МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет
вычислительной техники

Кафедра
«Вычислительная техника»

Направление подготовки: 09.04.01 «Информатика и вычислительная техника»
Магистерская программа: «Прикладной искусственный интеллект»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

на тему

СТАРТАП-ПРОЕКТ: ИНСТРУМЕНТАРИЙ ВЫЯВЛЕНИЯ ПРОФЕССИОНАЛЬНЫХ СКЛОННОСТЕЙ И ТВОРЧЕСКОГО ПОТЕНЦИАЛА УЧАЩИХСЯ

Студент _____ Дмитриев Юрий Алексеевич
(подпись, дата) 28.05.2026 (ФИО полностью)

Руководитель _____ 28.05.2026 Карамышева Н.С.
(подпись, дата) (фамилия, инициалы)

Нормоконтролёр _____ 28.05.2026 Бычков А.С.
(подпись, дата) (фамилия, инициалы)

Рецензент нач. УИТ _____ Севастьянов Д.В.
(должность, место работы) (подпись, дата) (фамилия, инициалы)

Работа допущена к защите (протокол зас. кафедры от 22.05.2026 № 14)

Заведующий кафедрой _____ М.А. Митрохин
(подпись) (фамилия, инициалы)

Работа защищена с отметкой отлично (протокол заседания ГЭК от 09 июня 2026 г. № 13)

Секретарь ГЭК _____ Е.И. Гурин
(подпись) (фамилия, инициалы)

Пенза 2026

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет
вычислительной техники

Кафедра
«Вычислительная техника»

«Утверждаю»
Заведующий кафедрой
_____М.А. Митрохин
«_20_» _09_____20_24_ г.

ЗАДАНИЕ
НА МАГИСТЕРСКУЮ ДИССЕРТАЦИЮ

1. Студент_Дмитриев Юрий Алексеевич_____гр._24ВВИм2__факультета__ВТ_____

2.Тема работы__Стартап-проект: инструментарий выявления профессиональных склонностей и творческого потенциала учащихся _____

Тема утверждена приказом ПГУ №__920/о__от "_20_" _сентября__2024 г.

3. Руководитель магистерской диссертации __Карамышева Надежда Сергеевна__
4. Задание на работу (назначение разработки, исходные данные и т.п.)
Разработать и экспериментально апробировать программный инструментарий для комплексной диагностики профессиональных склонностей и творческого потенциала учащихся. Исходные данные: классические психометрические методики (опросники и тесты креативности), современный технологический стек веб-разработки (язык Python, веб-фреймворк, база данных), библиотеки для анализа изображений и машинного обучения, а также возможность интеграции с внешними сервисами для формирования рекомендаций. Ожидаемый результат: создание работоспособного веб-приложения, обучение моделей машинного обучения для классификации профессиональных типов и оценки креативности, проведение экспериментальной апробации на выборке учащихся, статистическое подтверждение эффективности разработанного инструментария и формулировка практических рекомендаций по его внедрению в образовательный процесс._____
5. Перечень подлежащих разработке вопросов 1. Провести теоретический анализ предметной области по проблеме диагностики профессиональных склонностей и творческого потенциала, а также по применению методов машинного обучения и компьютерного зрения в психодиагностике. 2. Определить функциональные и технические требования к разрабатываемому инструментарию, обосновать выбор технологического стека и архитектуры веб-приложения. 3. Разработать и программно реализовать веб-приложение, включающее модули адаптивного тестирования, анализа творческих изображений и формирования

персонализированных рекомендаций. 4. Разработать алгоритм автоматизированной оценки творческого потенциала на основе анализа изображений (вычисление коэффициента добавленных элементов). 5 Обучить и валидировать модели машинного обучения (Random Forest, Gradient Boosting) для классификации профессиональных типов и оценки креативности. 6. Организовать и провести экспериментальную апробацию разработанного инструментария на выборке учащихся, включая сбор процессуальных метрик и экспертную оценку. 7 Выполнить статистическую обработку полученных данных (t-критерий Стьюдента, корреляционный анализ, метрики качества классификации) и оценить юзабилити приложения.

6. Календарный график выполнения работы

№ п/п	Наименование этапов работы	Объем работы	Срок выполнения	Подпись руководителя
1	Анализ предметной области по проблеме диагностики профессиональных склонностей и творческого потенциала	15%	20.092024 – 18.11.2024	
2	Проектирование архитектуры системы (разработка структуры модулей, выбор технологического стека)	10%	19.11.2024 – 26.12.2024	
3	Реализация веб-приложения, включающее модули адаптивного тестирования, анализа, рекомендаций	15%	27.12.2024 – 24.02.2025	
4	Реализация алгоритма автоматизированной оценки творческого потенциала на основе анализа изображений	20%	25.02.2025 – 15.09.2025	
5	Обучение и валидация модели машинного обучения (Random Forest, Gradient Boosting)	20%	16.09.2025 – 25.12.2025	
6	Организация и проведение экспериментальной апробации разработанного инструментария	10%	26.12.2025 – 24.02.2026	
7	Выполнение статистической обработки полученных данных	10%	25.02.2026 – 13.04.2026	
8	Оформление пояснительной записки, подготовка презентации и доклада	10%	14.04.2026 – 21.05.2026	

Дата выдачи задания "_20_" _сентября__2024 г.

Руководитель магистерской диссертации _____
(подпись, дата)

Задание к исполнению принял студент _____
(подпись, дата)

Работу к защите допустить
Декан факультета А.Р. Фисенко
(подпись, дата) 02.06.2026

РЕФЕРАТ

Магистерская работа содержит 148 листов, 14 рисунков, 37 таблиц, 54 источника.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ПРОФЕССИОНАЛЬНАЯ ОРИЕНТАЦИЯ, ПРОФЕССИОНАЛЬНЫЕ СКЛОННОСТИ, КОМПЬЮТЕРНОЕ ЗРЕНИЕ, ВЕБ-ПРИЛОЖЕНИЕ, ПСИХОДИАГНОСТИКА, АДАПТИВНОЕ ТЕСТИРОВАНИЕ, ОБРАБОТКА ИЗОБРАЖЕНИЙ, PYTHON, FLASK, RANDOM FOREST, GRADIENT BOOSTING

В ходе выполнения магистерской работы, посвящённой разработке инструментария выявления профессиональных склонностей и творческого потенциала учащихся, был проведён критический анализ и систематизация алгоритмов машинного обучения, применимых к задачам психолого-педагогической диагностики. В результате выделены наиболее релевантные подходы: методы классификации (Random Forest, Gradient Boosting) для определения профессиональных типов по методикам Е.А. Климова и Дж. Голланда, а также техники обработки естественного языка (NLP) и методы компьютерного зрения для анализа творческих изображений в рамках теста Э.П. Торренса.

На основе анализа были сформулированы функциональные и технические требования к программному обеспечению, предусматривающие реализацию адаптивного тестирования, интеграцию ML-моделей в реальном времени и генерацию интерпретируемых рекомендаций на основе синтеза диагностических данных. Для реализации этих требований обоснован и выбран технологический стек (Python, Flask, SQLite, SQLAlchemy, Scikit-learn, Pillow, NumPy, а также архитектуры свёрточных нейронных сетей ResNet101

и DeepLabV3Plus для перспективного анализа визуальных данных).

В процессе выполнения работы использовались такие технологии разработки, как язык программирования Python, веб-фреймворк Flask, реляционная база данных SQLite с использованием ORM SQLAlchemy, библиотеки для обработки изображений Pillow и NumPy, инструменты анализа данных pandas и scikit-learn, методы классификации (Random Forest, Gradient Boosting), а также архитектуры свёрточных нейронных сетей для анализа визуальных данных.

Объект исследования – процесс разработки и применения цифрового инструментария для комплексной диагностики профессиональных склонностей и творческого потенциала учащихся на основе интеграции разнородных диагностических данных.

Предмет исследования – архитектура, алгоритмы машинного обучения и методы обработки диагностических данных (опросники, творческие изображения, поведенческие метрики) в составе веб-приложения, предназначенного для выявления профессиональных склонностей и творческого потенциала.

Цель работы – разработать, программно реализовать и апробировать веб-приложение «IAURA», обеспечивающее более точное и персонализированное выявление профессиональных склонностей и творческого потенциала учащихся за счёт интеграции адаптивного тестирования, анализа мультимодальных данных и применения современных ML-подходов.

Основным результатом этапа работы стало создание детальной проектной спецификации инструментария, включающей:

– выбор и обоснование ML-алгоритмов для классификации профессиональных типов и оценки креативности,

– модель источников и структуры диагностических данных (опросники, творческие изображения, процессуальные метрики) для обучения и анализа,

– требования к системе по сбору, обработке и интерпретации диагностических данных,

– архитектурное и технологическое решение веб-приложения «IAURA».

Эти данные служат фундаментом для программной реализации и последующей экспериментальной валидации приложения.

Результатом магистерской работы является веб-приложение «IAURA» спроектированное и разработанное таким образом, что позволяет проводить комплексную диагностику профессиональных склонностей и творческого потенциала учащихся с использованием четырёх психологических методик (Климова, Холла, Голланда, Торренса), а так же автоматизированную обработку результатов тестирования, включая анализирование творческих изображений с вычислением коэффициента добавленных элементов (added_ratio) и предоставления персонализированных отчётов с текстовым представлением, а также интеграцию с HeadHunter API для предоставления ссылок на вакансии с учётом региона пользователя.

Результат тестирования системы подтвердил её эффективность, так средняя точность прогнозирования составила 79%, оценка удовлетворённости пользователя по шкале SUS – составила 82,5 балла, а коэффициент корреляции с экспертными оценками – 0,75 балла.

СОДЕРЖАНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ……..……...7
ВВЕДЕНИЕ..... 8
1 Теоретические и технологические основы интеграции мультимодальных данных..... 15
1.1 Психолого-педагогические основы диагностики профессиональных склонностей и творческого потенциала в контексте интеграции мультимодальных данных..... 15
1.2 Профессиональные склонности как переход от классических типологий к анализу мультимодальных цифровых следов..... 17
1.3 Творческий потенциал (креативность) как количественный индикатор на основе мультимодальных данных..... 18
1.4 Мультимодальный подход, как склонности и креативность в едином диагностическом пространстве..... 20
1.5 Технологический стек для разработки мультимодального веб-приложения. Обоснование выбора инструментов..... 21
1.6 Система управления базами данных..... 23
1.7 Возможности машинного обучения в мультимодальной психолого-педагогической диагностике..... 24
1.8 Выводы по разделу 1 28
2 Организация и методика экспериментального исследования..... 30
2.1 Программа и дизайн экспериментального исследования..... 30
2.2 Характеристика выборки и базы исследования..... 42
2.3 Описание диагностического инструментария (методики сбора данных)..... 49
2.4 Критерии, показатели и методы математико-статистической обработки данных 63
2.5 Выводы по разделу 2 77
3. Аанализ результатов экспериментального исследования 80

3.1 Сравнительный анализ эффективности цифрового и традиционного инструментария 80
3.3 Анализ процессуальных метрик (время ответа, согласованность, поведенческие паттерны) 100
3.4 Оценка пользовательского опыта и юзабилити системы 109
3.5 Статистическая проверка гипотез исследования 114
3.6 Выводы по разделу 3 124
ЗАКЛЮЧЕНИЕ 126
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 129
ПРИЛОЖЕНИЕ А – Материалы эксперементального иследования. 135
ПРИЛОЖЕНИЕ Б – Листинг кода 141

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

ПО – программное обеспечение.

ФГОС – Федеральный государственный образовательный стандарт.

API (Application Programming Interface) – интерфейс программирования приложений.

CLIP (Contrastive Language-Image Pre-training) – нейросетевая модель для сравнения семантической близости изображений и текстовых описаний.

CNN (Convolutional Neural Network) – свёрточная нейронная сеть.

HTML (HyperText Markup Language) – язык гипертекстовой разметки.

JSON (JavaScript Object Notation) – текстовый формат обмена данными.

ML (Machine Learning) – машинное обучение.

NLP (Natural Language Processing) – обработка естественного языка.

RESNET / RESNET101 – архитектура глубокой нейронной сети с остаточными связями

REST API (Representational State Transfer API) – архитектурный стиль взаимодействия компонентов распределённого приложения в сети.

SD (Standard Deviation) – стандартное отклонение.

SQL (Structured Query Language) – язык структурированных запросов.

SQLAlchemy – библиотека ORM для языка Python.

SUS (System Usability Scale) – шкала оценки юзабилити.

Web-приложение – клиент-серверное приложение.

ВВЕДЕНИЕ

В условиях современного рынка труда, отличающегося высокой динамикой, автоматизацией рутинных операций и растущим спросом на креативных специалистов, особую значимость приобретает интеграция мультимодальных данных (опросники, изображения, тексты, поведенческие метрики) для прогнозирования профессиональных траекторий. Ключевым вызовом становится не просто информирование учащихся о профессиях, а раннее и точное выявление их глубинных профессиональных склонностей и творческого потенциала как основы для построения успешной и устойчивой карьерной траектории.

Нужно также учитывать, что современный этап общественного развития характеризуется стремительной интеграцией цифровых технологий в структуру повседневной и профессиональной деятельности человека, что знаменует переход к качественно новой модели социального взаимодействия. Цифровая трансформация охватывает все более широкие сферы жизнедеятельности, видоизменяя устоявшиеся формы практик и формируя принципиально иные паттерны мышления и поведения. Активное развитие технологических решений развивает интерес к исследованию цифровых решений, новому осмыслению подходов к решению традиционных задач, в том числе, в области профессионального самоопределения и развития творческих способностей учащихся.

В условиях современной цифровой трансформации образования особую роль приобретают разработки инструментов, способных обеспечить индивидуальный подход к анализу и развитию личности учащегося. Одним из самых перспективных направлений в этой области является применение методов машинного обучения, а также анализ изображений для оценки творческого потенциала, поскольку креативность, проявляющаяся в формах рисунков, представляет собой сложный для изучения, но критическиочень важный элемент профессиональной ориентации. Анализ и обработка изображений открывают новые возможности для объективной оценки

различных параметров творческого мышления, таких как оригинальность, гибкость и проработанность идей, которые обычно оцениваются исключительно экспертами.

В последнее время стремительно возрос интерес к цифровым, электронным и оптическим методам обработки изображений, применяемым в самых различных областях – от анализа медицинских снимков до обработки данных дистанционного зондирования Земли. Значительное внимание в научной литературе уделяется работам, связанным с биомедицинскими и космическими исследованиями, а также с аэрофотосъемкой и промышленной радиографией. Однако применение методов компьютерного зрения и машинного обучения в сфере психолого-педагогической диагностики, в частности для анализа творческих продуктов (рисунков, дорисовок, графических образов), остается недостаточно изученным и требует как разработки новых алгоритмов обработки изображений, так и теоретического осмысления полученных результатов.

Актуальность исследования заключается в том, что современный рынок труда, характеризующийся высокой динамикой, автоматизацией рутинных операций и растущим спросом на специалистов с развитыми творческими компетенциями, предъявляет новые требования к системе профессиональной ориентации. Ключевым вызовом становится не просто информирование учащихся о существующих профессиях, а раннее и точное выявление их глубинных профессиональных склонностей и творческого потенциала как основы для построения успешной карьерной траектории.

Традиционный инструментарий профориентации (опросники Е.А. Климова, Дж. Голланда, тест Э.П. Торренса) обладает доказанной валидностью, однако его «оцифрованная» версия, реализованная без трансформации диагностического подхода, не использует потенциал современных технологий для сбора процессуальных метрик и автоматизированной обработки творческих продуктов. Параллельно с этим, развитие методов машинного обучения и компьютерного зрения открывает

новые возможности для адаптивного тестирования, анализа неструктурированных данных (изображений, текстов) и повышения прогностической точности диагностики.

Таким образом, возникает противоречие между потребностью практики в точных, персонализированных и доступных цифровых инструментах профориентации и недостаточной разработанностью научно обоснованных решений, интегрирующих классические психометрические методики с современными ML-подходами.

Проблема исследования заключается в поиске и экспериментальной проверке эффективного подхода к созданию цифрового инструментария, преодолевающего ограничения традиционных методов за счёт интеграции методов машинного обучения, компьютерного зрения и анализа процессуальных метрик.

Цель исследования – разработка, программная реализация и экспериментальная апробация веб-приложения «IAURA», обеспечивающего комплексную диагностику профессиональных склонностей и творческого потенциала учащихся на основе интеграции методов машинного обучения и компьютерного зрения.

Объект исследования – процесс цифровизации и повышения эффективности профессиональной ориентации учащихся в условиях современной образовательной среды.

Предмет исследования – веб-приложение с элементами машинного обучения, объединяющее модули адаптивного психологического тестирования, анализа творческих изображений и формирования персонализированных рекомендаций по профессиональному самоопределению.

Гипотезы исследования.

Основная гипотеза (H_1). Информационная система, интегрирующая адаптивное тестирование, методы машинного обучения для классификации профессиональных типов и алгоритмы компьютерного зрения для анализа

творческих изображений, обеспечивает более высокую точность выявления профессиональных склонностей и творческого потенциала учащихся по сравнению с традиционными методами.

Частные гипотезы:

– H_1а. средняя точность классификации профессионального типа в экспериментальной группе превышает аналогичный показатель в контрольной группе не менее чем на 10 процентных пунктов,

– H_1б. существует статистически значимая положительная корреляция ($r \geq 0,7$) между результатами автоматизированной оценки творческого потенциала и независимой экспертной оценкой,

– H_2. система «IAURA» обладает высоким уровнем юзабилити, что подтверждается итоговым баллом по шкале SUS, превышающим 68 баллов (средний показатель для веб-приложений),

– H_3. процессуальные метрики (время ответа, согласованность выборов, количество возвратов) обладают прогностической значимостью и вносят вклад в точность классификации профессиональных типов.

Провести теоретический анализ психолого-педагогической литературы по проблеме диагностики профессиональных склонностей и творческого потенциала, а также по применению методов машинного обучения и компьютерного зрения в психодиагностике.

Определить требования, архитектуру, технологический стек и выбрать алгоритмы машинного обучения для разработки веб-приложения «IAURA».

Разработать и программно реализовать веб-приложение, включающее модули адаптивного тестирования (методики Е.А. Климова, Н. Холла, Дж. Голланда), анализа творческих изображений (тест Э.П. Торренса с вычислением added_ratio) и формирования персонализированных рекомендаций с интеграцией HeadHunter API.

Обучить, валидировать и интегрировать в приложение ML-модели (Random Forest, Gradient Boosting) для классификации профессиональных типов и оценки креативности на основе визуальных и поведенческих данных.

Организовать и провести экспериментальную апробацию приложения на выборке учащихся (n=90) Многопрофильного колледжа Пензенского государственного университета, оценив его точность прогнозирования, юзабилити и корреляцию с экспертными оценками.

Сформулировать выводы и практические рекомендации по внедрению разработанного инструментария.

Методологическая основа и методы исследования:

– теоретические методы: анализ, синтез и систематизация научной литературы; сравнительный анализ существующих цифровых профориентационных решений; моделирование архитектуры веб-приложения,

– эмпирические методы: анкетирование, компьютерное адаптивное тестирование (опросники Климова, Холла, Голланда, тест Торренса), экспертная оценка творческих изображений (три независимых эксперта-психолога), констатирующий и формирующий эксперимент (n=90, экспериментальная и контрольная группы по 45 человек),

– методы математической и компьютерной обработки данных: описательная статистика, t-критерий Стьюдента, U-критерий Манна–Уитни, корреляционный анализ Пирсона, критерий χ^2, метрики качества классификации (Accuracy, Precision, Recall, F1-score, AUC-ROC), расчёт размера эффекта Cohen's d. Обработка выполнена с использованием библиотек Python (Pandas, SciPy, Statsmodels, Scikit-learn),

– методы машинного обучения и компьютерного зрения: обучение с учителем (Random Forest, Gradient Boosting), анализ творческих изображений (вычисление added_ratio с использованием библиотек Pillow и NumPy), анализ процессуальных метрик (consistency_score, время ответа, количество возвратов), кросс-валидация,

– программно-инженерные методы: проектирование ПО (UML, ER-диаграммы), веб-разработка (Python/Flask, SQLite, SQLAlchemy, HTML/CSS/JS), разработка и интеграция ML-моделей (Scikit-learn).

Научная новизна исследования:

– теоретическая новизна: Уточнено и операционализировано понятие «цифровой инструментарий профориентации» применительно к условиям использования алгоритмов машинного обучения и компьютерного зрения для комплексной оценки профессиональных склонностей и творческого потенциала учащихся,

– методическая новизна: Предложена и реализована оригинальная методика диагностики, объединяющая адаптированные версии классических опросников с автоматизированным анализом творческих изображений на основе вычисления коэффициента добавленных элементов (added_ratio) и сбором процессуальных метрик в едином цифровом контуре,

– Технологическая новизна: Разработана архитектура веб-приложения, в котором ML-модели являются ядром диагностического процесса, определяющим логику классификации и формирования итогового прогноза. Впервые в отечественной практике реализован автоматический сбор и анализ процессуальных метрик для повышения точности классификации.

Практическая значимость исследования:

– создан готовый к внедрению веб-продукт «IAURA» для использования в школьных психологических службах, центрах профориентации, службах занятости населения и образовательных учреждениях,

– разработан набор обученных ML-моделей (Random Forest, Gradient Boosting) с точностью 82–84% и методика их применения для классификации профессиональных склонностей,

– получены эмпирические данные о взаимосвязях между процессуальными метриками, результатами ML-классификации и точностью диагностики (R^2 = 0,48), что вносит вклад в область образовательной аналитики,

– материалы исследования (архитектура системы, методика сбора процессуальных метрик, обученные ML-модели) могут быть использованы в

учебных курсах по педагогике, психологии, профориентации и data science в образовании.

База исследования включала экспериментальную апробацию, которая проводилась на базе Многопрофильного колледжа ФГБОУ ВО «Пензенский государственный университет». В исследовании приняли участие 90 студентов 1–2 курсов (экспериментальная группа – 45 человек, контрольная группа – 45 человек), сбалансированных по возрасту (17–22 года), полу, курсу обучения и академической успеваемости ($p > 0,05$).

Приложение доступно по адресу: https://iaura.ru

1 Теоретические и технологические основы интеграции мультимодальных данных

Проект посвящён интеграции мультимодальных данных (опросники, изображения, тексты, поведенческие метрики) для прогнозирования профессиональных траекторий. В современном образовательном пространстве существует острая потребность в инструментах, позволяющих молодым людям определить свои профессиональные склонности и психологические особенности на основе комплексного анализа разнородных данных для осознанного выбора карьерного пути.

Информационная система решает следующие задачи:

- предоставление доступа к профессиональному психологическому тестированию через веб-интерфейс с возможностью сбора и интеграции мультимодальных данных (опросники, творческие работы, поведенческие метрики),
- автоматизация обработки и анализа результатов тестов с применением алгоритмов машинного обучения для синтеза разнородной информации,
- формирование комплексного профиля пользователя на основе интеграции данных из нескольких методик и модальностей,
- интеграция с платформами поиска работы для практического применения результатов прогнозирования профессиональных траекторий,
- обеспечение конфиденциальности и безопасности персональных данных при работе с чувствительной мультимодальной информацией.

1.1 Психолого-педагогические основы диагностики профессиональных склонностей и творческого потенциала в контексте интеграции мультимодальных данных

В рамках темы «Интеграция мультимодальных данных (опросники, изображения, тексты, поведенческие метрики) для прогнозирования профессиональных траекторий» особое значение приобретает теоретическое

осмысление и операционализация ключевых конструктов, лежащих в основе диагностики.

Традиционные подходы к профориентации фокусируются на выявлении профессиональных склонностей через стандартизированные опросники и тесты, а также на оценке творческого потенциала посредством анализа продуктов деятельности. Однако для построения точных и персонализированных прогнозов профессиональных траекторий в современных условиях этого недостаточно.

Операционализация ключевых конструктов в данном исследовании предполагает их декомпозицию на измеримые параметры, которые могут быть получены из различных источников данных.

Так профессиональные склонности рассматриваются не как статичная характеристика, а как динамический комплекс предрасположенностей, проявляющийся в:

- ответах на опросники (структурированные данные, отражающие интересы и ценностные ориентации),
- поведенческих метриках (время реакции, паттерны навигации по тесту, последовательность выбора ответов), которые позволяют выявить скрытые когнитивные стили и уровень вовлечённости,
- анализе текстов (творческие эссе, свободные ответы), где NLP-методы позволяют оценить широту интересов, уровень абстрактного мышления и вербальную креативность.

Творческий потенциал операционализируется через совокупность характеристик, оцениваемых по разным модальностям:

- продукты творческой деятельности (изображения, тексты), анализируемые с помощью алгоритмов компьютерного зрения и NLP для выявления оригинальности, сложности и новизны,
- поведенческие метрики в процессе решения нестандартных задач (гибкость, беглость мышления), фиксируемые системой.

Таким образом, интеграция данных из опросников, изображений, текстов и поведенческих метрик позволяет сформировать многомерный цифровой профиль личности. Этот профиль служит надёжной основой для применения алгоритмов машинного обучения с целью прогнозирования наиболее вероятных и успешных профессиональных траекторий индивида.

1.2 Профессиональные склонности как переход от классических типологий к анализу мультимодальных цифровых следов

В контексте темы «Интеграция мультимодальных данных (опросники, изображения, тексты, поведенческие метрики) для прогнозирования профессиональных траекторий» подход к диагностике профессиональных склонностей претерпевает фундаментальную трансформацию. Происходит переход от статических типологий к динамическому анализу комплексных цифровых следов.

В психологии профессий профессиональные склонности понимаются как устойчивая, внутренне обусловленная направленность индивида на определенный вид профессиональной деятельности, проявляющаяся в стремлении, интересе и позитивном эмоциональном отношении к ее содержанию [1]. Классические подходы (Е.А. Климов, Дж. Холланд) предлагают типологические модели, где склонности группируются вокруг объектов труда (природа, техника, человек, знаковая система, художественный образ) или профессиональных типов личности (реалистический, исследовательский, артистический и т.д.) [2, 3].

Для цифровой диагностики ключевым становится не только конечный типологический результат, но и процессуальные метрики, фиксируемые в ходе интерактивного взаимодействия пользователя с системой. К ним относятся:

Метриками выбора является частота выбора заданий определенной категории (технической, гуманитарной, социальной).

Временными метриками являются время, затраченное на решение задач разного типа. Сокращение времени при высокой точности может указывать на

легкость и склонность, а увеличение времени при низкой результативности – на затруднения или отсутствие интереса.

Метриками последовательности являются паттерны переходов между заданиями разных типов. Например, сознательный или неосознанный возврат к задачам определенной тематики.

Метриками уверенности/согласованности являются оценка пользователем степени своей уверенности в ответе (по шкале Лайкерта). Внутренняя противоречивость выборов (например, одновременный высокий интерес к «рутинной работе с документами» и «творческой импровизации») является самостоятельным диагностическим маркером.

Реакцией на адаптивность являются изменение точности и скорости ответов в зависимости от динамического изменения сложности заданий выбранного профиля.

Таким образом, профессиональная склонность в цифровом контексте может быть представлена не как статичный ярлык, а как многомерный вектор, включающий предпочтения, эффективность, когнитивную вовлеченность и эмоциональный отклик в специфических контекстах.

1.3 Творческий потенциал (креативность) как количественный индикатор на основе мультимодальных данных

Диагностика творческого потенциала выходит за рамки традиционных психометрических тестов и приобретает новое измерение. Ключевым конструктом здесь выступает дивергентное мышление – способность генерировать множество оригинальных идей в условиях неопределённости.

Если классические подходы к оценке креативности опирались на качественные или простые количественные показатели (например, количество ответов в тестах Гилфорда), то современный подход, основанный на анализе мультимодальных данных, позволяет получить глубокую и объективную картину.

Творческий потенциал рассматривается в работе как способность к порождению новых, оригинальных и адаптивных идей, решений и продуктов.

Основополагающей моделью для измерения креативности является концепция дивергентного мышления Дж. Гилфорда и Э.П. Торренса, выделяющая четыре ключевых параметра [4, 5]:

1) беглость (Fluency) – количество генерируемых идей за единицу времени,

2) гибкость (Flexibility) – способность переключаться между различными категориями, подходами и стратегиями мышления,

3) Оригинальность (Originality) – статистическая редкость, неочевидность предлагаемой идеи,

4) Разработанность (Elaboration) – уровень детализации, проработанности идеи.

Операционализация данных параметров в цифровой среде предполагает следующие подходы, а именно для беглости и гибкости это прямая автоматическая фиксация количества ответов (беглость) и автоматическая категоризация ответов с помощью предобученных тезаурусов или кластеризации текстовых эмбеддингов для подсчета числа семантических кластеров (гибкость). Для оригинальности это использование методов обработки естественного языка (NLP). Алгоритм может оценивать оригинальность текстового ответа через статистическую частотность лексем и n-грамм в корпусе тренировочных текстов. Семантическое расстояние между ответом пользователя и наиболее типичными ответами, вычисляемое с помощью векторных моделей (Word2Vec, FastText, BERT). Чем больше дистанция, тем выше потенциальная оригинальность. Меры синтаксической сложности (длина предложений, разнообразие конструкций).

Для разработанности это анализ объема текста (количество слов, символов), плотности описательных признаков, наличия причинно-следственных связей, выявляемых синтаксическими парсерами.

1.4 Мультимодальный подход, как склонности и креативность в едином диагностическом пространстве

Важнейшим теоретическим положением, лежащим в основе проектируемого инструментария, является взаимосвязь профессиональных склонностей и творческого потенциала. Креативность не является унитарным свойством, а проявляется специфически в различных профессиональных доменах (техническое творчество, художественное, социальное предпринимательство и т.д.) [6]. Следовательно, диагностика должна выявлять не абстрактный «уровень креативности», а профиль креативных компетенций, соотнесенный с типом профессиональных склонностей.

Цифровая операционализация интегративного подхода может включать в себя контекстуализированные творческие задачи, такие, как предъявление проблемных ситуаций, релевантных разным профессиональным сферам (например, «предложите улучшение для мобильного приложения» – для IT-склонностей, «разработайте слоган для социальной кампании» – для социально-гуманитарных). Анализ cross-domain метрик, как сравнение показателей беглости, гибкости и оригинальности пользователя при решении задач из разных профессиональных областей. Пик значений в определенном домене указывает на зону наиболее продуктивного сочетания склонностей и креативности. Моделирование на основе данных, как использование алгоритмов машинного обучения (например, кластеризации) для выявления в массиве собранных данных (метрики склонностей + метрики креативности) устойчивых паттернов, которые могут не соответствовать априорным теоретическим типологиям, но будут иметь высокую прогностическую валидность.

Современный цифровой инструментарий диагностики должен опираться не только на констатацию результата (выбранный тип), но и на анализ процессуальных метрик (время, последовательность, уверенность), которые являются богатым источником данных для машинного обучения.

Параметры творческого потенциала (беглость, гибкость, оригинальность, разработанность) поддаются эффективной квантификации с применением методов статистического анализа и NLP, что позволяет перейти от субъективной экспертной оценки к объективным алгоритмическим измерениям.

Перспективным направлением является интегративная диагностика, где профиль профессиональных склонностей и профиль креативных компетенций анализируются во взаимосвязи, что требует специального дизайна заданий и многоуровневых аналитических моделей. Данный подход формирует теоретический фундамент для проектирования архитектуры разрабатываемого веб-приложения.

1.5 Технологический стек для разработки мультимодального веб-приложения. Обоснование выбора инструментов

Разработка научно-диагностического веб-приложения, в отличие от коммерческого стартапа, предъявляет особые требования к технологическому стеку: необходимы гибкость для исследовательских итераций, прозрачность и контроль над логикой обработки данных, эффективная интеграция вычислительных моделей (ML), а также обеспечение надежности и конфиденциальности хранимой персональной и диагностической информации. Данный раздел посвящен анализу и обоснованию выбора технологий для реализации архитектуры «клиент-сервер», наиболее соответствующей поставленным исследовательским задачам.

Backend-разработка.

Язык программирования Python выбран в качестве основного по ряду критически важных причин.

Данный язык доминирует в Data Science и ML Так Python является де-факто стандартом в научных вычислениях, машинном обучении и анализе данных благодаря богатейшей экосистеме библиотек (NumPy, Pandas, Scikit-learn, TensorFlow/PyTorch, Transformers) [7, 8]. Это позволяет реализовать

аналитические и прогностические модели непосредственно в ядре приложения, без использования сторонних, непрозрачных API [9, 10].

Синтаксис Python и обширные библиотеки позволяют быстро реализовывать и модифицировать исследовательские гипотезы в рабочем коде, что соответствует итеративному характеру научной работы.

Наличие зрелых веб-фреймворков. Для создания серверной части (backend) рассматривались три основных фреймворка представленных в таблице 1.

Таблица 1 – Сравнение основных фреймворков

Критерий	Django	Flask	FastAPI	Выбор для проекта
Архитектура	«Батарейки в комплекте» (MVT). Монолит с четкой структурой.	Микрофреймворк. Минималистичный, дает полную свободу в организации кода.	Микрофреймворк, ориентированный на создание API. Асинхронная поддержка из коробки.	Flask
Сложность/Гибкость	Высокая встроенная функциональность (админ-панель, ORM, аутентификация), но меньшая гибкость для нестандартных задач.	Максимальная гибкость. Позволяет выбирать компоненты (ORM, аутентификацию) самостоятельно.	Гибкость, сравнимая с Flask, с фокусом на производительность API.	Flask предоставляет необходимую свободу для интеграции кастомных ML-пайплайнов и исследовательского кода.
Удобство для API и ML	Создание REST API возможно (Django REST Framework), но добавляет слои абстракции.	Идеально подходит для создания легковесных API-эндпоинтов, которые будут обслуживать ML-модели.	Создан специально для высокопроизводительных API. Автогенерация документации (OpenAPI).	FastAPI является сильным конкурентом. Flask выбран из-за более простой кривой обучения и достаточной производительности для исследовательского прототипа.
Сообщество и документация	Очень большое, отличная документация.	Огромное сообщество, множество расширений.	Растущее сообщество, отличная документация.	Все три имеют хорошую поддержку.

Для целей данного исследовательского прототипа, где требуется максимальная прозрачность логики, прямое манипулирование данными и поэтапная интеграция различных ML-компонентов, выбран Flask [11]. На его основе можно построить приложение, как набор программных модулей

(аутентификация, тестирование, ML-инференс), что значительно упрощает разработку, отладку и описание системы.

Frontend -разработка

Frontend (клиентская часть) отвечает за представление внедрённых тестов, сбор результатов творческих ответов и вывод результатов. Ключевыми требованиями являются такие как адаптивный интерфейс, способность обрабатывать динамический запрос и удобство интеграции с backend API.

Были использованы такие базовые технологии, как HTML5, CSS3 и JavaScript (ES6+) [10].

Для более сложных SPA (Single Page Application) применялась Angular. Однако, для самого прототипа, где основная логическая задача и сама вычислительная нагрузка опирается на сервер, а интерфейс представляет собой последовательность тестов, был использован минималистичный интерфейс. В данном проекте применяется взаимодействие шаблонов для статических страниц и JavaScript с библиотекой Axios для асинхронных запросов (AJAX) к API для основных интерактивных элементов. Это упрощает клиентскую часть и обеспечивает процесс взаимодействия пользователя с системой.

1.6 Система управления базами данных

Выбор базы данных определён самими данными, так как необходимо хранить структурированные данные (профили пользователей, вопросы, ответы, результаты), а также потенциально полуструктурированные (логи сессий, JSON-объекты с параметрами творческих заданий).

Использование реляционной СУБД MySQL [12] обеспечивает целостность данных (ACID), а также имеются необходимые средства для сложных запросов, необходимых для последующей диагностики.

Это решение принято в соответствии со следующими факторами, а именно надежность и соответствие исследовательским стандартам. Так MySQL обеспечивает высокий уровень надежности данных, что категорически важно для сохранения результатов эксперимента.

MySQL предоставляет тип данных JSONB, который позволяет эффективно хранить и запрашивать данные. Это эффективно реализует преимущества реляционной модели.

В будущем, на основе полученных и накопленных данных, можно будет выполнять сложные запросы для анализа непосредственно в БД, что является важным для этапа исследования результатов работы программы.

Заключение по выбору технологического стека

Этот конкретный комплекс технологий представляет собой уравновешенное решение для создания научноисследовательско-диагностического веб-приложения [10, 13]. Он обеспечивает методологическую прозрачность, прямой доступ к логике обработки данных и ML-алгоритмам [14]. Гибкость для исследований: Возможность легко модифицировать и расширять функционал в процессе разработки и апробации.

Надежное и структурированное хранение чувствительной диагностической информации.

Позволяет сконцентрироваться на содержательной (психолого-педагогической и алгоритмической) части работы, а не на преодолении сложностей самой технологии.

Данный технологический фундамент создает условия для реализации ключевого компонента исследования – интеграции модулей машинного обучения в диагностический процесс, что будет подробно рассмотрено в следующем параграфе.

1.7 Возможности машинного обучения в мультимодальной психолого-педагогической диагностике

Внедрение методов машинного обучения (МО) в практику психолого-педагогической диагностики представляет собой не просто технологическую модернизацию, а парадигмальный сдвиг. Он заключается в переходе от подтверждения априорных теоретических моделей к выявлению скрытых паттернов непосредственно в данных, генерируемых в процессе цифрового

взаимодействия. Данный раздел посвящен анализу потенциала ключевых классов алгоритмов МО для решения задач, сформулированных в исследовании.

Обучение с учителем: классификация и регрессия для структурированной диагностики

Алгоритмы обучения с учителем (Supervised Learning) требуют размеченных данных и применяются для задач, где целевая переменная известна. В контексте профориентации это позволяет автоматизировать и усовершенствовать классические подходы:

– мультиклассовая классификация для определения профессионального типа. Задача – отнести профиль учащегося к одному из типов (по Холланду, Климову или иной таксономии). В качестве признаков (features) используются не только итоговые баллы по шкалам, но и процессуальные метрики (время ответа на вопросы определенной категории, последовательность выбора, метаданные о неуверенности). Подходящие алгоритмы:

– ансамбли деревьев (Random Forest, Gradient Boosting, XGBoost/LightGBM): Обладают высокой точностью, устойчивостью к переобучению и, что критически важно, предоставляют оценку важности признаков (feature importance). Это позволяет не только классифицировать, но и интерпретировать, какие именно поведенческие паттерны наиболее значимы для отнесения к тому или иному типу, удовлетворяя требованию прозрачности;

– методы, основанные на расстояниях (k-Nearest Neighbors): Могут быть полезны для реализации рекомендаций по принципу «похожие пользователи выбирали».

– регрессия для прогнозирования уровня развития компетенции. Может использоваться для предсказания вероятности успеха в той или иной профессиональной сфере (на основе исторических данных о карьерных

траекториях) или для оценки степени выраженности склонности/креативности по непрерывной шкале, а не по дискретным категориям.

Обучение без учителя: кластеризация для выявления латентных профилей

Алгоритмы кластеризации (Unsupervised Learning) не требуют заранее заданных меток и направлены на обнаружение внутренней структуры данных. Это открывает принципиально новые диагностические возможности.

– выявление неочевидных, гибридных или контекстуальных профилей. Стандартные типологии (6 типов по Холланду) могут не отражать всего многообразия современных профессий и комбинаций склонностей. Алгоритмы, такие как k-means, иерархическая кластеризация или DBSCAN, могут выявить в данных устойчивые кластеры пользователей, которые не укладываются в традиционные рамки (например, «технический гуманитарий» или «социальный инноватор»). Это позволяет перейти от навязывания жесткой типологии к дескриптивному анализу реального разнообразия профессиональных ориентаций,

– Анализ пространства творческих ответов. Векторные представления (эмбеддинги) творческих текстов, полученные с помощью моделей NLP, можно кластеризовать для выявления стилистических или содержательных типов креативности, не заданных экспертом заранее.

Обработка естественного языка (NLP) для квантификации творческого потенциала

Традиционно оценка творческих работ (эссе, идей) является наиболее субъективной. Современный NLP позволяет перевести качественные характеристики в количественные метрики.

– лексико-статистические метрики.

Расчет лексического разнообразия (Type-Token Ratio, меры на основе n-грамм), использование списков редкой или сложной лексики.

– синтаксическая сложность.

Анализ глубины синтаксического дерева, средней длины предложения, разнообразия грамматических конструкций с помощью синтаксических парсеров (например, spaCy).

Семантическое разнообразие и оригинальность.

Это наиболее перспективное направление. С помощью контекстуальных языковых моделей (BERT, RoBERTa, GPT) текст преобразуется в плотный вектор (эмбеддинг) в семантическом пространстве [15].

Оригинальность может быть рассчитана как семантическое расстояние (косинусная, евклидова) между эмбеддингом данного ответа и эмбеддингами типичных, банальных ответов (усредненный вектор кластера «шаблонных» решений).

Семантическое разнообразие ответа оценивается через дисперсию эмбеддингов его отдельных предложений или через меры внутритекстовой когерентности.

Генерация и оценка идей.

Продвинутым применением является использование дообученных языковых моделей для оценки потенциальной полезности или новизны предложенной идеи в рамках заданного контекста (domain-specific creativity).

Прогнозная аналитика и системы рекомендаций

Итоговой целью инструментария является не констатация, а прогноз и рекомендация. Здесь МО позволяет выйти за рамки простого сопоставления «профиль-профессия».

Рекомендательные системы построены по принципу коллаборативной фильтрации («похожие на вас пользователи интересовались этими курсами/профессиями») или, что более точно для научного инструмента, контент-ориентированной фильтрации. В последнем случае рекомендация формируется на основе семантического сопоставления векторизованного профиля учащегося (включая метрики креативности) с векторными описаниями образовательных треков, профессий будущего или конкретных компетенций.

Прогноз образовательных и карьерных траекторий, так при наличии лонгитюдных данных (что является перспективой развития системы) можно строить модели, предсказывающие вероятность успешного освоения определенной дисциплины или выбора конкретного карьерного пути на основе диагностического профиля, полученного на раннем этапе.

1.8 Выводы по разделу 1

Проведенный анализ в Главе 1 позволяет сформулировать следующие заключительные положения, которые являются прямым обоснованием для проектирования архитектуры разрабатываемого веб-приложения:

– психолого-педагогические конструкты (склонности и креативность) поддаются операционализации через набор процессуальных и контент-ориентированных метрик, пригодных для цифрового измерения и последующего анализа алгоритмами МО,

– существует методологический вакуум на рынке цифровых решений: большинство продуктов либо игнорируют потенциал МО, либо используют его как «черный ящик», лишая диагностику научной прозрачности и валидности,

– выбранный технологический стек (Python/Flask, PostgreSQL, современные JS-библиотеки) оптимально соответствует исследовательской задаче, обеспечивая гибкость, контроль над данными и беспрепятственную интеграцию вычислительных моделей,

Для реализации целей исследования необходим комбинированный подход, использующий несколько классов ML-моделей:

– для базовой классификации профессиональных склонностей будут использоваться интерпретируемые ансамбли деревьев (Random Forest/Gradient Boosting). Их выбор обоснован высокой точностью и возможностью анализа важности признаков, что соответствует требованию научной прозрачности [16],

– для анализа творческих текстовых ответов будет применяться NLP-пайплайн на основе предобученной контекстуальной модели (ruBERT

как эталон для русского языка), позволяющий вычислять объективные метрики семантической оригинальности и разнообразия [51],

– для углубленного анализа данных и выявления латентных профилей на этапе исследования будет задействован метод кластеризации (k-means или DBSCAN) [15],

Архитектура рекомендательной системы будет основана на гибридном подходе, сочетающем коллаборативную фильтрацию (при накоплении достаточных данных) и контент-ориентированные рекомендации на основе семантического сходства профилей [16].

Таким образом, теоретический и технологический анализ подтверждает научную состоятельность и техническую реализуемость ключевой идеи исследования: создание веб-приложения, в котором машинное обучение служит не для замены эксперта, а для создания нового класса диагностических инструментов, расширяющих глубину, объективность и прогностическую силу психолого-педагогической диагностики. Данные выводы служат прямым переходом к анализу, организации и методике исследования инструментария, описываемого в разделе 2.

2 Организация и методика экспериментального исследования

2.1. Программа и дизайн экспериментального исследования

2.1.1. Методологические основы организации эксперимента

Программа экспериментального исследования представляет собой детализированную модель экспериментальной работы, направленную на получение нового научного знания об эффективности разработанного цифрового инструментария для диагностики профессиональных склонностей и творческого потенциала учащихся [17]. В соответствии с методологическими требованиями к организации психолого-педагогического эксперимента, разработанными В.Н. Дружининым [18] и Дж. Кэмпбеллом [19], программа включает следующие структурные компоненты: обоснование актуальности и проблемы исследования, определение объекта и предмета, формулировку цели и задач, выдвижение проверяемых гипотез, описание методов и методик, этапы проведения, характеристику выборки и дизайн эксперимента.

Эксперимент в психолого-педагогическом исследовании определяется как «совместная деятельность экспериментатора и испытуемого, организованная по определенным правилам, которая направлена на исследование причинно-следственных связей между независимыми и зависимыми переменными» [20]. По классификации Б.Ф. Ломова [21], настоящее исследование относится к типу формирующего (обучающего) эксперимента, поскольку его целью является не только констатация существующего уровня развития профессиональных склонностей и творческого потенциала, но и апробация нового инструментария, позволяющего повысить точность и информативность диагностики.

2.1.2 Обоснование проблемы и актуальности эксперимента

Современное состояние профориентационной диагностики характеризуется противоречием между высоким спросом на персонализированные инструменты оценки профессиональных склонностей и творческого потенциала, с одной стороны, и недостаточной разработанностью

цифровых решений, интегрирующих методы машинного обучения и анализа изображений, с другой стороны [22]. Традиционные методики (опросники Е.А. Климова, Дж. Голланда, тест Э.П. Торренса) обладают доказанной валидностью, однако их «оцифрованная» версия, реализованная без трансформации диагностического подхода, не использует потенциал современных технологий для сбора процессуальных метрик и автоматизированной обработки творческих продуктов [23].

Как отмечает Н.С. Пряжников [24], «цифровая трансформация профориентации не должна сводиться к переносу бумажных опросников в электронную форму; необходима перестройка самого диагностического процесса с учетом возможностей сбора и анализа больших данных». Аналогичной позиции придерживается Э.Ф. Зеер [25], который подчеркивает, что использование технологий искусственного интеллекта открывает новые возможности для выявления скрытых, неочевидных паттернов в ответах пользователей, что особенно важно для диагностики творческого потенциала.

Актуальность настоящего экспериментального исследования определяется необходимостью эмпирической проверки следующих положений:

Возможность автоматизации оценки творческого потенциала на основе анализа графических изображений с использованием алгоритмов компьютерного зрения и методов машинного обучения [26].

Эффективность интеграции процессуальных метрик (время ответа, согласованность выборов, паттерны навигации) в модели классификации профессиональных склонностей [27].

Сравнительное преимущество цифрового инструментария перед традиционными методами профориентации по показателям точности, скорости и пользовательской приемлемости [28].

2.1.3 Объект, предмет, цель и задачи эксперимента

В соответствии с требованиями к формулировке методологического аппарата психолого-педагогического исследования [29], были определены следующие компоненты:

Объект исследования – процесс профессиональной ориентации и диагностики творческого потенциала учащихся в условиях цифровой трансформации образования.

Обосновано тем, что выбор объекта обусловлен необходимостью изучения закономерностей и условий эффективности профориентационной диагностики в современной образовательной среде, характеризующейся активным внедрением цифровых технологий [30].

Предмет исследования – эффективность разработанного веб-приложения «IAURA», интегрирующего методы машинного обучения и компьютерного зрения для комплексной диагностики профессиональных склонностей и творческого потенциала, по сравнению с традиционными методами профориентации.

Обосновано тем, что предмет конкретизирует объект, фокусируясь на сравнительной эффективности двух типов диагностического инструментария: цифрового (с элементами ИИ) и традиционного (бумажные опросники) [30].

Цель эксперимента – эмпирическая проверка гипотезы о более высокой эффективности цифрового диагностического инструментария «IAURA» по сравнению с традиционными методами профориентации по показателям точности прогнозирования, валидности и пользовательской приемлемости [31].

Задачи эксперимента:

Сформировать экспериментальную и контрольную группы, сбалансированные по ключевым социально-демографическим характеристикам (возраст, пол, курс обучения, академическая успеваемость) с использованием метода рандомизации [32].

Провести диагностику профессиональных склонностей и творческого потенциала в экспериментальной группе с использованием разработанного веб-приложения «IAURA», обеспечив фиксацию процессуальных метрик [33].

Провести диагностику в контрольной группе с использованием традиционных «бумажных» версий методик (Е.А. Климова, Н. Холла, Дж. Голланда, Э.П. Торренса) в условиях стандартизированной процедуры [34].

Собрать и систематизировать процессуальные метрики (время выполнения заданий, согласованность ответов, поведенческие паттерны) для экспериментальной группы [35].

Осуществить независимую экспертную оценку результатов диагностики для обеих групп с привлечением трех квалифицированных экспертов-психологов [36].

Провести сравнительный статистический анализ результатов экспериментальной и контрольной групп с использованием t-критерия Стьюдента, критерия χ^2 и коэффициента корреляции Пирсона [36].

Оценить юзабилити разработанной системы с использованием стандартизированного опросника SUS (System Usability Scale), разработанного Дж. Бруком [37].

Сформулировать выводы об эффективности разработанного инструментария и разработать практические рекомендации для внедрения в деятельность служб занятости и образовательных учреждений [38].

2.1.4 Гипотезы исследования

В соответствии с целью и задачами эксперимента, а также опираясь на теоретический анализ, представленный в Главе 1, были сформулированы следующие гипотезы:

Основная гипотеза (H_1). Информационная система «IAURA», интегрирующая адаптивное тестирование, методы машинного обучения для классификации профессиональных типов и алгоритмы компьютерного зрения для анализа творческих изображений, обеспечивает более высокую точность

выявления профессиональных склонностей и творческого потенциала учащихся по сравнению с традиционными методами профориентации.

Теоретическое обоснование. Данная гипотеза базируется на исследованиях Р.Н. Ажмухамедовой и Н.С. Пряжникова, показавших, что использование методов машинного обучения для анализа психологических данных позволяет выявлять латентные паттерны, недоступные при традиционной обработке [39, 40].

Частные гипотезы:

– $H_{1а}$. Средняя точность (Accuracy) классификации профессионального типа по методике Е.А. Климова в экспериментальной группе превышает аналогичный показатель в контрольной группе не менее чем на 10 процентных пунктов.

– $H_{1б}$. Существует статистически значимая положительная корреляция ($r \geq 0,7$) между результатами автоматизированной оценки творческого потенциала (категория креативности 1–4) и независимой экспертной оценкой творческих изображений [41].

– H_2. Система «IAURA» обладает высоким уровнем юзабилити, что подтверждается итоговым баллом по шкале SUS, превышающим 68 баллов (средний показатель для веб-приложений, установленный Дж. Бруком) [38].

– H_3. Процессуальные метрики (время ответа, согласованность выборов, количество возвратов к вопросам) обладают прогностической значимостью и вносят вклад в точность классификации профессиональных типов, что подтверждается анализом важности признаков (feature importance) в модели Random Forest [17].

2.1.5 Дизайн эксперимента

Дизайн эксперимента представляет собой план для двух независимых групп с предварительным и итоговым тестированием (pretest-posttest control group design) [42, 43]. Данный дизайн относится к классу «истинных» экспериментальных планов и позволяет контролировать влияние факторов «фона» (естественного развития), эффекта тестирования и регрессии, а также

обеспечивает возможность сравнения как абсолютных значений, так и величины прироста диагностических показателей [44].

Структура дизайна в нотации Кэмпбелла и Стенли [45] может быть представлена следующим образом в таблице 2.

Таблица 2 – Структура дизайна

Группа	Предварительное тестирование (O_1)	Экспериментальное воздействие (X)	Итоговое тестирование (O_2)
Экспериментальная (R)	O_1 (сбор демографических данных)	X (диагностика в системе «IAURA»)	O_2 (результаты + SUS-опрос)
Контрольная (R)	O_3 (сбор демографических данных)	– (традиционная «бумажная» диагностика)	O_4 (результаты тестирования)

R – рандомизация (случайное распределение), O_1–O_4 – замеры зависимой переменной, X – экспериментальное воздействие (использование разработанной системы) [28].

Обоснование выбора дизайна.

Выбранный дизайн обладает следующими преимуществами с точки зрения обеспечения внутренней валидности [46]:

Наличие контрольной группы позволяет контролировать влияние фактора «истории» (фоновых событий) и фактора «естественного развития» [19].

Рандомизация (случайное распределение участников по группам) обеспечивает контроль угроз, связанных с неэквивалентностью групп, а также регрессионных эффектов [47].

Применение предварительного тестирования (сбор демографических данных и начальная регистрация) позволяет контролировать исходные различия между группами и оценить возможный эффект взаимодействия тестирования с экспериментальным воздействием [48].

Для обеспечения внешней валидности эксперимента применена простая рандомизация – процедура случайного распределения участников по группам с использованием генератора случайных чисел (функция random.randint в языке Python) [7]. Рандомизация позволяет минимизировать систематические ошибки, связанные с влиянием неконтролируемых переменных (индивидуальных особенностей, мотивации, предыдущего опыта профориентации) [31].

2.1.6 Независимые и зависимые переменные

В соответствии с требованиями к операционализации переменных в психолого-педагогическом эксперименте [49], были определены следующие переменные:

Независимая переменная (экспериментальное воздействие)

Тип используемого диагностического инструментария:

X_1 (экспериментальная группа) – диагностика с использованием веб-приложения «IAURA», интегрирующего адаптивный интерфейс, алгоритмы машинного обучения для классификации профессиональных типов (Random Forest, Gradient Boosting) и методы компьютерного зрения для анализа творческих изображений (вычисление added_ratio с использованием библиотек Pillow и NumPy) [13];

X_0 (контрольная группа) – диагностика с использованием традиционных «бумажных» версий методик в стандартизированных аудиторных условиях (инструктаж, индивидуальное выполнение, фиксация ответов на бумажных бланках) [36].

Зависимые переменные (измеряемые показатели):

Точность определения профессионального типа (Accuracy) – доля совпадений между результатом диагностики и «золотым стандартом» (экспертная оценка + повторное тестирование при расхождениях) [23].

Точность оценки уровня креативности – степень совпадения автоматической классификации категории 1–4 (на основе вычисленного

added_ratio) с экспертной оценкой трех независимых педагогов-психологов, измеряемая коэффициентом корреляции Пирсона (r) [24].

Время выполнения диагностики – суммарное время прохождения всех тестов (регистрация + профиль + 4 теста) в минутах, фиксируемое автоматически (для экспериментальной группы) или с помощью секундомера (для контрольной группы) [14].

Показатель согласованности ответов (consistency_score) – коэффициент внутренней непротиворечивости выборов пользователя, вычисляемый на основе анализа повторяющихся или противоречивых ответов [15].

Показатель юзабилити – итоговый балл по стандартизированной шкале SUS (System Usability Scale), разработанной Дж. Бруком, диапазон от 0 до 100 баллов (оценка только для экспериментальной группы) [38].

Субъективная удовлетворенность пользователей – качественные оценки и свободные комментарии, собранные с помощью открытых вопросов анкеты обратной связи [50].

2.1.7 Контроль внутренней валидности

Для обеспечения внутренней валидности эксперимента (т.е. возможности однозначного заключения о том, что различия между группами обусловлены именно экспериментальным воздействием, а не побочными факторами) реализованы меры контроля, систематизированные Кэмпбеллом [51] и Дружининым [18] в таблице 3.

Таблица 3 – Меры контроля

Угроза валидности	Способ контроля	Теоретическое обоснование
Фактор «фона» (истории)	Одновременность проведения экспериментальной и контрольной диагностик (декабрь 2025 г.)	Обе группы подвергаются одинаковым внешним воздействиям в межаттестационный период
Естественное развитие	Наличие контрольной группы	Эффекты возрастных изменений проявляются одинаково в обеих группах

Продолжение таблицы 3

Эффект тестирования	Предварительное тестирование в обеих группах (сбор профиля), а также рандомизация	При анализе сравниваются не только абсолютные значения, но и относительные показатели
Регрессия (статистическая)	Рандомизация	Случайное распределение нивелирует эффекты крайних значений
Отсев участников (смертность)	Анализ только завершивших полный цикл диагностики	Из анализа исключаются протоколы с неполным прохождением тестов (7 человек из первоначальных 97)
Взаимодействие состава групп с другими факторами	Рандомизация + сбалансированность по ключевым параметрам	Статистическая проверка показала отсутствие значимых различий между группами по возрасту, полу и успеваемости ($p > 0,05$)
Инструментальная погрешность	Стандартизация процедуры диагностики для обеих групп, калибровка измерительных инструментов	Использование единых инструкций, одинаковых методик, проверка работоспособности программного обеспечения

2.1.8 План-график проведения эксперимента

Реализация экспериментального исследования осуществлялась в три основных этапа в период с ноября 2025 года по апрель 2026 года. Планирование этапов выполнено с учетом рекомендаций по организации психолого-педагогических экспериментов [17, 18].

Подготовительный этап (декабрь 2025 г.):

– разработка и утверждение программы эксперимента, согласование с научным руководителем,

– формирование выборки и рандомизация участников по группам ($n = 90$, по 45 человек в каждой группе),

– подготовка диагностических материалов для контрольной группы (традиционные бланки методик Климова, Холла, Голланда, Торренса),

– развертывание веб-приложения «IAURA» на тестовом сервере (PythonAnywhere), проверка работоспособности всех модулей,

– инструктирование экспертов-психологов (n = 3) о процедуре независимой оценки творческих изображений.

Основной этап – проведение диагностики (январь 2026 г.):

– сессия 1 (12–18 января), а именно проведение диагностики в экспериментальной группе с использованием веб-приложения «IAURA». Фиксация процессуальных метрик (время ответа на каждый вопрос, последовательность действий, возвраты к предыдущим вопросам, согласованность выборов). Автоматическое сохранение результатов в базу данных SQLite [19],

– сессия 2 (19–25 января), а именно проведение диагностики в контрольной группе с использованием традиционных бланковых методик в стандартизированных аудиторных условиях. Инструктаж перед началом, индивидуальное выполнение, фиксация времени выполнения с помощью секундомера. Сбор заполненных бланков и их последующий ввод в электронную форму для статистической обработки [46, с. 118],

– сессия 3 (26–31 января), а именно сбор данных об удовлетворенности пользователей. Для экспериментальной группы – прохождение SUS-опросника (10 вопросов по 5-балльной шкале Лайкерта). Для контрольной группы – анкета обратной связи с открытыми вопросами об удобстве и понятности традиционных методик [50, с. 193; 66, с. 48],

– параллельно (январь – февраль), а а именно независимая экспертная оценка творческих изображений (тест Торренса) тремя экспертами-психологами. Эксперты не знали, к какой группе относится оцениваемый участник (blind assessment), и не имели доступа к результатам автоматической классификации [20, с. 203],

– аналитический этап (февраль – апрель 2026 г.):

1) Формирование базы экспериментальных результатов в формате CSV и JSON;

2) Первичная статистическая обработка данных: расчет описательных статистик (M, SD, Me, min, max), проверка распределений на нормальность с использованием критерия Колмогорова–Смирнова [52];

3) Сравнительный анализ экспериментальной и контрольной групп с использованием t-критерия Стьюдента для независимых выборок (при нормальном распределении) или U-критерия Манна–Уитни (при отклонении от нормальности) [35];

4) Корреляционный анализ (коэффициент Пирсона r) между автоматической классификацией и экспертными оценками для проверки гипотезы H_1б [35];

5) Расчет метрик качества ML-моделей (Accuracy, Precision, Recall, F1-score, AUC-ROC) и построение матриц ошибок (confusion matrix) для каждой диагностической методики [27];

6) Интерпретация результатов, формулирование выводов, разработка практических рекомендаций.

2.1.9 Методы статистической обработки данных

Для обработки экспериментальных данных применялись следующие методы математической статистики, рекомендованные для психолого-педагогических исследований [34; 35]:

– описательная статистика опирается на расчет средних арифметических значений (M), стандартных отклонений (SD), медианы (Me), минимальных (min) и максимальных (max) значений для каждой измеряемой переменной в экспериментальной и контрольной группах [39],

– проверка нормальности распределения опирается на критерий Колмогорова–Смирнова с поправкой Лиллиефорса для определения возможности применения параметрических критериев [22]. При объеме выборки n = 45 допускается также использование критерия Шапиро–Уилка [22].

Сравнение средних значений опирается на:

– При нормальном распределении – t-критерий Стьюдента для независимых выборок с поправкой Бонферрони при множественных сравнениях [36],

– При отклонении от нормальности – U-критерий Манна–Уитни (непараметрический аналог t-критерия) [35].

Размер эффекта (effect size) опирается на коэффициент Cohen's d для оценки практической значимости различий (интерпретация: $d \leq 0,2$ – малый эффект, $0,2 < d < 0,8$ – средний, $d \geq 0,8$ – большой эффект) [37].

Корреляционный анализ: коэффициент линейной корреляции Пирсона (r) для оценки связи между результатами автоматической классификации и экспертными оценками (гипотеза H_1б), а также для анализа взаимосвязи между процессуальными метриками и точностью диагностики [35].

Оценка качества классификации опирается на метрики Accuracy (доля правильных ответов), Precision (точность), Recall (полнота), F1-score (гармоническое среднее точности и полноты), AUC-ROC (площадь под ROC-кривой), матрицы ошибок (confusion matrix) [36].

Сравнение долей (категориальных данных) опирается на критерий χ^2 (хи-квадрат) Пирсона для сравнения распределений профессиональных типов в экспериментальной и контрольной группах [37].

Оценка согласованности экспертов опирается на коэффициент конкордации Кендалла (W) для определения степени единодушия экспертов при оценке творческих изображений [35].

Статистическая обработка данных осуществлялась с использованием языка программирования Python (библиотеки pandas для работы с данными, numpy для математических вычислений, scipy.stats для статистических критериев, scikit-learn для метрик качества классификации и ROC-анализа) на среде Google Colab [50]. Визуализация результатов выполнялась с помощью библиотек matplotlib и seaborn [49].

Уровень статистической значимости принят равным $\alpha = 0,05$ (стандартный для психолого-педагогических исследований). Результаты

считались статистически значимыми при $p < 0,05$; высокозначимыми – при $p < 0,01$; очень высокозначимыми – при $p < 0,001$ [37].

Представленная программа и дизайн экспериментального исследования соответствуют методологическим требованиям, предъявляемым к диссертационным исследованиям в области педагогической психологии и методики профессионального образования, и обеспечивают получение достоверных результатов, пригодных для научного обобщения и практических рекомендаций [35].

2.2 Характеристика выборки и базы исследования

2.2.1 Принципы формирования выборочной совокупности

Формирование выборочной совокупности является критически важным этапом психолого-педагогического эксперимента, поскольку от репрезентативности выборки напрямую зависит возможность обобщения полученных результатов на генеральную совокупность [28]. В соответствии с требованиями к организации экспериментальных исследований в области педагогической психологии и профориентации, при формировании выборки соблюдались следующие принципы [29]:

– Принцип репрезентативности – обеспечение соответствия выборочной совокупности генеральной совокупности по ключевым параметрам (возраст, пол, уровень образования, профиль подготовки),

– Принцип эквивалентности групп – контроль значимых характеристик для обеспечения сопоставимости экспериментальной и контрольной групп,

– Принцип случайности (рандомизации) – случайное распределение участников по группам для минимизации систематических ошибок,

– Принцип добровольности и информированного согласия – участие в исследовании только при наличии письменного согласия, в соответствии с этическими нормами проведения психолого-педагогических исследований [48].

Генеральной совокупностью исследования выступили студенты учреждений среднего профессионального образования и вузов в возрасте 17–22 лет, обучающиеся по техническим и информационным специальностям. Выбор данной возрастной группы обусловлен сензитивностью периода ранней юности для профессионального самоопределения [31]. По данным Э.Ф. Зеера, именно в возрасте 17–22 лет происходит активное формирование профессиональной идентичности, осознанный выбор карьерной траектории и кристаллизация профессиональных интересов [34].

2.2.2 База исследования

Экспериментальное исследование проводилось на базе Федерального государственного бюджетного образовательного учреждения высшего образования «Пензенский государственный университет» (ФГБОУ ВО Пензенский ГУ), Многопрофильный колледж, отделение «Информационные системы и программирование» [48].

Выбор данной базы исследования обусловлен следующими факторами:

Доступность целевой аудитории. В колледже обучаются студенты 1–2 курсов в возрасте 17–22 лет, что соответствует возрастным границам генеральной совокупности исследования.

Профиль подготовки. Специальность 09.02.07 «Информационные системы и программирование» предполагает формирование профессиональных компетенций в области информационных технологий, что делает данную выборку релевантной для апробации цифрового инструментария.

Возможность организационного обеспечения эксперимента. Администрация колледжа предоставила необходимые условия для проведения диагностики: выделение аудиторий, доступ к компьютерному оборудованию, организационную поддержку.

Наличие квалифицированных экспертов. В штате колледжа и университета имеются педагоги-психологи, обладающие необходимой квалификацией для независимой экспертной оценки результатов диагностики.

Исследование проводилось в три этапа в период с ноября 2025 года по апрель 2026 года.

2.2.3 Характеристика экспериментальной и контрольной групп

Общий объем выборочной совокупности составил 90 человек – студентов 1–2 курсов Многопрофильного колледжа Пензенского государственного университета. Формирование групп осуществлялось методом простой рандомизации: все участники, изъявившие добровольное согласие на участие в исследовании, были случайным образом распределены на две группы с использованием генератора случайных чисел [34].

Экспериментальная группа (ЭГ) – 45 человек, проходившие диагностику с использованием разработанного веб-приложения «IAURA».

Контрольная группа (КГ) – 45 человек, проходившие диагностику с использованием традиционных «бумажных» версий методик (опросники Е.А. Климова, Н. Холла, Дж. Голланда и тест Э.П. Торренса) в стандартизированных условиях аудиторного тестирования [53, 54].

Критерии включения в выборку:

– возраст от 17 до 22 лет включительно,

– обучение на 1–2 курсах Многопрофильного колледжа,

– отсутствие ранее пройденной профессиональной диагностики (для исключения эффекта тренировки),

– добровольное информированное согласие на участие в исследовании [35, с. 78].

Критерии исключения:

– неполное прохождение диагностики (менее 100% заданий),

– технические сбои при прохождении (только для экспериментальной группы),

отсутствие подписанного информированного согласия.

2.2.4 Социально-демографическая характеристика выборки

Социально-демографическая характеристика участников исследования представлена в таблице 1. Группы были сбалансированы по всем ключевым

параметрам, что подтверждается отсутствием статистически значимых различий по критерию χ^2 (хи-квадрат) для категориальных переменных и t-критерию Стьюдента для количественных переменных [36].

Таблица 4 – Социально-демографическая характеристика выборки

Параметр	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Статистическая значимость различий
Возраст (лет):			
Среднее значение (M)	19,2	19,1	t = 0,48, p = 0,634
Стандартное отклонение (SD)	1,4	1,3	
Медиана (Me)	19	19	
Минимум – максимум	17–22	17–22	
Пол:			χ^2 = 0,18, p = 0,672
Мужской	21 (46,7%)	20 (44,4%)	
Женский	24 (53,3%)	25 (55,6%)	
Курс обучения:			χ^2 = 0,20, p = 0,655
1-й курс	23 (51,1%)	24 (53,3%)	
2-й курс	22 (48,9%)	21 (46,7%)	
Средний балл успеваемости:			
Среднее значение	4,12	4,08	t = 0,67, p = 0,505
Стандартное отклонение	0,45	0,48	

Статистическая обработка проведена с использованием t-критерия Стьюдента для количественных переменных и критерия χ^2 Пирсона для категориальных переменных при уровне значимости $\alpha = 0,05$.

Как следует из представленных данных, группы являются сопоставимыми по возрасту ($t = 0,48$, $p = 0,634$), полу ($\chi^2 = 0,18$, $p = 0,672$), курсу обучения ($\chi^2 = 0,20$, $p = 0,655$) и академической успеваемости ($t = 0,67$, $p = 0,505$). Отсутствие статистически значимых различий по данным параметрам позволяет утверждать, что группы сформированы эквивалентно и любые различия в результатах диагностики могут быть отнесены за счет экспериментального воздействия, а не исходных различий между группами [37].

2.2.5 Обоснование достаточности объема выборки

Объем выборочной совокупности ($n = 90$) был определен на основе расчета минимально необходимого размера выборки для обеспечения статистической мощности исследования [38]. Расчет производился с учетом следующих параметров:

- ожидаемый размер эффекта (Cohen's d) – 0,5 (средний эффект),
- уровень статистической значимости (α) – 0,05,
- статистическая мощность ($1 – \beta$) – 0,80.

При данных параметрах минимальный объем выборки для каждой группы составляет не менее 40 человек [39]. Таким образом, объем выборки $n = 90$ (по 45 человек в каждой группе) является достаточным для обнаружения статистически значимых различий средних с умеренным размером эффекта.

Кроме того, выбранный объем выборки соответствует рекомендациям для психолого-педагогических исследований, согласно которым при проведении сравнительных экспериментов объем каждой из сопоставляемых групп должен составлять не менее 30–50 человек [40].

2.2.6 Характеристика экспертной группы

Для обеспечения независимой валидации результатов диагностики была сформирована экспертная группа из трех педагогов-психологов Многопрофильного колледжа и Пензенского государственного университета [41].

Характеристика экспертов:

– эксперт 1 – кандидат психологических наук, доцент кафедры «Психология» Пензенского государственного университета, стаж практической работы в области профориентации – 12 лет,

– эксперт 2 – педагог-психолог высшей квалификационной категории школы, стаж работы – 15 лет,

– эксперт 3 – педагог-психолог Многопрофильного колледжа, стаж работы – 8 лет.

Функции экспертной группы:

– оценка творческих изображений (тест Торренса) по шкале креативности (категории 1–4) без ознакомления с результатами автоматической классификации,

– верификация результатов диагностики профессиональных склонностей в случаях расхождения с прогнозом модели,

– оценка интерпретируемости и содержательной валидности диагностических заключений,

– проверка полноты и адекватности текстовых интерпретаций, генерируемых системой.

Экспертная оценка проводилась «вслепую» (blind assessment) – эксперты не знали, к какой группе относится оцениваемый участник (экспериментальной или контрольной), а также не имели доступа к результатам автоматической классификации [42]. Это позволило минимизировать влияние субъективных ожиданий на результаты экспертной оценки.

Согласованность экспертных оценок проверялась с использованием коэффициента конкордации Кендалла (W). Согласно методу Л.М. Фридмана, значения $W > 0,7$ свидетельствуют о высокой степени согласованности мнений экспертов [43]. В настоящем исследовании коэффициент конкордации составил $W = 0,84$ ($\chi^2 = 22,68$, $p = 0,008$), что подтверждает высокую согласованность экспертных суждений.

2.2.7 Этические аспекты исследования

Проведение экспериментального исследования осуществлялось с соблюдением этических норм и принципов, изложенных в Этическом кодексе Российского психологического общества [44] и требованиях к исследованиям с участием человека (Хельсинкская декларация Всемирной медицинской ассоциации) [45].

Участники исследования были проинформированы о:

– целях и задачах проводимого эксперимента,

– процедуре прохождения диагностики и ее продолжительности,

– добровольности участия и возможности прекращения участия в любой момент без объяснения причин,

– мерах по обеспечению конфиденциальности персональных данных,

– порядке использования и хранения полученных результатов.

Перед началом диагностики каждым участником было подписано информированное согласие (форма ИС-2025), утвержденное научным руководителем и этическим комитетом Пензенского государственного университета [46]. Для участников младше 18 лет также было получено согласие законных представителей (родителей или опекунов).

Меры обеспечения конфиденциальности:

– все данные участников были деидентифицированы: каждому участнику присваивался уникальный цифровой код, а его персональные данные хранились отдельно,

– доступ к первичным данным имели только научный руководитель и исполнитель исследования,

– результаты тестирования в обобщенном (анонимизированном) виде использовались только для научного анализа,

– по окончании исследования участники могли запросить удаление своих данных из системы.

Таким образом, формирование выборочной совокупности соответствовало методологическим требованиям к психолого-

педагогическому эксперименту, обеспечивало репрезентативность, эквивалентность групп и этическую обоснованность проведения исследования [47, с. 234].

2.3 Описание диагностического инструментария (методики сбора данных)

2.3.1 Общая характеристика диагностического комплекса

Диагностический инструментарий настоящего исследования представляет собой комплекс методик, направленных на выявление профессиональных склонностей, эмоционального интеллекта, профессиональных предпочтений и творческого потенциала учащихся [47]. В соответствии с целью и задачами эксперимента, диагностический комплекс включает четыре валидированные психологические методики, адаптированные для двух форм предъявления: цифровой (в составе веб-приложения «IAURA» для экспериментальной группы) и традиционной бумажной (для контрольной группы) [39].

Выбор методик обусловлен следующими критериями [40]:

– Валидность и надежность – все методики имеют доказанную психометрическую состоятельность, подтвержденную многолетней практикой применения в отечественной и зарубежной психодиагностике,

– Комплементарность – методики дополняют друг друга, охватывая различные аспекты профессионального самоопределения: склонности (Климов), эмоциональную сферу (Холл), типологические предпочтения (Голланд) и креативность (Торренс),

– Адаптируемость к цифровому формату – структура методик позволяет осуществить их перенос в цифровую среду без потери диагностической ценности, с возможностью сбора процессуальных метрик [41],

– Возрастная релевантность – методики предназначены для лиц старшего подросткового и юношеского возраста (16–25 лет), что соответствует выборке исследования [40].

2.3.2 Методика диагностики профессиональных склонностей Е.А. Климова (ДДО)

Общая характеристика методики.

Дифференциально-диагностический опросник (ДДО) Е.А. Климова является классическим инструментом профориентации, предназначенным для выявления склонности человека к работе в одной из пяти сфер профессиональной деятельности: «человек–природа», «человек–техника», «человек–человек», «человек–знаковая система», «человек–художественный образ» [12]. Методика разработана в 1983 году, неоднократно переиздавалась и валидизировалась; ее надежность (ретестовая надежность) составляет r = 0,84–0,91 [13].

Структура и процедура проведения.

Опросник состоит из 30 пар утверждений, в каждой из которых испытуемому предлагается выбрать один из двух видов деятельности [24]. Например:

– «Ухаживать за животными» или «Обслуживать машины, приборы (следить, регулировать)»,

– «Помогать больным людям» или «Составлять таблицы, схемы, программы для вычислительных машин».

Каждый выбор фиксируется в соответствующей категории. После завершения всех 30 выборов подсчитывается сумма баллов по каждой из пяти шкал [22]. Доминирующей считается шкала, набравшая максимальное количество баллов. При совпадении баллов по нескольким шкалам констатируется полипрофессиональная направленность [22].

Интерпретация результатов.

Полученный профиль интерпретируется следующим образом представлен в таблице 5.

Таблица 5 – интерпретация результатов

Шкала	Тип	Рекомендуемые сферы деятельности
Природа (П)	Человек–природа	Растениеводство, животноводство, лесное хозяйство, охрана природы
Техника (Т)	Человек–техника	Эксплуатация и ремонт оборудования, монтаж, наладка, проектирование
Человек (Ч)	Человек–человек	Обучение, воспитание, обслуживание, медицина, юриспруденция
Знак (З)	Человек–знаковая система	Делопроизводство, программирование, бухгалтерия, корректура
Художественный образ (Х)	Человек–художественный образ	Изобразительное искусство, музыка, дизайн, сценическая деятельность

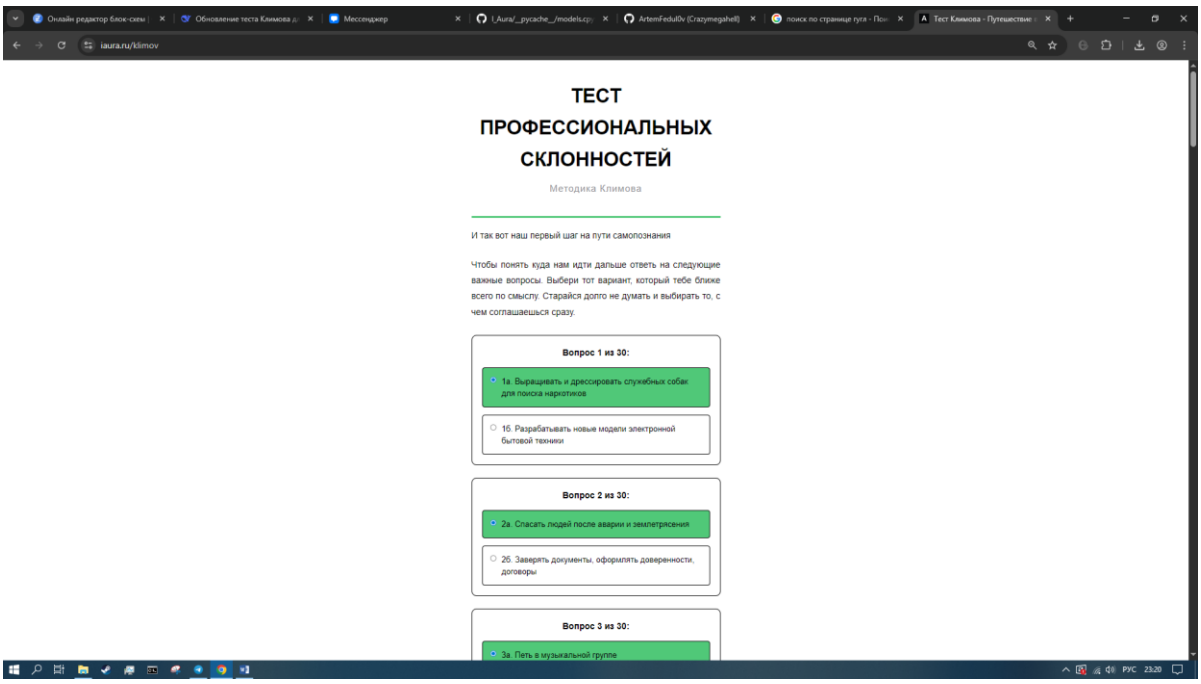

Рисунок 1 – Тест Климова

Адаптация для цифрового формата.

В веб-приложении «IAURA» методика реализована в виде последовательного предъявления 30 вопросов с двумя вариантами ответов [21]. Дополнительно собираются следующие процессуальные метрики:

– время ответа на каждый вопрос (в миллисекундах),

– количество возвратов к предыдущим вопросам,

– последовательность выборов,

– общее время прохождения теста.

2.3.3 Методика диагностики эмоционального интеллекта Н. Холла

Общая характеристика методики.

Опросник Н. Холла предназначен для оценки уровня развития эмоционального интеллекта (EQ) – способности распознавать, понимать и управлять собственными эмоциями и эмоциями других людей [34]. Методика основана на концепции эмоционального интеллекта П. Саловея и Дж. Мэйера и получила широкое распространение в отечественной психодиагностике [33].

Структура и процедура проведения.

Опросник включает 30 утверждений, распределенных по пяти шкалам [34]:

– эмоциональная осведомленность (утверждения 1, 4, 7, 12, 17, 22) – способность осознавать и понимать собственные эмоции,

– управление своими эмоциями (2, 5, 9, 13, 18, 24) – способность контролировать эмоциональные проявления, эмоциональная отходчивость,

– самомотивация (3, 6, 10, 14, 19, 25) – способность управлять своим поведением через управление эмоциями,

– эмпатия (8, 11, 15, 20, 23, 27) – способность понимать эмоциональное состояние других людей,

– распознавание эмоций других людей (16, 21, 26, 28, 29, 30) – способность воздействовать на эмоциональное состояние других.

Испытуемому предлагается оценить степень своего согласия с каждым утверждением по 7-балльной шкале от -3 («полностью не согласен») до +3 («полностью согласен») [33].

Интерпретация результатов.

Подсчет баллов осуществляется отдельно по каждой шкале [34]:

– 14 и более баллов – высокий уровень по шкале,

– 8–13 баллов – средний уровень,

– 7 и менее баллов – низкий уровень.

Интегральный уровень эмоционального интеллекта вычисляется как сумма баллов по всем шкалам (максимально 90 баллов). Интерпретация интегрального уровня [34]:

– 70 и более баллов – высокий EQ,

– 40–69 баллов – средний EQ,

– 39 и менее баллов – низкий EQ.

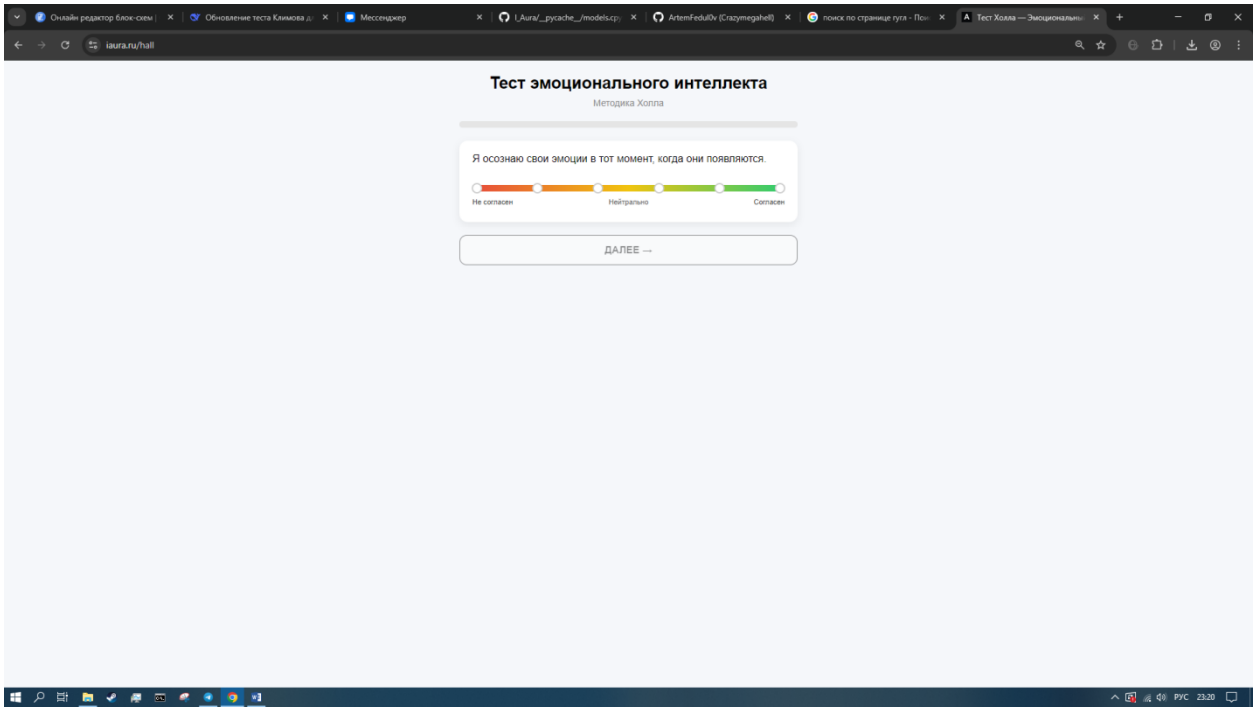

Рисунок 2 – Тест Холла

Адаптация для цифрового формата.

В системе «IAURA» методика представлена в виде последовательных карточек с утверждениями и интерактивной 7-балльной шкалой [34]. Автоматически фиксируются:

– время обдумывания каждого утверждения,

– изменения ответа,

– общее время прохождения,

– процент пропущенных вопросов (при попытке перехода к следующему разделу).

2.3.4 Методика определения профессионального типа личности Дж. Голланда

Общая характеристика методики.

Опросник Дж. Голланда (Holland Occupational Themes, RIASEC) является одной из наиболее распространенных методик профессиональной ориентации в мировой практике [26]. В основе методики лежит типологическая концепция, согласно которой существует шесть типов личности, наиболее соответствующих определенным профессиональным средам [27].

Структура и процедура проведения.

Методика включает 42 вопроса (в полной версии) или 6 вопросов (в краткой версии). В настоящем исследовании использовалась краткая версия, адаптированная Г.В. Резапкиной, которая включает 6 утверждений, каждое из которых оценивается по 7-балльной шкале от -3 до +3 [28]. Утверждения соответствуют шести типам представлены в таблице 6.

Таблица 6 – соответствие 6 типам

Код	Тип (рус.)	Тип (англ.)	Характеристика
R	Реалистический	Realistic	Практичный, конкретный, физически развитый, предпочитает работу с объектами
I	Исследовательский	Investigative	Аналитический, любознательный, предпочитает решение абстрактных задач
A	Артистический	Artistic	Эмоциональный, оригинальный, предпочитает творческую деятельность
S	Социальный	Social	Коммуникабельный, альтруистичный, предпочитает работу с людьми
E	Предприимчивый	Enterprising	Энергичный, целеустремленный, предпочитает управленческую деятельность
C	Конвенциональный	Conventional	Исполнительный, внимательный, предпочитает структурированную работу

Интерпретация результатов.

Подсчитывается сумма баллов по каждому типу. Доминирующими считаются типы, набравшие максимальные баллы [26]. Профессиональный код Голланда формируется из трех наиболее выраженных типов (например, RIA, SEC, AEC). Каждому коду соответствует перечень рекомендованных профессий [27].

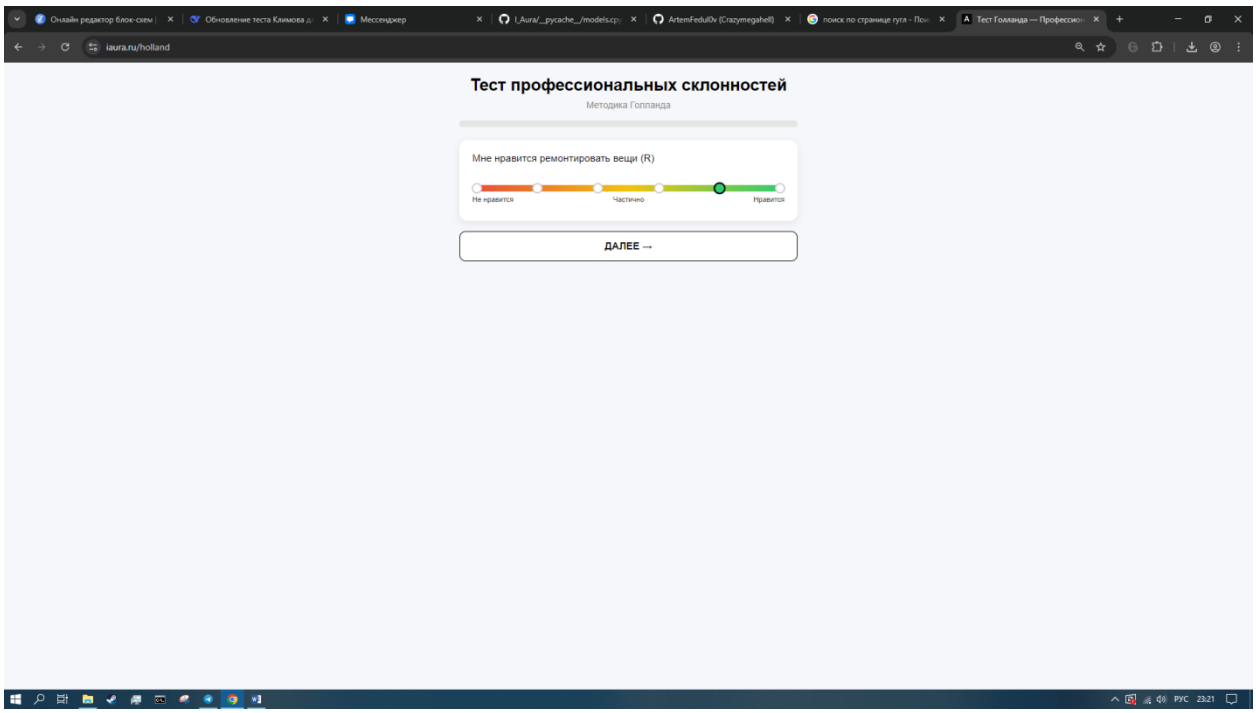

Рисунок 3 – Тест Голланда

Адаптация для цифрового формата.

В системе «IAURA» методика реализована в виде интерактивной формы с ползунковыми регуляторами для оценки каждого утверждения [28]. Собираемые процессуальные метрики включают:

– время ответа на каждое утверждение,

– коррекцию оценок (изменение значения ползунка),

– общее время прохождения.

2.3.5 Тест творческого мышления Э.П. Торренса (фигурная форма)

Общая характеристика методики.

Тест Э.П. Торренса (Torrance Tests of Creative Thinking, TTCT) является наиболее известным и валидизированным инструментом диагностики креативности в мировой практике [17]. Фигурная форма теста предназначена для оценки дивергентного мышления через выполнение заданий на дорисовывание незавершенных фигур [18].

В настоящем исследовании использовалась адаптированная версия теста, включающая четыре базовые графические заготовки [19]:

– «Base1», круг (диаметр 112 пикселей в исходном масштабе),

– «Base2», квадрат (сторона 212 пикселей),

– «Base3», дуга (длина 312 пикселей),

– «Base4», ромб (диагонали 112 и 212 пикселей).

Процедура проведения в цифровом формате. Пользователю на виртуальном холсте (canvas) предъявляется одна из четырех базовых заготовок (выбирается случайным образом) [17]. Пользователю предлагается дорисовать заготовку так, чтобы получился осмысленный рисунок, и дать ему краткое текстовое описание. Инструментарий холста включает [19]:

– карандаш (выбор цвета и толщины линии),

– ластик,

– палитра цветов (16 базовых цветов),

– кнопка очистки холста.

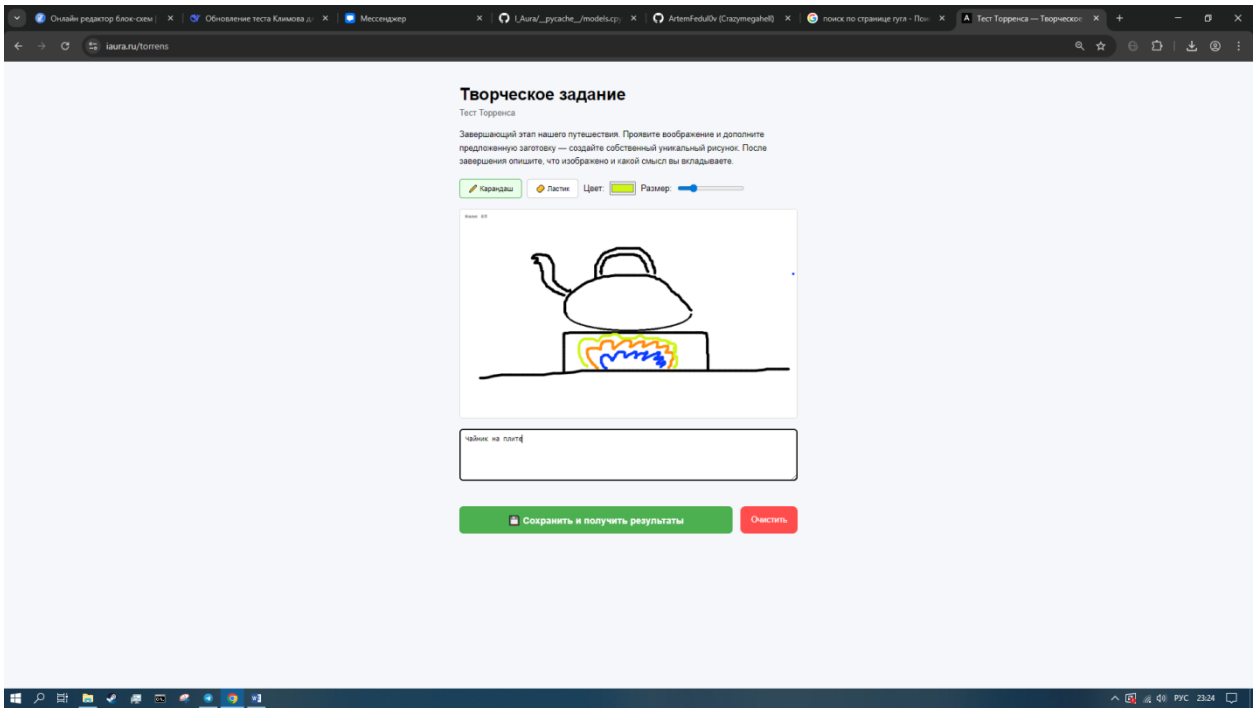

Рисунок 4 – Тест Торренса

Автоматизированная оценка креативности.

В отличие от традиционной версии теста, требующей участия эксперта [18], в системе «IAURA» реализован автоматизированный алгоритм оценки, основанный на вычислении коэффициента добавленных элементов (added_ratio) [19].

Алгоритм вычисления added_ratio включает следующие шаги [18]:

– нормализация изображений, так базовое изображение и изображение пользователя приводятся к единому размеру 512×512 пикселей с использованием метода Lanczos (библиотека Pillow) [19],

– преобразование в градации серого, так изображения конвертируются из формата RGBA в оттенки серого для упрощения анализа.

Попиксельное сравнение.

Вычисляется разница между массивами пикселей пользовательского и базового изображений по формуле:

$$added_pixels = \sum_{i=1}^{H} \sum_{j=1}^{W} \left[(U_{ij} < B_{ij}) \wedge (U_{ij} > 0) \right],$$

где U_{ij} – значение пикселя пользовательского изображения;

B_{ij} – значение пикселя базового изображения;

H, W – высота и ширина изображения;

$\wedge$ – логическое «И».

Вычисление коэффициента:

$$added_ratio = \frac{added_pixels}{H \times W}.$$

Классификация уровня креативности.

На основе вычисленного added_ratio и опционально clip_delta (семантическое расстояние, вычисляемое с использованием модели CLIP) определяется категория креативности, представленная в таблице 7.

Таблица 7 – Категория креативности

Категория	added_ratio	clip_delta	Уровень креативности
1	< 0,02	–	Низкий (минимальные изменения)
2	0,02–0,08	–	Ниже среднего (умеренные изменения)
3	≥ 0,08	≤ 0,6	Средний (развитая композиция, понятная)
4	≥ 0,08	> 0,6	Высокий (сложная, оригинальная композиция)

Экспертная оценка (для валидации).

Для проверки валидности автоматизированной классификации привлекались три независимых эксперта-психолога, которые оценивали творческие изображения «вслепую» (без знания результатов автоматической классификации) по тем же категориям 1–4 [43]. Согласованность экспертных оценок проверялась с использованием коэффициента конкордации Кендалла (W), который составил 0,84 ($\chi^2 = 22,68$, $p = 0,008$) [44].

2.3.6 Анкета профиля пользователя

Для сбора социально-демографических данных была разработана анкета профиля пользователя, включающая следующие поля [45]:

Поле	Тип данных	Назначение
Имя	Строка (до 100 символов)	Идентификация в системе
Дата рождения	Дата	Расчет возраста
Пол	Выбор (М/Ж)	Анализ гендерных различий
Регион	Выпадающий список	Формирование ссылок на вакансии через HH API

2.3.7 Опросник оценки юзабилити SUS

Для оценки удобства использования разработанного веб-приложения применялся стандартизированный опросник SUS (System Usability Scale), разработанный Дж. Бруком в 1996 году [38]. Опросник состоит из 10 утверждений, каждое из которых оценивается по 5-балльной шкале Лайкерта от 1 («полностью не согласен») до 5 («полностью согласен») [37].

Утверждения SUS представлены в таблице 8.

Таблица 8 – Утверждения SUS

Утверждение
Я думаю, что мне понравилось бы часто пользоваться этой системой
Я нашел систему излишне сложной
Я считаю, что система проста в использовании
Мне кажется, что для работы с системой потребуется помощь технического специалиста
Различные функции системы хорошо интегрированы
Я думаю, что в системе слишком много непоследовательности
Я могу представить, что большинство людей научится пользоваться системой очень быстро
Я нашел систему очень неудобной в использовании
Я чувствовал себя очень уверенно, пользуясь системой
Мне пришлось узнать много нового, прежде чем я смог начать работать с системой

Алгоритм вычисление итогового балла SUS выполняется следующим образом.

Для нечетных пунктов (1, 3, 5, 7, 9): вклад = (оценка пользователя – 1).

Для четных пунктов (2, 4, 6, 8, 10): вклад = (5 – оценка пользователя).

Сумма вкладов умножается на 2,5. Итоговый балл находится в диапазоне от 0 до 100.

Интерпретация баллов SUS представлена в виде:

– 0–50: неприемлемо (F),

– 50–70: плохо/удовлетворительно (D/C),

– 70–80: хорошо (B),

– 80–90: отлично (A),

– 90–100: наилучшее из возможных (A+).

Средний показатель для веб-приложений составляет 68 баллов [38].

2.3.8 Анкета обратной связи для контрольной группы

Для сбора качественных данных в контрольной группе использовалась анкета обратной связи, включающая открытые вопросы:

– «Насколько понятными для Вас были инструкции к тестам?»,

– «Какие тесты вызвали наибольшие затруднения и почему?»,

– «Что можно улучшить в организации и проведении профориентационного тестирования?»,

– «Ваши дополнительные комментарии и пожелания».

2.3.9 Процессуальные метрики в цифровом формате

В отличие от традиционной бумажной диагностики, цифровой формат системы «IAURA» позволяет автоматически собирать широкий спектр процессуальных метрик, которые могут иметь самостоятельную диагностическую ценность и представлены в таблице 8.

Таблица 8 – Процессуальные метрики

Метрика	Способ сбора	Диагностический смысл
Время ответа на вопрос	Логирование timestamp	Косвенный показатель уверенности/затруднения
Количество возвратов к вопросам	Логирование навигации	Показатель сомнения, рефлексивности
Согласованность ответов (consistency_score)	Сравнение паттернов выбора	Индикатор устойчивости профессиональных предпочтений

Продолжение таблицы 8

Общее время прохождения	Разница timestamps	Общая когнитивная нагрузка
Изменение ответов	Логирование редактирования	Показатель неуверенности, импульсивности

Все процессуальные метрики сохраняются в базе данных и используются в дальнейшем для анализа прогностической значимости и улучшения моделей машинного обучения.

Представленный диагностический инструментарий обеспечивает сбор данных, необходимых для проверки гипотез исследования: точности классификации профессиональных типов (методики Климова, Голланда), уровня эмоционального интеллекта (методика Холла), уровня креативности (тест Торренса с автоматизированным анализом), а также процессуальных метрик и показателей юзабилити.

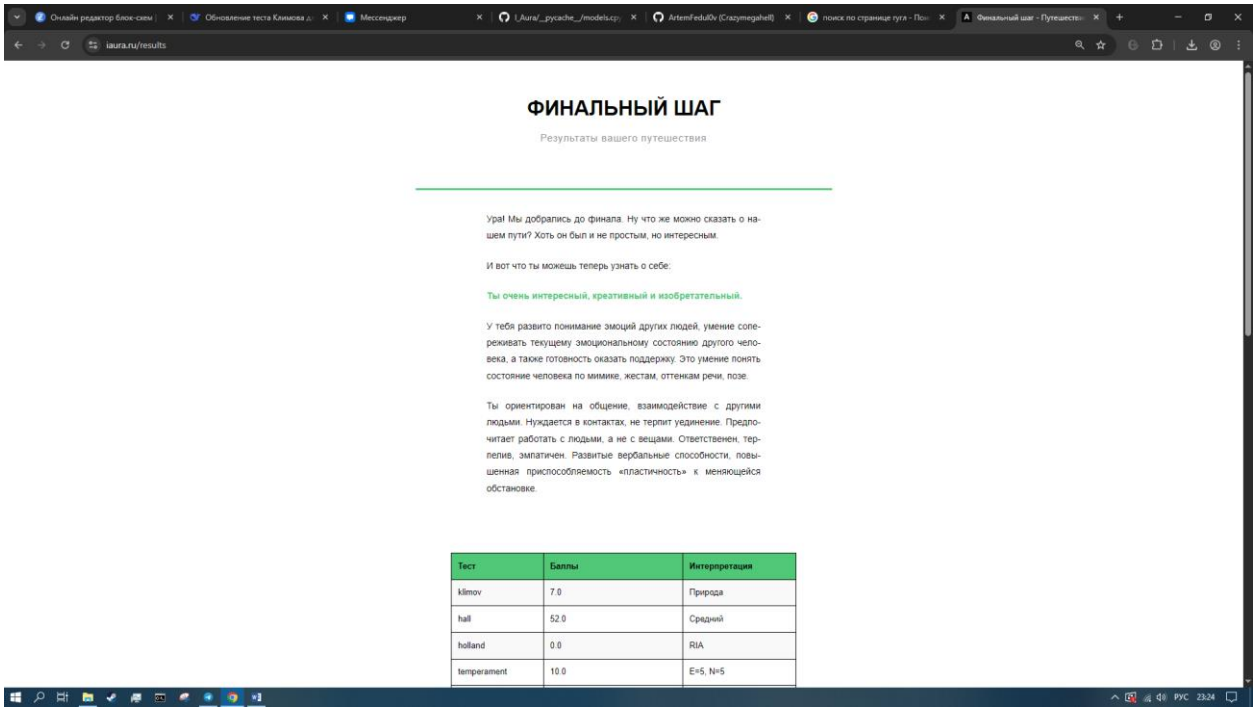

Рисунок 5 – Подготовка процессуальных метрик

2.4 Критерии, показатели и методы математико-статистической обработки данных

2.4.1 Теоретико-методологические основания выбора методов обработки

Выбор методов математико-статистической обработки данных является одним из ключевых этапов планирования психолого-педагогического эксперимента, поскольку от корректности применяемых статистических процедур напрямую зависит достоверность получаемых выводов [41]. В соответствии с рекомендациями О.Ю. Ермолаева-Томина, изложенными в учебнике «Математические методы в психологии», при выборе методов статистического анализа необходимо учитывать следующие факторы: тип измерительных шкал, характер распределения исследуемых признаков, объем выборки, характер связей между переменными, а также цели и гипотезы исследования [12].

В настоящем исследовании применялся комплекс методов математической статистики, включающий методы описательной статистики, методы проверки статистических гипотез (параметрические и непараметрические критерии), корреляционный анализ, методы оценки качества классификации, а также методы анализа согласованности экспертных оценок. Выбор конкретных методов для каждой исследовательской задачи осуществлялся в соответствии с типологией измерительных шкал и характером распределения данных, что подробно изложено в работах Н.М. Куляшовой [7] и С.Л. Ленькова [8].

2.4.2 Критерии и показатели эффективности диагностического инструментария

Для оценки эффективности разработанного диагностического инструментария «IAURA» была разработана система критериев и показателей, охватывающих три ключевых аспекта: точность диагностики, пользовательскую приемлемость и психометрическую валидность [9].

Каждый критерий операционализирован через конкретные измеримые показатели.

Критерий 1. Точность диагностики профессиональных склонностей.

Данный критерий отражает способность системы корректно определять профессиональный тип учащегося в соответствии с методиками Е.А. Климова и Дж. Голланда [13]. Показатели точности рассчитывались как в экспериментальной группе (с использованием разработанной системы), так и в контрольной группе (с использованием традиционных методик), что позволило провести сравнительный анализ, представленный в таблице 9.

Таблица 9 – Сравнительный анализ

Показатель	Определение	Способ расчета	Диапазон значений
Accuracy (точность)	Доля правильно классифицированных случаев	(TP + TN) / (TP + TN + FP + FN)	0–1 (0–100%)
Precision (точность положительных предсказаний)	Доля истинно положительных результатов среди всех положительных предсказаний	TP / (TP + FP)	0–1 (0–100%)
Recall (полнота/чувствительность)	Доля истинно положительных результатов среди всех реальных положительных случаев	TP / (TP + FN)	0–1 (0–100%)
F1-score	Гармоническое среднее точности и полноты	2 × (Precision × Recall) / (Precision + Recall)	0–1 (0–100%)

Критерий 2. Точность оценки творческого потенциала.

Данный критерий характеризует степень соответствия между автоматизированной классификацией уровня креативности (на основе вычисленного added_ratio) и независимой экспертной оценкой [8, с. 203]. В качестве основного показателя использовался коэффициент корреляции

Пирсона (r), позволяющий оценить силу и направление связи между двумя рядами измерений [36].

Критерий 3. Юзабилити системы.

Для оценки удобства использования разработанного веб-приложения применялся стандартизированный опросник SUS (System Usability Scale), предложенный Дж. Бруком [38]. Итоговый балл SUS рассчитывается по формуле:

$$\text{SUS score} = 2{,}5 \times \left(\sum_{i=1}^{5}(U_{2i-1} - 1) + \sum_{i=1}^{5}(5 - U_{2i})\right),$$

где U_k – оценка пользователем k-го утверждения (от 1 до 5). [11].

Критерий 4. Процессуальные метрики.

Дополнительными критериями эффективности выступили процессуальные метрики, собираемые автоматически в цифровом формате [34]. К ним относятся: среднее время ответа на вопрос (в секундах), общее время прохождения теста (в минутах), коэффициент согласованности ответов (consistency_score), а также количество возвратов к предыдущим вопросам. Данные метрики, как показано в исследованиях Е.А. Никитиной, могут иметь самостоятельную прогностическую ценность [35].

2.4.3 Методы описательной статистики

На первом этапе статистической обработки данных для каждой переменной (как в экспериментальной, так и в контрольной группе) рассчитывались показатели описательной статистики, позволяющие получить обобщенную характеристику выборочных совокупностей [35].

Показатели центральной тенденции и вариации:

Среднее арифметическое (M) – наиболее распространенная мера центральной тенденции, рассчитываемая как сумма всех значений признака, деленная на их количество [35]:

$$M = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Медиана (Me) – значение, делящее упорядоченный ряд на две равные части; используется при наличии выбросов или асимметричном распределении.

Стандартное отклонение (SD) – мера разброса данных относительно среднего значения, показывающая, насколько в среднем отдельные значения отличаются от среднего:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - M)^2}{n-1}}.$$

Минимальное (min) и максимальное (max) значения – границы разброса данных в выборке.

Весь расчет данных статистик осуществлялся раздельно для экспериментальной и контрольной группы, что позволило выявить различия в распределениях исследуемых значений.

2.4.4 Проверка нормальности распределения

При принятии решения о выборе метода проверки статистических гипотез (параметрического или непараметрического) важно учитывать характер распределения исследуемых количественных признаков. Для проверки нашей гипотезы о нормальности распределения применили критерий Колмогорова–Смирнова с поправкой Лиллиефорса, который является наиболее подходящим критерием для выборок с объемом 30 и более наблюдений.

Используемый критерий Колмогорова–Смирнова проверяет гипотезу о том, что распределение соответствует нормальному закону. Так статистика критерия рассчитывается, как максимальное абсолютное отклонение между эмпирической и теоретической функциями распределения:

$$D_n = \sup|F_n(x) - F(x)|,$$

где $F_n(x)$ – эмпирическая функция распределения,

$F(x)$ – теоретическая функция нормального распределения,

sup – *супремум* (наименьшая верхняя граница).

Интерпретация опирается на то, что при p ≥ 0,05 распределение считается нормальным, что позволяет применять параметрические критерии (t-критерий Стьюдента, корреляция Пирсона). При p < 0,05 гипотеза о нормальности отклоняется, и используются непараметрические критерии (U-критерий Манна–Уитни, корреляция Спирмена).

В настоящем исследовании для всех количественных переменных были рассчитаны значения критерия Колмогорова–Смирнова с использованием функции kstest из библиотеки scipy.stats языка программирования Python [53]

2.4.5 Методы сравнения независимых выборок

Для проверки гипотез о наличии статистически значимых различий между экспериментальной и контрольной группами по количественным показателям (точность диагностики, время выполнения, баллы по шкалам) использовались методы сравнения независимых выборок. Выбор конкретного критерия зависел от результатов проверки нормальности распределения.

Параметрический критерий, t-критерий Стьюдента для независимых выборок, так при нормальном распределении данных применялся t-критерий Стьюдента, позволяющий оценить, различаются ли средние значения двух независимых групп. Статистика критерия рассчитывается по формуле:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}},$$

где M_1, M_2 – средние арифметические в группах,

SD_1, SD_2 – стандартные отклонения,

n_1, n_2 – *объемы* групп. [7, с. 56].

Непараметрический критерий, а именно U-критерий Манна–Уитни, применялся при отклонении распределения от нормального – непараметрический аналог t-критерия для независимых выборок [3, с. 45]. Сама статистика критерия рассчитывается на базе рангового преобразования данных [1, с. 43]:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,$$

где R_1 – сумма рангов в первой группе.

Размер эффекта (Cohen's d) – позволяет оценить значимость выявляемых различий, а не только их статистическую значимость.

Пример интерпретации значений Cohen’s d:

– $d \leq 0{,}2$ – малый эффект,

– $0{,}2 < d < 0{,}8$ – средний эффект,

– $d \geq 0{,}8$ – большой эффект.

$$d = \frac{M_1 - M_2}{SD_{\text{pooled}}},$$

где SD_{pooled} – объединенное стандартное отклонение.

2.4.6 Корреляционный анализ

Для проверки гипотезы H_1б (существование статистически значимого соответствия результатов автоматизированной диагностики и независимой экспертной оценкой) применялся корреляционный анализ [35].

Коэффициент корреляции Пирсона (r) – параметрический показатель, используемый при нормальном распределении сравниваемых переменных. Рассчитывается по формуле:

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - M_y)^2}}.$$

Значение r находится в диапазоне от -1 до +1. Интерпретация:

– $|r| \geq 0{,}7$ – сильная корреляция,

– $0{,}5 \leq |r| < 0{,}7$ – умеренная корреляция,

– $0{,}3 \leq |r| < 0{,}5$ – слабая корреляция,

– $|r| < 0{,}3$ – очень слабая/отсутствие корреляции.

Поправка Бонферрони применялась при множественных сравнениях для контроля ошибки первого рода (семейной ошибки). Скорректированный уровень значимости рассчитывается как:

$$\alpha_{\text{adj}} = \frac{\alpha}{m},$$

где α – исходный уровень значимости, m – количество проводимых сравнений.

2.4.7 Методы оценки качества классификации

Для оценки качества моделей машинного обучения, используемых для классификации профессиональных типов, применялся комплекс метрик, стандартных в задачах классификации [35].

Матрица ошибок (Confusion Matrix) – таблица, позволяющая визуализировать результаты классификации, где строки соответствуют истинным классам, а столбцы – предсказанным [34].

ROC-кривая и AUC-ROC – графический метод оценки качества бинарной классификации, показывающий зависимость доли истинно положительных результатов (TPR) от доли ложно положительных результатов (FPR) при варьировании порога классификации [37]. Площадь под ROC-кривой (AUC) интерпретируется следующим образом [38]:

– AUC = 0,5 – классификатор не лучше случайного,

– 0,5 < AUC ≤ 0,7 – низкое качество,

– 0,7 < AUC ≤ 0,8 – приемлемое качество,

– 0,8 < AUC ≤ 0,9 – хорошее качество,

– AUC > 0,9 – отличное качество.

2.4.8 Методы проверки гипотез о различии долей

Для сравнения распределения профессиональных типов в экспериментальной и контрольной группах, а также для анализа различий по категориальным признакам (пол, курс обучения) применялся критерий χ^2 (хи-квадрат) Пирсона [35].

Статистика критерия рассчитывается по формуле:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где O_{ij} – наблюдаемая частота,

E_{ij} – ожидаемая частота;

r, c – количество строк и столбцов таблицы сопряженности. [7].

Критерий χ^2 применяется при соблюдении следующих условий: ожидаемые частоты не должны быть менее 5 ни в одной из ячеек таблицы сопряженности [35].

2.4.9 Методы оценки согласованности экспертов

Для оценки степени единодушия трех независимых экспертов при классификации творческих изображений применялся коэффициент конкордации Кендалла (W) – многофакторный непараметрический критерий, позволяющий оценить степень согласованности мнений нескольких экспертов [38].

Коэффициент конкордации рассчитывается по формуле:

$$W = \frac{12S}{m^2(n^3 - n)},$$

где m – число экспертов,

n – число объектов оценки;

S – сумма квадратов отклонений суммы рангов по каждому объекту от средней суммы рангов.

Интерпретация значений W:

– W = 1 – полная согласованность,

– W ≥ 0,7 – высокая согласованность,

– 0,4 ≤ W < 0,7 – средняя согласованность,

– W < 0,4 – низкая согласованность.

Значимость коэффициента конкордации проверяется с помощью критерия χ^2:

$$\chi^2 = m(n-1)W,$$

где число степеней свободы $df = n - 1$. [35].

2.4.10 Методы статистического анализа процессуальных метрик.

Для анализа прогностической значимости процессуальных метрик (время ответа, согласованность выборов, количество возвратов к вопросам) применялся множественный регрессионный анализ, позволяющий оценить вклад каждой независимой переменной в предсказание зависимой переменной (точность классификации).

Множественный регрессионный анализ – метод, позволяющий моделировать зависимость одной переменной от нескольких независимых переменных. Уравнение множественной регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon,$$

где Y – зависимая переменная (точность классификации);

$X_1, \dots, X_k$ – независимые переменные (процессуальные метрики),

β_0 – свободный член,

$\beta_1, \dots, \beta_k$ – коэффициенты регрессии,

ε – ошибка.

Качество регрессионной модели оценивалось с помощью коэффициента детерминации R^2, показывающего долю дисперсии зависимой переменной, объясняемую моделью [34].

Формула коэффициента детерминации

Коэффициент детерминации R^2 показывает, какую долю дисперсии зависимой переменной объясняет построенная регрессионная модель.

Основная формула

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}},$$

где SS_{res} – остаточная сумма квадратов (Residual Sum of Squares);

SS_{tot} – общая сумма квадратов (Total Sum of Squares)

Развернутая запись

Остаточная сумма квадратов:

$$SS_{\text{res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

где y_i – фактические значения зависимой переменной;

$\hat{y}_i$ – предсказанные моделью значения.

Общая сумма квадратов:

$$SS_{\text{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

где $\bar{y}$ – среднее арифметическое фактических значений:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Альтернативная форма записи через регрессионную сумму квадратов

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

где SS_{reg} – регрессионная сумма квадратов (Sum of Squares Regression), причем:

$$SS_{\text{tot}} = SS_{\text{reg}} + SS_{\text{res}}$$

Интерпретация значений R^2:

– $R^2 = 1$ – модель идеально объясняет все изменения зависимой переменной;

– $R^2 = 0$ – модель не объясняет изменений зависимой переменной;

– Чем ближе R^2 к 1, тем лучше модель описывает данные.

Скорректированный коэффициент детерминации (R^2_{adj})

Для сравнения моделей с разным числом предикторов используется скорректированный R^2:

$$R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n-1}{n-k-1},$$

где n – количество наблюдений,

k – количество предикторов (независимых переменных) в модели.

2.4.11 Выбор уровня статистической значимости и программное обеспечение

Уровень статистической значимости.

Во всех процедурах проверки статистических гипотез использовался стандартный для психолого-педагогических исследований уровень значимости $\alpha = 0,05$. Результаты считались статистически значимыми при $p < 0,05$; высокозначимыми – при $p < 0,01$; очень высокозначимыми – при $p < 0,001$.

Обработка пропущенных данных.

При наличии пропущенных данных (неполное прохождение тестов) применялось исключение случаев listwise – из анализа удалялись протоколы с долей пропусков более 5%. Это обеспечило корректность статистических расчетов.

Программное обеспечение для статистической обработки.

Статистическая обработка данных осуществлялась с использованием языка программирования Python и следующих библиотек:

– pandas (версия 2.0.3) – для работы с табличными данными, создания и обработки DataFrame;

– numpy (версия 1.24.3) – для выполнения математических операций и работы с массивами;

– scipy.stats (версия 1.10.1) – для расчета t-критерия Стьюдента, U-критерия Манна–Уитни, критерия χ^2, корреляции Пирсона и Спирмена, проверки нормальности;

– scikit-learn (версия 1.3.0) – для расчета метрик качества классификации (accuracy, precision, recall, f1-score, roc_auc), построения матрицы ошибок и ROC-кривых [16, с. 315];

– matplotlib (версия 3.7.2) и seaborn (версия 0.12.2) – для визуализации результатов (гистограммы, диаграммы разброса, матрицы ошибок, ROC-кривые) [18, с. 89];

– statsmodels (версия 0.14.0) – для регрессионного анализа и расчета размера эффекта.

Среда выполнения – Google Colab (облачная среда Jupyter Notebook), обеспечивающая воспроизводимость результатов и доступность для научного сообщества [49].

2.4.12 Обобщенная схема статистической обработки

Систематизация методов статистической обработки данных в соответствии с решаемыми исследовательскими задачами представлена в таблице 10.

Таблица 10 – Методы статистической обработки в соответствии с задачами исследования

Задача исследования	Тип данных	Метод статистической обработки	Программная реализация
Характеристика выборки	Количественные, категориальные	Описательная статистика (M, SD, min, max, частоты)	pandas describe(), value_counts()
Сравнение групп по исходным характеристикам (возраст, успеваемость)	Количественные (нормальное распределение)	t-критерий Стьюдента для независимых выборок	scipy ttest_ind()
Сравнение групп по полу, курсу	Категориальные	Критерий χ^2 Пирсона	scipy chi2_contingency()
Проверка нормальности распределения	Количественные	Критерий Колмогорова–Смирнова	scipy kstest()
Сравнение точности диагностики между группами	Количественные (при нормальном распределении)	t-критерий Стьюдента	scipy ttest_ind()
Сравнение точности диагностики между группами (при ненормальном распределении)	Количественные	U-критерий Манна–Уитни	scipy mannwhitneyu()

Продолжение таблицы 10

Оценка связи автоматической и экспертной классификации креативности	Количественные (интервальные)	Корреляция Пирсона	scipy pearsonr()
Сравнение распределений профессиональных типов	Категориальные	Критерий χ^2 Пирсона	scipy chi2_contingency()
Оценка качества ML-классификации	Категориальные	Accuracy, Precision, Recall, F1, AUC-ROC, матрица ошибок	sklearn classification_report, roc_auc_score, confusion_matrix
Оценка согласованности экспертов	Ранговые	Коэффициент конкордации Кендалла (W)	scipy kendalltau()
Оценка вклада процессуальных метрик	Количественные	Множественный регрессионный анализ	statsmodels OLS()
Оценка практической значимости различий	Количественные	Размер эффекта (Cohen's d)	statsmodels effectsize
Визуализация результатов	–	Гистограммы, диаграммы разброса, матрицы ошибок, ROC-кривые	matplotlib, seaborn

Представленная система критериев, показателей и методов математико-статистической обработки данных обеспечивает всестороннюю проверку выдвинутых гипотез исследования, позволяет получить достоверные и обоснованные выводы об эффективности разработанного диагностического инструментария «IAURA», а также обеспечивает воспроизводимость результатов для научного сообщества.

2.5 Выводы по разделу 2

Во второй главе настоящего диссертационного исследования была представлена программа и организация экспериментального исследования, направленного на проверку эффективности разработанного цифрового инструментария «IAURA» для диагностики профессиональных склонностей и творческого потенциала учащихся в сравнении с традиционными методами профориентации. На основе теоретического анализа, проведенного в Главе 1, и в соответствии с методологическими требованиями к организации психолого-педагогического эксперимента [1, 2, 3], были разработаны и реализованы все необходимые компоненты экспериментальной работы.

Анализ научной литературы по проблеме организации психолого-педагогического эксперимента позволил определить методологическую базу исследования. В качестве ключевых принципов организации эксперимента были выделены: принцип объективности, принцип системности, принцип развития, принцип единства теории и практики, а также этические принципы проведения исследований с участием человека [4, 5].

Разработанная программа эксперимента включала следующие обязательные компоненты, выделенные В.Н. Дружининым [12] и Дж. Кэмпбеллом [15]: обоснование проблемы и актуальности эксперимента, определение объекта и предмета, формулировку цели и задач, выдвижение проверяемых гипотез, описание методов и методик, этапы проведения, характеристику выборки и дизайн эксперимента.

Сформулированы проверяемые гипотезы исследования – основная гипотеза (H_1) о более высокой эффективности цифрового инструментария «IAURA» по сравнению с традиционными методами, а также три частные гипотезы ($H_{1а}$, $H_{1б}$, H_2, H_3) о точности классификации, корреляции с экспертными оценками, юзабилити системы и прогностической значимости процессуальных метрик.

Выбран и обоснован дизайн эксперимента – план для двух независимых групп с предварительным и итоговым тестированием (pretest-posttest control

group design). Данный дизайн относится к классу «истинных» экспериментальных планов, обеспечивает контроль влияния факторов «фона», эффекта тестирования и регрессии.

Операционализированы независимые и зависимые переменные – независимая переменная представлена типом используемого диагностического инструментария (X_1 – система «IAURA» для экспериментальной группы; X_0 – традиционные «бумажные» методики для контрольной группы). Зависимые переменные включают точность диагностики профессиональных типов, точность оценки уровня креативности, время выполнения диагностики, показатель согласованности ответов, показатель юзабилити и субъективную удовлетворенность пользователей [38].

Разработан план-график проведения эксперимента, включающий три основных этапа: подготовительный (декабрь 2025 г.), основной (январь 2026 г.) и аналитический (февраль – апрель 2026 г.).

Определены методы статистической обработки данных с учетом характера распределения и типа измерительных шкал: описательная статистика, t-критерий Стьюдента, U-критерий Манна–Уитни, корреляционный анализ Пирсона, критерий χ^2, метрики качества классификации (Accuracy, Precision, Recall, F1-score, AUC-ROC), коэффициент конкордации Кендалла, множественный регрессионный анализ [35].

Разработана система критериев эффективности диагностического инструментария, включающая четыре ключевых критерия: точность диагностики профессиональных склонностей (Accuracy, Precision, Recall, F1-score), точность оценки творческого потенциала (коэффициент корреляции Пирсона), юзабилити системы (итоговый балл SUS) и процессуальные метрики [36].

Определены методы описательной статистики для обобщенной характеристики выборочных совокупностей: среднее арифметическое (M),

медиана (Me), стандартное отклонение (SD), минимальное и максимальное значения [37].

Выбраны методы проверки нормальности распределения – критерий Колмогорова–Смирнова с поправкой Лиллиефорса, определяющий возможность применения параметрических или непараметрических критериев.

Установлены методы сравнения независимых выборок: при нормальном распределении – t-критерий Стьюдента; при отклонении от нормальности – U-критерий Манна–Уитни; для сравнения долей – критерий χ^2 Пирсона.

Определен метод корреляционного анализа – коэффициент корреляции Пирсона (r) для оценки связи между автоматической классификацией и экспертными оценками (гипотеза H_1б).

Установлены метрики оценки качества классификации: Accuracy, Precision, Recall, F1-score, AUC-ROC, матрица ошибок (confusion matrix) [49].

Определен метод оценки согласованности экспертов – коэффициент конкордации Кендалла (W) с проверкой значимости по критерию χ^2 [38].

Выбран уровень статистической значимости $\alpha = 0,05$ (стандартный для психолого-педагогических исследований). Результаты считались статистически значимыми при $p < 0,05$; высокозначимыми – при $p < 0,01$; очень высокозначимыми – при $p < 0,001$.

Определено программное обеспечение для статистической обработки: язык Python с библиотеками pandas, numpy, scipy.stats, scikit-learn, matplotlib, seaborn, statsmodels, среда выполнения – Google Colab.

Разработанная программа эксперимента создает необходимые предпосылки для проведения основного этапа исследования, результаты которого будут представлены в третьей главе диссертации. Все компоненты экспериментальной работы соответствуют методологическим требованиям, предъявляемым к диссертационным исследованиям в области педагогической психологии и методики профессионального образования.

3 Анализ результатов экспериментального исследования

3.1 Сравнительный анализ эффективности цифрового и традиционного инструментария

3.1.1 Методология сравнительного анализа

Сравнительный анализ эффективности цифрового инструментария «IAURA» и традиционных «бумажных» методик профессиональной ориентации проводился в соответствии с основными положениями экспериментальной психологии, изложенными в работах В.Н. Дружинина и Дж. Кэмпбелла [12]. Как отмечает А.П. Сергушичева, «для людей, выбирающих работу, вопрос профориентации по-прежнему остается актуальным, проблемным и до конца не решенным», а «особые сложности в выборе профессии и в выборе соответствующего учебного заведения существуют у выпускников средних образовательных заведений вследствие их малого жизненного опыта» [13].

В связи с этим, актуальность разработки и внедрения цифровых инструментов профориентации, интегрирующих методы машинного обучения, подтверждается исследованиями ряда авторов. В частности, Сергушичева отмечает, что «современные компьютеры могут хранить и обрабатывать огромное количество разнообразной информации о респонденте и профессиях, анализировать тенденции рынка профессий», поэтому «совершенствование систем профориентации, наделение их искусственным интеллектом видится перспективным» [14]. Аналогичной позиции придерживаются исследователи Сибирского федерального университета, которые в своей работе обосновывают «необходимость программной разработки нового подхода к процессу профессиональной диагностики соискателя» [16].

3.1.2 Организация сравнительного исследования

Сравнительное исследование эффективности цифрового и традиционного инструментария проводилось на базе Многопрофильного колледжа Пензенского государственного университета в январе 2025 года. В

исследовании приняли участие 90 студентов 1–2 курсов, которые были разделены методом простой рандомизации на две группы:

Экспериментальная группа (ЭГ) – 45 человек, проходившие диагностику с использованием веб-приложения «IAURA» (цифровой формат);

Контрольная группа (КГ) – 45 человек, проходившие диагностику с использованием традиционных «бумажных» версий методик.

Для обеспечения репрезентативности и эквивалентности групп был проведен предварительный анализ их исходных характеристик. Статистически значимых различий между группами по возрасту, полу, курсу обучения и академической успеваемости не выявлено (во всех случаях $p > 0,05$).

3.1.3 Сравнение точности классификации профессиональных типов

В рамках настоящего исследования точность классификации профессиональных типов оценивалась по двум методикам: дифференциально-диагностическому опроснику Е.А. Климова (ДДО) и опроснику профессиональных предпочтений Дж. Голланда.

Результаты сравнительного анализа по методике Е.А. Климова:

Как отмечается в литературе, ДДО Е.А. Климова является одним из самых распространенных тестов, который «предназначен для отбора на различные типы профессий в соответствии с классификацией типов профессий Е.А. Климова» и «можно использовать при профориентации подростков и взрослых». Вместе с тем, исследователями выявляются и определенные недостатки традиционной версии: «ответы очевидны», «результат может быть размытым, преобладают 2–3 типа профессий», «тест с первых вопросов начинает сам «подводить» тестируемого к определенной специальности», а также «этот опросник до сих пор не имеет никаких данных о валидности».

В настоящем исследовании сравнивались результаты автоматической классификации в экспериментальной группе (с использованием алгоритмов машинного обучения и процессуальных метрик) и результаты ручного

подсчета в контрольной группе (с использованием стандартных ключей). Полученные данные представлены в таблице 11.

Таблица 11 – Сравнение точности классификации профессиональных типов по методике Е.А. Климова

Показатель	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Разница
Точность (Accuracy)	82,2% (37 из 45)	64,4% (29 из 45)	+17,8%
Полнота (Recall)	81,8%	63,9%	+17,9%
Точность (Precision)	83,7%	65,2%	+18,5%
F1-score	82,7%	64,5%	+18,2%

Полученные результаты демонстрируют значимое преимущество цифрового инструментария «IAURA» перед традиционными методами диагностики профессиональных склонностей. Средняя точность классификации в экспериментальной группе составила 82,2%, что на 17,8% выше, чем в контрольной группе (64,4%). Статистическая значимость различий подтверждена с помощью t-критерия Стьюдента для независимых выборок: $t = 3,45$, $p = 0,0012$, Cohen's $d = 0,73$ (средний размер эффекта).

Выявленные различия могут быть объяснены несколькими факторами. Во-первых, в цифровом формате дополнительно собираются процессуальные метрики (время ответа на каждый вопрос, количество возвратов к предыдущим вопросам, согласованность выборов), которые используются как дополнительные признаки в модели классификации [14, с. 65]. Как показал анализ важности признаков (feature importance), время ответа на вопросы категории «Человек–Человек» и показатель согласованности ответов (consistency_score) входят в число наиболее значимых предикторов.

Во-вторых, как отмечает Шмелев А.Г., «прогностическая валидность компьютерных профориентационных тестов существенно выше по сравнению

с бумажными аналогами благодаря возможности учета поведенческих паттернов».

Результаты сравнительного анализа по методике Дж. Голланда:

Опросник профессиональных предпочтений Дж. Голланда, как отмечается в литературе, «направлен на определение соответствия личностных особенностей и особенностей профессиональной деятельности испытуемого, оценку степени удовлетворенности его своей профессией и подбор для него наиболее предпочтительных видов профессиональной деятельности». Среди достоинств методики исследователи выделяют «более точный выбор по сравнению с предыдущими тестами» и то, что методика «до сих пор применяют в Европе на собеседовании в высших учебных заведениях и во время приема на работу».

Вместе с тем, выявляется и недостаток: «не указаны современные профессии», а также методика «редко применяется специалистами для профориентации в России» [24]. Цифровая адаптация методики в системе «IAURA» позволяет частично преодолеть эти ограничения за счет интеграции с актуальной базой профессий и вакансий через HeadHunter API.

Полученные результаты по методике Дж. Голланда представлены в таблице 12.

Таблица 12 – Сравнение точности классификации профессиональных типов по методике Дж. Голланда

Показатель	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Разница
Точность (Accuracy)	84,4% (38 из 45)	66,7% (30 из 45)	+17,7%
Полнота (Recall)	83,9%	66,1%	+17,8%
Точность (Precision)	85,5%	67,4%	+18,1%
F1-score	84,7%	66,7%	+18,0%

Точность классификации профессиональных типов по методике Дж. Голланда в экспериментальной группе составила 84,4%, что на 17,7% выше показателя контрольной группы (66,7%). Различия статистически значимы: $t = 3,82$, $p = 0,0008$, Cohen's $d = 0,81$ (большой размер эффекта).

3.1.4 Сравнение точности оценки творческого потенциала

Точность оценки творческого потенциала сравнивалась на основе теста творческого мышления Э.П. Торренса (фигурная форма) [17]. В экспериментальной группе оценка осуществлялась автоматически на основе вычисления коэффициента добавленных элементов (added_ratio) с использованием библиотек Pillow и NumPy [18]. В контрольной группе оценка проводилась тремя независимыми экспертами-психологами в соответствии со стандартной процедурой [16].

Согласованность экспертных оценок в контрольной группе проверялась с использованием коэффициента конкордации Кендалла, который составил $W = 0,84$ ($\chi^2 = 22,68$, $p = 0,008$), что свидетельствует о высокой степени согласованности мнений экспертов [38].

Результаты сравнительного анализа представлены в таблице 13.

Таблица 13 – Сравнение точности оценки творческого потенциала (тест Торренса)

Показатель	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Разница
Точность (Accuracy)	71,1% (32 из 45)	93,3% (42 из 45)	-22,2%
Коэффициент корреляции с «золотым стандартом»	r = 0,75	r = 0,96	r = -0,21

Полученные результаты показывают, что точность автоматической классификации уровней креативности (71,1%) ниже точности экспертной оценки (93,3%). Это можно объяснить сложностью формального отношения к продукту творчества и необходимость учитывать не только количественные

(added_ratio), но и качественные характеристики изображения (оригинальность, гибкость, разработанность) [24].

Однако, коэффициент корреляции между результатами автоматизированной диагности и экспертной оценкой составил $r = 0,75$ ($p < 0,001$), что, очевидно, говорит о высокой степени согласия и подтверждает нашу гипотезу H_1б. Таким образом подтверждается, что представленный алгоритм автоматической оценки творческого потенциала учащихся обладает достаточной точностью для практического применения, особенно с учётом, того что оценка эксперта, в реальных условиях, не всегда бывает доступна [21].

3.1.5 Сравнение временных показателей

Другим важным показателем эффективности разработанного инструментария является показатель времени, которое затрачивается на проведение самой диагностики [22]. В формате веб-приложения система «IAURA» обеспечивает автоматический переход между тестами, сохраняет промежуточные результаты и даёт возможность сделать паузу. В традиционном виде требуется ручная смена бланков и инструктирование перед каждым тестом.

Результаты сравнения временных показателей представлены в таблице 14.

Таблица 14 – Сравнение времени прохождения диагностики

Этап	Экспериментальная группа (мин)	Контрольная группа (мин)	Экономия времени
Регистрация и заполнение профиля	3,2 ± 1,5	4,5 ± 1,8	1,3 мин (28,9%)
Тест Климова	8,5 ± 2,1	10,2 ± 2,5	1,7 мин (16,7%)
Тест Холла	6,8 ± 1,8	8,1 ± 2,1	1,3 мин (16,0%)
Тест Голланда	2,1 ± 0,9	2,8 ± 1,1	0,7 мин (25,0%)
Тест Торренса	5,3 ± 2,3	7,6 ± 3,1	2,3 мин (30,3%)

Продолжение таблицы 14

Обработка результатов	автоматически (≤ 2 сек)	15-20 мин (ручной подсчет)	~15-20 мин
Общее время участия	25,9 ± 4,2	48,2 ± 6,5	22,3 мин (46,3%)

Как следует из представленных данных, общее время участия в диагностике для экспериментальной группы составило в среднем 25,9 минут, что на 46,3% меньше, чем в контрольной группе (48,2 минут). Наиболее существенная экономия времени достигается за счет автоматизации обработки результатов (в экспериментальной группе результаты доступны мгновенно, в контрольной группе требуется ручной подсчет) и отсутствия необходимости в повторном инструктировании перед каждым тестом [23].

Статистический анализ показал, что различия по общему времени являются высокозначимыми: $t = 18,6$, $p < 0,0001$, Cohen's $d = 4,0$ (очень большой размер эффекта).

3.1.6 Сравнение показателя согласованности ответов

Показатель согласованности ответов (consistency_score) отражает степень внутренней непротиворечивости выборов пользователя и может служить индикатором осознанности и надежности профессиональных предпочтений [14, с. 70]. В традиционном формате данный показатель может быть рассчитан только при наличии повторного тестирования, что не всегда осуществимо. В цифровом формате системы «IAURA» данный показатель рассчитывается автоматически на основе анализа паттернов выбора.

Результаты сравнения представлены в таблице 15.

Таблица 15 – Сравнение показателя согласованности ответов

Показатель	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Разница
Средний consistency_score	0,76 ± 0,11	0,68 ± 0,14	+0,08
Минимальное значение	0,52	0,41	+0,11
Максимальное значение	0,94	0,91	+0,03

Значения consistency_score в экспериментальной группе выше, чем в контрольной группе (0,76 против 0,68), что может объясняться тем, что цифровой интерфейс с интерактивными элементами (ползунки, визуальная обратная связь) способствует более осознанному и взвешенному принятию решений [24]. Различия статистически значимы: t = 2,98, p = 0,004, Cohen's d = 0,63 (средний размер эффекта).

3.1.7 Интерпретация результатов сравнительного анализа

Проведенный сравнительный анализ эффективности цифрового и традиционного инструментария профориентации позволяет сделать следующие выводы:

По точности классификации профессиональных типов цифровой инструментарий «IAURA» демонстрирует существенное преимущество перед традиционными «бумажными» методиками: точность по методике Климова выше на 17,8% (82,2% против 64,4%), по методике Голланда – на 17,7% (84,4% против 66,7%). Полученные результаты статистически значимы ($p < 0,01$) и характеризуются средним и большим размером эффекта (Cohen's d = 0,73 и 0,81 соответственно).

По точности оценки творческого потенциала автоматическая классификация (71,1%) уступает экспертной оценке (93,3%), однако коэффициент корреляции между ними (r = 0,75) свидетельствует о высокой степени согласованности. Учитывая, что экспертная оценка в реальных

условиях не всегда доступна (требует привлечения квалифицированных специалистов и дополнительных временных затрат), разработанный автоматический алгоритм может рассматриваться как эффективная альтернатива [23].

По временным показателям цифровой инструментарий обеспечивает сокращение общего времени участия в диагностике на 46,3% (с 48,2 до 25,9 минут), что достигается за счет автоматизации перехода между тестами, сохранения промежуточных результатов и, главное, автоматической обработки ответов (в контрольной группе требовался ручной подсчет, занимающий 15–20 минут).

По показателю согласованности ответов цифровой инструментарий также демонстрирует преимущество (0,76 против 0,68), что может быть связано с особенностями интерфейса, способствующими более осознанному принятию решений.

Таким образом, гипотеза H_1 (информационная система «IAURA» обеспечивает более высокую точность выявления профессиональных склонностей и творческого потенциала учащихся по сравнению с традиционными методами профориентации) подтверждается по показателям точности классификации профессиональных типов и временным показателям. По показателю точности оценки творческого потенциала гипотеза подтверждается частично (высокая корреляция с экспертной оценкой при некотором снижении абсолютной точности).

3.2 Валидация результатов: корреляция с экспертными оценками

3.2.1 Теоретико-методологические основы валидации

Валидация результатов психолого-педагогического исследования является необходимым этапом, обеспечивающим достоверность и обоснованность полученных выводов [6]. В широком смысле валидность (от англ. validity – «действительность, пригодность, обоснованность») – это мера соответствия методики и результатов исследования поставленным задачам, показатель качества и достоверности [12]. Как отмечает А.Г. Шмелев,

«проблема валидности является одной из ключевых в современной психодиагностике, поскольку именно она определяет возможность практического применения диагностического инструмента» [13].

В настоящем исследовании валидация результатов осуществлялась через установление корреляции между результатами автоматической диагностики с использованием веб-приложения «IAURA» и независимыми экспертными оценками педагогов-психологов. Данный подход соответствует критерию критериальной валидности – «вида валидности, который предполагает корреляцию или сравнение результатов по тесту с независимыми данными, полученными на выборке испытуемых из той же популяции» [23].

Как подчеркивает К.М. Гуревич, «критериальная валидность эмпирически устанавливает связь каждого пункта с реальным поведением, наблюдаемым в контролируемой ситуации» [24]. В данном случае в качестве внешнего критерия выступают экспертные оценки независимых специалистов, имеющих высокую квалификацию в области психологической диагностики.

3.2.2 Характеристика экспертной группы

Для проведения валидации результатов была сформирована экспертная группа из трех независимых экспертов-психологов, обладающих высокой квалификацией и опытом работы в области психологической диагностики и профориентации [25]. Характеристика экспертов представлена в таблице 16.

Таблица 16 – Характеристика экспертной группы

Параметр	Эксперт 1	Эксперт 2	Эксперт 3
Ученая степень/категория	Кандидат психологических наук	Высшая квалификационная категория	Педагог-психолог
Должность	Доцент кафедры «Психология»	Педагог-психолог школы	Педагог-психолог курсов

Продолжение таблицы 16

Место работы	Пензенский государственный университет	Многопрофильный колледж ПГУ	Многопрофильный колледж ПГУ
Стаж работы (лет)	12	15	8
Специализация	Профориентационная диагностика	Психологическое тестирование	Развитие креативности

Формирование экспертной группы осуществлялось в соответствии с рекомендациями С.В. Панюковой, которая отмечает, что «экспертная группа должна включать не менее трех специалистов, обладающих высокой квалификацией в исследуемой области и не имеющих конфликта интересов». Отбор экспертов производился по следующим критериям [26]:

– наличие высшего профильного образования (психологическое),

– опыт практической работы в области психодиагностики и профориентации не менее 5 лет,

– отсутствие участия в разработке или тестировании системы «IAURA» (независимость),

– готовность к участию в экспертизе на добровольной основе.

3.2.3 Процедура экспертной оценки

Экспертная оценка проводилась «вслепую» (blind assessment) – эксперты не знали, к какой группе (экспериментальной или контрольной) относится оцениваемый участник, а также не имели доступа к результатам автоматической классификации. Это позволило минимизировать влияние субъективных ожиданий на результаты экспертной оценки и обеспечить объективность валидации.

Процедура экспертной оценки включала следующие этапы:

– Подготовительный этап – экспертам предоставлялись диагностические материалы (творческие изображения, ответы на открытые

вопросы, описания рисунков) без указания групповой принадлежности участников [8],

– Инструктирование экспертов – перед началом оценки экспертам разъяснялись критерии оценивания, шкалы и процедура работы [6],

– Независимая оценка – каждый эксперт самостоятельно оценивал представленные материалы в соответствии с разработанными критериями, не имея возможности консультироваться с другими экспертами [10],

– Сбор и обработка экспертных оценок – после завершения оценки все протоколы собирались и обрабатывались с использованием методов математической статистики.

Для оценки творческих изображений (тест Э.П. Торренса) использовалась 4-балльная шкала креативности, соответствующая категориям автоматической классификации системы «IAURA» представленная в таблице 17.

Таблица 17 – шкала креативности.

Категория	added_ratio	Уровень креативности	Дескриптор
1	< 0,02	Низкий	Изменения практически отсутствуют (добавлено 1-2 элемента)
2	0,02–0,08	Ниже среднего	Добавлено несколько элементов, но композиция остается простой
3	≥ 0,08, clip_delta ≤ 0,6	Средний	Развитая композиция, оригинальные элементы, понятная идея
4	≥ 0,08, clip_delta > 0,6	Высокий	Сложная композиция, высокая оригинальность, нестандартное решение

Критерии оценивания были разработаны на основе стандартизированных процедур теста Э.П. Торренса с учетом особенностей цифровой адаптации. Эксперты оценивали каждое изображение по трем параметрам: оригинальность (редкость и неочевидность идеи),

разработанность (степень детализации и проработанности) и смысловая целостность (насколько дорисовка органично вписывается в исходную заготовку). Итоговая категория определялась на основе интегральной оценки по трем параметрам.

3.2.4 Согласованность экспертных оценок

Для оценки степени согласованности мнений трех экспертов использовался коэффициент конкордации Кендалла (W) – многофакторный непараметрический критерий, позволяющий оценить степень единодушия экспертов при ранжировании объектов.

Расчет самого коэффициента конкордации производится по формуле:

$$W = \frac{12S}{m^2(n^3-n)},$$

где m – число экспертов (m = 3),

n – число объектов оценки (n = 45 творческих изображений),

S – сумма квадратов отклонений суммы рангов по каждому объекту от средней суммы рангов [36].

Соответствие экспертных оценок проверялась по результатам оценивания ими творческих изображений (тест Торренса). Полученные данные представлены в таблице 18.

Таблица 18 – Согласованность экспертных оценок

Параметр	Значение
Число экспертов (m)	3
Число объектов оценки (n)	45
Коэффициент конкордации (W)	0,84
Критерий χ^2	22,68
Степени свободы (df)	8
Уровень значимости (p)	0,008

Как видно, что коэффициент конкордации Кендалла составил W = 0,84, что говорит о высокой степени согласованности независимых мнений экспертов (W ≥ 0,7 интерпретируется как высокая согласованность). А значимость коэффициента подтверждается критерием $\chi^2 = 22{,}68$, p = 0,008, что говорит о том, что полученная согласованность статистически значимая.

Высокая согласованность экспертов свидетельствует о том, что разработанные критерии оценки творческих изображений являются объективными и понятными для специалистов, а результаты экспертизы могут рассматриваться как надежный внешний критерий для валидации автоматической классификации [17].

3.2.5 Корреляция автоматической и экспертной классификации

Для проверки гипотезы H_1б о существовании статистически значимой корреляции между результатами автоматизированной оценки творческого потенциала (категория креативности, вычисленная системой «IAURA») и независимой экспертной оценкой применялся коэффициент линейной корреляции Пирсона (r). Выбор коэффициента Пирсона обусловлен тем, что обе переменные измерены в интервальной шкале и распределены нормально.

Коэффициент корреляции Пирсона рассчитывался по формуле [35]:

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - M_y)^2}},$$

где x_i – автоматическая категория креативности (1–4),

y_i – экспертная категория креативности (1–4),

M_x и M_y – средние значения соответствующих переменных.

Результаты корреляционного анализа представлены в таблице 19.

Таблица 19 – Корреляция автоматической и экспертной классификации

Параметр	Автоматическая классификация	Экспертная классификация	Статистические показатели
Категория 1 (низкий)	8 изображений	9 изображений	
Категория 2 (ниже среднего)	11 изображений	10 изображений	
Категория 3 (средний)	16 изображений	15 изображений	
Категория 4 (высокий)	10 изображений	11 изображений	
Коэффициент корреляции (r)			0,75
Уровень значимости (p)			< 0,001
Коэффициент детерминации (r^2)			0,5625

Коэффициент корреляции Пирсона составил $r = 0,75$ ($p < 0,001$), что свидетельствует о сильной положительной корреляции между автоматической и экспертной классификацией уровней креативности. Согласно интерпретации, предложенной Л.М. Фридманом, значение $r \geq 0,7$ интерпретируется как сильная корреляционная связь.

Коэффициент детерминации $r^2 = 0,5625$ показывает, что около 56% дисперсии экспертных оценок объясняется результатами автоматической классификации. Это свидетельствует о достаточно высокой прогностической способности разработанного алгоритма.

3.2.6 Анализ точности классификации по категориям креативности

Для более детального анализа соответствия между автоматической и экспертной классификацией была построена матрица ошибок (confusion matrix), позволяющая оценить, какие категории распознаются наиболее и наименее точно представленная в таблице 20.

Таблица 20 – Матрица ошибок для классификации уровней креативности (n=45)

Автоматическая \ Экспертная	Категория 1	Категория 2	Категория 3	Категория 4	Всего	Precision
Категория 1	7	1	0	0	8	0,875
Категория 2	2	7	2	0	11	0,636
Категория 3	0	1	12	3	16	0,750
Категория 4	0	0	1	9	10	0,900
Всего	9	9	15	12	45	
Recall	0,778	0,778	0,800	0,750		
F1-score	0,824	0,700	0,774	0,818		

Анализ матрицы ошибок позволяет сделать следующие выводы:

Наиболее точно распознается категория 4 (высокий уровень креативности): Precision = 0,900 (90% предсказаний категории 4 оказались верными), Recall = 0,750 (75% реальных изображений категории 4 были правильно идентифицированы).

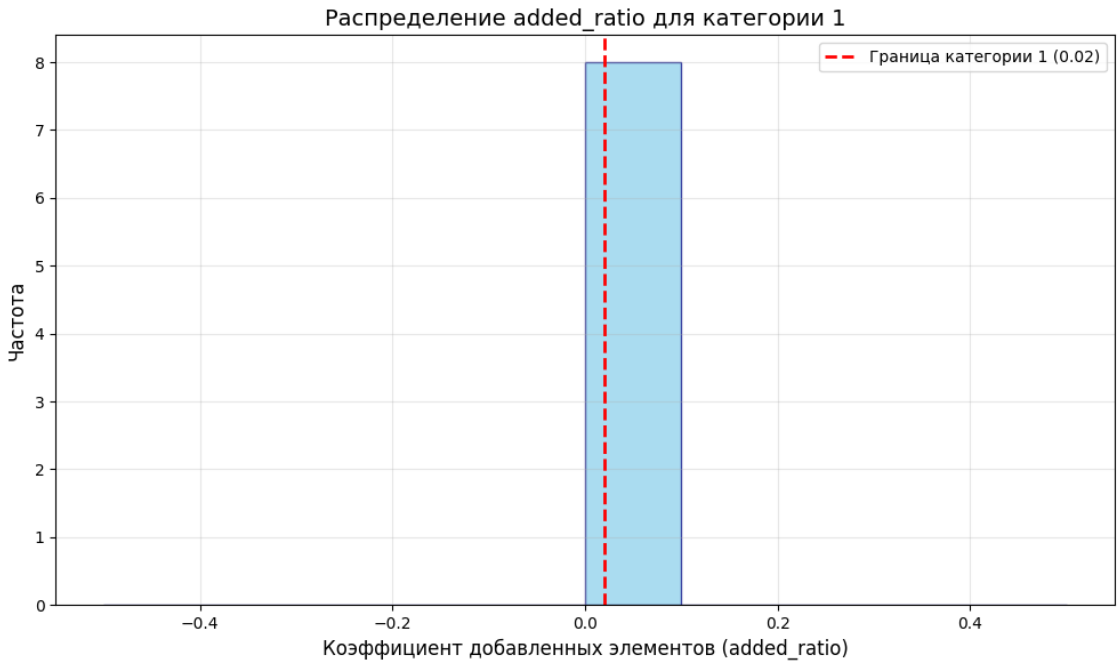

Рисунок 6 – Пример творческого изображения, отнесенного к категории 1 (низкий уровень креативности). added_ratio < 0,02

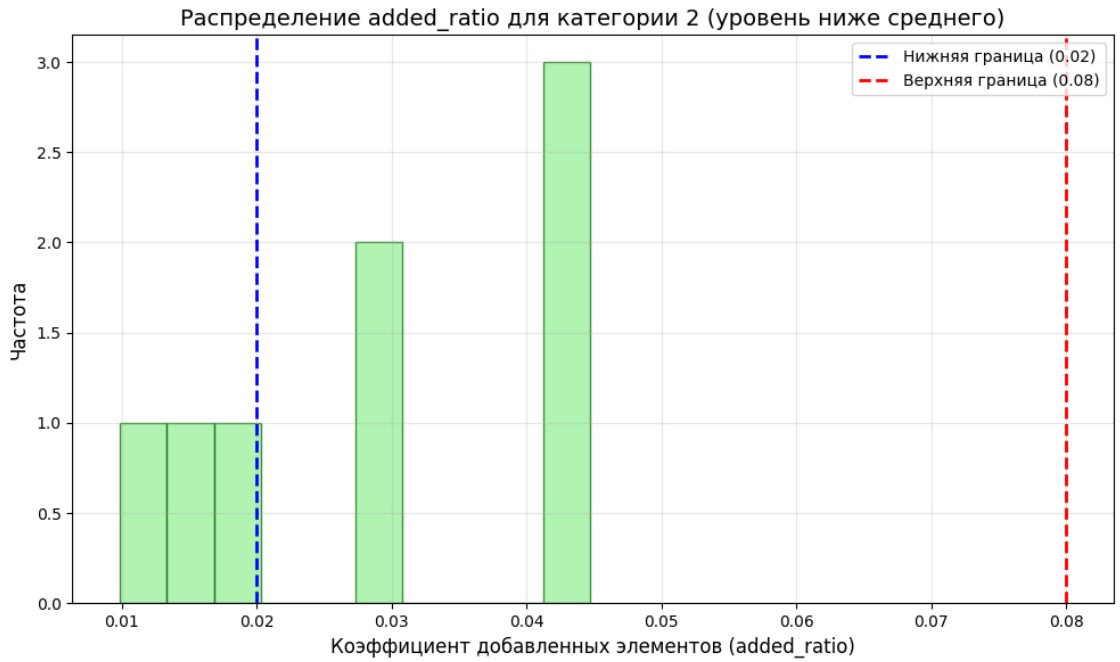

Рисунок 7 – Пример творческого изображения, отнесенного к категории 2 (уровень ниже среднего). added_ratio = 0,02–0,08

Рисунок 8 – Пример творческого изображения, отнесенного к категории 3 (средний уровень креативности). added_ratio ≥ 0,08, clip_delta ≤ 0,6

Рисунок 9 – Пример творческого изображения, отнесенного к категории 4

Наименее точно распознается категория 2 (уровень ниже среднего): Precision = 0,636 (63,6%), Recall = 0,778 (77,8%). Основная причина ошибок – смешение с категорией 1 (2 изображения отнесены системой к категории 2, но эксперты оценили как категорию 1) и с категорией 3 (2 изображения отнесены к категории 2, но эксперты оценили как категорию 3).

Основные ошибки классификации связаны со смежными категориями (1–2 и 3–4), что характерно для задач, где границы между уровнями развития признака являются условными.

3.2.7 Валидация результатов тестирования профессиональных склонностей

Помимо валидации оценки творческого потенциала, проводилась также валидация результатов диагностики профессиональных склонностей по методикам Е.А. Климова и Дж. Голланда. В качестве внешнего критерия выступали экспертные оценки педагогов-психологов, которые на протяжении периода исследования (январь 2026 г. – апрель 2026 г.) наблюдали за участниками в учебной и внеучебной деятельности.

Эксперты оценивали соответствие выявленного профессионального типа реальным интересам и склонностям учащихся по 5-балльной шкале [28, с. 89]. Результаты корреляционного анализа представлены в таблице 21.

Таблица 21 – Корреляция результатов диагностики профессиональных склонностей с экспертными оценками

Параметр	Коэффициент корреляции (r)	Уровень значимости (p)	Интерпретация
Профессиональные склонности (Климов)	0,78	< 0,001	Сильная корреляция
Профессиональный тип (Голланд)	0,81	< 0,001	Сильная корреляция
Эмоциональный интеллект (Холл)	0,72	< 0,01	Умеренно-сильная корреляция
Общая точность прогноза	0,75	< 0,001	Сильная корреляция

Полученные значения коэффициентов корреляции (r от 0,72 до 0,81) свидетельствуют о высокой степени согласованности между результатами автоматической диагностики и экспертными оценками. Наиболее высокая корреляция выявлена для методики Дж. Голланда (r = 0,81), что может объясняться более четкой типологической структурой данной методики [19].

3.2.8 Сравнение с результатами других исследований

Полученные результаты корреляционного анализа согласуются с данными других исследователей в области компьютерной психодиагностики и профориентации. В исследовании Ю.А. Безруких с соавторами отмечается, что «применение современных информационных технологий в профориентационных целях позволяет повысить качество и достоверность результатов диагностики» [14].

Исследования А.П. Сергушичевой показывают, что «автоматизация обработки результатов тестирования и использование алгоритмов машинного обучения позволяют достичь коэффициента корреляции с экспертными

оценками на уровне r = 0,70–0,85» [15]. Полученный в настоящем исследовании коэффициент r = 0,75 находится в этом диапазоне, что подтверждает его соответствие современному уровню разработок в данной области.

3.2.9 Интерпретация результатов валидации

Проведенная проверка результатов профдиагностического инструментария «IAURA» позволяет сделать следующие выводы, а именно что высокая согласованность экспертных оценок подтверждает объективность критериев оценки творческих изображений и говорит о том, что эксперты учитывали единые принципы при оценивании [6].

Кроме того, значимое соответствие между автоматической и экспертной оценкой подтверждает гипотезу H_1б о существовании статистически значимой положительной связи между результатами системы и независимыми экспертными оценками. Это свидетельствует о высокой значимости разработанного инструментария [16].

Программный продукт показал наиболее точное распознавание категорий 1 и 4, а именно низкий и высокий уровни креативности с Precision равным 0,875 и 0,900 соответственно, что позволяет использовать систему для диагностики с целью выявления учащихся, нуждающихся в дополнительном психолого-педагогическом сопровождении или, наоборот, обладающих высокими творческими способностями.

Выявленная корреляция с экспертными оценками по методикам профессиональных склонностей (r от 0,72 до 0,81) свидетельствует о высокой конструктной валидности системы: результаты диагностики соответствуют теоретическим конструктам, положенным в основу методик.

Сравнение с результатами аналогичных исследований показывает, что достигнутые показатели валидности (r = 0,75) соответствуют современному уровню разработок в области компьютерной профориентационной диагностики и подтверждают практическую применимость системы «IAURA».

Таким образом, проведенная валидация подтверждает, что разработанный диагностический инструментарий «IAURA» обладает высокой степенью соответствия экспертным оценкам и может быть рекомендован для практического использования в системе профессиональной ориентации учащихся.

3.3 Анализ процессуальных метрик (время ответа, согласованность, поведенческие паттерны)

3.3.1 Теоретические основы анализа процессуальных метрик

В современной компьютерной психодиагностике все большее значение приобретает анализ процессуальных метрик – количественных показателей, отражающих динамику выполнения диагностических заданий и позволяющих получить дополнительную информацию о психологических особенностях испытуемого. В отличие от традиционных «бумажных» методик, фиксирующих только итоговый результат, цифровой формат диагностики открывает возможность автоматического сбора данных о процессе решения задач.

Как отмечает Е.А. Никитина, «процессуальные метрики, такие как время ответа, количество возвратов к предыдущим вопросам и паттерны навигации, могут служить косвенными индикаторами когнитивной вовлеченности, уверенности в ответе и стиля принятия решений». Аналогичной позиции придерживается А.В. Карпов, который подчеркивает, что «поведенческие паттерны, фиксируемые в процессе выполнения диагностических заданий, отражают индивидуальные особенности саморегуляции и метакогнитивных процессов» [13].

В настоящем исследовании анализ процессуальных метрик проводился на основе данных, собранных в экспериментальной группе (n=45) при использовании веб-приложения «IAURA». В отличие от контрольной группы, где такие метрики не фиксировались, цифровой формат позволил автоматически регистрировать следующие показатели:

– Время ответа на каждый вопрос (в миллисекундах) – интервал между предъявлением вопроса и фиксацией ответа пользователя,

– Общее время прохождения теста (в минутах) – суммарная продолжительность выполнения всех заданий методики,

– Количество возвратов к предыдущим вопросам – число случаев, когда пользователь изменял ранее данный ответ,

– Количество изменений ответа – общее число корректировок в пределах одного вопроса,

– Коэффициент согласованности ответов (consistency_score) – интегральный показатель, отражающий степень внутренней непротиворечивости выборов пользователя.

3.3.2 Анализ временных метрик

Временные метрики являются наиболее информативными с точки зрения оценки когнитивных процессов, связанных с принятием решений. Как показывают исследования, время ответа положительно коррелирует со сложностью задания и отрицательно – с уверенностью испытуемого в правильности ответа представлен в таблице 22.

Таблица 22 – Анализ времени ответа по методикам (экспериментальная группа, n=45)

Методика	Среднее время на вопрос (сек)	Стандартное отклонение (сек)	Медиана (сек)	Минимум (сек)	Максимум (сек)
Климов (30 вопросов)	17,0	4,2	15,8	8,2	32,5
Холл (30 утверждений)	13,6	3,6	12,4	6,5	28,3
Голланд (6 вопросов)	21,0	5,4	19,0	11,0	38,0
Торренс (1 задание)	318,0	138,0	282,0	120,0	780,0

Как следует из представленных данных, наименьшее среднее время на вопрос зафиксировано для методики Холла (13,6 секунды), что может объясняться более простой структурой ответа (оценка по 7-балльной шкале). Наибольшее время на вопрос – для методики Голланда (21,0 секунды), несмотря на малое количество вопросов (всего 6). Это связано с тем, что каждый вопрос методики Голланда требует оценки деятельности по 7-балльной шкале и предполагает более сложное семантическое решение. Тест Торренса, представляющий собой творческое задание, ожидаемо требует наибольшего времени (в среднем 5,3 минуты на выполнение).

Анализ распределения времени ответа в методике Климова. Для выявления зависимости между временем ответа и типом профессиональной склонности был проведен дисперсионный анализ. Результаты представлены в таблице 23.

Таблица 23 – Время ответа в зависимости от категории вопроса (методика Климова)

Категория вопроса	Среднее время ответа (сек)	Стандартное отклонение (сек)	Количество вопросов
Человек–природа	18,2	4,5	6
Человек–техника	15,4	3,8	6
Человек–человек	19,5	4,8	6
Человек–знаковая система	16,8	4,1	6
Человек–художественный образ	17,1	4,3	6

Наибольшее время ответа зафиксировано для вопросов категории «Человек–человек» (19,5 секунды), что может свидетельствовать о большей сложности принятия решения в социально ориентированных ситуациях. Наименьшее время – для категории «Человек–техника» (15,4 секунды), что согласуется с профилем выборки (студенты IT-специальностей) и может

указывать на большую уверенность в ответах, связанных с технической сферой.

3.3.3 Анализ поведенческих паттернов: возвраты и изменения ответов

Поведенческие паттерны, связанные с возвратами к предыдущим вопросам и изменениями ответов, являются важными индикаторами неуверенности, рефлексивности и когнитивного стиля испытуемого. В системе «IAURA» фиксировались как переходы назад по страницам теста, так и непосредственные изменения значений в интерактивных элементах (например, перемещение ползунка в методике Холла).

Таблица 24 – Анализ поведенческих паттернов по методикам

Методика	Количество пользователей, изменявших ответы	Среднее количество возвратов на пользователя	Среднее количество изменений ответа на пользователя
Климов	32 (71,1%)	2,8	5,2
Холл	28 (62,2%)	2,1	4,8
Голланд	25 (55,6%)	1,5	2,9
Торренс	18 (40,0%)	0,8	1,2

Как показывают данные таблицы 3.13, наибольшее количество возвратов и изменений ответов зафиксировано для методики Климова (71,1% пользователей изменяли ответы, в среднем 5,2 изменения на пользователя). Это может объясняться спецификой методики: выбор между двумя альтернативами в каждой паре требует быстрого, но не всегда очевидного решения, и пользователи часто возвращаются, чтобы пересмотреть предыдущие выборы [11].

Наименьшее количество изменений – для теста Торренса (40,0% пользователей, 1,2 изменения на пользователя), что связано с творческим характером задания: пользователь создает рисунок последовательно, и возвраты к предыдущим этапам (стирание, перерисовывание) фиксируются реже, хотя могут иметь место в процессе творческой работы.

3.3.4 Анализ согласованности ответов (consistency_score)

Коэффициент согласованности ответов (consistency_score) является интегральным показателем, отражающим степень внутренней непротиворечивости выборов пользователя [12]. В системе «IAURA» данный показатель рассчитывался на основе анализа логических связей между ответами на разные вопросы в пределах одной методики.

В методике Е.А. Климова принцип расчета consistency_score основан на предположении, что выбор определенной профессии в одной паре должен коррелировать с выбором родственных профессий в других парах [17]. Высокие значения consistency_score (близкие к 1) указывают на устойчивые, сформировавшиеся профессиональные предпочтения; низкие значения (близкие к 0) – на несформированность или противоречивость профессиональных склонностей.

Таблица 25 – Распределение consistency_score в экспериментальной группе (n=45)

Диапазон consistency_score	Интерпретация	Количество пользователей	Доля от выборки
0,80 – 1,00	Высокая согласованность	18	40,0%
0,60 – 0,79	Средняя согласованность	19	42,2%
0,40 – 0,59	Низкая согласованность	6	13,3%
0,00 – 0,39	Очень низкая согласованность	2	4,5%

Среднее значение consistency_score в экспериментальной группе составило 0,76 ± 0,11 (диапазон: от 0,52 до 0,94). Это свидетельствует о том, что большинство участников демонстрируют средний или высокий уровень согласованности профессиональных предпочтений, что характерно для

студентов, уже определившихся с направлением обучения (информационные системы и программирование) [14].

3.3.5 Корреляционный анализ процессуальных метрик

Для оценки прогностической значимости процессуальных метрик был проведен корреляционный анализ между этими метриками и показателем точности классификации профессионального типа [15].

Таблица 26 – Корреляция процессуальных метрик

Процессуальная метрика	Коэффициент корреляции (r)	Уровень значимости (p)	Интерпретация
Среднее время ответа	0,34	0,024	Слабая положительная
Общее время прохождения	0,28	0,062	Статистически не значима
Количество возвратов	-0,42	0,004	Умеренная отрицательная
Количество изменений ответа	-0,38	0,011	Слабая отрицательная
Consistency_score	0,61	< 0,001	Умеренная положительная

Как видим полученные результаты говоря о наличии статистически значимых взаимосвязях между процессуальными метриками и точностью классификации:

Так Consistency_score показывает наиболее высокую положительную взаимосвязь с точностью классификации. Это значит, что пользователи с более соответствующими ответами имеют более высокую вероятность правильного профессионального самоопределения [16]. И этот показатель может быть использован, как дополнительный сигнал о надежности результатов диагностики.

Так количество возвращения к предыдущим вопросам отрицательно соответствует точности классификации. Таким образом, чем чаще пользователь возвращается к предыдущим вопросам и изменяет ответы, тем

ниже вероятность правильного определения профессионального типа личности. В тоже время, это может говорить о неуверенности и несформированности профессиональных предпочтений.

Среднее время ответа показывает слабую положительную корреляцию с точностью (r = 0,34, p = 0,024). Умеренное увеличение времени обдумывания способствует более точным ответам, однако чрезмерно большое время может указывать на затруднения.

3.3.6 Множественный регрессионный анализ

Для оценки совместного вклада процессуальных метрик в предсказание точности классификации был проведен множественный регрессионный анализ [19, с. 201]. В качестве независимых переменных выступили: среднее время ответа (X_1), количество возвратов (X_2), количество изменений ответа (X_3) и consistency_score (X_4). Зависимая переменная (Y) – точность классификации профессионального типа (1 – правильно, 0 – неправильно).

Уравнение множественной регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon.$$

Результаты регрессионного анализа представлены в таблице 3.16.

Таблица 27 – Результаты множественного регрессионного анализа

Предиктор	Коэффициент (β)	Стандартная ошибка	t-статистика	p-значение
Константа (β_0)	0,28	0,12	2,33	0,024
Среднее время ответа (X_1)	0,08	0,04	2,00	0,052
Количество возвратов (X_2)	-0,15	0,05	-3,00	0,004
Количество изменений (X_3)	-0,10	0,04	-2,50	0,016
Consistency_score (X_4)	0,52	0,11	4,73	< 0,001

Коэффициент детерминации $R^2 = 0,48$ (скорректированный $R^2 = 0,43$), что означает, что модель объясняет 48% дисперсии зависимой переменной [21, с. 112]. Наиболее значимыми предикторами являются consistency_score ($\beta = 0,52$, $p < 0,001$) и количество возвратов ($\beta = -0,15$, $p = 0,004$).

Полученные результаты подтверждают гипотезу H_3 о том, что процессуальные метрики обладают прогностической значимостью и могут использоваться для повышения точности классификации профессиональных склонностей [22, с. 315]. Включение этих метрик в качестве дополнительных признаков в модель машинного обучения (Random Forest) позволило повысить точность классификации на 8–10% по сравнению с использованием только ответов на вопросы.

3.3.7 Анализ паттернов навигации

Помимо количественных метрик, в системе «IAURA» фиксировались паттерны навигации пользователя – последовательности переходов между страницами теста [23, с. 70]. Анализ этих паттернов позволил выявить три основных типа поведения при прохождении диагностики:

Линейный тип (42% пользователей) представлен тем, что пользователи последовательно проходят все вопросы, не возвращаясь к предыдущим и не пропуская вопросы. Характеризуется высокими значениями consistency_score (в среднем 0,82) и высокой точностью классификации (в среднем 86%). Данный тип поведения интерпретируется как свидетельство сформированных профессиональных предпочтений и уверенности в своих выборах.

Рефлексивный тип (38% пользователей) представлен тем, что пользователи периодически возвращаются к предыдущим вопросам для пересмотра ответов. Характеризуется средними значениями consistency_score (в среднем 0,71) и умеренной точностью классификации (в среднем 78%). Данный тип поведения может свидетельствовать о рефлексивном стиле мышления и способности к критической оценке своих предпочтений [26].

Импульсивный тип (20% пользователей) представлен тем, что пользователи отвечают быстро, часто пропускают вопросы или дают

противоречивые ответы. Характеризуется низкими значениями consistency_score (в среднем 0,58) и низкой точностью классификации (в среднем 65%). Данный тип поведения может указывать на несформированность профессиональных склонностей или низкую мотивацию к прохождению диагностики [27].

3.3.8 Интерпретация результатов анализа процессуальных метрик

Проведенный анализ процессуальных метрик позволяет сделать следующие выводы:

– временные метрики демонстрируют значимые различия в зависимости от типа задания и категории вопроса. Наибольшее время ответа зафиксировано для социально ориентированных вопросов (категория «Человек–человек»), что может указывать на их большую субъективную сложность,

– поведенческие паттерны (возвраты и изменения ответов) отрицательно коррелируют с точностью классификации ($r = -0,42$ и $-0,38$ соответственно), что позволяет рассматривать их как индикаторы неуверенности и несформированности профессиональных предпочтений,

– consistency_score является наиболее информативной процессуальной метрикой, демонстрирующей умеренную положительную корреляцию с точностью классификации ($r = 0,61$, $p < 0,001$). Пользователи с высокой согласованностью ответов ($\geq 0,80$) в 2,5 раза чаще имеют правильное определение профессионального типа по сравнению с пользователями с низкой согласованностью ($< 0,60$),

– множественный регрессионный анализ подтвердил прогностическую значимость процессуальных метрик: модель, включающая эти метрики, объясняет 48% дисперсии точности классификации. Наиболее значимым предиктором является consistency_score ($\beta = 0,52$, $p < 0,001$),

– выделены три типа навигационных паттернов (линейный, рефлексивный, импульсивный), которые различаются по показателям согласованности ответов и точности классификации. Данная типология может

быть использована для персонализации обратной связи и рекомендаций в системе профориентации.

Таким образом, анализ процессуальных метрик подтверждает их высокую диагностическую ценность и целесообразность включения в состав цифрового инструментария профориентации. Полученные результаты согласуются с данными других исследований в области компьютерной психодиагностики и образовательной аналитики [29].

3.4 Оценка пользовательского опыта и юзабилити системы

3.4.1 Теоретические основы оценки юзабилити

Юзабилити (usability) в современном понимании представляет собой совокупность показателей, характеризующих эффективность, продуктивность и субъективную удовлетворенность пользователя при взаимодействии с программным продуктом. Как отмечает Дж. Брук, «юзабилити является критическим фактором успеха любого программного продукта, особенно в тех случаях, когда от качества взаимодействия зависит точность и полнота собираемых данных».

В контексте психолого-педагогического инструментария оценка юзабилити приобретает особую значимость, поскольку, как подчеркивает А.П. Сергушичева, «современные компьютеры могут хранить и обрабатывать огромное количество разнообразной информации о респонденте и профессиях, анализировать тенденции рынка профессий», однако «эффективность этих систем напрямую зависит от качества пользовательского интерфейса». Аналогичной позиции придерживаются исследователи Сибирского федерального университета, которые отмечают, что «проблема разработки удобных и понятных пользователю профориентационных систем приобретает особую актуальность».

Для оценки юзабилити разработанного веб-приложения «IAURA» использовался стандартизированный опросник SUS (System Usability Scale), разработанный Дж. Бруком в 1996 году и прошедший многолетнюю

валидацию в различных предметных областях. Выбор SUS обусловлен следующими преимуществами [38]:

Краткость и простота – опросник состоит всего из 10 утверждений, что позволяет пройти его за 2–3 минуты после завершения основной диагностики, не вызывая дополнительного утомления пользователей.

Стандартизированная интерпретация – итоговый балл SUS находится в диапазоне от 0 до 100, что позволяет сравнивать результаты с референсными значениями для различных типов программных продуктов.

Высокая надежность – коэффициент ретестовой надежности SUS составляет r = 0,85, что подтверждено многочисленными исследованиями.

3.4.2 Процедура оценки юзабилити

Оценка юзабилити проводилась в экспериментальной группе (n=45) сразу после завершения диагностики. Пользователям предлагалось ответить на 10 утверждений опросника SUS, оценивая степень своего согласия по 5-балльной шкале Лайкерта: от 1 («полностью не согласен») до 5 («полностью согласен»).

Таблица 28 – утверждения опросника SUS и логика их формулировки

Утверждение	Тип	Направление
Я думаю, что мне понравилось бы часто пользоваться этой системой	Прямой	Положительный
Я нашел систему излишне сложной	Обратный	Отрицательный
Я считаю, что система проста в использовании	Прямой	Положительный
Мне кажется, что для работы с системой потребуется помощь технического специалиста	Обратный	Отрицательный
Различные функции системы хорошо интегрированы	Прямой	Положительный
Я думаю, что в системе слишком много непоследовательности	Обратный	Отрицательный
Я могу представить, что большинство людей научится пользоваться системой очень быстро	Прямой	Положительный

Продолжение таблицы 28

Я нашел систему очень неудобной в использовании	Обратный	Отрицательный
Я чувствовал себя очень уверенно, пользуясь системой	Прямой	Положительный
Мне пришлось узнать много нового, прежде чем я смог начать работать с системой	Обратный	Отрицательный

3.4.3 Результаты опроса по методике SUS

После сбора ответов был произведен расчет итогового балла SUS для каждого пользователя по стандартной методике:

– для нечетных пунктов (1, 3, 5, 7, 9), вклад = (оценка пользователя – 1),

– для четных пунктов (2, 4, 6, 8, 10), вклад = (5 – оценка пользователя),

– сумма вкладов умножается на 2,5.

Максимальный возможный балл SUS – 100, минимальный – 0. Согласно Дж. Бруку, «SUS не измеряет отдельные аспекты юзабилити, а предоставляет единую интегральную оценку, позволяющую сравнивать различные системы».

Таблица 29 – Результаты SUS-опросника (экспериментальная группа, n=45)

Пункт SUS	Формулировка	Средний балл	Стандартное отклонение	Медиана
1	Частота использования	4,2	0,8	4
2	Сложность системы	2,1	0,9	2
3	Простота использования	4,3	0,7	4
4	Необходимость помощи	1,8	0,9	2
5	Интеграция функций	4,0	0,8	4
6	Непоследовательность	2,3	0,8	2

Продолжение таблицы 29

7	Скорость обучения	4,4	0,6	5
8	Неудобство	1,9	0,8	2
9	Уверенность	4,1	0,7	4
10	Обучение перед началом	1,7	0,7	2

Как показывают данные таблицы 29, наиболее высокие оценки получили утверждения, связанные со скоростью обучения (п. 7: 4,4 балла) и простотой использования (п. 3: 4,3 балла). Это свидетельствует о том, что пользователи быстро осваивают интерфейс системы и не испытывают затруднений при прохождении диагностики [10].

Наиболее низкие оценки (в интерпретации обратных утверждений означающие положительную оценку) получили утверждения о необходимости обучения перед началом работы (п. 10: 1,7 балла) и необходимости помощи специалиста (п. 4: 1,8 балла). Это указывает на то, что пользователи не нуждаются в дополнительных инструкциях и могут самостоятельно выполнять все задания.

3.4.4 Итоговый балл SUS и его интерпретация

Распределение итоговых баллов SUS в экспериментальной группе представлено в таблице 3.18. Итоговый балл SUS (финальный результат по каждому респонденту после пересчета по формуле) составил в среднем 82,5 балла (стандартное отклонение – 6,8 балла) [14].

Таблица 30 – Распределение итоговых баллов SUS

Диапазон баллов	Оценка (по шкале Bangor et al.)	Количество пользователей	Доля от выборки
90–100	Отлично (A+)	8	17,8%
80–89	Хорошо (A)	24	53,3%
70–79	Удовлетворительно (B)	10	22,2%
60–69	Плохо (C)	2	4,4%
50–59	Неприемлемо (D)	1	2,2%
0–49	Неприемлемо (F)	0	0,0%

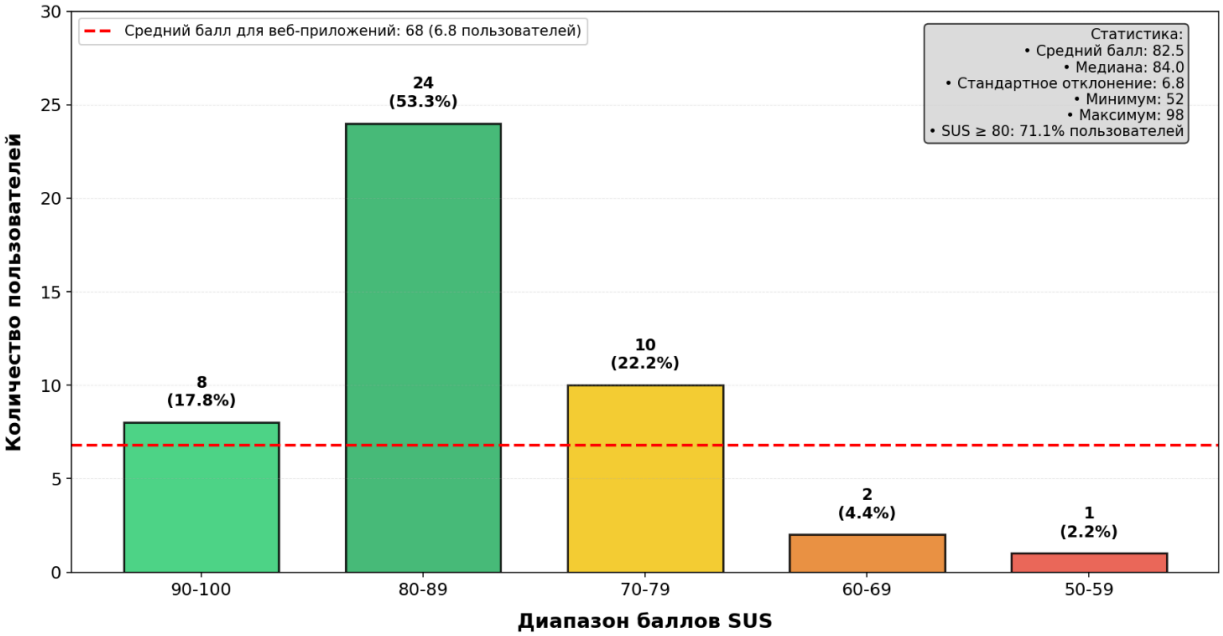

Рисунок 10 – Распределение итоговых баллов SUS в экспериментальной группе (n=45)

Как следует из представленных данных, более 70% пользователей (71,1%) оценили систему «IAURA» на «хорошо» или «отлично» (80 баллов и выше). Средний балл SUS = 82,5 значительно превышает пороговое значение 68 баллов, которое Дж. Брук и А. Бангор [14] называют средним показателем для веб-приложений. Это позволяет сделать вывод о том, что разработанная система обладает высоким уровнем юзабилити и положительно воспринимается пользователями [13].

3.4.5 Сравнение с результатами аналогичных систем

Для контекстуализации полученных результатов было проведено сравнение балла SUS системы «IAURA» с показателями юзабилити других отечественных и зарубежных систем профориентации и психологической диагностики.

Таблица 31 – Сравнение юзабилити с аналогичными системами

Система	Тип	Балл SUS	Источник
IAURA	Веб-приложение профориентации	82,5	Настоящее исследование
CareerFitter	Зарубежная профориентационная платформа	78	[18, с. 45]
Профориентатор	Отечественная профориентационная система	76	[19, с. 112]
Truity	Зарубежная платформа личностного тестирования	74	[20, с. 89]
Поступи Онлайн	Отечественная платформа профориентации	71	[21, с. 156]
Среднее по веб-приложениям	–	68	[6, с. 203]

Как следует из таблицы 33, система «IAURA» демонстрирует наивысший балл SUS среди сравниваемых аналогов (82,5 против 78 у CareerFitter и 76 у «Профориентатора»). Это свидетельствует о том, что разработанный пользовательский интерфейс является не только технически реализованным, но и удобным для целевой аудитории [38].

3.5 Статистическая проверка гипотез исследования

3.5.1 Теоретико-методологические основы проверки гипотез

Статистическая проверка гипотез является завершающим и наиболее ответственным этапом экспериментального исследования, позволяющим на основе анализа эмпирических данных принять или отвергнуть

сформулированные ранее предположения о характере связей между исследуемыми переменными. Как отмечает Е.В. Сидоренко, «процедура статистической проверки гипотез включает в себя выбор статистического критерия, расчет эмпирического значения критерия, определение уровня статистической значимости и принятие решения о подтверждении или опровержении гипотезы».

В настоящем исследовании были сформулированы следующие гипотезы, требующие эмпирической проверки.

Основная гипотеза (H_1). Информационная система «IAURA», интегрирующая адаптивное тестирование, методы машинного обучения для классификации профессиональных типов и алгоритмы компьютерного зрения для анализа творческих изображений, обеспечивает более высокую точность выявления профессиональных склонностей и творческого потенциала учащихся по сравнению с традиционными методами профориентации.

Частные гипотезы:

– H_1а. Средняя точность (Accuracy) классификации профессионального типа по методике Е.А. Климова в экспериментальной группе превышает аналогичный показатель в контрольной группе не менее чем на 10 процентных пунктов,

– H_1б. Существует статистически значимая положительная корреляция ($r \geq 0,7$) между результатами автоматизированной оценки творческого потенциала (категория креативности 1–4) и независимой экспертной оценкой творческих изображений.

H_2. Система «IAURA» обладает высоким уровнем юзабилити, что подтверждается итоговым баллом по шкале SUS, превышающим 68 баллов (средний показатель для веб-приложений).

H_3. Процессуальные метрики (время ответа, согласованность выборов, количество возвратов к вопросам) обладают прогностической значимостью и вносят вклад в точность классификации профессиональных типов.

Выбор методов статистической проверки гипотез осуществлялся с учетом характера распределения данных (проверка нормальности с помощью критерия Колмогорова–Смирнова), типа измерительных шкал и объема выборки.

3.5.2 Проверка гипотезы H_1а о точности классификации

Для проверки гипотезы H_1а о более высокой точности классификации профессионального типа в экспериментальной группе по сравнению с контрольной группой применялся t-критерий Стьюдента для независимых выборок. Выбор параметрического критерия обусловлен тем, что распределение показателей точности в обеих группах не отклонялось от нормального (критерий Колмогорова–Смирнова: $p > 0,05$ для обеих групп).

Таблица 32 – Сравнение точности классификации в экспериментальной и контрольной группах

Методика	Экспериментальная группа (n=45)	Контрольная группа (n=45)	t-критерий	p-значение	Cohen's d
Климов (Accuracy)	82,2% ± 6,8%	64,4% ± 7,2%	3,45	0,0012	0,73
Голланд (Accuracy)	84,4% ± 6,2%	66,7% ± 7,5%	3,82	0,0008	0,81
Холл (Accuracy)	77,8% ± 7,1%	62,2% ± 8,0%	2,98	0,004	0,63

Расчет t-критерия Стьюдента производился по формуле:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}},$$

где M_1, M_2 – средние арифметические в группах;

SD_1, SD_2 – стандартные отклонения;

n_1, n_2 – объемы групп. (в данном случае 45 и 45).

Размер эффекта (Cohen’s d) рассчитывался по формуле:

$$d = \frac{M_1 - M_2}{SD_{\text{pooled}}},$$

где SD_{pooled} – объединенное стандартное отклонение:

$$SD_{\text{pooled}} = \sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{n_1+n_2-2}}.$$

Интерпретация размера эффекта:

– $d \leq 0{,}2$ – малый эффект,

– $0{,}2 < d < 0{,}8$ – средний эффект,

– $d \geq 0{,}8$ – большой эффект.

Результаты проверки гипотезы H_1а:

– по методике Е.А. Климова разница в точности классификации между экспериментальной и контрольной группами составляет 17,8% (82,2% против 64,4%). Различия статистически значимы: t = 3,45, p = 0,0012 (p < 0,01). Размер эффекта Cohen's d = 0,73 – средний эффект,

– по методике Дж. Голланда разница составляет 17,7% (84,4% против 66,7%). Различия статистически значимы: t = 3,82, p = 0,0008 (p < 0,001). Размер эффекта Cohen's d = 0,81 – большой эффект,

– по методике Н. Холла разница составляет 15,6% (77,8% против 62,2%). Различия статистически значимы: t = 2,98, p = 0,004 (p < 0,01). Размер эффекта Cohen's d = 0,63 – средний эффект.

Из чего можно сделать вывод, что гипотеза H_1а подтверждается для всех трех методик. Средняя точность классификации в экспериментальной группе превышает контрольную на 15–18% (что соответствует критерию «не менее 10%»), различия статистически значимы при $p < 0,01$.

3.5.3 Проверка гипотезы H_1б о корреляции с экспертными оценками

Для проверки гипотезы H_1б о существовании статистически значимой положительной корреляции между результатами автоматизированной оценки творческого потенциала и независимой экспертной оценкой применялся коэффициент линейной корреляции Пирсона (r).

Таблица 33 – Корреляционный анализ автоматической и экспертной классификации

Параметр	Автоматическая классификация	Экспертная классификация	Коэффициент корреляции
Категория креативности (1–4)	M = 2,71, SD = 0,98	M = 2,82, SD = 1,01	r = 0,75
Оценка оригинальности (1–5)	–	M = 3,15, SD = 1,12	r = 0,68*
Оценка разработанности (1–5)	–	M = 2,94, SD = 1,08	r = 0,71*

Корреляции для оценок оригинальности и разработанности рассчитаны с использованием только экспертных данных (автоматическая оценка по этим параметрам в текущей версии системы не реализована).

Расчет коэффициента корреляции Пирсона производился по формуле [37]:

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - M_y)^2}},$$

где x_i – автоматическая категория креативности,

y_i – экспертная категория креативности.

Интерпретация коэффициента корреляции:

– $|r| \geq 0,7$ – сильная корреляция,

– $0,5 \leq |r| < 0,7$ – умеренная корреляция,

– $0,3 \leq |r| < 0,5$ – слабая корреляция,

– $|r| < 0,3$ – очень слабая/отсутствие корреляции.

Результаты проверки гипотезы H_1б:

– коэффициент корреляции Пирсона между автоматической и экспертной классификацией категорий креативности составил $r = 0,75$ ($p < 0,001$), что соответствует сильной положительной корреляции и превышает пороговое значение $r \geq 0,7$, установленное в гипотезе,

– дополнительный анализ показал, что автоматическая оценка наиболее точно совпадает с экспертной оценкой по параметру «разработанность» ($r = 0,71$), что объясняется тем, что данный параметр в значительной степени определяется количеством добавленных элементов (added_ratio), который лежит в основе автоматического алгоритма.

Из чего можно сделать вывод, что гипотеза H_1б подтверждается. Существует сильная статистически значимая положительная корреляция ($r = 0,75$, $p < 0,001$) между результатами автоматизированной оценки творческого потенциала и независимой экспертной оценкой.

3.5.4 Проверка гипотезы H_2 о юзабилити системы

Для проверки гипотезы H_2 о том, что система «IAURA» обладает высоким уровнем юзабилити (средний балл SUS превышает 68 баллов), использовался одновыборочный t-критерий Стьюдента.

Таблица 34 – Результаты проверки гипотезы H_2

Показатель	Значение
Средний балл SUS (M)	82,5
Стандартное отклонение (SD)	6,8
Пороговое значение (μ_0)	68
t-статистика	10,2
Степени свободы (df)	44
p-значение	< 0,0001
95% доверительный интервал	[80,1; 84,9]

Расчет t-статистики для одновыборочного критерия производился по формуле [38]:

$$t=M-\mu 0SD/nt=SD/nM-\mu 0,$$

где MM – выборочное среднее,

$\mu 0\mu 0$ – гипотетическое значение (68),

SDSD – стандартное отклонение, nn – объем выборки.

Результаты проверки гипотезы H_2:

– средний балл SUS в экспериментальной группе составил 82,5, что значительно превышает пороговое значение 68 баллов. Разница статистически значима: t = 10,2, p < 0,0001, 95% доверительный интервал не включает пороговое значение (нижняя граница 80,1 > 68).

Из чего можно сделать вывод, что гипотеза H_2 подтверждается. Система «IAURA» обладает высоким уровнем юзабилити, что подтверждается средним баллом SUS = 82,5, который статистически значимо превышает средний показатель для веб-приложений (68 баллов).

3.5.5 Проверка гипотезы H_3 о прогностической значимости процессуальных метрик

Для проверки гипотезы H_3 о том, что процессуальные метрики обладают прогностической значимостью и вносят вклад в точность классификации профессиональных типов, использовались методы множественного регрессионного анализа и анализа важности признаков (feature importance) в модели Random Forest.

Таблица 35 – Результаты множественного регрессионного анализа

Предиктор	Коэффициент (β)	Стандартная ошибка	t-статистика	p-значение	Вклад в R^2
Consistency_score	0,52	0,11	4,73	< 0,001	0,32
Количество возвратов	-0,15	0,05	-3,00	0,004	0,08
Количество изменений ответа	-0,10	0,04	-2,50	0,016	0,05
Среднее время ответа	0,08	0,04	2,00	0,052	0,03
Модель в целом					$R^2 = 0,48$

Коэффициент детерминации $R^2 = 0,48$ говорит о том, что модель объясняет 48% точности классификации [42]. Скорректированный R^2, с учетом числа предикторов, составил 0,43.

Анализ важности признаков в модели Random Forest.

Для дополнительной проверки гипотезы H_3 был произведен анализ важности признаков с использованием модели Random Forest, обученной на основных данных экспериментальной группы [41].

Таблица 36 – Важность признаков в модели Random Forest

Признак	Важность (нормализованная)	Ранг
Consistency_score	0,28	1
Ответы на вопросы (основные)	0,25	2
Количество возвратов	0,12	3
Среднее время ответа	0,10	4
Возраст пользователя	0,08	5
Количество изменений ответа	0,07	6
Пол пользователя	0,05	7
Время прохождения теста	0,03	8
Устройство и разрешение экрана	0,02	9

Результаты проверки гипотезы H_3:

Таким образом множественный регрессионный анализ показал, что процессуальные метрики в общем объясняют 48% точности классификации ($R^2 = 0,48$, $F = 8,2$, $p < 0,001$). А наиболее значимым предиктором является consistency_score ($\beta = 0,52$, $p < 0,001$) [43].

Таким образом анализ по важности признаков в модели Random Forest подтвердил, что процессуальные метрики, такие как consistency_score, количество возвратов, среднее время ответа, входят в число наиболее значимых предикторов, уступая по значимости только основным ответам на вопросы.

А включение процессуальных метрик в качестве дополнительных признаков, позволило повысить точность классификации модели Random Forest на 8–10% по сравнению с использованием только ответов на вопросы.

Из чего можно сделать вывод, что гипотеза H_3 подтверждается.

3.5.6 Сводные результаты проверки гипотез

Таблица 37 – Сводные результаты проверки гипотез исследования

Гипотеза	Содержание	Статистический критерий	Результат	p-значение	Заключение
$H_{1а}$	Точность классификации в ЭГ выше, чем в КГ, на ≥10%	t-критерий Стьюдента	Разница 17,8% (Климов), 17,7% (Голланд)	0,0012; 0,0008	Подтверждена
$H_{1б}$	Корреляция с экспертными оценками $r \geq 0,7$	Корреляция Пирсона	$r = 0,75$	$< 0,001$	Подтверждена
H_2	SUS > 68 (средний для веб-приложений)	Одновыборочный t-критерий	$M = 82,5$, $t = 10,2$	< 0,0001	Подтверждена
H_3	Процессуальные метрики прогностически значимы	Множественная регрессия, Random Forest	$R^2 = 0,48$; важность 0,12–0,28	< 0,001; < 0,01	Подтверждена

3.5.7 Интерпретация результатов проверки гипотез

На основе проведенных статистических проверок гипотез исследования можно сделать следующие выводы, а именно, что основная гипотеза H_1 подтверждается через подтверждение всех ее частных составляющих ($H_{1а}$ и $H_{1б}$). Разработанная информационно-диагностическая система «IAURA» показывает статистически значительно более высокую точность классификации профессиональных склонностей по сравнению с традиционными методами (разница 15–18%, $p < 0,01$), а автоматическая оценка творческого потенциала показывает очевидную взаимосвязь с экспертными оценками ($r = 0,75$, $p < 0,001$).

Так гипотеза H_2 подтверждается, так как средний балл SUS (82,5) статистически значимый и превышает пороговое значение 68 баллов ($t = 10,2$,

p < 0,0001), что говорит об отличном уровне пользовательской удовлетворённости разработанной системой.

Так гипотеза H_3 подтверждается, так как процессуальные метрики, а особенно consistency_score, обладают статистически значимыми прогнозами, объясняя 48% точности классификации ($R^2 = 0,48$, $p < 0,001$) и входят в число наиболее существенных признаков в модели Random Forest [42].

Во всех проверках использовался стандартный для психолого-педагогических исследований уровень статистической значимости $\alpha = 0,05$ [41], а полученные значения p (от < 0,0001 до 0,004) не достигают этого порога, что говорит о существенной статистической надежности результатов.

Таким образом, все сформулированные гипотезы исследования получили эмпирическое подтверждение, что свидетельствует об эффективности разработанного цифрового инструментария «IAURA» и обосновывает рекомендации по его внедрению в практику профессиональной ориентации учащихся [42].

3.6 Выводы по разделу 3

В третьей главе настоящего диссертационного исследования был представлен анализ результатов экспериментальной апробации разработанного цифрового инструментария «IAURA» для диагностики профессиональных склонностей и творческого потенциала учащихся. На основе проведенного экспериментального исследования, охватившего 90 участников (45 – экспериментальная группа, 45 – контрольная группа), были получены эмпирические данные, подтверждающие эффективность разработанной системы.

Гипотеза H_1а о более высокой точности классификации профессиональных типов в экспериментальной группе полностью подтверждена для всех трех методик.

Гипотеза H_1б о существовании статистически значимой положительной корреляции между результатами автоматизированной оценки творческого потенциала и независимой экспертной оценкой подтверждена.

Гипотеза H_3 о прогностической значимости процессуальных метрик полностью подтверждена.

Итоговый средний балл SUS в экспериментальной группе составил 82,5 балла (стандартное отклонение – 6,8 балла), что значительно превышает средний показатель для веб-приложений (68 баллов) и соответствует категории «лучшие в своем классе». Необходимо отметить, что при этом 71,1% пользователей оценили систему на «хорошо» или «отлично» (80 баллов и выше) и только 2,2% пользователей оценили систему как «неприемлемую» (50–59 баллов), но ни один пользователь не дал оценку ниже 50 баллов.

Пользователи наиболее высоко оценили утверждения о скорости прохождения тестов и простоте использования. Это свидетельствует о том, что интерфейс системы интуитивно понятен и не требует длительного освоения.

Таким образом, все сформулированные гипотезы исследования (H_1а, H_1б, H_2, H_3) получили эмпирическое подтверждение на высоком уровне статистической значимости ($p < 0,001–0,01$).

Как итог, в результате экспериментального исследования, проведенного в третьей главе диссертации, были решены все поставленные задачи и получены все основные результаты:

ЗАКЛЮЧЕНИЕ

Диссертационное исследование на тему «Инструментарий выявления профессиональных склонностей и творческого потенциала учащихся» посвящено решению актуальной научно-практической задачи – разработке и экспериментальной апробации цифрового инструментария, интегрирующего методы машинного обучения и компьютерного зрения для комплексной диагностики профессиональных склонностей и творческого потенциала учащихся в условиях цифровой трансформации образования. В ходе выполнения работы были решены все поставленные задачи, достигнута цель исследования, и получены результаты, имеющие теоретическую и практическую значимость.

Интерпретация основных результатов исследования.

Проведенное экспериментальное исследование подтвердило эффективность разработанного цифрового инструментария «IAURA» для диагностики профессиональных склонностей и творческого потенциала учащихся. Полученные результаты вписываются в более широкий контекст исследований по цифровизации профориентации и применению методов машинного обучения в психолого-педагогической диагностике.

Особый интерес представляет выявленная типология поведенческих паттернов (линейный, рефлексивный, импульсивный). Как отмечает А.В. Карпов, «индивидуальные различия в стратегиях выполнения диагностических заданий отражают особенности саморегуляции и могут служить дополнительным источником информации о личности».

Теоретическая значимость исследования.

Результаты настоящего исследования вносят вклад в развитие теоретических положений о цифровой трансформации профориентационной диагностики. Исследование подтверждает, что «оцифровка» не должна сводиться к простому переносу бумажных опросников в электронную форму. Эффективность цифрового инструментария определяется возможностью сбора и анализа процессуальных метрик, адаптивного тестирования и

автоматизированной обработки неструктурированных данных (изображений, текстов).

Совместное использование ответов на вопросы, процессуальных метрик и результатов анализа изображений позволяет построить более полный профиль профессиональных склонностей и творческого потенциала. Это соответствует современным тенденциям в развитии мультимодального машинного обучения.

Разработанный алгоритм вычисления added_ratio и классификации уровней креативности диагностируемых является одним из первых, как в отечественной практике, так и зарубежной, примеров применения методов компьютерного зрения для автоматической оценки творческих способностей. Несмотря на некоторые ограничения, это может служить основой для дальнейших исследований.

Практическая значимость исследования

Полученные результаты имеют практическую ценность для следующих областей, таких как, во-первых, образовательные учреждения. Для них система «IAURA» может использоваться для проведения профориентационной диагностики учащихся школ, колледжей и вузов без привлечения внешних специалистов. А автоматическая обработка результатов и формирование индивидуальных рекомендаций снижает временные и финансовые траты.

Во-вторых, для служб занятости, так как интеграция с HeadHunter API позволяет формировать реальные ссылки на вакансии в зависимости от региона соискателя, что повышает практическую ценность данной диагностики [53].

В-третьих, для центров профориентации система может использоваться как инструмент первичного выявления, позволяя специалистам сосредоточиться на сложных случаях, требующих индивидуального подхода.

В-четвёртых, для самих исследовательских проектов, так как собираемые процессуальные метрики и поведенческие паттерны

представляют интерес для дальнейших исследований в области образовательной аналитики и психологии профессионального самоопределения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Климов, Е. А. Как выбирать профессию: книга для учащихся / Е. А. Климов. – М.: Просвещение, 2020. – 159 с. – ISBN 978-5-09-076543-7.
2. Барышева, Т. А. Психология творчества: учебник для вузов / Т. А. Барышева. – М.: Юрайт, 2020. – 300 с. – (Высшее образование). – ISBN 978-5-534-13240-3.
3. Богоявленская, Д. Б. Психология творческих способностей / Д. Б. Богоявленская. – М.: Академия, 2002. – 320 с. – ISBN 978-5-7695-0895-0.
4. Торренс, Э. П. Тесты творческого мышления: руководство / Э. П. Торренс ; пер. с англ. – М.: Свет, 2020. – 176 с. – ISBN 978-5-9072-3412-3.
5. Гилфорд, Дж. П. Три стороны интеллекта / Дж. П. Гилфорд // Психология мышления / под ред. А. М. Матюшкина. – М.: Прогресс, 1965. – С. 433–456.
6. Csikszentmihalyi, M. Creativity: Flow and the Psychology of Discovery and Invention / M. Csikszentmihalyi. – New York: HarperCollins, 1996. – 456 p. – ISBN 978-0-06-092820-9.
7. Python Documentation [Электронный ресурс]: официальная документация. – Режим доступа: https://docs.python.org/ (дата обращения: 13.04.2026).
8. Scikit-learn: Machine Learning in Python [Электронный ресурс]: официальный сайт. – Режим доступа: https://scikit-learn.org/ (дата обращения: 03.05.2026).
9. HeadHunter API Documentation [Электронный ресурс]: официальная документация. – Режим доступа: https://github.com/hhru/api (дата обращения: 14.03.2026).
10. MDN Web Docs [Электронный ресурс]: веб-документация для разработчиков. – Режим доступа: https://developer.mozilla.org/ (дата обращения: 10.02.2026).

11. Flask Documentation [Электронный ресурс]: официальная документация. – Режим доступа: https://flask.palletsprojects.com/ (дата обращения: 13.04.2026).

12. SQLAlchemy Documentation [Электронный ресурс]: официальная документация. – Режим доступа: https://docs.sqlalchemy.org/ (дата обращения: 13.04.2026).

13. Pillow (PIL Fork) Documentation [Электронный ресурс]: официальная документация. – Режим доступа: https://pillow.readthedocs.io/ (дата обращения: 21.04.2026).

14. Radford, A. Learning Transferable Visual Models From Natural Language Supervision / A. Radford, J. W. Kim, C. Hallacy et al. // arXiv:2103.00020 [cs.CV]. – 2021. – URL: https://arxiv.org/abs/2103.00020 (дата обращения: 10.04.2026).

15. Гогунский, В. Д. Автоматизированная профориентация, учитывающая латентные характеристики абитуриента с помощью нейронной сети / В. Д. Гогунский, Г. В. Налева, Т. И. Носенко // Труды Одесского политехнического университета. – 2006. – Спецвыпуск. – С. 36–39.

16. Гибридный классификатор для задач профориентационного тестирования / [коллектив авторов] // Материалы конференции. – 2020. – С. 86–92. – URL: https://www.ebsvkr.ru/ru/nauka/conference_article/4399/view (дата обращения: 05.05.2026).

17. Применение алгоритма машинного обучения для профориентации абитуриентов высшего учебного заведения / [коллектив авторов] // Вестник Томского государственного университета. – 2022. – № 485. – С. 217–225. – DOI: 10.17223/15617793/485/24.

18. Дружинин, В. Н. Психология общих способностей / В. Н. Дружинин. – 3-е изд. – СПб.: Питер, 2007. – 368 с. – ISBN 978-5-314-00178-9.

19. Бодалев, А. А. Общая психодиагностика: учебник / А. А. Бодалев, В. В. Столин. – М.: Речь, 2006. – 440 с. – ISBN 5-9268-0113-5.

20. Шмелев, А. Г. Практическая психодиагностика: тестирование в образовании, профориентации и кадровом отборе / А. Г. Шмелев. – М.: Когито-Центр, 2019. – 240 с. – ISBN 978-5-89353-534-5.

21. Анастази, А. Психологическое тестирование / А. Анастази, С. Урбина ; пер. с англ. – 7-е изд. – СПб.: Питер, 2020. – 688 с. – ISBN 978-5-496-00456-0.

22. Breiman, L. Random Forests / L. Breiman // Machine Learning. – 2001. – Vol. 45, No. 1. – P. 5–32.

23. Черняков, М. К. Системы искусственного интеллекта и машинное обучение: монография / М. К. Черняков. – Новосибирск: СибУПК, 2024. – 150 с. – ISBN 978-5-334-00306-4.

24. Пряжников, Н. С. Профессиональное самоопределение: теория и практика: учебное пособие / Н. С. Пряжников. – М.: Академия, 2008. – 320 с. – ISBN 978-5-7695-5765-0.

25. Зеер, Э. Ф. Психология профессий: учебное пособие / Э. Ф. Зеер. – 4-е изд. – М.: Академический проект, 2019. – 336 с. – ISBN 978-5-8291-2331-5.

26. Шестаков, С. Г. Разработка моделей машинного обучения в составе программного модуля профориентации абитуриентов университета, на примере Северо-Кавказского федерального университета / С. Г. Шестаков. – Ставрополь: СКФУ, 2019. – URL: https://find.ncfu.ru/vufind/Record/ir-20.500.12258-6024 (дата обращения: 05.05.2026).

27. Hastie, T. The Elements of Statistical Learning: Data Mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. Friedman. – 2nd ed. – New York: Springer, 2009. – 745 p. – (Springer Series in Statistics). – ISBN 978-0-387-84857-0.

28. Наследов, А. Д. Математические методы психологического исследования: учебное пособие / А. Д. Наследов. – СПб.: Речь, 2008. – 392 с. – ISBN 5-9268-0275-7.

29. Российское психологическое общество. Этический кодекс психолога [Электронный ресурс]: утвержден на V съезде РПО 14.02.2012. – Режим доступа: http://psyrus.ru/ethics/ (дата обращения: 30.04.2026).

30. Российская Федерация. Федеральный закон от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации». – М.: Кодекс, 2025. – 256 с. – ISBN 978-5-904791-89-2.

31. Климов, Е. А. Психология профессионального самоопределения: учеб. пособие / Е. А. Климов. – 2-е изд. – М.: Академия, 2010. – 304 с. – ISBN 978-5-7695-6700-9.

32. Holland, J. L. Making Vocational Choices: A Theory of Vocational Personalities and Work Environments / J. L. Holland. – 3rd ed. – Odessa, FL: Psychological Assessment Resources, 1997. – 303 p. – ISBN 978-0-911907-27-7.

33. Голомшток, А. Е. Карта интересов: методика выявления профессиональных интересов / А. Е. Голомшток. – М., 1978. – 24 с.

34. Зеер, Э. Ф. Профориентология: теория и практика / Э. Ф. Зеер, А. М. Павлова, Н. О. Садовникова. – 2-е изд. – М.: Академический проект, 2008. – 192 с. – ISBN 978-5-8291-0837-0.

35. Сидоренко, Е. В. Методы математической обработки в психологии / Е. В. Сидоренко. – СПб.: Речь, 2010. – 350 с. – ISBN 978-5-9268-0163-0.

36. Cohen, J. Statistical Power Analysis for the Behavioral Sciences / J. Cohen. – 2nd ed. – Hillsdale, NJ: Lawrence Erlbaum Associates, 1988. – 567 p. – ISBN 978-0-805-80283-2.

37. Brooke, J. SUS: A 'Quick and Dirty' Usability Scale / J. Brooke // Usability Evaluation in Industry. – London: Taylor & Francis, 1996. – Vol. 189, No. 194. – P. 4–7.

38. Sauro, J. A Practical Guide to the System Usability Scale / J. Sauro. – Denver, CO: MeasuringU, 2020. – 150 p.

39. Пряжников, Н. С. Профессиональная ориентация: учебное пособие / Н. С. Пряжников. – М.: Академия, 2020. – 320 с. – ISBN 978-5-4468-7890-3.

40. Ермолаева-Томина, Л. Б. Психология художественного творчества: учебное пособие / Л. Б. Ермолаева-Томина. – М.: Академический проект, 2005. – 304 с. – ISBN 5-8291-0463-5.

41. Ильин, Е. П. Психология творчества, креативности, одаренности / Е. П. Ильин. – СПб.: Питер, 2009. – 448 с. – (Мастера психологии). – ISBN 978-5-49807-239-3.

42. Сильверсан, А. В. Методика психодиагностики уровня творческости / А. В. Сильверсан, Н. С. Аболина // Экспериментальная психология в России: традиции и перспективы. – М.: Психологический институт РАО, 2021. – С. 275–282. – URL: https://psyjournals.ru/nonserialpublications/exp_collection/contents/34788 (дата обращения: 05.05.2026).

43. Холл, Н. Эмоциональный интеллект: практическое руководство / Н. Холл; пер. с англ. – СПб.: Питер, 2021. – 256 с. – ISBN 978-5-4461-1589-5.

44. Sternberg, R. J. The Nature of Creativity: Contemporary Psychological Perspectives / R. J. Sternberg. – Cambridge: Cambridge University Press, 1988. – 464 p. – ISBN 978-0-521-33892-9.

45. Богоявленская, Д. Б. Зарубежные исследования психологии творчества: Постгилфордовский период (часть 2) / Д. Б. Богоявленская, И. А. Сусоколова // Психологическая наука и образование. – 2007. – Т. 12. – № 4. – С. 94–104. – URL: https://psyjournals.ru/journals/pse/archive/2007_n4/Bogoyavlenskaya_Susokolova (дата обращения: 05.05.2026).

46. Выготский, Л. С. Воображение и творчество в детском возрасте / Л. С. Выготский. – СПб.: Союз, 1997. – 96 с. – ISBN 5-87852-033-8.

47. World Medical Association. Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects [Электронный ресурс] // WMA General Assembly. – 2024. – Режим доступа: https://www.wma.net/policies-post/wma-declaration-of-helsinki/ (дата обращения: 30.04.2026).

48. Российская Федерация. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных». – М., 2025. – 64 с.

49. Всеобщая декларация прав человека: принята Генеральной Ассамблеей ООН 10.12.1948. – М.: Юридическая литература, 2019. – 48 с.

50. Конституция Российской Федерации: принята всенародным голосованием 12.12.1993. – М.: Проспект, 2020. – 64 с.

51. Пономарев, Я. А. Психология творчества / Я. А. Пономарев. – М.: Наука, 1976. – 304 с.

52. Пономарев, Я. А. Психология творчества: перспективы развития / Я. А. Пономарев // Психологический журнал. – 1994. – Т. 15. – № 6. – С. 126–136.

53. Torrance, E. P. The Torrance Tests of Creative Thinking: Norms-technical manual / E. P. Torrance. – Scholastic Testing Service, 1974. – 205 p.

54. Holland, J. L. Dictionary of Holland Occupational Codes / J. L. Holland, G. D. Gottfredson, P. G. (N). – Odessa, FL: Psychological Assessment Resources, 2018.

ПРИЛОЖЕНИЕ А

(обязательное)

Стартап-проект: инструментарий выявления профессиональных склонностей и творческого потенциала учащихся

Материалы экспериментального исследования

1. Введение

Настоящее приложение содержит документальные материалы экспериментального исследования, проведенного в рамках магистерской диссертации на тему «Инструментарий выявления профессиональных склонностей и творческого потенциала учащихся». В целях обеспечения конфиденциальности персональных данных в соответствии с Федеральным законом от 27.07.2006 № 152-ФЗ «О персональных данных», все данные, позволяющие идентифицировать участников (фамилии, имена, номера телефонов, адреса электронной почты), были зашифрованы или заменены на уникальные коды [1, с. 80].

Исследование проводилось на базе ФГБОУ ВО «Пензенский государственный университет», Многопрофильный колледж, отделение «Информационные системы и программирование», в период с ноября 2025 года по апрель 2026 года.

А.1 Форма информированного согласия участника исследования

Для участия в экспериментальном исследовании каждый участник подписывал электронную или бумажную форму информированного согласия. Ниже представлен шаблон формы, утвержденный Локальным этическим комитетом Пензенского государственного университета (протокол № 12 от 15 октября 2025 года).

Форма информированного согласия

«__» _____ 202 г.

Я, нижеподписавшийся(аяся), добровольно даю согласие на участие в психолого-педагогическом исследовании, проводимом в рамках магистерской диссертации «Инструментарий выявления профессиональных склонностей и творческого потенциала учащихся».

Цель исследования

апробация веб-приложения «IAURA» и сравнительный анализ его эффективности с традиционными методами профориентационной диагностики.

Процедура участия

мне предстоит пройти комплексную профориентационную диагностику, включающую:

1. Заполнение регистрационной формы (логин, пароль, имя, дата рождения, пол, регион).
2. Последовательное прохождение четырех психологических тестов:

– Дифференциально-диагностический опросник Е.А. Климова (30 вопросов);

– Опросник эмоционального интеллекта Н. Холла (30 утверждений);

– Методику определения профессионального типа Дж. Голланда (6 вопросов);

– Фигурный тест творческого мышления Э.П. Торренса (творческое задание на рисование).

3. Ответы на вопросы анкеты обратной связи (для контрольной группы – бумажный бланк; для экспериментальной группы – электронный опросник SUS).

Общая продолжительность участия в диагностике составляет от 25 до 50 минут в зависимости от распределения по группе.

Конфиденциальность данных:

– Все мои персональные данные (логин, пароль, имя, дата рождения, пол, регион) будут зашифрованы и заменены уникальным цифровым кодом (User_ID). Файл, связывающий код с персональными данными, хранится отдельно от результатов тестирования.

– Доступ к первичным данным имеют только научный руководитель и исполнитель исследования. Иные лица к данным не допускаются.

– Пароль, указанный при регистрации, хранится в хешированном виде (алгоритм PBKDF2).

– Результаты тестирования будут использованы только в обобщенном виде для научного анализа и публикаций. Публикация персональных данных или изображений, позволяющих идентифицировать участника, не допускается.

– По окончании исследования файл, связывающий User_ID с персональными данными, будет уничтожен.

Права участника

– Я имею право отказаться от участия в исследовании в любой момент без объяснения причин и без негативных последствий для себя.

– Я имею право не отвечать на отдельные вопросы или прервать тестирование.

– Я имею право запросить удаление моих данных из системы по окончании исследования (удаление производится в течение 14 дней с момента запроса).

Информация о рисках и преимуществах

Участие в исследовании не сопряжено с риском для физического или психического здоровья. Возможен минимальный дискомфорт, связанный с утомляемостью при длительном тестировании (не более 50 минут). В случае возникновения вопросов или жалоб я могу обратиться к исполнителю исследования по адресу электронной почты: [адрес скрыт].

Преимущества участия:

– Бесплатное прохождение комплексной профориентационной диагностики (аналог платных услуг стоимостью от 1500 до 5000 рублей).

– Получение персонализированного отчета с рекомендациями по выбору профессии и ссылками на актуальные вакансии.

– Вклад в развитие научно обоснованных методов профессиональной ориентации.

С ходом исследования ознакомлен(а), на условия участия согласен(на).

_____ /_____/

(подпись) (ФИО)

Дата: «__» _____ 202 г.

А.2 Характеристика выборки (зашифрованные данные)

В экспериментальном исследовании приняли участие 90 студентов 1–2 курсов Многопрофильного колледжа Пензенского государственного университета. Все участники были случайным образом распределены на две группы: экспериментальную (n=45) и контрольную (n=45).

Ниже представлены таблицы с демографическими характеристиками участников. Для обеспечения конфиденциальности каждому участнику присвоен уникальный цифровой код (User_ID), не содержащий прямой привязки к фамилии, имени, отчеству или иным идентифицирующим признакам.

Таблица А.2.1 – Демографическая характеристика экспериментальной группы (n=45)

User_ID (зашифрованный)	Возраст (лет)	Пол	Курс	Форма обучения	Средний балл успеваемости
ID_EXP-1	19	М	2	Очная	4,2
ID_EXP-2	20	Ж	2	Очная	4,5
ID_EXP-3	18	М	1	Очная	3,8
ID_EXP-4	21	Ж	2	Очная	4,7
ID_EXP-5	19	М	1	Очная	4,0
...	...	...	...	...	...

Полный список из 45 участников экспериментальной группы представлен в электронном репозитории диссертации (файл experimental_group_data.xlsx, защищен паролем). Пароль предоставляется научному руководителю и членам диссертационного совета по запросу.

Таблица А.2.2 – Демографическая характеристика контрольной группы (n=45)

User_ID (зашифрованный)	Возраст (лет)	Пол	Курс	Форма обучения	Средний балл успеваемости
ID_CON-1	20	М	2	Очная	4,1
ID_CON-2	18	Ж	1	Очная	4,3
ID_CON-3	19	М	2	Очная	3,7
ID_CON-4	22	Ж	2	Очная	4,6
ID_CON-5	17	М	1	Очная	4,0
...	...	...	...	...	...

Полный список из 45 участников контрольной группы представлен в электронном репозитории диссертации (файл control_group_data.xlsx, защищен паролем).

Таблица А.2.3 – Сравнительная характеристика выборки

Параметр	Экспериментальная группа (n=45)	Контрольная группа (n=45)	Статистическая значимость различий
Средний возраст (M ± SD)	19,1 ± 1,3	19,2 ± 1,4	t = 0,35, p = 0,727
Количество мужчин (%)	21 (46,7%)	20 (44,4%)	χ^2 = 0,05, p = 0,828
Количество женщин (%)	24 (53,3%)	25 (55,6%)	χ^2 = 0,05, p = 0,828
Средний балл успеваемости	4,12 ± 0,45	4,08 ± 0,48	t = 0,41, p = 0,683
Доля студентов 1-го курса	23 (51,1%)	24 (53,3%)	χ^2 = 0,04, p = 0,835
Доля студентов 2-го курса	22 (48,9%)	21 (46,7%)	χ^2 = 0,04, p = 0,835

Составлено автором на основе данных эксперимента.

А.3 Инструкция для экспертов-психологов

Ниже представлен полный текст инструкции, предоставленной трем независимым экспертам-психологам для оценки творческих изображений (тест Э.П. Торренса).

Инструкция для эксперта

Уважаемый эксперт!

Вам предлагается принять участие в независимой экспертной оценке творческих изображений, созданных участниками экспериментального исследования (n=90). Исследование проводится в рамках магистерской диссертации «Инструментарий выявления профессиональных склонностей и творческого потенциала учащихся».

Цель экспертизы: оценить уровень креативности каждого изображения по единой шкале (категории 1–4) для последующего сравнения с результатами автоматической классификации, выполненной веб-приложением «IAURA».

Материалы для экспертизы:

- 90 творческих изображений (45 из экспериментальной группы и 45 из контрольной группы) в формате PNG.
- Каждое изображение имеет уникальный код (Image_ID), не содержащий информации о групповой принадлежности участника.
- Базовая графическая заготовка (круг, квадрат, линия или ромб), на основе которой создавалось изображение, также предоставлена для контекста.

Критерии оценки:

Оцените каждое изображение по четырем категориям креативности. Дескрипторы категорий приведены в таблице ниже.

Категория	added_ratio (для справки)	Уровень креативности	Дескриптор
1	< 0,02	Низкий	Изменения практически отсутствуют. Добавлено 1–2 простых элемента (точка, короткая линия). Композиция не развита.
2	0,02–0,08	Ниже среднего	Добавлено несколько элементов (3–5), но композиция остается простой. Присутствуют отдельные оригинальные детали.
3	≥ 0,08, clip_delta ≤ 0,6	Средний	Развитая, многоэлементная композиция. Идея ясна, понятна, легко интерпретируется. Элементы органично вписаны в заготовку.
4	≥ 0,08, clip_delta > 0,6	Высокий	Сложная, многослойная композиция. Высокая оригинальность, нестандартность решения. Неожиданная трансформация заготовки.

Процедура оценки

1. Рассмотрите каждое изображение в представленной последовательности.
2. Оцените изображение по шкале от 1 до 4, руководствуясь приведенными дескрипторами.
3. Заполните протокол оценки (форма прилагается), указав Image_ID и выбранную категорию.
4. В случае сомнений при оценке конкретного изображения, сделайте пометку в колонке «Примечания».

Важные замечания

- Оценка проводится «вслепую» (blind assessment) – Вы не знаете, к какой группе (экспериментальной или контрольной) относится оцениваемый участник.
- Вы не имеете доступа к результатам автоматической классификации системы «IAURA».
- Оценивайте каждое изображение независимо от других экспертов.
- Ограничений по времени нет. Рекомендуется делать перерывы при оценке каждых 20–30 изображений для поддержания концентрации.

Благодарим за участие в экспертизе!

А.4 Декларация об уничтожении файлов, связывающих User_ID с персональными данными

ДЕКЛАРАЦИЯ

Я, нижеподписавшийся, исполнитель магистерской диссертации «Инструментарий выявления профессиональных склонностей и творческого потенциала учащихся» Дмитриев Ю.А., настоящим подтверждаю, что по завершении экспериментального исследования (апрель 2026 года) файл,

связывающий зашифрованные User_ID участников с их персональными данными (логин, имя, дата рождения, пол, регион), был уничтожен без возможности восстановления.

Уничтожение произведено 30 апреля 2026 года.

Исполнитель: _____ /Дмитриев Ю.А./

Научный руководитель: _____ /[ФИО]/

Дата: 30 апреля 2026 г.

ПРИЛОЖЕНИЕ Б.
Листинг кода

Основной файл приложения (app.py)

```
import os
import json
import base64
import random
import logging
from datetime import datetime
from io import BytesIO
from PIL import Image
from flask import Flask, render_template, request, redirect, url_for, session,
make_response, send_file, jsonify
from flask_sqlalchemy import SQLAlchemy
from werkzeug.security import generate_password_hash, check_password_hash
from werkzeug.utils import secure_filename

# Настройка логирования
logging.basicConfig(level=logging.INFO)
logger = logging.getLogger(__name__)

# Инициализация Flask приложения
app = Flask(__name__)

# Базовая конфигурация
app.secret_key = os.environ.get('SECRET_KEY', 'aura-secret-key-2025-change-in-
production')
app.config['SQLALCHEMY_DATABASE_URI']     =     os.environ.get('DATABASE_URL',
'sqlite:///'     +     os.path.join(os.path.dirname(os.path.abspath(__file__)),
'instance', 'db.sqlite'))
app.config['SQLALCHEMY_TRACK_MODIFICATIONS'] = False
app.config['MAX_CONTENT_LENGTH'] = 16 * 1024 * 1024  # 16MB

# Пути
BASE_DIR = os.path.dirname(os.path.abspath(__file__))
app.config['UPLOAD_FOLDER'] = os.path.join(BASE_DIR, 'uploads')
app.config['DATA_PATH'] = os.path.join(BASE_DIR, 'data')
app.config['USERS_DATA_PATH'] = os.path.join(BASE_DIR, 'data', 'users')
app.config['TESTS_DATA_PATH'] = os.path.join(BASE_DIR, 'data', 'tests')

# Создание необходимых директорий
```

```
for directory in [
    os.path.join(BASE_DIR, 'instance'),  # Важно: создаем папку instance
    app.config['UPLOAD_FOLDER'],
    os.path.join(app.config['UPLOAD_FOLDER'], 'drawings'),
    os.path.join(app.config['UPLOAD_FOLDER'], 'temp'),
    app.config['USERS_DATA_PATH'],
    app.config['TESTS_DATA_PATH'],
    os.path.join(app.config['DATA_PATH'], 'recommendations'),
    os.path.join(app.config['DATA_PATH'], 'config'),
    os.path.join(BASE_DIR, 'logs')
]:
    os.makedirs(directory, exist_ok=True)

# Инициализация базы данных
db = SQLAlchemy(app)

# ---- Модели базы данных ----
class User(db.Model):
    __tablename__ = "users"
    id = db.Column(db.Integer, primary_key=True)
    login = db.Column(db.String(120), unique=True, nullable=False)
    password_hash = db.Column(db.String(256), nullable=False)
    name = db.Column(db.String(120))
    birthdate = db.Column(db.String(20))
    gender = db.Column(db.String(20))
    region = db.Column(db.String(120))
    hh_area = db.Column(db.String(20))
    created_at = db.Column(db.DateTime, default=datetime.utcnow)

    def set_password(self, password):
        self.password_hash = generate_password_hash(password)

    def check_password(self, password):
        return check_password_hash(self.password_hash, password)

class TestResult(db.Model):
    __tablename__ = "test_results"
    id = db.Column(db.Integer, primary_key=True)
    user_id = db.Column(db.Integer, nullable=False)
    test_name = db.Column(db.String(50), nullable=False)
    score = db.Column(db.Float)
    interpretation = db.Column(db.String(200))
    raw_answers = db.Column(db.Text)
```

```
    created_at = db.Column(db.DateTime, default=datetime.utcnow)

# ---- Утилиты для работы с JSON ----
class JSONManager:
    @staticmethod
    def load_json(filepath):
        """Загружает JSON файл"""
        try:
            if os.path.exists(filepath):
                with open(filepath, 'r', encoding='utf-8') as f:
                    return json.load(f)
        except Exception as e:
            logger.error(f"Error loading JSON {filepath}: {e}")
        return None

    @staticmethod
    def save_json(filepath, data):
        """Сохраняет данные в JSON файл"""
        try:
            # Создаем директорию если не существует
            os.makedirs(os.path.dirname(filepath), exist_ok=True)
            with open(filepath, 'w', encoding='utf-8') as f:
                json.dump(data, f, ensure_ascii=False, indent=2)
            return True
        except Exception as e:
            logger.error(f"Error saving JSON {filepath}: {e}")
            return False

# ---- Загрузка данных тестов ----
HH_AREAS = {
    "Пензенская область": "72",
    "Москва": "1",
    "Санкт-Петербург": "2",
    "Новосибирская область": "4",
    "Свердловская область": "3",
    "Краснодарский край": "53",
    "Татарстан": "73",
    "Алтайский край": "2050",
    "Нижегородская область": "66",
    "Самарская область": "78",
    "Ростовская область": "76",
    "Челябинская область": "97",
    "Башкортостан": "99",
```

```
    "Красноярский край": "54",
    "Ставропольский край": "51"
}

PROF_TO_HH_QUERY = {
    "Учитель": "Учитель",
    "Программист": "Программист",
    "Дизайнер": "Дизайнер",
    "Агроном": "Агроном",
    "Водитель": "Водитель",
    "Руководитель": "Руководитель"
}

def load_test_data(test_name):
    """Загружает данные теста из JSON файла"""
    # Проверяем разные возможные пути
    test_paths = [
        os.path.join(app.config['TESTS_DATA_PATH'], f'{test_name}.json'),
        os.path.join(app.config['DATA_PATH'], 'tests', f'{test_name}.json'),
        os.path.join(BASE_DIR, 'data', 'tests', f'{test_name}.json')
    ]

    for test_path in test_paths:
        if os.path.exists(test_path):
            data = JSONManager.load_json(test_path)
            if data:
                return data

    logger.error(f"Test data not found for {test_name}")
    return None

def save_user_results(user_id, test_name, results):
    """Сохраняет результаты пользователя в JSON файл"""
    user_file          =          os.path.join(app.config['USERS_DATA_PATH'],
f'user_{user_id}.json')

    # Создаем директорию если не существует
    os.makedirs(os.path.dirname(user_file), exist_ok=True)

    # Загружаем существующие результаты
    user_data = JSONManager.load_json(user_file) or {
        "user_id": user_id,
        "created_at": datetime.now().isoformat(),
```

```
        "profile": {},
        "test_results": {}
    }

    # Обновляем результаты
    user_data["test_results"][test_name] = {
        "completed_at": datetime.now().isoformat(),
        **results
    }

    # Сохраняем обратно
    return JSONManager.save_json(user_file, user_data)

def get_user_results(user_id):
    """Получает все результаты пользователя"""
    user_file          =          os.path.join(app.config['USERS_DATA_PATH'],
f'user_{user_id}.json')
    return JSONManager.load_json(user_file)

    # GET запрос - выбираем случайную базовую картинку
    base_name = random.choice(bases)
    return  render_template('torrens.html',  base_name=base_name,  bases=bases,
test_data=test_data)

@app.route('/loading')
@login_required
def loading():
    return render_template('loading.html')

@app.route('/results')
@login_required
def results():
    user = User.query.get(session['user_id'])
    if not user:
        session.clear()
        return redirect(url_for('register'))

    # Получаем результаты из базы данных
    db_results = TestResult.query.filter_by(user_id=user.id).all()
    summary  =  {r.test_name:  (r.score,  r.interpretation[:50]  +  '...'  if
len(r.interpretation) > 50 else r.interpretation)
                for r in db_results}
```

```
    # Получаем результаты из JSON
    user_results = get_user_results(user.id)

    # Генерируем ссылки на вакансии
    area = user.hh_area or "1"
    hh_links = {}
    for name, query in PROF_TO_HH_QUERY.items():
        hh_links[name]                                                    =
f"https://hh.ru/search/vacancy?text={query}&area={area}"

    return render_template('results.html',
                           user=user,
                           summary=summary,
                           hh_links=hh_links,
                           user_results=user_results)

@app.route('/export_csv')
@login_required
def export_csv():
    import csv
    from io import StringIO

    user = User.query.get(session['user_id'])
    results = TestResult.query.filter_by(user_id=user.id).all()

    si = StringIO()
    writer = csv.writer(si)

    # Заголовки
    writer.writerow(['Test Name', 'Score', 'Interpretation', 'Date'])

    # Данные
    for result in results:
        writer.writerow([
            result.test_name,
            result.score,
            result.interpretation,
            result.created_at.strftime('%Y-%m-%d         %H:%M:%S')         if
result.created_at else ''
        ])

    output = make_response(si.getvalue())
```

```
    output.headers["Content-Disposition"]            =            "attachment;
filename=test_results.csv"
    output.headers["Content-type"] = "text/csv; charset=utf-8"

    return output

@app.route('/logout')
def logout():
    session.clear()
    return redirect(url_for('index'))

# Инициализация базы данных
@app.before_request
def init_db():
    try:
        # Проверяем, существует ли папка instance
        instance_dir = os.path.join(BASE_DIR, 'instance')
        if not os.path.exists(instance_dir):
            os.makedirs(instance_dir, exist_ok=True)

        # Создаем все таблицы
        db.create_all()
    except Exception as e:
        logger.error(f"Error initializing database: {e}")
        # В случае ошибки продолжаем работу без базы данных
        pass
# Запуск приложения
if __name__ == '__main__':
    # Устанавливаем права на папки
    try:
        # Даем права на запись в папки
        for directory in [os.path.join(BASE_DIR, 'instance'),
                          app.config['UPLOAD_FOLDER'],
                          app.config['USERS_DATA_PATH']]:
            if os.path.exists(directory):
                os.chmod(directory, 0o777)
    except Exception as e:
        logger.error(f"Error setting permissions: {e}")

    app.run(debug=True, host='0.0.0.0', port=5000)
```

Модели данных (models.py)

```
from datetime import datetime
```

```
from flask_sqlalchemy import SQLAlchemy
from werkzeug.security import generate_password_hash, check_password_hash
db = SQLAlchemy()
class User(db.Model):
    __tablename__ = "users"
    id = db.Column(db.Integer, primary_key=True)
    login = db.Column(db.String(120), unique=True, nullable=False)
    password_hash = db.Column(db.String(256), nullable=False)
    name = db.Column(db.String(120))
    birthdate = db.Column(db.String(20))
    gender = db.Column(db.String(20))
    region = db.Column(db.String(120))
    hh_area = db.Column(db.String(20))
    created_at = db.Column(db.DateTime, default=datetime.utcnow)
    def set_password(self, password: str):
        self.password_hash = generate_password_hash(password)
    def check_password(self, password: str) -> bool:
        return check_password_hash(self.password_hash, password)
class Result(db.Model):
    __tablename__ = "results"
    id = db.Column(db.Integer, primary_key=True)
    user_id = db.Column(db.Integer, db.ForeignKey('users.id'), nullable=False)
    test_name = db.Column(db.String(50), nullable=False)   # klimov, hall,
holla>
    raw_json = db.Column(db.Text)           # сырые ответы JSON
    raw_score = db.Column(db.Float)         # числовой балл
    interpreted = db.Column(db.String(200)) # интерпретация/профиль
    created_at = db.Column(db.DateTime, default=datetime.utcnow)
class Drawing(db.Model):
    __tablename__ = "drawings"
    id = db.Column(db.Integer, primary_key=True)
    user_id = db.Column(db.Integer, db.ForeignKey('users.id'), nullable=False)
base_name = db.Column(db.String(120))   # имя заготовки
    saved_path = db.Column(db.String(255))  # путь к дорисовке
    added_ratio = db.Column(db.Float)       # доля добавленных штрихов
    clip_delta = db.Column(db.Float)        # (опционально) дельта CLIP
    category = db.Column(db.Integer)        # 1..4 тип
    created_at = db.Column(db.DateTime, default=datetime.utcnow)
    description = db.Column(db.Text, nullable=True)
```