

1. Заказчик

TOO NEOCODE, участник Astana Hub. Разрабатывает и эксплуатирует финтех-продукты и голосовых AI-агентов для финансовых организаций Казахстана.

2. Назначение

Синтез речи для голосового агента, который ведёт исходящие звонки клиентам финансовой организации на казахском и русском языках: напоминает о платеже, называет сумму и срок, отвечает на типовые вопросы. Модель заменяет текущий модуль синтеза и подключается к уже работающему агенту.

3. Текущая ситуация

Агент собран из трёх частей: распознавание речи (казахский и русский), языковая модель для разбора реплик и формулировок, синтез речи. Ход разговора и все решения задаёт код, языковая модель только формулирует готовый шаг. Распознавание и диалог работают на собственных серверах.

Проблемы текущего синтеза:

- нет мужского голоса нужного качества, хотя агент по сценарию мужского рода и называет себя мужским именем;
- ошибки в числительных: падеж и род в суммах («двадцать одна тысяча тенге», «жиырма бір мың теңге»), датах и процентах;
- запинки и смена акцента при переходе с русского на казахский внутри фразы, например «Здравствуйте! Сәлеметсіз бе!»;
- неверное ударение в казахских именах и в нашей лексике (названия компаний, «ЧСИ», «Kassa24»);
- качество заметно падает после передачи в телефонный канал 8 кГц.

4. Функциональные требования

1. **Языки.** Казахский и русский, включая смешанные фразы с переключением языка внутри предложения без паузы и смены тембра.
2. **Голоса.** Не меньше двух голосов в поставке: мужской и женский. Каждый голос говорит на обоих языках.
3. **Нормализация текста.** Суммы в тенге, даты, время, проценты, номера договоров и телефонов читаются словами в правильном падеже и роде для каждого языка. Правила нормализации отделены от модели и редактируются без переобучения.
4. **Словарь произношения.** Пользовательский словарь ударений и транскрипций для

имён, названий и аббревиатур. Изменения применяются без переобучения.

5. **Управление речью.** Скорость, паузы и ударение на отдельных словах — через SSML или эквивалентную разметку.
6. **Потоковый синтез.** Звук отдаётся фрагментами по мере генерации, агент начинает воспроизведение до окончания синтеза всей фразы.
7. **Форматы.** PCM 16 бит 8 кГц и 16 кГц, μ -law/A-law 8 кГц для телефонии, WAV 24 кГц для проверки качества.
8. **API.** HTTP (REST) и потоковый интерфейс (WebSocket или gRPC), описание в OpenAPI или proto. Совместимость с текущей интеграцией агента согласуется на старте работ.
9. **Дообучение — новый голос.** Создание нового голоса по 1–3 часам студийных записей одного диктора на двух языках. Скрипты подготовки данных (нарезка, выравнивание текста, фильтрация брака) входят в поставку.
10. **Дообучение — домен.** Адаптация произношения и интонации к лексике заказчика по 30–60 минутам записей фраз из сценариев без потери качества на общих фразах.
11. **Самостоятельность заказчика.** Дообучение запускается командой заказчика по инструкции, без участия исполнителя.

5. Нефункциональные требования

- Работа полностью в контуре заказчика, без обращений к внешним сервисам во время синтеза и обучения.
- Инференс: один сервер с GPU уровня NVIDIA L4 / RTX 4090 (24 ГБ) — не меньше 20 одновременных потоков.
- Задержка первого звукового фрагмента не больше 300 мс, коэффициент реального времени (RTF) не больше 0,1 на один поток.
- Обучение и дообучение: один GPU уровня NVIDIA A100 / H100 (80 ГБ) или эквивалент.
- Поставка в Docker, Linux x86_64. Развёртывание по инструкции не дольше 2 часов.
- Метрики сервиса для мониторинга (задержка, очередь, ошибки) в формате Prometheus.

6. Критерии приёмки

Контрольные наборы формирует исполнитель и передаёт заказчику вместе с результатом. Заказчик вправе повторить измерения на своих фразах.

Показатель	Требование
Естественность (MOS)	Не ниже 4,0 по каждому языку и каждому голосу; оценивают не меньше 20 носителей каждого языка, звук после кодирования в 8 кГц

Показатель	Требование
Разборчивость	Доля ошибок в символах (CER) при распознавании синтезированной речи не больше 3% по каждому языку
Числа, даты, имена	На контрольном наборе из 500 фраз сценариев (суммы, даты, проценты, казахские имена) не меньше 99% фраз без ошибок в чтении
Смешанные фразы	На 200 фразах с переключением kk ↔ ru не больше 2% фраз с ошибкой языка или произношения
Задержка	Первый фрагмент не позже 300 мс при 20 одновременных потоках на одном GPU из раздела 5
Новый голос	По 2 часам записей обучается не дольше 24 часов; сходство с диктором и MOS — не ниже 4,0
Устойчивость	72 часа непрерывной работы под нагрузкой без перезапусков и утечек памяти

7. Данные, права и безопасность

- Модель, веса и компоненты — с лицензией, допускающей коммерческое использование и дообучение. Исполнитель передаёт перечень лицензий всех компонентов и наборов данных.
- Записи дикторов — с письменным согласием диктора на создание и коммерческое использование синтезированного голоса. Права на голос оформляются договором.
- Заказчик не передаёт исполнителю записи реальных звонков и персональные данные клиентов. Обучающие данные исполнитель собирает сам; тексты фраз сценариев без персональных данных заказчик предоставляет.
- Обработка в соответствии с Законом РК «О персональных данных и их защите», данные хранятся на территории РК.
- Модель не должна клонировать голос по короткому образцу без явного запуска процедуры дообучения администратором.

8. Этапы

1. Прототип: один голос, оба языка, отчёт о метриках на контрольных наборах — 4–6 недель.
2. Второй голос, пайплайн дообучения, нагрузочные испытания — 4 недели.
3. Пилот на реальных звонках агента в контуре заказчика — 2–4 недели.
4. Передача исходного кода, весов, скриптов обучения, инструкций по развёртыванию и дообучению, описания API.

9. Что не входит

Логика диалога, распознавание речи, языковая модель, телефония и интеграция с АТС, закупка оборудования, доступ к системам заказчика и к записям реальных

ЗВОНКОВ.

10. Формат взаимодействия

Рассматриваем готовые продукты с локальным развёртыванием и разработку или дообучение под заказ. Права на результат и порядок использования определяются договором. Условия — по итогам рассмотрения технических и коммерческих предложений. Ответственное лицо: Андрей Водяников.