

КРУГЛЫЙ СТОЛ «Вызовы, риски и проблемы системы образования РФ
в среднесрочной перспективе»
19 декабря 2023 года

О подготовке кадров для цифровой экономики

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
зав. лаб. Машинного обучения и семантического анализа
Института Искусственного Интеллекта МГУ,
зав. кафедрой Математических методов прогнозирования ВМК МГУ,
зав. кафедрой Машинного обучения и цифровой гуманитаристики МФТИ,
зав. кафедрой Интеллектуальных систем МФТИ,
г.н.с. ФИЦ «Информатика и управление» РАН

Национальная стратегия развития ИИ в РФ

10 октября 2019 г.



УКАЗ

ПРЕЗИДЕНТА РОССИЙСКОЙ ФЕДЕРАЦИИ

**О развитии искусственного интеллекта
в Российской Федерации**

В целях обеспечения ускоренного развития искусственного интеллекта в Российской Федерации, проведения научных исследований в области искусственного интеллекта, повышения доступности информации и вычислительных ресурсов для пользователей, совершенствования системы подготовки кадров в этой области **п о с т а н о в л я ю**:

1. Утвердить прилагаемую Национальную стратегию развития искусственного интеллекта на период до 2030 года.

2. Правительству Российской Федерации:

а) до 15 декабря 2019 г. обеспечить внесение изменений в национальную программу "Цифровая экономика Российской Федерации", в том числе разработать и утвердить федеральный проект "Искусственный интеллект";

б) представлять Президенту Российской Федерации ежегодно доклад о ходе реализации Национальной стратегии развития искусственного интеллекта на период до 2030 года;

в) предусматривать при формировании в 2020 - 2030 годах проектов федеральных бюджетов на очередной финансовый год и на плановый период бюджетные ассигнования на реализацию настоящего Указа.

УТВЕРЖДЕНА
Указом Президента
Российской Федерации
от 10 октября 2019 г. № 490

**НАЦИОНАЛЬНАЯ СТРАТЕГИЯ
развития искусственного интеллекта
на период до 2030 года**

I. Общие положения

1. Настоящей Стратегией определяются цели и основные задачи развития искусственного интеллекта в Российской Федерации, а также меры, направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технологического развития.

2. Правовую основу настоящей Стратегии составляют Конституция Российской Федерации, Федеральный закон от 28 июня 2014 г. № 172-ФЗ "О стратегическом планировании в Российской Федерации", указы Президента Российской Федерации от 7 мая 2018 г. № 204 "О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года", от 9 мая 2017 г. № 203 "О Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы", от 1 декабря 2016 г. № 642 "О Стратегии научно-технологического развития Российской Федерации" и иные нормативные правовые акты Российской Федерации, определяющие направления применения информационных технологий в Российской Федерации.

3. Настоящая Стратегия является основой для разработки (корректировки) государственных программ Российской Федерации, государственных программ субъектов Российской Федерации, федеральных и региональных проектов, плановых и программно-целевых документов государственных корпораций, государственных компаний, акционерных обществ с государственным участием,

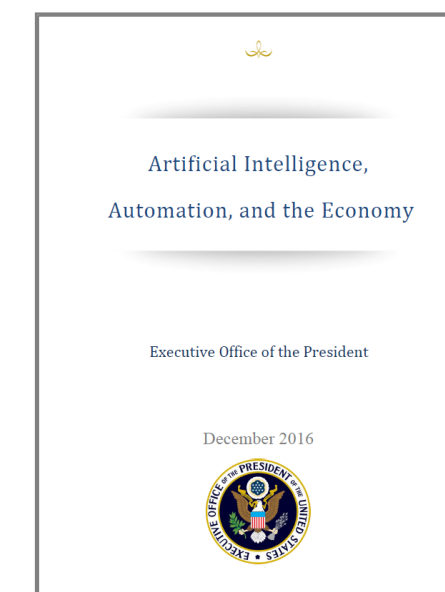
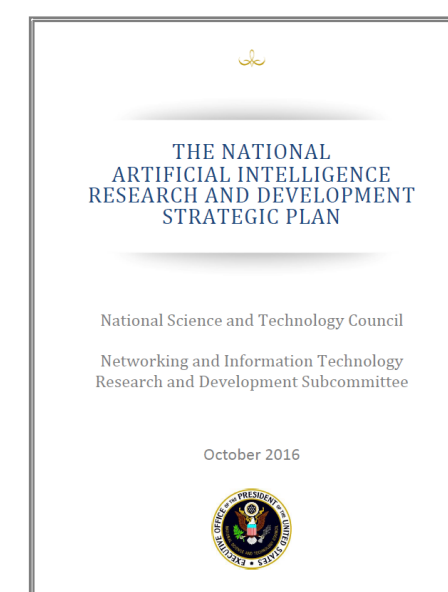
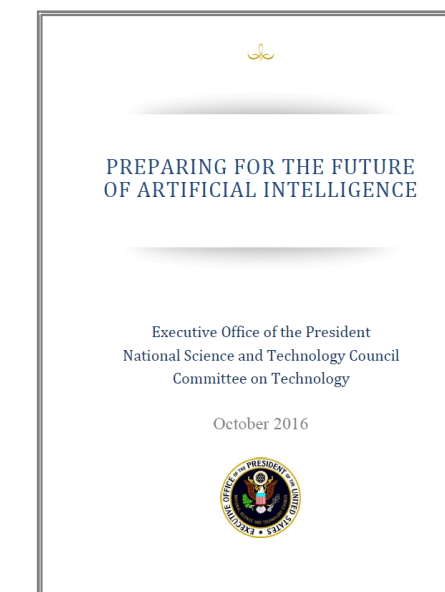
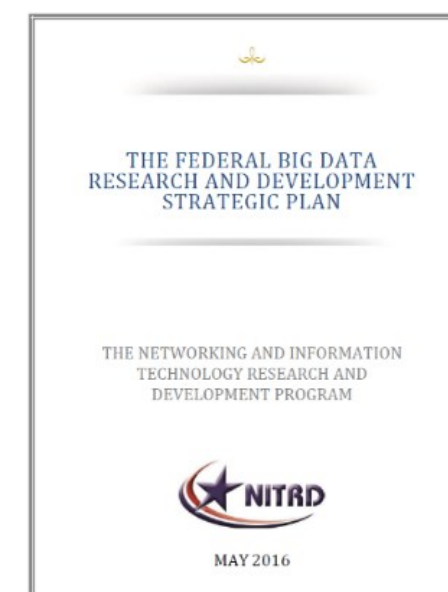


2 100011 98860 6

Отчёты Белого дома США, май-октябрь 2016

«Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future»

Цифровая и распределённая экономика
Автоматизация и сокращение издержек
Автономный транспорт и роботизация
Оптимизация логистики и цепей поставок
Оптимизация энергетических сетей (Energy Tech)
Автоматизация банковских услуг (Fin Tech)
Автоматизация юридических услуг (Legal Tech)
Автоматизация образовательных услуг (Ed Tech)
Автоматизация работы с кадрами (HR Tech)
Персональная медицина (Med Tech)
Автоматизация в сельском хозяйстве (Agro Tech)
Автономные системы вооружений (Mil Tech)



«Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, *искусственном интеллекте* и *машинном обучении*» (январь 2016)

Клаус Мартин Шваб,
президент Всемирного
экономического форума



Бум искусственного интеллекта

1956: Дартмутский семинар, введено понятие AI

1997: IBM Deep Blue обыграл чемпиона мира по шахматам

2005: Беспилотный автомобиль: DARPA Grand Challenge

2006: Google Translate – статистический машинный перевод

2011: 40 лет DARPA CALO привели к созданию Apple Siri

2011: IBM Watson победил в ТВ-игре «Jeopardy!»

2015: ImageNet: превзойдён человеческий уровень ошибок распознавания 5%

2015: Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана

2016: DeepMind, OpenAI: динамическое обучение играм Atari

2016: Google DeepMind обыграл чемпиона мира по игре го

2017: OpenAI обыграл чемпиона мира по компьютерной игре Dota 2

2020: Модель GPT-3 от OpenAI синтезирует тексты, неотличимые от человеческих

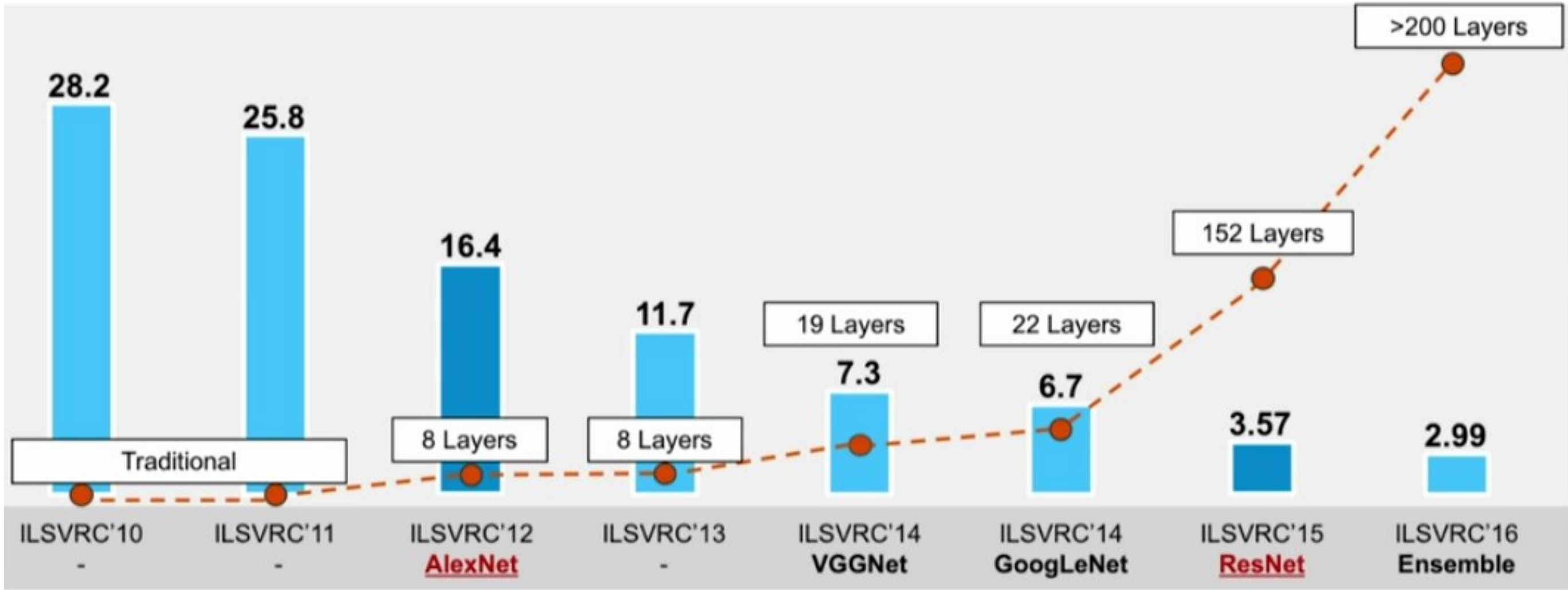
2023: GPT-4 от OpenAI демонстрирует «проблески общего искусственного интеллекта»



Большие данные и глубокие нейронные сети

ImageNet: открытая выборка 14М изображений, 20К категорий

IMAGENET



Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

Конкурсы (контесты) анализа данных как драйверы цифрового развития отраслей

Netflix Prize (2006—2009): \$1 000 000
рекомендации фильмов пользователям сервиса



ImageNet (2007—2016):
распознавание 22К классов объектов на 15М изображениях



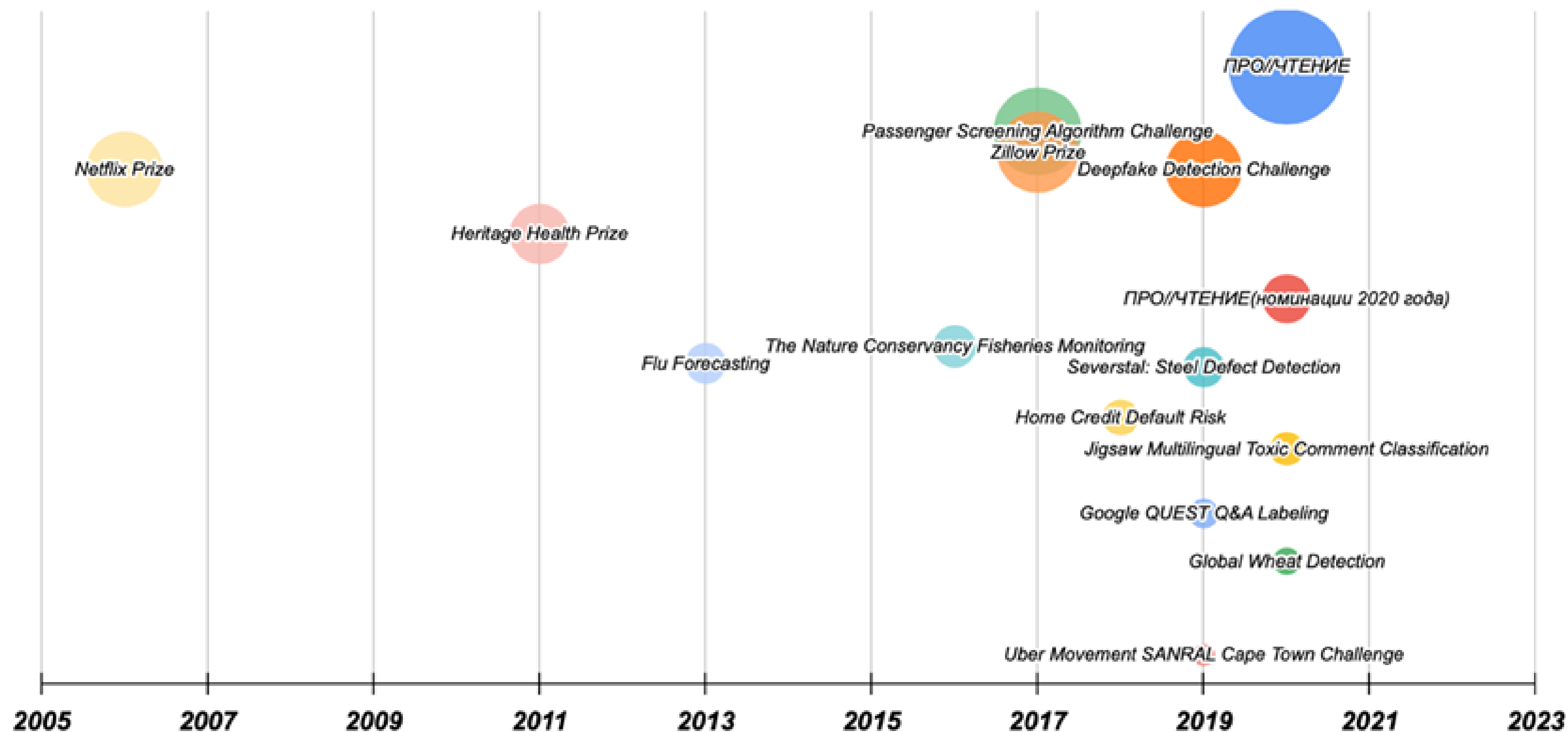
ПРО//ЧТЕНИЕ (2019—2022): 100+100 млн. руб.
Поиск смысловых ошибок в школьных сочинениях ЕГЭ



Конкурсные платформы



Конкурсы (контесты) анализа данных как драйверы цифрового развития отраслей



Рынок труда в области анализа данных

Инженер по данным (Data Engineer)

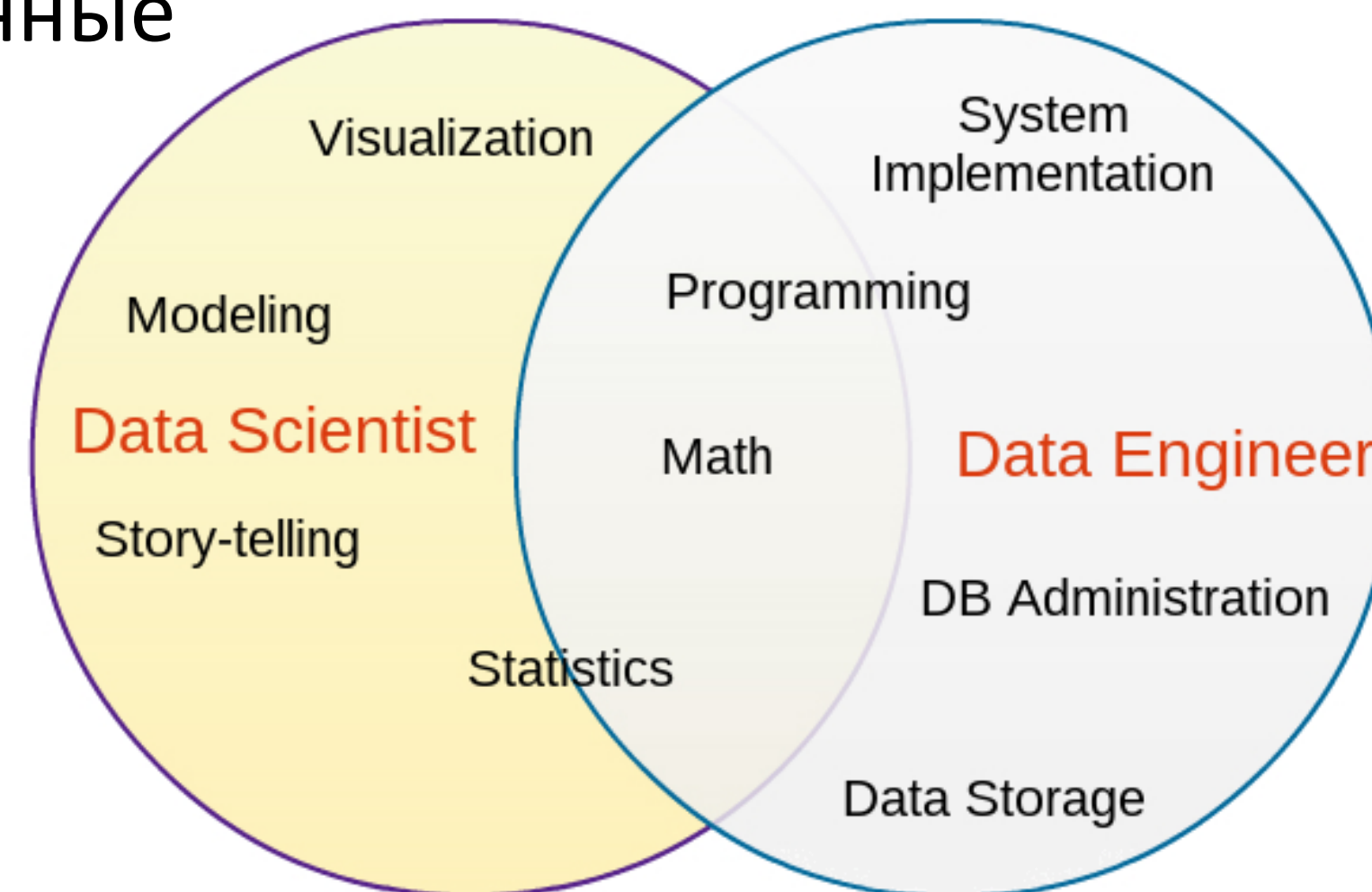
- Понимает бизнес-процессы, порождающие данные
- Работает с данными в различных форматах
- Визуализирует, понимает, готовит данные

Исследователь данных (Data Scientist)

- Строит модели, подбирает методы
- Оценивает решения, организует тестирование

Менеджер проектов по анализу данных

- Организует бизнес-процессы сбора и очистки данных
- Понимает бизнес задачи и формализует их в терминах «Дано-Найти-Критерий»
- Организует открытые конкурсы и пилотные проекты
- Оценивает сложность задач и трудозатраты



Проблемы и пути решения

Проблемы:

- для системной цифровизации всех отраслей экономики
- нужны сотни тысяч специалистов по DS/AI/ML,
- обладающих сочетанием технических и предметных знаний,
- а также тысячи преподавателей с опытом решения прикладных задач

Пути решения:

- создание общедоступного депоzitария прикладных задач и
- межотраслевого стандарта образовательных DS/AI/ML модулей

Структура образовательных DS/AI/ML модулей

1. Научно-популярное введение, история успеха, мотивации
2. Ссылки на образовательный контент (лекции, видео, слайды)
3. Ссылки на источники научных и профессиональных знаний
4. Ссылки на связанные задачи
5. **Открытый набор данных**
6. **Открытый код базовых решений и метрик качества**
7. **Активный контеcт на открытой конкурсной площадке**
8. Методики и системы автоматизированного контроля знаний
9. Методические материалы, открытое экспертное сообщество
10. Постановки открытых исследовательских задач для НИР/НИОКР

Проблемы и пути решения

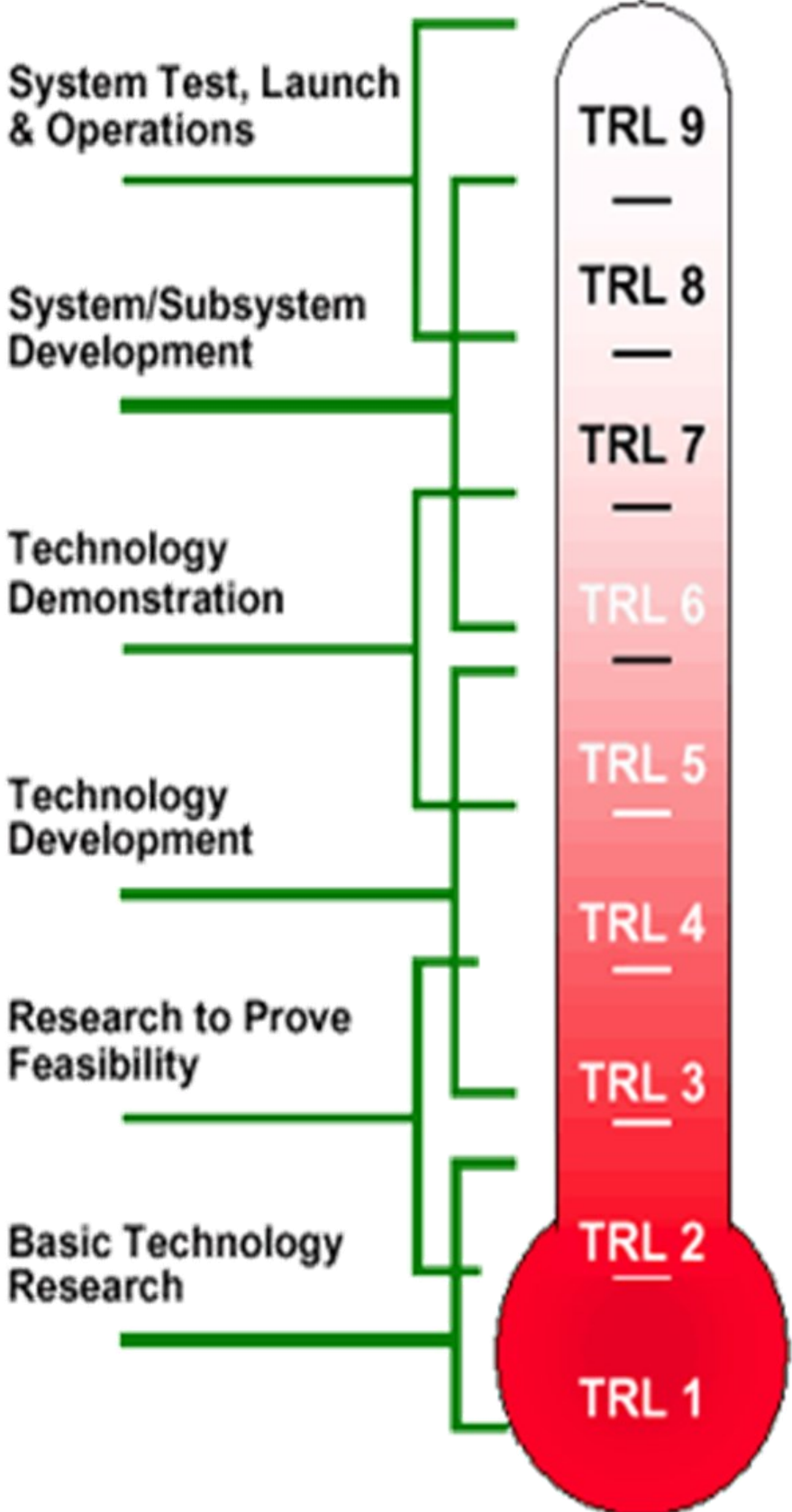
Проблемы:

- бизнес слабо заинтересован в предоставлении открытых данных для образовательных целей,
- имеет потребительское отношение к образованию и науке,
- слабо способствует устранению TRL-разрыва «предпосевной стадии»

Пути решения:

- создание **бирж данных** (при государственной поддержке)
- создание (при государственной поддержке) инфраструктуры для развития технологических компаний, устраняющих TRL-разрыв
- создание нормативной базы

Концепция TRL (Technology Readiness Level)



TRL - 9	1) Постановка на производство 2) Запуск в серию 3) Коммерциализация	SEED (Посевной)
TRL - 8	Финальный прототип (pre-production) готовый к опытной/малой серии	
TRL - 7	Пилотные/полевые испытания и тесты прототипов	PRE - SEED (Предпосевной)
TRL - 6	Построение модели, прототипа, «похожего на ожидания»	
TRL - 5	Этап конструирования узлов и модулей	
TRL - 4	Лабораторный прототип, валидация методики в лабораторных условиях	Фаза научных исследований и изысканий
TRL - 3	Разработка базовой технологии в лаборатории	
TRL - 2	Техническая проработка, технологический концепт	
TRL - 1	Идея, описание принципов работы	

Биржи данных: китайский опыт

Шанхайская биржа данных

первая сделка: 25 ноября 2021 г.



Государственная поддержка

«Ускорить развитие единого рынка технологий и данных. Ускорить развитие **рынка данных как фактора производства**, создать комплексную базовую систему, стандарты и нормативы в области безопасности данных, защиты прав, управления трансграничными перемещениями, сделок и оборота данных, открытия и совместного использования данных и сертификации безопасности.»

«Устранить проблемы с монополизацией данных платформенными компаниями, предотвращать использование данных, алгоритмов и технологических средств для исключения или ограничения конкуренции.»

*«Мнения об ускоренном создании единого национального рынка данных»,
25 марта 2022 г., ЦК КПК, Госсовет КНР*

Выводы

Цифровизация экономики требует огромного количества (сотен тысяч) специалистов в области ИИ / МО / АД;

в пропорции, ориентировочно, *инженеров : математиков $\approx 20 : 1$*

Для организации такого обучения необходима централизация усилий:

- единый депозитарий практических задач
- сетевое сообщество преподавателей (например, на базе ODS)
- доступные отечественные платформы и вычислительные ресурсы

Воронцов Константин Вячеславович

voron@mlsa-iai.ru