

Геоинформатика. 2024. № 4. С. 83–92.  
*Geoinformatika*. 2024;(4):83–92.

### Искусственный интеллект в прикладных областях знаний

Научная статья  
 УДК 004.8, 303.732, 550.8.01  
<https://doi.org/10.47148/1609-364X-2024-4-83-92>

## Базовые задачи искусственного интеллекта на примере геологоразведки

© 2024 г. — Евгения Наумовна Черемисина<sup>1, 2, a)</sup>, Елена Юрьевна Кирпичева<sup>2, b)</sup>, Надежда Александровна Токарева<sup>2, c)</sup>, Анна Александровна Миловидова<sup>2, d)</sup>

<sup>1</sup>ФГБУ «Всероссийский научно-исследовательский геологический нефтяной институт» (ФГБУ «ВНИГНИ»); Москва, Россия

<sup>2</sup>Государственный университет «Дубна»; Дубна, Россия

<sup>a)</sup>chere@uni-dubna.ru, <sup>b)</sup>kirphel@mail.ru, <sup>c)</sup>tokareva@uni-dubna.ru, <sup>d)</sup>milanna@uni-dubna.ru

**Аннотация:** Статья посвящена исследованию базовых задач искусственного интеллекта, возникающих в процессе мультидисциплинарных исследований, и методов их решения с применением машинного обучения и нейронных сетей. Авторы акцентируют внимание на важности системного подхода, который позволяет корректно формулировать задачи, особенно в слабоформализованных науках, таких как геология и экология. В статье рассматриваются ключевые задачи искусственного интеллекта, включая ретрогноз, прогноз, поиск и конструирование, а также предлагаются методы их решения, такие как кластеризация и регрессионный анализ. Важное место отведено рассмотрению использования методов объяснимого искусственного интеллекта (XAI), который повышает интерпретируемость моделей и способствует более глубокому пониманию сложных процессов в междисциплинарных исследованиях. Особое внимание уделено применению голотипных алгоритмов, которые эффективно решают задачи классификации и распознавания объектов на основе анализа многомерных данных. Работа подчеркивает значимость использования искусственного интеллекта для автоматизации и повышения эффективности научных исследований в междисциплинарных областях.

**Ключевые слова:** *искусственный интеллект; мультидисциплинарные исследования; объяснимый искусственный интеллект; машинное обучение; нейронные сети; системный подход.*

Для цитирования: Черемисина Е.Н., Кирпичева Е.Ю., Токарева Н.А., Миловидова А.А. Базовые задачи искусственного интеллекта на примере геологоразведки // *Геоинформатика*. — 2024. — № 4. — С. 83–92. <https://doi.org/10.47148/1609-364X-2024-4-83-92>.

### Artificial intelligence in applied fields of knowledge

Original article

## Basic artificial Intelligence tasks in context of geological prospecting

© 2024 — Eugeniya N. Cheremisina<sup>1, 2, a)</sup>, Elena Y. Kirpicheva<sup>2, b)</sup>, Nadezhda A. Tokareva<sup>2, c)</sup>, Anna A. Milovidova<sup>2, d)</sup>

<sup>1</sup>All-Russian Research Geological Petroleum Institute (VNIGNI); Moscow, Russia

<sup>2</sup>Dubna State University; Dubna, Russia

<sup>a)</sup>chere@uni-dubna.ru, <sup>b)</sup>kirphel@mail.ru, <sup>c)</sup>tokareva@uni-dubna.ru, <sup>d)</sup>milanna@uni-dubna.ru

**Abstract:** The article focuses on the fundamental challenges of artificial intelligence (AI) encountered in multidisciplinary research and methods for addressing them using machine learning and neural networks. The authors emphasize the importance of a systematic approach, which enables accurate problem formulation, especially in poorly formalized domains such as geology and ecology. The study examines key AI tasks, including retrodiction, prediction, search, and design, while proposing solutions such as clustering and regression analysis. Special attention is given to the application of explainable artificial intelligence (XAI) methods, which enhance model interpretability and foster a deeper understanding of complex processes in interdisciplinary studies. The use of holotypic algorithms, which effectively address classification and object recognition tasks based on multidimensional data analysis, is highlighted. The paper underscores the significance of AI in automating and improving the efficiency of scientific research in interdisciplinary fields.

**Key words:** *artificial Intelligence; multidisciplinary research; explainable artificial intelligence; machine learning; neural networks; systematic approach.*

For citation: Cheremisina E.N., Kirpicheva E.Y., Tokareva N.A., Milovidova A.A. A Basic artificial Intelligence tasks in context of geological prospecting. *Geoinformatika*. 2024;(4):83–92. <https://doi.org/10.47148/1609-364X-2024-4-83-92>. In Russ.

### Введение

Все науки делятся на **описательные**, которые описывают некоторое явления, законы и следствия, и **предсказательные**, которые на основании выявленных закономерностей предсказывают те или

иные явления или объекты. Каждая наука в своем историческом развитии движется от статуса описательной к статусу предсказательной.

Все предсказательные науки изучают **модели** природы и общества, а не изучают саму природу

или общество. Точность предсказаний зависит от адекватности и вероятности предсказаний представленной модели. Если модель объекта исследования охватывает достаточно обширный круг процессов и явлений, рассматриваемых данной наукой, наука называется «**формализованной**». В современной физике фигурируют 2-3 модели, описывающие явления и процессы, поэтому и является примером **формализованной** науки. Если модель должна создаваться отдельно для каждого объекта, явления или процесса, то такая наука называется «**слабоформализованной**». К таким наукам, в частности, относится геология, экология, медицина и др.

В формализованной науке для решения задачи необходимо собрать информацию, подобрать подходящую модель и перевести проблему на язык математики с дальнейшим решением в этой области. В слабоформализованных науках получить решение задачи, как правило, намного сложнее, так как не существует модели рассматриваемого объекта или явления, поэтому для решения таких задач необходимо разработать схему **постановки и решения** [1].

Однако необходимо отметить, что разные ученые используют различные представления о задачах, их постановке и решении, по-разному оценивают полученные результаты. Это заставляет вернуться к постановке и решению мультидисциплинарных задач.

### 1. О системном подходе и мультидисциплинарных исследованиях

Известны различные толкования системного подхода [2–4]. Его сутью являются такие представления о научной деятельности, которые позволяют использовать компьютерные распределенные средства для решения мультидисциплинарных научных проблем [5–7].

Первой особенностью системного подхода в науке является фиксация значимости метанаучных вопросов.

Второй особенностью системного подхода в науке является понимание необходимости пересмотра соотношений между целями, проблемами, задачами, моделями и теориями.

Потребности общества формируют исходные представления об общественно-значимых целях. Научное сообщество с учетом этих целей и возможностей контроля степени их достижения порождает исходные проблемы. С учетом возможностей и условий разрешения этих проблем ученые по-разному формулируют постановку задачи. В зависимости от специфики рассматриваемой проблемы выбирается соответствующая модель.

Кратко говоря, системный подход в науке — это такой подход к проблемам, когда главными считаются ответы на вопросы: на каких основаниях, как и кто сформулировал проблемы, кто показал возможности их разрешения и перспективы использования

полученных результатов, кто и как сформулировал и обосновал постановки задач, отвечающих этим проблемам.

Таким образом, системный подход к проблемам в науке — последовательное проведение их методологизации, формализации, теоретизации, математизации и компьютеризации.

В качестве приоритетных рассматриваются проблемы, связанные с управлением качеством населения, уровнем его жизни, использованием окружающей среды, которые являются мультидисциплинарными.

К мультидисциплинарным научным проблемам относятся такие проблемы, постановка которых не может быть реализована с помощью модели, используемой в одной из наук. Они требуют согласованного участия ученых из разных научных сфер.

В рамках данного исследования искусственный интеллект является основой для постановки и решения мультидисциплинарных задач, которые могут быть формально представлены последовательностью задач: ретрогноз, поиск, разведка, конструирование и др. Если для прикладной математики главное — эффективное решение уже поставленных задач, то для задач искусственного интеллекта главное — их эффективная постановка, а также проектирование подходящих программных средств.

Существенно, что искусственный интеллект может помочь любому ученому сделать свои задачи понятными любому другому ученому.

### 2. Развитие представлений о задачах искусственного интеллекта

Зафиксируем общий перечень мультидисциплинарных задач, опирающихся на комплексные данные:

1. Ретрогноз — задача направлена на анализ происхождения, а также условий и процессов возникновения объекта исследования.
2. Прогноз — предсказание будущего состояния или поведения объекта исследования.
3. Поиск — идентификация объекта исследования, определение его категории, вида и т. д.
4. Разведка — поиск объекта исследования и оценка методов его изучения и использования.
5. Строение — формирование пространственно-временной структуры свойств, характеристик рассматриваемого объекта.
6. Конструирование — создание нового объекта.

На рис. 1 представлено описание мультидисциплинарных задач и применяемых методов их решения.

Как следует из опыта решения этих задач, базовыми формальными задачами являются следующие:

- выделение объектов исследования;

Рис. 1. Классификация мультидисциплинарных задач и методы их решения

Fig. 1. Classification of multidisciplinary problems and methods for their solution



- описание и сравнение объектов исследования;
  - разделение объектов исследования на перспективные и неперспективные;
  - упорядочение объектов исследования по степени перспектив;
  - оценивание объектов исследования;
  - прогнозирование объектов исследования [4].
- Эти задачи в дальнейшем будем называть базовыми задачами искусственного интеллекта.

Главные трудности в проведении мультидисциплинарных исследований обусловлены тем, что их участники имеют существенно разные представления о задачах, опираются на различные методолого-теоретические установки, используют разные языки, по-разному относятся к уже имеющимся эмпирическим данным и способам их получения, изолированно обучаются и не имеют опыта совместной работы.

Важная особенность мультидисциплинарных задач заключается в возможности получить для них много различных математических постановок. При этом выбор предпочтительной заведомо находится вне отдельных дисциплин и вне математики.

Рассмотрим следующую схему постановки и решения мультидисциплинарных задач:

1. Математическая постановка задачи и ее обоснование.
2. Построение семейства алгоритмов решения задачи и его обоснование.
3. Создание инструментария для решения задачи (например, программных комплексов).
4. Получение решений задач в конкретных ситуациях.
5. Популяризация полученных результатов решения задачи и построенного инструментария [9].

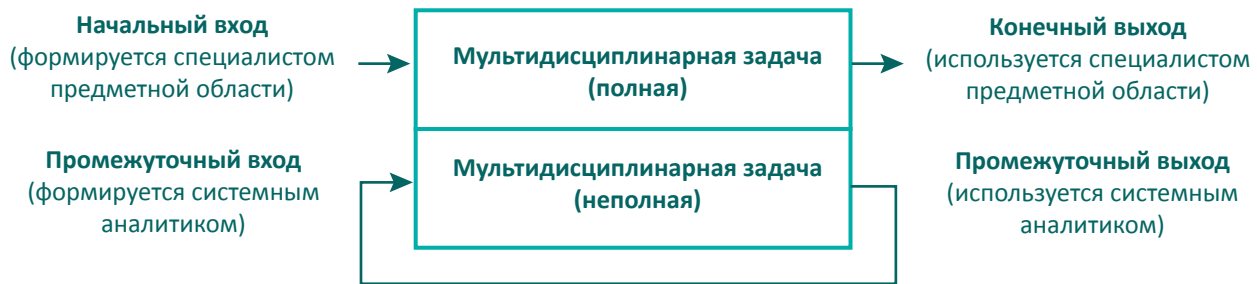
Те, кто реализует отдельные этапы, должны составлять единую команду, которая владеет языком общения, согласованными этическими и методолого-теоретическими нормами, общей и специальной подготовкой. Необходимо совместно рассматривать формирование научных команд постановки и решения мультидисциплинарных задач. При этом в отдельных научных дисциплинах важно одинаково фиксировать представление иерархии задач. В задачах всегда можно выделить их вход и выход (рис. 2).

Вход и выход задачи могут быть объективно или субъективно верифицируемыми.

Задачу, у которой вход является начальным, а выход — конечным, будем считать полной, все другие задачи — неполными [1].

Рис. 2. Модель мультидисциплинарной задачи вход–выход

Fig. 2. Model of multidisciplinary problem “input–output”



Для успешного решения междисциплинарных задач критически важным фактором является коллаборация специалистов различных дисциплин. С позиций предметно-ориентированного подхода должен быть сформирован единый язык предметной области, который называется «ядром предметной области». Ядро включает термины и понятия, которые используются во всех научных дисциплинах междисциплинарной задачи, а также специализированную терминологию, адаптированную в рамках команды для удобства взаимодействия. Помимо языка, необходима единая методологическая база, которая регламентирует, как будут ставиться задачи, осуществляться работа с данными и оцениваться результат.

Для успешного освоения мультидисциплинарных задач необходимо наличие общепринятой схемы, которая фиксирует содержание ее составляющих и позволяет говорить о разделении функций и ответственности его участников. Ниже предложены контуры такой схемы.

Можно считать, что она должна содержать следующее:

#### 1. Истоки задачи

- Краткое описание исходной проблемы, породившей задачу, примеры ее толкования и рассмотрения.

- Краткое описание основных компонентов задачи: что дано, что требуется и каковы ограничения.

- Возможности модификации основных компонентов задачи и основания для различения постановок задачи.

#### 2. Описание теоретической цели (ТЦ)

- Множество допустимых прямых способов достижения ТЦ:

- допустимых, т. е. отвечающих некоторым методолого-теоретическим требованиям;

- прямых, т. е. точных, простых, многозатратных.

- Оптимальный прямой способ достижения ТЦ:

- оптимальный, т. е. выбранный из множества допустимых прямых способов по некоторому

критерию и позволяющий точно достичь ТЦ при приемлемых затратах.

- Множество допустимых косвенных способов достижения ТЦ:

- допустимых, т. е. отвечающих некоторым методолого-теоретическим требованиям, возможно, более сильным чем в множествах допустимых прямых способов достижения ТЦ;

- косвенных, т. е. неточных, возможно, сложно реализуемых, но малозатратных и практически всегда реализуемых.

- Условия рациональности допустимых косвенных способов достижения ТЦ:

- рациональных, т. е. приемлемых по точности, сложности и затратам, пригодных к массовому использованию.

- Условия и критерии эффективности косвенных способов достижения ТЦ.

- эффективных, т. е. обеспечивающих соизмеримость проигрыша в точности и выигрыша в затратах с учетом направления и способа использования факта достижения ТЦ, а также требований пользователей.

- Априорные предположения для выделения, минуя полный перебор, из допустимых всех рациональных и из рациональных одного эффективного или предпочтительного косвенного способа достижения ТЦ.

- Исходные эмпирические данные, связанные с достижениями ТЦ в некоторых конкретных ситуациях оптимальным прямым и эффективным косвенным способами [9].

Под формулировкой мультидисциплинарной задачи будем понимать указание цели, исходные данные, представление о модели изучаемых объектов, ожидаемого результата и критериев его оценки. Для трансформации постановки задачи из содержательной формы в формальную используется язык теории множеств и анализа данных.

Когда для формальной задачи определен математический метод решения, будем говорить о математической постановке задачи. Если при решении

математической задачи требуется выбрать один из нескольких возможных вариантов постановки, будем говорить о задаче искусственного интеллекта.

При этом решение задачи проводится при ее уже заданной постановке, учитывающей возможности последующего использования результатов решения, а решение реализуется с учетом соответствия эмпирических и вычисленных данных.

### 3. Возможности машинного обучения и нейронных сетей для решения базовых задач искусственного интеллекта

Машинное обучение и нейронные сети открывают широкие возможности для решения фундаментальных задач искусственного интеллекта в различных областях науки и техники. Эти технологии позволяют автоматизировать сложные процессы анализа данных, что приводит к новым открытиям и повышению качества решений в междисциплинарных исследованиях. На этапе выбора алгоритма и модели необходимо учитывать специфические особенности решаемой задачи. Рассмотрим в качестве примера задачу анализа структуры месторождения. Глубокое понимание предметной области, в нашем случае классификаций пород месторождения на основе сейсмических исследований, взаимодействия тектонических плит, процесса эрозии, обеспечивает успешную интеграцию этих технологий. Таким образом, глубокое понимание предметной области позволяет правильно подобрать параметры модели для оценки месторождений.

Рассмотрим на примерах, как системный подход находит отражение в применении методов машинного обучения, нейронных сетей в междисциплинарных задачах, описанных выше, в частности в области геологии.

Одной из особенностей системного подхода является иерархическая упорядоченность задач, иными словами, разделение сложных на более простые задачи сверху вниз. В задаче идентификации мест залежей нефти задача делится на такие этапы, как:

- анализ сейсмических данных;
- кластеризация геологических структур в соответствии с типом пород;
- определение вероятности нахождения нефти в соответствии с глубинами залегания и тектонической активностью в рассматриваемом районе.

Для решения каждой из вышеперечисленных задач используются разные методы, такие как кластеризация, регрессионный анализ или нейронные сети.

Одним из ключевых принципов системного анализа является многоаспектность рассмотрения сложных систем. Этот принцип позволяет при прогнозировании месторождений полезных ископаемых учитывать не только геофизические и геохимические данные, но и влияние климатических региональных условий, а также, что особенно

важно, исторические данные о геологических процессах. Комплексный междисциплинарный подход позволяет учесть важнейшие аспекты и необходимые переменные при построении предиктивной модели.

Еще одна существенная особенность применения системного подхода в решении задач с помощью методов машинного обучения и нейронных сетей — это возможность адаптации моделей с учетом обратной связи. Такие интеллектуальные решения итеративно совершенствуются за счет пополнения данных, переобучения моделей в ранее рассматриваемой задаче прогнозирования месторождений, например, с помощью дополнительных данных о буровых работах.

В таблице представлены базовые формальные задачи и основные методы их решения. Остановимся подробнее на конкретных примерах решения таких задач.

Процесс выделения объектов хорошо иллюстрирует задача обнаружения идентификации зон с высоким уровнем загрязнения на основе данных о концентрациях различных вредных веществ в рамках экологического мониторинга территории. В частности, метод *K*-Means здесь позволит сгруппировать регионы по уровням загрязнения и выделять наиболее критичные области с самыми высокими концентрациями вредных веществ. Следует отметить, что на качество полученных результатов влияет число выбранных кластеров. Для визуализации многомерных данных на трехмерных прямоугольных картах с гексагональной решеткой популярным методом являются самоорганизующиеся карты Кохонена.

В качестве примера описания и сравнения объектов исследования рассмотрим задачу классификации загрязнений водоемов. Набор данных содержит информацию о концентрациях тяжелых металлов. SVM и логистическая регрессия позволят классифицировать водоемы по уровням опасности. В свою очередь глубокие нейронные сети позволят решить задачу мониторинга динамики и выявления источников вредных веществ на основе спутниковых снимков.

В предиктивной аналитике эффективным инструментом решения задачи являются рекуррентные нейронные сети (RNN). Например, простейшая линейная регрессия позволяет прогнозировать урожайность на основе таких данных, как погодные условия, состав почвы и используемые удобрения.

Линейная регрессия проста в использовании и хорошо работает с линейными данными, но ограничена в случаях сложных нелинейных зависимостей. Поэтому, когда данные включают временные паттерны, например, сезонные изменения, то более эффективный результат дают рекуррентные нейронные сети (например, LSTM). Недостаток применения LSTM состоит в высокой ресурсоемкости и за-

Таблица. Методы решения базовых задач

Table. Methods for solving basic problems

<b>1. Выделение объектов исследования</b> В рамках задачи выделения объектов исследования осуществляется идентификация объектов или областей в большом наборе данных	Кластеризация методом <i>K</i> -средних ( <i>K</i> -Means) разбивает данные на кластеры для минимизации различий внутри кластеров. В методе <i>K</i> -средних используется в задачах разделения данных на несколько групп на основе их сходства
	Иерархическая кластеризация позволяет строить иерархическую структуру кластеров, начиная с каждого объекта как отдельного кластера и последовательно объединяя их на основе заданного критерия сходства
	Карты самоорганизации Кохонена представляют собой тип нейронных сетей, которые применяются для визуализации и кластеризации высокоразмерных данных, сохраняя топологические отношения между данными
<b>2. Описание и сравнение объектов исследования</b> Описание и сравнение объектов исследования позволяет лучше понять структуру и взаимосвязи в данных	Логистическая регрессия — это простой, но мощный метод классификации, который часто используется для двухклассовых задач
	Для разделения данных на классы, максимизируя границу разделения между ними применяется метод опорных векторов (SVM)
	В задачах классификации сложных данных, таких как изображения, текст, речь хороший результат показывают глубокие нейронные сети (DNN)
<b>3. Прогнозирование объектов исследования</b> Прогнозирование — это одна из наиболее сложных задач искусственного интеллекта, особенно в условиях неопределенности и изменчивости данных. В случае, когда между переменными наблюдается линейная зависимость, применяется линейная регрессия. Для прогнозирования данных, имеющих временную структуру, эффективным инструментом решения задачи являются рекуррентные нейронные сети (RNN)	
<b>4. Упорядочение объектов по степени уникальности</b> Упорядочение объектов позволяет ранжировать объекты по их значимости или уникальности, что может быть полезно в ряде случаев. Для решения задачи ранжирования используется метод главных компонент (PCA). Метод позволяет уменьшить размерность данных с сохранением наиболее значимых признаков	
<b>5. Оптимизация признакового пространства</b> Для оптимизации признакового пространства, снижения сложности модели, улучшения ее интерпретируемости и предотвращения переобучения применяются методы на основе деревьев решений	

тухании градиента, что может затруднять обучение на длинных последовательностях данных.

Упорядочение объектов позволяет ранжировать объекты по их значимости или уникальности, что может быть полезно в ряде случаев. Для решения задачи ранжирования используется метод главных компонент (PCA). Метод позволяет уменьшить размерность данных с сохранением наиболее значимых признаков.

В задаче поиска перспективных месторождений полезных ископаемых PCA может использоваться для выделения уникальных характеристик различных территорий и их последующего ранжирования по перспективности освоения. PCA, как и факторный анализ, полезен для работы с высокоразмерными данными, однако результаты могут быть трудноинтерпретируемыми. Это особенно важно, когда компоненты или факторы не имеют явной физической интерпретации, что затрудняет понимание результатов.

Для оптимизации признакового пространства, снижения сложности модели, улучшения ее интерпретируемости и предотвращения переобучения применяются методы на основе деревьев решений.

Выбор методов оптимизации зависит от характера задачи и данных. Методы отбора признаков позволяют автоматически исключить ненужные переменные, однако требуют тщательной настройки для предотвращения потери важной информации. Факторный анализ, в свою очередь, помогает выявить скрытые связи между признаками, но может быть трудным для интерпретации и внедрения в модели.

Таким образом, машинное обучение и нейронные сети предоставляют мощные инструменты для решения базовых задач искусственного интеллекта, особенно в контексте мультидисциплинарных исследований, что обеспечивает возможность системного подхода к исследованию и решению сложных научных проблем и позволяет эффективно формулировать, моделировать и решать междисциплинарные задачи, основанные на данных.

Однако использование сложных моделей искусственного интеллекта порождает новые вызовы, связанные с интерпретируемостью и доверием к результатам. Чтобы преодолеть эти трудности и обеспечить высокую степень прозрачности, возникает необходимость в применении объяснимого искусственного интеллекта (Explainable AI, XAI).

#### 4. Применение объяснимого искусственного интеллекта в мультидисциплинарных исследованиях

Современные научные исследования, особенно в мультидисциплинарных областях, все чаще сталкиваются с необходимостью использования сложных моделей искусственного интеллекта. Эти модели, часто обладающие высокой прогностической способностью, нередко воспринимаются как «черные ящики», поскольку их внутренние механизмы остаются неясными для пользователя. В ответ на эту проблему развивается направление объяснимого искусственного интеллекта (Explainable AI, XAI), основной целью которого является обеспечение интерпретируемости и прозрачности моделей машинного обучения. Объяснимый искусственный интеллект становится особенно важным в тех случаях, когда от его решений зависят критически значимые выводы и управленческие решения. Это особенно актуально для таких дисциплин, как геология, экология и медицина. В этих областях модели не только должны демонстрировать высокую точность, но и предоставлять понятные объяснения своих прогнозов, чтобы специалисты могли их использовать для анализа процессов, разработки гипотез и подтверждения научных теорий.

Одним из эффективных инструментов XAI является использование методов, основанных на прототипах. Эти методы выделяют характерные примеры из данных, которые модель использует в качестве эталонов для классификации и прогнозирования. Например, в задачах классификации геологических объектов прототипы помогают установить, какие параметры сыграли ключевую роль в принятии решения моделью. Это значительно облегчает понимание сложных взаимосвязей в многомерных данных [10].

Важным дополнением к методам XAI является использование стратегий агрегации признаков. В геонауках, где часто работают с высокоразмерными и коррелированными данными (например, с растровыми картами), такие стратегии помогают структурировать и упрощать анализ. Это позволяет исследователям выделять ключевые факторы, влияющие на процессы, и минимизировать влияние шума данных. Упрощенные представления признаков обеспечивают более детализированные и точные объяснения работы моделей [11].

Еще одной перспективной технологией в рамках XAI является нейросимволический искусственный интеллект (NSAI). Этот подход сочетает пре-

имущества глубинного обучения, которое хорошо справляется с анализом больших объемов данных и символических методов, обеспечивающих логическую интерпретацию результатов. NSAI позволяет учитывать разнородные данные, что особенно важно в задачах предсказания месторождений полезных ископаемых. Символическая часть модели может объяснять, как геофизические, геохимические и пространственные данные взаимодействуют между собой, а также как они влияют на прогноз [12].

Применение XAI значительно расширяет возможности междисциплинарных исследований, обеспечивая понимание сложных процессов и доверие к результатам анализа. В геонауках XAI используется для анализа климатических явлений, прогнозирования природных катастроф, выявления зон повышенного риска, а также для моделирования процессов в недрах Земли. Эти модели не только повышают точность прогнозов, но и предоставляют ученым важные объяснения, которые помогают корректировать и развивать научные гипотезы [13].

Таким образом, развитие XAI открывает новые горизонты в мультидисциплинарных исследованиях, обеспечивая ученых инструментами для более глубокого анализа данных и повышения прозрачности результатов. Эффективное использование XAI позволяет интегрировать знания из различных научных областей и повышает доверие к результатам моделирования.

#### 5. Голотипные алгоритмы и их применение в междисциплинарных задачах

Опираясь на рассмотренные выше примеры, можно сделать вывод, что для более точной работы с разнородными выборками данных в задачах таксономии, распознавания различных объектов на основе многомерных данных необходимы специализированные алгоритмы. Одним из них является семейство голотипных алгоритмов.

Семейство голотипных алгоритмов включает в себя голотипную таксономию, частичную таксономию, распознавание на 1 образ (голотип 1), распознавание на  $N$  образов (голотип  $N$ ) и алгоритм поиска самых нетипичных объектов в выборке (направление опробования). Все эти алгоритмы основаны на мере сходства и построении кратчайшего графа, соединяющего все объекты выборки, о чем будет сказано ниже.

Мерой сходства называется функция  $\mu$ , заданная на множестве пар объектов и удовлетворяющая следующим условиям:

1.  $0 \leq \mu(a, b) \leq 1$ ;
2.  $\mu(a, b) = \mu(b, a)$ ;
3.  $\mu(a, a) = 1$ .

Из определения видно, что класс мер очень широк. Если наши объекты — точки в  $N$ -мерном

пространстве признаков, то чаще всего в качестве меры сходства используют следующую функцию:

$$\mu(a, b) = 1 - \sum_i v_i \mu_i(a_i, b_i),$$

где  $v_i$  — вес  $i$ -го свойства;  $\mu_i(a_i, b_i)$  — мера сходства между  $i$ -ми координатами объектов  $a$  и  $b$ .

Если эта координата измерена в количественной шкале, то обычно

$$\mu_i(a_i, b_i) = |a_i - b_i| / e_i,$$

где  $e_i$  — экстремальная разность, т. е. разность между максимальным и минимальным значением в данном свойстве.

Веса должны быть положительны, а их сумма равна 1.

На базе меры сходства между объектами можно решать классификационные задачи. К ним относятся задача районирования, т. е. разбиения множества объектов на группы близких между собой, задача распознавания, т. е. разделение всех объектов на классы, соответствующие известным эталонным объектам (объектам обучения), и занимающая промежуточное место задача частичной таксономии, т. е. задача, когда эталонный материал не представляет всего множества объектов и, соответственно, некоторые выделенные при работе алгоритма группы не могут быть сопоставлены ни с каким эталонным классом.

Естественно было бы рассчитать матрицу мер сходства между всеми объектами и на основании работы с этой матрицей решать отмеченные выше задачи, однако при современных запросах, когда число объектов зачастую превосходит 1 млн, держать квадратную матрицу (и даже ее половину, поскольку она симметричная) в оперативной памяти совершенно нереально, а работа с диском не даст возможности получить решение в разумное время. Для того чтобы обойти эту трудность, строится скелет полного графа, имеющий минимальную длину. Это производится с помощью итерационной процедуры: выбирается какой-то объект, находится ближайший к нему и присоединяется ребром, далее находится ближайшей к паре, тройке, четверке и т. д. Каждый раз новый объект присоединяется ребром до исчерпания всего множества объектов.

Можно доказать, что полученный таким образом граф имеет минимальную длину из всех графов, содержащих все объекты и не зависит от выбора первого объекта.

Теперь задача районирования решается разрыванием нескольких самых длинных ребер графа. При этом он распадется на подграфы, и объекты, входящие в один подграф, будут составлять группу близких между собой объектов. Таков принцип работы **голотипной таксономии**.

Если у нас имеется несколько эталонных классов, то мы должны по очереди разрывать наиболее длинные ребра до тех пор, пока не окажется, что в каждом полученном подграфе имеются эталонные

объекты только одного класса. Если объекты обучения были представительны, то в каждом подграфе окажутся эталонные объекты и тем самым задача разделения будет решена (**голотип N**). Если же эталонные объекты не представляли всю совокупность объектов, то некоторые подграфы не будут содержать каких-либо эталонных объектов и, следовательно, не будут отнесены ни к какому из эталонных классов (**частичная таксономия**).

В предположении, что у нас имеются эталоны только одного (продуктивного) класса, необходимо разрывать самые длинные ребра в полученном нами графе до тех пор, пока все эталонные объекты остаются в одном подграфе. Процесс останавливается, когда следующее разделение приведет к тому, что эталоны окажутся в двух разных подграфах. Таким образом, объекты, входящие в тот же подграф, что и эталоны, будут отнесены к продуктивному классу, а объекты, входящие в другие подграфы — к непродуктивному. Этот алгоритм называется **«голотип 1»**.

Предположим теперь, что мы работаем на неисследованной площади и у нас имеются только косвенные свойства. Пусть нашей целью является скорейшее опознавание хотя бы одного продуктивного объекта. Можно сделать следующие предположения:

1. Продуктивных объектов мало, и они не похожи друг на друга и на непродуктивные.
2. Непродуктивных объектов много, и они похожи друг на друга [14].

В этом случае решение предоставляется алгоритмом **Направление опробования**. В процессе его работы при делении графа на компоненты нас будут интересовать только подграфы, включающие малое число объектов. Эти объекты упорядочиваются в том же порядке как соответствующие компоненты отделяются от основного графа. Легко заметить, что вышеописанные предположения при этом выполняются. Можно ожидать, что исследование выделенных нами объектов в указанном порядке приведет к открытию нового продуктивного объекта.

### Заключение

В статье рассмотрены базовые задачи искусственного интеллекта в контексте мультидисциплинарных исследований с акцентом на системный подход и методы решения слабоформализованных задач. Использование машинного обучения и нейронных сетей позволяет значительно повысить точность и автоматизировать процессы анализа данных, что особенно актуально в таких областях, как геология и экология. Отдельно подчеркнута роль ХАИ, который обеспечивает прозрачность моделей и облегчает интерпретацию их результатов, что важно для доверия и принятия решений в междисциплинарных исследованиях. Голотипные алгоритмы показали свою эффективность в задачах класси-

фикации и распознавания, демонстрируя гибкость при работе с многомерными данными и большими выборками. Таким образом, применение методов искусственного интеллекта способствует более быстрому и качественному решению сложных меж-

дисциплинарных задач, обеспечивая интеграцию знаний из различных научных областей и улучшение результатов исследований.

*Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124092700007-4 Применение объяснительного искусственного интеллекта для интерпретации алгоритмов машинного обучения).*

#### Список источников

1. Воронин А.Ю., Воронин Ю.А. Задачи искусственного интеллекта и управление персоналом геологоразведочных организаций. – Новосибирск: ИВМИМГ, 2005. – 259 с.
2. Поleshchuk И.А. Системный подход и понятие системы // Наука, техника и образование. – 2015. – № 10 (16). – С. 165–168.
3. Спивак В.А. Системный подход и системное мышление как универсальная компетенция специалиста и руководителя. – Чебоксары: Среда, 2022. – 136 с. DOI 10.31483/a-10424.
4. Далбаева В.Ю., Нюренберг М.Н. К вопросу о трактовке понятия «Системный подход» // Известия Иркутской государственной экономической академии. – 2013. – № 5. – С. 14.
5. Ядав Н. Этика искусственного интеллекта и робототехники: ключевые проблемы и современные способы их решения // Journal of Digital Technologies and Law. – 2023. – Т. 1. – № 4. – С. 955–972. DOI: 10.21202/jdtl.2023.41.
6. Матвеев А.Н., Семенов А.Н. Использование data science и машинного обучения в интеллектуальных системах // E-Scio. – 2023. – № 6 (81). – С. 64–70.
7. Вешнева И.В. Технологии искусственного интеллекта: классификация, ограничения, перспективы и угрозы // Известия Саратовского университета. Новая серия. Серия: Экономика. Управление. Право. – 2023. – Т. 23 – № 4. – С. 428–438. DOI: 10.18500/1994-2540-2023-23-4-428-438.
8. Воронин А.Ю. Районирование территорий на основе искусственного интеллекта и распознавания образов в задачах природопользования : дис. ... докт. техн. наук : 25.00.35. – М., 2007. – 213 с.
9. Чугунов В.С. Ступени контроллинга и структурная устойчивость систем управления организацией // Контроллинг. – 2011. – № 39. – С. 18–22.
10. Narayanan A., Bergen K.J. Prototype-Based Methods in Explainable AI and Emerging Opportunities in the Geosciences // arXiv. – 2024. – arXiv:2410.19856. DOI: 10.48550/arXiv.2410.19856.
11. Krell E., Kamangir H., Collins W., King S.A., Tissot P. Aggregation strategies to improve XAI for geoscience models that use correlated, high-dimensional rasters // Environmental Data Science. – 2023. – Vol. 2. – e45. DOI: 10.1017/eds.2023.39.
12. Chen W., Ma X., Wang Z., Li W., Fan C., Zhang J., Que X., Li C. Exploring neuro-symbolic AI applications in geoscience: implications and future directions for mineral prediction // Earth Science Informatics. – 2024. – Vol. 17. – P. 1819–1835. DOI: 10.1007/s12145-024-01278-7.
13. Jiang S., Sweet L.-B., Blougouras G., Brenning A., Li W., Reichstein M., Denzler J., Shangguan W., Yu G., Huang F., Zscheischler J. How interpretable machine learning can benefit process understanding in the geosciences // Earth's Future. – 2024. – Vol. 12. – Iss. 7. – e2024EF004540. DOI: 10.1029/2024EF004540.
14. Черемисина Е.Н., Костылева Т.В., Курпичева Е.Ю. Методика постановки и решения слабоформализованных геологических задач // Системный анализ в науке и образовании. – 2024. – № 1. – С. 1–13.

#### References

1. Voronin A.Yu., Voronin Yu.A. Zadachi iskusstvennogo intellekta i upravlenie personalom geologorazvedochnykh organizatsii [Tasks of artificial intelligence and personnel management of geological exploration organizations]. Novosibirsk: IVMIMG; 2005. 259 p.
2. Poleshchuk I.A. Sistemnyi podkhod i ponyatie sistemy [Systems approach and the concept of a system]. *Science, Technology and Education*. 2015;(10):165–168.
3. Spivak V.A. Sistemnyi podkhod i sistemnoe myshlenie kak universal'naya kompetentsiya spetsialista i rukovoditelya [System approach and system thinking as a universal competence of a specialist and manager]. Cheboksary: Sreda; 2022. 136 p. DOI 10.31483/a-10424.
4. Dalbaeva V.Yu., Nyurenberg M.N. On interpreting the "system approach" concept. *Baikal Research Journal*. 2013;(5):14.
5. Yadav N. Ethics of artificial intelligence and robotics: key issues and modern ways to solve them. *Journal of Digital Technologies and Law*. 2023;1(4):955–972. DOI: 10.21202/jdtl.2023.41.
6. Matveev A.N., Semenov A.N. Ispol'zovanie data science i mashinnogo obucheniya v intellektual'nykh sistemakh [Using data science and machine learning in intelligent systems]. *E-Scio*. 2023;(6):64–70.
7. Veshneva I.V. Artificial intelligence technologies: classification, limitations, prospects and threats. *Izvestiya of Saratov University. Economics. Management. Law*. 2023;23(4):428–438. DOI: 10.18500/1994-2540-2023-23-4-428-438.
8. Voronin A.Yu. Raionirovanie territorii na osnove iskusstvennogo intellekta i raspoznavaniya obrazov v zadachakh prirodopol'zovaniya [Zoning of territories based on artificial intelligence and pattern recognition in environmental management tasks]: dissertation of Doctor of Technical Sciences : 25.00.35. Moscow, 2007. 213 p.

9. *Chugunov V.S.* Stupeni kontrolirovaniya i strukturnaya ustoichivost' sistem upravleniya organizatsiei [Stages of controlling and structural stability of organization management systems]. *Controlling*. 2011;(39):18–22.
10. *Narayanan A., Bergen K.J.* Prototype-Based Methods in Explainable AI and Emerging Opportunities in the Geosciences. 2024. arXiv:2410.19856. DOI: 10.48550/arXiv.2410.19856.
11. *Krell E., Kamangir H., Collins W., King S.A., Tissot P.* Aggregation strategies to improve XAI for geoscience models that use correlated, high-dimensional rasters *Environmental Data Science*. 2023;2:e45. DOI: 10.1017/eds.2023.39.
12. *Chen W., Ma X., Wang Z., Li W., Fan C., Zhang J., Que X., Li C.* Exploring neuro-symbolic AI applications in geoscience: implications and future directions for mineral prediction. *Earth Science Informatics*. 2024;17:1819–1835. DOI: 10.1007/s12145-024-01278-7.
13. *Jiang S., Sweet L.-B., Blougouras G., Brenning A., Li W., Reichstein M., Denzler J., Shangguan W., Yu G., Huang F., Zscheischler J.* How interpretable machine learning can benefit process understanding in the geosciences. *Earth's Future*. 2024;12(7):e2024EF004540. DOI: 10.1029/2024EF004540.
14. *Cheremisina E.N., Kostyleva T.V., Kirpicheva E.Yu.* Methodology for setting and solving underformalized geological problems. *System analysis in science and education*. 2024;(1):1–13.

Статья поступила в редакцию 18.11.2024 г., одобрена после рецензирования 19.11.2024 г., принята к публикации 25.11.2024 г.  
The article was submitted 18.11.2024; approved after reviewing 19.11.2024; accepted for publication 25.11.2024.

### Информация об авторах

#### Черемисина Евгения Наумовна

Доктор технических наук, профессор  
Заведующий отделением Геоинформатики  
ФГБУ «Всероссийский научно-исследовательский геологический  
нефтяной институт» (ФГБУ «ВНИГНИ»)  
117105 Москва, Варшавское ш., д. 8  
Заведующий кафедрой системного анализа и управления  
Государственный университет «Дубна»  
141980 Дубна, Московская обл., ул. Университетская, д. 19  
e-mail: e.cheremisina@geosys.ru  
ORCID: 0000-0002-6041-8359  
SPIN-код: 1472-2283  
AuthorID: 119359

#### Кирпичева Елена Юрьевна

Кандидат технических наук, доцент  
Директор института системного анализа и управления  
Государственный университет «Дубна»  
141980 Дубна, Московская обл., ул. Университетская, д. 19  
e-mail: kirphel@mail.ru

#### Токарева Надежда Александровна

Кандидат физико-математических наук, доцент  
Заведующий кафедрой информационных технологий  
Государственный университет «Дубна»  
141980 Дубна, Московская обл., ул. Университетская, д. 19  
e-mail: tokareva@uni-dubna.ru

#### Миловидова Анна Александровна

Кандидат технических наук  
Доцент кафедры системного анализа и управления  
Государственный университет «Дубна»  
141980 Дубна, Московская обл., ул. Университетская, д. 19  
e-mail: milanna@uni-dubna.ru

### Information about authors

#### Eugenia N. Cheremisina

Doctor of Technical Sciences, Professor  
Head of the Geoinformatics Department  
All-Russian Research Geological Petroleum Institute (VNIIGNI)  
8, Varshavskoe shosse, 117105, Moscow, Russia  
Head of the Department of Systems Analysis and Management  
Dubna State University  
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia  
e-mail: e.cheremisina@geosys.ru  
ORCID: 0000-0002-6041-8359  
SPIN-код: 1472-2283  
AuthorID: 119359

#### Elena Yu. Kirpicheva

Candidate of Technical Sciences  
Director of the Institute of Systems Analysis and Management  
Dubna State University  
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia  
e-mail: kirphel@mail.ru

#### Nadezhda A. Tokareva

Candidate of Physical and Mathematical Sciences,  
Associate Professor, Head of the Department of Information  
Technologies  
Dubna State University  
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia  
e-mail: tokareva@uni-dubna.ru

#### Anna A. Milovidova

Candidate of Technical Sciences  
Associate Professor of the Department of Systems Analysis and  
Management  
Dubna State University  
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia  
e-mail: milanna@uni-dubna.ru