

Нейросети в юридическом консалтинге

Опыт, проблемы и перспективы



o₂ consulting

Поднимайтесь выше всех
Мы будем рядом

Нейросети – что, как и зачем?

Разберемся в терминологии

- ИИ - это общая концепция (для типа задач)
- Машинное обучение – это способ обучения
- Нейросети - это инструменты ML, имитирующие функции человеческого мозга
- Большие языковые модели (LLM), такие как Gemini, ChatGPT, - это конкретные реализации нейросетей (трансформеров), обученные на огромных объемах текстовых данных

Не LLM: Google Cloud Vision, Google Speech-to-Text/Text-to-Speech

Мультимодальные LLM – интеграция нескольких видов нейросетей

Возможности и вызовы

Внедрение нейросетей в юриспруденцию открывает перспективы для автоматизации рутинных задач, улучшения качества анализа данных и прогнозирования, а также для разработки новых подходов в правоприменении.

Однако это также влечёт за собой необходимость адаптации юристов к новым технологиям, этические вопросы и риски, связанные с точностью и надёжностью получаемых результатов.

Ограничения технологии

- Зависимость от качества и количества данных
- Неспособность объяснить свои решения (проблема «черного ящика»)
- Отсутствие критического мышления
- Неспособность учитывать контекст в полной мере. Риск галлюцинаций
- Отсутствие «здорового смысла»



«Они среди нас!» - 1

NLP (Natural Language Processing) – это область искусственного интеллекта, которая занимается анализом, пониманием и генерацией человеческого языка. *LLM, такие как GPT-4, Claude, Gemini, представляют собой наиболее продвинутый класс NLP-моделей, обученных на огромных массивах текстовых данных.*

Машинный перевод (старый-добрый), чат-боты поддержки, анализ тональности текста. Системы ЭДО. Справочно-правовые системы!

Системы распознавания лиц, номерных знаков автомобилей.

Системы анализа сканов документа (сверточные нейросети) позволяют сортировать входящие/исходящие документы, выявить следы подчистки, различия в чернилах, признаки монтажа.

Автоэнкодеры (поиск аномалий) – поиск отклонений в типовых документах. Он «учится» сжимать текст в компактное векторное представление и восстанавливать его с минимальными потерями. Если есть отклонение – выдаст ошибку восстановления (сигнал юристу).

Системы оптического распознавания символов (OCR) эффективно распознают текст на изображениях (сканы документов, фотографии, скриншоты). Это позволяет автоматизировать ввод данных, оцифровывать архивы и интегрировать информацию из различных источников.

Всеми любимый OCR в продуктах АБВУУ

** OCR используется в LLM (интегрируется в архитектуру)*



«Они среди нас!» - 2

Тип нейросети	Основные задачи в юридическом консалтинге	Примечания
Сверточные нейронные сети (CNN)	Анализ изображений документов: классификация по типу (скан паспорта, договора, счета), выявление графических элементов (печати, подписи), поиск аномалий (подчистки, исправления). CNN лучше всего подходят для работы с изображениями.	LLM тут не нужны. Пример: Vision Labs. <i>Не видят текст, смотрят только на изображение</i>
Рекуррентные нейронные сети (RNN, LSTM, GRU)	Обработка последовательностей текста: анализ динамики изменений в договорах (сравнение версий), выявление паттернов в судебных решениях (например, последовательность аргументов, приводящих к определенному решению). Распознавание речи (аудиозаписи судебных заседаний)	<i>Для сложного анализа текста сейчас чаще используются трансформеры или гибридные системы.</i> Yandex SpeechKit, Tinkoff VoiceKit, Descript - гибридные системы с использованием RNN)
Трансформеры (не LLM)	Распознавание объектов на изображениях (включая текст – OCR), Обработка последовательностей (альтернатива RNN) Созданы для обработки последовательностей.	Различные реализации OCR на основе Transformer (вместо связки CNN+RNN). Пример: Whisper
Полносвязные нейронные сети (FNN/MLP)	Классификация документов по структурированным признакам (тип, предмет, риск – если эти признаки уже извлечены). Прогнозная аналитика на основе структурированных данных (вероятность просрочки платежа, сроки типовых задач)	Внутренние разработки юридических фирм и компаний (часто – не отдельные продукты, а модули в CRM, ERP, СЭД). <i>Нужны табличные данные</i>
Автоэнкодеры (Autoencoders)	Выявление аномалий в текстах договоров (нестандартные формулировки), в транзакциях (признаки мошенничества).	DataRobot, H2O.ai, внутренние разработки компаний. <i>Просто сжимают и восстанавливают данные</i>
Большие языковые модели (LLM)	Анализ текста: извлечение информации, суммаризация, ответы на вопросы, поиск по смыслу. Генерация текста: составление проектов документов. Машинный перевод.	Популяризировали технологию

Важно понимать, что это не исчерпывающий список. Есть много внутренних разработок компаний, которые не афишируются.

Кроме того, многие системы являются гибридными, то есть используют комбинацию разных типов нейросетей и других алгоритмов.

Разрушаем мифы о магии LLM



1. LLM вычисляет вероятность того или иного ответа и выбирает вариант, который статистически чаще соответствует подобным запросам в её обучающих данных. Это упрощение – но разрушающее иллюзии.
2. Данные попадают в LLM после обработки (токенизация и эмбединги) и выходные данные также постобрабатываются

LLM не заменят юристов, а станут их помощниками

LLM не захватят мир

LLM не обладают магией



Работа с LLM

Как правильно работать с LLM?

Правильное использование LLM требует базового понимания концепции промптинга.

Промптинг (prompt engineering) – это искусство и наука формулирования запросов (промптов) к LLM таким образом, чтобы получить желаемый ответ. От качества промпта напрямую зависит качество результата.

Поскольку LLM генерирует ответ, базируясь на вероятностных распределениях, структура промпта всегда начинается с вводных данных по задаче и содержит указание на решаемую задачу.

«Как спросишь, так и ответят»



Промпты для LLM

Опрос по результатам и ссылка на
скачивание методички по промптингу тут ->
И можно поставить оценку 😊



Создание промптов – 1

1. Укажите роль для ИИ. Например, «юрист по IP».
2. Чётко сформулируйте задачу для ИИ. Например: «Проанализируй договор на предмет наличия условий о неустойке».
! Разбейте комплексные задачи на подзадачи.
! Используйте ясные и емкие формулировки.
3. Приведите релевантные факты. Даты и т.д.
4. Просите ИИ сделать самопромтинг. Например, «подумай, как правильно было бы сформулировать структуру промпта для ответа».
5. Укажите формат результата. Например: «Составь таблицу с основными проблемами, отсортированными по уровню риска для стороны».
6. Попросите ИИ оценить свой ответ по шкале и ответить снова, если ответ не «идеален».
7. Не останавливайтесь на первой итерации.



Создание промптов - 2

ВАЖНО: вся информация устаревает очень быстро, т.к. технология очень быстро развивается. Используйте всю информацию критически!

Не стоит нагружать передовые модели типа Open AI o1 чрезвычайно сложными примерами и промптами с цепочкой рассуждений (Chain-of-Thoughts), лучше начать с простого.

А вот для простых и «старых» или опенсорсных моделей детальный промпт может быть решением (поскольку предусматривает эту «цепочку рассуждений»).

Если у вас есть сомнение, стоит ли покупать подписку на модели типа ChatGPT, Claude (вне рамок базовой подписки), **развернутый промпт может помочь потестировать возможности бесплатных подписок.**



Примеры простейших промптов

Ниже иллюстративные примеры:

- **Анализ договора:** "Проанализируй этот договор аренды и выяви все риски для арендатора".
- **Составление документа:** "Составь правила внутреннего трудового распорядка".
- **Написание делового письма:** «Переформулируй текст в деловой манере коммерческого предложения».
- **Поиск информации:** "Найди в Интернете ключевую судебную практику по вопросам новации обязательств".
- **Подготовка к встрече с клиентом:** "Напиши список вопросов, которые нужно задать клиенту, чтобы понять его проект в сфере толлинга".
- **Оценка перспектив дела:** «Проанализируй фабулу и предположи, какое решение должен вынести судья".
- **Фактчекинг:** "Проверь, имеются ли противоречия в изложенном документе со следующими фактами, известными мне достоверно".
- **Толкование положения договора:** «Разъясни значение данной формулировки в документе в общем контексте".
- **Улучшение текста:** "Изложи этот юридический текст без воды и канцеляризмов".
- **Составление таймлайна:** "Проанализируй текст и составь хронологию событий".

Какой из промптов кажется вам неудачным и почему?

Кейсы – проектная работа

№	Кейс	Уровень сложности
1	<u>Интерпретация норм права</u>	Легкий
2	Составление юридических писем (Автоматизация деловой переписки)	Легкий
3	<u>Анализ договорных рисков (Экспертиза договоров)</u>	Средний
4	<u>Создание краткого содержания (Саммаризация юридической информации)</u>	Легкий
5	Юридический редактор (Проверка и коррекция юридических документов)	Средний
6	Составление проектов договоров (Автоматизация договорной работы)	Сложный
7	Интерпретатор (Перевод с юридического на «человеческий» язык)	Средний
8	Прогнозирование исхода судебного спора (Анализ судебной практики)	Сложный
9	<u>Удаление слов-паразитов (Редактирование текста)</u>	Легкий
10	Анализ переписки (Извлечение информации из неструктурированных данных)	Очень сложный
11	Изучение отраслевой специфики (Погружение в контекст)	Средний
12	<u>Анализ судебного акта (Извлечение фактов и позиций сторон)</u>	Легкий
13	Создание схем процессов (Визуализация информации)	Средний
14	Аналитика данных (Преобразование данных в выводы)	Средний
15	Интеллектуальный поиск по юридической базе (Семантический поиск)	Средний/Сложный
16	Автоматическая классификация юридических документов (Рубрикация)	Легкий/Средний
17	Выявление аномалий в юридических документах (Детектор аномалий)	Средний/Сложный
18	Сравнение версий юридических документов (Автоматическое сравнение)	Легкий/Средний
19	Генерация ответов на типовые юридические вопросы (Юридический FAQ)	Средний
20	Сравнение версий НПА	Средний

Пример кейса

Опрос по результатам и ссылка на
скачивание методички по промптингу тут ->
И можно поставить оценку 😊



Кейсы – непроектная работа

№	Кейс для непроектных задач	Уровень сложности
1	<u>Создание внутренних документов (Разработка ЛНА и инструкций)</u>	Средний
2	<u>Описание бизнес-процессов (Автоматизация и оптимизация)</u>	Сложный
3	Метапромтинг (Создание инструкций для работы с ИИ)	Сложный
4	<u>Структурирование выступлений (Подготовка к публичным выступлениям)</u>	Средний
5	Эмоциональный анализ текста (Оценка тональности)	Средний
6	<u>Low-code / No-code программирование</u>	Очень сложный
7	Разработка методических материалов (Создание базы знаний)	Сложный
8	Создание новостных сводок (Мониторинг изменений)	Легкий
9.	Генерация идей	Легкий
10.	Написание статей и аннотаций	Средний
11	Планирование и управление задачами	Легкий/Средний
12	Контент-план для юриста/фирмы	Легкий
13	Рерайтинг юридического текста (Перефразирование)	Легкий/Средний
14	Подготовка к юридическим переговорам (Моделирование ситуации)	Сложный
15	Обучение и развитие (создание планов, подбор материалов)	Средний
16	Генерация изображений по текстовому описанию	Средний
17	<u>Транскрибация (дословная расшифровка) аудио и видео</u>	Легкий

А судьи кто?

Разделение кейсов на уровни сложности является субъективным мнением (оценочным суждением) и может не совпадать с мнением других пользователей. Какого-либо четкого алгоритма подсчета баллов выработать пока что не получилось, оценка производилась опытным путем.

Был проведен эксперимент с Gemini 2.0 Pro Experimental (февраль 2025 г.), в результате которого получились следующие критерии.

Уровень сложности каждого кейса оценивается на основе комбинации следующих факторов:

- **Теоретическая сложность:** Насколько сложна сама по себе задача для юриста, даже без использования нейросетей? Требует ли она глубоких знаний в узкой области права, большого опыта, умения анализировать большие объемы информации, учитывать множество нюансов?
- **Практическая сложность:** Насколько трудоемка задача? Сколько времени и усилий обычно тратит юрист на ее решение?
- **Степень неопределенности:** Насколько однозначен результат? Есть ли «правильный» ответ, или возможны разные варианты?
- **Ответственность:** Насколько высока цена ошибки? Какие риски и последствия могут возникнуть, если задача решена неправильно?
- **Наличие готовых решений:** Существуют ли уже готовые инструменты/плагины/расширения на основе нейросетей, специально предназначенные для решения этой задачи?
- **Качество генерации текста:** Насколько хорошо нейросеть справляется с генерацией текста, необходимого для решения задачи (формулировки, юридическая точность, стиль)?
- **Понимание контекста:** Насколько хорошо нейросеть понимает контекст юридической задачи, учитывает нюансы, распознает скрытые смыслы?
- **Необходимость проверки человеком:** Насколько обязательна проверка и корректировка результата работы нейросети человеком? Можно ли полностью полагаться на нейросеть?

Ограничения и риски



ИИ-гигиена и NDA

Юридическая работа неразрывно связана с обработкой конфиденциальной информации и персональных данных, подпадающих под действие соглашений о неразглашении (NDA).

Загрузка таких данных в сервисы ИИ создает риск утечки, компрометации или неправомерного использования.

Для обеспечения безопасности при использовании технологий необходимо применять специализированные меры защиты информации.

При наличии возможности любое применение нелокальных LLM следует согласовывать с клиентом.

«Очистка» - удаление ПДн и замену их на обобщённые обозначения или «синтетические» данные, что помогает предотвратить идентификацию ситуации или клиента.

Удалению подлежит и любая чувствительная по содержанию информация.

Использование временных сессий для обработки данных, которые автоматически удаляются после завершения работы, проверка настроек чата. Запрет на обучение LLM на ваших данных в настройках аккаунта.

Развертывание ИИ-моделей локально, чтобы обеспечить полный контроль над данными.



Галлюцинации LLM

«Галлюцинации» LLM могут привести к серьёзным ошибкам, поэтому необходимо тщательно перепроверять все результаты, полученные от LLM.

Чтобы минимизировать риск ошибок, не следует использовать LLM без достаточного наполнения модели контекстом.

LLM может генерировать неточную, вымышленную или вводящую в заблуждение информацию, что связано с неполнотой, устареванием или противоречивостью данных в базе.



Доверие клиентов и использование ИИ



Использование искусственного интеллекта в юридической практике можно сравнить с применением других инструментов, таких как справочно-правовые системы. Юрист не обязан подробно информировать клиента о том, какие технологии он использует в работе. Но использование ИИ – не такой случай.

ВАЖНО: Клиент должен иметь возможность самостоятельно принять решение о том, согласен ли он на использование искусственного интеллекта в своём деле. Это позволит ему сделать осознанный выбор и почувствовать контроль над ситуацией.

Ответственность юриста и использование ИИ

Юрист несёт полную ответственность за результаты своей работы, и это включает в себя учёт возможных предубеждений клиентов относительно использования ИИ, а также принятие ответственности за соблюдение NDA.

Решение об информировании клиента о применении ИИ должно зависеть от обстоятельств дела, характера взаимоотношений с клиентом и этических соображений. В будущем это может повлиять и на ценообразование услуг.

Важно объяснить, как именно эта технология используется, какие преимущества она даёт и какие риски может снизить.



ТГ-канал по праву->



Связаться со мной->



Вопросы?



Софья Смирнова,
советник корпоративной практики
O2 Consulting

Опрос по результатам и ссылка на
скачивание методички тут ->



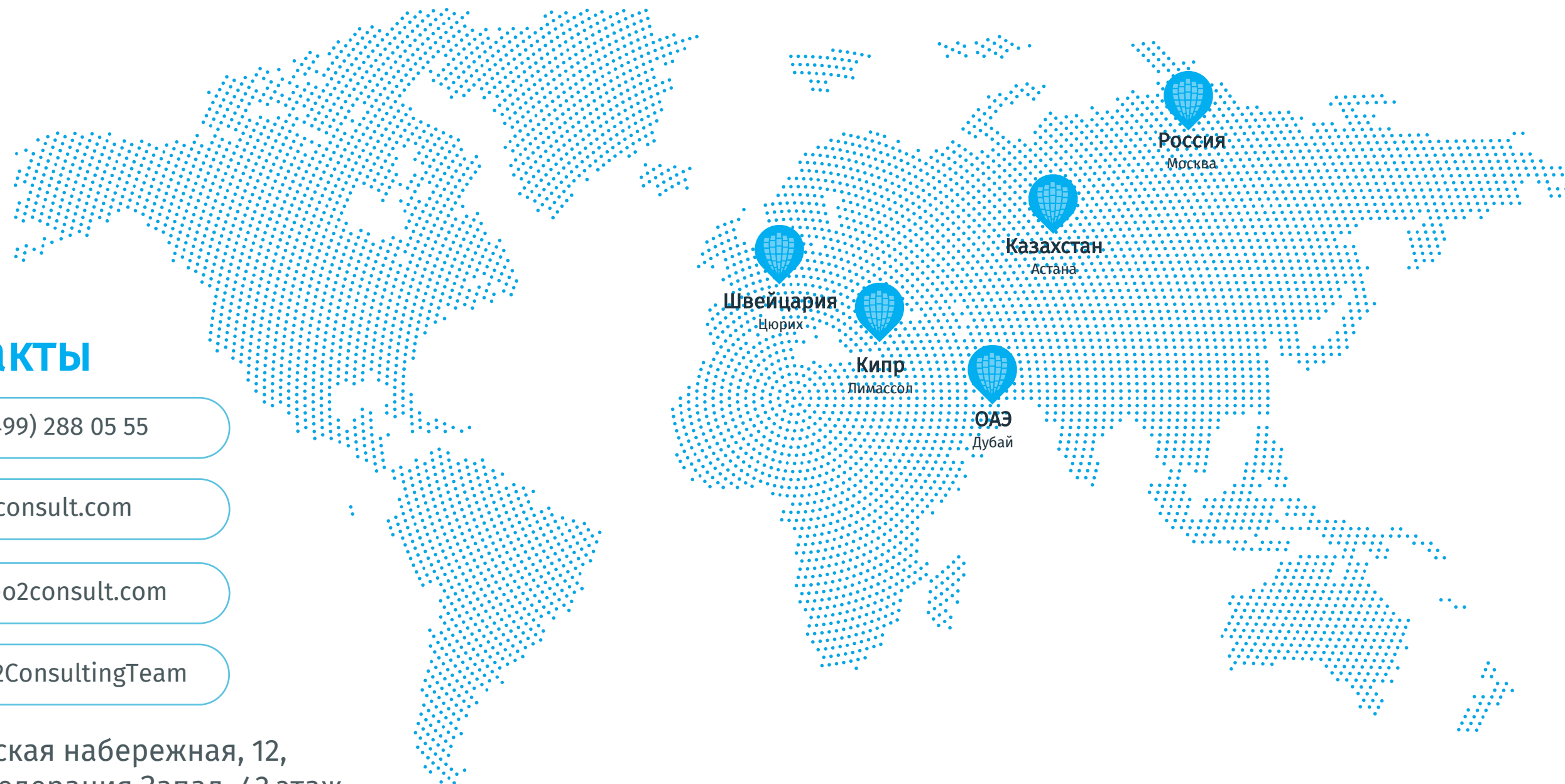
Россия – Москва

ОАЭ – Дубай

Казахстан – Астана

Швейцария – Цюрих

Кипр – Лимассол



Контакты

+7 (499) 288 05 55

o2consult.com

info@o2consult.com

t.me/O2ConsultingTeam

Пресненская набережная, 12,
Башня Федерация Запад, 43 этаж,
Москва, Россия, 123112