Yandex Cloud

Проекты машинного обучения: проблемы и решения

Machine Learning

Алёна Дробышевская

01

Основные тренды в ML & AI

ML-решения — основа ИТ завтрашнего дня

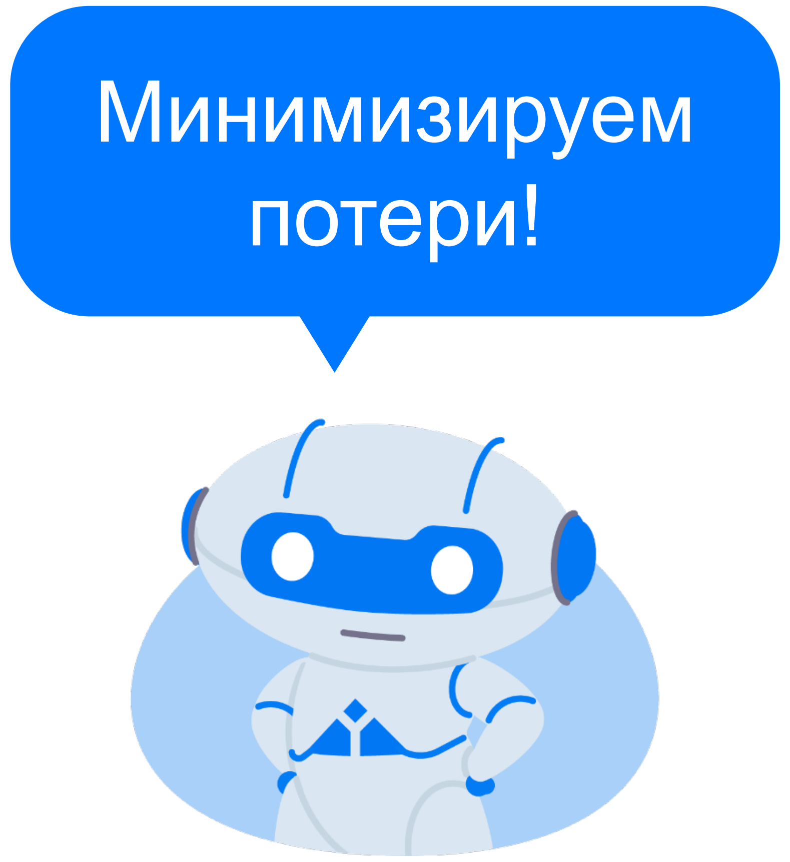
Неизбежный ИИ:

- К 2025 году по меньшей мере 90% новых корпоративных приложений будут встроены в ИИ
- К 2024 году более 50% взаимодействий с пользовательским интерфейсом будут использовать компьютерное зрение с поддержкой ИИ, речь, обработку естественного языка и AR/VR

IDC FutureScape: Worldwide IT Industry 2020 Predictions

Рынок ML растёт, и даже COVID-19 стал фактором роста

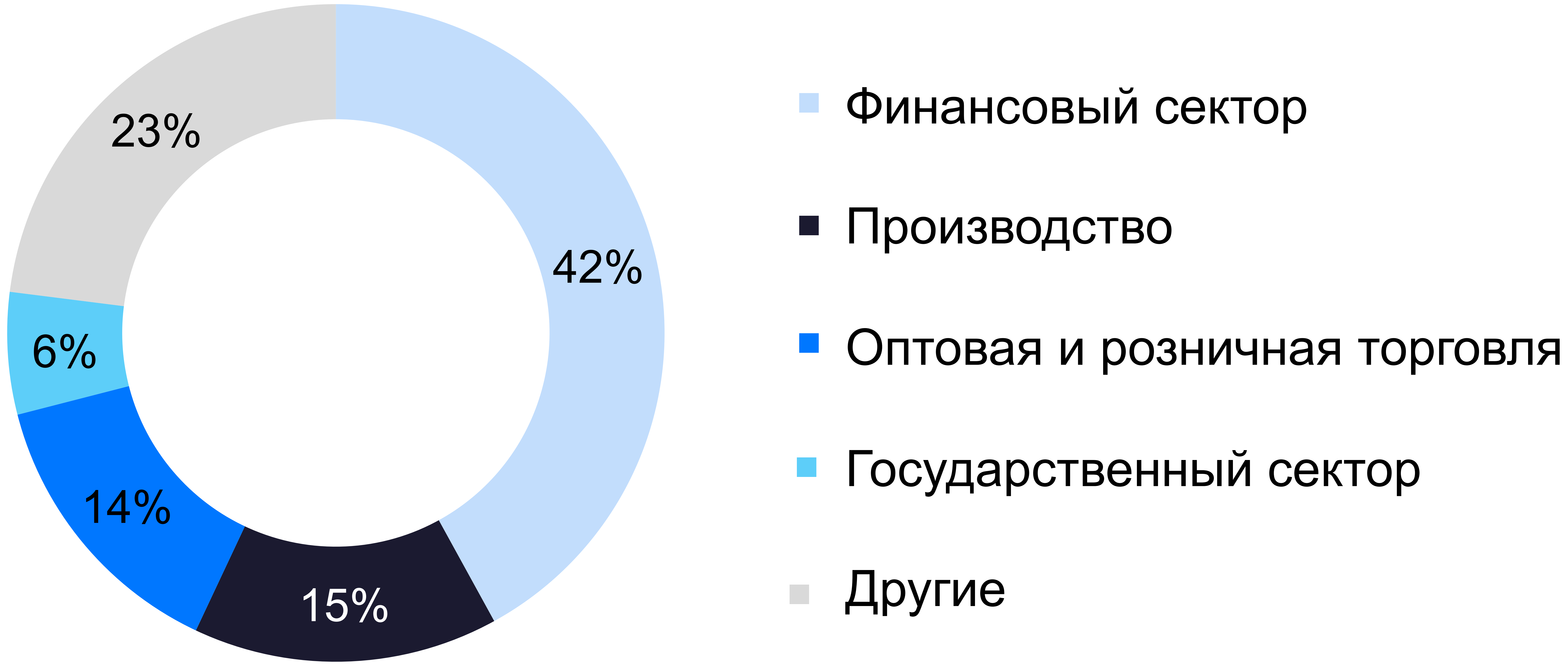
- Глобальные расходы на искусственный интеллект (ИИ), по прогнозам IDC, удвоятся в течение следующих четырёх лет, увеличившись с $50,1 млрд в 2020 году до более чем $110 млрд в 2024 году
- Среднегодовой рост в период до 2024 года составит 20,1%
- В конце марта IDC опросили 180+ компаний в Европе, и 16% респондентов заявили, что видят в автоматизации при помощи искусственного интеллекта и новых технологий средство для минимизации последствий пандемии



IDC: What is the Impact of COVID-19 on the European IT Market? (IDC #EUR146175020, April 2020)
IDC Worldwide Artificial Intelligence Spending Guide 2020

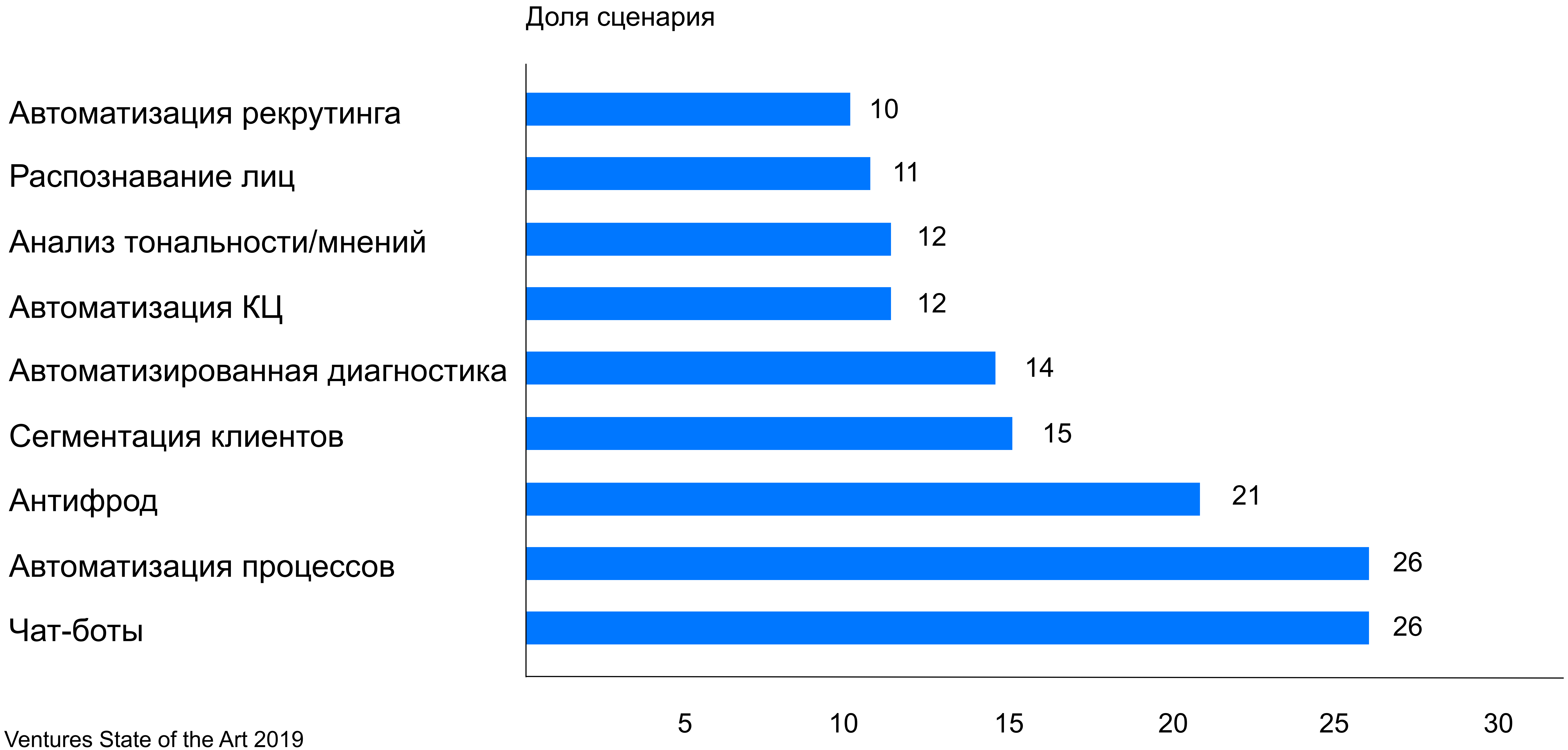
Наиболее активные отрасли

Доля индустрии в структуре расходов



Проникновение ML в индустрию

Одна из десяти крупных компаний использует не менее десяти приложений ML/AI

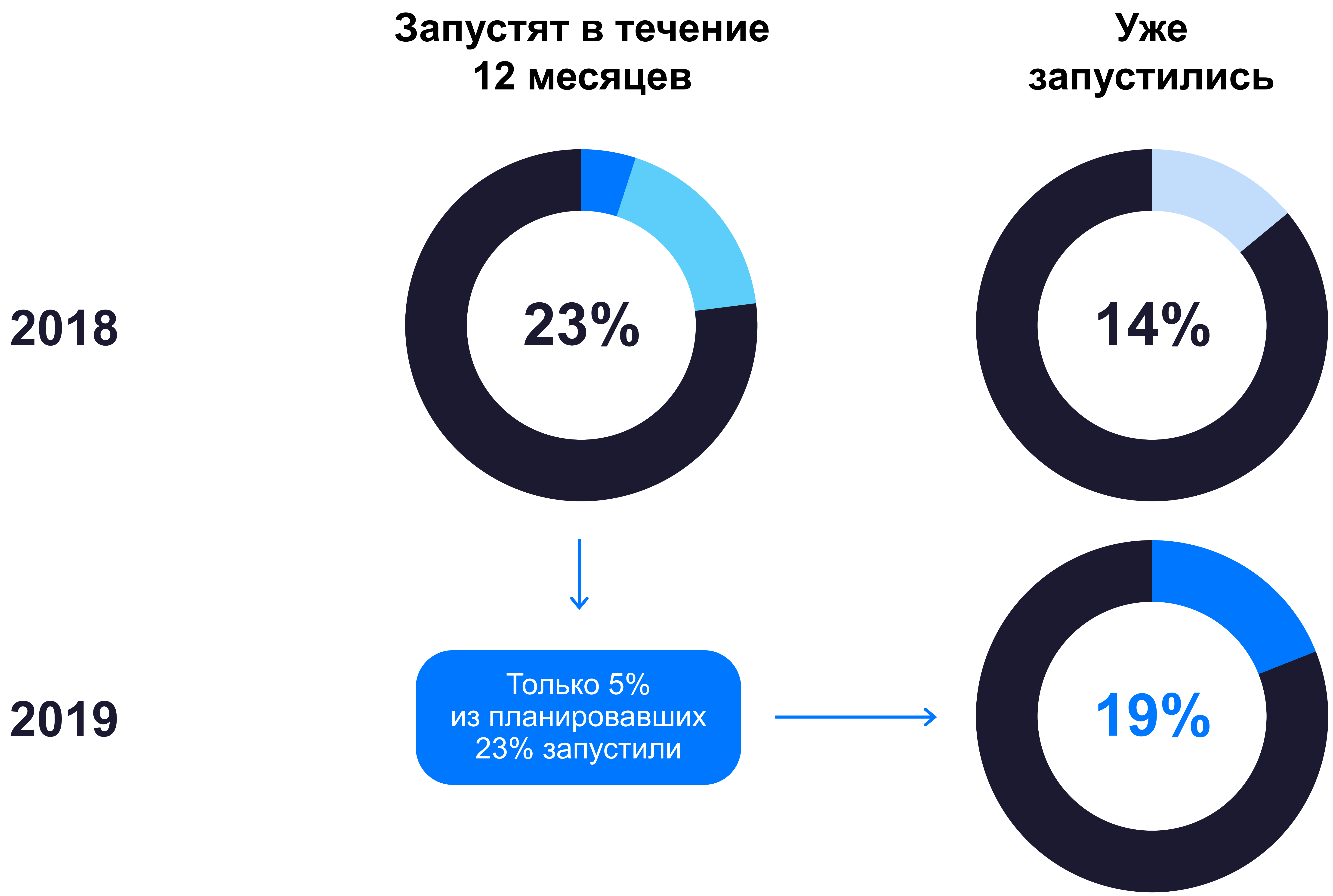


MMC Ventures State of the Art 2019
Gartner, 2019 CIO Survey: CIOs Have Awoken to the Importance of AI, figure 1, 3 January 2019

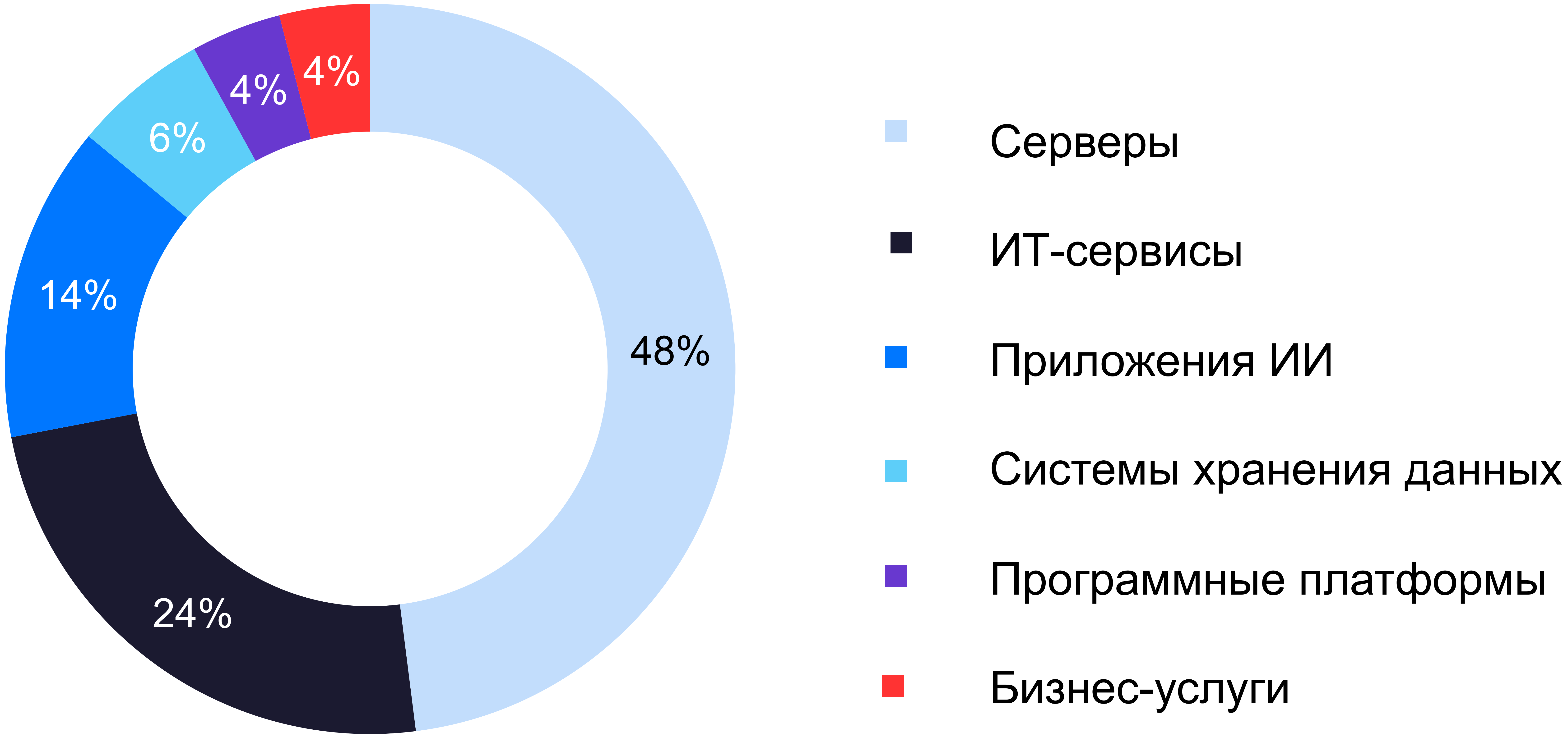
02

Основные сложности внедрения ML-технологий

Рынок ML-технологий



Статьи расходов на ML-рынке России



IDC Worldwide Artificial Intelligence Spending Guide 2020

Наблюдаемые сложности внедрения ML

- Отсутствие ИТ-инфраструктуры
- Недостаток экспертных знаний в ML
- Нехватка бюджетов

Кроме затрат, есть риски

При переходе из прототипа в продакшн многое может пойти «не так»

- Данные в источниках не соответствуют тем, что были на этапе прототипирования
- Метрики оказались неэффективны
- Цели изменились и требуют перестройки всего проекта
- Реализованная модель не соответствует прототипу

Эксперимент — это только начало пути!

STAND BACK!

I'M RUNNING
MACHINE LEARNING
EXPERIMENTS!

Ряд проблем возникает при переводе решений в промышленную эксплуатацию

- Масштабирование решения Data Science на полном объеме данных
- Контуры эксперимента и промышленной эксплуатации различаются
- Использование разнородных сред разработки (Python, R, Java и др.)
- Ограничения инструментов

Business Broadway «Top 10 Challenges to Practicing Data Science at Work»

Требование к окружению прототипирования

У программиста на компьютере всё работает!

- Максимальная близость данных эксперимента к реальным

  ⟶ Сделать выборку данных частью исследования

- Обеспечение консистентности модели

  ⟶ Переиспользование кода / автоматизация тестирования

- Полная история исследования и его повторяемость

  ⟶ Сквозное версионирование всех этапов исследования

03

Что может предложить Yandex.Cloud

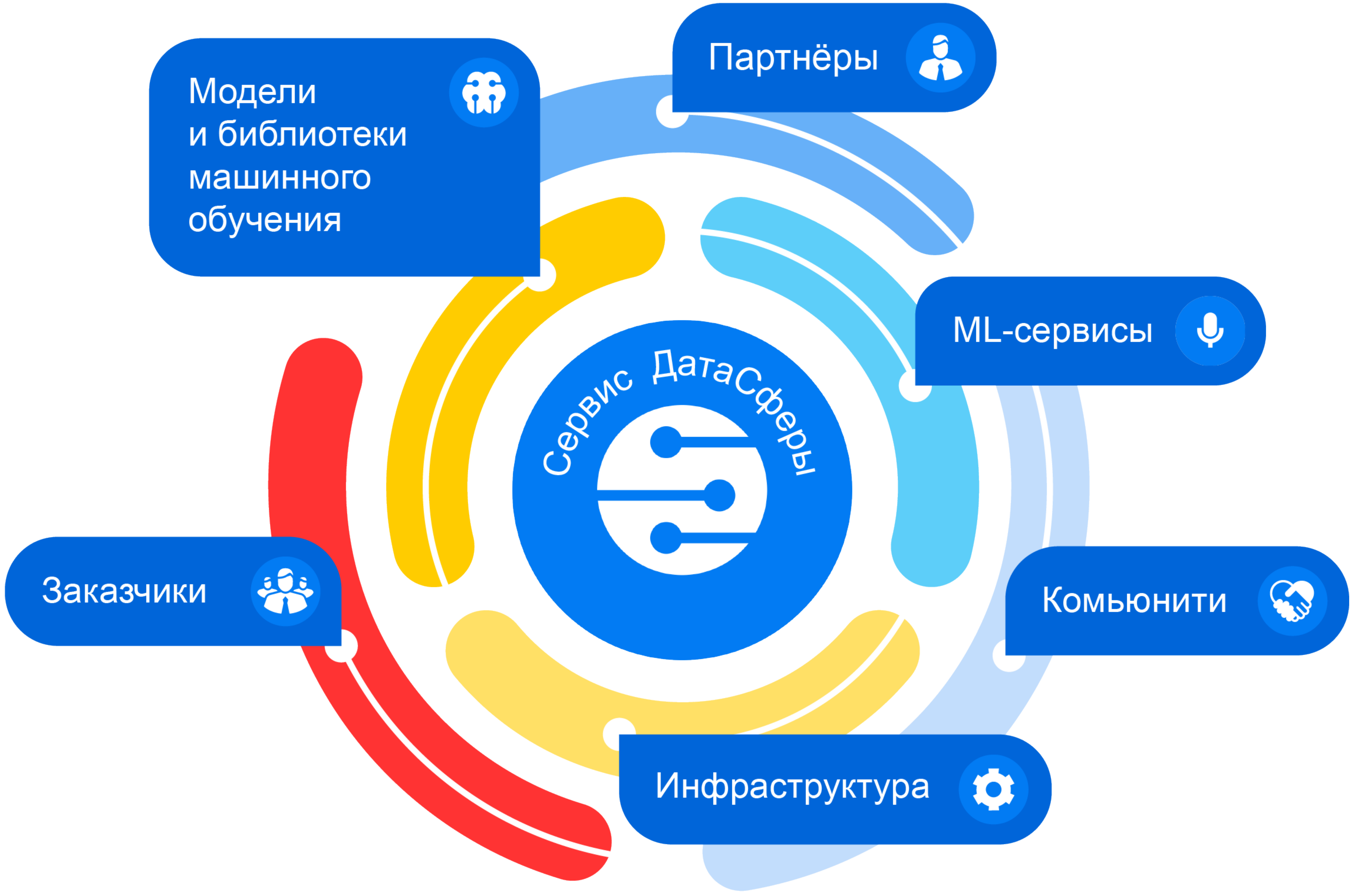
Основные направления работы ML-команды Yandex.Cloud

- Голосовые технологии
- Машинный перевод
- Компьютерное зрение
- Инструменты разработки машинного обучения

ML и Data Science в Яндексе — почему мы

- Яндекс.Поиск — по сути начало промышленного Data Science в России
- Яндекс — «инкубатор» Data Science с большим опытом и большой командой
- Яндекс стал готовить ресурсы не только для себя, но и для рынка: Школа анализа данных, курсы Яндекса на Coursera, сотрудничество с ведущими вузами
- Yandex.Cloud дал возможность делиться своими наработками и экспертными знаниями не только в формате знаний и курсов, но и в формате технологий
- Yandex DataSphere — среда для ML-разработки в Yandex.Cloud, мы делимся нашим опытом, нашей экспертизой, нашими технологическими наработками

Yandex DataSphere — интегрированная среда для облачной ML-разработки «всё в одном»



Yandex.Cloud для внедрения ML

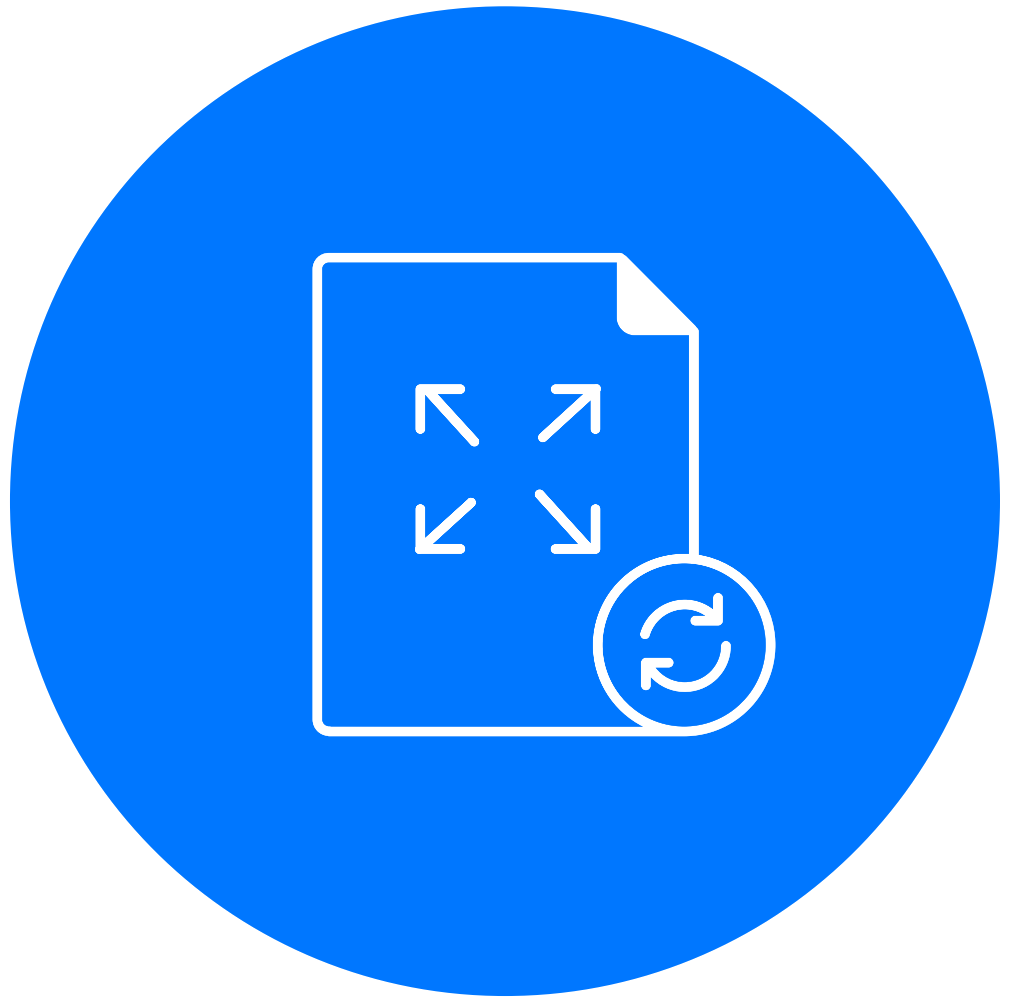
Новые начинания проще делать в облаках

- Облачная инфраструктура позволяет обойтись без закупок специализированного железа для ML
- DataSphere существенно понижает порог входа и стоимость прототипирования
- DataSphere существенно упрощает переход от прототипа в эксплуатацию
- Мы создаём экосистему партнёров для упрощения поиска экспертов в ML

Yandex DataSphere – интегрированная среда для разработки ML "все в одном"

- Serverless computing на мощностях Яндекс.Облака
- Среда исследования и разработки: привычный интерфейс Jupyter Notebook
- Бесшовное переключение разных типов ресурсов прямо во время работы, без перезапуска проекта
- Распределенное исполнение (SPARK, Нирвана)
- Сквозное версионирование хода разработки
- Интеграция с сетью пользователя и единое управление правами доступа
- Inference и регулярное исполнение

DataSphere: среда разработки и эксплуатации



Возможность смешивать продуктовый и экспериментальный контуры

- Интеграция источников данных и стандартного инструментария
- Возможность использовать код прототипа в эксплуатации
- Оптимизация производительности решения «по частям»
- Повторное использование разработанного инструментария в разных проектах

Уменьшение расходов

ML-разработка в распределённой разделяемой среде

- Готовая среда для начала разработки на базе актуальных технологий
- Оплата только потреблённых вычислений
- Бесшовное масштабирование потребляемых ресурсов
- Возможность откатиться к любому этапу исследования

Разнообразие доступных инструментов

Оптимальный инструмент для решения каждого этапа исследования

- Смена среды исполнения без потери состояния:
  - Отключенное состояние
  - Одиночные машины различной конфигурации: S, L (GPU v100), XL (4x GPU v100)
  - Data Proc (Spark)
  - Nirvana
- Единое версионируемое состояние

Yandex Cloud

Спасибо!

Алёна Дробышевская

Руководитель направления по развитию бизнеса ML-сервисов

alenadrob@yandex-team.ru

+7 (985) 761-26-24