

ГИБРИДНЫЙ ПОДХОД К АВТОНОМНОМУ УПРАВЛЕНИЮ РОБОТИЗИРОВАННЫМ УСТРОЙСТВОМ, СОЧЕТАЮЩИЙ ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ И ОБУЧЕНИЕ НА ОСНОВЕ ДЕМОНСТРАЦИЙ ОПЕРАТОРА

Р. Т. Исхаков¹ [0009-0004-9262-7296], В. В. Кугуракова² [0000-0002-1552-4910]

^{1, 2}Казанский (Приволжский) федеральный университет, г. Казань, Россия

¹iskhakovrobert42@gmail.com, ²vlada.kugurakova@gmail.com

Аннотация

Предложен и экспериментально верифицирован воспроизводимый гибридный подход к обучению автономного управления роботизированным устройством, реализованный на общедоступном инструментарии без специализированного вычислительного оборудования. Подход сочетает обучение на основе демонстраций оператора, обучение с подкреплением, учебный план и доменную рандомизацию в единой последовательной схеме. На основе сравнительного анализа пяти прототипов, представляющих ключевые классы архитектурных решений, выявлены применимые подходы для ресурсно-ограниченных условий. Три контролируемых эксперимента в Unity ML-Agents показали превосходство гибридной комбинации в сравнении с изолированными подходами по скорости обучения и устойчивости к изменениям среды. Обозначены ограничения, связанные с катастрофическим забыванием, и выявлены перспективы интеграции больших языковых моделей для генерации синтетических обучающих данных.

Ключевые слова: телеуправление, удаленное управление, робот, виртуальная реальность, виртуальная среда, машинное обучение, машинное обучение с подкреплением, машинное обучение на основе демонстраций, демонстрация оператором, Unity, ML-agents.

ВВЕДЕНИЕ

Индустрия роботизированных устройств переживает устойчивый переход от телеуправляемых систем к автономным. Работы авторов по VR-телеуправлению [1–7], генерации вариаций движений [8] и распознаванию объектов в прикладных задачах [9] создали основу, позволяющую замкнуть следующую цепочку: оператор записывает демонстрации через VR¹-интерфейс, после этого роботизированное устройство дообучается самостоятельно. Ключевым препятствием остается обучение: ни один классический подход машинного обучения не обеспечивает одновременно быстрый старт, устойчивость к изменениям среды и способность обобщать навыки. Обучение предложено проводить с цифровой моделью прототипа роботизированного устройства (см. рис. 1). Важно учесть, что применение обученной модели ограничено использованием того роботизированного устройства, цифровая модель которого использовалась в процессе обучения. Применение обученной модели для других роботизированных устройств требует провести процесс обучения заново, но уже с другой соответствующей цифровой моделью этих устройств.

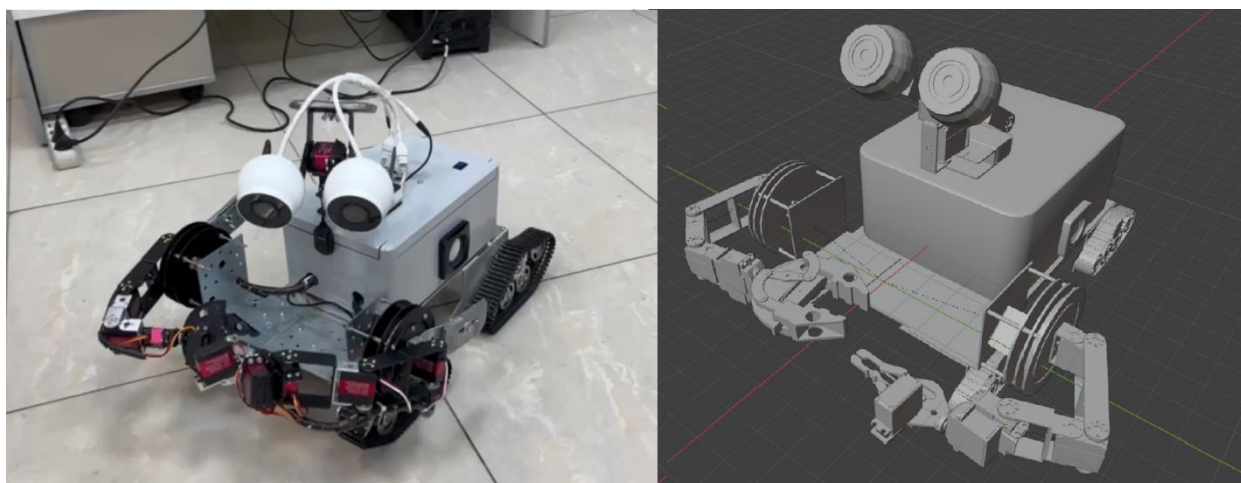


Рис. 1. Слева – физический прототип, справа – цифровая модель роботизированного устройства.

Практика ведущих разработчиков свидетельствует о повсеместном переходе к гибридным подходам. Однако все успешные решения либо опираются на

¹ VR (Virtual Reality) – виртуальная реальность; технология, создающая эффект присутствия в смоделированном трехмерном окружении.

корпоративную инфраструктуру, либо являются закрытыми проприетарными системами без воспроизводимого протокола, поэтому исследовательские группы, работающие с ограниченными ресурсами, лишены практического ориентира.

Цель настоящей работы – разработать воспроизводимый гибридный подход к обучению автономного управления роботизированным устройством в условиях ограниченных ресурсов и экспериментально подтвердить, что сочетание обучения на основе демонстраций оператора с обучением с подкреплением и доменной рандомизацией превосходит каждый из подходов по отдельности.

В результате проведенных исследований предложен и верифицирован воспроизводимый гибридный подход на общедоступном инструментарии (Unity ML-Agents, VR-гарнитура): обучение на основе демонстраций – инициализирующий этап; RL^2 – этап обобщения; доменная рандомизация совместно с учебным планом – механизм формирования устойчивости. Поставлены и решены следующие задачи:

- 1) провести сравнительный анализ прототипов, представляющих ключевые классы гибридных архитектур, с целью выявления применимых решений для ресурсно-ограниченных условий;
- 2) обосновать выбор виртуальной среды для обучения и инструментов захвата движений оператора;
- 3) разработать и реализовать гибридный подход в Unity ML-Agents;
- 4) экспериментально сравнить эффективность изолированных и гибридного подходов;
- 5) определить ограничения и обозначить направления развития, включая интеграцию больших языковых моделей (Large Language Model, LLM) для генерации синтетических данных.

1. АНАЛИЗ ПОДХОДОВ ОБУЧЕНИЯ РОБОТИЗИРОВАННЫХ УСТРОЙСТВ

Подходы к обучению роботизированных устройств принято делить на три фундаментальных класса: обучение с учителем, обучение с подкреплением и обучение без учителя.

При обучении с учителем роботизированное устройство воспроизводит

² RL (Reinforcement Learning) – обучение с подкреплением; агент обучается методом проб и ошибок, максимизируя числовую награду.

действия оператора по записанным демонстрациям. Такой подход обеспечивает быстрый старт и особенно эффективен для задач, которые сложно описать с помощью математической функции награды. Принципиальное ограничение состоит в том, что модель не способна превзойти учителя: накапливающиеся отклонения от траектории демонстрации приводят к ошибкам в ситуациях, не представленных в обучающем наборе [10].

Обучение с подкреплением (RL) позволяет агенту самостоятельно формировать оптимальную стратегию методом проб и ошибок, максимизировав числовую награду. Достоинством является принципиальная возможность превзойти оператора, недостатком – высокая вычислительная стоимость: в условиях разреженной награды агент совершает огромное число случайных действий до первого успеха [11, 12].

Обучение без учителя применяют преимущественно для анализа сенсорных данных; в чистом виде для управления роботизированными устройствами его используют редко, как правило, в составе гибридных схем для предварительного формирования признакового пространства [13].

Ни один из фундаментальных подходов в изолированном применении не решает достаточно эффективно задачу автономного управления в изменяющейся среде. Это обусловило развитие следующих прикладных стратегий:

- обучение с учебным планом (curriculum learning) – задача разбивается на уроки возрастающей сложности, ускоряя обучение в условиях разреженной награды [14, 15];
- преобразование симуляции в реальность (sim-to-real transfer) – обучение в виртуальной среде позволяет запускать сотни параллельных экземпляров и ускоряет процесс в тысячи раз по сравнению с реальным миром [16];
- доменная рандомизация (domain randomization) – создается множество вариаций виртуальной среды с различными параметрами (освещение, цвет, положение объектов), что повышает устойчивость обученной модели к реальным условиям [17];
- состязательное обучение (adversarial training) – в процессе обучения задействуется вторая нейронная сеть, намеренно усложняющая задачу основному агенту; метод выявляет и устраняет слабые места модели [18];
- обучение на основе крупных предобученных моделей (large-scale

pretraining + fine-tuning) – вместо обучения с нуля используют готовые мультимодальные модели с обобщенным пониманием мира; дообучение на небольшом наборе роботизированных демонстраций резко снижает требования к объему специализированных данных [19].

Поскольку фундаментальные подходы редко применяются в чистом виде, именно их комбинации с прикладными стратегиями образуют гибридные схемы, демонстрирующие наилучшие результаты на практике.

2. ГИБРИДНЫЕ ПОДХОДЫ В ПРОМЫШЛЕННЫХ И АКАДЕМИЧЕСКИХ ПРОТОТИПАХ

Для формирования аналитической базы нами рассмотрены пять прототипов, каждый из которых представляет отдельный класс архитектурных решений:

- Tesla Optimus воплощает подход на основе маховика данных (data-flywheel): корпоративный флот устройств обеспечивает непрерывный поток обучающих данных;
- HumanPlus представляет маломасштабную академическую разработку (low-cost academic) с акцентом на эффективность алгоритмов при минимальном бюджете;
- XPeng Iron демонстрирует перенос технологий автономного вождения на роботизированную платформу (automotive transfer);
- Figure AI реализует стратегию тонкой настройки предобученной базовой модели (foundation model fine-tuning);
- GREEN от Сбера является примером системы на основе LLM, адаптированной к локализованному рынку (LLM-based, localized).

Совокупность этих классов охватывает основное пространство современных гибридных архитектур: практические решения, как правило, представляют собой вариации или комбинации этих паттернов. Каждый кейс анализируется по схеме: проблема → решение → ключевое ограничение (сводные данные см. в табл. 1).

2.1. Tesla Optimus Gen2

Tesla обладает уникальным преимуществом – флотом из более четырех миллионов автомобилей, собирающих видеоданные в теновом режиме: система FSD³ непрерывно сравнивает предполагаемые действия нейросети с действиями водителя, а момент расхождения автоматически поступает в обучающую выборку [20]. Optimus Gen2 наследует этот механизм напрямую.

Архитектура построена на едином Bot Brain SoC⁴, восьми камерах без LiDAR⁵ и принципе End-to-end⁶: входные сигналы камер напрямую преобразуются в команды двигателям; VLA⁷-модуль добавляет поддержку команд на естественном языке (см. рис. 2).

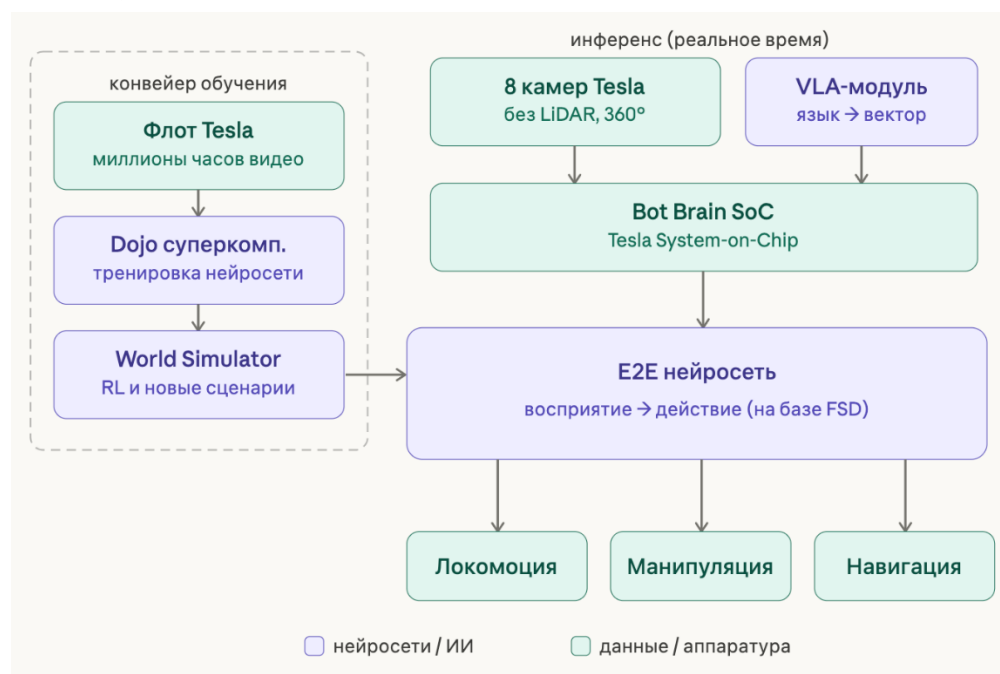


Рис. 2. Архитектура роботизированного устройства Tesla Optimus Gen2.

³ FSD (Full Self-Driving) – система полностью автономного вождения, разработанная компанией Tesla.

⁴ Bot Brain SoC (System-on-Chip) – система на кристалле; интегральная схема, объединяющая процессор, память и периферийные модули.

⁵ LiDAR (Light Detection and Ranging) – технология дистанционного зондирования, которая измеряет расстояние до объектов с помощью лазерных лучей.

⁶ End-to-end (E2E) – сквозная нейросетевая архитектура: входные данные напрямую преобразуются в выходные действия без промежуточных символических модулей.

⁷ VLA (Vision-Language-Action) – мультимодальная архитектура, объединяющая зрительное восприятие, понимание языковых команд и генерацию физических действий.

Neural World Simulator генерирует фотореалистичные сценарии на основе действий самого агента, обеспечивая замкнутые симуляции и RL на синтетических данных. Каждый час работы робота генерирует данные для улучшения нейросети, создавая нарастающий эффект, недоступный без корпоративной инфраструктуры [21].

2.2. HumanPlus (Стэнфордский университет)

Проект HumanPlus решал задачу, обратную задаче Tesla: достичь значимого результата при минимальных ресурсах. За основу был взят серийный робот Unitree H1; вся разработка сосредоточена на методологии обучения [22] (см. рис. 3).

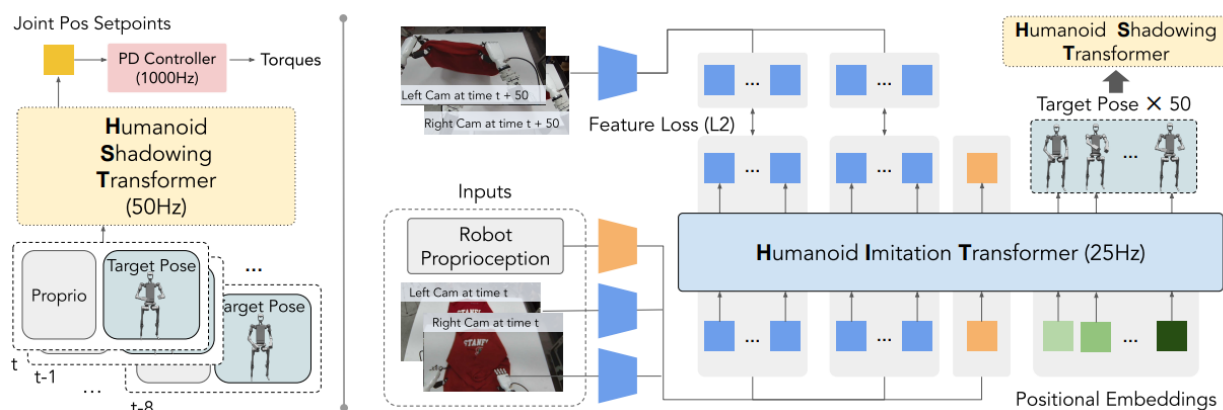


Рис. 3. Архитектура роботизированного устройства HumanPlus.

Гибридная схема состоит из трех этапов. Первый этап — обучение с подкреплением (RL) в симуляторе NVIDIA Isaac Gym для освоения базовой локомоции и баланса (что дает ускорение в 1000 раз по сравнению с реальным миром). Второй этап включает теневое копирование с системой WHAM⁸, которая восстанавливает трехмерное движение оператора из обычных камер и формирует парные данные «наблюдение → команда мотору». Третий этап — нейросетевое обучение на собранном датасете с переходом к автономии, где за 40 часов видеозаписей робот осваивает сложные двигательные навыки. Ключевое ограни-

⁸ WHAM (World-grounded Humans with Accurate Motion) — система восстановления трёхмерного движения человека в глобальной системе координат по обычным видеокерам.

чение состоит в том, что датасет покрывает только продемонстрированные действия.

2.3. XPeng Iron

XPeng решал проблему задержки в модульных архитектурах. Ответ – трех-слойная сквозная архитектура: VLM⁹ (слой понимания сцены) → VLT¹⁰ (слой внимания и планирования) → VLA (слой трансляции плана в команды двигателям) (см. рис. 4). Вход – видеопоток и текстовая команда; выход – управляющие воздействия без промежуточных символических представлений [23]. Архитектура развивается на базе семилетней автопилотной платформы XPeng. Актуальным вопросом остается задержка (latency) многослойной сети в задачах реального времени.

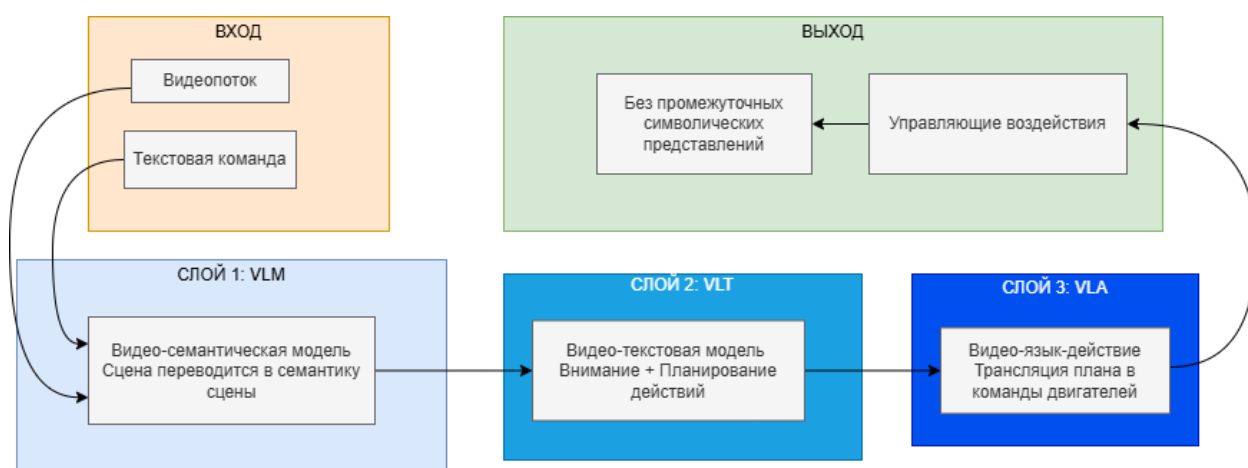


Рис. 4. Архитектура роботизированного устройства Xpeng Iron.

2.4. Figure AI

Система Figure AI основана на стратегии ресурсно-рационального стартапа: тонкая настройка готового предобученного решения вместо разработки базовой модели с нуля (см. рис. 5). Основой послужили модели OpenAI, на базе которых создана VLA-модель Helix [24]. Гибридная схема: имитационное обучение через

⁹ VLM (Vision-Language Model) – мультимодальная нейросетевая модель, обрабатывающая одновременно визуальные данные и текст.

¹⁰ VLT (Vision-Language Transformer) – трансформерная архитектура для совместной обработки зрительных и языковых данных; отвечает за механизм внимания и планирование действий.

телеоперацию реальных роботов (Helix Lab) + RL. Принципиальным ограничением является зависимость от внешней модели: архитектурные решения провайдера не находятся под контролем команды разработки.

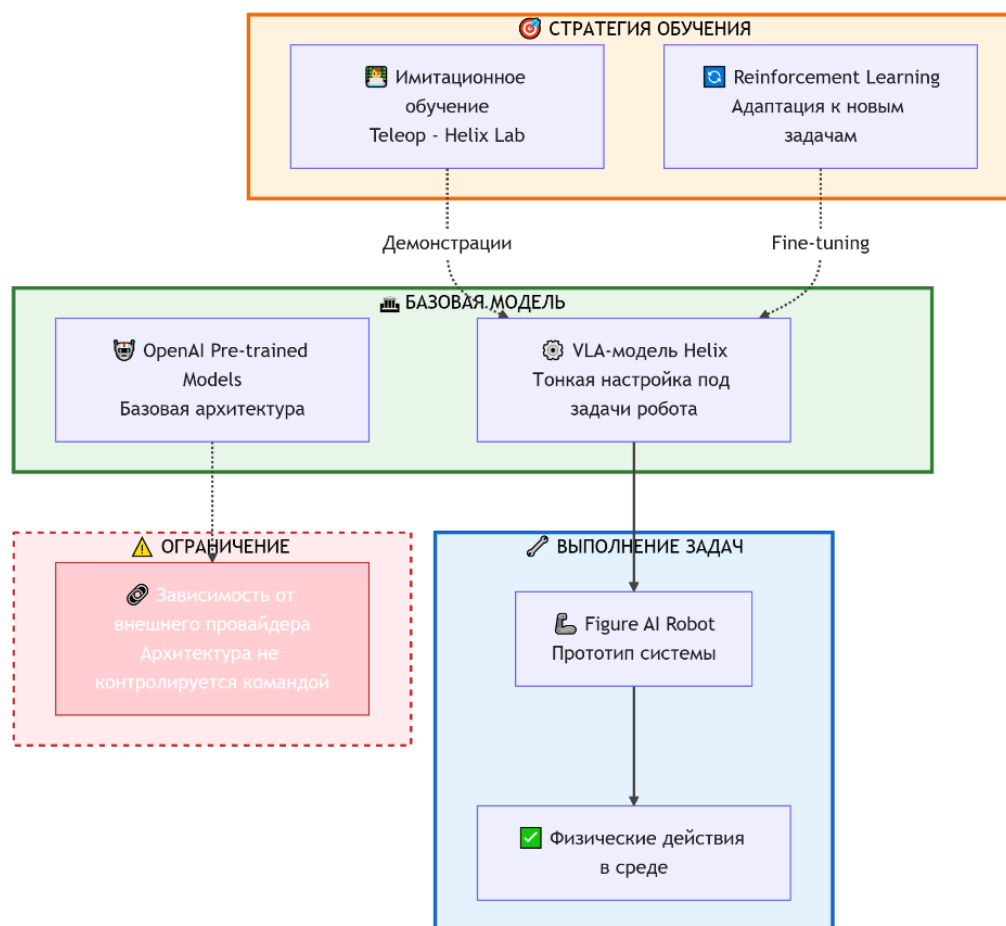


Рис. 5. Архитектура роботизированного устройства компании Figure AI.

2.5. GREEN (Сбер)

Система GREEN представляет нишу сервисных ассистентов с акцентом на диалог. Гибридная схема: GigaChat¹¹ в роли LLM-ядра (понимание команд, русскоязычный семантический контекст) + RL (адаптация к новым задачам) + VLA (связь зрительного восприятия с языковыми командами) [25].

Использование GigaChat – это обоснованное рыночное решение для русскоязычной среды. Ограничение состоит в том, что проект находится на ранней стадии, независимых метрик производительности недостаточно (см. рис. 6).

¹¹ GigaChat – большая языковая модель от Сбера. URL: <https://giga.chat/>

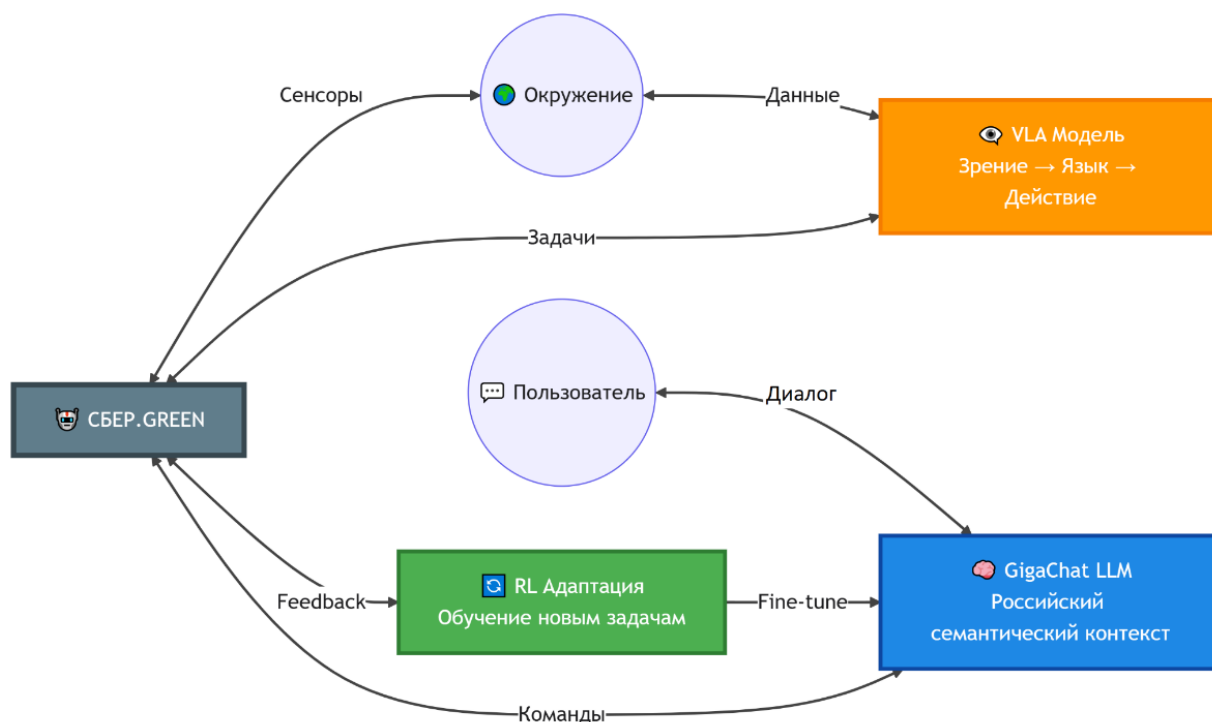


Рис. 6. Схема взаимодействия компонентов роботизированного устройства Tesla Optimus Gen2.

2.5. Сравнительный анализ и позиционирование работы

Представленные в обзоре роботизированные платформы и программные комплексы (подп. 2.1–2.5) демонстрируют разнообразие подходов к интеграции мультимодальных моделей и алгоритмов обучения с подкреплением. Каждый из рассмотренных проектов ориентирован на решение специфических задач автономного управления — от имитации действий человека до навигации в неструктурированной среде. Вместе с тем сопоставление их ключевых характеристик (см. табл. 1) выявляет общие технологические ограничения существующих решений. Как следует из табл. 1, ни одна из рассмотренных систем не предлагает воспроизводимого протокола, пригодного для ресурсно-ограниченных условий развёртывания. Система HumanPlus, наиболее близкая к постановке настоящей работы, демонстрирует высокую точность имитации, однако её применение сдерживается зависимостью от объёмного датасета демонстраций и ограниченной автономностью политика управления. Указанный исследовательский разрыв — отсутствие лёгковесного и воспроизводимого подхода к автономному

управлению на основе демонстраций оператора — определяет позиционирование настоящей работы.

Табл. 1. Сравнительный анализ гибридных подходов к обучению роботизированных устройств.

Критерий	Tesla Optimus	HumanPlus	XPeng Iron	Figure AI	GREEN
Класс	Маховик данных (data-flywheel)	Малобюджетный академический проект	Перенос автopilота (automotive transfer)	Настройка базовой модели (fine-tuning)	На основе LLM, локализованная
Проблема	Масштабирование данных без ручных демонстраций	Тяжело достичь значимого результата без дорогой инфраструктуры	Задержка в модульных системах зрения/движения	Высокая стоимость обучения с нуля	Диалог и русскоязычный контекст
Решение	E2E FSD + World Simulator + RL	RL (симуляция) + теневое копирование + имитация	VLM + VLT + VLA (трехслойный E2E)	Предобученные модели OpenAI + Helix VLA + RL	GigaChat LLM + RL + VLA
Источник данных	Флот 4+ млн авто (теневой режим)	40 ч видео оператора (WHAM)	Данные автopilота XPeng + оператор	Предобученные модели + телеоперация	GigaChat pretrain + симуляция
Ключевое ограничение	Корпоративная инфраструктура; непрозрачность автономности	Только продемонстрированные задачи	Задержка многослойной сети	Зависимость от внешней модели	Ранняя стадия; метрик нет
Воспроизводимость	Нет	Частичная (открытый код)	Нет	Нет	Нет

3. ВЫБОР ИНСТРУМЕНТАЛЬНОЙ БАЗЫ

3.1. Виртуальные среды

Виртуальные среды позволяют обучать роботизированные устройства в тысячи раз быстрее реального мира без риска повреждения физического прототипа. Были рассмотрены четыре платформы.

Webots¹² – профессиональный открытый симулятор с высокой точностью физики и готовыми моделями реальных роботов. Недостатки: слабый графический движок затрудняет обучение с компьютерным зрением; параллельный запуск сотен симуляций требует сложной серверной настройки.

NVIDIA Isaac Sim¹³ – наиболее мощная платформа, которая использует видеокарты для параллельного обучения тысячи роботов, также включает фотореалистичный рендеринг, специализированный модуль Isaac Gym для RL. Принципиальное ограничение – это минимальное требование к графическому процессору поколения RTX 3080.

Gazebo¹⁴ – стандарт де-факто для работы с реальными роботами: нативная интеграция с ROS¹⁵, физически корректные модели датчиков. Ограничена масштабируемость параллельного обучения.

Unity ML-Agents – сотни параллельных симуляций, широкий набор сенсорных моделей и нативная интеграция с VR-гарнитурами [26, 27]. Средние требования к аппаратуре. Недостаток: физический движок PhysX¹⁶ уступает специализированным решениям по точности. Для целей данной работы выбран Unity ML-Agents: доступность и поддержка VR-интерфейса обеспечивают полный цикл от записи демонстраций до автономного обучения на единой платформе.

3.2. Физическая симуляция в Unity

В процессе разработки были сравнены компоненты Rigidbody и ArticulationBody. Rigidbody обеспечивает базовую симуляцию столкновений и гравитации, однако без корректной симуляции трения суставов; применение FixedJoint как обходного решения приводило к нереалистичному «прилипанию»

¹² Webots – открытый профессиональный симулятор робототехники.

URL: <https://cyberbotics.com>

¹³ NVIDIA Isaac Sim – фотореалистичный симулятор роботизированных устройств на базе Omniverse. URL: <https://developer.nvidia.com/isaac/sim>

¹⁴ Gazebo – открытая среда симуляции робототехники, стандарт де-факто в экосистеме ROS. URL: <https://gazebo.org>

¹⁵ ROS (Robot Operating System) – открытая программная платформа для разработки и управления роботизированными системами.

¹⁶ PhysX – это кроссплатформенный физический движок, который имитирует законы реального мира в компьютерных играх и симуляторах.

захватываемых объектов. Компонент ArticulationBody (2022) разработан специально для кинематических цепочек роботизированных устройств и обеспечивает физически корректное поведение суставов без артефактов. Для задач настоящей работы был выбран ArticulationBody.

3.3. Роль VR-гарнитуры и захват движений оператора

VR-гарнитура выполняет двойную функцию. Во-первых, обеспечивает оператору стереоскопическое восприятие от камер роботизированного устройства, что повышает качество телеуправления и записываемых демонстраций. Во-вторых, контроллеры гарнитуры служат непосредственным инструментом захвата движений для первого этапа гибридного обучения. VR-интерфейс является ключевым звеном, соединяющим предыдущие работы авторов по телеуправлению с предлагаемым подходом автономного обучения.

Среди альтернатив рассматривались классические RGB-камеры (как, например, в HumanPlus – низкая стоимость, но ограниченная точность и проблема окклюзии) и маркерные мокап-костюмы (максимальная точность, однако требуют специальной комнаты и остаются недоступными для большинства групп). VR-гарнитуры предлагают оптимальный баланс: высокая точность, практически полное отсутствие окклюзии, доступная цена. Ограничениями являются: возможный кинетоз при длительных сессиях и отсутствие тактильной обратной связи.

4. РАЗРАБОТКА И ВЕРИФИКАЦИЯ ГИБРИДНОГО ПОДХОДА

4.1. Логика гибридной схемы

Предлагаемый нами подход построен на принципе последовательной компенсации недостатков: каждый следующий этап устраняет ограничение предыдущего (см. рис. 7). Ни один из компонентов не является достаточным сам по себе – их ценность реализуется именно в цепочке.

На первом этапе оператор управляет роботизированным устройством через VR-интерфейс: он видит мир глазами робота благодаря стереоизображению с камер устройства и управляет им контроллерами. Все действия записываются в виде парных данных «наблюдение с камеры → управляющее воздействие».

Нейросеть, обученная на этом датасете, сразу получает направленное поведение. Однако она ограничена «потолком» оператора: если в датасете нет примера данной ситуации, модель ошибается.

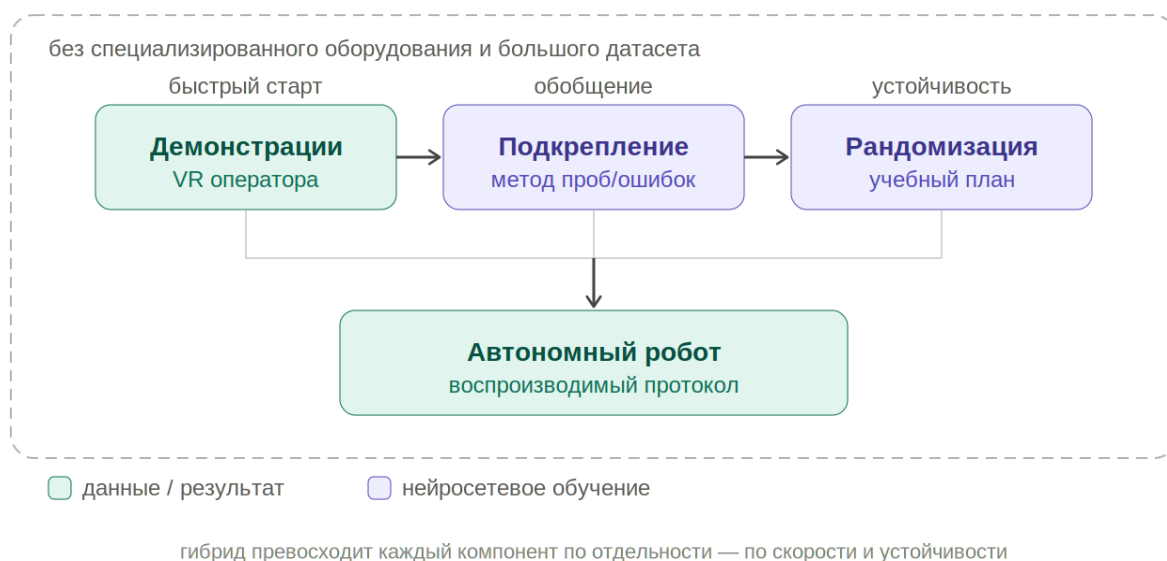


Рис. 7. Логика гибридного подхода: последовательная компенсация ограничений каждого этапа.

Второй этап снимает это ограничение. RL позволяет агенту исследовать пространство состояний самостоятельно, находить решения, которых оператор не демонстрировал, и превзойти его. Инициализация весами из первого этапа резко сокращает начальную фазу хаотичного поиска.

Третий этап формирует устойчивость. Доменная рандомизация совместно с учебным планом обеспечивает обучение на множестве вариаций среды с прогрессивно нарастающей сложностью. Схема на рис. 1 иллюстрирует логику цепочки и ее позиционирование: весь конвейер реализован на общедоступном инструментарии без специализированного оборудования.

4.2. Постановка задачи и условия экспериментов

В качестве тестовой задачи взята навигация роботизированного устройства к целевому объекту при наличии препятствия. Обучающая сцена содержит три элемента с изменяемым положением: агент, целевой объект (зеленая сфера) и препятствие (красная стена). Были проведены три контролируемых эксперимента на идентичной сцене с фиксированной функцией награды. По осям графиков: X – количество шагов (действий агента), Y – накопленная награда за эпизод.

4.3. Эксперимент 1: обучение на основе демонстраций

Средствами ML-Agents записан датасет демонстраций с вариациями взаимного расположения объектов. На рис. 8 представлен результат: награда стабильно растет с первых эпизодов и достигает плато к ~100–110 тыс. шагов – агент сразу имеет направленное поведение, унаследованное от демонстраций. Принципиальный недостаток – это низкая устойчивость: при изменении положений объектов за пределами вариаций датасета поведение деградирует.

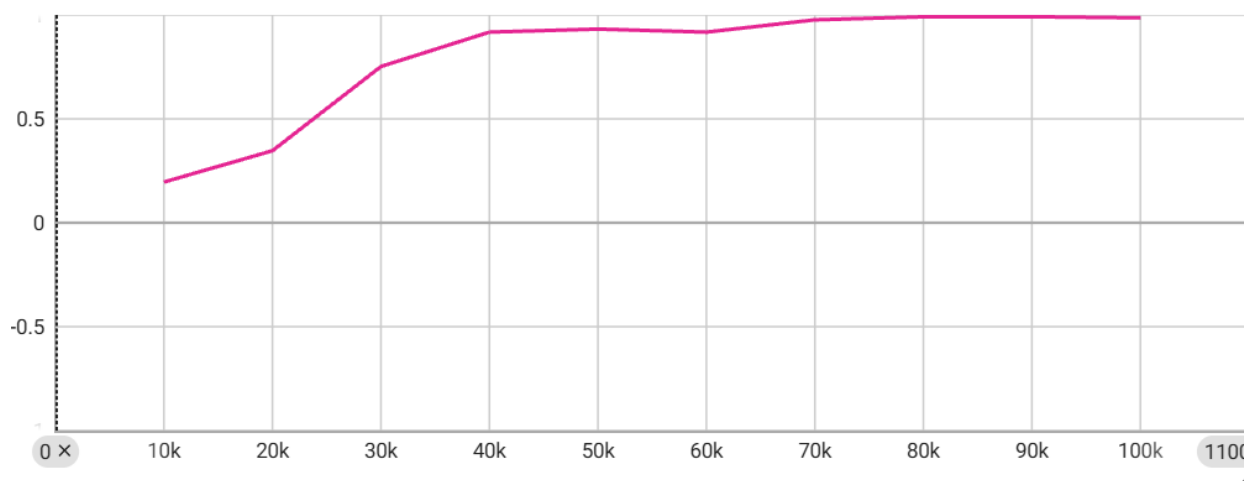


Рис. 8. Результат имитационного подхода обучения.

4.4. Эксперимент 2: обучение с подкреплением

Агент начинает с полностью случайных действий. На рис. 9 представлен результат: длительная фаза нулевой или отрицательной награды до ~250 тыс. шагов, затем резкий рост до плато к ~350–400 тыс. шагов. Достоинством является возможность превзойти оператора. Недостаток аналогичен: агент освоил конкретный маршрут, а не общую стратегию обхода.

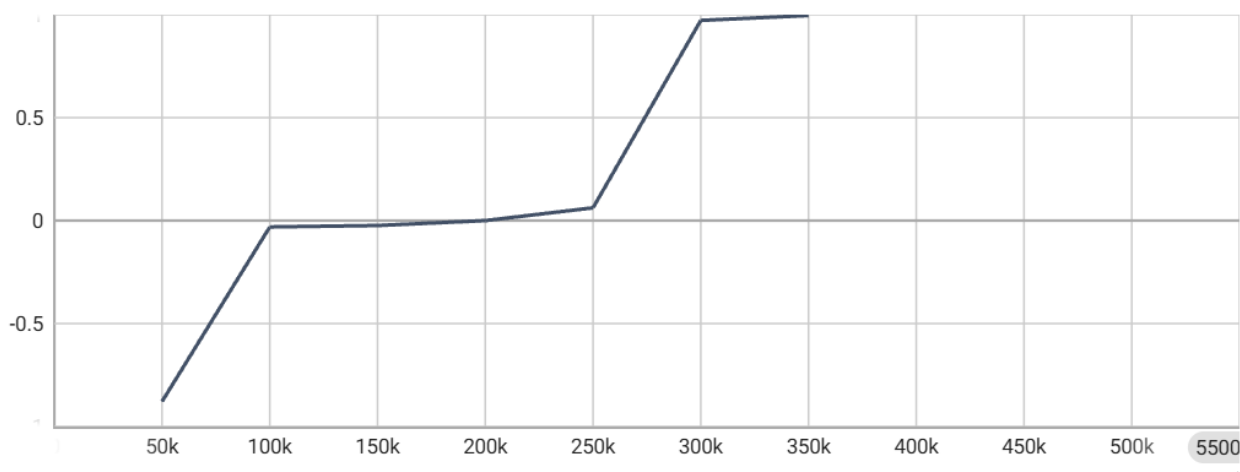


Рис. 9. Результат подхода обучения с подкреплением.

4.5. Эксперимент 3: гибридный подход

Гибридная схема была реализована в виде последовательного учебного плана.

1. Обучение на основе демонстраций в статичной сцене – инициализация направленного поведения.
2. RL с узким диапазоном рандомизации положений препятствия.
3. RL со средним диапазоном рандомизации всех ключевых объектов.
4. RL с рандомизацией по всей доступной обучающей среде.

График финального этапа показан на рис. 10: быстрый выход на положительную награду к ~75–100 тысячам шагов (от инициализации демонстрациями), повышаются устойчивость к изменениям окружающей среды и вариативность траекторий перемещения за счет незначительного снижения точности модели. Результирующая модель корректно выполняет навигацию при произвольном расположении всех объектов. Гибридный подход объединяет преимущества обоих компонентов: быстрый старт от демонстраций и устойчивость от RL с рандомизацией.

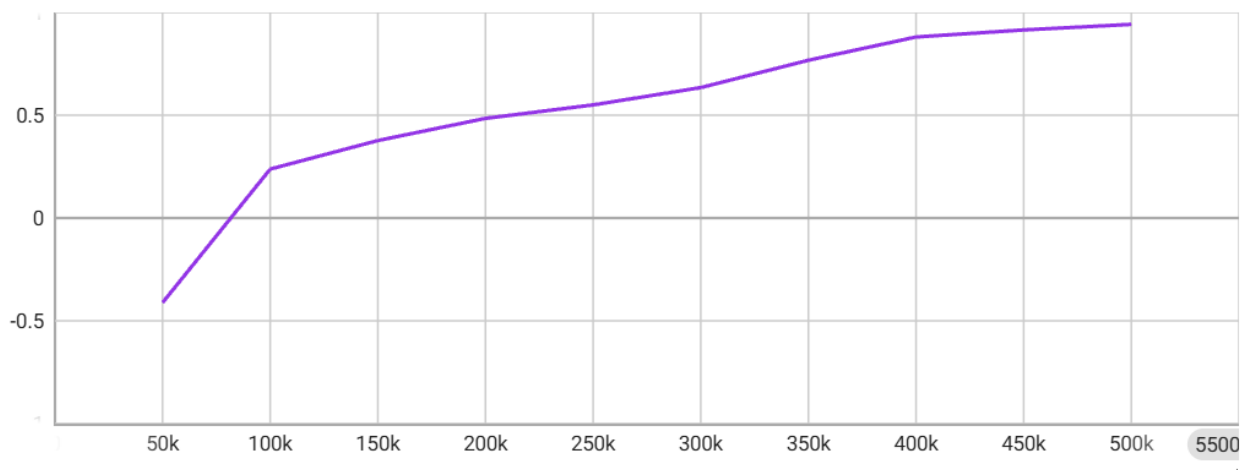


Рис. 10. Результат гибридного подхода обучения.

5. ОГРАНИЧЕНИЯ ПОДХОДА И НАПРАВЛЕНИЯ РАЗВИТИЯ

5.1. Катастрофическое забывание

Предложенный подход эффективен для обучения одному навыку с устойчивостью к изменениям среды. При переходе к мультизадачному обучению возникает системное ограничение – катастрофическое забывание: обучение новому навыку перезаписывает веса нейросети, разрушая ранее приобретенные [28]. Существуют три основных стратегии противодействия:

- буферы воспроизведения (Replay Buffers) [29] – встроенный механизм ML-Agents: данные о ранее освоенных задачах «подмешиваются» в обучающую выборку при освоении новых, постоянно напоминая сети о прошлых навыках. Не требует изменения архитектуры;
- мультизадачное обучение [30] – единая модель обучается на гетерогенном датасете, где текстовая команда задает условие текущей задачи. VLA-архитектура естественно поддерживает этот режим, однако требует значительного объема размеченных данных;
- иерархическое обучение [31] – навыки хранятся в отдельных модулях (подсетях), которыми управляет главная планировочная сеть по принципу «начальник – подчиненный». Позволяет добавлять новые навыки без перезаписи существующих, однако требует более сложной архитектуры.

Соответственно, для большей универсальности и многофункциональности можно комбинировать другие различные подходы, преимущества которых

можно объединить с другими для более эффективной работы в виде гибридного подхода.

5.2. Интеграция LLM и AI-агентов для генерации синтетических данных

Предложенный подход открывает перспективу применения LLM на этапе формирования обучающего датасета. LLM-агент может получать текстовое описание задачи и автоматически генерировать вариации демонстраций в виртуальной среде: сгенерированные последовательности действий проверяются функцией награды и при успехе добавляются в обучающую выборку.

Такой подход снял бы ключевое ограничение HumanPlus и настоящей работы – зависимость от объема реальных демонстраций оператора. Генерация синтетических демонстраций через LLM совместно с доменной рандомизацией потенциально позволяет покрыть значительно более широкое пространство состояний. Проверка этой гипотезы является следующим шагом исследования.

ЗАКЛЮЧЕНИЕ

Предложен и верифицирован воспроизводимый гибридный подход к обучению автономного управления роботизированным устройством на общедоступном инструментарии без специализированного оборудования. Сравнительный анализ пяти прототипов, представляющих ключевые классы гибридных архитектур: маховик данных, малобюджетная академическая разработка, перенос автопилота, настройка предобученной базовой модели, система на основе LLM – показал: все успешные системы применяют гибридные схемы, однако ни одна не предоставляет воспроизводимого протокола для ресурсно-ограниченных условий.

Три контролируемых эксперимента в Unity ML-Agents подтвердили основной тезис: гибридная комбинация обучения на основе демонстраций, RL, учебного плана и доменной рандомизации превосходит изолированные подходы по скорости выхода на целевую награду и устойчивости к изменениям среды.

Предложенный подход формирует замкнутую цепочку с нашими предыдущими работами: от VR-телеуправления и записи стереоизображения к автономному обучению на основе собранных демонстраций. VR-интерфейс выступает не просто инструментом управления, а источником высококачественных обучающих данных.

Обозначим перспективы дальнейшей работы: применение эвристики эскалации для преодоления катастрофического забывания; использование LLM-агентов для автоматической генерации синтетических демонстраций; перенос обученной модели на физический прототип с оценкой sim-to-real gap.

Благодарности

Публикация выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения проекта № FZSM-2026-0019 «Телеуправление удаленными устройствами с использованием виртуальной реальности».

СПИСОК ЛИТЕРАТУРЫ

1. Кугуракова В.В., Хафизов М.Р., Кадыров С.А., Зыков Е.Ю. Удаленное управление роботизированным устройством с использованием технологий виртуальной реальности // Программные продукты и системы. 2022. Т. 35, № 3. С. 348–361. <https://doi.org/10.15827/0236-235X.139.348-361>
2. Кугуракова В.В. и др. VR-телеуправление «многорукими» устройствами: проблемы, гипотезы, постановка задачи // Электронные библиотеки. 2022. Т. 25, № 5. С. 441–471. <https://doi.org/10.26907/1562-5419-2022-25-5-441-471>
3. Зыкова А.Е., Зыков Е.Ю., Кугуракова В.В. Применение технологий виртуальной реальности для удаленного управления роботизированными устройствами // Итог. науч.-практ. конф. КФУ. Казань, 2023. С. 87. EDN GDBPQR.
4. Сергунин И.Д., Зыков Е.Ю. «РУСАК» – модульное решение системы управления удаленными устройствами с интегрированными защищенными каналами связи // Состояние и перспективы развития современной науки по направлению «Робототехника». 2023. С. 265–271.
5. Кугуракова В.В., Исхаков Р.Т., Низамутдинов М.М. и др. Точечное телеуправление для сельскохозяйственной техники, выполняющей задачи на основе ИИ // Международный форум KAZAN DIGITAL WEEK – 2025: сборник материалов, Часть 1. 2025. С. 1964–1971.
6. Программа для синхронизации стереоизображения со стереопары, размещенной на удаленном двигающемся устройстве: Свидетельство о гос. рег.

программы для ЭВМ № 2022684277 / Кугуракова В.В., Хафизов М.Р., Габдуллина Д.Р.; КФУ. Зарег. 13.12.2022.

7. Программа адаптивного управления типовыми маршрутами и рутинными действиями модели робота-манипулятора в виртуальной среде: Свидетельство о гос. рег. программы для ЭВМ № 2024689377 / Кугуракова В.В., Исхаков Р.Т., Костюк Д.И. и др.; КФУ. Зарег. 05.12.2024.

8. *Исхаков Р.Т.* Разработка генератора большого количества вариаций движений роботизированного устройства на основе базовых записанных действий оператора: выпускная квалификационная работа: 09.03.04 «Программная инженерия» / Р.Т. Исхаков; Казанский (Приволжский) федеральный университет. Казань, 2024. 41 с.

9. *Guzaerov D.A., Sergunin I.D., Kugurakova V.V.* Infrared Video-Based Object Recognition for Agricultural Research Field Uses // BIO Web of Conferences. EDP Sciences, 2024. Vol. 141. No. 01025. <https://doi.org/10.1051/bioconf/202414101025>

10. *Zare M. et al.* A survey of imitation learning: Algorithms, recent developments, and challenges // IEEE Transactions on Cybernetics. 2024. Vol. 54, No. 12. P. 7173–7186. <https://doi.org/10.48550/arXiv.2309.02473>

11. *Arulkumaran K. et al.* Deep Reinforcement Learning: A Brief Survey // IEEE Signal Processing Magazine. 2017. Vol. 34, No. 6. P. 26–38. <https://doi.org/10.1109/MSP.2017.2743240>

12. *Kober J., Bagnell J.A., Peters J.* Reinforcement Learning in Robotics: A Survey // The International Journal of Robotics Research. 2013. Vol. 32, No. 11. P. 1238–1274. <https://doi.org/10.1177/0278364913495721>

13. *Chebotar Y. et al.* Actionable models: Unsupervised offline reinforcement learning of robotic skills // arXiv preprint arXiv:2104.07749. 2021.

14. *Bengio Y. et al.* Curriculum Learning // Proceedings of ICML. 2009. P. 41–48. <https://doi.org/10.1145/1553374.1553380>

15. *Graves A. et al.* Automated Curriculum Learning for Neural Networks // Proceedings of ICML. 2017. P. 1311–1320. <https://doi.org/10.48550/arXiv.1704.03003>

16. *Chukwurah N., Adebayo A. S., Ajayi O.O.* Sim-to-real transfer in robotics: Addressing the gap between simulation and real-world performance // International Journal of Robotics and Simulation. 2024. Vol. 6, No. 1. P. 89–102.

<https://doi.org/10.54660/.IJFMR.2024.5.1.33-39>

17. *Tobin J. et al.* Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017. P. 23–30. doi: 10.1109/IROS.2017.8202133

18. *Pinto L. et al.* Robust Adversarial Reinforcement Learning // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017.

<https://doi.org/10.48550/arXiv.1703.02702>

19. *Abnar S. et al.* Exploring the limits of large scale pre-training // arXiv preprint arXiv:2110.02095. 2021.

20. Tesla Optimus // teslarobots.ru. 2025. URL: <https://teslarobots.ru/> (дата обращения: 05.02.2026).

21. *Brohan A. et al.* RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control // arXiv preprint arXiv:2307.15818. 2023.

22. *Fu Z. et al.* Humanplus: Humanoid shadowing and imitation from humans // arXiv preprint arXiv:2406.10454. 2024.

23. Xpeng Shares Achievements in Physical AI: Unveils XPeng VLA 2.0, Next-Gen Iron // xpeng.com. 2025. URL: <https://www.xpeng.com/pressroom/news/019a56f54fe99a2a0a8d8a0282e402b7> (дата обращения: 04.03.2026).

24. Figure AI Opens New Helix Lab to Accelerate Learning for Its Humanoid Robots // mlq.ai. 2025. URL: <https://mlq.ai/news/figure-ai-opens-new-helix-lab-to-accelerate-learning-for-its-humanoid-robots/> (дата обращения: 04.03.2026).

25. *Apanasevich I. et al.* Green-VLA: Staged Vision-Language-Action Model for Generalist Robots // arXiv preprint arXiv:2602.00919. 2026.

26. *Yu A., Palefsky-Smith R., Bedi R.* Deep reinforcement learning for simulated autonomous vehicle control // Course Project Reports: Winter. 2016. Vol. 2016. P. 1–7. URL: https://cs231n.stanford.edu/reports/2016/pdfs/112_Report.pdf

27. *Done C. et al.* Reinforcement Learning-Driven Prosthetic Hand Actuation in a Virtual Environment Using Unity ML-Agents // Virtual Worlds. MDPI. 2025. Vol. 4, No. 4. P. 53. <https://doi.org/10.3390/virtualworlds4040053>

28. *Kirkpatrick J. et al.* Overcoming Catastrophic Forgetting in Neural Networks // Proceedings of the National Academy of Sciences. 2017. Vol. 114, No. 13. P. 3521–3526. <https://doi.org/10.1073/pnas.1611835114>

29. Fedus W. et al. Revisiting fundamentals of experience replay // International conference on machine learning. PMLR. 2020. P. 3061–3071. <https://doi.org/10.48550/arXiv.2007.06700>
30. Caruana R. Multitask Learning // Machine Learning. 1997. Vol. 28, No. 1. P. 41–70. <https://doi.org/10.1023/A:1007379606734>
31. Pateria S. et al. Hierarchical reinforcement learning: A comprehensive survey // ACM Computing Surveys (CSUR). 2021. Vol. 54, No. 5. P. 1–35. <https://doi.org/10.1145/3453160>
-

A HYBRID APPROACH TO AUTONOMOUS CONTROL OF A ROBOTIC DEVICE THAT COMBINES REINFORCEMENT LEARNING AND OPERATOR-BASED DEMONSTRATION LEARNING

R. T. Iskhakov¹ [0009-0004-9262-7296], V. V. Kugurakova² [0000-0002-1552-4910]

^{1, 2}Kazan Federal University, Kazan, Russia

¹vlada.kugurakova@gmail.com, ²iskhakovrobert42@gmail.com

Abstract

A reproducible hybrid approach to learning autonomous control of a robotic device has been proposed and experimentally verified, implemented using publicly available tools without specialized computing hardware. The approach combines learning from operator demonstrations, reinforcement learning, curriculum learning, and domain randomization into a single sequential framework. Based on a comparative analysis of five prototypes representing key classes of architectural solutions, applicable approaches for resource-constrained environments are identified. Three controlled experiments in Unity ML-Agents demonstrated the superiority of the hybrid combination over isolated approaches in terms of both learning speed and robustness to environmental changes. Limitations associated with catastrophic forgetting are outlined, and prospects for integrating large language models to generate synthetic training data are identified.

Keywords: teleoperation, remote control, robot, virtual reality, virtual environment, machine learning, reinforcement learning, learning from demonstrations, operator demonstration, Unity, ML-Agents.

Acknowledgments

This work was funded by a grant from the Academy of Sciences of the Republic of Tatarstan (Agreement No. 12/2025-PD-KFU dated December 22, 2025).

REFERENCES

1. Kugurakova V.V., Khafizov M.R., Kadyrov S.A., Zykov E.Yu. Remote control of a robotic device using virtual reality // *Software & Systems*. 2022. Vol. 35, No. 3. P. 348–361. <https://doi.org/10.15827/0236-235X.139.348-361>
2. Kugurakova V.V. et al. VR-Telecontrol of Multi-Arm Devices: Problems, Hypotheses, Problem Statement // *Russian Digital Libraries Journal*. 2022. Vol. 25, No. 5. P. 441–471. <https://doi.org/10.26907/1562-5419-2022-25-5-441-471>
3. Zykova A.E., Zykov E.Yu., Kugurakova V.V. *Primenenie tekhnologiy virtualnoy realnosti dlya udalennogo upravleniya robotizirovannymi ustroystvami* // *Itogovaya nauchno-prakticheskaya konferentsiya KFU. Kazan, 2023*. P. 87. EDN GDBPQP.
4. Sergunin I.D., Zykov E.Yu. «RUSAK» – modul’noe reshenie sistemy upravleniya udalonnymi ustroystvami s integrirovannymi zashchishchyonnymi kanalami svyazi // *Sostoyanie i perspektivy razvitiya sovremennoj nauki po napravleniyu «Robototekhnika»*. 2023. P. 265–271.
5. Kugurakova V.V., Iskhakov R.T. et al. *Tochechnoe teleupravlenie dlya sel’skokhozyajstvennoj tekhniki, vypolnyayushchej zadachi na osnove II* // *International Forum KAZAN DIGITAL WEEK – 2025: conference papers, volume 1*. 2025. P. 1964–1971.
6. *Programma dlya sinkhronizatsii stereoizobrazheniya so stereopary, razmeshchyonnoj na udalennom dvigayushchemsya ustrojstve: Svidetel’stvo o gos. reg. programmy dlya EVM № 2022684277* / Kugurakova V.V., Khafizov M.R. Gabdullina D.R.; KFU. Date of registration: 13.12.2022.
7. *Programma adaptivnogo upravleniya tipovymi marshrutami i rutinnyimi dejstviyami modeli robota-manipulyatora v virtual’noj srede: Svidetel’stvo o gos. reg. programmy dlya ЭБМ № 2024689377* / Kugurakova V.V., Iskhakov R.T. et al.; KFU. Date of registration: 05.12.2022.
8. Iskhakov R.T. *Razrabotka generatora bol’shogo kolichestva variatsij dvizhenij robotizirovannogo ustrojstva na osnove bazovykh zapisannykh dejstvij op-*

eratora: vypusknaya kvalifikatsionnaya rabota: 09.03.04 «Programmnaya inzheneriya» / R.T. Iskhakov; Kazan (Volga Region) Federal University. Kazan, 2024. 41 p.

9. *Guzaerov D.A., Sergunin I.D., Kugurakova V.V.* Infrared Video-Based Object Recognition for Agricultural Research Field Uses // BIO Web of Conferences. EDP Sciences, 2024. Vol. 141. No. 01025. <https://doi.org/10.1051/bioconf/202414101025>
10. *Zare M. et al.* A survey of imitation learning: Algorithms, recent developments, and challenges // IEEE Transactions on Cybernetics. 2024. Vol. 54, No. 12. P. 7173–7186. <https://doi.org/10.48550/arXiv.2309.02473>
11. *Arulkumaran K. et al.* Deep Reinforcement Learning: A Brief Survey // IEEE Signal Processing Magazine. 2017. Vol. 34, No. 6. P. 26–38. <https://doi.org/10.1109/MSP.2017.2743240>
12. *Kober J., Bagnell J.A., Peters J.* Reinforcement Learning in Robotics: A Survey // The International Journal of Robotics Research. 2013. Vol. 32, No. 11. P. 1238–1274. <https://doi.org/10.1177/0278364913495721>
13. *Chebatar Y. et al.* Actionable models: Unsupervised offline reinforcement learning of robotic skills // arXiv preprint arXiv:2104.07749. 2021.
14. *Bengio Y. et al.* Curriculum Learning // Proceedings of ICML. 2009. P. 41–48. <https://doi.org/10.1145/1553374.1553380>
15. *Graves A. et al.* Automated Curriculum Learning for Neural Networks // Proceedings of ICML. 2017. P. 1311–1320. <https://doi.org/10.48550/arXiv.1704.03003>
16. *Chukwurah N., Adebayo A.S., Ajayi O.O.* Sim-to-real transfer in robotics: Addressing the gap between simulation and real-world performance // International Journal of Robotics and Simulation. 2024. Vol. 6, No. 1. P. 89–102. <https://doi.org/10.54660/IJFMR.2024.5.1.33-39>
17. *Tobin J. et al.* Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017. P. 23–30. <https://doi.org/10.1109/IROS.2017.8202133>
18. *Pinto L. et al.* Robust Adversarial Reinforcement Learning // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. <https://doi.org/10.48550/arXiv.1703.02702>

19. *Abnar S. et al.* Exploring the limits of large scale pre-training // arXiv preprint arXiv:2110.02095. 2021.
20. Tesla Optimus // teslarobots.ru. 2025. URL: <https://teslarobots.ru/> (accessed at 05.02.26)
21. *Brohan A. et al.* RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control // arXiv preprint arXiv:2307.15818. 2023.
22. *Fu Z. et al.* Humanplus: Humanoid shadowing and imitation from humans // arXiv preprint arXiv:2406.10454. 2024.
23. Xpeng Shares Achievements in Physical AI: Unveils XPeng VLA 2.0, Next-Gen Iron // xpeng.com. 2025. URL: <https://www.xpeng.com/pressroom/news/019a56f54fe99a2a0a8d8a0282e402b7> (accessed: March 4, 2026)
24. Figure AI Opens New Helix Lab to Accelerate Learning for Its Humanoid Robots // mlq.ai. 2025. URL: <https://mlq.ai/news/figure-ai-opens-new-helix-lab-to-accelerate-learning-for-its-humanoid-robots/> (accessed: March 4, 2026)
25. *Apanasevich I. et al.* Green-VLA: Staged Vision-Language-Action Model for Generalist Robots // arXiv preprint arXiv:2602.00919. 2026.
26. *Yu A., Palefsky-Smith R., Bedi R.* Deep reinforcement learning for simulated autonomous vehicle control // Stanford University Course Project Reports: Winter. 2016. P. 1–7.
URL: https://cs231n.stanford.edu/reports/2016/pdfs/112_Report.pdf
27. *Done C. et al.* Reinforcement Learning-Driven Prosthetic Hand Actuation in a Virtual Environment Using Unity ML-Agents // Virtual Worlds. MDPI. 2025. Vol. 4, No. 4. P. 53. <https://doi.org/10.3390/virtualworlds4040053>
28. *Kirkpatrick J. et al.* Overcoming Catastrophic Forgetting in Neural Networks // Proceedings of the National Academy of Sciences. 2017. Vol. 114, No. 13. P. 3521–3526. <https://doi.org/10.1073/pnas.1611835114>
29. *Fedus W. et al.* Revisiting fundamentals of experience replay // International conference on machine learning. 2020. P. 3061–3071.
<https://doi.org/10.48550/arXiv.2007.06700>
30. *Caruana R.* Multitask Learning // Machine Learning. 1997. Vol. 28. P. 41–70. <https://doi.org/10.1023/A:1007379606734>
31. *Pateria S. et al.* Hierarchical reinforcement learning: A comprehensive survey // ACM Computing Surveys (CSUR). 2021. Vol. 54, No. 5. P. 1–35.

<https://doi.org/10.1145/3453160>

СВЕДЕНИЯ ОБ АВТОРАХ



ИСХАКОВ Роберт Тагирович – лаборант-исследователь кафедры индустрии разработки видеоигр Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, магистр. Зона научных интересов: формальные методы верификации и переноса обученных моделей на разнородные агенты.

Robert Tagirovich ISKHAKOV – research assistant at the department of Video Game Development Industry at the Institute of Information Technology and Intelligent Systems (ITIS), Kazan Federal University (KFU), Master's degree in Software Engineering. Research interests include the formal methods for verification and transfer of trained models to heterogeneous agents.

email: iskhakovrobert42@gmail.com

ORCID: 0009-0004-9262-7296



КУГУРАКОВА Влада Владимировна – кандидат технических наук, доцент, зав. кафедрой индустрии разработки видеоигр (КИРВИ) Института информационных технологий и интеллектуальных систем (ИТИС) Казанского федерального университета (КФУ). Область научных интересов – формальные методы верификации видеоигр.

Vlada Vladimirovna KUGURAKOVA – PhD, Head of the Department of Video Game Development Industry (GD Chair) at the Institute of Information Technology and Intelligent Systems (ITIS) of Kazan Federal University (KFU). Research interests include the formal methods of video game verification.

email: vlada.kugurakova@gmail.com

ORCID: 0000-0002-1552-4910

Материал поступил в редакцию 25 июня 2026 года