

УДК 004.89, 376.33

## **ОБЗОР И АНАЛИЗ СЕРВИСОВ И ИНСТРУМЕНТОВ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ ДЛЯ ОРГАНИЗАЦИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ЛИЦ С НАРУШЕНИЯМИ СЛУХА**

Савенко А.Г.

*Институт информационных технологий БГУИР, г. Минск, Республика Беларусь,  
[savenko@bsuir.by](mailto:savenko@bsuir.by)*

В статье рассмотрены современные сервисы и инструменты с искусственным интеллектом, которые могут быть полезны при реализации инклюзивного образовательного процесса с обучаемыми с нарушениями слуха. Определены основные проблемы, которые могут возникнуть при реализации такого процесса и проанализированы существующие высокотехнологичные инструменты их решения: как программные, так и программно-аппаратные.

Ключевые слова: искусственный интеллект, средства обучения, инклюзивный образовательный процесс, нарушения слуха, глухота, обучение глухонемых и слабослышащих.

Современные тенденции и возможности образовательного процесса направлены на расширение инклюзивности обучения. Однако, обучение различных групп лиц с особыми потребностями накладывают на образовательный процесс свои уникальные ограничения: различные группы лиц с различными особыми потребностями требуют своих уникальных решений, в том числе и технических.

Одной из универсальных технологий, применимой в образовательном процессе как для нормотипичных обучаемых, так и для лиц с особыми образовательными потребностями, стал искусственный интеллект. Это мощный инструмент, который в совокупности с другими технологиями и методиками (педагогическими, управленческими и иными) может дать хороший результат в области инклюзивного обучения [1-4].

Одной из непростых категорий обучаемых с точки зрения реализации инклюзивного образовательного процесса являются лица с нарушениями слуха. Менее сложно реализуемым будет являться электронный формат обучения, однако очное (аудиторное) обучение совместно с нормотипичными обучаемыми требует использования специальных инструментов и может быть очень затратным.

Рассмотрим и проанализируем существующие сервисы и инструменты с искусственным интеллектом, которые могут стать простым и не затратным решением данной задачи.

При инклюзивном обучении лиц с нарушением слуха могут возникнуть следующие проблемы, имеющие разный характер:

### **1. Проблема 1: восприятие цифрового видео- и аудио- образовательного контента.**

В современном мире роль библиотек, как источника знаний была замещена (со снижением качества знаний, но повышением их доступности) сервисами и ресурсами глобальной сети Интернет. Всё чаще обучаемые интересуются у преподавателя не какую книгу стоит прочесть по теме занятия, а какой видеоролик посмотреть в сервисе видеохостинга YouTube. Стоит отметить, что в настоящее время там действительно увеличилось количество образовательного контента. Многие преподаватели сами создают и размещают на YouTube свой образовательный контент.

Однако, такой формат контента ещё недавно был недоступен для лиц с нарушениями слуха. В настоящее же время, с применением искусственного интеллекта, YouTube способен генерировать в режиме реального времени автоматические субтитры к любому из своих загруженных роликов. Стоит отметить, что качество и точность такой автоматической генерации субтитров зависит от используемой лексики, чёткости дикции и скорости речи говорящего. Пример использования автоматических субтитров в видеохостинге YouTube приведён на рисунке 1.

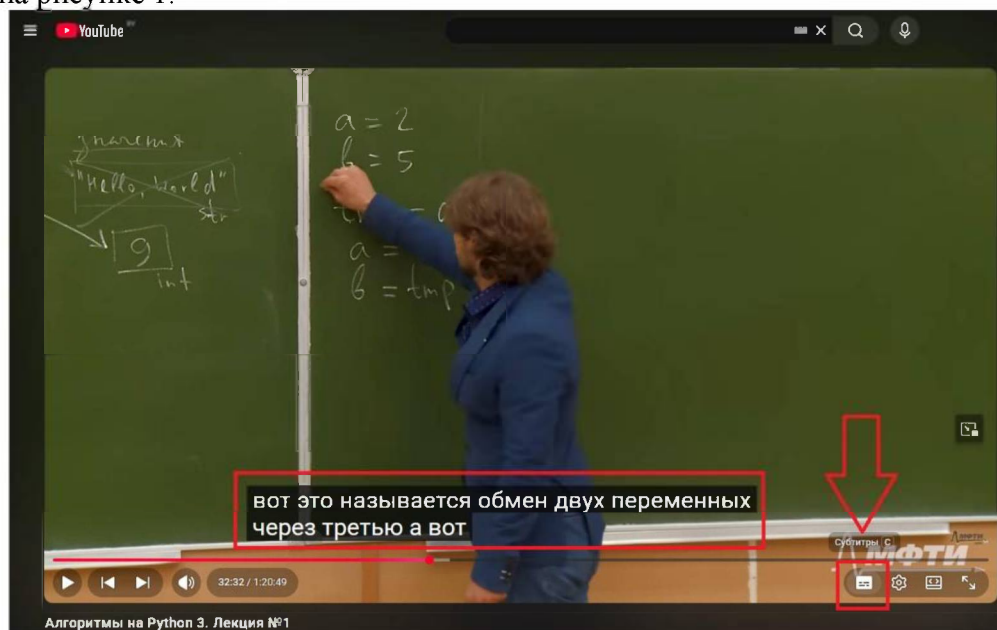


Рисунок 1 – Пример использования автоматических субтитров на сервисе видеохостинга YouTube [5]

Таким образом, само наличие функции автоматического субтитрования несомненно является большим достоинством, однако в качестве недостатка можно отметить зависимость качества субтитров от особенностей речи спикера;

**2. Проблема 2: восприятие речи преподавателя (собеседника) лицом с нарушениями слуха.** При реализации инклюзивного образовательного процесса преподаватель должен проводить занятие одинаково информативно как для нормотипичных обучаемых, так и для лиц с особыми образовательными потребностями, что на первый взгляд может показаться нерешаемой задачей.

В настоящее время существует множество как встроенных в различные другие инструменты, так и сторонних сервисов с искусственным интеллектом (технология Speech-to-Text).

Технология Speech-to-Text (STT) или, иными словами, транскрипция – это процесс преобразования устной речи в текст. Основой этой технологии являются алгоритмы машинного обучения и нейронные сети, которые анализируют оцифрованные звуковые волны и преобразуют их в текст.

Транскрипционные сервисы позволяют автоматически и часто очень быстро преобразовывать аудиозаписи или живые выступления в текстовые документы, что значительно упрощает обработку интервью, выступлений и подкастов.

Так, например, при реализации дистанционного обучения можно использовать популярную платформу видеоконференцсвязи ZOOM. Она имеет встроенный сервис транскрибирования речи говорящего в режиме реального времени. Пример использования сервиса транскрибирования речи на платформе ZOOM представлен на рисунке 2.

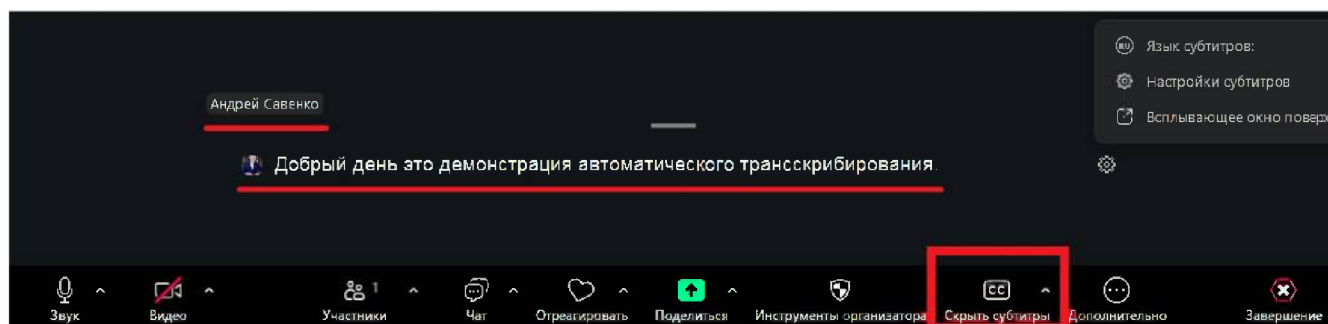


Рисунок 2 – Пример использования транскрибирования речи на платформе ZOOM

Встроенный сервис платформы ZOOM поддерживает десятки языков (в том числе и русский) и имеет настройки их отображения.

Помимо платформы ZOOM, аналогичные встроенные сервисы есть и в других популярных платформах видеоконференцсвязи, таких как Google Meet и Microsoft Teams.

Помимо это существует ряд сторонних сервисов для транскрибирования речи. Среди них можно выделить следующие основные:

- Sonix. Этот сервис способен обрабатывать аудио на множестве языков, автоматически идентифицируя говорящих (диаризация). Уникальная функция «встроенный редактор Sonix» позволяет настраивать интерфейс, редактировать и форматировать распознанный текст. Sonix также предлагает интеграцию с ZOOM для автоматической записи и транскрибирования встреч и делает это лучше встроенных инструментов ZOOM;

- Rev. Этот сервис уже давно на рынке услуг, хорошо работает с большим объемом аудиофайлов. Также Rev предлагает, как ИИ-транскрипцию, так и услуги человека-транскрибера, обеспечивая гибкость в зависимости от требований к точности. Отличительная черта Rev – это возможность создания глоссария для специфических терминов, что повышает точность при работе с узкоспециализированным контентом;

- Riverside. Уникальная особенность этого сервиса – это возможность редактирования текста транскрипции и автоматическая синхронизация с аудио/видео файлом. Это может быть особенно полезным для создания подкастов, где требуется точная синхронизация текста и медиа. Таким образом, можно работать с подкастом как с текстовым документом: удаляя,

добавляя или перемещая слова в транскрипте они автоматически изменяются в аудиофайле;

– Whisper. Это сервис автоматического распознавания речи, разработанный компанией OpenAI с открытым исходным кодом, что означает свободный доступ к её коду и моделям для использования, модификации и исследований. В основе Whisper лежит архитектура Transformer (кодировщик-декодировщик), аналогичная той, что используется в других известных моделях OpenAI, таких как GPT. Это обеспечивает высокую эффективность обработки последовательностей данных. Входное аудиосообщение разбивается на 30-секундные фрагменты, которые преобразуются в мел-спектрограммы – визуальное представление звука. Эти данные проходят через кодировщик, а декодировщик предсказывает соответствующий текстовый вариант. Модель обучена выполнять несколько задач одновременно, включая распознавание речи, перевод на английский, идентификацию языка и расстановку временных меток. Эти задачи объединены в единый процесс с помощью специальных токенов. Whisper обеспечивает более детальную и точную расшифровку технического контента по сравнению со встроенными решениями YouTube, поддерживает десятки языков и различные аудио форматы.

– Gladia. Достоинствами этого сервиса является бесплатный тарифный план (10 часов обработки речи в месяц) и качество распознавания речи. Сервис автоматически определяет язык и распределяет роли говорящих, что критично для групповых и многоязычных занятий. Gladia использует усовершенствованную версию модели Whisper, названную Whisper-Zero, которая практически полностью устраняет проблему «галлюцинаций» при транскрипции (ошибки распознавания речи), обеспечивая повышенную точность на 10-15% по сравнению с оригинальным Whisper. Также Gladia работает с видеофайлами (не больше 500 Мб) и ссылками на видеохостинг YouTube. Для использования этого сервиса можно воспользоваться google-аккаунтом.

Пример использования сервиса транскрибирования речи Gladia представлен на рисунке 3.

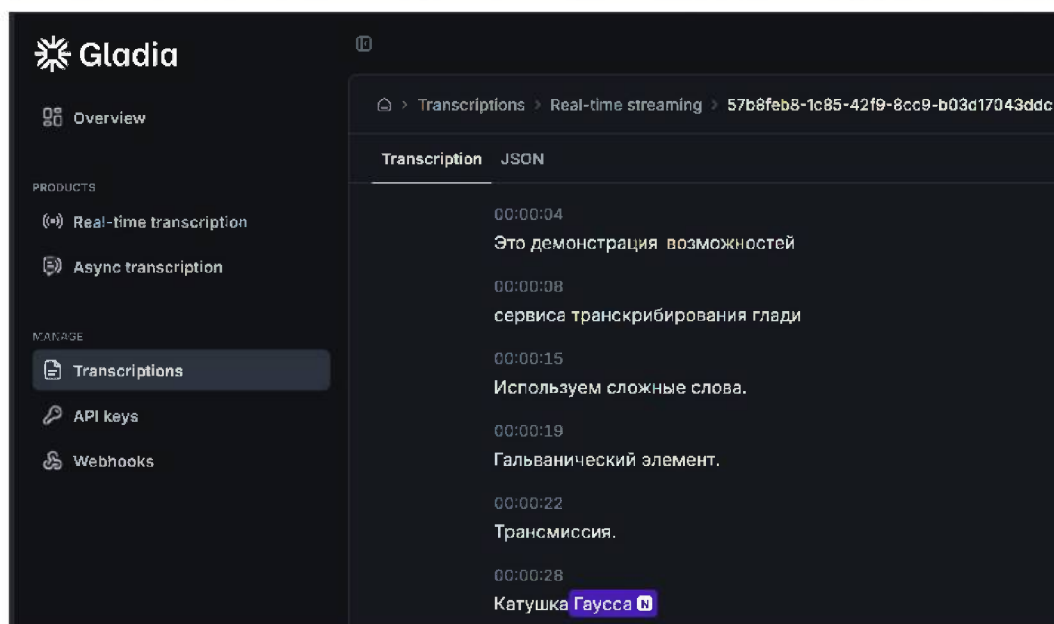


Рисунок 3 – Пример использования транскрибирования речи с помощью сервиса Gladia

Наглядное сравнение рассмотренных сервисов транскрибирования представлено в таблице 1.

Таблица 1 – Сравнение различных сервисов транскрибирования

Сервис	Объём бесплатного тарифного плана	Стоимость платного тарифного плана	Поддерживаемые языки
Sonix	30 минут речи/аудио	10\$ / час или 22\$ / месяц	50+ языков (в том числе русский)
Rev	Нет	1,5\$ / минута	35+ языков (в том числе русский)
Riverside	Нет	15\$ / месяц	100+ языков (в том числе русский)
Whisper	Open source	Open source	~100 языков (в том числе русский)
Gladia	10 часов	0,612\$ / час	~100 языков (в том числе русский)



Помимо программных средств транскрибирования существуют и программно-аппаратные средства. Наиболее продвинутым среди них является разработка XRAI Glass.

Аппаратная платформа этого устройства – умные очки Ray-Ban Meta (камера, микрофон, дисплей). Программная составляющая – это технология ASR (распознавание речи искусственным интеллектом в режиме реального времени). Программное обеспечение преобразует звук в субтитры, которые выводятся на дисплей (линзы) очков. Работа приложения интуитивна и ненавязчива, что помогает стирать физические различия между участниками диалога. Распознавание голоса может определить личность говорящего, а перевод в режиме реального времени выдаёт разные варианты транскрипции на девяти языках, включая английский, китайский, французский, немецкий, итальянский, японский, корейский, португальский и испанский. В настоящий момент ведётся работа над другими языковыми пакетами.

XRAI Glass продаётся по многоуровневой подписке, в которой пользователям доступны планы различные тарифные планы:

- Essentials – бесплатный и предлагает базовые функции, включая перевод с девяти языков и однодневную историю разговоров;
- Premium стоит порядка 25\$ в месяц и включает перевод с девяти языков, 30-дневную историю разговоров, режим дополненной реальности с поддержкой 3D и стабилизацией движения субтитров, 10-часовое облачное транскрибирование с идентификатором говорящего и настройку пользовательского интерфейса;
- Ultimate стоит порядка 65\$ в месяц и добавляет к Premium неограниченное хранение истории разговоров, неограниченное облачное транскрибирование с идентификатором говорящего, а также личного ИИ-помощника с несколькими вариантами вывода ответов.

Иллюстрация работы XRAI Glass представлена на рисунке 4.

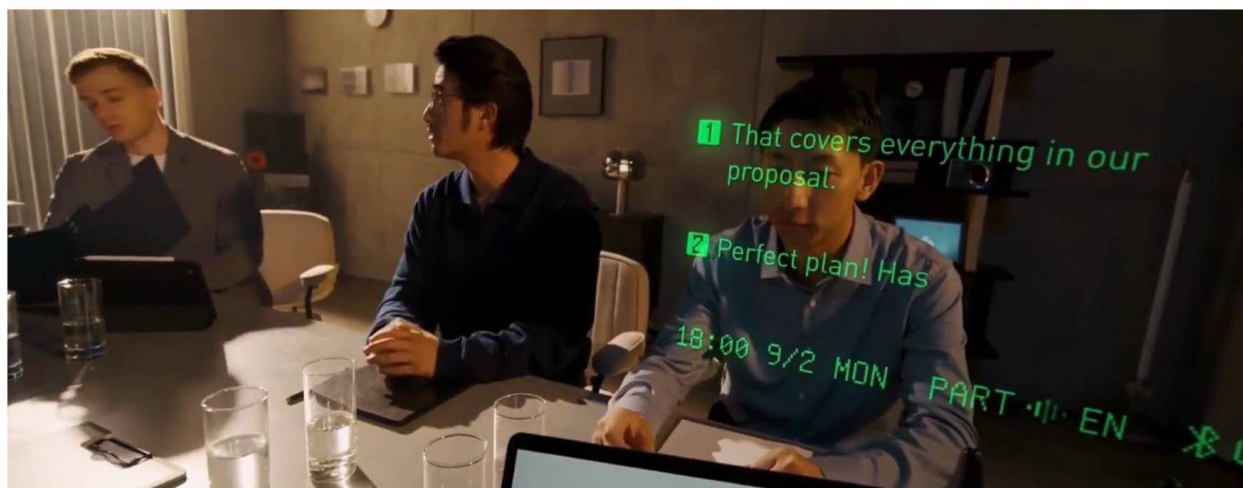


Рисунок 4 – Пример использования транскрибирования речи с помощью очков XRAI Glass

**3. Проблема 3: понимания речи или текста лицами с нарушениями слуха, которые не умеют читать.** Некоторые глухие от рождения люди не понимают также и письменной речи на привычном нормотипичным людям языке. Такие люди с детства учатся общаться на языке жестов.

Существует ряд инструментов, для перевода устной речи в язык жестов. Среди них можно отметить следующие:

- мобильное приложение и веб-сервис «Сурдопомощь»;
- сервис «Адаптис»;
- сервис «Мой РЖЯ» (русский жестовый язык).

Все эти сервисы анализируют текстовое или аудиосообщение преобразует его в жесты, используя трёхмерную анимацию и алгоритмы машинного обучения. Виртуальный персонаж-переводчик по показываает движения рук, мимику и позы, соответствующие русскому жестовому языку. Также многие сервисы имеют словарь популярных жестов и фраз

с пояснениями и примерами.

Ключевым недостатком данных сервисов на данный момент является то, что удаётся выполнить корректный перевод только отдельных слов, или отдельных фраз с сохранением грамматической логики жестового языка. Передача интонации, эмоции и контекста общения – пока является перспективой развития данных сервисов.

#### **4. Проблема понимания лица с нарушением слуха, разговаривающего на языке жестов.**

Аналогичным образом, человек не знающий жестового языка не сможет понять собеседника, разговаривающего на нём. Решение этой проблемы пока находится на ранних этапах. Так, например, существует разработка студентов Вашингтонского университета – умные перчатки, которые могут переводить американский язык жестов (ASL) в речь или текст. За свое изобретение два студента, Навид Азоди и Томас Прайор, выиграли студенческую премию Lemelson-MIT.

Разработанные перчатки SignAloud, включают различные датчики, отслеживающие малейшие движения пальцев, запястья и рук, имеющие беспроводной интерфейс Bluetooth для передачи собранных данных на компьютер. Эту информацию обрабатывает специально разработанное программное обеспечение. Перчатки способны улавливать малейшие движения рук пользователя, считывать в движениях знаки языка жестов и после расшифровки полученных данных, преобразовывать их в устную речь посредством специального синтезатора.

Устройство может быть эффективно использовано и в ситуации, когда нет возможности услышать полученный перевод – знаки языка жестов конвертируются в письменный текст на английском языке. По Bluetooth-каналу перчатки пересылают информацию о жестах к компьютеру, где с помощью статистической обработки данных определяется, какое именно слово было «произнесено» в жесте. Анализ данных производится на основе обработки последовательной статистической регрессии.

Прототип разработанного устройства представлен на рисунке 5.



Рисунок 5 – Прототип перчаток SignAloud для перевода с языка жестов

Таким образом, рассмотрены современные сервисы и инструменты с искусственным интеллектом, которые могут быть полезны при реализации инклюзивного образовательного процесса с обучаемыми с нарушениями слуха. Определены основные проблемы, которые могут возникнуть при реализации такого процесса и проанализированы существующие инструменты их решения.

#### **Литература**

1. Суский, А. А. Преимущества и перспективы внедрения нейронных сетей в образовательный процесс, как инструмент повышения качества подготовки специалистов / А. А. Суский, А. Г. Савенко // Высшее техническое образование: проблемы и пути развития = Engineering education: challenges and developments: материалы IX Международной научно-методической конференции, Минск, 1–2 ноября 2018 года / Белорусский государственный университет информатики и радиоэлектроники; редкол.: В.

А. Богуш [и др.]. – Минск, 2018. – С. 454–456.

2. Савенко, А. Г. Искусственный интеллект, как инструмент повышения качества образовательного процесса лиц с особыми потребностями / А. Г. Савенко // Информационные системы и технологии: материалы 57-й научной конференции аспирантов, магистрантов и студентов, Минск, 19–23 апреля 2021 г. / Белорусский университет информатики и радиоэлектроники, Институт информационных технологий; редкол.: А. А. Охрименко. – Минск, 2021. – С. 41–43.

3. Савенко, А. Г. Адаптивная система управления обучением с искусственным интеллектом для инклюзивного образовательного процесса / Савенко А. Г. // Новые горизонты-2021: сборник материалов VIII Белорусско-Китайского молодежного инновационного форума, Минск, 11–12 ноября 2021 года / Белорусский национальный технический университет. – Минск, 2021. – Т. 1. – С. 185–187.

4. Савенко, А. Г. Элементы искусственного интеллекта в адаптивных системах управления обучением для реализации инклюзивного образовательного процесса / А. Г. Савенко // Непрерывное профессиональное образование лиц с особыми потребностями: сборник статей IV Международной научно-практической конференции, Минск, 9–10 декабря 2021 / Белорусский государственный университет информатики и радиоэлектроники; редкол.: А. А. Охрименко [и др.]. – Минск, 2021. – С. 237–241.

5. Видеохостинг YouTube [Электронный ресурс]. Режим доступа: <https://www.youtube.com/watch?v=KdZ4HF1SrFs&list=PLRDzFCPr95fK7tr47883DFUbm4GeOjjc0>. Дата доступа: 21.11.2025.

## **REVIEW AND ANALYSIS OF SERVICES AND TOOLS WITH ARTIFICIAL INTELLIGENCE FOR ORGANIZING THE EDUCATIONAL PROCESS OF INDIVIDUALS WITH HEARING IMPAIRMENTS**

Savenko A.G.

*Institute of Information Technologies BSUIR, Minsk, Republic of Belarus*

This article examines modern artificial intelligence-powered services and tools that can be useful in implementing inclusive education for students with hearing impairments. It identifies the key challenges that may arise during the implementation of such a process and analyzes existing high-tech solutions, both software and hardware.

Key words: artificial intelligence, teaching aids, inclusive educational process, hearing impairments, deafness, education of the deaf and hard of hearing.