

УДК 004.934

<http://doi.org/10.35854/1998-1627-2026-4-496-505>

Информационные системы аналитической поддержки управления на основе многоязычной текстовой и речевой аналитики

Аркадий Анатольевич Дроздов^{1✉}, Азат Булатович Галиев²,
Александр Николаевич Дмитриев³

^{1, 2, 3} *Институт проблем передачи информации имени А. А. Харкевича Российской академии наук, Москва, Россия*

¹ adrozdv.aa@iitp.ru ✉, <https://orcid.org/0000-0002-8115-2211>

² galiev-ab@iitp.ru, <https://orcid.org/0000-0002-4026-5914>

³ dmitriev-an@iitp.ru, <https://orcid.org/0009-0006-6353-4773>

Аннотация

Цель. Разработать и обосновать воспроизводимую рамку построения информационных систем аналитической поддержки управления, интегрирующих многоязычную текстовую аналитику и технологии обработки речевых сигналов, с акцентом на их экономическую и управленческую ценность.

Задачи. Определить роль текстовых и речевых коммуникаций как источника ранних управленческих сигналов; рассмотреть архитектуру объединения распознавания речи и текстовой аналитики в едином контуре принятия решений; охарактеризовать систему метрик для оценки технического качества моделей и качества управленческих решений; показать эффект мультимодальной интеграции, а также сформулировать требования к устойчивости, интерпретируемости и мониторингу дрейфа.

Методология. Исследование сочетает в себе анализ открытых данных и опубликованных ориентиров качества для речевых и текстовых моделей с воспроизводимым синтетическим мультимодальным экспериментом. В эксперименте использованы прокси-признаки (эмбединги речи и текста), моделирующие управленческий сценарий риск-классификации. Авторы сравнивают различные стратегии интеграции модальностей (раздельную и объединенную обработку) по метрикам AUC и F1. Оценка проведена с учетом сегментации по языкам и каналам коммуникации.

Результаты. Согласно авторской позиции, интеграция речевой и текстовой модальностей способна повышать качество управленческих решений, по сравнению с использованием одной модальности, если каждая из них несет комплементарную информацию. Выявлены реалистичные ориентиры технического качества распознавания речи, подтверждена практическая применимость открытых ресурсов для построения воспроизводимых прототипов управленческих систем.

Выводы. Экономическая эффективность интеграции многоязычного NLP и ASR определена не только точностью распознавания, но и вкладом системы в снижение управленческих рисков, ускорение реакции и повышение прозрачности процессов. Для устойчивого внедрения необходимы стратифицированная оценка качества, мониторинг изменений данных во времени и встроенные механизмы интерпретации результатов.

Ключевые слова: *предсказательное моделирование, социально-экономические процессы, интеллектуальный анализ данных, машинное обучение, временные ряды, интерпретируемость*

Для цитирования: Дроздов А. А., Галиев А. Б., Дмитриев А. Н. Информационные системы аналитической поддержки управления на основе многоязычной текстовой и речевой аналитики // *Экономика и управление*. 2026. Т. 32. № 4. С. 496–505. <http://doi.org/10.35854/1998-1627-2026-4-496-505>

© Дроздов А. А., Галиев А. Б., Дмитриев А. Н., 2026

Данная статья является статьей открытого доступа, распространяемой в соответствии с условиями лицензии Creative Commons Attribution 4.0 License (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>).

Information systems for analytical decision support based on multilingual textual and speech analytics

Arkadiy A. Drozdov^{1✉}, Azat B. Galiev², Alexander N. Dmitriev³

^{1, 2, 3} *Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia*

¹ drozdov.aa@iitp.ru ✉, <https://orcid.org/0000-0002-8115-2211>

² galiev-ab@iitp.ru, <https://orcid.org/0000-0002-4026-5914>

³ dmitriev-an@iitp.ru, <https://orcid.org/0009-0006-6353-4773>

Abstract

Aim. To develop and validate a reproducible framework for building information systems for analytical decision support that integrate multilingual textual analytics and speech signal processing technologies, with an emphasis on their economic and managerial value.

Objectives. To determine the role of textual and speech communications as a source of early managerial signals; to examine the architecture for combining speech recognition and textual analytics within a unified decision-making loop; to characterize a system of metrics for assessing the technical quality of models and the quality of managerial decisions; to demonstrate the effect of multimodal integration; and to formulate requirements for robustness, interpretability, and drift monitoring.

Methods. The study combines the analysis of open data and published quality benchmarks for speech and text models with a reproducible synthetic multimodal experiment. The experiment uses proxy features (speech and text embeddings) simulating a managerial risk classification scenario. The authors compare different strategies for modality integration (separate vs. joint processing) using AUC and F1 metrics. The evaluation is conducted with consideration of segmentation by language and communication channel.

Results. According to the authors' position, the integration of speech and text modalities can improve the quality of managerial decisions compared to using a single modality, provided that each modality carries complementary information. Realistic benchmarks for the technical quality of speech recognition are identified, and the practical applicability of open resources for building reproducible prototypes of managerial systems is confirmed.

Conclusion. The economic efficiency of integrating multilingual NLP and ASR is determined not only by recognition accuracy but also by the system's contribution to reducing managerial risks, accelerating response times, and increasing process transparency. Sustainable implementation requires stratified quality assessment, monitoring of data drift over time, and built-in mechanisms for interpreting results.

Keywords: *predictive modeling, socioeconomic processes, data mining, machine learning, time series, interpretability*

For citation: Drozdov A.A., Galiev A.B., Dmitriev A.N. Information systems for analytical decision support based on multilingual textual and speech analytics. *Ekonomika i upravlenie = Economics and Management*. 2026;32(4):496-505. (In Russ.). <http://doi.org/10.35854/1998-1627-2026-4-496-505>

Введение

Информационные системы аналитической поддержки управления традиционно развивались вокруг структурированных источников данных, представленных в виде бухгалтерских и управленческих показателей, транзакций, KPI и отчетности. Однако цифровизация коммуникаций сделала тексты и речь самостоятельным классом управленческих данных, способных объяснять причины изменений в метриках и предупреждать о рисках раньше, чем они проявятся в агрегированных показателях.

Письма, чаты, регламентные документы, записи звонков и совещаний отражают не только факт проблемы, но и контекст: что именно произошло, на каком участке процесса возник сбой, как реагирует клиент или сотрудник, какие формулировки указывают на нарушение регламента или нарастающий конфликт. В экономике управления именно этот слой причинности превращает «данные» в «информацию» для принятия решения.

В прикладной управленческой аналитике ценность коммуникационных данных проявляется через ускорение цикла обратной

связи: сигнал, зафиксированный в обращениях или звонках, может быть выявлен и интерпретирован значительно раньше, чем его эффект станет видимым в финансовой отчетности. Это влияет на эффективность распределения ресурсов, контроль качества, управление рисками и продуктовые решения, особенно в клиентских сервисах и комплаенс-процессах. В крупных интегрированных управленческих контурах задача консолидации разнородных источников, аналитической обработки и поддержки нерегламентированных запросов закреплена в официальных описаниях информационно-аналитических подсистем¹. Тем самым коммуникационные данные становятся легитимной частью «управленческого наблюдения» наряду с традиционными формализованными показателями.

Сегодня управленческую ценность продукта определяют не только качеством распознавания речи, но и тем, какие функции поверх него доступны бизнесу (поиск сценариев, контроль соблюдения скриптов, выявление причин обращений, интеграции с CRM/BI и управленческая отчетность). В исследовании по ценообразованию на рынке речевой аналитики показано, что при формировании цены фирмы ориентируются в большей степени на наличие дополнительных опций, чем на точность распознавания как на единственный критерий качества [1]. Для экономиста-управленца это означает, что «качество модели» следует измерять инженерными метриками, а также вкладом в управленческие решения и сокращение потерь.

Отдельный вызов современных информационных систем аналитической поддержки управления связан с многоязычностью. Трансграничные цепочки поставок, международные команды, миграционные сегменты клиентов и омниканальные коммуникации делают многоязычную поддержку в большей мере нормой, чем экзотикой. Система, устойчиво работающая на русском и английском языках, а также сохраняющая качество на смешанных регистрах и акцентах, получает преимущества в наблюдаемости процессов и управляемости рисков.

Поэтому в настоящей статье многоязычность рассмотрена не как вспомогательная функция, а как базовое требование к архитектуре и оценке качества.

Цель статьи состоит в том, чтобы рассмотреть воспроизводимую рамку построения информационных систем аналитической поддержки управления, объединяющую распознавание речи и многоязычную текстовую аналитику в едином контуре принятия решений. Для достижения этой цели мы попытались раскрыть три исследовательских вопроса о том, как устроить управленческий пайплайн «речь — текст — аналитика» и когда целесообразна мультимодальная интеграция; как корректно сравнивать решения по метрикам распознавания и по управленческим метрикам, учитывая дрейф и сегменты; каким образом экономическая ценность решения зависит от функциональности и интерпретируемости, а также от прироста точности распознавания.

Материалы и методы

Исследование, основанное в том числе на данных, выявленных рядом отечественных и зарубежных авторов [2; 3; 4; 5; 6; 7; 8; 9], организовано в логике воспроизводимости, которая важна для академической проверки и практической переносимости результатов. В работе использована комбинация двух эмпирических форматов. Первый формат опирается на открытые ресурсы и опубликованные ориентиры качества, что позволяет зафиксировать реалистичные уровни метрик на доступных корпусах без необходимости повторять вычислительно тяжелое обучение «с нуля». Второй формат представлен воспроизводимым синтетическим экспериментом, позволяющим контролируемо продемонстрировать управленческий эффект мультимодальной интеграции, при которой каждая модальность несет комPLEMENTARную информацию.

В контексте данных для репликационного формата использованы общедоступные речевые и текстовые ресурсы. Многоязычный корпус Common Voice описан как масшОво-многоязычная база речи, собранная

¹ Отчет о выполнении первого этапа научно-исследовательской работы для официального использования Евразийской экономической комиссии по теме «Рекомендации по взаимодействию государств — членов ЕАЭС в области регулирования, развития и обеспечения безопасности критической цифровой инфраструктуры с перспективой до 2025 года» (промежуточный) // Евразийская экономическая комиссия. 2018. URL: <https://eec.eaeunion.org/upload/iblock/05e/Otchet-o-NIR-EEK-OBI-TSI.-Etap-1.pdf?ysclid=mn90fs74n832429155> (дата обращения: 18.11.2025).

и валидированная через краудсорсинговый механизм [10], а информация об актуальных релизах и доступности публикуется на официальных страницах проекта¹. Для русскоязычной речи рассмотрен открытый датасет Golos, представленный на платформе OpenSLR как крупный корпус с фиксируемым масштабом и инфраструктурой загрузки², его рецензируемое описание, при этом приведены ориентиры качества распознавания в терминах WER в разных условиях [11]. Для текстовой компоненты и нормализации русского языка применены НКРЯ 2.0 как современная корпусная платформа с расширенными возможностями анализа [12] и обновленными корпусами³, а также RusVectōrēs как инфраструктура русскоязычных семантических моделей и эмбедингов, релевантных задачам семантического поиска и прикладной аналитики⁴.

Методологически архитектура информационных систем аналитической поддержки управления представлена как последовательность преобразований и решений. Речевой канал начинается с обработки аудио и распознавания речи (ASR), затем транскрипты поступают в слой многоязычной текстовой аналитики (NLP). Такой каскад служит наиболее распространенным вариантом внедрения, поскольку позволяет использовать зрелые текстовые модели и обеспечивает прозрачность контроля качества. Одновременно признан тот факт, что каскадная схема неизбежно теряет часть информации, присутствующей в аудио как сигнале, включая просодические признаки и характеристики эмоциональной окраски, а также может страдать эффектом «распространения ошибок», при котором ошибочный транскрипт ухудшает качество последующей аналитики.

Поэтому в процессе исследования дополнительно охарактеризована мультимодальная интеграция признаков. При указанном подходе текст и речь трактуются как две модальности, каждая из которых может содержать комплементарные управленческие сигналы. Текст чаще отражает содержание

и тему обращения, а речь может несоразмерно сильнее отражать напряжение, конфликтность и уровень эскалации, которые важны для риск-классификации и контроля качества. Для многоязычных управленческих систем релевантны унифицированные мультимодальные архитектуры, при этом речь и текст согласуются в общем представлении, как в концепции SpeechT5 [13], а также масштабные подходы, ориентированные на многоязычный перевод и унификацию коммуникаций, включая SeamlessM4T [14]. Переход к мультимодальности в настоящей статье рассмотрен как управленческое решение, оправданное не «технологической модой», а наличием дополнительной информации и измеримого выигрыша в метриках верхнего уровня.

Протокол оценки качества отражает разделение инженерных и управленческих метрик. Для ASR использованы WER и CER, поскольку они передают точность транскрипции и критичны для downstream-задач [11]. В целях реализации управленческих задач поверх транскриптов и исходных текстов использованы метрики, связанные с классификацией и качеством решений, прежде всего AUC и F1 для риск-классификации и задач контроля сценариев, а также ROUGE для суммаризации, при которой система применена для управленческой отчетности [15]. Принципиально важным видится то, что в реальных управленческих процессах средние значения метрик часто маскируют проблемы в отдельных сегментах. Поэтому оценка должна быть стратифицированной по языкам, каналам, подразделениям и временным периодам, так как дрейф данных служит системным свойством коммуникационных потоков.

В синтетическом мультимодальном эксперименте использованы прокси-признаки в форме эмбедингов, имитирующих выход современных моделей речи и текста. Такой дизайн нужен для того, чтобы не смешивать задачу сравнения стратегий интеграции модальностей с задачей обучения крупных моделей с нуля. В эксперименте задают

¹ Технологии, говорящие на вашем языке // Common Voice. URL: <https://commonvoice.mozilla.org/> (дата обращения: 18.11.2025).

² Golos (SLR114): Russian Speech Recognition Dataset // Open Speech and Language Resources (OpenSLR). 2021. URL: <https://www.openslr.org/114/> (дата обращения: 15.12.2025).

³ Обзор возможностей // Национальный корпус русского языка. URL: <https://ruscorpora.ru/> (дата обращения: 02.11.2025).

⁴ Кутузов А., Кузьменко Е. RusVectōrēs: семантические модели для русского языка // RusVectōrēs. URL: <https://rusvectors.org/ru/> (дата обращения: 14.11.2025).

бинарную управленческую метку «высокий риск / высокая значимость», а затем сравнивают стратегии text-only и speech-only с интеграционными схемами раннего и позднего объединения. Выбор порога классификации проводят по критерию F1 как приближению к практической логике балансировки пропусков и ложных срабатываний в риск-процессах. Главный исследовательский интерес — показать, что даже простая интеграция может давать выигрыш, если модальности несут комплементарную информацию.

Результаты

Результаты репликационного формата фиксируют ориентиры качества на открытых ресурсах и свидетельствуют о том, какие уровни метрик следует считать реалистичными для воспроизводимых прототипов. В рецензируемом описании датасета Golos приведены параметры корпуса и ориентиры точности распознавания речи в терминах WER при определенных условиях декодирования, что задает практический масштаб достижимого качества на открытых русскоязычных данных [11]. Официальная карточка на OpenSLR подтверждает доступность датасета и его масштаб, важный для построения воспроизводимых экспериментов¹. Эти ориентиры особенно значимы для управленческих информационных систем, поскольку помогают понимать, где именно возникает граница «инженерного качества», после которой главным становится качество решений верхнего уровня.

Многоязычный корпус Common Voice предоставляет принципиальную возможность строить многоязычный ASR-контур на открытых данных, хотя для управленческих доменов обычно требуется адаптация под реальную речь контакт-центров и переговоров. Для текстовой части использование корпусных и семантических ресурсов русского языка повышает качество нормализации, поиска и извлечения смыслов, что в итоге снижает шум в управленческих признаках, повышает переносимость моделей между подразделениями [12] и каналами коммуникации².

Синтетический мультимодальный эксперимент передает управленческий смысл

интеграции. При раздельном использовании текста и речи каждая модальность «видит» часть сигнала: текстовая — содержание и тему, речевая — дополнительные прокси-признаки напряжения и эскалации. Интеграция модальностей через объединение признаков или через объединение вероятностей улучшает качество риск-классификации по метрикам AUC и F1 по сравнению с использованием одной модальности. Этот результат следует трактовать не как «универсальный закон», а как иллюстрацию экономически важного механизма: дополнительная модальность приносит эффект, если несет независимую и полезную для решения информацию, а не просто дублирует имеющийся сигнал.

На рисунке 1 представлена схема, на которой показаны два взаимосвязанных блока результатов исследования. Первый блок отражает ориентиры технического качества, зафиксированные на открытых многоязычных и русскоязычных ресурсах распознавания речи, которые задают реалистичный уровень достижимых метрик для воспроизводимых прототипов ASR-контуров. Второй блок демонстрирует результаты синтетического мультимодального эксперимента по риск-классификации, показывающего прирост управленческих метрик (AUC, F1) при интеграции текстовой и речевой модальностей по сравнению с их раздельным использованием.

Схема, представленная на рисунке 1, визуализирует переход от уровня инженерного качества к уровню управленческого качества решений, отражает условие экономической эффективности мультимодальности, в частности наличие независимой и комплементарной информации в каждой модальности.

Обсуждение

Полученные результаты подтверждают, что в информационных системах аналитической поддержки управления важна двухуровневая логика оценки. Инженерные метрики, такие как WER и CER, определяют качество преобразования речи в текст и задают базовый уровень надежности входных данных. Однако конечная ценность

¹ Golos (SLR114): Russian Speech Recognition Dataset // Open Speech and Language Resources (OpenSLR). 2021. URL: <https://www.openslr.org/114/> (дата обращения: 15.12.2025).

² Обзор возможностей // Национальный корпус русского языка. URL: <https://ruscorpora.ru/> (дата обращения: 02.11.2025); Кутузов А., Кузьменко Е. Указ. соч.

Мультимодальный эксперимент по классификации рисков

Интеграция текста и речи улучшает классификацию рисков

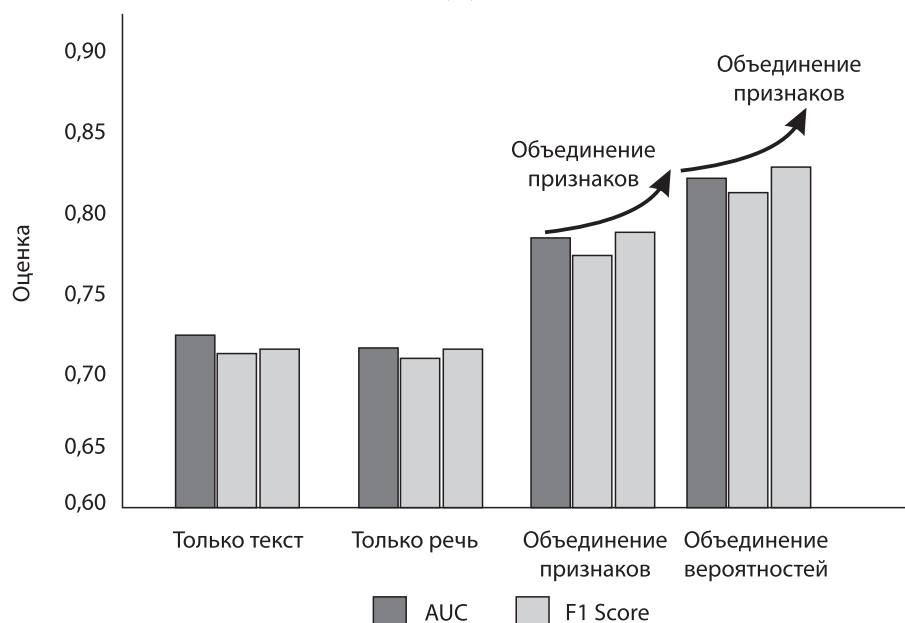


Рис. 1. Репликационные ориентиры качества и эффект мультимодальной интеграции
Fig. 1. Replication quality benchmarks and the effect of multimodal integration

Источник: составлено авторами.

для управленца обусловлена метриками верхнего уровня, связанными с управленческими решениями: насколько точно система выделяет рискованные случаи, насколько надежно выявляет нарушения сценариев, как стабильно формирует краткие резюме и отчеты. Экономическая интерпретация этих метрик должна быть выполнена через стоимость ошибок, поскольку ложные пропуски в риск-процессах могут приводить к несоразмерным потерям по сравнению с умеренным ростом нагрузки от ложных срабатываний.

С этой точки зрения особенно значим вывод о том, что рыночная ценность решений речевой аналитики может зависеть от функциональности сильнее, чем от точности распознавания. Эмпирический результат исследования по ценообразованию на рынке речевой аналитики прямо указывает на роль «опций» и дополнительных функций в формировании цены и привлекательности продукта [1]. Для экономиста-управленца это означает, что инвестиционное решение о внедрении должно оценивать не только прирост WER, но и то, какие управленческие контуры будут закрыты: комплаенс-контроль, мониторинг качества сервиса, аналитика причин обращений, интеграция с CRM и управленческой отчетностью.

Многоязычность усиливает и потенциал эффекта, и риски внедрения. Сегментные провалы качества — по языкам, акцентам и каналам — могут разрушать доверие к системе, создавать управленческие и репутационные риски. Современные обзоры по качеству многоязычных речевых датасетов свидетельствуют о том, что перекосы данных и неоднородность разметки — системная проблема, которую нельзя компенсировать одной «средней метрикой» [16]. В управленческой практике это требует постоянного мониторинга качества на срезах и регламентов обновления моделей.

Отдельным блоком следует рассматривать вопросы приватности и организационного управления рисками. Речевые данные могут содержать персональные сведения и встраиваться в процессы, затрагивающие клиентов и сотрудников, что требует строгих политик доступа, хранения, маскирования и аудита. Рискоориентированный подход к управлению ИИ-системами на жизненном цикле, включающий в себя надежность, управляемость и ответственность, систематизирован в AI RMF [17], и он полезен как универсальная рамка для построения внутренней «модели управления моделью» в организациях.

Вместе с тем в исследовании находят отражение опубликованные ориентиры каче-

ства на открытых данных без полного переобучения крупных моделей в кейсе репликации, поскольку воспроизведение таких пайплайнов требует значительных вычислительных ресурсов и полного совпадения параметров обучения [10; 11]. Синтетический эксперимент опирается на эмбединги как прокси-признаки и демонстрирует эффект интеграции модальностей, но не заменяет end-to-end сравнение архитектур на «сыром» аудио в конкретном домене. Экономическая оценка эффекта приведена в логике управленческой интерпретации, однако расчет ROI/TCO для конкретной организации требует учета инфраструктурных затрат, интеграций и комплаенс-процедур.

Выводы

Интеграция многоязычной текстовой аналитики и технологий обработки речевых сигналов в информационные системы аналитической поддержки управления экономически оправдана в организациях с интенсивными коммуникационными потоками и повышенной чувствительностью к операционным и репутационным рискам. Коммуникационные данные помогают выявлять причины отклонений и сигналы эскалации до того, как они проявятся в агрегированных показателях, тем самым повышая скорость реакции и качество управленческих решений.

Экономическая ценность таких решений определена не только точностью распознавания речи, но и функциональностью ана-

литического слоя, глубиной интеграции с управленческими контурами и интерпретируемостью результатов. Исследования рынка речевой аналитики показывают, что рыночная стоимость продукта во многом зависит от набора управленческих опций, а не исключительно от показателя WER. Следовательно, инженерные метрики служат лишь базовым уровнем оценки, а ключевой эффект формируется на уровне управленческих метрик и бизнес-результатов.

Открытые ресурсы, такие как Common Voice и Golos, обеспечивают основу для построения воспроизводимых многоязычных ASR-контуров и сопоставления качества по проверяемым ориентирам [18]¹. Корпусные и семантические инструменты русского языка поддерживают нормализацию и семантический анализ текстов, что повышает устойчивость аналитических решений. Мультимодальная интеграция речи и текста усиливает управленческий эффект в случаях, если модальности несут комплементарные сигналы, что подтверждается результатами эксперимента по метрикам AUC и F1.

Для устойчивого внедрения необходимы стратифицированная оценка качества, мониторинг дрейфа и встроенные механизмы управления рисками на жизненном цикле ИИ-систем [17]. В этих условиях интеграция многоязычного NLP и ASR становится не просто технологическим расширением, а стратегическим элементом современной управленческой инфраструктуры.

Список источников

1. Салютин Т. Ю., Гумеров М. Ф., Каберова А. Р., Платунина Г. П. Принятие решений в управлении ценовой политикой на рынке речевой аналитики // BENEFICIUM. 2023. № 2. С. 20–27. [https://doi.org/10.34680/BENEFICIUM.2023.2\(47\).20-27](https://doi.org/10.34680/BENEFICIUM.2023.2(47).20-27)
2. Graves A., Fernández S., Gomez F., Schmidhuber J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks // Proc. of the 23rd Int. conf. on machine learning (ICML 2006) / ed. by W. Cohen, A. Moore. New York: ACM, 2006. P. 369–376. <https://doi.org/10.1145/1143844.1143891>
3. Baevski A., Zhou Y., Mohamed A., Auli M. Wav2Vec 2.0: A framework for self-supervised learning of speech representations // Proc. 34th Int. conf. on neural information processing systems (NIPS'20). New York, NY: ACM, 2020. P. 12449–12460. (Advances in neural information processing systems (NeurIPS 2020). Vol. 33). <https://doi.org/10.48550/arXiv.2006.11477>
4. Conneau A., Khandelwal K., Goyal N. et al. Unsupervised cross-lingual representation learning at scale // Proc. of the 58th Annu. meet. Association for Computational Linguistics (ACL 2020). Kerrville, TX: Association for Computational Linguistics, 2020. P. 8440–8451. <https://doi.org/10.18653/v1/2020.acl-main.747>

¹ Golos (SLR114): Russian Speech Recognition Dataset // Open Speech and Language Resources (OpenSLR). 2021. URL: <https://www.openslr.org/114/> (дата обращения: 15.12.2025); Технологии, говорящие на вашем языке // Common Voice. URL: <https://commonvoice.mozilla.org/> (дата обращения: 18.11.2025).

5. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019). Kerrville, TX: Association for Computational Linguistics, 2019. Vol. 1. P. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
6. Xue L., Constant N., Roberts A. et al. mT5: A massively multilingual pre-trained text-to-text transformer // Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021). Minneapolis, MN Association for Computational Linguistics (ACL 2020). Kerrville, TX: Association for Computational Linguistics, 2021. P. 483–498. <https://doi.org/10.18653/v1/2021.naacl-main.41>
7. Kudo T., Richardson J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing // Proc. of the 2018 Conf. on empirical methods in natural language processing (EMNLP). Kerrville, TX: Association for Computational Linguistics, 2018. P. 66–71. <https://doi.org/10.18653/v1/D18-2012>
8. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages // Analysis of images, social networks and texts (AIST 2015) /eds. M. Khachay, N. Konstantinova, A. Panchenko, D. Ignatov, V. Labunets. Cham: Springer, 2015. P. 320–332. (Communications in computer and information science. Vol. 542). https://doi.org/10.1007/978-3-319-26123-2_31
9. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: A method for automatic evaluation of machine translation // Proc. of the 40th Annu. meet. Association for Computational Linguistics (ACL 2002). Kerrville, TX: Association for Computational Linguistics, 2002. P. 311–318. <https://doi.org/10.3115/1073083.1073135>
10. Ardila R., Branson M., Davis K. et al. Common voice: A massively-multilingual speech corpus // Proc. of the 12th Language resources and evaluation conf. (LREC 2020). Paris: European Language Resources Association (ELRA), 2020. P. 4218–4222. <https://doi.org/10.48550/arXiv.1912.06670>
11. Karpov N., Denisenko A., Minkin F. Golos: Russian dataset for speech research // Proc. of Interspeech 2021. Brno: International Speech Communication Association (ISCA), 2021. P. 1419–1423. <https://doi.org/10.21437/Interspeech.2021-462>
12. Савчук С. О., Архангельский Т. А., Бонч-Осмоловская А. А. и др. Национальный корпус русского языка 2.0: новые возможности и перспективы развития // Вопросы языкознания. 2024. № 2. С. 7–34. <https://doi.org/10.31857/0373-658X.2024.2.7-34>
13. Ao J., Wang R., Li P. et al. SpeechT5: Unified-modal encoder-decoder pre-training for spoken language processing // Proc. of the 60th Annu. meet. Association for Computational Linguistics (ACL 2022). Kerrville, TX: Association for Computational Linguistics, 2022. P. 5723–5738. <https://doi.org/10.18653/v1/2022.acl-long.393>
14. Barrault L., Chung Y.-A., Chiu C.-C. et al. SeamlessM4T — massively multilingual and multimodal machine translation // Proc. of the 61st Annu. meet. Association for Computational Linguistics (ACL 2023). Kerrville, TX: Association for Computational Linguistics, 2023. P. 15082–15096. <https://doi.org/10.18653/v1/2023.acl-long.831>
15. Lin C.-Y. ROUGE: A package for automatic evaluation of summaries // Text summarization branches out: Proc. of the ACL-04 Workshop. Kerrville, TX: Association for Computational Linguistics, 2004. P. 74–81. URL: <https://aclanthology.org/W04-1013.pdf> (дата обращения: 24.02.2026).
16. Lau M., Chen Q., Fang Y. et al. Data quality issues in multilingual speech datasets: The need for sociolinguistic awareness and proactive language planning // Proc. of the 63rd Annu. meet. Association for Computational Linguistics (ACL). Kerrville, TX: Association for Computational Linguistics. 2025. Vol. 1. P. 7466–7492. <https://doi.org/10.18653/v1/2025.acl-long.370>
17. Tabassi E. Artificial intelligence risk management framework (AI RMF 1.0) // National Institute of Standards and Technology (NIST). January 26. 2023. <https://doi.org/10.6028/NIST.AI.100-1>
18. Radford A., Kim J. W., Xu T. et al. Robust speech recognition via large-scale weak supervision // Proc. of the 40th Int. conf. on machine learning (ICML 2023). (Honolulu, HI, July 23–29, 2023). Cambridge, MA: MLResearchPress, 2023. P. 28492–28518. URL: <https://proceedings.mlr.press/v202/radford23a.html> (дата обращения: 26.02.2026).

References

1. Salutina T.Yu., Gumerov M.F., Kaberova A.R., Platunina G.P. Decision-making in price policy management in the speech analytics market. *Beneficium*. 2023;(2):20-27. (In Russ.). [https://doi.org/10.34680/BENEFICIUM.2023.2\(47\).20-27](https://doi.org/10.34680/BENEFICIUM.2023.2(47).20-27)
2. Graves A., Fernández S., Gomez F., Schmidhuber J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: Cohen W.,

- Moore. A., eds. Proc. 23rd Int. conf. on machine learning (ICML 2006). New York, NY: ACM; 2006:369-376. <https://doi.org/10.1145/1143844.1143891>
3. Baevski A., Zhou Y., Mohamed A., Auli M. Wav2Vec 2.0: A framework for self-supervised learning of speech representations. In: Proc. 34th Int. conf. on neural information processing systems (NIPS'20). New York, NY: ACM; 2020:12449-12460. (Advances in neural information processing systems (NeurIPS 2020). Vol. 33). <https://doi.org/10.48550/arXiv.2006.11477>
4. Conneau A., Khandelwal K., Goyal N., et al. Unsupervised cross-lingual representation learning at scale. In: Proc. 58th Annu. meet. Association for Computational Linguistics (ACL 2020). Kerrville, TX: Association for Computational Linguistics; 2020:8440-8451. <https://doi.org/10.18653/v1/2020.acl-main.747>
5. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019). Kerrville, TX: Association for Computational Linguistics; 2019:1:4171-4186. <https://doi.org/10.18653/v1/N19-1423>
6. Xue L., Constant N., Roberts A., et al. mT5: A massively multilingual pre-trained text-to-text transformer. In: Proc. 2021 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021). Kerrville, TX: Association for Computational Linguistics; 2021:483-498. <https://doi.org/10.18653/v1/2021.naacl-main.41>
7. Kudo T., Richardson J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In: Proc. 2018 Conf. on empirical methods in natural language processing (EMNLP). Kerrville, TX: Association for Computational Linguistics; 2018:66-71. <https://doi.org/10.18653/v1/D18-2012>
8. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages. In: Khachay M., Konstantinova N., Panchenko A., Ignatov D., Labunets V., eds. Analysis of images, social networks and texts (AIST 2015). Cham: Springer; 2015:320-332. (Communications in computer and information science. Vol. 542). https://doi.org/10.1007/978-3-319-26123-2_31
9. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: A method for automatic evaluation of machine translation. In: Proc. 40th Annu. meet. Association for Computational Linguistics (ACL 2002). Kerrville, TX: Association for Computational Linguistics; 2002:311-318. <https://doi.org/10.3115/1073083.1073135>
10. Ardila R., Branson M., Davis K., et al. Common voice: A massively-multilingual speech corpus. In: Proc. 12th Language resources and evaluation conf. (LREC 2020). Paris: European Language Resources Association (ELRA); 2020:4218-4222. <https://doi.org/10.48550/arXiv.1912.06670>
11. Karpov N., Denisenko A., Minkin F. Golos: Russian dataset for speech research. In: Proc. Interspeech 2021. Brno: International Speech Communication Association (ISCA); 2021: 1419-1423. <https://doi.org/10.21437/Interspeech.2021-462>
12. Savchuk S.O., Arkhangelskiy T.A., Bonch-Osmolovskaya A.A., et al. Russian National Corpus 2.0: New opportunities and development prospects. *Voprosy yazykoznaniya*. 2024;(2): 7-34. (In Russ.). <https://doi.org/10.31857/0373-658X.2024.2.7-34>
13. Ao J., Wang R., Li P., et al. SpeechT5: Unified-modal encoder-decoder pre-training for spoken language processing. In: Proc. 60th Annu. meet. Association for Computational Linguistics (ACL 2022). Kerrville, TX: Association for Computational Linguistics; 2022: 5723-5738. <https://doi.org/10.18653/v1/2022.acl-long.393>
14. Barrault L., Chung Y.-A., Chiu C.-C., et al. SeamlessM4T — massively multilingual and multimodal machine translation. In: Proc. 61st Annu. meet. Association for Computational Linguistics (ACL 2023). Kerrville, TX: Association for Computational Linguistics; 2023: 15082-15096. <https://doi.org/10.18653/v1/2023.acl-long.831>
15. Lin C.-Y. ROUGE: A package for automatic evaluation of summaries. In: Text summarization branches out. Proc. ACL-04 workshop. Kerrville, TX: Association for Computational Linguistics; 2004:74-81. URL: <https://aclanthology.org/W04-1013.pdf> (accessed on 24.02.2026).
16. Lau M., Chen Q., Fang Y., et al. Data quality issues in multilingual speech datasets: The need for sociolinguistic awareness and proactive language planning. In: Proc. 63rd Annu. meet. Association for Computational Linguistics (ACL). Kerrville, TX: Association for Computational Linguistics; 2025;1:7466-7492. DOI: 10.18653/v1/2025.acl-long.370
17. Tabassi E. Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology (NIST). Jan. 26. 2023. <https://doi.org/10.6028/NIST.AI.100-1>
18. Radford A., Kim J.W., Xu T., et al. Robust speech recognition via large-scale weak supervision. In: Proc. 40th Int. conf. on machine learning (ICML 2023). (Honolulu, HI, July 23-29, 2023). Cambridge, MA: MLResearchPress; 2023:28492-28518. (Proceedings of machine learning research (PMLR). Vol. 202). URL: <https://proceedings.mlr.press/v202/radford23a.html> (accessed on 26.02.2026).

Информация об авторах**Аркадий Анатольевич Дроздов**

кандидат физико-математических наук,
старший научный сотрудник лаборатории
информационных процессов в сложных
социальных системах

Институт проблем передачи информации
имени А. А. Харкевича Российской академии
наук

127051, Москва, Большой Каретный пер., д. 19,
стр. 1

Азат Булатович Галиев

научный сотрудник лаборатории
информационных процессов в сложных
социальных системах

Институт проблем передачи информации
имени А. А. Харкевича Российской академии
наук

127051, Москва, Большой Каретный пер., д. 19,
стр. 1

Александр Николаевич Дмитриев

младший научный сотрудник лаборатории
обработки и передачи информации
в когнитивных системах

Институт проблем передачи информации
имени А. А. Харкевича Российской академии
наук

127051, Москва, Большой Каретный пер., д. 19,
стр. 1

Поступила в редакцию 25.02.2026
Прошла рецензирование 11.03.2026
Подписана в печать 27.04.2026

Information about the authors**Arkadiy A. Drozdov**

PhD in Physical and Mathematical Sciences,
senior researcher of the Laboratory
of Information Processes in Complex Social
Systems

Institute for Information Transmission Problems
of the Russian Academy of Sciences
(Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051,
Russia

Azat B. Galiev

researcher of the Laboratory of Information
Processes in Complex Social Systems

Institute for Information Transmission Problems
of the Russian Academy of Sciences
(Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051,
Russia

Alexander N. Dmitriev

junior researcher of the Laboratory
for Information Processing and Transmission
in Cognitive Systems

Institute for Information Transmission Problems
of the Russian Academy of Sciences
(Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051,
Russia

Received 25.02.2026
Revised 11.03.2026
Accepted 27.04.2026

Конфликт интересов: авторы декларируют отсутствие конфликта интересов,
связанных с публикацией данной статьи.

Conflict of interest: the authors declare no conflict of interest
related to the publication of this article.