 Спикеры:



Александр Прокопенко
Лид тестирования



Ольга Сердобинцева
Эксперт по обезличиванию
данных, владелец продукта
«Маскировщик»



Ника Сулова
Директор банковской
практики в HFLabs



**Как безопасно работать с ИИ:
обезличивание данных в текстах
и документах**

О чем расскажем

Кто мы	01
Где нужна защита данных в свободном тексте	02
Зачем маскировать чувствительную информацию	03
Требования к системе маскирования	04
Подводные камни при работе с текстом	05
Как мы в HFLabs решаем эти задачи	06

Н F Labs

Лидер российского рынка обработки клиентских данных

Повышаем качество, доступность
и ценность клиентских данных
для эффективной работы бизнеса

2005

год основания

160

человек в штате



Полностью российская компания



Продукты входят в Реестр отечественного ПО

10

собственных
IT-решений

90+

проектов по управлению
качеством клиентских
данных

 ВТБ

 БМ
БАНК

 АЛЬФА
СТРАХОВАНИЕ

 СБЕР

OZON

РОСГОССТРАХ

 СОВКОМБАНК
СТРАХОВАНИЕ ЖИЗНИ

 лента

СОГАЗ

ИНГОССТРАХ

 Альфа-Банк

Нам доверяют лучшие

 Т БАНК

 Ренессанс
Банк

 отпбанк

 МТС БАНК

 ПОЧТА
БАНК

 ПОЧТА
РОССИИ

ГОСУСЛУГИ

 X5Group

 Ростелеком

М Т
С

 МЕГАФОН

М.ВидеоЭльдорадо



ИИ шагает по планете...

Обезличивание данных в свободном тексте: сферы и возможности

H F Labs



Улучшение фактуры для аналитики, тестирования и обучения моделей



Расчет потенциальных рисков кредитного конвейера



Работа с профилем клиента



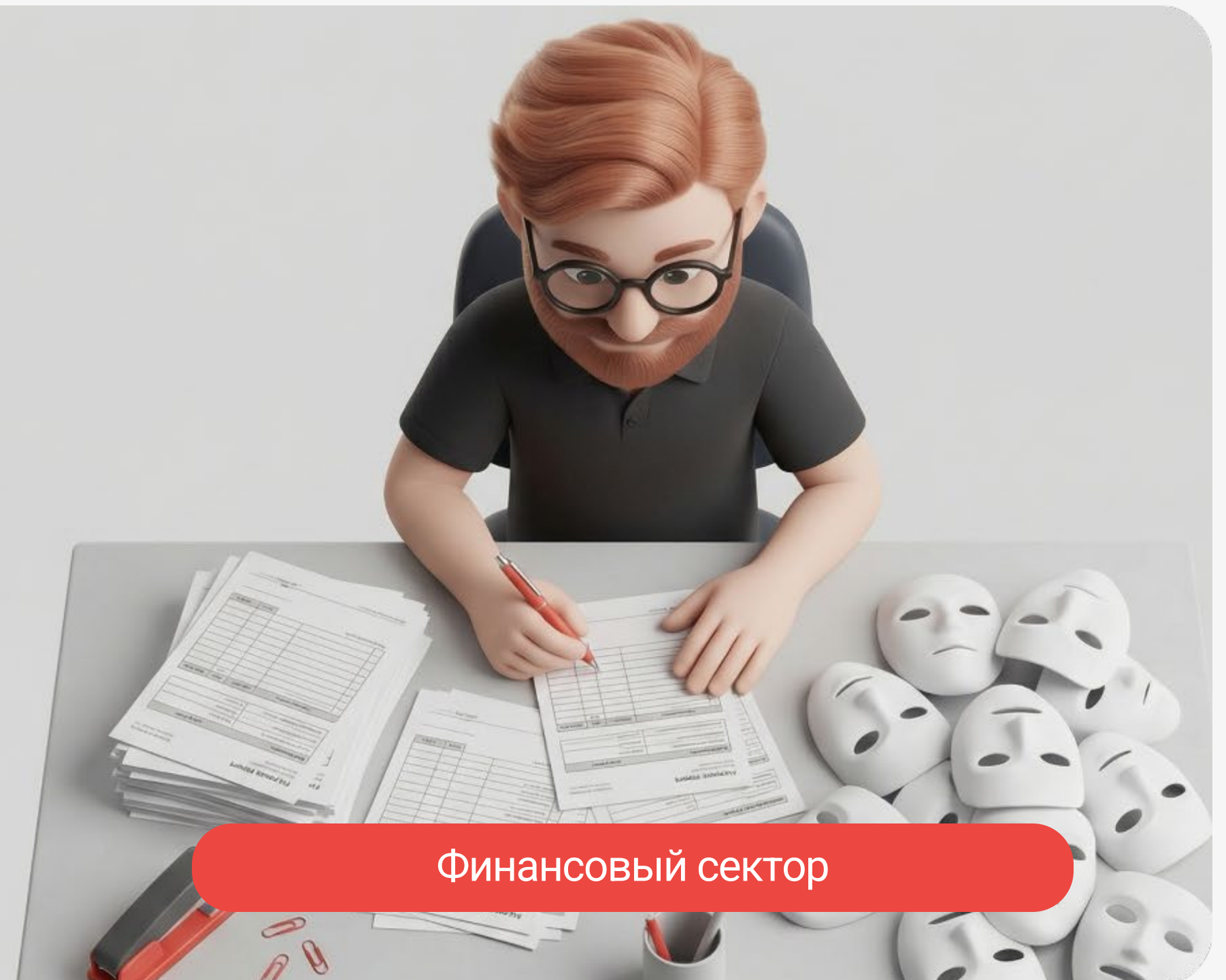
Противодействие мошенничеству (anti-fraud)



Выявление устойчивых тенденций оттока



Оцифровка согласий





Финансовый сектор


Обезличивание данных в свободном тексте: сферы и возможности

H F Labs

Рекламные технологии и розничная торговля

 Анализ потребительского спроса и таргетирования

 Образование экосистем и апсейл-схем, объединяющих дочерние компании





 Исследование спроса, таргетирование и анализ социально-демографических привычек

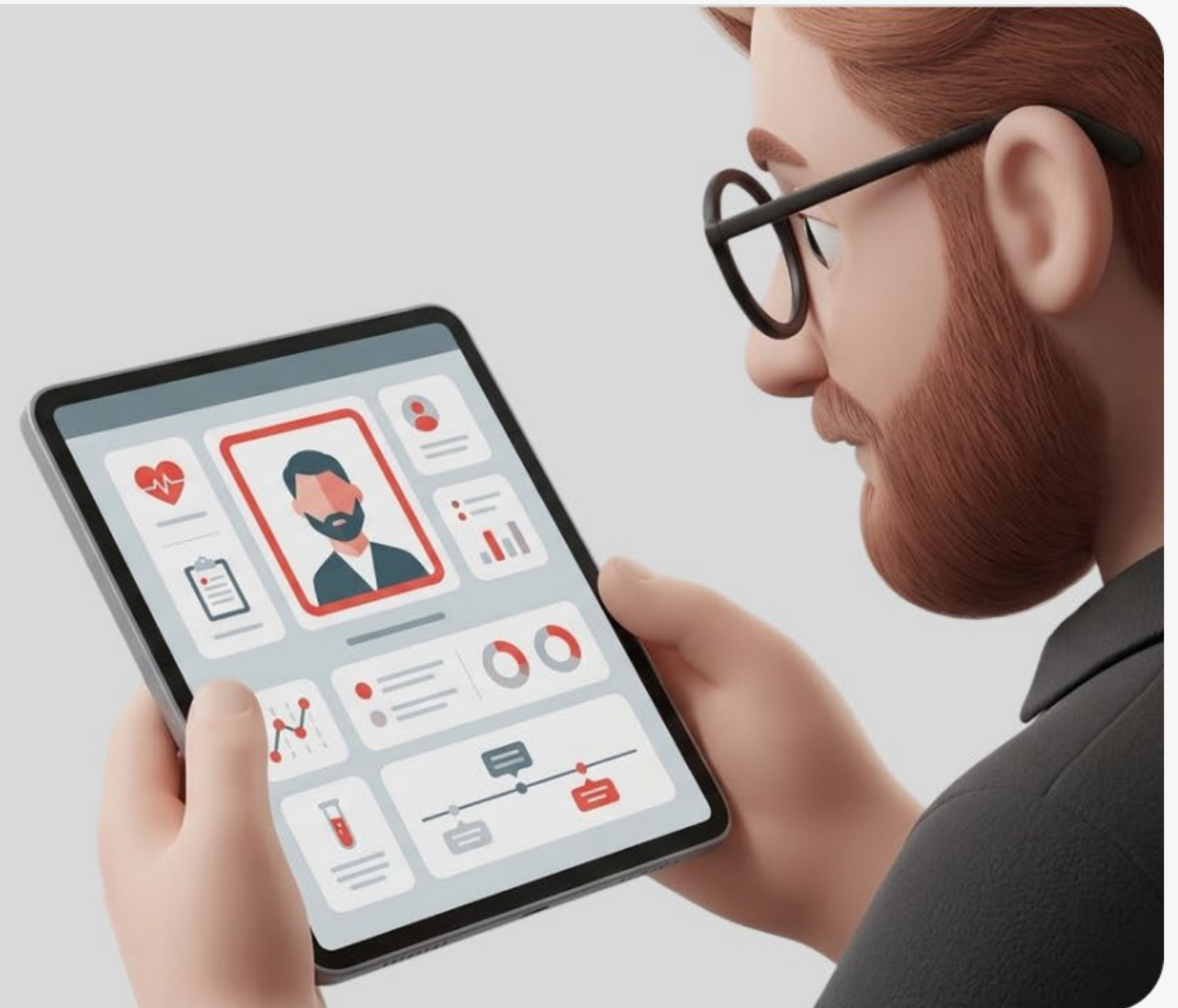


Обезличивание данных в свободном тексте: сферы и возможности

H F Labs

Медицинские исследования и здравоохранение

-  Агрегированная аналитика заболеваний по регионам и социальным группам
-  Валидация ведения и анализ индивидуальных историй болезни
-  Массовая проверка диагнозов и формирование второго мнения
-  Обмен статистикой при исследованиях заболеваний и разработке лекарств



Вместе с задачей ИИ-агент получает доступ к чувствительной информации

H F Labs

Любой документ с персональными данными — потенциальный источник утечки

✳️ Риски оборотных штрафов

🎮 Проблемы с соблюдением NDA

😞 Угроза утраты доверия

ДОГОВОР КУПЛИ-ПРОДАЖИ АВТОМОБИЛЯ

г. Санкт-Петербург, Пятое июля две тысячи двадцать третья

Продавец Трушкин Сергей Владиславович, паспорт № 47 01 № 501122, выдан Миграционной службой ФРГ У Санкт-Петербургу 20 июня 2006 года, проживающий по адресу: г. Санкт-Петербург, ул. Ботаническая, д. 14, кв. 58

Покупатель Филатов Павел Дмитриевич, паспорт № 41 03 № 274511, выдан ОВД Калининского района Санкт-Петербургу 12 октября 2009 года, проживающий по Санкт-Петербургу, пр. Солидарности, д. 20, кв. 31

Павел Дмитриевич отдал мне документы на подпись

Понял, благодарю вас, что написали!

Задача: найти и заменить без потери смысла и консистентности

H F Labs

Безопасно

- Кратно снизить риски нарушения требований 152-ФЗ.
- Использовать внешние модели в рутинных задачах.
- Дообучать модели без риска компрометации.



Задача: найти и заменить без потери смысла и консистентности

H F Labs

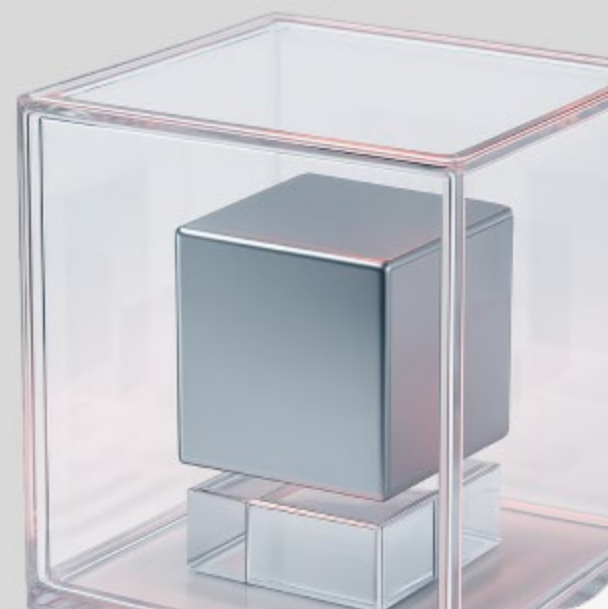
Безопасно

- Кратно снизить риски нарушения требований 152-ФЗ.
- Использовать внешние модели в рутинных задачах.
- Дообучать модели без риска компрометации.



Консистентно

- Удерживать контекст и профиль.
- Маскировать исходные данные одинаково во всех запросах.
- Сохранять связи при маскировании разных источников, баз и документов.



Задача: найти и заменить без потери смысла и консистентности

H F Labs

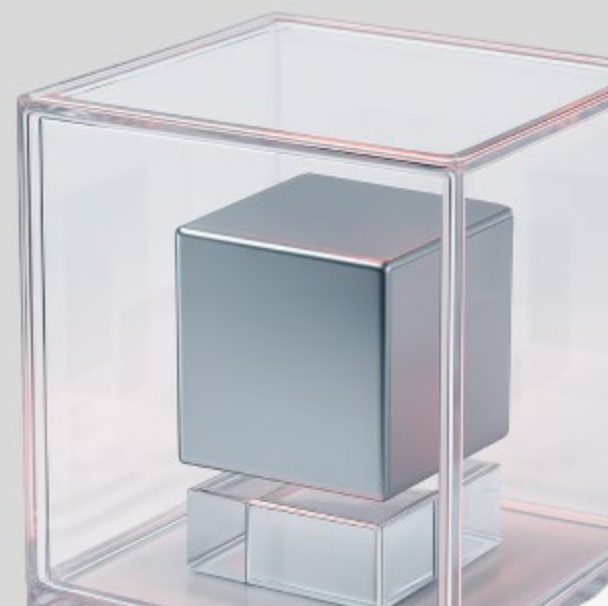
Безопасно

- Кратно снизить риски нарушения требований 152-ФЗ.
- Использовать внешние модели в рутинных задачах.
- Дообучать модели без риска компрометации.



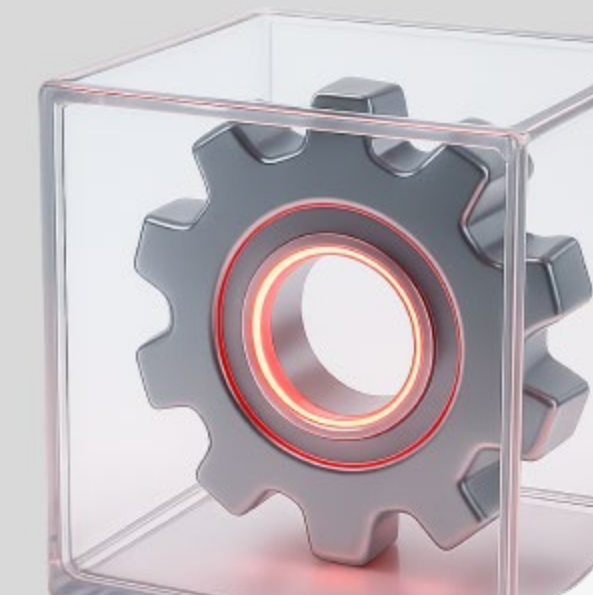
Консистентно

- Удерживать контекст и профиль.
- Маскировать исходные данные одинаково во всех запросах.
- Сохранять связи при маскировании разных источников, баз и документов.



Полезно

- Сохранять диапазоны соцдема.
- Проводить аналитику без потери смысла.
- Обучать модели на эквивалентных данных.



Технические особенности

Подводные камни при реализации



Контекст меняет смысл — модель должна это учитывать



Язык богат на омонимию



Нужно научить систему различать, что важно маскировать, а что — оставлять как есть



Сами замены должны быть консистентны и сохранять смысл

Консистентное маскирование ФИО в латинице и кириллице

H F Labs

Smirnovu_Evgeniyu Kirillovichu_.docx ▾

Поиск

Макет Ссылки Рассылки Рецензирование Вид Справка

ООО «ТехПродукт»
ул. Рогова, д. 47, оф. 310
г. Москва, 123456
тел.: +7 (495) 123-45-67
info@tehproduct.ru

Генеральному директору
ООО «СтройГрупп»
Смирнову Евгению Кирилловичу

Исх. № 47/КП от 31 марта 2026 г.

КОММЕРЧЕСКОЕ ПРЕДЛОЖЕНИЕ

на поставку программного обеспечения и сопровождение

Консистентное маскирование ФИО в латинице и кириллице

H F Labs

ООО «МедЛаб»

Центр лабораторной диагностики
г. Москва, пр. Вернадского, д. 45, стр. 2 | тел.: +7 (495) 222-33-44 | medlab.ru

РЕЗУЛЬТАТЫ КЛИНИЧЕСКОГО АНАЛИЗА КРОВИ

Patient information (auto)

Patient name:

Sokolova Irina Vladimirovna

Date:

31.03.2026 09:14

ОПИСАНИЕ

Пациентка **Соколова Ирина Владимировна**, 40 лет. Направлена в рамках профилактического осмотра. Жалоб на момент забора крови не предъявляла.

Замена «ложных» СИМВОЛОВ в сторону кириллицы или латиницы

H F Labs

Распознавание ФИО с латинскими
буквами внутри кириллицы

Введите текст

Баркова Лариса Андреевна

Нечеткий поиск

Обработать

Обработка опечаток и слитного написания

H F Labs

Уверенность выделения сущности.
Как это влияет на маскирование

Введите текст

Иван0в

Нечеткий поиск

Обработать

Поиск простых типов

На примере паспорта

H F Labs

1 Поиск полных паспортов
множественными шаблонами

☰ Базовые варианты

✂ С оборотами речи, вариантами
перестановок триггерных слов,
шумом

✂ Без триггерных слов
«паспорт», «серия», и пр.

Поиск простых типов

На примере паспорта

H F Labs

1 Поиск полных паспортов
множественными шаблонами

Базовые варианты

С оборотами речи, вариантами
перестановок триггерных слов,
шумом

Без триггерных слов
«паспорт», «серия», и пр.

2 Поиск фрагментов — если
человек написал только часть

Только серия или только номер

Начало / конец значения

Упоминание в живой фразе,
а не в шаблонном виде

Поиск простых типов

На примере паспорта

H F Labs

1 Поиск полных паспортов
множественными шаблонами

Базовые варианты

С оборотами речи, вариантами
перестановок триггерных слов,
шумом

Без триггерных слов
«паспорт», «серия», и пр.

2 Поиск фрагментов — если
человек написал только часть

Только серия или только номер

Начало / конец значения

Упоминание в живой фразе,
а не в шаблонном виде

3 Поддержка значительного
разрыва между триггером
и значением

Добрый день,
чем можем помочь?

Здравствуйте!

Менял свой **паспорт**. Старый
потерял в проруби, дело было
в прошлый четверг. Вроде бы
старый оканчивается на ***27150**,
могли бы проверить?

Пересечение контекста

Ложные срабатывания

H F Labs

Правильное маскирование невозможно без анализа окружения

Блохина

Блохина

Пересечение контекста

Ложные срабатывания

H F Labs

ФИО считываются
в фрагменте адреса

Арендодатель передает во временное пользование нежилое помещение площадью 47 кв. м, расположенное по адресу: г. Санкт-Петербург ул. Блохина, д. 14, офис 7. Срок аренды – 12 месяцев с момента подписания

Общество с ограниченной ответственностью «Сервис-Плюс», именуемое в дальнейшем «Исполнитель», в лице генерального директора Королева Игоря Владимировича, действующего на основании Устава, с одной стороны, и Блохина Анна Сергеевна, именуемый в дальнейшем «Заказчик», с другой стороны, заключили настоящий договор о нижеследующем

Пересечение контекста

Ложные срабатывания

H F Labs

Обычные слова
выглядят как ФИО

Для оформления заграничного паспорта необходимо предоставить копию внутреннего паспорта и фотографии установленного образца.

На основании доверенности №15 от 15.02.2026 представитель Заграничного Михаила Сергеевича вправе подписывать документы от его имени.

Пересечение контекста

Ложные срабатывания

H F Labs

Обычные слова
выглядят как ФИО

Всероссийское общество глухих проводит ежегодный съезд делегатов.
В работе форума примут участие представители из 74 регионов страны.

Протокол подписан председателем Вороновой Е. И. и секретарем
Глухих Д. Ю., действующим на основании доверенности № 47 от 12.01.2026.

Пересечение контекста

Ложные срабатывания

H F Labs

Обычные слова
выглядят как ФИО

В счет включено: роза чайно-гибридная (25 штук), упаковка подарочная, открытка. Итого к оплате — 4400 руб.

Покупатель Роза Андреевна Кленова подтверждает получение товара в полном объеме и претензий к качеству не имеет. Подпись:

Пересечение контекста

Ложные срабатывания

H F Labs

Медицинские термины
похожи на ФИО

Критерии Вильямса, Ди Джорджи Нунана отрицательные

Синдром Боткина отрицательный

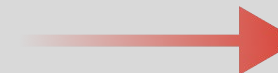
Симптомы Прадер–Вилли, Элерса–Данлоса и Лойса–Дитца у **Кляузера Романа**

К каким рискам это ведет?



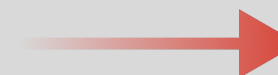
1 Маскируется лишнее — теряется контекст и смысл

Марка автомобиля



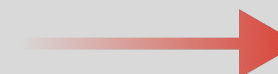
Евгения автомобиля

Сахарный диабет



Скеленчев диабет

Синдром Паркинсона



Синдром Недобойкина


2 Персональные данные остаются
без маскирования


Мой номер паспорта оканчивается на *17534270

«Маскировщик. Вуаль»


H F Labs

Инструмент для обезличивания чувствительной информации в текстах

 Находит в тексте персональные данные

 Заменяет их синтетическими с сохранением смысла

 Отправляет безопасный запрос в нейросеть

 Опционально демаскирует данные





Мониторинг отправленных
вовне чувствительных данных

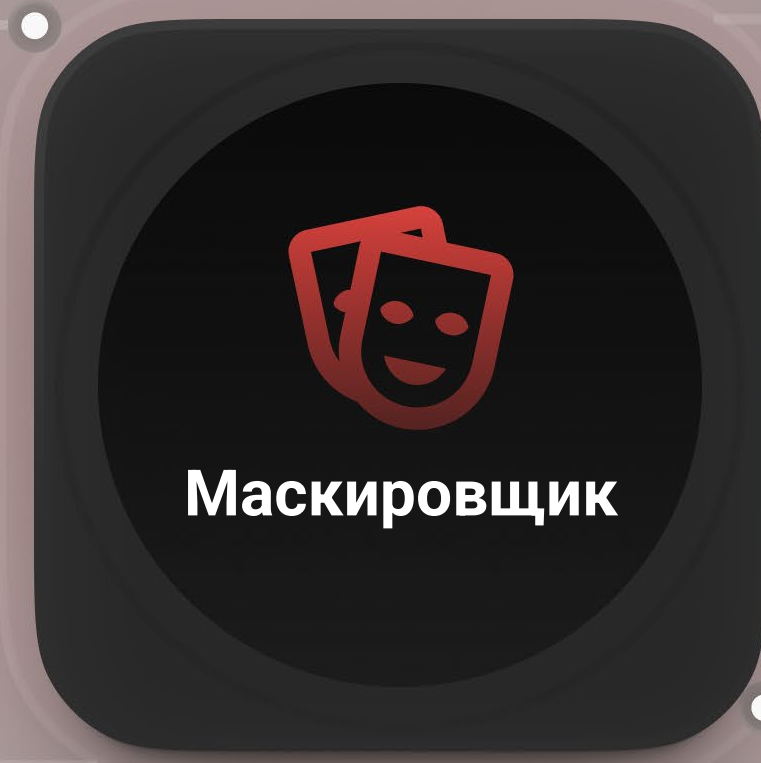
Безопасная работа с GPT-агентами

H F Labs

Вопрос с оригинальными данными



1



2

Вопрос с замененными ПД



3

4

Ответ с возвращенными оригинальными значениями

Ответ может содержать замененные значения

Пример работы с GPT

Замаскировали

Заменяли на оригинальные

20 марта 2025

О чем текст ниже? Я , Орлов Игорь Владимирович, документ, удостоверяющий личность 7010 № 256423, выдан ТП УФМС Москвы, зарегистрированный (ая) по адресу: Москва ул Мира дом 13 кв 33, согласен (а) на обработку моих персональных данных: (фамилия, имя, отчество; дата рождения; контактный телефон (дом., мобильный, рабочий); адрес проживания; место работы (учебы), должность, ученая степень, ученое звание, адрес организации, информацию о банковских счетах) Администрацией Томской области с целью обработки материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры.

Маскирование завершено

О чем текст ниже? Я , Волков Григорий Айдарович, документ, удостоверяющий личность 5463 № 787580, выдан ТП УФМС Москвы, зарегистрированный (ая) по адресу: РОССИЯ, ГОРОД МОСКВА, ПРОЕЗД ТРОЕКУРОВСКИЙ, согласен (а) на обработку моих персональных данных: (фамилия, имя, отчество; дата рождения; контактный телефон (дом., мобильный, рабочий); адрес проживания; место работы (учебы), должность, ученая степень, ученое звание, адрес организации, информацию о банковских счетах) Администрацией Томской области с целью обработки материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры.

Отправка в chat gpt завершена

Автор текста, Волков Григорий Айдарович, даёт согласие на обработку своих персональных данных Администрацией Томской области. Цель обработки — рассмотрение материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры. В согласии указаны персональные данные, которые будут обработаны: Ф. И. О., дата рождения, контактные телефоны, адрес проживания, место работы (учёбы), должность, а также информация об организации и банковских счетах автора.

Демаскирование завершено

Автор текста, Орлов Игорь Владимирович, даёт согласие на обработку своих персональных данных Администрацией Томской области. Цель обработки — рассмотрение материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры. В согласии указаны персональные данные, которые будут обработаны: Ф. И. О., дата рождения, контактные телефоны, адрес проживания, место работы (учёбы), должность, а также информация об организации и банковских счетах автора.

Нашли персональные данные

Нашли замаскированные данные



Попробуйте «Маскировщик»
в бесплатном пилоте



«Маскировщик»
на сайте HFLabs



Александр Прокопенко
Лид тестирования
✉ alexanderp@hflabs.ru



Ольга Сердобинцева
Эксперт по обезличиванию
данных, владелец продукта
«Маскировщик»
✉ olgaser@hflabs.ru



Ника Сулова
Директор банковской
практики в HFLabs
✉ nikas@hflabs.ru