подзадачи и выбирает подходящую модель ИИ в зависимости от особенностей каждой подзадачи. Далее происходит агрегация всех решений и ChatGPT формирует резюме для ответа на запрос.

Заключение

Проведённый анализ литературы показывает, что методы искусственного интеллекта становятся перспективным инструментом изучения почвенных микробиомов и оценки состояния почв. В отличие от традиционных статистических подходов, модели машинного и глубокого обучения способны одновременно учитывать большое количество биотических и абиотических факторов, выявлять сложные нелинейные взаимосвязи между составом микробных сообществ и свойствами почвы, а также использовать данные метагеномного анализа для прогнозирования показателей её качества и здоровья. Это делает возможным выявление ранее неочевидных закономерностей в организации почвенных микробных сообществ, а также построение прогностических моделей состояния почвы, пригодных для мониторинга и поддержки принятия решений в сфере сельского хозяйства.

Использование технологий ИИ в изучении и оценке почвенных ресурсов может произвести революцию в нашем понимании подходов к управлению плодородием. Одним из существенных преимуществ моделей ИИ является их способность постоянно обучаться и адаптироваться по мере поступления новых данных. Высокая адаптивность ИИ имеет решающее значение для изучения почвенных микробиомов, отличающихся постоянной изменчивостью под воздействием окружающей среды. По мере поступления новых данных модели могут быть усовершенствованы для построения более точных прогнозов состояния почвы.

Учёным, фермерам и представителям аграрной промышленности необходимо со-

вместно разработать стандартизированные протоколы по сбору и обработке данных, связанных с почвой. Стандартные протоколы, большой объём данных и качественная методологическая база приблизят нас к пониманию того, насколько хорошо ИИ прогнозирует состояние почвы.

Работа выполнена в рамках Государственных заданий ФАНЦ Северо-Востока имени Н.В. Рудницкого (тема № FNWE-2025-0005) и ФИЦ «Коми научный центр Уральского отделения Российской академии наук» (по теме «Оценка состояния трансформированных экосистем подзоны южной тайги, методические подходы к их биоремедиации», номер государственной регистрации в ЕГИСУ № 125021402208-5).

References

1. Dobrovolskii G.V. Dokuchaev and present-day natural science // Eurasian Soil Sc. 1996. V. 29. No. 2. P. 105–110.
2. Dokuchaev V.V. Essays: in 9 volumes. Moskva; Leningrad: Publishing House of the USSR Academy of Sciences, 1949–1961. (in Russian).
3. Montgomery D.R. Soil erosion and agricultural sustainability // Proc. Natl. Acad. Sci. USA. 2007. V. 104. No. 33. P. 13268–13272. doi: 10.1073/pnas.0611508104
4. Dobrovolskiy G.V. Soil degradation as a threat of global ecological crisis // Vek globalizatsii. 2008. No. 2. P. 54–65 (in Russian).
5. Giller K.E., Delaune T., Silva J.V., Descheemaeker K., van de Ven G., Schut A.G.T., van Wijk M., Hammond J., Hochman Z., Taulya G., Chikowo R., Narayanan S., Kishore A., Bresciani F., Teixeira H.M., Andersson J.A., van Ittersum M.K. The future of farming: who will produce our food? // Food Sec. 2021. V. 13. P. 1073–1099. doi: 10.1007/s12571-021-01184-6
6. Kurmanbayev A.A., Syndet T.R. Soil health concept and modern soil health indicators // Soil Science and Agrichemistry. 2023. No. 2. P. 91–106 (in Russian). doi: 10.51886/1999-740X_2023_2_91
7. Nikitin D.A., Semenov M.V., Chernov T.I., Ksenofontova N.A., Zhelezova A.D., Ivanova E.A., Khitrov N.B., Stepanov A.L. Microbiological indicators of ecological soil functions: a review // Eurasian Soil Sc. 2022. V. 55. P. 221–234. doi: 10.1134/S1064229322020090
8. Golubev A.S., Berestetskiy A.O. Future directions for use of biological and biorational herbicides in Russia (review) // Agricultural Biology. 2021. V. 56. No. 5. P. 868–884 (in Russian). doi: 10.15389/agrobiol.2021.5.868rus
9. Mohanty L.K., Singh N.K., Raj P., Prakash A., Tiwari A.K., Singh V., Sachan P. Nurturing crops, enhancing soil health, and sustaining agricultural prosperity worldwide through agronomy // J. Exp. Agric. Int. 2024. V. 46. No. 2. P. 46–67. doi: 10.9734/jeai/2024/v46i22308
10. Ryazanov S.S., Kulagina V.I. Soil health: an overview of concepts and monitoring methods // Moscow Univ. Soil Sci. Bull. 2026. V. 81. P. 16–31. doi: 10.3103/S0147687425700711
11. Slobodkin A.I., Slobodkina G.B. Diversity of sulfur-disproportionating microorganisms // Microbiology. 2019. V. 88. P. 509–522. doi: 10.1134/S0026261719050138
12. Zavarzina D.G. The role of dissimilatory Fe(III)-reducing bacteria in transformation of iron minerals // Paleontol. J. 2004. V. 38. No. 3. P. 231–237.
13. Vecherskii M.V., Semenov M.V., Lisenkova A.A., Stepankov A.A. Metagenomics: a new direction in ecology // Biol. Bull. Russ. Acad. Sci. 2022. V. 48. Suppl. 3. P. S107–S117. doi: 10.1134/S1062359022010150
14. Semenov M.V. Metabarcoding and metagenomics in soil ecological research: achievements, problems and opportunities // Biol. Bull. Rev. 2021. V. 11. P. 40–53. doi: 10.1134/S2079086421010084
15. Nesme J., Achouak W., Agathos S.N., Bailey M., Baldrian P., Brunel D., Frostegård Å., Heulin T., Jansson J.K., Jurkevitch E., Kruus K.L., Kowalchuk G.A., Lagares A., Lappin-Scott H.M., Lemanceau P., Le Paslier D., Mandic-Mulec I., Murrell J.C., Myrold D.D., Nalin R., Nannipieri P., Neufeld J.D., O'Gara F., Parnell J.J., Pühler A., Py-lro V., Ramos J.L., Roesch L.F.W., Schlöter M., Schleper C., Sczyrba A., Sessitsch A., Sjöling S., Sørensen J., Sørensen S.J., Tebbe C.C., Topp E., Tsiamis G., van Elsas J.D., van Keulen G., Widmer F., Wagner M., Zhang T., Zhang X., Zhao L., Zhu Y.G., Vogel T.M., Simonet P. Back to the future of soil metagenomics // Front. Microbiol. 2016. V. 7. Article No. 73. doi: 10.3389/fmicb.2016.00073
16. Leite M.F.A., van den Broek S.W.E.B., Kuramae E.E. Current challenges and pitfalls in soil metagenomics // Microorganisms. 2022. V. 10. No. 10. Article No. 1900. doi: 10.3390/microorganisms10101900
17. Wydro U. Soil microbiome study based on DNA extraction: a review // Water. 2022. V. 14. No. 24. Article No. 3999. doi: 10.3390/w14243999
18. Veselovsky V., Romanov M., Zoruk P., Larin A., Babenko V., Morozov M., Strokach A., Zakharevich N., Khamidova S., Danilova A., Vatlin A., Pavshintsev V., Chenguang F., Tsybizov D., Mitkin N., Galanova O., Zakharenko A., Golokhvast K., Klimina K. Comparative evaluation of sequencing platforms: Pacific Biosciences, Oxford Nanopore Technologies, and Illumina for 16S rRNA-based soil microbiome profiling // Front. Microbiol. 2025. V. 16. Article No. 1633360. doi: 10.3389/fmicb.2025.1633360
19. D'Argenio V. The high-throughput analyses era: are we ready for the data struggle? // High Throughput. 2018. V. 7. No. 1. Article No. 8. doi: 10.3390/ht7010008
20. Sharpton T.J. An introduction to the analysis of shotgun metagenomic data // Front. Plant Sci. 2014. V. 5. Article No. 209. doi: 10.3389/fpls.2014.00209

21. Panciu B.-I., Barbu Dragne G.-L. A.I. in agriculture – a bibliometric approach // *Proceedings of the International Conference on Business Excellence*. 2025. V. 19. No. 1. P. 5189–5199. doi: 10.2478/picbe-2025-0397
22. Zhao Y., Liang J., Chen B., Deng X., Zhang Y., Xiong Z., Pan M., Meng X. Applications research progress and prospects of multi-agent large language models in agricultural // *Smart Agric*. 2025. V. 7. No. 5. P. 37–51. doi: 10.12133/j.smartag.SA202503026
23. Sirilertmekasakul C., Rattanawong W., Gongvatana A., Srikiatkachorn A. The current state of artificial intelligence-augmented digitized neurocognitive screening test // *Front. Hum. Neurosci*. 2023. V. 17. Article No. 1133632. doi: 10.3389/fnhum.2023.1133632
24. Saied M., Mokhtar N., Badr A., Adel M., Boles P., Khoriba G. AI in literature reviews: a survey of current and emerging methods // *Proceedings of the 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. Cairo, Egypt, 2024. P. 61–65. doi: 10.1109/MIUCC62295.2024.10783597
25. Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls // *AI & Soc*. 2020. V. 35. No. 3. P. 761–765. doi: 10.1007/s00146-020-00978-0
26. Gerlich M. AI tools in society: impacts on cognitive offloading and the future of critical thinking // *Societies*. 2025. V. 15. No. 1. Article No. 6. doi: 10.3390/soc15010006
27. Schmidt L., Cree I., Campbell F., WCT EVI MAP group. Digital tools to support the systematic review process: an introduction // *J. Eval. Clin. Pract.* 2025. V. 31. No. 3. Article No. e70100. doi: 10.1111/jep.70100
28. Filimonov V. Yu. Large language models and their role in modern scientific discoveries // *Philosophical Problems of IT and Cyberspace (PhilIT&C)*. 2024. V. 25. No. 1. P. 42–57 (in Russian). doi: 10.17726/philIT.2024.1.3
29. Mtotywa M.M., Mowers J.L.J., Ndou W., Moleko T.V.Q., Ledwaba M.J. Artificial intelligence in literature review synthesis: a step-by-step methodological approach for researchers and academics // *Informatics*. 2026. V. 13. No. 3. Article No. 43. doi: 10.3390/informatics13030043
30. Sa I., Ge Z., Dayoub F., Upcroft B., Perez T., McCool C. DeepFruits: a fruit detection system using deep neural networks // *Sensors*. 2016. V. 16. No. 8. Article No. 1222. doi: 10.3390/s16081222
31. Fallico N. Exploring Research Rabbit: your new favourite reference manager // *Medical Writing*. 2025. V. 34. No. 1. P. 67–69. doi: 10.56012/wxzm3433
32. Zhang Y., Long J., Zhao W. The curvilinear relationships between relational embeddedness and dynamic capabilities: the mediating effect of ambidextrous learning // *Front. Psychol*. 2022. V. 13. Article No. 830377. doi: 10.3389/fpsyg.2022.830377
33. Correa-Garcia S., Constant P., Yergeau E. The forecasting power of the microbiome // *Trends Microbiol*. 2022. V. 31. No. 5. P. 444–452. doi: 10.1016/j.tim.2022.11.013
34. De Andrade V.H.G.Z., Redmile-Gordon M., Barbosa B.H.G., Andreote F.D., Roesch L.F.W., Pylro V.S. Artificially intelligent soil quality and health indices for ‘next generation’ food production systems // *Trends Food Sci. Technol*. 2021. V. 107. P. 195–200. doi: 10.1016/j.tifs.2020.10.018
35. Wilhelm R.C., van Es H.M., Buckley D.H. Predicting measures of soil health using the microbiome and supervised machine learning // *Soil Biol. Biochem*. 2022. V. 164. Article No. 108472. doi: 10.1016/j.soilbio.2021.108472
36. Breiman L. Random forests // *Mach. Learn*. 2001. V. 45. P. 5–32. doi: 10.1023/A:1010933404324
37. Cortes C., Vapnik V. Support-vector networks // *Mach. Learn*. 1995. V. 20. P. 273–297. doi: 10.1007/BF00994018
38. Hermans S.M., Buckley H.L., Case B.S., Curran-Cournane F., Taylor M., Lear G. Using soil bacterial communities to predict physico-chemical variables and soil quality // *Microbiome*. 2020. V. 8. No. 1. Article No. 79. doi: 10.1186/s40168-020-00858-1
39. Chang H.X., Haudenshield J.S., Bowen C.R., Hartman G.L. Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity // *Front. Microbiol*. 2017. V. 8. Article No. 519. doi: 10.3389/fmicb.2017.00519
40. Sadeghi S., Petermann B.J., Steffan J.J., Brevik E.C., Gedeon C. Predicting microbial responses to changes in soil physical and chemical properties under different land management // *Appl. Soil Ecol*. 2023. V. 188. Article No. 104878. doi: 10.1016/j.apsoil.2023.104878
41. Tajik S., Ayoubi S., Nourbakhsh F. Prediction of soil enzymes activity by digital terrain analysis: comparing artificial neural network and multiple linear regression models // *Environ. Eng. Sci*. 2012. V. 29. No. 8. P. 798–806. doi: 10.1089/ees.2011.0313
42. Larsen P.E., Field D., Gilbert J.A. Predicting bacterial community assemblages using an artificial neural network approach // *Nat. Methods*. 2012. V. 9. No. 6. P. 621–625. doi: 10.1038/nmeth.1975
43. Santos E.C., Armas E.D., Crowley D., Lambais M.R. Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil // *Soil Biol. Biochem*. 2014. V. 69. P. 101–109. doi: 10.1016/j.soilbio.2013.10.049
44. Shirokih I.G., Bokov N.A., Shirokih A.A., Ashikhmina T. Ya. Antibiotic resistance profiles of streptomycetes from different soils of the Vyatka Prikamye // *Theoretical and Applied Ecology*. 2026. No. 2. P. 183–191 (in Russian). doi: 10.25750/1995-4301-2026-2-183-191
45. Langley P. The changing science of machine learning // *Mach. Learn*. 2011. V. 82. P. 275–279. doi: 10.1007/s10994-011-5242-y
46. LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. 2015. V. 521. No. 7553. P. 436–444. doi: 10.1038/nature14539
47. Ghannam R.B., Techtmann S.M. Machine learning applications in microbial ecology, human microbiome

studies, and environmental monitoring // *Comput. Struct. Biotechnol. J.* 2021. V. 19. P. 1092–1107. doi: 10.1016/j.csbj.2021.01.028

48. Pace R., Schiano Di Cola V., Monti M.M., Affinito A., Cuomo S., Loreto F., Ruocco M. Artificial intelligence in soil microbiome analysis: a potential application in predicting and enhancing soil health – a review // *Discov. Appl. Sci.* 2025. V. 7. Article No. 85. doi: 10.1007/s42452-024-06381-4

49. Iadanza E., Fabbri R., Bašić-Čičak D., Amedei A., Telalovic J.H. Gut microbiota and artificial intelligence approaches: a scoping review // *Health Technol.* 2020. V. 10. P. 1343–1358. doi: 10.1007/s12553-020-00486-7

50. Gewers F.L., Ferreira G.R., de Arruda H.F., Silva F.N., Comin C.H., Amancio D.R., Costa L.F. Principal component analysis: a natural approach to data exploration // *ACM Comput. Surv.* 2021. V. 54. No. 4. Article No. 70. doi: 10.1145/3447755

51. Van der Maaten L., Hinton G. Visualizing data using t-SNE // *J. Mach. Learn. Res.* 2008. V. 9. P. 2579–2605.

52. Becht E., McInnes L., Healy J., Dutertre C.A., Kwok I.W.H., Ng L.G., Ginhoux F., Newell E.W. Dimensionality reduction for visualizing single-cell data using UMAP // *Nat. Biotechnol.* 2019. V. 37. P. 38–44. doi: 10.1038/nbt.4314

53. Zeng T., Yu X., Chen Z. Applying artificial intelligence in the microbiome for gastrointestinal diseases: a review // *J. Gastroenterol. Hepatol.* 2021. V. 36. No. 4. P. 832–840. doi: 10.1111/jgh.1550

54. Friedman J.H. Stochastic gradient boosting // *Comput. Stat. Data Anal.* 2002. V. 38. No. 4. P. 367–378. doi: 10.1016/S0167-9473(01)00065-2

55. LeCun Y., Boser B., Denker J.S., Henderson D., Howard R.E., Hubbard W., Jackel L.D. Backpropagation applied to handwritten zip code recognition // *Neural Comput.* 1989. V. 1. No. 4. P. 541–551. doi: 10.1162/neco.1989.1.4.541

56. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // *Proc. Natl. Acad. Sci. USA.* 1982. V. 79. No. 8. P. 2554–2558. doi: 10.1073/pnas.79.8.255

57. Liou C.Y., Cheng W.C., Liou J.W., Liou D.R. Autoencoder for words // *Neurocomputing.* 2014. V. 139. P. 84–96. doi: 10.1016/j.neucom.2013.09.055

58. Oh M., Zhang L. DeepMicro: deep representation learning for disease prediction based on microbiome data // *Sci. Rep.* 2020. V. 10. No. 1. Article No. 6026. doi: 10.1038/s41598-020-63159-5

59. Cuomo S., de Rosa M., Piccialli F., Pompameo L., Vocca V. A numerical approach for soil microbiota growth prediction through physics-informed neural network // *Appl. Numer. Math.* 2024. V. 207. P. 97–110. doi: 10.1016/j.apnum.2024.08.025

60. Papoutsoglou G., Tarazona S., Lopes M.B., Klammersteiner T., Ibrahim E., Eckenberger J., Novielli P., Tonda A., Simeon A., Shigdel R., Béreux S., Vitali G., Tangaro S., Lahti L., Temko A., Claesson M.J., Berland M. Machine learning approaches in microbiome research: challenges

and best practices // *Front. Microbiol.* 2023. V. 14. Article No. 1261889. doi: 10.3389/fmicb.2023.1261889

61. Deng Z., Zhang J., Li J., Zhang X. Application of deep learning in plant–microbiota association analysis // *Front. Genet.* 2021. V. 12. Article No. 697090. doi: 10.3389/fgene.2021.697090

62. Zhao B., Jin W., Del Ser J., Yang G. ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification // *Neurocomputing.* 2023. V. 557. Article No. 126708. doi: 10.1016/j.neucom.2023.126708

63. Popova L.V., Gorshkova N.V., Shaldokhina S.Yu. The introduction of Agriculture technology 4.0: conditions and forecasts // *Bulletin of the Adyghe State University, Series “Economics”.* 2019. No. 1. P. 83–89 (in Russian).

64. Zayats O.A., Nazarova Yu.N., Strizhakova E.A., Penkova R.I. Big Data technologies in agriculture // *Fundamental research.* 2022. No. 7. P. 35–40 (in Russian). doi: 10.17513/fr.43280

65. Cravero A., Pardo S., Sepúlveda S., Muñoz L. Challenges to use machine learning in agricultural Big Data: a systematic literature review // *Agronomy.* 2022. V. 12. No. 3. Article No. 748. doi: 10.3390/agronomy12030748

66. Raevskaya E.G. Introducing artificial intelligence in Chinese agriculture (review) // *Agrarnaya nauka Euro-Severo-Vostoka.* 2024. V. 25. No. 5. P. 739–753 (in Russian). doi: 10.30766/2072-9081.2024.25.5.739-753

67. Fenu G., Mallocci F.M. DSS LANDS: a decision support system for agriculture in Sardinia // *HighTech and Innovation Journal.* 2020. V. 1. No. 3. P. 129–135. doi: 10.28991/HIJ-2020-01-03-05

68. Folorunso O., Ojo O., Busari M., Adebayo M., Joshua A., Folorunso D., Ugwunna C.O., Olabanjo O., Olabanjo O. Exploring machine learning models for soil nutrient properties prediction: a systematic review // *Big Data Cogn. Comput.* 2023. V. 7. No. 2. Article No. 113. doi: 10.3390/bdcc7020113

69. Shen Y., Song K., Tan X., Li D., Lu W., Zhuang Y. HuggingGPT: solving AI tasks with ChatGPT and its friends in Hugging Face // *Advances in Neural Information Processing Systems: proceedings of the 37th conference on neural information processing systems.* 2024. V. 36. P. 38154–38180. doi: 10.52202/075280-1657

70. Zhu H., Qin S., Su M., Lin C., Li A., Gao J. Harnessing large vision and language models in agriculture: a review // *Front. Plant Sci.* 2025. V. 16. Article No. 1579355. doi: 10.3389/fpls.2025.1579355

71. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision // *Proceedings of the 40th International Conference on Machine Learning.* Honolulu, 2023. V. 202. P. 28492–28518.

72. Ren Y., Ruan Y., Tan X., Qin T., Zhao S., Zhao Z., Liu T.Y. FastSpeech: Fast, robust and controllable text to speech // *Advances in neural information processing systems: proceedings of the 33rd conference on neural information processing systems.* Vancouver, 2019. V. 32. No. 285. P. 3171–3180.