ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК

программы для ЭВМ «Версиум AI»

1. Общие сведения

Наименование	Версиум AI (Versium AI)
Правообладатель	ООО «ВЕРСИУМ», ИНН 6670527467, ОГРН 1246600061014
Тип программы	Самостоятельно функционирующая программа для ЭВМ, серверное приложение с веб-интерфейсом
Класс программы	Шлюз и средство управления искусственным интеллектом в организации
Режимы работы	Веб-платформа; прокси для ИИ-трафика
Размещение	В контуре заказчика (on-premise), на его физических или виртуальных серверах
Пользовательский интерфейс	Веб-интерфейс, доступный из браузера; язык интерфейса — русский (также доступны английский, немецкий и испанский)
Программный интерфейс	API шлюза для приложений, сервисов и AI-агентов организации; программный интерфейс управления

2. Назначение и область применения

Программа предназначена для организации единой контролируемой точки доступа сотрудников, приложений и AI-агентов организации к языковым моделям. Программа обнаруживает и обезличивает чувствительные данные до их передачи в модель, останавливает передачу информации, запрещённой политикой организации, фиксирует обращения к моделям и лиц, получивших доступ к данным, и позволяет поручать AI-агентам выполнение задач на серверах организации в рамках строгих рабочих процессов и заданных прав.

Программа обеспечивает:

- использование языковых моделей и программных агентов на основе искусственного интеллекта в веб-чате;
- единый шлюз доступа к внешним, внутренним и локальным провайдерам языковых моделей;
- обнаружение и маскирование чувствительной информации, блокировку передачи данных;
- фиксацию запросов, ответов и лиц, получивших доступ к немаскированным данным;
- автономное выполнение задач AI-агентами по утверждённым рабочим процессам и расписаниям;
- формирование и сопровождение конфигураций внутренних и внешних систем, в том числе платформы Versium, посредством взаимодействия с языковой моделью.

Область применения – подразделения ИТ, информационной безопасности, разработки и аналитики организаций, применяющие языковые модели при работе со служебной информацией без её неконтролируемой передачи внешним провайдерам.

3. Режимы работы

- **Веб-платформа** – сотрудники работают с языковыми моделями и AI-агентами в веб-чате программы. Маскирование, политики доступа, лимиты и аудит применяются к каждому диалогу, шагу рабочего процесса и вызову инструмента.
- **Прокси для ИИ-трафика** – приложения, сервисы и AI-агенты организации направляют запросы к языковым моделям через API шлюза программы. Трафик маскируется, запросы с запрещённой информацией останавливаются и не передаются провайдеру, обращения фиксируются в журнале аудита.

Режимы работают одновременно и используют общие настройки провайдеров, правил маскирования, ролей и журналов.

4. Пользователи и роли

Доступ к функциям программы разграничивается ролями. Пользователю может быть назначено несколько ролей; при входе он получает объединение их полномочий.

Роль	Полномочия
Администратор	Настройка провайдеров и моделей, AI-агентов, навыков, MCP-серверов, исполнителей, скриптов, рабочих процессов, расписаний, сценариев запуска, пространств, правил маскирования и словарей; управление пользователями, группами, способами входа и лицензиями; перенос конфигурации; просмотр аудита и сводных показателей
Пользователь	Работа в веб-чате с доступными ему AI-агентами, запуск сценариев и рабочих процессов, работа в пространствах, подключение личных исполнителей, ведение личной базы знаний агентов, просмотр собственного расхода лимитов
ИБ-аудитор	Просмотр журнала аудита диалогов, журнала доступа к оригиналам и журнала файлов пространств; раскрытие немаскированных значений с обязательной фиксацией факта раскрытия

Пользователи объединяются в группы. Группам выдаются права на AI-агентов, рабочие процессы, исполнители, сценарии запуска и пространства; объект, право на который не выдано, пользователю не виден. Для интеграций с внешними системами применяются служебные (технические) учётные записи с ролью, ограниченной задачей интеграции.

Вход выполняется по локальной учётной записи, через Active Directory (в том числе по LDAPS) или по протоколу OpenID Connect с сопоставлением внешних групп ролям.

5. Функциональные возможности

5.1. Работа с языковыми моделями в веб-чате

Функция	Описание
Диалог с AI-агентом	Ведение диалога с выбранным AI-агентом в реальном времени с потоковым выводом ответа, остановкой ответа и повторной отправкой сообщения

Функция	Описание
История чатов	Перечень чатов пользователя с поиском, фильтрами и сортировкой; раздел недавних чатов и запусков процессов
Вложения и ссылки на файлы	Прикрепление изображений по политике провайдера; вставка ссылок на файлы и папки рабочей области исполнителя
Ход работы агента	Отображение шагов агента, вызываемых инструментов и списка задач с отметкой выполнения
Подтверждение действий	Запрос подтверждения пользователя перед вызовом инструмента, если это требует право агента; разрешение однократно или на весь сеанс
Уточняющие вопросы	Вывод вопросов агента с вариантами ответа и возможностью ввести собственный вариант
Управление контекстом	Автоматическое уплотнение истории беседы при приближении к пределу контекста модели; показ заполнения контекста, числа токенов и стоимости сеанса

5.2. AI-агенты и навыки

AI-агент программы задаётся как сочетание основной языковой модели, системной инструкции, навыков, инструментов и MCP-серверов и прав их использования.

- **Профиль агента** – название, описание, системная инструкция, основная модель; доступ к агенту выдаётся группам пользователей.
- **Права на инструменты** – для каждого инструмента (файловые операции, выполнение команд, инструменты MCP-серверов) задаётся режим: «Разрешено», «Подтверждать каждый раз», «Подтвердить один раз на сессию», «Запрещено», «Скрыто». Для MCP-сервера задаётся режим по умолчанию и, при необходимости, отдельный режим для конкретного инструмента.
- **Политика выполнения команд** – правила по командам, перечень дополнительно разрешённых путей, контроль выхода за рабочую папку и опасных конструкций командной оболочки.
- **Навыки (скиллы)** – переиспользуемые наборы инструкций в формате SKILL.md, загружаемые администратором текстом, файлом или архивом. Навык выдаётся конкретному агенту; агент вызывает навык, когда задача соответствует его описанию, и получает текст инструкции в контекст. Навык, не разрешённый агенту, агенту не виден.
- **База знаний агентов (LLM-вики)** – персональная база знаний, которую агенты ведут между чатами пользователя; включается пользователем, доступ регулируется политикой администратора («разрешено всем», «всем, кроме списка», «никому, кроме списка»).

5.3. Сценарии запуска и пространства

- **Сценарии запуска** – сохранённые связки «агент + исполнитель» (чат) или рабочий процесс, отображаемые пользователю плитками на главной странице. Видимость сценария: личная, для групп, для всей организации. Сценарий может иметь шаблон первого сообщения с параметрами.

- **Открытие сценария по внешней ссылке** – внешняя система (например, карточка инцидента платформы Versium) открывает сценарий по ссылке: программа создаёт новый чат и отправляет первое сообщение, собранное из шаблона. В ссылке передаются только значения параметров ограниченного формата; ход выполняется от имени вошедшего пользователя, с его правами, лимитами и аудитом.
- **Пространства** – рабочие области, объединяющие чаты, файлы, базу знаний и состав агентов. Поддерживаются личные пространства и общие пространства с доступом для групп. Для пространства задаются домашний исполнитель, допустимые агенты, доступ, архивирование; операции с файлами пространства фиксируются в журнале.

5.4. Рабочие процессы и планировщик

Функция	Описание
Визуальный редактор процессов	Построение процесса на холсте перетаскиванием узлов: начало, AI-агент, скрипт, параллельное выполнение, ожидание параллельных шагов, начало и конец цикла
Переходы по результату	Выбор следующего шага по коду результата предыдущего шага; агент выполняет только утверждённые шаги процесса
Циклы	Повтор участка процесса с ограничением числа итераций и досрочным выходом по признаку результата
Скрипты	Каталог скриптов (путь, аргументы) для выполнения на исполнителе с настраиваемым тайм-аутом
Проверка процесса	Проверка структуры процесса до сохранения и запуска
Запуск и наблюдение	Ручной запуск с начальным вводом, отображение состояния узлов и попыток, повтор упавшего шага в интерактивном запуске
Планировщик	Автоматический запуск процессов по расписанию в формате cron с учётом часового пояса; политики пропущенных запусков и наложения, ограничение длительности, запуск вне расписания, включение и отключение, история запусков

5.5. Исполнители

Исполнитель – отдельно устанавливаемая программа, выполняющая инструменты AI-агентов (чтение, запись, правка, поиск и удаление файлов, выполнение команд и скриптов) на сервере организации. Модели не выполняют действий напрямую: вызов инструмента проходит через шлюз программы, который восстанавливает исходные значения аргументов перед передачей исполнителю и маскирует результат перед возвратом модели.

- исполнитель устанавливается на серверы под управлением Linux и Windows из дистрибутива, загружаемого в веб-интерфейсе;
- исполнитель подключается к программе исходящим соединением по токену подключения, входящий порт на сервере исполнителя не требуется;

- исполнитель работает в пределах заданной рабочей папки; права на исполнитель выдаются группам;
- пользователь может подключить собственную машину как личный исполнитель для личных сценариев.

5.6. Подключение MCP-серверов

Программа подключает внешние MCP-серверы (Model Context Protocol), через которые AI-агенты получают доступ к информационным системам организации: системам контроля версий, системам управления задачами и ИТ-услугами, CMDB, SIEM, учётным системам, CRM и внутренним API.

- регистрация MCP-сервера: адрес, область действия (системный или на конкретном исполнителе), способ авторизации (без авторизации, токен, API-ключ, логин и пароль, OAuth2 Client Credentials), проверка сертификата сервера;
- автоматическое получение перечня инструментов сервера и их входных схем, контроль состояния подключения и последней ошибки;
- использование инструментов агентами только в пределах прав, заданных в профиле агента (п. 5.2);
- встроенный управляющий MCP-сервер, позволяющий администратору выполнять настройку программы из чата: MCP-серверы, агенты и их права, провайдеры и модели, правила маскирования, исполнители, сценарии запуска, расписания.

5.7. Защита данных: обнаружение и маскирование

Чувствительные данные в сообщениях, системных инструкциях, вложенных текстах и результатах инструментов обнаруживаются до передачи в модель и заменяются масками. В ответе модели маски заменяются исходными значениями, поэтому пользователь и исполнитель работают с настоящими данными, а провайдер получает только замаскированный текст. Внутри одного диалога одно и то же значение получает одну и ту же маску.

Способ обнаружения	Описание
Регулярное выражение	Собственный шаблон для предметных форматов: внутренние идентификаторы, артикулы и т. п.
Словарь (список слов)	Точное совпадение с заданным перечнем значений: логины, названия продуктов, кодовые имена проектов
Шаблон	Готовые шаблоны распространённых типов данных: адрес электронной почты, телефон, URL, IP-адрес, номер банковской карты
Диапазон чисел	Целые числа в заданных границах: числовые идентификаторы известного порядка
Алгоритм Луна	Номера банковских карт, проходящие проверку контрольной суммы
LLM NER	Распознавание именованных сущностей (имена, адреса, организации) по смыслу с помощью языковой модели

Способы замены: константа с уникальной нумерацией (например, «[REDACTED]-1») или подстановка правдоподобного значения из словаря замен. Словари замен ведутся вручную или наполняются с помощью AI-агента с предварительным просмотром. Набор и порядок механизмов

маскирования назначаются каждому провайдеру отдельно: например, для локальной модели в контуре организации маскирование может быть ослаблено, а для внешнего провайдера — применяться в полном объёме.

5.8. Блокировка передачи данных

- запрос, содержащий информацию, запрещённую к передаче, останавливается и не передаётся провайдеру; пользователь или приложение получает сообщение о причине блокировки;
- для каждого провайдера задаётся политика передачи изображений и двоичного содержимого, к которым не применяется маскирование: запрет, передача с предупреждением или фиксацией в аудите, разрешение;
- для прокси задаётся политика использования серверных инструментов провайдера модели: блокировать, только фиксировать в аудите, разрешать;
- запрещённые инструменты агентов и команды, не прошедшие политику выполнения, не исполняются.

5.9. Политики, роли и лимиты

- доступ к агентам, процессам, исполнителям, сценариям и пространствам по группам;
- выбор разрешённых провайдеров и моделей через состав агентов и маршрутов прокси;
- лимиты числа сообщений за 5 часов и за неделю для пользователя и для группы (на участника и на общий пул группы);
- недельный потолок расходов группы в денежном выражении;
- учёт стоимости запросов по ценам, заданным для моделей;
- безопасные настройки по умолчанию: запрет двоичного содержимого, скрытие навыков и инструментов, не выданных агенту явно.

5.10. Маршрутизация провайдеров

Программа ведёт единый список провайдеров языковых моделей – внешних, внутренних моделей организации и локальных моделей в контуре заказчика – и перечень моделей каждого провайдера. Ключи доступа к провайдерам хранятся в шлюзе в зашифрованном виде и не показываются в интерфейсе (допускается только перезапись). Приложения и агенты обращаются к модели по её имени в шлюзе, поэтому смена провайдера или модели выполняется в настройках программы без изменения приложений.

5.11. Прокси для ИИ-трафика

- приём запросов приложений, сервисов и AI-агентов организации через API шлюза в форматах распространённых программных интерфейсов языковых моделей, включая потоковую передачу ответа;
- аутентификация по персональным токенам доступа с ограниченным сроком действия, выпускаемым пользователю или технической учётной записи в веб-интерфейсе; отзыв токенов и напоминание об истечении срока;
- маршрутизация запроса на модель, разрешённую пользователю, с использованием ключа организации, хранящегося в шлюзе, или собственного ключа пользователя, который проходит через шлюз без сохранения;
- маскирование запроса и восстановление исходных значений в ответе, блокировка запросов по политикам п. 5.8;
- учёт токенов и стоимости запросов, фиксация обращений в журнале аудита;

- формирование готовых настроек подключения клиентского приложения при выпуске токена.

5.12. Генерация конфигураций платформы Versium

Программа обеспечивает формирование и сопровождение конфигураций внутренних и внешних систем в диалоге с языковой моделью. При совместном использовании с платформой Versium AI-агент с соответствующими навыками и MCP-сервером платформы настраивает объекты, формы, процессы и отчёты платформы по описанию задачи на естественном языке и действует через рабочие процессы и роли платформы.

- запуск сценария программы из карточки объекта платформы Versium по внешней ссылке (кнопка «Запрос к ИИ», п. 5.3);
- интеграция с Versium SOAR: по новому инциденту платформа создаёт чат по сценарию в общем пространстве и передаёт агенту события (решения по заявкам, новые инциденты группы) по защищённому обратному каналу с выпускаемым токеном;
- пакеты конфигурации: перенос настроек (провайдеры, агенты, навыки, процессы, правила маскирования, роли и др.) между установками программы с предварительным просмотром, защитой архива паролем и получением обновлений поставщика.

Версиум AI функционирует и без платформы Versium.

5.13. Аудит и журнал доступа

Функция	Описание
Аудит диалогов	Все обращения к моделям: чаты, шаги рабочих процессов, запросы через прокси, служебные запросы; пользователь, агент, модель, провайдер, применённая политика, длительность, токены, стоимость, состояние
Фактически переданные данные	Просмотр запроса в том виде, в котором он ушёл в модель (с масками), и перечня замаскированных сущностей
Журнал внешних обращений	Фиксация запросов, фактически отправленных провайдеру модели и внешним MCP-серверам
Раскрытие оригиналов	Просмотр исходных значений и восстановленного ответа после явного подтверждения; доступно администратору и ИБ-аудитору
Журнал доступа	Фиксация лиц, получивших доступ к немаскированным данным: пользователь, диалог, агент, действие, время; поиск и фильтры
Журнал файлов пространств	Фиксация операций с файлами пространств

5.14. Администрирование

- управление пользователями: учётные данные, активность, роли, группы, личные лимиты;
- управление группами пользователей и их составом;
- настройка входа через Active Directory и OpenID Connect, сопоставление внешних групп с ролями;
- сводная панель (дашборд): расходы, число запросов и токенов, доля успешных запросов, длительность обработки, число активных пользователей, расход лимитов групп, состав настроенных объектов;

- личные настройки пользователя: язык интерфейса, часовой пояс, смена пароля.

5.15. Лицензирование

- загрузка файла лицензии в веб-интерфейсе, проверка электронной подписи ключа и его принадлежности продукту встроенным модулем лицензирования локально, без обращения к внешним серверам;
- отображение заказчика, адреса установки, даты выдачи и срока действия лицензии, предупреждение о скором истечении срока;
- применение лицензионных ограничений по количеству пользователей, количеству MCP-серверов и сроку действия; поддержка срочных и бессрочных лицензий;
- при отсутствии действующей лицензии доступны только вход и управление лицензиями.

6. Интерфейсы взаимодействия

Интерфейс	Назначение
Веб-интерфейс	Работа пользователей в чате, сценариях и пространствах; администрирование, аудит, лицензирование (поверх программного интерфейса управления)
API шлюза (прокси)	Приём запросов приложений, сервисов и AI-агентов организации к языковым моделям и возврат ответов
MCP-интерфейс программы	Предоставление AI-агентам инструментов программы с маскированием аргументов и результатов, управляющий MCP-сервер
Канал исполнителей	Исходящее защищённое соединение исполнителя с программой для выполнения инструментов
Интерфейсы провайдеров	Подключение к внешним, внутренним и локальным языковым моделям по их программным интерфейсам
Внешние ссылки и интеграции	Открытие сценариев из внешних систем, обратный канал интеграции с платформой Versium

7. Входные и выходные данные

Вид	Состав
Входные данные	Сообщения пользователей и вложения; запросы приложений и AI-агентов через API шлюза; ответы языковых моделей; результаты инструментов исполнителей и MCP-серверов; параметры внешних ссылок и события интегрированных систем; настройки программы; файлы лицензии, навыков и пакетов конфигурации

Вид	Состав
Выходные данные	Ответы моделей с восстановленными исходными значениями; замаскированные запросы провайдерам; результаты действий AI-агентов и рабочих процессов; записи журналов аудита и доступа; сводные показатели использования, расходов и лимитов; пакеты конфигурации, дистрибутивы исполнителей, настройки подключения клиентских приложений