

Создание AI-агентов с использованием low-code BPM-системы Platform V Flow для
обработки клиентских обращений
Сбер 2026

АННОТАЦИЯ

ПАО «Сбербанк» — крупнейший банк России с развитой системой процессного управления. На сегодняшний день 100% продуктов и сервисов банка охвачены портфелем BPM-инициатив: все процессы включены в реестр, описаны, обеспечены непрерывным циклом редизайна и мониторингом метрик. Ключевым инструментом автоматизации процессов с 2020 года выступает BPM-система Platform V Flow (разработка СберТех). Цифровая трансформация банка затрагивает как внутренние операции, так и клиентские сервисы, а дополнительными драйверами выступают импортозамещение и активное внедрение технологий искусственного интеллекта (AI).

На конкурс представлены два флагманских проекта 2025 года в сфере клиентского сервиса (2-я линия поддержки), реализованных на базе Platform V Flow: автоматизация обработки обращений по комиссиям за дебетовые карты и консультирование клиентов по задолженности кредитных карт.

Данные проекты демонстрируют передовой опыт практического применения AI: решение по обращениям принимает мультиагентная система, в которой задействованы AI-агентов с различными ролями, реализованными на связке Platform V Flow и LLM GigaChat. Внедрение Platform V Flow позволило мгновенно анализировать данные из различных источников. Благодаря использованию данных проектов удалось повысить удовлетворенность внешних клиентов, сократить трудозатраты и время обработки обращений клиентов при сохранении качества сервиса по более чем **30 тыс. обращений клиентов в месяц**.

ВВЕДЕНИЕ

Предпосылкой создания представленных проектов стало принятое в 2020 году решение о поэтапном переходе от монолитной архитектуры к микросервисной и процессной модели на базе low-code BPM-системы Platform V Flow. На первом этапе, с 2020 по 2024 годы, был реализован базовый функционал автоматизации с детерминированной логикой принятия решений: процессы описывались в виде BPM-моделей, настраивались продуктовые интеграции, а ключевые решения оставались в зоне ответственности сотрудников.

С 2025 года стартовал новый этап развития — переход к недетерминированной логике с использованием LLM GigaChat для обработки текстовой информации. Это позволило автоматизировать те части бизнес-процесса, где решения ранее могли приниматься только сотрудниками на основе свободного текстового обращения клиента.

БИЗНЕС-КОНТЕКСТ

Проект автоматизации обработки обращений по комиссиям по дебетовым картам.

В рамках развития процессного управления и повышения уровня автоматизации клиентского сервиса реализован проект по переводу обработки обращений с тематикой «комиссии по дебетовым картам» на мультиагентное AI-решение на базе Platform V Flow и GigaChat. Проект реализуется на 2-й линии клиентской поддержки и направлен на снижение доли ручных операций в отложенной обработке обращений при сохранении качества сервиса и соблюдении регламентных сроков.

До запуска проекта значимая часть цепочки обработки обращений выполнялась сотрудниками на сценарных этапах специализированного рабочего места: требовалось вручную вычитывать обращение, сопоставлять описанную клиентом ситуацию с транзакциями по карте, определять тип комиссии, потребность клиента и составлять ответ клиенту.

Цель проекта — заменить сотрудников на ряде типовых сценарных этапов за счет применения мультиагентной AI-системы, интегрированной в BPM-модель процесса. Решение призвано: увеличить долю полностью автоматической обработки обращений, сократить трудозатраты на операции по вычитке и анализу запросов клиентов, снизить среднее время обработки при сохранении или росте удовлетворенности клиентов. Дополнительным эффектом является накопление унифицированной логики принятия

решений, которую можно тиражировать на смежные тематики и процессы.

Проект консультирования клиентов по задолженности кредитных карт.

Проект ориентирован на автоматизацию обработки отложенных обращений 2-й линии по вопросам клиента о задолженности (структура долга, размер платежа, просрочка, зачисления, операции и т.д.).

Особенности данного направления:

- широкий и неоднозначный пул вопросов клиентов;
- сезонный характер потока;
- широкая вариативность кейсов;
- обработка неструктурированной текстовой информации из различных источников.

Комбинация данных факторов делает невозможным реализацию классической детерминированной автоматизации без участия сотрудника.

БИЗНЕС-ПРОЦЕСС

Проект автоматизации обработки обращений по комиссиям по дебетовым картам.

Для реализации целевой модели процесса использована low-code-платформа Platform V Flow, в которой описана бизнес-логика обработки обращений, настроены интеграции с продуктовыми системами и реализовано взаимодействие с GigaChat. Запуск процесса инициируется при регистрации обращения оператором 1-й линии клиентской поддержки. После регистрации обращение попадает в Platform V Flow, где создается экземпляр BPM-модели.

После запуска экземпляра BPM-модель, вместо передачи обращения на ручной сценарный этап, инициирует работу мультиагентской системы, состоящей из четырех AI-агентов, выполняющих роли «Классификатор», «Планировщик», «Исполнитель» и «Контроль качества».

Агент-Классификатор анализирует потребность клиента, уточняет тематику обращения и выделяет ключевые параметры для дальнейшей обработки. На этом шаге происходит приведение свободного текстового описания к формализованной категории и набору признаков, необходимых для принятия решения в рамках BPM-модели.

Агент-Планировщик на основе результата классификации и доступного набора действий формирует пошаговый пронумерованный план обработки обращения. Он выступает «мозгом» системы: определяет, какие проверки и в каком порядке выполнить, когда требуется запрос транзакционной информации, в каких случаях возможна автоматическая обработка, а в каких — необходимо подключение сотрудника.

Агент-Исполнитель является «руками» системы: последовательно исполняет шаги плана, инициирует вызовы интеграций для получения и анализа данных по операциям, выбирает релевантные транзакции и выполняет другие предусмотренные действия. При необходимости он может вызывать отдельные действия несколько раз, пока план не будет полностью отработан или пока не станет очевидна необходимость передачи обращения человеку. При необходимости Исполнитель привлекает эксперта-сотрудника, задача на сотрудника формируется при помощи TaskList Platform V Flow.

Контроль качества проверяет работу остальных агентов: оценивает корректность классификации, обоснованность плана и выбранных действий, а также соответствие решения регламентам. В случае выявления ошибок или неоднозначностей он возвращает обращение в цикл Агенту который допустил ошибку с уточняющим комментарием, тем самым обеспечивая дополнительный уровень контроля.

Проект консультирования клиентов по задолженности кредитных карт.

В отличие от сценария по дебетовым картам, фокус агента — формирование точного консультационного ответа по индивидуальной финансовой ситуации клиента за счёт доменных знаний по кредитным картам и контроля корректности формулировок.

Этапы обработки:

- AI-Классификатор: Определяет потребность клиента, нормализует её до списка формализованных вопросов.
- AI-Планировщик (диспетчер): Формирует план для формирования ответа клиента. Определяет кто требуется: доменный агент или сотрудник-эксперт.
- Доменный агент по кредитной карте: Объединяет исполнительную логику, извлечение знаний из базы и данные из профильных систем. Преобразует вопрос на естественном языке в проверяемый ответ: уточняет контекст, подбирает правила, сопоставляет с данными карты и потребности клиента.
- AI-Коммуникатор: Собирает единый текст ответа в официальной тональности. Коммуникатор состоит 4 отдельных агентов с доменными

знаниями по конкретным вопросам клиента. Выбор нужного коммуникатора определяется Планировщиком.

- AI-Контроль качества: Проверяет корректность и полноту; при проблемах возвращает на Планировщику на доработку с комментарием.
- Опциональный шаг. При необходимо готовый ответ передается сотруднику для проверки. Задача на сотрудника формируется при помощи TaskList Platform V Flow
- Отправка ответа клиенту

ИННОВАЦИОННОСТЬ

Инновационность определяется следующими аспектами:

Снижение time-to-market. Проект демонстрирует кратное сокращение time-to-market по сравнению с классической разработкой: для процесса обработки обращений по комиссиям по дебетовым картам от старта до вывода MVP в промышленную среду прошло 2 недели, полноценный функционал ещё 2 недели. Второй проект — консультирование клиентов по задолженности кредитной карты, более сложный по составу интеграций и количеству агентов был полностью реализован за 6 недель. Такое ускорение стало возможным благодаря переносу бизнес-логики из программного кода в промпты: описание логики текстом и визуальными схемами выполняется значительно быстрее, чем традиционная разработка.

Мультиагентная AI-архитектура в промышленном процессе. Впервые в процессе рассмотрения обращений клиентов реализована мультиагентная система, где AI-агенты с выделенными ролями совместно обрабатывают клиентские обращения в рамках единого BPM-процесса. Каждый агент решает свою задачу, а их взаимодействие оркестрируется Platform V Flow, что обеспечивает прозрачность и управляемость каждого шага принятия решения.

Недетерминированная логика на реальном потоке. Проекты демонстрируют переход от классической детерминированной автоматизации к работе с неструктурированной текстовой информацией — свободными описаниями обращений, суммаризациями диалогов, комментариями операторов. Именно возможность анализировать и интерпретировать естественный язык клиента позволила автоматизировать те этапы процесса, которые ранее были доступны исключительно

сотрудникам.

Тиражируемость решения. Архитектура представленных проектов изначально спроектирована как универсальный шаблон, пригодный для быстрого тиражирования на другие процессы клиентской поддержки. Мультиагентная система отделена от конкретной предметной области: для адаптации под новый процесс достаточно заменить промпты AI-агентов и настроить необходимые продуктовые интеграции. Это сокращает трудозатраты и сроки запуска новых решений на базе AI, превращая каждый последующий проект из полноценной разработки в конфигурирование готового решения.

ТРУДНОСТИ

Основные трудности при реализации проекта были связаны с освоением новых технологических инструментов и интеграцией их в существующие архитектурные решения. Команда впервые использовала LLM GigaChat, что потребовало дополнительного времени на изучение принципов её работы, проработку архитектуры взаимодействия и выстраивание безопасных каналов обмена данными.

Отсутствие ранее накопленной экспертизы в использовании LLM добавило сложности при проектировании процессов и настройке окружения для моделей. Для минимизации рисков команда провела несколько итераций прототипирования и внутреннего тестирования, что позволило выработать подходы к эффективной интеграции GigaChat с другими компонентами платформы.

Кроме того, значительное внимание уделялось вопросам безопасности и соответствия внутренним требованиям организации. Каждая конфигурация и новая версия сервисов, использующих LLM, проходила обязательную проверку, что увеличивало сроки вывода в эксплуатацию, но обеспечивало надежность и стабильность конечного решения.

В результате команда не только успешно преодолела технические сложности, но и сформировала компетенции, которые могут быть использованы при последующих проектах с применением LLM-технологий.

РЕЗУЛЬТАТЫ

Благодаря внедрению проекта автоматизации обработки обращений по комиссиям по дебетовым картам мы достигли следующих результатов:

- Старт GenAI трансформации клиентского сервиса;
- Сокращение трудозатрат. Доля обращений, требующих участия сотрудника,

снизилась до **13%**;

- Рост уровня автоматизации. Уровень автоматизации обработки обращений по тематике увеличился на **30%**;
- Сокращение времени обработки в **5 раз**;
- Снижение количества ошибок. Исключение человеческого фактора из цепочки принятия типовых решений привело к уменьшению количества ошибок.

Проект консультирования клиентов по задолженности кредитных карт показал результаты:

- Сокращение трудозатрат на анализ обращения и подготовку ответа клиенту
- Рост уровня автоматизации. Уровень автоматизации обработки обращений по тематике увеличился на **15%**;
- Сокращение времени обработки в **2 раз**;

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

В рамках проектов используется LLM-модель GigaChat 2 Max и инструмент low-code - Platform V Flow. Инструмент состоит из четырех компонентов:

- Designer – для моделирования процессов и настройки атрибутов;
- Engine – для исполнения процессов, отображение шагов выполнения;
- TaskList является универсальной системой взаимодействия с пользовательскими задачами, одновременно представляет интерфейс для работы с задачами, API для взаимодействия с задачами из сторонних backend/frontend сервисов и событийную систему доступную для потребителей.
- Designer предоставляет интерфейс для работы с построением процессом. Для моделирования можно использовать весь инструментарий нотации BPMN 2.0: устанавливать промежуточные события с различными типами, устанавливать типы задач и в зависимости от типа задач будет или вызваться внешний сервис, получать выполняются скрипт или вызвать вложенный процесс, экземпляр которого будет запущен в рамках выполнения процесса.

- Monitor предоставляет широкий функционал взаимодействия с собранными событиями, имеется возможность создавать дашборды которые отслеживают техническое состояние и стабильность экземпляров процессов. Дополнительно имеется возможность настроить поиск отклонений в событиях и в случае, если такое событие будет найдено - отправить уведомление необходимой группе пользователей на почту.

Каждый запуск бизнес-процесса создавал «Экземпляр» процесса, который выполняется в компоненте Engine. «Экземпляр» – это единичное исполнение процесса. На странице экземпляра в компоненте Engine выделяется зеленым цветом ветка выполнения процесса и текущий этап. Есть возможность просматривать журнал выполнения работ и ошибки.

Monitor предоставляет широкий функционал взаимодействия с собранными событиями, имеется возможность создавать дашборды которые отслеживают техническое состояние и стабильность экземпляров процессов. Дополнительно имеется возможность настроить поиск отклонений в событиях и в случае, если такое событие будет найдено - отправить уведомление необходимой группе пользователей на почту.

Frontend

- React - библиотека для разработки пользовательских интерфейсов;
- Zustand, Redux, MobX – менеджер управления состоянием;
- Jest - библиотека для UNIT тестирования;
- Библиотеки готовых UI интерфейсов
- React-query, Jest, Cypress, React Testing Library - библиотека для получения и управления данными в приложении.
- Библиотеки готовых UI интерфейсов

Backend

- Java 17, Java 11 - язык программирования;
- Spring Core, Spring MVC, Spring Data JPA, Spring Security - фреймворки;
- Hibernate - ORM библиотека;
- Maven - система сборки;
- Kubernetes, Openshift, Docker, Jenkins, nexus – CI/CD;
- JUnit, Mockito - unit-тестирование;

- OpenAPI - спецификация описания взаимодействия;
- Liquibase - библиотека миграций БД;
- SonarQube - статический анализ кода;
- PostgreSQL – БД;
- Hadoop, Apache Spark - инструменты построения витрин данных;
- Hive, Impala, Spark SQL - доступ к большим данным;
- Oozie, ELK - инструменты логирования и мониторинга;
- Apache Kafka – брокер сообщений.

Рисунок 2. Модель процесса консультирования клиентов по задолженности кредитных карт из Designer Platform V Flow.

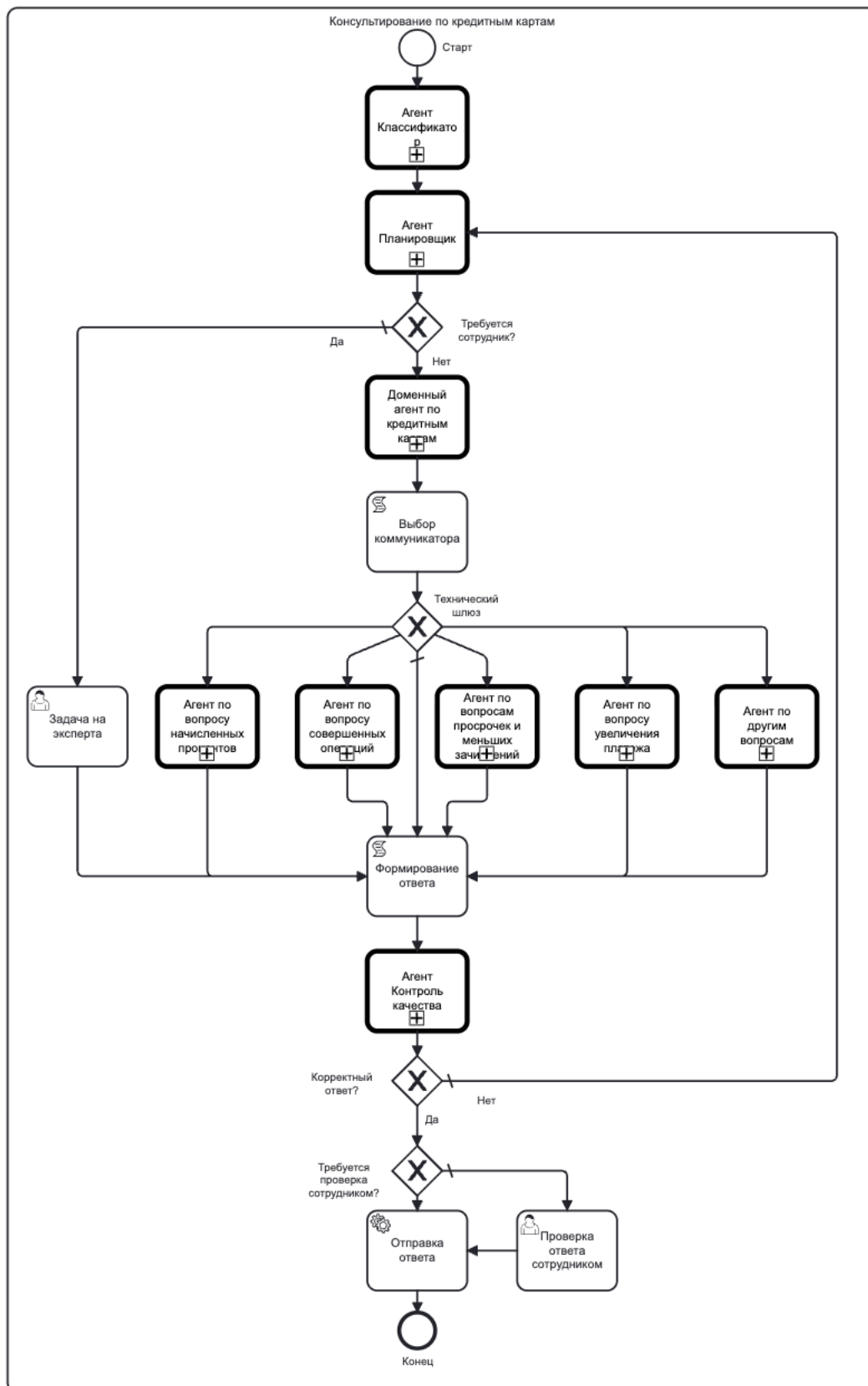


Рисунок 3. Пример построения дашбордов в компоненте Monitor. Приведённые данные подобраны случайным образом.

