

## ДИАГНОСТИКА УРОВНЯ КРЕАТИВНОСТИ ПО РИСУНОЧНОМУ ТЕСТУ ТОРРАНСА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

© Панфилова А.С.

Кандидат технических наук, заведующая лабораторией технологий искусственного интеллекта  
в психологии, Институт психологии РАН, Москва, Россия,  
panfilova87@gmail.com; ORCID: 0000-0003-1892-5901

© Валуева Е.А.

Кандидат психологических наук, научный сотрудник лаборатории психологии  
и психофизиологии творчества, Институт психологии РАН, Москва, Россия,  
ekval@mail.ru; ORCID: 0000-0003-3637-287X

© Дмитриева Ю.А.

Кандидат психологических наук, доцент кафедры психологии управления и служебной  
деятельности, Южно-Уральский государственный университет (НИУ), Челябинск, Россия,  
dmitrieva.julia.86@mail.ru; ORCID: 0000-0002-0331-4684

© Грязева-Добшинская В.Г.

Доктор психологических наук, профессор, заведующий лабораторией психологии  
и психофизиологии стрессоустойчивости и креативности, Южно-Уральский государственный  
университет (НИУ), Челябинск, Россия,  
vdobshinya@mail.ru; ORCID: 0000-0002-9986-4073

© Коробова С.Ю.

Кандидат психологических наук, научный сотрудник лаборатории психологии  
и психофизиологии стрессоустойчивости и креативности, Южно-Уральский государственный  
университет (НИУ), Челябинск, Россия,  
k.svetlana-1991@mail.ru; ORCID: 0000-0002-8633-7231

© Колтунов Е.И.

Научный сотрудник лаборатории психологии и психофизиологии стрессоустойчивости  
и креативности, Южно-Уральский государственный университет (НИУ), Челябинск, Россия,  
aspiratingle@gmail.com; ORCID: 0000-0003-4836-7859

Исследование выполнено за счет гранта Российского научного фонда № 23–18–01059  
«Универсальные факторы массовой креативности: диагностика на основе методов  
машинного обучения», <https://rscf.ru/project/23-18-01059/>

### Для цитирования:

Панфилова А.С., Валуева Е.А.,  
Дмитриева Ю.А., Грязева-Добшинская В.Г.,  
Коробова С.Ю., Колтунов Е.И.  
Диагностика уровня креативности по рису-  
ночному тесту Торранса с использованием  
методов машинного обучения // Ученые  
записки Института психологии Российской  
академии наук. 2025. Т.5. №4(17). С.76-85.  
DOI: 10.38098/proceedings\_2025\_05\_04\_09  
Panfilova A.S., Valueva E.A.,  
Dmitrieva Yu.A., Gryazeva-Dobshinskaya V.G.,  
Korobova S.Yu., Koltunov E.I.  
Machine learning methods for assessing cre-  
ativity levels using the Torrance drawing test.  
Proceedings of the Institute of Psychology of  
the Russian Academy of Sciences. 2025, Vol.5,  
No4(17), Pp.76-85.  
DOI: 10.38098/proceedings\_2025\_05\_04\_09

В статье представлен подход по созданию моделей для автоматизации  
процедуры оценки креативности по тесту Торранса. Проведенное ис-  
следование продемонстрировало наилучшие результаты по дообуче-  
нию модели на базе Swin-base с использованием изображений субте-  
ста 2. Достигнутая точность составила 0,74–0,88 для оригинальности  
и 0,69–0,82 для разработанности по отдельным типам исходных рисун-  
ков. Обучение модели предсказанию уровня оригинальности по назва-  
нию изображений субтеста 2 посредством дообучения модели на базе  
Bert продемонстрировало точность 0,79.

**Ключевые слова:** креативность, психологическая диагностика, диагно-  
стика креативности, компьютерное зрение, машинное обучение, анализ  
естественного языка, тест Торренса (ТТСТ), уровень креативности, ори-  
гинальность, разработанность

## Введение

Автоматизация оценки творческого потенциала по рисункам из фигурной части теста Торренса (ТТСТ-F) — относительно новая, активно развивающаяся область. В ней используются как традиционные методы машинного обучения с извлечением признаков, так и глубокие модели, включая сверточные нейронные сети и трансформеры. Кроме того, применяются подходы объяснимого искусственного интеллекта для анализа решений моделей.

Одним из подходов является использование классических алгоритмов машинного обучения с признаками, извлеченными из изображений. Например, в работе С. Акар и др. (S. Acar et al.) мультимодальная модель CLIP (вариант ViT-B/32) применялась для получения 512-мерных эмбедингов изображений и текстовых описаний рисунков ТТСТ-F (Acar et al., 2025). Эти эмбединги использовались для обучения ансамблей классификаторов (Random Forest, AdaBoost и XGBoost). Random Forest обеспечил точность классификации 78–85% (в зависимости от комбинации изображения и текста) и корреляцию с экспертными оценками 0,54–0,65. На датасете MTCI (Multi-Trial Creative Ideation, более 11 000 рисунков) все три ансамбля (AdaBoost, Random Forest и XGBoost), обученные на CLIP-эмбедингах, показали корреляции с экспертной оценкой в диапазоне  $r=0,77$ – $0,80$  (AdaBoost  $r=0,80$ , Random Forest  $r=0,79$ , XGBoost  $r=0,77$ ). Эти результаты демонстрируют, что традиционные методы машинного обучения, усиленные эмбедингами, достигают приемлемой точности, близкой к простым эвристическим правилам.

Развитие сверточных нейронных сетей (CNN) расширило возможности автоматизации. Д.Х. Кропли и Р.Л. Марроне (D.H. Cropley, R.L. Marrone) обучили CNN на тесте К. Урбана — Test of Creative Thinking — Drawing Production (TCT-DP) и получили точность около 90% на небольшом наборе данных (414 изображения) (Cropley, Marrone, 2025). В последующем расширенном исследовании авторов (Cropley et al., 2024) на корпусе из 1594 изображений модель достигла точности 80,4% и корреляции 0,81 с экспертными оценками.

Дж.Д. Паттерсон и др. (J.D. Patterson et al.) разработали платформу AuDrA на основе модифицированной ResNet с регрессионным выходом. Модель обучалась на 13 000 рисунках из задачи MTCI и показала корреляцию с экспертными оценками на новых данных от 0,64 до 0,93, устойчиво превосходя простой индикатор «количества чернил» на странице, что указывает на способность CNN захватывать сложные признаки, такие как композиция и оригинальные элементы. Авторы

разместили данные, модель и скрипты в открытом доступе на платформе OSF (Patterson et al., 2024).

Недавно в автоматизированную оценку креативности начали внедрять архитектуры Vision Transformer (ViT). С. Акар и др. дообучили модель BEiT на данных тестов ТТСТ-F и MTCI, достигнув точности 83% для BEiT-large на объединенном корпусе ТТСТ-F (обе формы теста) с корреляцией 0,64. На наборе MTCI BEiT-large показал корреляцию 0,85, что превосходит результаты ViT (Acar et al., 2025) и ResNet (Patterson et al., 2024).

Мультимодальные методы, сочетающие изображение и описывающий его текст, демонстрируют улучшенные результаты. В работе С. Акар и др. объединение изображения и текста через CLIP улучшало результаты по сравнению с только визуальным входом (Acar et al., 2025). При использовании признаков CLIP Random Forest классификаторы, обученные только на рисунках, достигали объединенной корреляции около  $r=0,56$ , тогда как текстовые названия рисунков обеспечивали  $r=0,60$ , а комбинация изображения и текста —  $r=0,62$ . При этом в одной для формы В текстовая и мультимодальная модели показывали одинаково высокие корреляции ( $r=0,65$ ). Аналогично Х. Чжан и др. (H. Zhang et al.) предложили модель для графического теста дивергентного мышления, где респонденты дополняют пятиугольник рисунками с описаниями (Zhang et al., 2025). Датасет включал 708 рисунков китайских старшеклассников. Модель объединяла ResNet50 (предобученную на ImageNet) для визуальных признаков и GloVe для текстовых эмбедингов. На тренировочном наборе (603 изображения) корреляция с экспертными оценками достигла 0,81 для мультимодальной версии, но при этом для текстовой модели корреляция составила 0,91, а для визуальной — 0,7. На валидационном наборе (100 изображений) показатели снизились, но их соотношение сохранилось — 0,561; 0,629 и 0,395 соответственно. При этом на уровне участника (усреднение 9 рисунков от каждого участника) преимущество сместилось в сторону текстовой модели —  $r$  составило 0,724; 0,602, и 0,135 соответственно.

Для валидации моделей стали применяться методы объяснимого ИИ. А.С. Панфилова и др. сравнивали разные CNN (MobileNet, ResNet18, AlexNet, DenseNet, EfficientNet, ViT) на рисунках теста Урбана (Panfilova et al., 2024). Лучшая модель, MobileNet, обеспечила точность 76% в предсказании баллов с высоким согласием с экспертами. Grad-CAM показал, что модель фокусируется на релевантных зонах аналогично экспертам. Подобные исследования показывают, что современные сети могут «учиться» признакам,

осмысленным для человека (например, оригинальным элементом рисунка), а методы объяснимого ИИ помогают проверить, что модель способна неявным образом усвоить экспертные критерии (Panfilova et al., 2024; Panfilova, Valueva, 2025).

Таким образом, современные методы машинного и глубокого обучения позволяют с приемлемой точностью оценивать креативность респондентов по рисуночным тестам. Ключевые модели включают сверточные сети (ResNet, MobileNet и др.), трансформеры (BEiT/ViT) и мультимодальные архитектуры (CLIP, ResNet + GloVe). Они извлекают характерные визуальные признаки (композиция, сложность рисунка и др.) и семантические признаки, приближаясь к экспертным оценкам. Результаты последних работ свидетельствуют о том, что автоматическая оценка образной креативности демонстрирует высокую согласованность с экспертными оценками и указывает на потенциал для масштабных исследований.

### Метод

Для диагностики креативности использовалась фигурная форма теста творческого мышления Е.П. Торранса. Фигурная форма теста Е.П. Торранса включает три субтеста («Создание рисунка», «Незавершенные фигуры» и «Повторяющиеся линии»), каждый из которых респонденту необходимо выполнить в течение 10 минут. По результатам прохождения теста рассчитываются такие показате

ли креативности, как *оригинальность*, *беглость*, *гибкость*, *разработанность* и *абстрактность названий* (Torrance, 2008). Показатель оригинальности характеризует способность респондента генерировать редкие, уникальные идеи и основывается на статистической редкости ответа; показатель разработанности характеризует способность респондента проявлять творческие способности через детальную проработку возникающих идей.

По субтесту 2 «Незавершенные фигуры» рассчитывались все перечисленные выше показатели креативности. Показатель *оригинальности* рассчитывался по шкале от 0 до 2 согласно частоте встречаемости одинаковых ответов: 2 балла для ответов с частотой менее 2%, 1 балл засчитывался за ответы, встречающиеся в 2–4,99% случаев, 0 баллов — для ответов с частотой 5% и более. При оценке показателя *разработанности* баллы давались за каждую значимую деталь (идею), дополняющую исходную стимульную фигуру, как в границах ее контура, так и за ее пределами. Один балл присваивался за каждую существенную деталь, цвет, специальную штриховку, каждую подробность в названии сверх необходимого минимума (Маркина, 2004).

В исследовании приняли участие 1707 респондентов трех возрастных категорий: дошкольники и учащиеся младшей школы, учащиеся средней и старшей школы, студенты и взрослые. Распределение участников по группам представлено в таблице 1.

**Таблица 1.** Возрастной состав выборки респондентов

Возрастная категория	Количество респондентов
Дошкольники и младшая школа	570
Средняя и старшая школа	510
Студенты и взрослые	627

Субтест № 2 теста Торранса включает в себя 10 отдельных изображений, которые респонденту предлагается дополнить, то есть по каждому респонденту формируется набор из 10 отдельных рисунков, которые были оценены по критериям оригинальности и разработанности. Рисунки, которые респондент не использовал, исключались из выборки.

Экспертная оценка по критерию оригинальности представлена в бинарном виде — рисунок получал оценку 0, если изначально оригинальность равнялась нулю, и оценку 1, если оригинальность была 1 и выше. В категории с низким уровнем оригинальности оказался 4491 рисунок, а с высоким — 8630 изображений. Экспертная оценка

по критерию разработанности также была переведена в бинарную систему. Для баланса выборки всем рисункам с оценкой разработанности два и ниже, был присвоен балл 0, остальным — балл 1. Таким образом, категория с низкой разработанностью включает в себя 7496 рисунков, а с высоким уровнем разработанности — 5353.

Помимо экспертных оценок, для каждого рисунка был доступен заголовок, который испытуемый давал своему изображению. Всего было доступно 12691 (некоторые рисунки не имели заголовков). Анализ заголовков в качестве предиктора оригинальности был произведен с помощью отдельной модели.

Решение задачи создания модели, определяющей уровень оригинальности и разработанности

по рисуночному тесту, предполагает дообучение какой-либо исходной модели на небольшом наборе данных. Исходная модель при этом уже обучена на задаче выделения объектов на общедоступных выборках большого размера (ImageNet). Модели, реализующие классификацию изображений, представлены различной архитектурой и числом параметров. Для поиска оптимального варианта исходной предобученной модели, которая способна обобщить информацию о рисунках в рамках теста Торранса, была проведена серия экспериментов по обучению модели на всем массиве изображений без учета номера изображения в рамках субтеста. Выборка была разделена на обучающую, тестовую и валидацион-

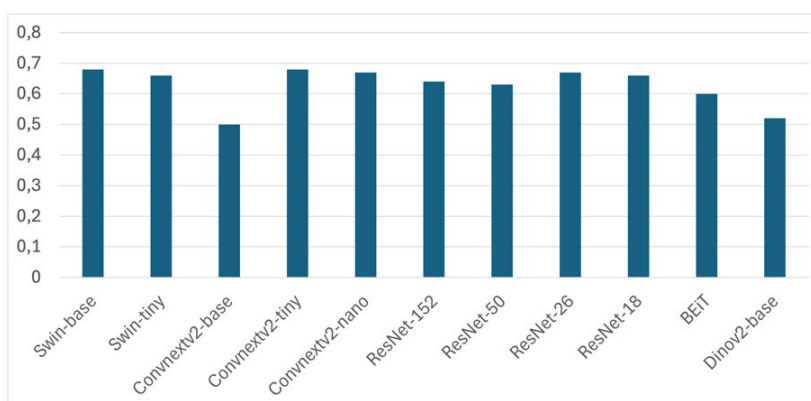
ную в отношении 70%, 15%, 15%, соответственно, и стратифицирована по оценке оригинальности. Отбор базовых моделей для проведения эксперимента опирается на общедоступные рейтинги их эффективности на выборках ImageNet (<https://paperswithcode.com>), содержащих 1000 или 22000 определяемых классов с учетом различного числа параметров модели. Таким образом был сформирован следующий набор базовых моделей, представленный в таблице 2. Предыдущие эксперименты продемонстрировали, что большее число параметров модели способствует ее переобучению, что означает снижение обобщающей способности модели (Panfilova et al., 2024).

**Таблица 2.** Базовые модели и их параметры

Модель	Количество параметров	Название модели в сервисе HF <sup>1</sup>
Swin-base (Liu et al., 2021)	109 млн	microsoft/swin-base-patch4-window7-224-in22k
Swin-tiny (Liu et al., 2022)	28,3 млн	microsoft/swinv2-tiny-patch4-window16-256
Convnextv2-base (Woo et al., 2023)	88,7 млн	facebook/convnextv2-base-22k-224
Convnextv2-tiny (Woo et al., 2023)	28,6 млн	facebook/convnextv2-tiny-22k-224
Convnextv2-nano (Woo et al., 2023)	15,6 млн	facebook/convnextv2-nano-22k-224
ResNet-152 (He et al., 2016)	60,3 млн	microsoft/resnet-152
ResNet-50 (He et al., 2016)	25,6 млн	microsoft/resnet-50
ResNet-26 (He et al., 2016)	16 млн	microsoft/resnet-26
ResNet-18 (He et al., 2016)	11,7 млн	microsoft/resnet-18
BEiT (Bao et al., 2021)	87 млн	microsoft/beit-base-patch16-224
Dinov2-base (Oquab et al., 2023)	86,6 млн	facebook/dinov2-base

Обучение проводилось с использованием батча размером 16, скорости обучения  $3e-5$  на 55 эпо-

хах. На рис. 1 представлена диаграмма точности классификации моделей на тестовой выборке.



**Рис. 1.** Точность классификации моделей на тестовой выборке по всем изображениям датасета

<sup>1</sup> Сервис общедоступных моделей и данных <https://huggingface.co>

Выявлено, что наибольшая точность наблюдается для модели Swin-base, однако, данная точность не считается приемлемой для получения требуемых оценок креативности. С целью повышения точности работы моделей далее проводилось дообучение модели Swin-base отдельно на каждую категорию изображений, то есть всего обучено 10 моделей для оценки уровня оригинальности и 10 моделей для оценки уровня разработанности. Наборы рисунков, соответствующих каждому из десяти номеров, были разделены на обучающую, тестовую и валидационную выборки в соотношении 70%, 15%, 15%. Параметры обучения схожи с параметрами обучением базовой модели.

Для предсказания оригинальности по названиям рисунков была дообучена модель Bert (Devlin et al., 2019). Модель была обучена на общей выборке из 12691 изображения без разделения по типу базового рисунка. Объединяя все изображения, мы опирались на допущение, что оригинальность в данном случае отражена прежде всего в семантическом содержании названия, а не в специфике конкретного образного шаблона. Поскольку модель использует только текст, тип базового изо-

бражения не выступает для нее отдельной модальностью, при этом разбиение данных по типам стимулов уменьшает объем обучающей выборки для каждой подгруппы. Объединение всех типов в одну модель повышает статистическую устойчивость оценки параметров и улучшает способность к обобщению. Выборка была разделена на обучающую, тестовую и валидационную в соотношении 70%, 15%, 15%. Обучение проводилось с использованием батча размером 32, скорости обучения  $2e-5$  на 10 эпохах.

### Результаты обучения моделей предсказания уровня оригинальности и разработанности по рисункам

Результаты обучения моделей бинарной классификации выборок изображений каждого типа по отдельности по критериям оригинальности и разработанности представлены в таблице 3. В качестве критерия оценки точности работы моделей использовался взвешенный F1 (weighted F1), учитывающий точность и полноту выделения высокого и низкого класса, а также представленность классов в выборке.

**Таблица 3.** Оценка точности моделей по определению уровня оригинальности и разработанности на тестовой выборке по критерию weighted F1

Номер рисунка	Оригинальность	Разработанность
1	0,83	0,82
2	0,84	0,79
3	0,82	0,72
4	0,74	0,73
5	0,84	0,79
6	0,87	0,75
7	0,88	0,78
8	0,83	0,78
9	0,79	0,77
10	0,82	0,69

Можно отметить, что наименьшая точность по оценке уровня оригинальности получена для изображения номер 4. Данный типа рисунка в ка-

честве исходного содержит фигуру, представляющую собой волнистую линию, пример представлен на рис. 2.



**Рис. 2.** Примеры рисунков респондентов для изображения № 4

Справедливо предположить, что в процессе обучения модель опосредованно усваивает базовые образы, связанные с тем или иным уровнем оригинальности (так как оценка оригинальности опирается на частоту базового образа). Однако, часть образов может иметь слишком широкую вариативность в своем представлении, что затрудняет усвоение их моделью.

Можно предположить, что в рисунках № 4 присутствуют в достаточной степени определенные базовые образы, которые слабо идентифицируются моделью. Оценка базовых образов, которые были использованы респондентами, на предмет точности классификации уровня оригинальности продемонстрировали, что образы ветра, волны и узора в наибольшей степени некорректно классифицируются в тот или иной уровень моделью.

Определение базовых образов, которые связываются с наибольшим количеством ошибок в целом по выборке для модели, которая была обучена на изображениях всех типов, продемонстрировал, что образы ветра, змеи, смайла, узора с наименьшей точностью классифицируются моделью в тот или иной уровень оригинальности.

Точность моделей оценки уровня разработанности по различным типам рисунков демонстрируют диапазон от 0,69 до 0,82.

Результаты обучения моделей предсказания уровня оригинальности по названию рисунков

Мультиязычная модель на основе Bert, обученная предсказывать уровень оригинальности на основании названия изображений всех типов, продемонстрировала результаты, представленные в таблице 4.

**Таблица 4.** Оценка точности модели по определению уровня оригинальности по названию изображений на тестовой выборке

Категория	Точность	Полнота	F1
0	0,64	0,71	0,67
1	0,87	0,83	0,85
Точность			0,79
Взвешенный F1	0,80	0,79	0,79

Можно отметить, что точность полученной модели ( $F1=0,79$ ) выше, чем модели, обученной на всех изображениях ( $F1=0,69$ ), но в большинстве случаев ниже, чем для моделей каждого базового рисунка ( $F1=0,74-0,87$ ). Данный результат может быть связан с тем, что оценка оригинальности изображений строится на основе частоты базового образа, что в большей степени отражается в названии изображения, а также с тем фактом, что названия изображений по выборке различных типов схожи, однако исходная графическая форма различается.

### Заключение

Проведенное исследование продемонстрировало, что современные архитектуры глубокого обучения могут с высокой точностью автоматизировать оценку креативности по рисуночному тесту Торренса (ТТСТ-F) на уровне отдельных стимулов и текстовых названий. Использование предобученных моделей, дообученных на специализированных данных, позволяет достичь высокой степени согласованности с экспертными оценками по критериям оригинальности и разработанности. Наиболее стабильные результаты

показала модель Swin-base, для которой при индивидуальном обучении на каждом типе изображения получены значения взвешенного F1 в диапазоне 0,74–0,88 по оригинальности и 0,69–0,82 по разработанности.

Анализ показал, что точность классификации существенно зависит от особенностей исходного изображения субтеста. В частности, рисунки с высокой вариативностью интерпретаций (например, изображение № 4 с волнистой линией) оказались наиболее сложными для автоматической классификации. Это может быть связано с тем, что модели сложнее усваивать слишком разнообразные базовые образы, не имеющие устойчивых визуальных паттернов.

Важно отметить, что значительный прогресс в автоматизации оценки креативности уже достигнут в области вербальных дивергентных заданий. Для задач вроде теста Необычное использование (Alternate Uses Test) и ассоциативных тестов показано, что модели, опирающиеся только на тексты ответов, достигают высокой согласованности с экспертами, сопоставимой с традиционными процедурами ручного оценивания (Beatty, Johnson, 2021; Organisciak, 2023). Это указывает на то, что значительная часть дисперсии оценок креативности может быть объяснена исключительно семантическими признаками. Наши результаты по предсказанию оригинальности по названиям рисунков фактически продолжают эту линию исследований, показывая, что даже в контексте фигурной части теста Торренса текстовые описания несут достаточно информации для надежного приближения экспертного балла (Acar et al., 2025; Zhang et al., 2025).

Так, в нашей работе предсказание уровня оригинальности по названиям изображений демонстрирует  $F1=0,79$ , сопоставимое с диапазоном 0,74–0,88 для отдельных моделей по изображениям и превышающее показатель общей визуальной модели по всем рисункам ( $F1=0,69$ ). Сходная (и несколько контринтуитивная) картина наблюдается и в ряде других исследований: при оценивании креативности по рисункам с подписями текстовые модели часто не уступают мультимодальным, а иногда и превосходят их. Так, в работе С. Акар комбинация изображения и текста лишь незначи-

тельно улучшает корреляцию относительно одного текста (Acar et al., 2025), а в сходном графическом тесте дивергентного мышления текстовая модель демонстрирует наилучшее обобщение как на уровне отдельных рисунков, так и на уровне суммарных баллов участника (Zhang et al., 2025). Как отмечают Х. Чжан и др. текстовая модель получает явный доступ к тем же семантическим признакам, на которые опираются эксперты при ручном оценивании. Визуальные и мультимодальные модели, напротив, вынуждены восстанавливать эту семантику по рисунку, что делает задачу сложнее и может снижать обобщающую способность. Таким образом, наши результаты, наряду с данными С. Акар и ХюЧжан (Acar et al., 2025; Zhang et al., 2025), показывают, что в задачах оценивания креативности по рисункам с подписями одного текста во многих случаях оказывается достаточно для приближенного воспроизведения экспертных оценок.

Наши результаты подтверждают, что нейросетевые методы способны не только воспроизводить экспертные оценки, но и выявлять закономерности, отражающие визуальные и семантические признаки, значимые для интерпретации творческих проявлений. Автоматизация оценки рисунков ТТСТ открывает перспективы для более масштабных психометрических исследований, снижения трудоёмкости экспертного анализа и стандартизации процедуры оценки креативности.

В дальнейшем развитие данного направления может включать:

- использование мультимодальных моделей, объединяющих визуальные и текстовые признаки;
- применение методов объяснимого ИИ для интерпретации решений моделей;
- расширение и балансировку обучающих выборок с целью повышения обобщающей способности моделей;
- исследование переносимости моделей между различными вариантами теста и культурными контекстами.

Таким образом, результаты работы демонстрируют, что глубокие сверточные и трансформерные архитектуры являются эффективным инструментом для автоматизированной оценки показателей креативности по данным теста Торренса, обеспечивая высокий уровень точности.

#### Литература:

- Маркина Н.В., Матвеева Л.Г. Миннесотские тесты творческого мышления (МТТМ графическая форма). Челябинск: ПсихРОН, 2004.
- Acar S., Organisciak P., Dumas D. Automated scoring of figural tests of creativity with computer vision // The Journal of Creative Behavior. 2025. V. 59. № 1. Art. e677.

- Bao H., Dong L., Piao S., Wei F. Beit: Bert pre-training of image transformers // arXiv preprint arXiv:2106.08254. 2021.
- Beaty R.E., Johnson D.R. Automating creativity assessment with SemDis: An open platform for computing semantic distance // Behavior Research Methods. 2021. V. 53. № 2. P. 757–780. DOI: <https://doi.org/10.3758/s13428-020-01453-w>.
- Cropley D.H., Marrone R.L. Automated Scoring of Figural Creativity using a Convolutional Neural Network // Psychology of Aesthetics Creativity and the Arts. 2025. V. 19. № 1. P. 77–86. DOI: 10.1037/aca0000510.
- Cropley D.H., Theurer C., Mathijssen A.S., Marrone R.L. Fit-For-Purpose Creativity Assessment: Automatic Scoring of the Test of Creative Thinking — Drawing Production (TCT-DP) // Creativity Research Journal. 2024. P. 1–16. DOI: <https://doi.org/10.1080/10400419.2024.2339667>.
- Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies. 2019. V. 1. P. 4171–4186.
- He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770–778.
- Liu Z., Hu H., Lin Y., Yao Z., Xie Z., Wei Y., Ning J., Cao Y., Zhang Z., Dong L., Wei F., Guo B. Swin transformer v2: Scaling up capacity and resolution // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. P. 12009–12019.
- Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. Swin transformer: Hierarchical vision transformer using shifted windows // Proceedings of the IEEE/CVF international conference on computer vision. 2021. P. 10012–10022.
- Oquab M., Darcet T., Moutakanni T., Vo H., Szafraniec M., Khalidov V., Fernandez P., Haziza D., Massa F., El-Nouby A., Assran M., Ballas N., Galuba W., Howes R., Huang P., Li S., Misra I., Rabbat M., Sharma V., Synnaeve G., Xu H., Jegou H., Mairall J., Labatut P., Joulin A., Bojanowski P. Dinov2: Learning robust visual features without supervision // arXiv preprint arXiv:2304.07193. 2023.
- Organisciak P., Acar S., Dumas D., Berthiaume K. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models // Thinking Skills and Creativity. 2023. V. 49. Art. 101356. DOI: 10.1016/j.tsc.2023.101356.
- Panfilova A.S., Valueva E.A., Ilyin I.Y. The application of explainable artificial intelligence methods to models for automatic creativity assessment // Frontiers in Artificial Intelligence. 2024. V. 7. Art. 1310518. DOI: <https://doi.org/10.3389/frai.2024.1310518>.
- Panfilova A.S., Valueva E.A. Deep Learning and Explainable AI for Creativity Scoring in TCT-DP Form B // Psychology. Journal of the Higher School of Economics. 2025. V. 22. № 4. P. 721–735. DOI: 10.17323/1813-8918-2025-4-721-735.
- Patterson J.D., Barbot B., Lloyd-Cox J., Beaty R.E. AuDrA: An automated drawing assessment platform for evaluating creativity // Behavior Research Methods. 2024. V. 56. № 4. P. 3619–3636. DOI: <https://doi.org/10.3758/s13428-023-02258-3>.
- Torrance E.P. The Torrance Tests of Creative Thinking Norms-Technical Manual Figural (Streamlined) Forms A & B. Bensenville, IL: Scholastic Testing Service, 2008.
- Woo S., Debnath S., Hu R., Chen X., Liu Z., Kweon I.S., Xie S. Convnext v2: Co-designing and scaling convnets with masked autoencoders // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. P. 16133–16142.
- Zhang H., Dong H., Wang Y., Zhang X., Yu F., Ren B., Xu J. Automated Graphic Divergent Thinking Assessment: A Multimodal Machine Learning Approach // Journal of Intelligence. 2025. V. 13. № 4. P. 45. DOI: 10.3390/jintelligence13040045.

## MACHINE LEARNING METHODS FOR ASSESSING CREATIVITY LEVELS USING THE TORRANCE DRAWING TEST

© Anastasia S. Panfilova

PhD in Computer science, Head of The Laboratory of AI Technologies in Psychology,  
Institute of Psychology Russian Academy of Sciences, Moscow, Russia  
[panfilova87@gmail.com](mailto:panfilova87@gmail.com); ORCID: 0000-0003-1892-5901

© **Ekaterina A. Valueva**

PhD in Psychology, Researcher at the Laboratory of Psychology and Psychophysiology of Creativity,  
Institute of Psychology Russian Academy of Sciences, Moscow, Russia  
ekval@mail.ru; ORCID: 0000-0003-3637-287X

© **Yulia A. Dmitrieva**

PhD in Psychology, Associate Professor, Department of Psychology of Management and Performance,  
South Ural State University, Chelyabinsk, Russia  
dmitrieva.julia.86@mail.ru; ORCID: 0000-0002-0331-4684

© **Vera G. Gryazeva-Dobshinskaya**

Doctor of Psychological Sciences, Professor, Head of the laboratory, Laboratory of Psychology  
and Psychophysiology of stress resistance and creativity, South Ural State University, Chelyabinsk, Russia  
vdobshinya@mail.ru; ORCID: 0000-0002-9986-4073

© **Svetlana Yu. Korobova**

PhD in Psychology, Researcher, Laboratory of Psychology and Psychophysiology of stress resistance  
and creativity, South Ural State University, Chelyabinsk, Russia  
k.svetlana-1991@mail.ru; ORCID: 0000-0002-8633-7231

© **Evgeniy I. Koltunov**

Researcher, Laboratory of Psychology and Psychophysiology of stress resistance and creativity,  
South Ural State University, Chelyabinsk, Russia  
aspiratingle@gmail.com; ORCID: 0000-0003-4836-7859

The reported study was funded by Russian Science Foundation (RSF) № 23–18–01059,  
<https://rscf.ru/project/23-18-01059/>

The article presents an approach to developing models for automating the Torrance test creativity assessment procedure. The research conducted demonstrated the best results in the further training of the Swin-base model using images of subtest 2. The accuracy achieved was 0,74–0,88 for originality indicator and 0,69–0,82 for elaboration indicator for individual types of source images. Model training to predict the level of originality based on the abstractness of titles indicator for subtest 2 images through retraining the Bert-based model demonstrated an accuracy of 0,79.

**Keywords:** creativity, psychological diagnostics, creativity diagnostics, computer vision, machine learning, natural language analysis, Torrance Test of Creative Thinking (TTCT), creativity level, originality, elaboration

## REFERENCES

- Markina N. V., Matveeva L. G. (2004). Minnesota tests of creative thinking (MTTM graphic form). Chelyabinsk: PsiKHRO N.
- Acar S., Organisciak P., Dumas D. (2025). Automated scoring of figural tests of creativity with computer vision // The Journal of Creative Behavior. V. 59. № 1. Art. e677.
- Bao H., Dong L., Piao S., Wei F. (2021). Beit: Bert pre-training of image transformers // arXiv preprint arXiv:2106.08254.
- Beaty R.E., Johnson D.R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance // Behavior Research Methods. V. 53. № 2. P. 757–780. DOI: <https://doi.org/10.3758/s13428-020-01453-w>.
- Cropley D.H., Marrone R.L. (2025). Automated Scoring of Figural Creativity using a Convolutional Neural Network // Psychology of Aesthetics Creativity and the Arts. V. 19. № 1. P. 77–86. DOI: [10.1037/aca0000510](https://doi.org/10.1037/aca0000510).
- Cropley D.H., Theurer C., Mathijssen A.S., Marrone R.L. (2024). Fit-For-Purpose Creativity Assessment: Automatic Scoring of the Test of Creative Thinking — Drawing Production (TCT-DP) // Creativity Research Journal. P. 1–16. DOI: [10.1080/10400419.2024.2339667](https://doi.org/10.1080/10400419.2024.2339667).

- Devlin J., Chang M.W., Lee K., Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies. V. 1. P. 4171–4186.
- He K., Zhang X., Ren S., Sun J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. P. 770–778.
- Liu Z., Hu H., Lin Y., Yao Z., Xie Z., Wei Y., Ning J., Cao Y., Zhang Z., Dong L., Wei F., Guo B. (2022). Swin transformer v2: Scaling up capacity and resolution // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. P. 12009–12019.
- Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows // Proceedings of the IEEE/CVF international conference on computer vision. P. 10012–10022.
- Oquab M., Darcet T., Moutakanni T., Vo H., Szafraniec M., Khalidov V., Fernandez P., Haziza D., Massa F., El-Nouby A., Assran M., Ballas N., Galuba W., Howes R., Huang P., Li S., Misra I., Rabbat M, Sharma V., Synnaeve G., Xu H., Jegou H., Mairall J., Labatut P., Joulin A., Bojanowski P. (2023). Dinov2: Learning robust visual features without supervision // arXiv preprint arXiv:2304.07193.
- Organisciak P., Acar S., Dumas D., Berthiaume K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models // Thinking Skills and Creativity. V. 49. Art. 101356. DOI: 10.1016/j.tsc.2023.101356.
- Panfilova A.S., Valueva E.A., Ilyin, I.Y. (2024) The application of explainable artificial intelligence methods to models for automatic creativity assessment // Frontiers in Artificial Intelligence. V. 7. Art. 1310518. DOI: 10.3389/frai.2024.1310518.
- Panfilova A.S., Valueva E.A. (2025). Deep Learning and Explainable AI for Creativity Scoring in TCT-DP Form B // Psychology. Journal of the Higher School of Economics. 2025. V. 22. № 4. P. 721–735. DOI: 10.17323/1813-8918-2025-4-721-735.
- Patterson J.D., Barbot B., Lloyd-Cox J., Beaty R.E. (2024). AuDrA: An automated drawing assessment platform for evaluating creativity. Behavior Research Methods. V. 56. № 4. P. 3619–3636. DOI: 10.3758/s13428-023-02258-3.
- Torrance E.P. (2008). The Torrance Tests of Creative Thinking Norms-Technical Manual Figural (Streamlined) Forms A & B. Bensenville, IL: Scholastic Testing Service.
- Woo S., Debnath S., Hu R., Chen X., Liu Z., Kweon I.S., Xie S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. P. 16133–16142.
- Zhang H., Dong H., Wang Y., Zhang X., Yu F., Ren B., Xu J. (2025). Automated Graphic Divergent Thinking Assessment: A Multimodal Machine Learning Approach // Journal of Intelligence. V. 13. № 4. P. 45. DOI: 10.3390/jintelligence13040045.