

## О ВОЗМОЖНОСТЯХ СОВРЕМЕННЫХ МУЛЬТИМОДАЛЬНЫХ ТРАНСФОРМЕРОВ В ОБЛАСТИ АНАЛИЗА ИЗОБРАЖЕНИЙ

П. А. Александров<sup>а</sup>, А. А. Прусаков<sup>б</sup>, Г. Н. Антонова<sup>в</sup>, М. Н. Шахов<sup>г</sup>, С. Е. Стельмак<sup>д</sup>,  
А. В. Беклемишева<sup>е</sup>, В. Г. Саркисов<sup>жк</sup>

НИИЦ «Курчатовский институт», г. Москва, Российская Федерация

<sup>а</sup> Maksimova\_SI@nrcki.ru, <sup>б</sup> prusakov.s@yandex.ru, <sup>в</sup> g\_n\_ant@mail.ru, <sup>г</sup> msha@bk.ru,

<sup>д</sup> Stelmak\_SE@nrcki.ru, <sup>е</sup> Beklemisheva\_AV@nrcki.ru, <sup>жк</sup> Galina\_Sarkisova@mail.ru

*Аннотация:* в работе представлены результаты исследования возможностей анализа изображений, которые широкому кругу пользователей предоставляют два нейросетевых сервиса: ChatGPT-5 mini и DeepSeek-3.1 Thinking. Для этих сервисов в соответствии с новой методикой и на уникальной экспериментальной базе оценивалось качество формирования признаков описаний и выявления аналогий при предъявлении всего четырех обучающих примеров для каждого из двух классов. В экспериментах из 93 предложенных уникальных и автоматически сгенерированных модифицированных тестов Бонгарда ChatGPT-5 mini успешно справился с 15 (16,1%), а DeepSeek-3.1 Thinking — с 17 (18,3%) тестами. Обосновывается утверждение, что, несмотря на несомненные достижения в области решения задачи обучения по малому числу примеров, современная технология контекстного обучения мультимодальных нейросетевых трансформеров имеет принципиальные ограничения.

*Ключевые слова:* компьютерное зрение, машинное обучение, системы искусственного интеллекта, искусственные нейронные сети, анализ изображений, ChatGPT, DeepSeek, трансформеры, большие языковые модели, обучение на малом числе примеров.

*Для цитирования:* Александров П. А., Прусаков А. А., Антонова Г. Н., Шахов М. Н., Стельмак С. Е., Беклемишева А. В., Саркисов В. Г. О возможностях современных мультимодальных трансформеров в области анализа изображений. *Успехи кибернетики*. 2026;7(1):93–103.

*Поступила в редакцию:* 23.11.2025.

*В окончательном варианте:* 18.12.2025.

## THE POTENTIAL OF CURRENT MULTIMODAL TRANSFORMERS FOR IMAGE ANALYSIS

P. A. Alexandrov<sup>a</sup>, A. A. Prusakov<sup>b</sup>, G. N. Antonova<sup>c</sup>, M. N. Shakhov<sup>d</sup>, S. E. Stelmak<sup>e</sup>,  
A. V. Beklemisheva<sup>f</sup>, V. G. Sarkisov<sup>g</sup>

National Research Centre “Kurchatov Institute”, Moscow, Russian Federation

<sup>a</sup> Maksimova\_SI@nrcki.ru, <sup>b</sup> prusakov.s@yandex.ru, <sup>c</sup> g\_n\_ant@mail.ru, <sup>d</sup> msha@bk.ru,

<sup>e</sup> Stelmak\_SE@nrcki.ru, <sup>f</sup> Beklemisheva\_AV@nrcki.ru, <sup>g</sup> Galina\_Sarkisova@mail.ru

*Abstract:* we studied the image analysis capabilities of two widely used neural network services: ChatGPT-5 mini and DeepSeek-3.1 Thinking. We measured the quality of feature generation and analogy matching using a new methodology and a unique experimental framework that employed all four training examples for each of two classes. In experiments with 93 proposed sounds and automatically generated Modified Bongard Tests, ChatGPT-5 mini completed 15 (16.1%) tests, and DeepSeek-3.1 Thinking completed 17 (18.3%). These results demonstrate that, despite clear progress in few-shot learning, current multimodal neural network transformers still face fundamental limitations in contextual learning.

*Keywords:* computer vision, machine learning, artificial intelligence systems, artificial neural networks, image analysis, ChatGPT, DeepSeek, transformers, large language models, few-shot learning.

*Cite this article:* Alexandrov P. A., Prusakov A. A., Antonova G. N., Shakhov M. N., Stelmak S. E., Beklemisheva A. V., Sarkisov V. G. The Potential of Current Multimodal Transformers for Image Analysis. *Russian Journal of Cybernetics*. 2026;7(1):93–103.

*Original article submitted:* 23.11.2025.

*Revision submitted:* 18.12.2025.

### Введение

Цели работы:

- выявить возможности и ограничения «контекстного обучения нейросетей» (в данной работе: ChatGPT-5 [1], DeepSeek-3.1 [2]) на малом числе примеров;
- предложить новую тестовую базу (benchmarks) алгоритмов машинного обучения, которые претендуют в своем развитии стать универсальной системой компьютерного зрения.

### Проблема

В эпоху классического компьютерного зрения [3] и с распространением сверточных нейросетей [4] организовывались различные соревнования распознающих алгоритмов.

С выходом мультимодальных рассуждающих трансформеров оказалось возможным нейросети дообучиться классификации в «новой области» буквально на 1-6 изображениях (few-shot learning) для каждого класса [5].

Проблема, решению которой посвящена настоящая работа, – выявить тесты (benchmarks), пригодные для оценки систем компьютерного зрения на основе машинного обучения, для которой заявлена способность дообучения по малому числу примеров в зрительных задачах.

### Наш вклад

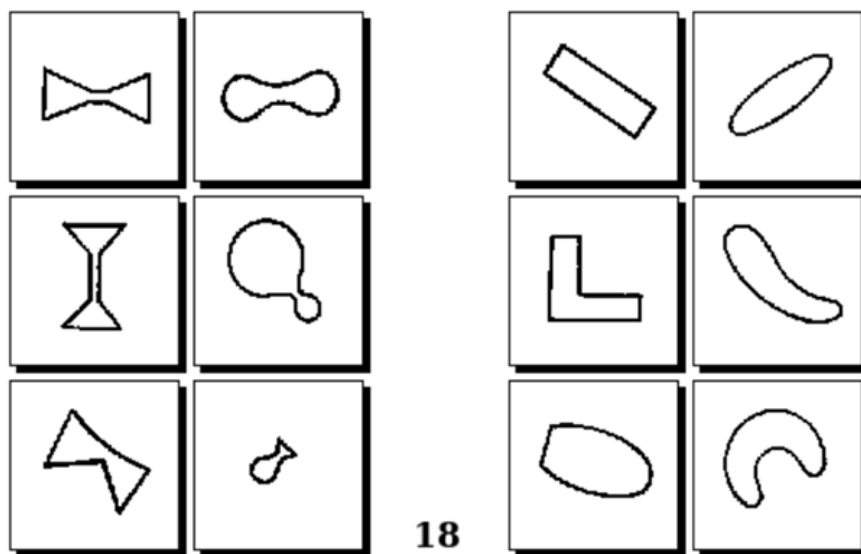
В настоящей работе демонстрируется, что автоматически сгенерированные модифицированные тесты Бонгарда (на этот набор тестов 05.08.2025 получено свидетельство о государственной регистрации базы данных № 2025623774) могут выступать материалом для измерения прогресса в области разработки алгоритмов по решению зрительных задач обучения на малом числе примеров.

### Существующие подходы по измерению «интеллектуальности» алгоритмов компьютерного зрения в части способности узнавания

#### Тесты Бонгарда

Головоломки Бонгарда для «устройства соревнования мыслящих машин» были предложены в 1967 году [6]. Эти головоломки стали широко известны после выхода в 1979 году книги Д. Хофштадтера [7].

Неоднократно предпринимались попытки создания автоматических генераторов тестов Бонгарда [8]. Из обзоров [9, 10] видно, что классические тесты Бонгарда заложили основу целого направления тестирования «интеллектуальных систем». Пример «классической» головоломки приводится на рис. 1.

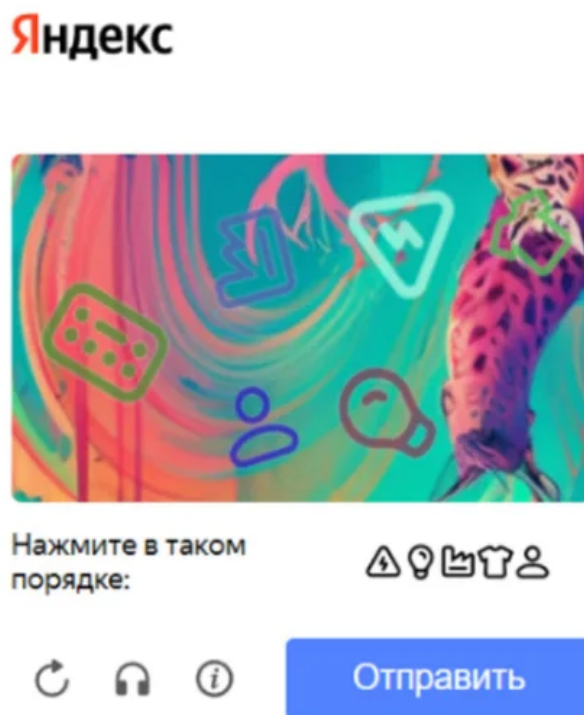


**Рис. 1.** Восемнадцатый из ста тестов Бонгарда, который приведен в приложении А к книге «Проблема узнавания». Для прохождения теста человеку или компьютерной программе предлагают словами описать принцип, по которому отличаются шесть картинок слева от шести картинок справа

#### Капчи

Большинство пользователей Интернета сталкивалось с тем, что в целях безопасности, сервер перед доступом к ресурсу требует от пользователя распознать несколько кривых букв и/или цифр. Капчи – это головоломки, которые трудны (а в идеале – непреодолимы) для компьютера, но легки для человека.

Капча, которая используется компанией Яндекс при проверке «Я не робот», представлена на рис. 2.



**Рис. 2.** Тест на «интеллектуальность» от компании Яндекс, который, как предполагается, пока не может быть пройден современной автоматической системой

Яндекс-капча дает представление о том, что современные методы распознавания пока не способны делать.

#### *Психометрические тесты на коэффициент интеллекта (IQ)*

Отчет о том, каких успехов добиваются современные нейросетевые трансформеры по решению визуальных IQ-тестов (например, по прохождению известнейшего набора тестов – Mensa [11]), можно получить на ресурсе [12].

#### *Тесты Шолле на общий искусственный интеллект (AGI)*

На рис. 3 приведен один из тестов Ф. Шолле на общий искусственный интеллект, который входит в его первый (от 2019 года) корпус тестов – ARC-AGI-1.

Аргументы, почему разработку алгоритма по успешному автоматическому решению набора тестов «The Abstraction and Reasoning Corpus» (ARC) следует считать шагом на пути к общему искусственному интеллекту, можно прочесть в работе Ф. Шолле «On the Measure of Intelligence» [13] или услышать в недавнем докладе «How We Get To AGI» [14].

Итоги соревнований за 2024 год по тестам ARC-AGI-1 представлены в отчете [15]. Наилучший официальный итог – 53,5% решенных тестов.

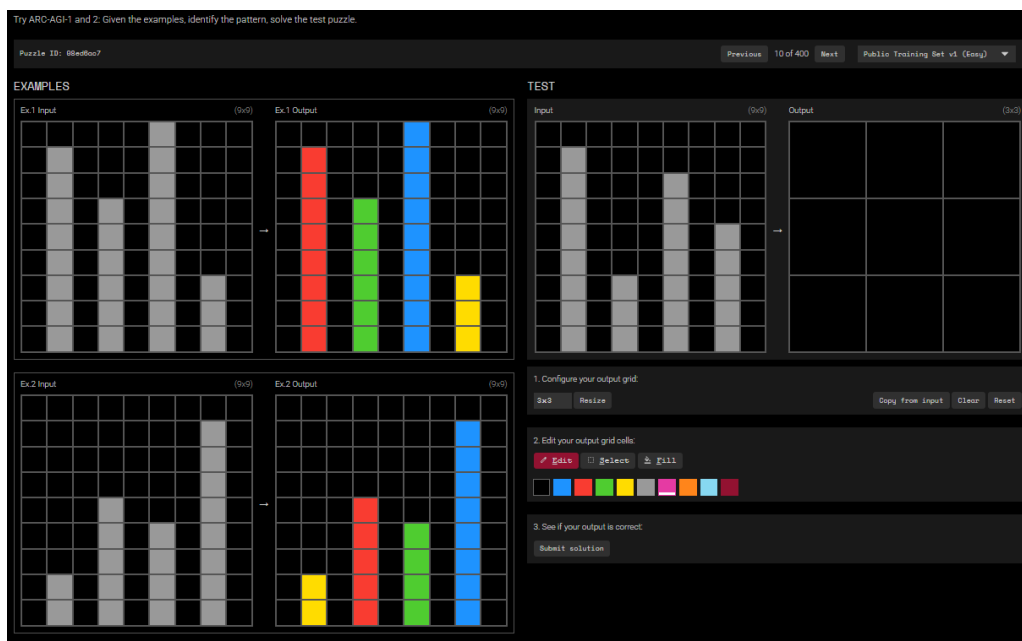
По состоянию на осень 2025 года самые успешные методы по автоматическому решению тестов Шолле применяют архитектуру нейросетевых трансформеров [16].

С 2020 по 2023 годы в данном соревновании побеждали системы «классического компьютерного зрения».

В 2024 году в призерах были только нейросетевые алгоритмы. Успех в конкурсе ARC-AGI-2024 для призеров во многом зависел от применения изошренных (не только сдвиги и повороты) техник «аугментации», иногда на основе предметно-ориентированного языка (Domain-Specific Language, DSL) [17].

#### **Метод**

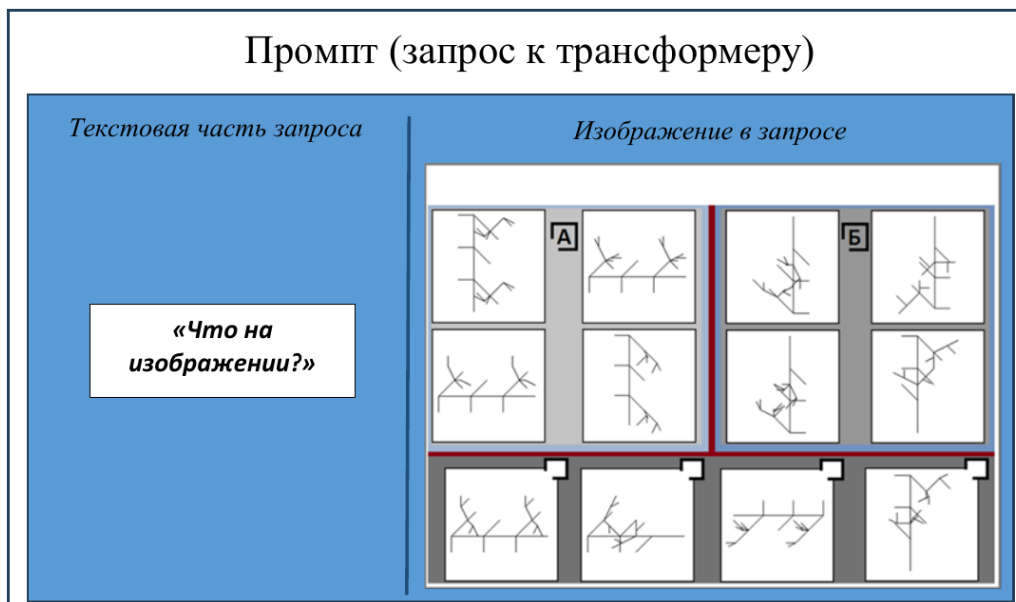
В первом предварительном эксперименте трансформеру демонстрировалось изображение одного из модифицированных тестов Бонгарда (см. рис. 4) в сопровождении краткого вопроса: «Что на



**Рис. 3.** Один из тестов на общий искусственный интеллект, предложенный Ф. Шолле. Данный тест включен в корпус тестов ARC-AGI-1 и использовался в обучающей выборке соревнований ARC-Prize в 2020-2024 годах. Для решения этого теста необходимо установить общий принцип преобразования картинку из 9x9 клеток в другую картинку 9x9

изображений?». Ответы, которые дали сервисы DeepSeek-3.1 Thinking и ChatGPT-5 mini, свидетельствовали о том, что они правильно «понимают» смысл и структуру изображения теста, см. рис. 5.

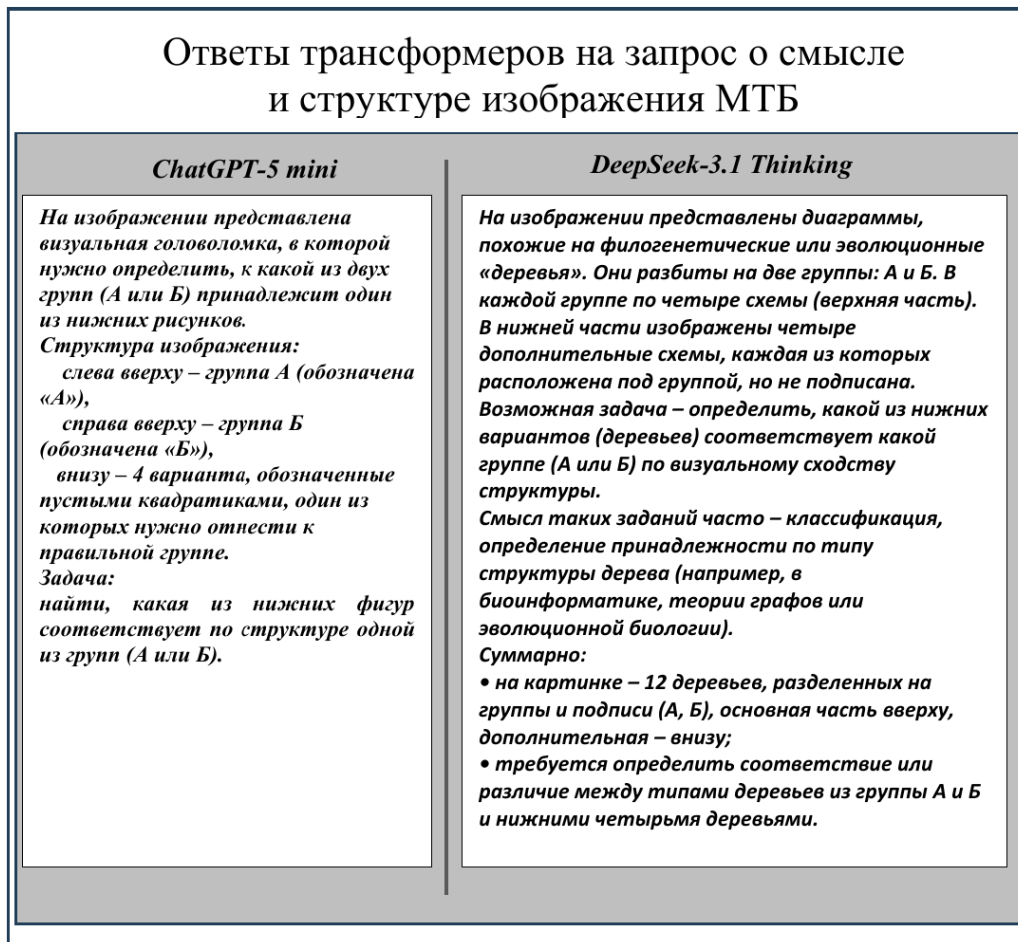
Таким образом, исследуемые сервисы получали на вход ровно то же задание (по форме и содержанию), что и человек.



**Рис. 4.** Изображение одного из автоматически сгенерированных модифицированных тестов Бонгарда и текст промпта, которые были поданы на вход DeepSeek-3.1 Thinking и ChatGPT-5 mini с предварительно удаленным контекстом в чате. Промпт содержал текст «Что на изображении?»

#### Предварительные исследования на реальных фотографиях

В начале опишем гипотетическую производственную задачу в области компьютерного зрения. На рис. 6 представлены обучающая выборка (по 4 примера для каждого из двух классов А и Б) и тест-



**Рис. 5.** Ответы тестируемых мультимодальных нейросетей DeepSeek-3.1 Thinking и ChatGPT-5 mini на вопрос «Что на изображении?». Это изображение представлено на рис. 4

вая выборка (4 изображения в нижнем ряду). Путем демонстрации примеров в зонах А и В делалась попытка обучить нейросетевой сервис отличать дефектные губки для мытья посуды от качественных.

В ходе предварительных экспериментов двум мультимодальным трансформерам ChatGPT-5 mini и DeepSeek-3.1 Thinking было продемонстрировано изображение, как оно представлено на рис. 6. Это изображение сопровождалось запросом (промптом) следующего содержания:

*«На изображении представлена визуальная головоломка на классификацию по аналогии.*

*Описание структуры задачи:*

- в верхней части изображены две группы по четыре рисунка каждая (группа А слева и группа В справа);*
- в нижней части изображения показаны четыре рисунка в ряд в отдельных квадратах;*
- требуется определить, к какой из двух групп (А или В) относится каждый из рисунков внизу, основываясь на закономерностях построения, аналогии формы.*

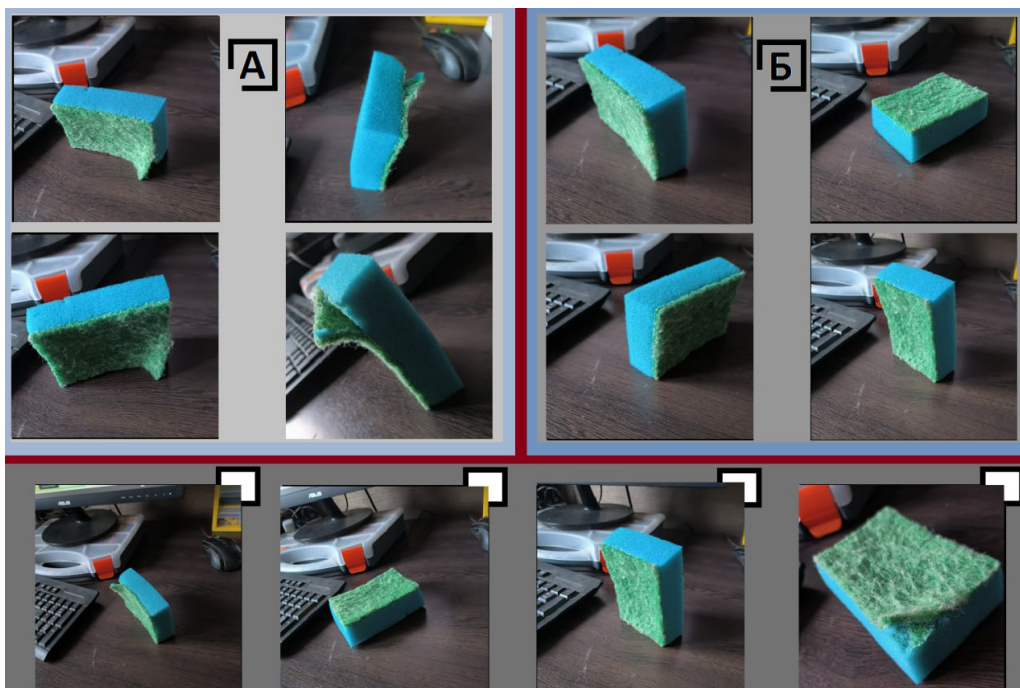
*Цель:*

*найди принцип, по которому разные рисунки относятся к группе А или В, и укажи правильный вариант соответствия для каждого из нижних рисунков».*

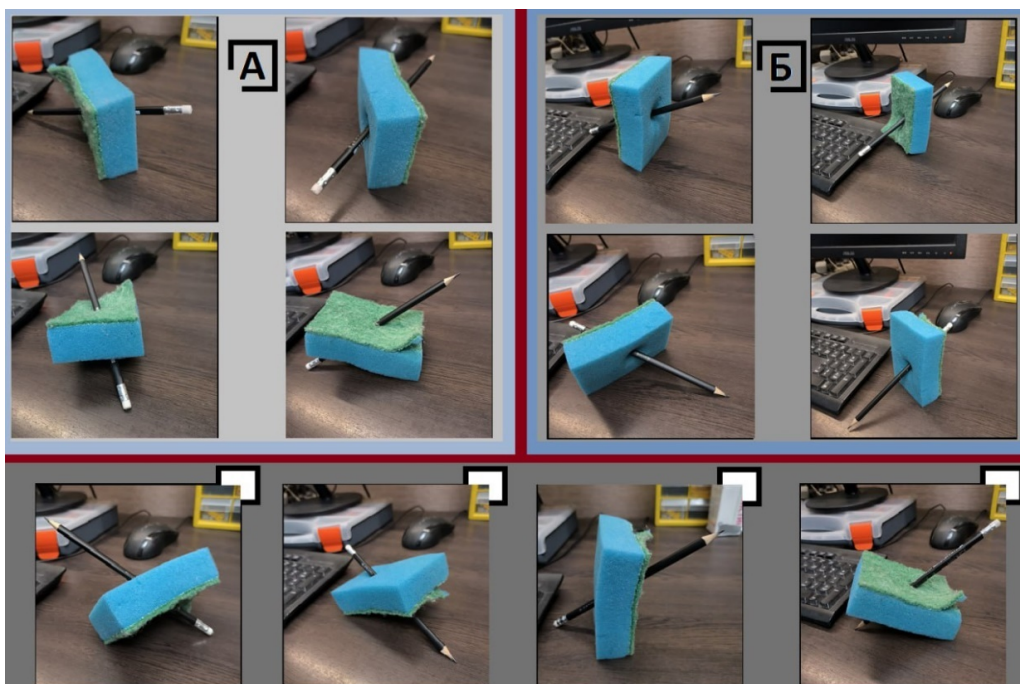
Оба трансформера справились с этой задачей, сформулировали принцип: «Группа А = поврежденные/рваные губки; группа В = целые/ровные губки» и дали правильное краткое решение: «А В В А».

Более сложная задача представлена на рис. 7.

Трансформеры с этой задачей не справились, но принцип отличия картинок зоны А от картинок зоны В DeepSeek-3.1 уловил верно. Вот его ответ: «Группа А: карандаш проходит через зеленую (жесткую) сторону губки; группа В: карандаш проходит через синюю (мягкую) сторону губки».



**Рис. 6.** Задача перед компьютерной системой – проанализировать обучающую выборку изображений в областях А и Б, далее выявить признак, отличающий объекты на этих изображениях, и на основе выявленного признака классифицировать четыре тестовых картинки в нижнем ряду (пометить каждую из них символом А или Б)



**Рис. 7.** В левой верхней области А четыре изображения губок, каждая губка проткнута карандашом, и острие выходит из зеленого слоя губки. В правой верхней области Б на изображениях четыре губки, где острие карандаша выходит из синего слоя каждой губки. Надо выявить это различие и правильно разнести картинки в нижней зоне теста по принципу принадлежности к области А или Б

### Создание «скрытого» набора тестов Бонгарда

В психометрии принято, чтобы испытуемый не был предварительно тренирован на том же типе тестов IQ, на которых он тестируется. Это было мотивом для создания большого набора концептуально

новых тестов Бонгарда, который ранее в Интернет не выкладывался.

### **Обеспечение однозначной интерпретации результата прохождения теста Бонгарда**

Классические тесты Бонгарда требуют дать решение в виде словесного описания принципа отличия левых картинок от правых. В настоящей работе применяются модифицированные тесты Бонгарда (МТБ), в качестве решения они подразумевают однозначно трактуемый ответ в виде четырех букв, например, «А,Б,Б,А».

#### **Предположения**

Предполагается следующий порядок действий мультимодального рассуждающего трансформера:

- 1) нейросетевой трансформер самостоятельно разбивает изображения МТБ на три зоны (А, Б и тестовую), в каждой зоне им выделяются четыре квадратные картинки;
- 2) каждой картинке трансформер сопоставляет текстовое описание;
- 3) рассуждающий трансформер (следуя требованию текстовой части запроса, промпту) анализирует текстовые описания для картинок двух верхних зон А и Б и далее, используя свою функцию «рассуждения», формулирует принцип (признак) отличия картинок в двух зонах;
- 4) к текстовым описаниям каждой из четырех картинок нижней тестовой зоны мультимодальный трансформер «применяет» принцип (признак) отнесения картинки либо к зоне А, либо к зоне Б;
- 5) в ответное сообщение трансформера добавляется краткое решение МТБ, в котором четырем картинкам тестовой зоны, расположенным в ряд, сопоставляются четыре буквы А или Б одной строкой.

#### **Анализ возможностей мультимодальных нейросетевых трансформеров с помощью синтетических тестов**

В отличие от классических тестов Бонгарда, которые были нарисованы вручную, метод синтеза изображений тестов компьютерными программами имеет ряд существенных преимуществ:

каждая картинка теста описывалась коротко и формально на специальном предметно-ориентированном языке (ПОЯ), что давало возможность при строго формальном описании изображения также задавать признак отличия картинок в зонах А и Б и описывать общую аналогию. Символьные описания картинок на ПОЯ преобразовывались в изображения с помощью программного интерпретатора;

формировать, при необходимости, сколько угодно большие обучающие и тестовые выборки; вносить деформации в объекты (повороты, масштабирование), добавлять отвлекающие шумовые детали.

#### **Эксперименты**

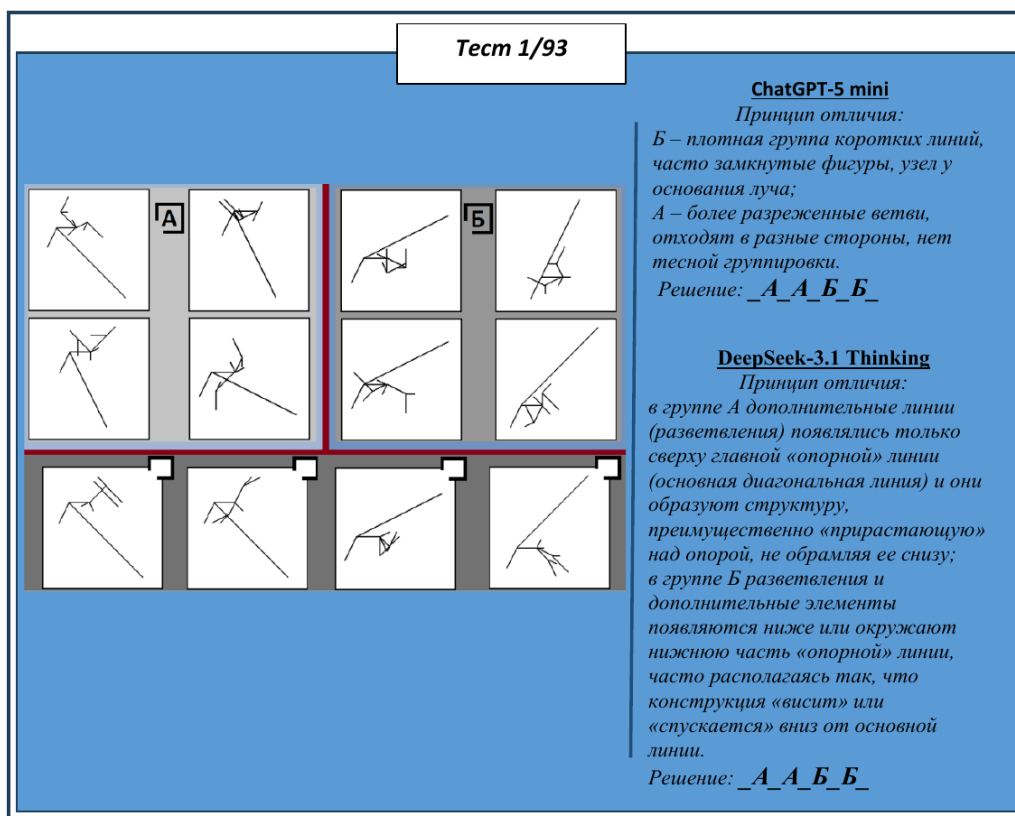
После автоматического синтеза 3000 изображений модифицированных тестов Бонгарда для основных экспериментов были отобраны 93 концептуально разных, которые решаются человеком без долгого разглядывания и размышлений.

Далее каждое из 93 изображений теста МТБ снабжалось текстовым описанием задания (текст задания, или промпт, приведен выше в разделе «Предварительные исследования на реальных фотографиях»). Всякий раз перед таким запросом контекст нейросетевого трансформера стирался.

На рис. 8 приводится первый тест МТБ и на этом же рисунке справа – текстовые ответы двух нейросетей. Ответы содержат краткое решение в виде последовательности символов А А Б Б и текст (см. над кратким решением), данный трансформером в обоснование этого решения. Тест, показанный на рис. 8, оба соревнующихся трансформера решили правильно!

На рис. 9 приведен второй тест МТБ, который решали нейросетевые трансформеры, и оба решили его неправильно.

Все 93 теста, которые участвовали в экспериментах, можно скачать по ссылке [18], в именах файлов всегда присутствует подстрока с правильным решением (вида АВВА). Промпт (текстовый запрос), которым сопровождалось изображение каждого теста, приведен выше в разделе «Предварительные исследования на реальных фотографиях». Таким образом, все эксперименты, выполненные в настоящей работе, можно повторить, а также выполнить сравнительное тестирование других мультимодальных трансформеров.



**Рис. 8.** В левой части расположено изображение теста, в правой – краткие решения от двух трансформеров в виде последовательности символов А А Б Б, а над ними – тексты в обоснование решения. Символ А или Б в ряду А А Б Б обозначает отнесение каждой из четырех картинок в нижнем ряду изображения к зоне А или В

В экспериментах в тексте ответа всегда выдавалось краткое решение в виде упорядоченных четырех символов, например: А А Б Б. Решение теста засчитывалось, если запрос посылался три раза и три раза выдавалось правильное решение.

На рис. 10 представлен самый «простой» тест из набора МТБ, однако оба трансформера провалили этот простейший тест.

По итогам испытаний двух мультимодальных рассуждающих нейросетевых трансформеров со счетом 17:15 победил DeepSeek-3.1 Thinking (см. рис. 11). При этом 12 тестов МТБ из 93 были решены обоими участниками соревнований.

### **Контекстное обучение анализу изображений на малом числе примеров с применением мультимодальных рассуждающих трансформеров (выводы)**

Можно ли на основе современных технологий построить достаточно надежную систему зрительного контроля изделий на конвейере (например, на целостность и дефекты), которая «понимала» бы, что от нее требуется, при предъявлении всего по 10 экземпляров качественных и дефектных изделий? Пожалуй, еще год назад представить себе, что задача компьютерного зрения может быть поставлена подобным образом, казалось фантастикой. И вот в 2025-ом появились мультимодальные рассуждающие трансформеры.

При оптимизации весов трансформера обрабатываются сотни миллионов изображений из Интернета и подписей к ним. Так как в процессе запросов к трансформеру нейросеть не меняет своих весов (при помещении в запрос нескольких обучающих примеров), то такое обучение называется «контекстным». Какие пределы у контекстного обучения?

В классическом компьютерном зрении распознавание строилось на придуманных инженерами признаках [19]. Большое внимание уделялось признакам, надежно выделяемым при преобразованиях плоскости: подобия, аффинным и проективным.

Примерно с 2010-го года сверточные сети стали автоматически формировать признаки и при-

Тест 2/93

**ChatGPT-5 mini**

Решение: Б Б Б А

**DeepSeek-3.1 Thinking**

Анализ групп

Группа А (слева):

1. В каждом рисунке три центральных «луча», от которых отходят еще по два ответвления (итого по три линии с каждой развилки, и на одном конце (или на двух) обязательно есть дополнительные ветви).
2. Центр расположен ближе к середине, рисунок напоминает разветвленную звезду.
3. Все рисунки стоят вертикально, и ни один не располагается горизонтально.

Группа Б (справа):

1. В центре соединяются 4 ветви (крестовина) либо разветвления иные, а сама структура вытянутая, горизонтальная.
2. Большинство рисунков явно вытянуты вдоль горизонтальной оси.
3. Конструкция похожа на дерево или основной хребет с ответвлениями.

Ключевая разница:

- А: центральная точка и три почти равных направления («луча»), симметрия, в основном, вертикальная композиция;
- Б: вытянутая, горизонтальная, большие развилки или четыре направления, структура напоминает основной хребет.

Решение: Б Б Б А

**Рис. 9.** В левой части расположено изображение теста, в правой – краткое решение от двух трансформеров в виде последовательности символов Б Б Б А. Для модели DeepSeek-3.1 Thinking приведен подробный текст рассуждений и принцип решения теста МТБ. Оба трансформера дали одинаковое и неправильное решение

знаковые описания, но достаточно далекие от «человеческих». Качество этих признаков с точки зрения устойчивости, например, к преобразованиям плоскости невысокое.

С внедрением мультимодальных трансформеров нейросети получили возможность применять эффективные «человеческие» признаки.

А с выходом в 2025 году нейросетевых трансформеров с функцией рассуждения стало возможным автоматическое формирование на естественном языке принципа отличия двух классов картинок и обучение на малом числе примеров.

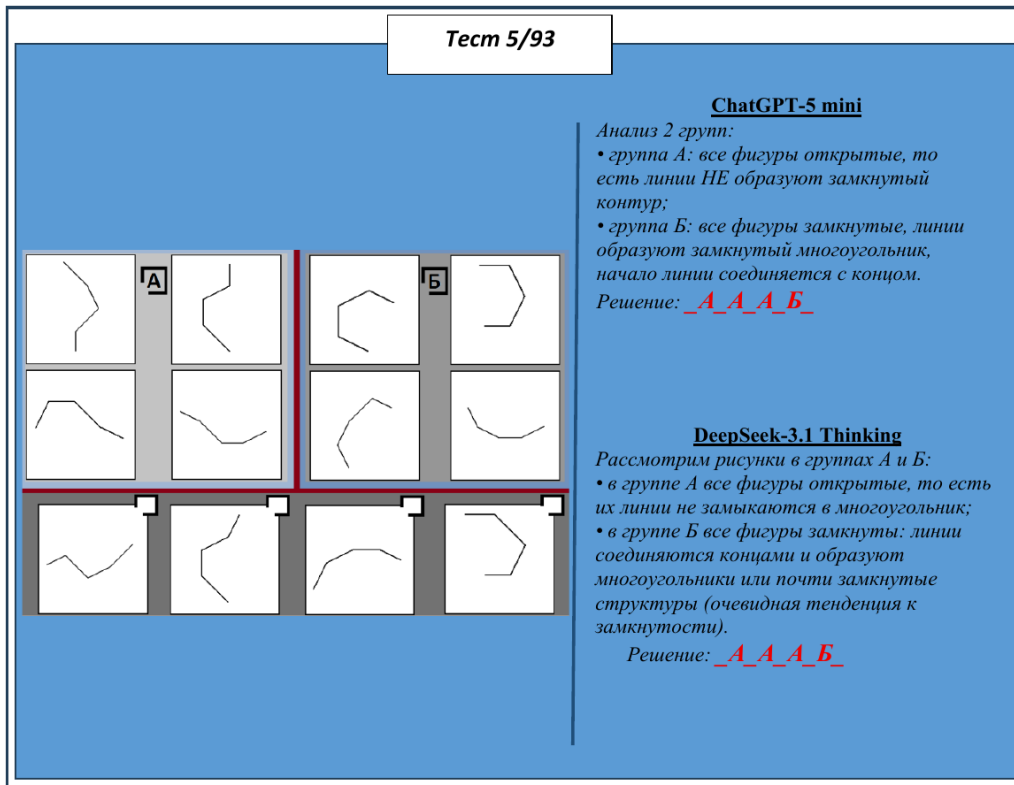
Однако способность вербализации человеком своего собственного восприятия крайне ограничена. На основе естественного языка невозможно обеспечить необходимое богатство признаков описаний. Действительно, более 80% тестов МТБ трансформеры решить не смогли.

Для наращивания возможностей систем контекстного обучения предлагается использовать не только словесные описания людей.

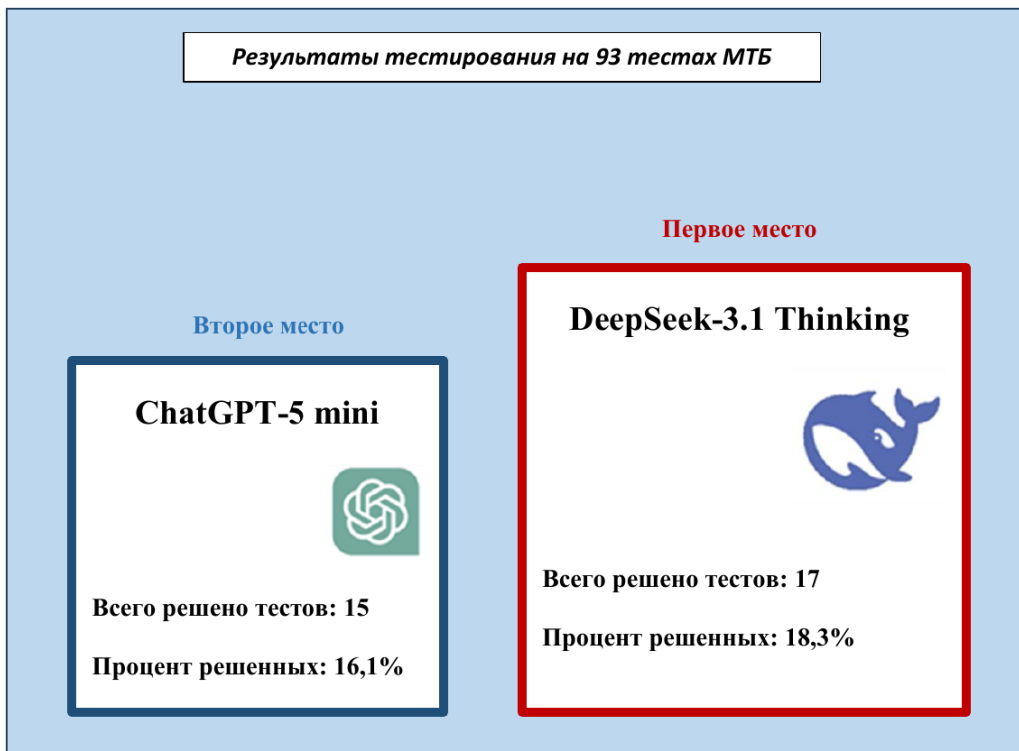
Например, обучение трансформеров можно проводить на изображениях и соответствующих им описаниях (в электронном формате) виртуальных 3D объектов. И это уже делается одним из нейросетевых трансформеров компании Microsoft [20].

Также следует обратить внимание на программные библиотеки по компьютерному зрению (например, на библиотеку OpenCV). Такие библиотеки могут послужить для формирования специальных языков описания изображений. Можно ли таким образом построить «универсальную систему компьютерного зрения» человеческого уровня?

Кант более двухсот лет назад определил интеллект как способность порождать понятия (абстракции). Мультимодальный трансформер при формировании принципа отличия классов пользуется только понятиями, которые используются в языке. И даже в случае, если признаки на изображении будут выделяться программными библиотеками, которые создали программисты, то и в этом случае «интеллект», скорее, будет человеческий. Таким образом, без создания подлинного автоматического



**Рис. 10.** В левой части рисунка расположено изображение теста, в правой части – большими буквами краткие решения, которые выдал каждый трансформер, и тексты описания аналогий и различий групп картинок в зонах А и Б. Оба трансформера не решили данный тест



**Рис. 11.** Из 93 тестов DeepSeek-3.1 Thinking решил 17, ChatGPT-5 mini решил 15

метода порождения понятий, обусловленных физической средой (включая те, которые использует человек, часто даже этого не осознавая), вряд ли удастся сделать серьезный прорыв в искусственном интеллекте.

## ЛИТЕРАТУРА

1. *GPT-5 is here – OpenAI*. Режим доступа: <https://openai.com/gpt-5>.
2. *DeepSeek*. Режим доступа: <https://www.deepseek.com>.
3. *Face Recognition Grand Challenge (FRGC)*. Режим доступа: <https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc>.
4. *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*. Режим доступа: <https://image-net.org/challenges/LSVRC/index.php>.
5. Radford A. et al. Learning Transferable Visual Models from Natural Language Supervision. *International Conference on Machine Learning*. 2021:8748-8763. DOI: <https://doi.org/10.48550/arXiv.2103.00020>.
6. Бонгард М. М. *Проблема узнавания*. М.: Физматгиз; 1967. 320 с.
7. Hofstadter D. R. *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic books; 1999.
8. Nie W. et al. Bongard-LOGO: A New Benchmark for Human-Level Concept Learning and Reasoning. *Advances in Neural Information Processing Systems*. 2020;33:16468-16480. Режим доступа: [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/bf15e9bbff22c7719020f9df4badc20a-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/bf15e9bbff22c7719020f9df4badc20a-Paper.pdf).
9. *Index of Bongard Problems*. Режим доступа: <https://www.foundalis.com/res/bps/bpidx.htm>.
10. Małkiński M., Pawlonka S., Mańdziuk J. Reasoning Limitations of Multimodal Large Language Models. A Case Study of Bongard Problems. 2024. *arXiv:2411.01173*. Режим доступа: <https://arxiv.org/abs/2411.01173>.
11. *IQ Test*. Режим доступа: <https://www.mensa.org/mensa-iq-challenge/#test>.
12. *Tracking AI. Monitoring Artificial Intelligence*. Режим доступа: <https://www.trackingai.org/home>.
13. Chollet F. On the Measure of Intelligence. 2019. *arXiv:1911.01547*. Режим доступа: <https://arxiv.org/pdf/1911.01547>.
14. Chollet F. How We Get To AGI. 2025. Режим доступа: <https://www.youtube.com/watch?v=5QcCeSsNRks>.
15. *ARC Prize 2024: Technical Report*. 2024. Режим доступа: <https://arcprize.org/competitions/2024/>.
16. Akyürek E. et al. The Surprising Effectiveness of Test-Time Training for Few-Shot Learning. 2024. *arXiv:2411.07279*. Режим доступа: <https://arxiv.org/html/2411.07279v2>.
17. *ARC Prize 2024*. Режим доступа: <https://arcprize.org/competitions/2024/>.
18. *База данных 93 изображений тестов МТБ 2025*. Режим доступа: <https://disk.yandex.ru/d/SDvvt4xqDh49ZQ>.
19. Мясников В. В. и др. *Методы обнаружения и распознавания объектов на цифровых изображениях*. Самара: Изд-во СГАУ; 2006. 168 с. Режим доступа: <https://repo.ssau.ru/handle/Uchebnye-posobiya/Metody-obnaruzheniya-i-raspoznavaniya-obektov-na-cifrovyyh-izobrazheniyah-Elektronnyi-resurs-ucheb-posobie-54225>.
20. *Copilot 3D Transforms an Image into a Usable 3D Model*. Режим доступа: <https://copilot.microsoft.com/labs/experiments/copilot-3d>.