

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ АУДИОДИСКРИПЦИИ КУЛЬТУРНОГО КОНТЕНТА ДЛЯ СЛАБОВИДЯЩИХ

Протченко Е.В., Колодич А.Д.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь,
protchenko@bsuir.by*

В статье рассматривается проблема ограниченного доступа незрячих и слабовидящих людей к аудиовизуальному контенту как к инструменту культурного и профессионального образования. Обосновывается необходимость применения технологий искусственного интеллекта (ИИ) для массового создания аудиодискрипции (тифлокомментирования). Описывается проект, направленный на разработку ИИ-алгоритмов для генерации описаний визуального ряда в фильмах, спектаклях и экранизациях, что позволит создать коллекцию адаптированного контента для повышения культурного интеллекта данной социальной группы. Проект включает в себя как техническую разработку, так и непосредственную работу с целевой аудиторией для выявления потребностей.

Ключевые слова: искусственный интеллект; инклюзивное образование; культурный интеллект; аудиодискрипт; алгоритмы компьютерного зрения; нейросети; люди с ограниченными возможностями по зрению.

Современный этап цифровизации, охвативший все сферы жизни, выдвигает новые требования к непрерывному культурному и профессиональному развитию. Однако для людей с нарушениями зрения этот процесс сопряжён с системными барьерами, ведущими к существенному неравенству возможностей. Основная причина заключается в аудиовизуальной природе значительного массива современного образовательного и культурного контента – фильмов, видеозаписей спектаклей, лекций. Для незрячего человека они зачастую сводятся к фрагментарному восприятию звуковой дорожки, лишённой ключевой визуальной информации: действий персонажей, смены обстановки, невербальной коммуникации и художественных приёмов. Возникающий «информационный вакуум» препятствует полноценному восприятию сюжета, пониманию авторского замысла и, как следствие, ограничивает возможности для образования и социальной интеграции.

Существующее решение – традиционная аудиодискрипция (тифлокомментирование) – представляет собой дорогостоящий и трудоёмкий ручной труд, что делает массовую адаптацию произведений, особенно современных, экономически нецелесообразной. В результате люди с нарушениями зрения оказываются исключёнными из актуального культурного дискурса. Перспективным решением данной проблемы является использование технологий искусственного интеллекта (ИИ), способных автоматизировать анализ и описание визуального ряда.

Актуальность проблемы и теоретическое обоснование. В рамках исследования культурный интеллект понимается как способность успешно взаимодействовать на границе культур, включая когнитивный, мотивационный и поведенческий компоненты [1, 2].

Кинематограф и театр для незрячего человека являются именно такой «границей культур», преодолеть которую без специального посредника практически невозможно. Аудиодискрипция, выступая этим посредником, не только делает контент доступным, но и способствует развитию культурного интеллекта, формируя понимание культурных кодов и стимулируя интерес к сложному контенту.

Доступ к культурным ресурсам напрямую связан с профессиональным образованием, которое рассматривается как непрерывное развитие компетенций личности [3, 4]. Многие современные профессии требуют понимания визуального языка медиа, что изначально малодоступно для незрячих специалистов. Кроме того, исторические фильмы, экранизации и записи спектаклей сами по себе являются важными образовательными ресурсами.

Следовательно, создание доступного контента является практической реализацией принципов инклюзивного образования [5, 6], расширяя его рамки за пределы формальных учреждений в сферу непрерывного культурного просвещения. Аудиодискрипт – это лаконичное вербальное описание визуальных элементов, специально созданное для передачи содержания, недоступного без словесных пояснений [7, 8]. Это не просто звуковая дорожка, а специализированная технология, которая, соблюдая принцип компрессии информации [8], органично встраивается в паузы между диалогами, «озвучивая» визуальный мир и обеспечивая доступ к культурным ценностям.

Проект по созданию коллекции фильмов с ИИ-аудиодискриптом: концепция и методология. На основе анализа научной литературы и результатов эмпирического исследования, проведенного в рамках диалоговой площадки в обществе слепых г. Пинск, был разработан проект по созданию коллекции фильмов с аудиодискриптом с помощью ИИ. Диалог с представителями целевой аудитории показал острую потребность в доступе к современному кинематографу и видеозаписям театральных постановок, которые практически не адаптируются традиционными методами.

Респонденты отмечали, что существующие адаптированные материалы часто ограничиваются устаревшим репертуаром или отдельными популярными фильмами, в то время как доступ к современным кинолентам, артхаусному кино и видеозаписям спектаклей остается крайне ограниченным. Это создает ощущение культурной изоляции и неравенства.

Цель проекта является повышение культурного интеллекта и обеспечение возможностей для непрерывного образования людей с ограниченными возможностями по зрению через массовое и оперативное создание качественной аудиодискрипции для фильмов, спектаклей и экранизаций с использованием алгоритмов искусственного интеллекта.

Проект реализуется в три взаимосвязанных этапа:

Первый этап – технологический – включает разработку и обучение алгоритмов ИИ для анализа визуального контента, и генерации текстовых описаний. На этом этапе применяются методы машинного обучения, компьютерного зрения и обработки естественного языка.

Второй этап – эмпирический – предполагает организацию фокус-групп и интервью с представителями целевой аудитории для выявления потребностей и получения обратной связи о качестве создаваемых материалов.

Третий этап – интеграционный – направлен на создание устойчивой модели распространения и использования адаптированного контента, а также на оценку эффективности проекта с точки зрения его влияния на культурное развитие участников.

Техническая реализация проекта. Техническое ядро проекта составляет программный комплекс, реализующий сквозной пайплайн обработки видео и генерации аудиодискрипции. Архитектура системы включает три ключевых модуля, последовательно обрабатывающих видеоконтент.

Модуль компьютерного зрения отвечает за анализ визуального ряда. На этом этапе используются современные алгоритмы детекции объектов (*YOLO*, *Faster R-CNN*), которые идентифицируют ключевые объекты в кадре: персонажей, предметы, элементы обстановки. Дополнительно применяются алгоритмы распознавания действий и активностей, позволяющие определить, что именно делают персонажи (идут, бегут, жестикулируют и т.д.).

Важной составляющей является анализ эмоционального состояния персонажей через

распознавание мимики. Это особенно значимо для передачи тонкостей актерской игры в театральных постановках. Также модуль отслеживает смену локаций, определяет время суток и погодные условия, если они *relevant* для сюжета, и распознает текстовые вставки (титры, надписи), которые часто не озвучиваются, но несут важную смысловую нагрузку.

Модуль обработки естественного языка (*NLP*) преобразует структурированные данные от модуля компьютерного зрения в связный текстовый описательный трек. На этом этапе используются передовые языковые модели на архитектуре *Transformer*, предварительно обученные на обширных корпусах текстов, включая специализированные коллекции профессиональной аудиодискрипции [8].

Ключевой задачей этого модуля является обеспечение лаконичности и релевантности описаний. Алгоритм должен отбирать наиболее значимые визуальные элементы для вербализации, опуская малозначительные детали. Важным аспектом является соблюдение принципа компрессии информации – описания должны уместиться в естественные паузы между диалогами, не нарушая ритма произведения.

Модуль также отвечает за обеспечение связности и плавности повествования, использование соответствующего стиля и лексики, адекватных жанру и тональности произведения.

Модуль синтеза речи (*Text-to-Speech*) преобразует итоговый текстовый описательный трек в озвучку. На этом этапе используются современные нейросетевые модели синтеза речи, способные генерировать естественное, человеческое звучание с правильной интонацией, расстановкой акцентов и эмоциональной окраской.

Качество синтезированной речи имеет критическое значение для комфортного восприятия контента. Современные TTS-системы позволяют достичь высокой степени натуральности звучания, что делает восприятие более комфортным и естественным.

Интеграция этих трех модулей в единый сквозной пайплайн позволяет автоматизировать процесс создания аудиодискрипции и значительно сократить время адаптации контента по сравнению с ручными методами. *Общий процесс обработки видео и генерации аудиодискрипции, реализуемый предлагаемой системой, представлен на рис. 1.*

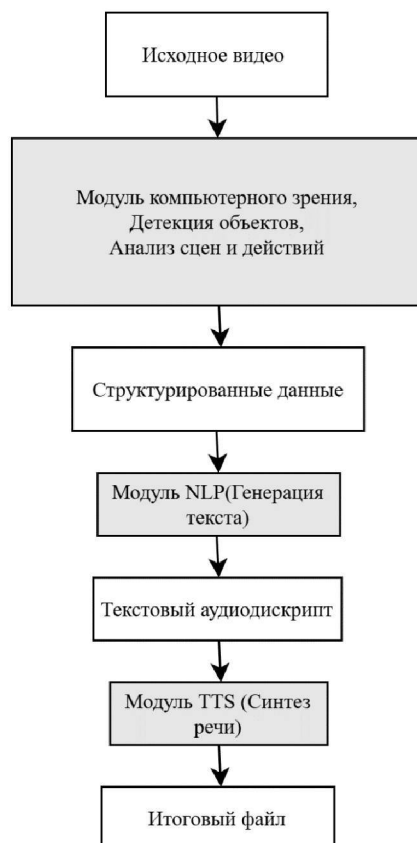


Рисунок 4 - Архитектура ИИ-пайплайна для создания аудиодискрипции

Техническая разработка ведется в тесном контакте с конечными пользователями – людьми с нарушениями зрения. Фокус-группы проводятся на регулярной основе для выявления потребностей и предпочтений целевой аудитории. В рамках этих обсуждений определяется, какие именно жанры, фильмы и спектакли наиболее востребованы для адаптации. Участники помогают определить, какие визуальные элементы являются наиболее значимыми для понимания сюжета, а какие могут быть опущены без ущерба для восприятия.

Индивидуальное тестирование прототипов позволяет получить детальную обратную связь о качестве сгенерированных описаний. Участникам предлагается прослушать фрагменты адаптированных произведений и оценить их по нескольким критериям: понятность и ясность описаний, их соответствие происходящему на экране, естественность темпа речи, отсутствие конфликта с диалогами и другими звуковыми элементами.

Важным аспектом тестирования является оценка эмоциональной составляющей описаний – насколько успешно они передают настроение сцены, эмоции персонажей, драматизм или комичность ситуации.

Участники, регулярно потребляющие материалы с ИИ-аудиодискрипцией, отмечают изменения в своем культурном опыте, расширение кругозора, повышение уверенности в обсуждении культурных тем в кругу зрячих друзей и коллег.

Обратная связь от пользователей является итерационной и позволяет непрерывно улучшать алгоритмы, корректируя их работу в соответствии с реальными потребностями и предпочтениями целевой аудитории.

В перспективе проект предполагает расширение функциональности системы. Планируется разработка модуля адаптации визуальных учебных материалов – графиков, диаграмм, схем – которые часто используются в образовательном процессе и в профессиональной деятельности. Это откроет дополнительные возможности для получения качественного образования и профессиональной реализации людей с нарушениями зрения.

Еще одним направлением развития является создание инструментов для адаптации живых театральных постановок в реальном времени с использованием технологий дополненной реальности и потоковой обработки видео. Это позволит людям с нарушениями зрения посещать театры и другие культурные мероприятия наравне со зрячими зрителями.

Предложенный проект демонстрирует практический путь реализации принципов инклюзивного образования в сфере культуры. Использование искусственного интеллекта для создания аудиодискрипции позволяет перейти от единичных ручных адаптаций к массовому производству доступного контента. Это не только расширяет возможности для досуга, но и предоставляет людям с ограниченными возможностями по зрению мощный инструмент для развития культурного интеллекта и непрерывного профессионального роста.

Важно подчеркнуть, что технология не рассматривается как полная замена работы профессиональных аудиодескрипторов. Наиболее перспективной видится гибридная модель, при которой ИИ выполняет первоначальную обработку контента, а человек-специалист осуществляет финальную редактуру и шлифовку.

Развитие подобных технологий способствует не только решению конкретной проблемы доступности культурного контента, но и формированию более инклюзивного общества, в котором каждый человек имеет равные возможности для культурного развития и профессиональной самореализации независимо от особенностей здоровья.

Литература

1. Ливермор, Д. Культурный интеллект. Почему он важен для успешного управления и как его развить / Д. Ливермор. – Москва: Манн, Иванов и Фербер, 2023. – 304 с.
2. Soldatova, G. U. Cultural Intelligence as a Socio-Psychological Phenomenon / G. U. Soldatova, S. V. Chigarkova // National Psychological Journal. – 2018. – № 4 (32). – С. 27–38.
3. Петров, А.Я. Профессиональное образование и обучение работников (персонала): правовые основы / А. Я. Петров. – Москва: Юрайт, 2020. – 305 с.
4. Белозерцев, Е. П. Педагогика профессионального образования / Е. П. Белозерцев. – Москва: Академия, 2010. – 256 с.
5. Bunch, G. Inclusive Education: The Basics / G. Bunch. – Toronto: Inclusion Press, 2015. – 178

р.

6. ЮНЕСКО. Политические рекомендации по инклюзивному образованию / UNESCO. – Париж: ЮНЕСКО, 2009. – 62 с.

7. Раренко, М.Б. Аудиодескрипция и тифлокомментирование как формы реализации аудиовизуального перевода / М. Б. Раренко // Известия РГПУ им. Герцена. – 2017. – № 184. – С. 56–62.

8. Шмидт, П.В. Приемы компрессии в аудиодескрипции / П.В. Шмидт. – Томск: Томский государственный университет, 2016. – 148 с.

ARTIFICIAL INTELLIGENCE FOR AUDIO DESCRIPTION IN CULTURAL CONTENT ACCESSIBILITY FOR THE VISUALLY IMPAIRED

Protchenko E.V., Kolodich A.D.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus,

The article addresses the problem of limited access for blind and visually impaired people to audiovisual content as a tool for cultural and professional education. The necessity of using artificial intelligence (AI) technologies for the mass creation of audio description (audio commentary) is substantiated. A project aimed at developing AI algorithms for generating descriptions of the visual component in films, performances, and screen adaptations is described, which will create a collection of adapted content to enhance the cultural intelligence of this social group. The project includes both technical development and direct work with the target audience to identify needs.

Key words: artificial intelligence; inclusive education; cultural intelligence; audio description; computer vision algorithms; neural networks; people with visual impairments.