

Нейросетевой машинный перевод и анализ текста

КАЧЕСТВО
СКОРОСТЬ
БЕЗОПАСНОСТЬ



О компании

ООО «ПРОМТ»

- Отечественный разработчик ПО в области ИИ
- Лучшие эксперты в области МП в России с 1991 года
- Один из ведущих вендоров МП в мире
- Уникальная команда разработчиков – более 40 программистов и лингвистов
- Все продукты компании внесены в Единый реестр отечественного ПО

Клиенты

amadeus



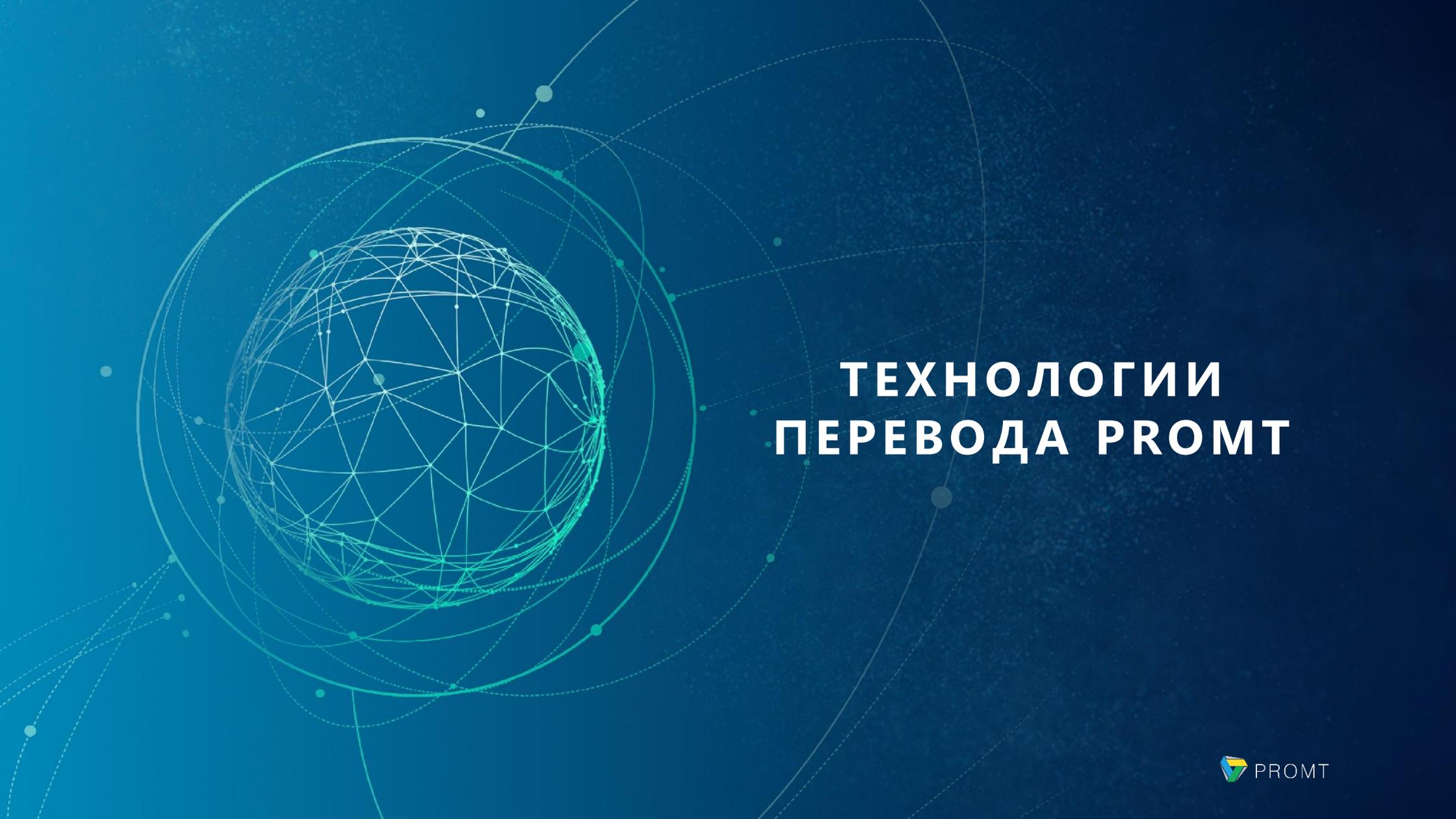
SIEMENS



ЛУК LUKOIL

SpanishD!ct





ТЕХНОЛОГИИ ПЕРЕВОДА PROMT

Определение ИИ: задачи, решаемые человеком – может решать машина

Машинный перевод – одна из самых успешных и широко распространенных реализаций Искусственного Интеллекта (ИИ)



Распознавание речи



Классификация текстовых сообщений в потоке



Распознавание образов (предметов, явлений, лиц)



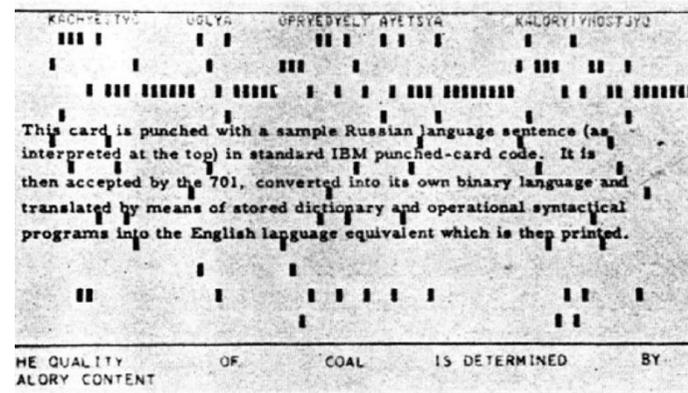
Идентификация людей по голосу



Перевод текста

7 января 1954 года
Georgetown University & IBM

Русско-английский переводчик:
6 грамматических правил, 250 лексических единиц; было переведено 200 фраз



О технологиях машинного перевода в мире

КРАТКАЯ ИСТОРИЯ МАШИННОГО ПЕРЕВОДА:

Rule-Based MT с 1950 гг.

Машинный акцент
в переводе,
предсказуемый
результат

ОБУЧЕНИЕ ЧЕРЕЗ
СЛОВАРИ

Statistical MT с 2000 гг.

Более гладкий
перевод чем в RBMT,
но непредсказуемый
результат

ОБУЧЕНИЕ ЧЕРЕЗ
ПАРАЛЛЕЛЬНЫЕ ДАННЫЕ

Neural MT с 2015 гг.

Точный и гладкий перевод,
результат на уровне
профессионального
переводчика

ОБУЧЕНИЕ ЧЕРЕЗ
ПАРАЛЛЕЛЬНЫЕ ДАННЫЕ

Гибридная технология PROMT Neural

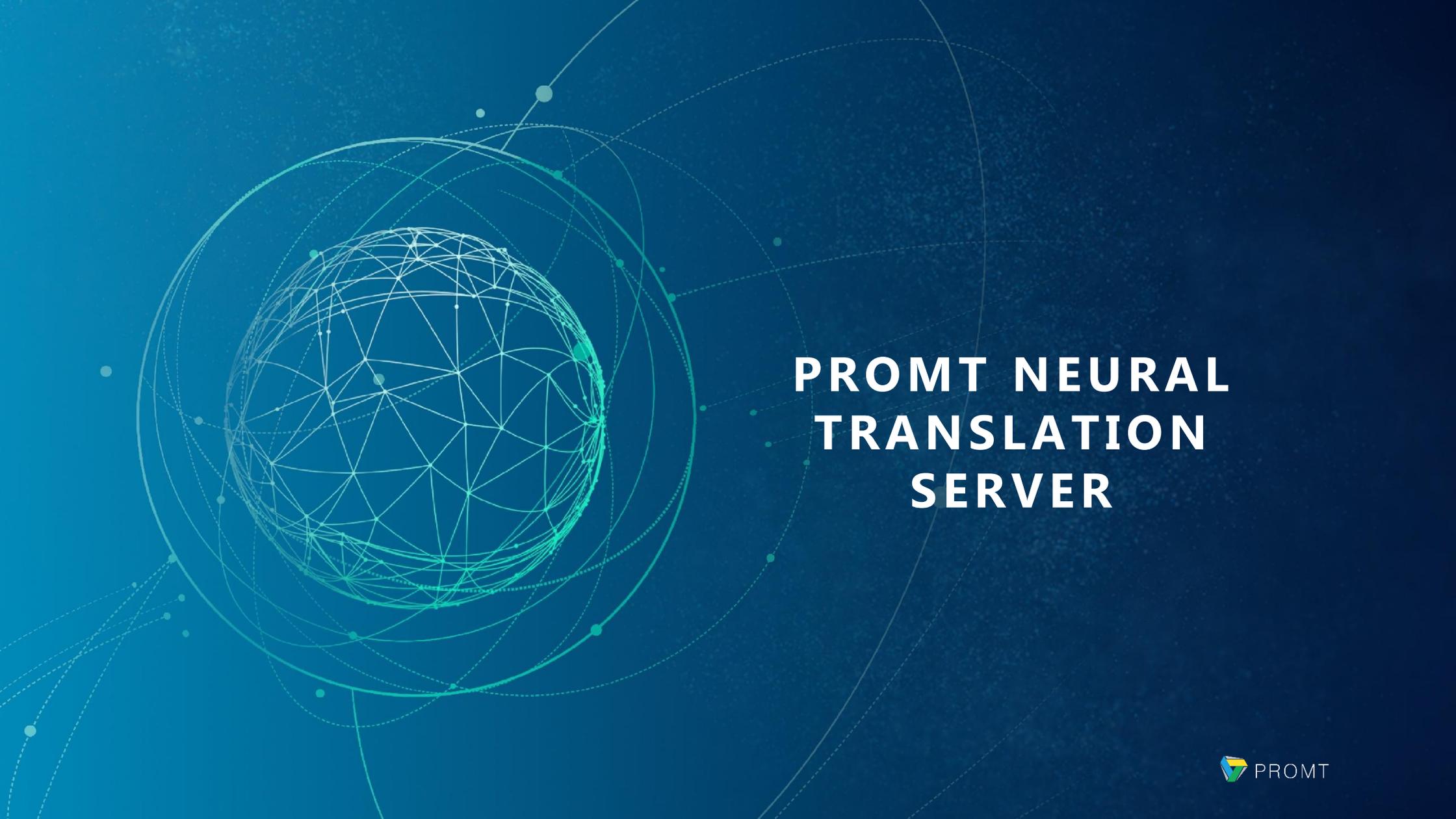
100% возможностей для решения бизнес-задач объединены в **PROMT Neural**

Гибридная технология PROMT Neural

Компания **PROMT** создала гибридную технологию перевода **PROMT Neural**, которая использует преимущества машинного обучения и алгоритмов семантико-синтаксического анализа для управления терминологией

- Высокая скорость перевода
- Высокое стартовое качество
- Предсказуемый результат
- Контроль за терминологией
- Обучаемость
- Постоянное развитие

3 ФАКТОРА УСПЕХА: ТЕХНОЛОГИЯ, ПРОДУКТ И ДАННЫЕ



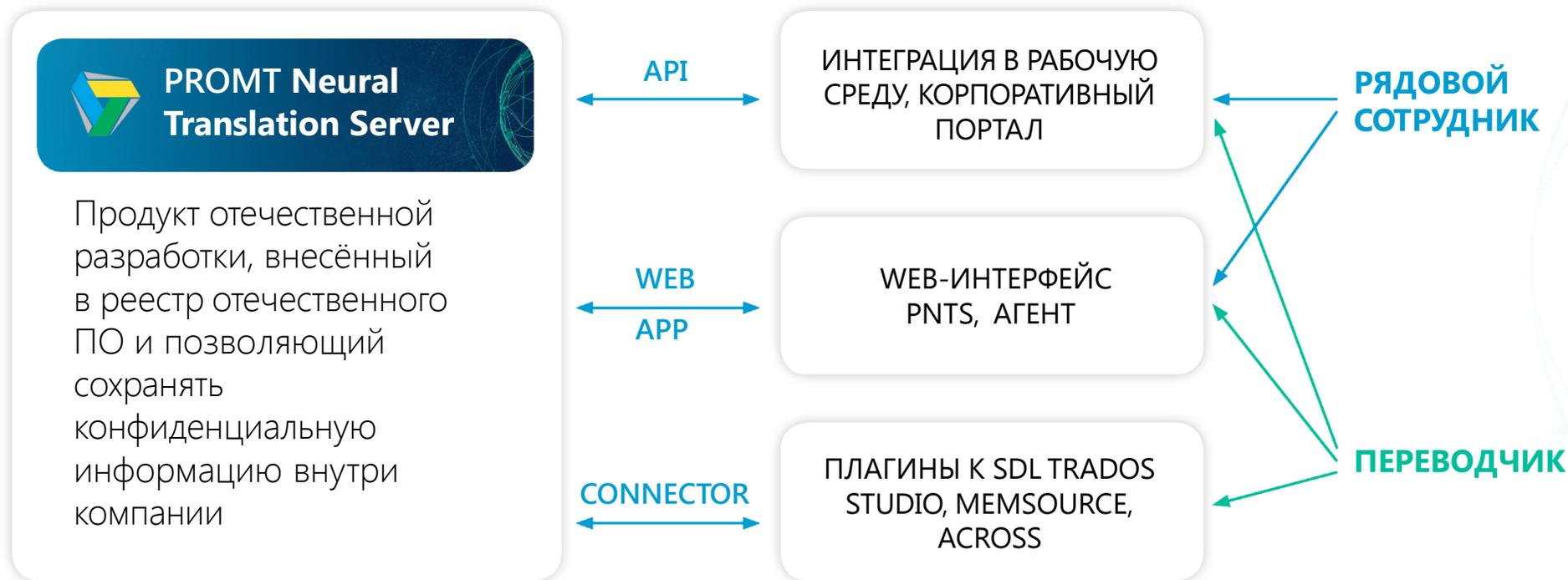
PROMT NEURAL TRANSLATION SERVER

PROMT Neural Translation Server 21

20+ ЛЕТ ПРОДУКТОВОЙ ЖИЗНИ



Варианты использования систем перевода PROMT



Сценарии применения

СЛУЖБЫ ПОДДЕРЖКИ

для оперативного ответа на запросы

E-COMMERCE

локализация контента для покупателей

КОРПОРАТИВНАЯ ПЕРЕПИСКА

общение внутри глобальных компаний

ДОКУМЕНТАЦИЯ К ПРОДУКЦИИ

документация к ПО,
техническая документация

ЛОКАЛИЗАЦИЯ САЙТОВ

перевод контента сайтов на лету

ПОИСК ИНФОРМАЦИИ

поиск информация в мультиязычных БД,
на сервисах

ИНТЕГРАЦИЯ В СЕРВИСЫ И ПРИЛОЖЕНИЯ

как дополнение к основным функциям

ОБРАЗОВАТЕЛЬНЫЕ СЕРВИСЫ

расширение возможностей для сервисов
по изучению иностранных языков



БЕЗОПАСНОСТЬ И МАШИННЫЙ ПЕРЕВОД

1 клик мыши отделяет от утечки информации

УГРОЗА!

Перевод через онлайн-сервис
создает для пользователя
иллюзию безопасности



Онлайн-сервис как канал утечки данных

2017 год. Скандал вокруг известной компании Statoil. Тысячи конфиденциальных документов попали в общий доступ.

Они были переведены американским сервисом онлайн-перевода **translate.com** и проиндексированы поисковыми системами.

Любой загруженный контент будет проанализирован с неочевидными последствиями как для сотрудника, так и для организации, в которой он работает.

бесплатная услуга = предоставление своих данных в свободное пользование сервиса



Условия использования сервисов онлайн-перевода



...не может и не гарантирует, что любая информация, предоставленная нам вами, не станет общедоступной ни при каких обстоятельствах.

Вы должны понимать, что вся информация, представленная на веб-сайте, потенциально может быть общедоступной.



When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content.

Какую информацию можно получить из отрывков текста?

ЧАСТОТНЫЙ И ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ

покажет частоту употребления тех или иных слов в связке с:

- событиями
- именами
- датами
- и любыми другими данными

КОНТЕНТ ФИЛЬТРУЕТСЯ ПО IP-АДРЕСУ И ID УСТРОЙСТВА

АНОНИМАЙЗЕР НЕ ПОВЛИЯЕТ НА РЕЗУЛЬТАТЫ



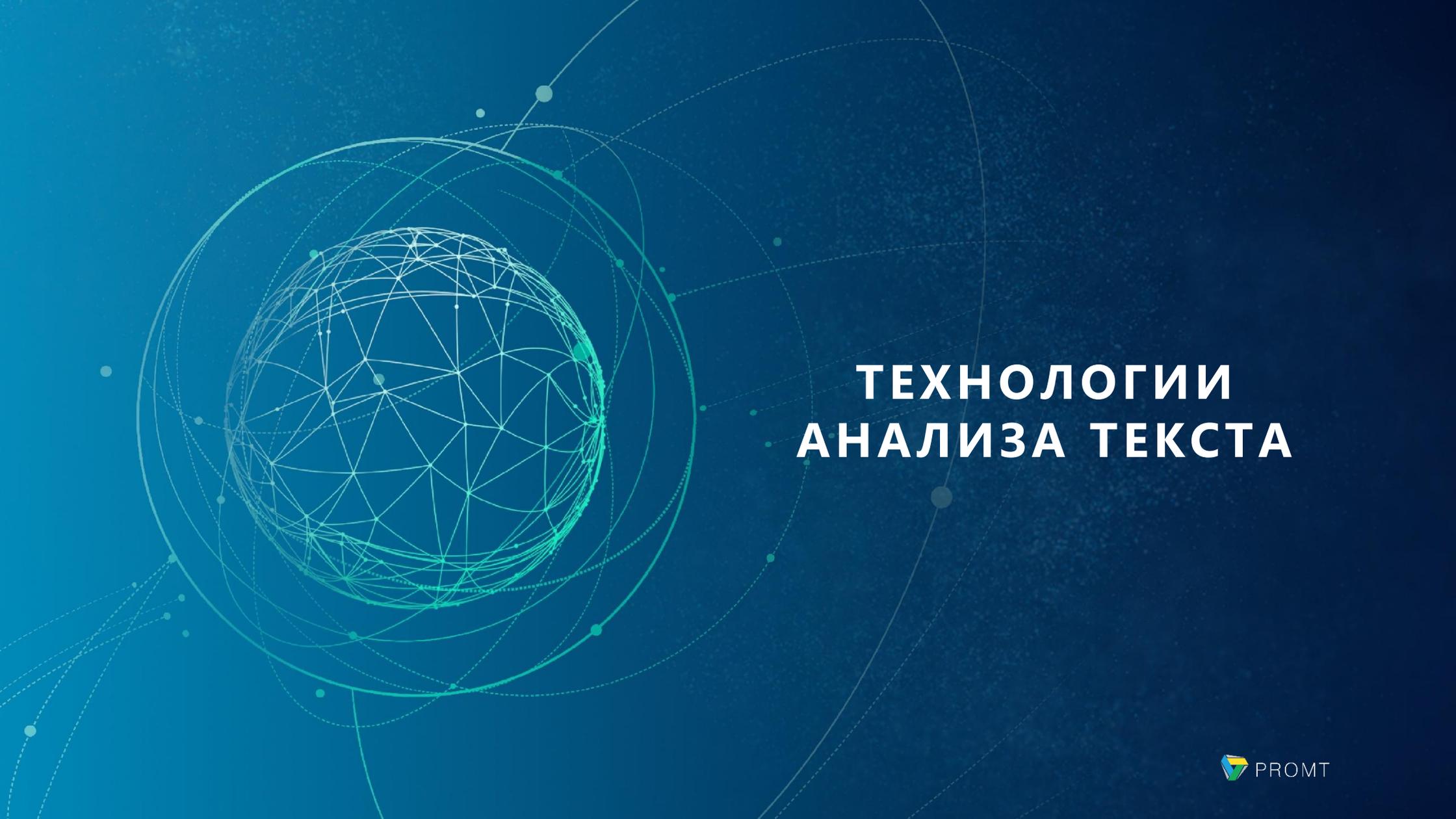
Экспертное мнение

PROMT как компания-эксперт в области технологий перевода хорошо осознает риски, возникающие при использовании общедоступных облачных сервисов при переводе конфиденциальной информации.

100% БЕЗОПАСНОСТЬ
КОНТЕНТА

обеспечивает только внутренний сервис машинного перевода, не соприкасающийся с внешней средой!





ТЕХНОЛОГИИ АНАЛИЗА ТЕКСТА

От Big Data к Smart Data

Анализ неструктурированных данных на разных языках – еще одна задача для искусственного интеллекта

Big Data – ключевой драйвер развития цифровой экономики

Оперативная реакция на изменения возможна только с использованием **Smart Data**



Что используется для аналитики?



Веб-ресурсы: блоги, СМИ, социальные сети



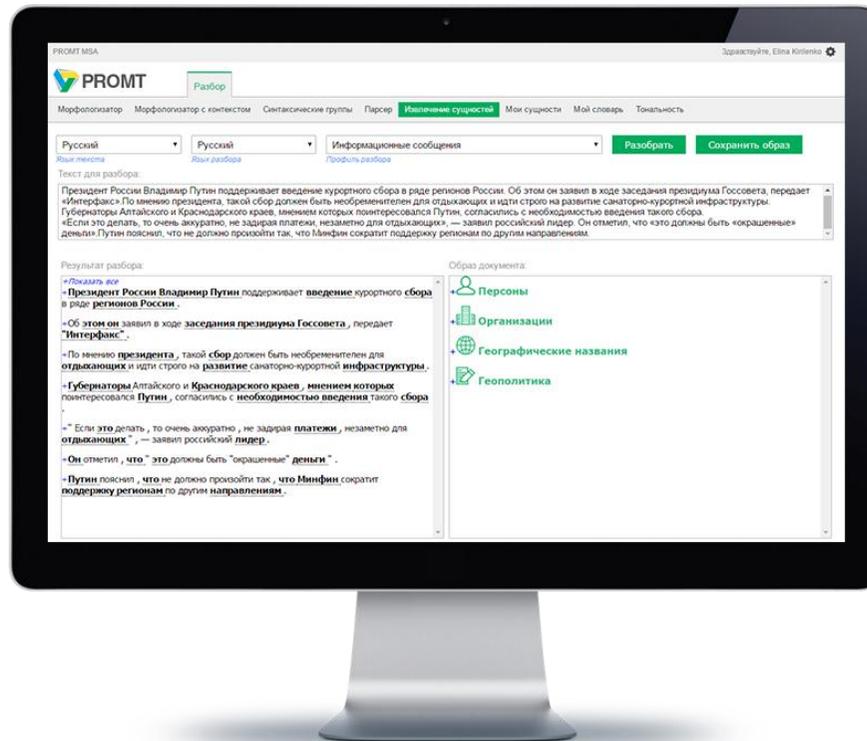
Архивы документов



Договоры, отчеты



Контент, сгенерированный пользователями



Лингвистический подход PROMT к анализу Big Data

- Выделение **сущностей** (в т.ч. именованных) и **фактов** (действий и высказываний) из текстов, в том числе на иностранных языках
- Определение **анафорических связей**, установление отношений между извлеченными сущностями и фактами и их учет при анализе
- Определение **синонимов**
- Определение **тональности** текста, отдельных высказываний, тональной характеристики сущностей

Функциональные возможности PROMT Analyzer Doc API

- извлечение фактографических данных из текстовых документов на естественных языках
- классификация документов
- интеллектуальный поиск

Выделяет ключевые объекты договора, сохраняет структурированную информацию в БД и может быть натренирован на выделение дополнительных объектов.



ТИП
ДОГОВОРА



ДАТА И МЕСТО
ЗАКЛЮЧЕНИЯ
ДОГОВОРА



СТОИМОСТЬ
ДОГОВОРА



ПРЕДМЕТ
ДОГОВОРА



СРОК
ДОГОВОРА



МЕСТО
РАЗРЕШЕНИЯ
СПОРОВ



СТОРОНЫ
ДОГОВОРА

названия, представители
с полномочиями, адреса/реквизиты
и обязательства сторон по договору

Функциональные возможности PROMT Analyzer News API

Позволяет собирать, индексировать и анализировать многоязычный контент из новостных источников

СБОР: важные новости с нужных ресурсов

АНАЛИЗ: выделение основных информационных объектов и автоматическое определение тональности

ПОИСК: база новостей с возможностью интеллектуального поиска по ней



ФАМИЛИИ И
ИМЕНА ЛЮДЕЙ



ГЕОГРАФИЧЕСКИЕ
НАЗВАНИЯ



ОРГАНИЗАЦИИ



АДРЕСА, Е-МАИЛ,
ТЕЛЕФОНЫ



ДАТЫ, ВРЕМЯ



ДЕНЕЖНЫЕ
СУММЫ

Какие задачи позволяет решать?

Docs API



Юридическая экспертиза



Проверка на соответствие



Уменьшение рисков потери капиталов



Предупреждение мошенничества



Поддержка принятия решений



Предупреждение ошибок



Интеллектуальный поиск для выявления договоров с похожими условиями и/или совпадающими данными (контрагенты, суммы....)

News API

Точный анализ информации, полученной от:



Сервисов для мониторинга СМИ и социальных сетей



Новостных агрегаторов и rss-ридеров



Инструментов для сбора данных



Инструментов аналитики СМИ и социальных сетей

Повышение эффективности аналитиков, маркетологов, PR специалистов

Преимущества



СКОРОСТЬ

Высокая скорость работы,
масштабируемость



ОБУЧАЕМОСТЬ

На стороне заказчика



ТОЧНОСТЬ

Точное выделение связей
между объектами



МНОГОЯЗЫЧНОСТЬ

Поддержка русского,
английского и других языков



ГИБКОСТЬ

Тонкая настройка под задачу заказчика

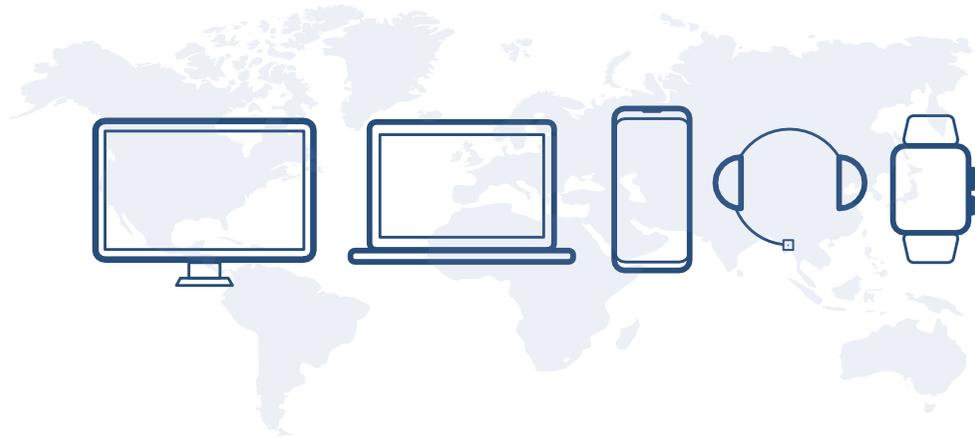


БЕЗОПАСНОСТЬ

Работа в закрытом контуре

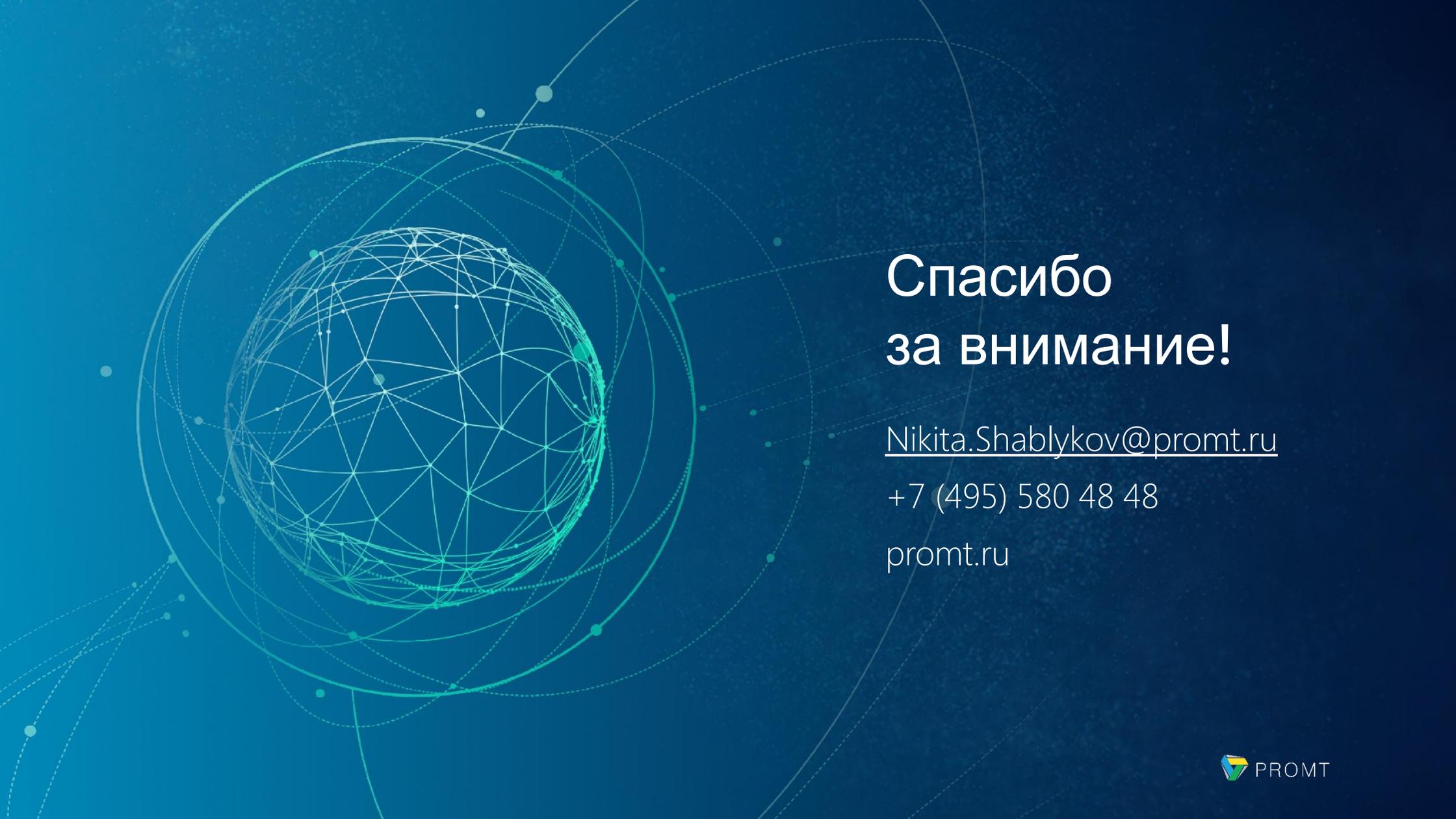
- Призер соревнований по автоматическому извлечению информации из русскоязычных текстов **FactRuEval**
- Финалист конкурса Аналитического центра при правительстве РФ
- Специальная награда конкурса **«Лучшие информационно-аналитические инструменты»**

Технологии **PROMT** работают на разных ОС (**Windows, Linux, iOS, Android, Mac**) и могут быть интегрированы в сторонние программные комплексы благодаря SDK.



Компания поддерживает перевод и анализ текстов для **40+** языков. Также возможна разработка новых языковых пар (сроки от 3 до 6 месяцев).





Спасибо
за внимание!

Nikita.Shablykov@promt.ru

+7 (495) 580 48 48

promt.ru