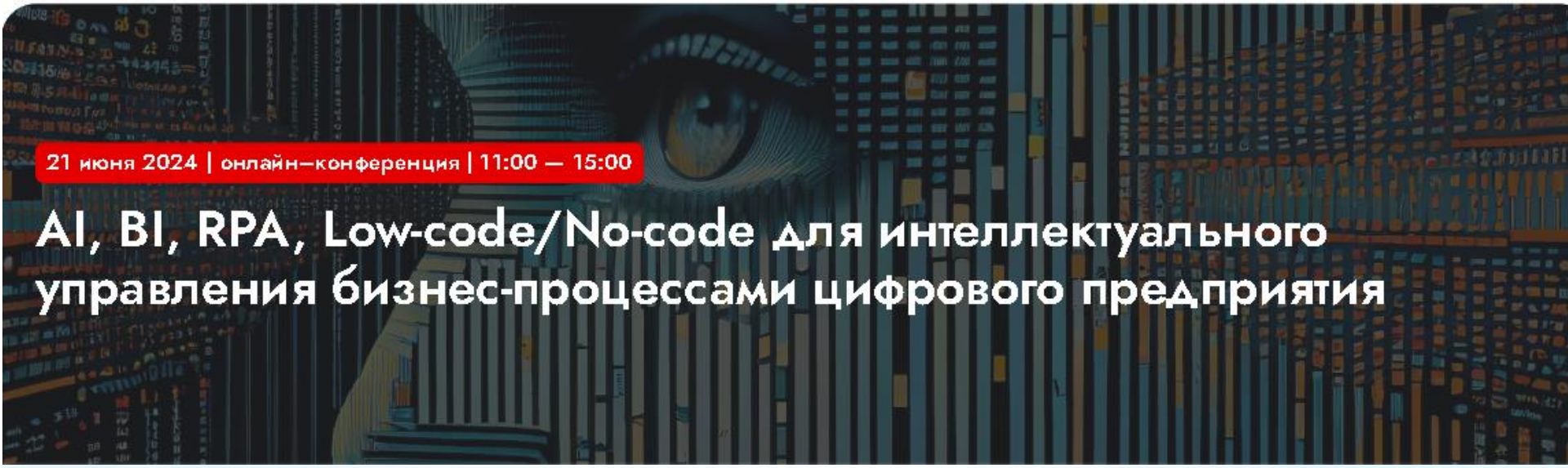


Особенности разработки, внедрения и эксплуатации систем искусственного интеллекта

Марина Аншина
Председатель Правления Российского Союза ИТ-Директоров
Член Совета по профессиональным квалификациям в области ИТ (СПК-ИТ)

ИСТОРИЯ И ТЕРМИНОЛОГИЯ

Термин и применение

- Автором термина «искусственный интеллект» (1956 г.) является Джон Маккарти, изобретатель языка Лисп, основоположник функционального программирования и лауреат премии Тьюринга за огромный вклад в область исследований искусственного интеллекта

Два основных подхода к разработке ИИ

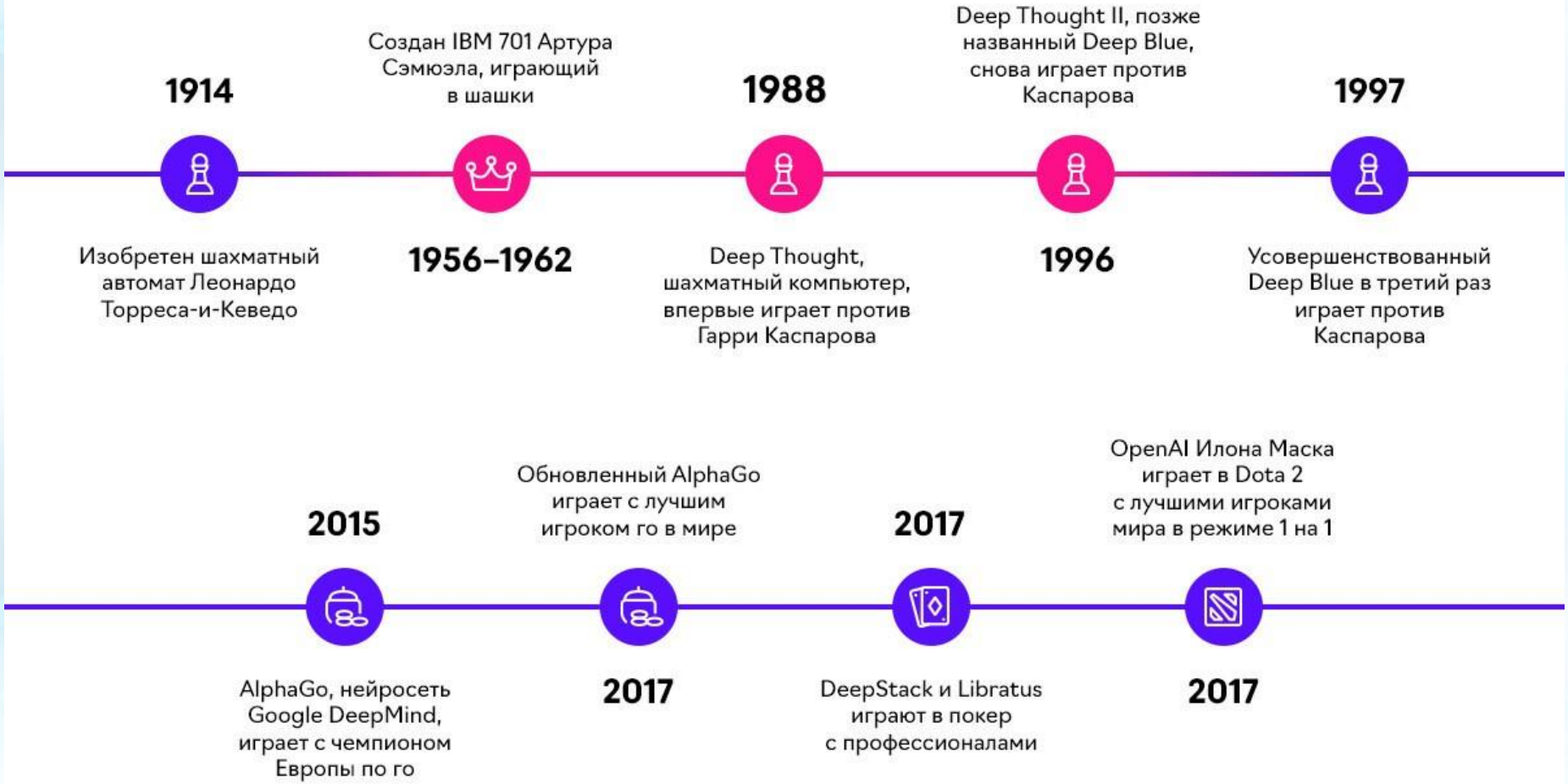
- нисходящий (*Top-Down*), семиотический — создание экспертных систем, баз знаний и систем логического вывода, имитирующих высокоуровневые психические процессы: мышление, рассуждение, речь, эмоции, творчество и т. д.;
- восходящий (*Bottom-Up*), биологический — изучение нейронных сетей и эволюционных вычислений, моделирующих интеллектуальное поведение на основе биологических элементов, а также создание соответствующих вычислительных систем, таких
как нейрокомпьютер или биокомпьютер.

Конференция по ИИ в Дартмуте (1956)

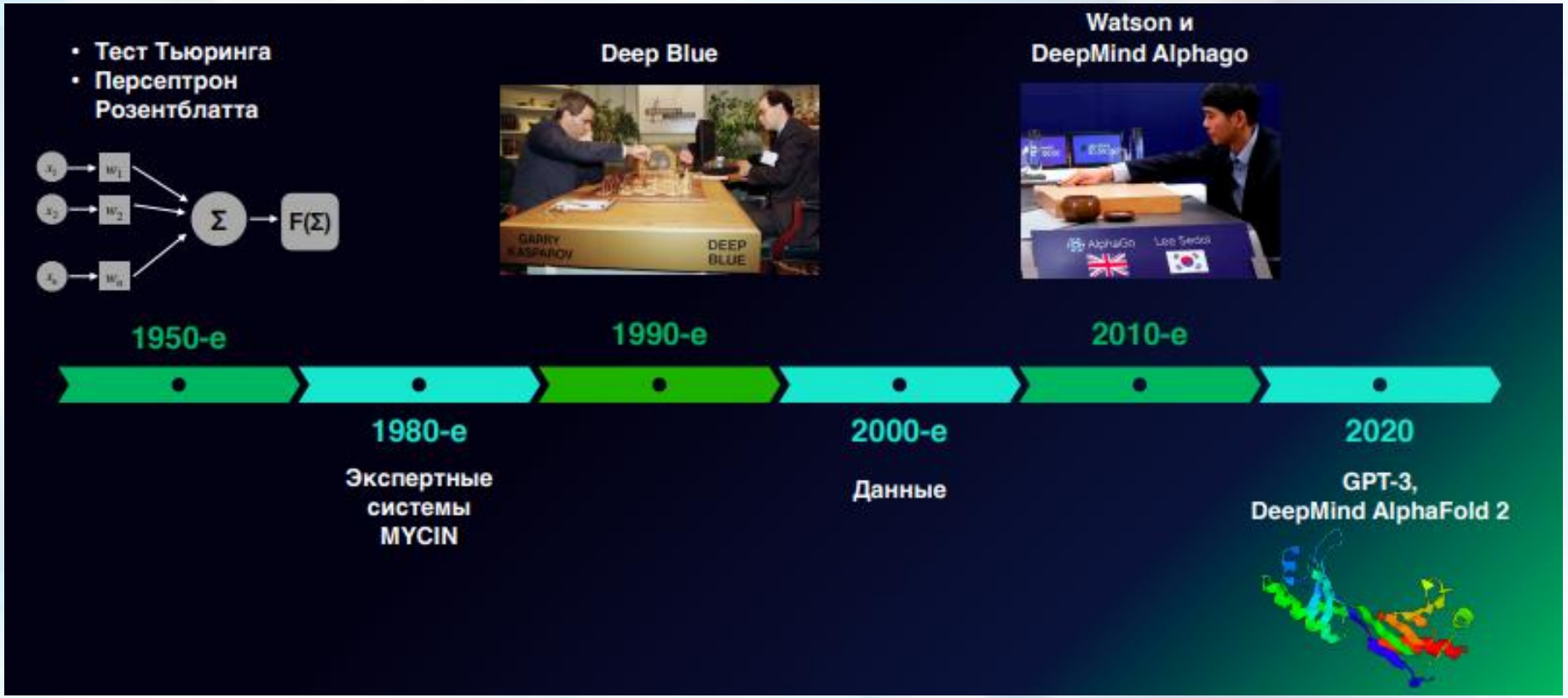
- К 1955 году учёные всего мира уже сформировали такие концепции, как нейросети и естественный язык, однако ещё не существовал объединяющих концепций, охватывающих различные разновидности машинного интеллекта.
- Профессор математики из Дартмутского колледжа, Джон Маккарти, придумал термин «искусственный интеллект», объединяющий их все.
- Маккарти руководил группой, подавшей заявку на грант для организации конференции по ИИ в 1956.
- На эту конференцию в Дартмут-холл летом 1956 были приглашены многие ведущие исследователи того времени.
- Учёные обсуждали различные потенциальные области изучения ИИ, включая обучение и поиск, зрение, логические рассуждения, язык и разум, игры (в частности, шахматы), взаимодействия человека с такими разумными машинами, как личные роботы.
- Общим консенсусом тех обсуждений стало то, что у ИИ есть огромный потенциал для того, чтобы принести пользу людям.

Человек VS искусственный интеллект

Победа человека **Победа искусственного интеллекта**

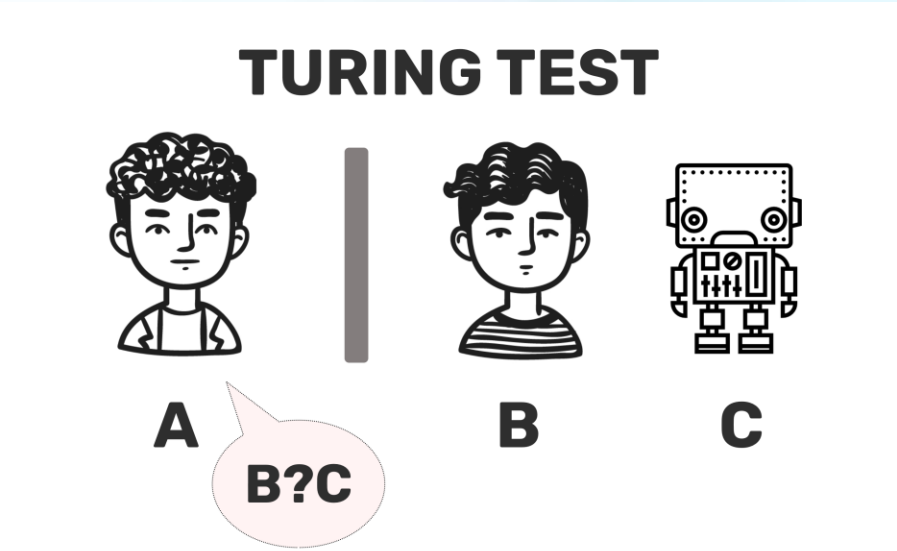


Основные этапы развития ИИ



Как проверить интеллект?

- В философии не решён вопрос о природе и статусе человеческого интеллекта.
- Нет точного критерия достижения компьютерами «разумности», хотя на заре искусственного интеллекта был предложен ряд гипотез, например, тест Тьюринга



ИИ – СВОЁ,
ЧЕЛОВЕКУ - СВОЁ

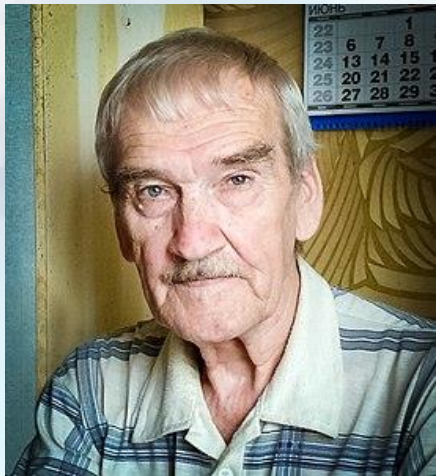
Искусственный интеллект

Способность компьютера обучаться, принимать решения и выполнять действия, свойственные человеческому интеллекту.

- Может ли ИИ делать что-то лучше человека?
- Можем ли мы доверять ИИ?

Человек vs ИИ

- **Станисла́в Евгра́фович Петро́в** (7.09.39 – 19.05.2017) - советский офицер, предотвративший 26 сентября 1983 года потенциальную ядерную войну после ложного срабатывания системы предупреждения о ракетном нападении со стороны США.



Право на ошибку

- Не бывает безошибочных программ, бывают недотестированные
- Не бывает здоровых людей, бывают недообследованные

Искусственный интеллект и естественный интеллект

Искусственный интеллект

Конвейер

- Действует по четко определенным алгоритмам и сценариям
- Умеет обрабатывать огромные объёмы информации в ограниченные сроки
- Может ошибиться при неожидаемом использовании
- Детерминированная система
- Интеграция
- Информационная безопасность

Естественный интеллект

Муравейник

- Умеет действовать в условиях неопределенности
- Может ошибаться при обработке информации
- Вероятностная система
- Коммуникации
- Безопасность

Интеллектуальный работник

работник индустриальной эпохи

Принципы Тейлора

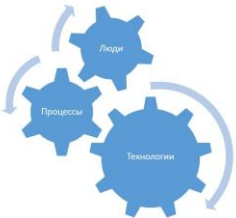
- проанализировать задание;
- записать все движения, проанализировать усилия и время, затраченное на каждое движение;
- исключить лишние движения, соединить рациональным образом оставшиеся так, чтобы минимизировать время выполнения задания и усилия работника, попутно рационализировать инструменты и орудия труда;
- сформировать описание необходимых действий и довести его до каждого работника.

Средство производства - капитал

интеллектуальный работник

- ответственность за эффективность перекладывается на плечи самого работника;
- неотъемлемой частью деятельности работника становятся инновации;
- работа включает в себя собственное обучение и обучение других (наставничество);
- результат измеряется не столько количественными, сколько качественными показателями;
- работник – это не затраты, а активы

Средство производства - информация



Развитие технологий по Гартнер

Период разочарования в облаках и больших данных.

Завышенные ожидания не оправдались

Вы просто не умеете их готовить

РАЗРАБОТКА, ВНЕДРЕНИЕ И ЭКСПЛУАТАЦИЯ

Machine learning

Машинное обучение: Процесс автоматического построения некоторого алгоритма решения задачи на основе данных

Направления:

Решить задачу:
подобрать такой алгоритм, который может ***по признакам*** *для каждого объекта* ***предсказать ответ***

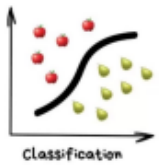
- **Обучение с учителем**
  - *Объекты*
  - *Признаки*
  - *Ответы*
- **Обучение с подкреплением**

- **Обучение без учителя**
  - *Объекты*
  - *Признаки*

Решить задачу:
подобрать алгоритм, который может ***по признакам:***
- *Разделять объекты*
- *Уменьшать количество исходных признаков*
- *Искать закономерности*

Класс задачи - от типа ответа

- **Регрессия**



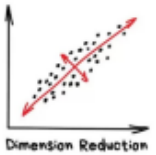
- **Классификация**

- **Ранжирование**

Класс задачи – от бизнес цели

- **Кластеризация**

- **Понижение размерности**

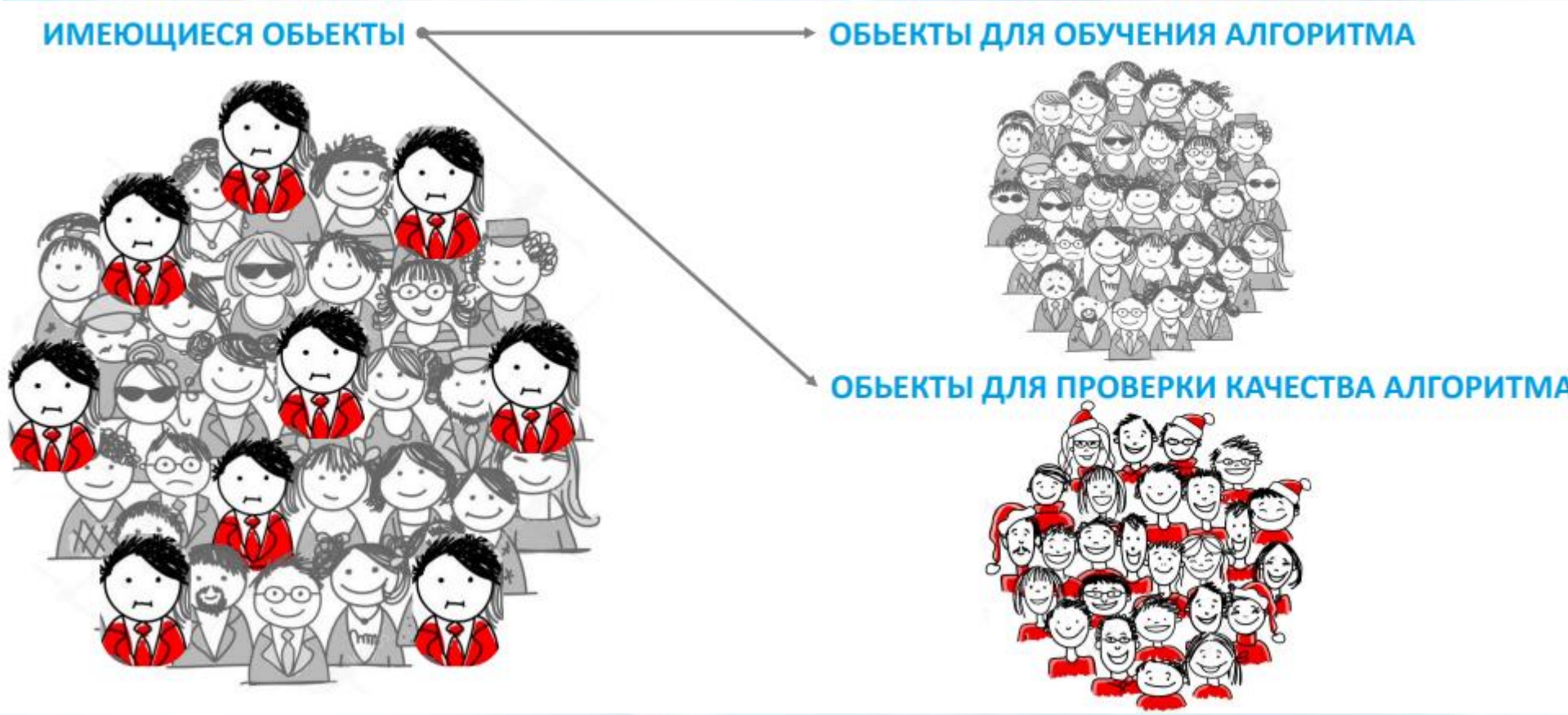


- **Ассоциации**

Типовой сценарий разработки, валидации и внедрения модели

Валидация vs верификация

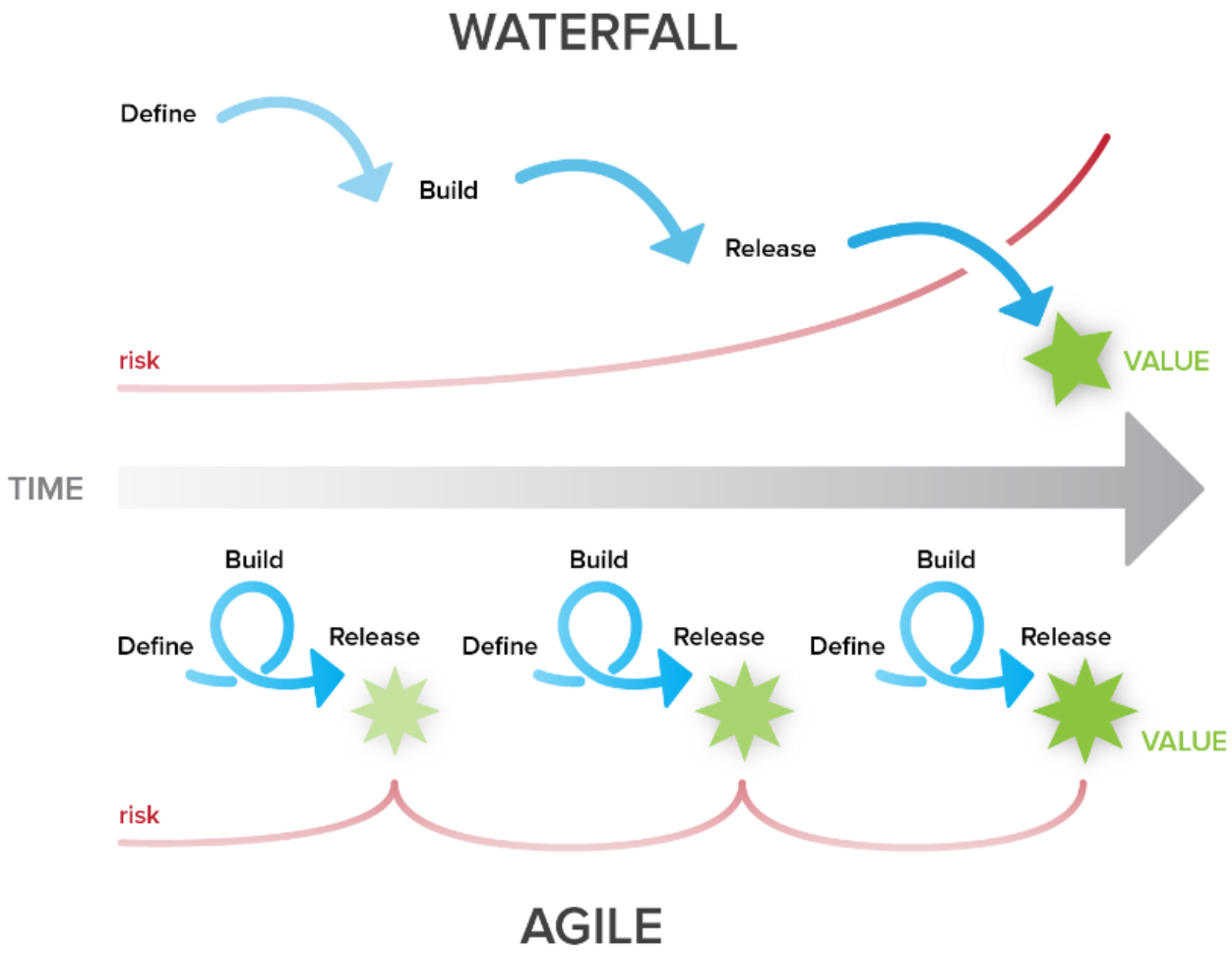
Предобработка данных
Обучающая и тестовая выборки



Особенности проекта

- Цель
- Качество данных
- Продвижение
- Пользователи
- Валидация
- Сопровождение

Водопад vs Agile

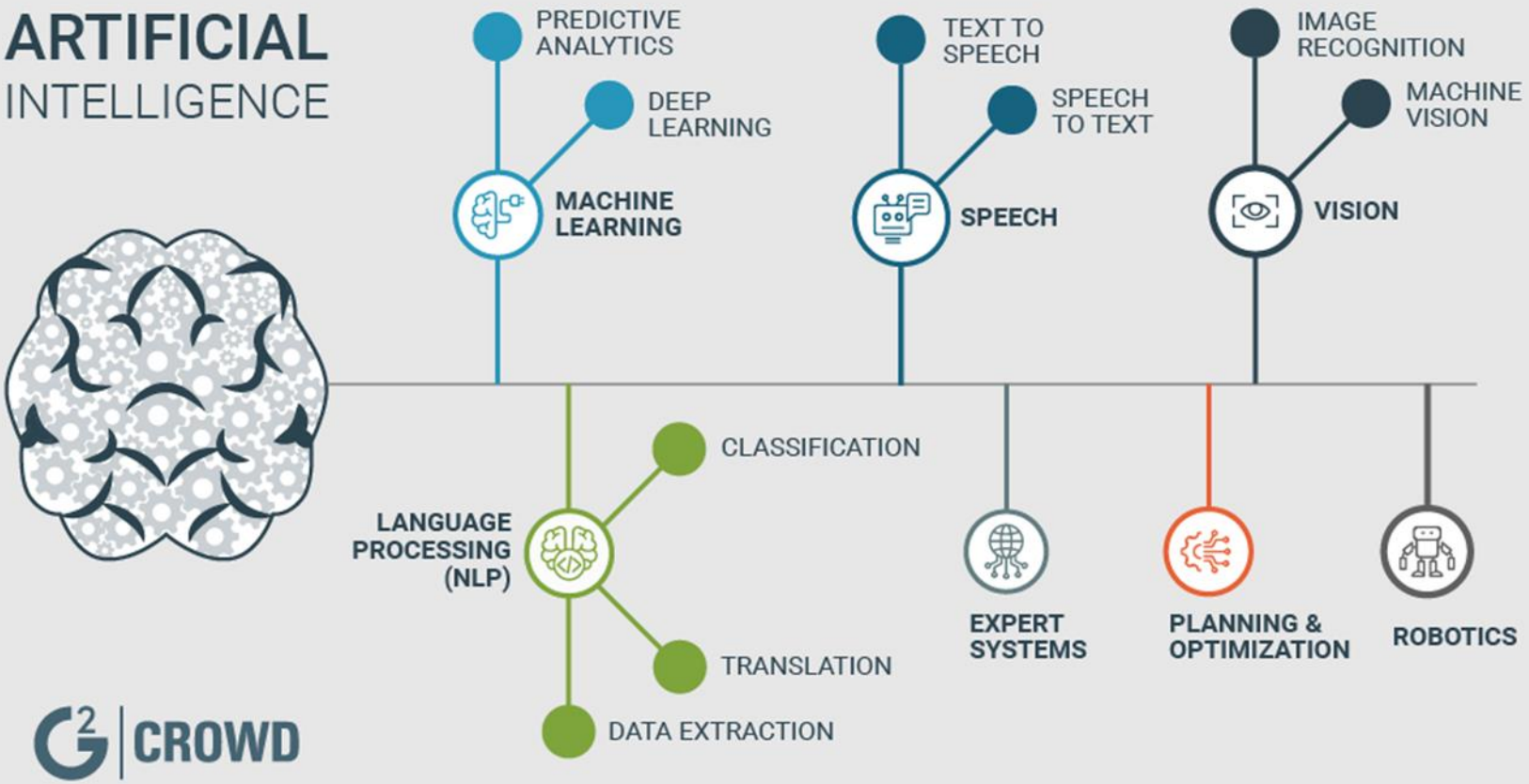


Верификация

Валидация

Внедрение ИИ

- Неприятие vs доверие



Применение ИИ

10 применений ИИ, которые изменят медицину

ИИ от IBM Watson for Oncology

- ИИ от IBM Watson for Oncology, предоставляющий рекомендации по лечению рака, так и не смог заслужить доверие онкологов.
- При взаимодействии с Watson врачи оказывались в двоякой ситуации. Когда указания программы совпадали с мнениями медиков, те не видели большой ценности в рекомендациях ИИ, поскольку они не меняли лечения. Скорее, врачи просто укреплялись в собственной правоте.
- Но подтвердить, что ИИ улучшает статистику выживаемости с раком, не удалось.

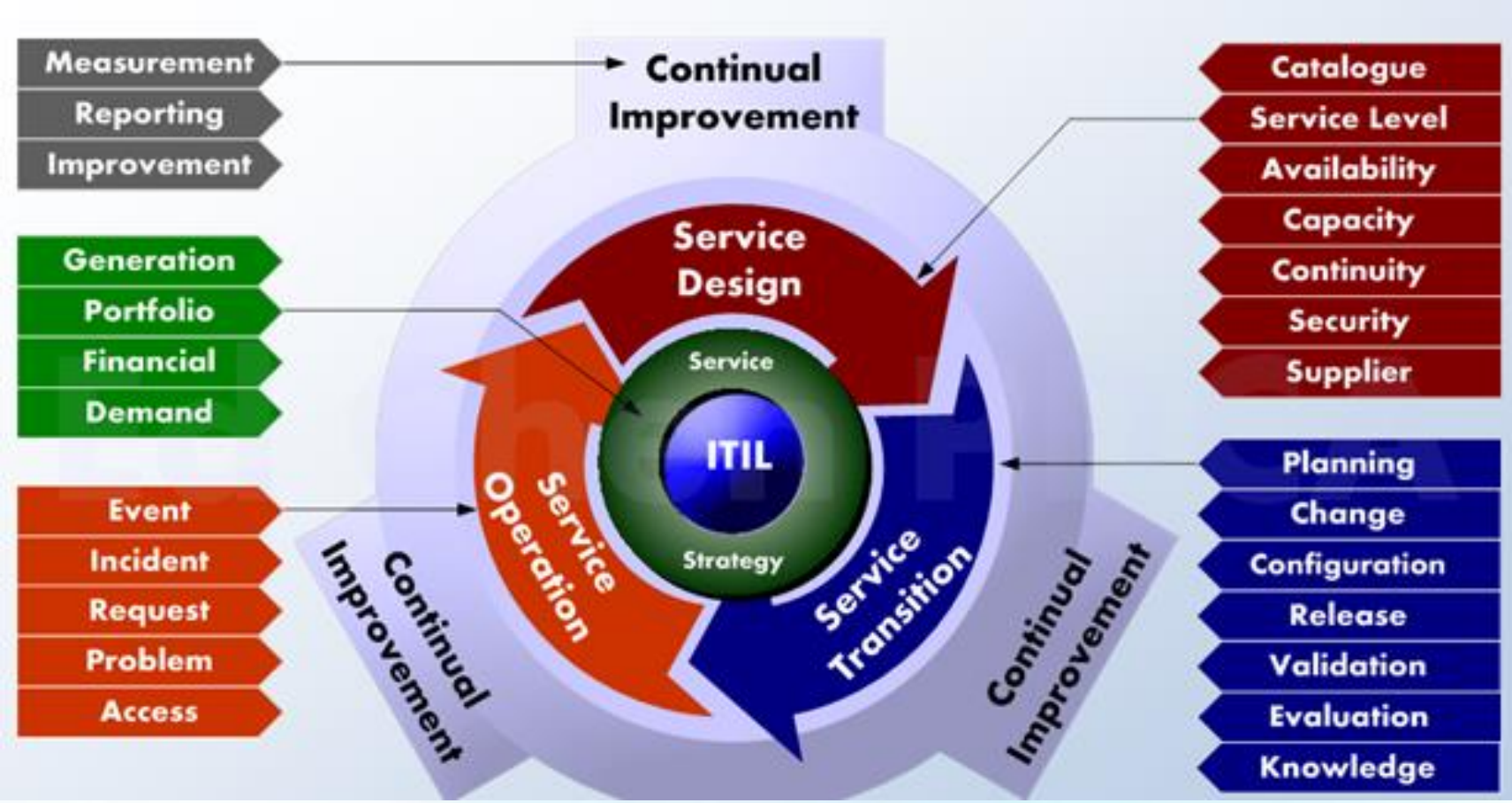
Эксплуатация

Программная система требует обслуживания и поддержки:

- Управление инцидентами и проблемами
- Управление конфигурациями и выпусками (релизами)
- Управление SLA
- Техническое обслуживание

ITIL - IT infrastructure library

- ITIL — это совокупность правил и методических указаний, благодаря которым компания может вывести IT-процессы на тот уровень, который соответствует её целям.



ЭТИКА ИИ

Три закона робототехники Айзека Азимова

- Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.
- Робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
- Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Законам.

Рассказ «Хоровод», 1942 г.

Уважение к человеческой жизни

- Одним из важнейших принципов в области этики ИИ является создание систем, способных сохранять уважение к человеческой жизни, достоинству и правам.
- В числе прочего, необходимо учитывать культурные различия во избежание нежелательных социальных последствий.

Философия (этика) становится точной наукой

- Алгоритм действий ИИ в критических ситуациях
- Обеспечение непрерывности бизнеса
- Качество информации и манипулирование людьми

Риск для общества
Отсутствие общих протоколов безопасности

- Некоммерческая организация Future of Life опубликовала письмо за подписью главы Tesla, SpaceX и Twitter Илона Маска, одного из создателей Apple Стива Возняка, и ещё более тысячи экспертов в области ИИ.
- Они настаивают на перерыве в исследованиях до появления общих протоколов безопасности.
- В обращении говорится, что системы с интеллектом, сравнимым с человеческим, представляют большой риск для общества.

Отличие правдивой информации от ложной

- К числу тех, кто видит в развитии ИИ опасность для человечества, присоединился после увольнения из Google один из основоположников нейросетей Джеффри Хинтон.
- По его мнению, созданный ИИ новостной контент (фото, видео, тексты) наводнит Интернет, и люди не смогут отличить правдивую информацию от ложной.

Присутствие ИИ в повседневности

- Среди наиболее заметных примеров – смертельные ДТП с участием самоуправляемых автомобилей Tesla и Uber.
- «Проблема вагонетки» – мысленный эксперимент, состоящий из набора ситуаций, когда необходимо выбрать одно из решений, чтобы обойтись минимальным числом жертв. В случае с беспилотным автомобилем алгоритмы ИИ должны просчитать, как лучше поступить в экстренных случаях на дороге: свернуть, создав угрозу для пассажиров, либо продолжать движение, подвергнув опасности нарушителей ПДД
- Люди в большинстве случаев стремятся указать наиболее рациональное решение, в основном принося в жертву пассажиров самоуправляемого автомобиля. Однако в роли пассажиров большинство выбирает тот вариант, который ставит их безопасность на первое место.

Манипулирование информацией

- Немаловажные проблемы в сфере использования ИИ касаются, в том числе, манипулирования информацией и различного рода дискриминацией.
- Завышенное доверие к ним со стороны неспециалистов
- Заниженное доверие к ним со стороны специалистов
- ИИ от IBM Watson for Oncology, предоставляющий рекомендации по лечению рака, так и не смог заслужить доверие онкологов.

Как занимается выборами Cambridge Analytica

- Фирма закупает персональные данные из всех возможных источников: кадастровые списки, бонусные программы, телефонные справочники, клубные карты, газетные подписки, медицинские данные.
- Затем Cambridge Analytica скрещивает эти данные со списками зарегистрированных сторонников [партий] и данными по лайкам-репостам в Facebook —получается личный профиль.
- Из цифровых данных возникают люди со страхами, стремлениями и интересами — и с адресами проживания.
- С июля 2016 года волонтеры кампании Трампа получили приложение, которое подсказывает политические предпочтения и личностные типы жителей того или иного дома.
- Волонтеры-агитаторы модифицировали свой разговор с жителями исходя из этих данных.
- Обратную реакцию волонтеры записывали в это же приложение — и данные отправлялись прямиком в аналитический центр.
- Трамп пустил три четверти рекламного бюджета в цифровую сферу.
- Facebook превратился в совершенное оружие и лучшего помощника на выборах, как написал в Twitter один из сподвижников Трампа.

Михаил Косинский: «Мы не заметим, как мир захватит искусственный интеллект»

- Основы технологий Cambridge Analytica лежат в исследованиях выпускника Кембриджика, выходца из Варшавы Михала Козинского.
- Начиная с 2008 год он с однокурсниками экспериментировал в Facebook с приложением MyPersonality.
- Модель Козинского позволяет: «лучше узнавать личность после десяти изученных лайков, нежели человека знают коллеги по работе.
- После 70 лайков — лучше, чем друг.
- После 150 лайков — лучше, чем родители.
- После 300 лайков — лучше, чем партнер.
- С еще большим количеством изученных действий можно было бы узнать о человеке лучше, чем он сам».

- «Трамп действует как идеальный оппортунистский алгоритм, который опирается лишь на реакцию публики», — отмечала в августе математик Кэти О’Нил.
- В день третьих дебатов между Трампом и Клинтон команда Трампа отправила в соцсети (преимущественно, Facebook) свыше 175 тыс. различных вариаций посланий.
- По сути, Козинский изобрел поисковую систему по людям.
- Но что случится, думал Козинский, если кто-то решит воспользоваться этой поисковой системой по людям, чтобы ими манипулировать?
- Он начинает ставить предупреждения на всех своих научных публикациях. Предупреждения о том, что его методы «могут нести угрозу благополучию, свободе или даже жизни людей».

10 Законов для искусственного интеллекта

- В числе первых в 2016 году были опубликованы «10 Законов для искусственного интеллекта» Microsoft, в которых от имени генерального директора компании Сатьи Наделлы указаны ключевые требования к развитию этики ИИ.

Правила ИИ

- Почти полсотни крупных ИТ-компаний по всему миру обладают собственными кодексами и правилами, основанные на этических принципах, относящихся к применению и развитию ИИ.
- В числе таких компаний российские ABBYY, Сбер и Яндекс.

What is AI Ethics? (6:10)

Этические нормы и ценности ИИ

- Этические нормы и ценности сформулированы в 13 из «23 принципов искусственного интеллекта» на Асиломарской конференции в 2017 году.
- В числе тех, кто подписал их, – Илон Маск, Стивен Хокинг, Рэй Курцвайл и другие. Эти принципы отразились в корпоративных нормах ряда компаний, чья деятельность связана с разработкой ИИ.

Этика и ценности

- 6) Безопасность: системы ИИ должны быть безопасны и защищены на протяжении всего срока эксплуатации, а в ситуациях, где это целесообразно, штатная работа ИИ должна быть легко верифицируема.
- 7) Открытость сбоев в системе: если система ИИ причиняет вред, должна быть возможность выяснить причину.
- 8) Открытость системе правосудия: любое участие автономной системы ИИ в принятии судебного решения должно быть удовлетворительным образом обосновано и доступно для проверки компетентным органами.
- 9) Ответственность: разработчики и создатели продвинутых систем ИИ напрямую заинтересованы в моральной стороне последствий использования, злоупотребления и действий ИИ, и именно на их плечах лежит ответственность за формирование подобных последствий.

Этика и ценности

- 10) Синхронизация ценностей: системы ИИ с высокой степенью автономности должны быть разработаны таким образом, чтобы их цели и поведение были согласованы с человеческими ценностями на всем протяжении работы.
- 11) Человеческие ценности: устройство и функционирование систем ИИ должно быть согласовано с идеалами человеческого достоинства, прав, свобод и культурного разнообразия.
- 12) Защита личных данных: люди должны иметь право на доступ к персональным данным, их обработку и контроль, при наличии у систем ИИ возможности анализа и использования этих данных.
- 13) Свобода и конфиденциальность: применение систем ИИ к персональным данным не должно безосновательно сокращать реальную или субъективно воспринимаемую свободу людей.
- 14) Совместная выгода: технологии ИИ должны приносить пользу максимально возможному числу людей.

Этика и ценности

- 15) Совместное процветание: экономические блага, созданные при помощи ИИ, должны получить широкое распространение ради принесения пользы всему человечеству.
- 16) Контроль ИИ человеком: люди должны определять процедуру и степень необходимости передачи системе ИИ функции принятия решений для выполнения целей, поставленных человеком.
- 17) Устойчивость систем: те, кто обладает влиянием, управляя продвинутыми системами ИИ, должны уважать и улучшать общественные процессы, от которых зависит здоровье социума, а не подрывать таковые.
- 18) Гонка вооружений в области ИИ: стоит избегать гонки вооружений в области автономного летального оружия.

123-ФЗ от 24 апреля 2020 года

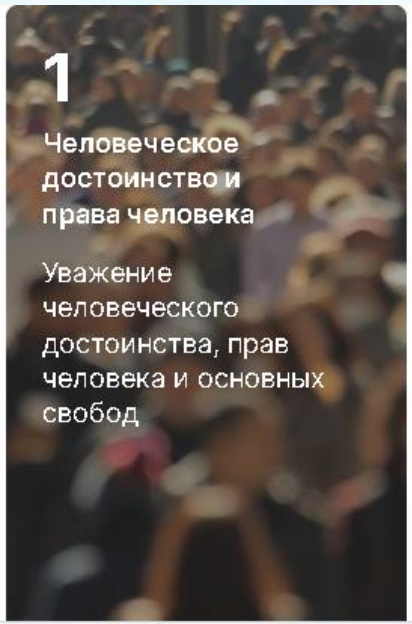
- Закон 123-ФЗ от 24 апреля 2020 года устанавливает в Москве специальный правовой режим в сфере ИИ. В нём указаны требования по защите персональных данных граждан и использования псевдоданных, собираемых в режиме анонимности. В частности, допускается обработка обезличенных персональных данных граждан для реализации эксперимента.
- Цель этого экспериментального правового режима – повышение качества жизни населения, эффективности госуправления и деятельности бизнеса в ходе внедрения технологий ИИ. А также формирование комплексной системы регулирования общественных отношений, возникающих в связи с развитием и использованием ИИ. Эксперимент позволит создать правовую базу для этого.

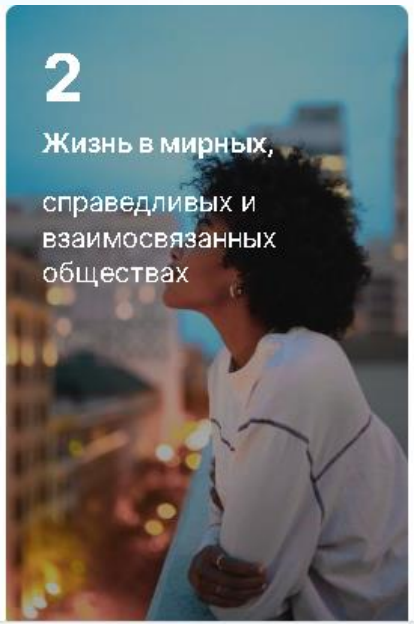
Группа ведущих экспертов по искусственному интеллекту (AI HLEG)

- В Европейском союзе была создана Группа ведущих экспертов по искусственному интеллекту (AI HLEG), которая занимается разработкой рекомендаций по этике. В частности, они предупреждают технологических гигантов о том, что алгоритмы не должны дискриминировать пользователей по признаку их возраста, расы или пола.
- В документе сказано, что ИИ должен соответствовать требованиям установленных норм и законов, этических принципов и ценностей, быть надёжным с технической и социальной точек зрения.
- Рекомендации включают необходимость учета этических вопросов на каждом этапе разработки и внедрения ИИ, прозрачности процессов, а также гарантии защиты конфиденциальности личных данных пользователей и другие меры.

UNESCO

- В ноябре 2021 года 193 государства-члена приняли **‘Рекомендацию по этическим аспектам искусственного интеллекта**’ — первый глобальный нормативный документ в этой области.
- https://www.unesco.org/ru/artificial-intelligence/recommendation-ethics





Кодекс этики в сфере ИИ

- В России была создана рабочая группа по разработке принципов этики ИИ.
- В 2021 году в рамках I международного форума «Этика искусственного интеллекта: начало доверия» был принят российский «Кодекс этики в сфере ИИ».
- На 2023 год документ подписали более 150 российских организаций.
- https://ethics.a-ai.ru

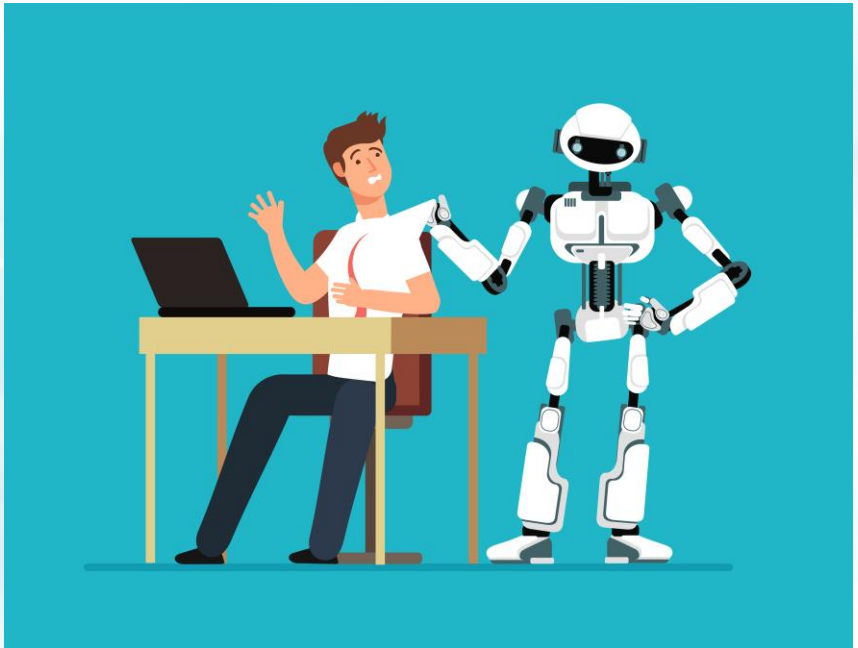
ОГЛАВЛЕНИЕ

Национальная стратегия развития ИИ до 2030 года

- Необходимость выработки этических норм и нормативного регулирования для ИИ прописана в «Национальной стратегии развития искусственного интеллекта на период до 2030 года».
- Ещё один важный документ – «Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 года».
- В ней сказано, что развитие технологий ИИ и РТ должно основываться на базовых этических нормах и предусматривать:
  - Цель обеспечения благополучия человека должна преобладать над иными целями разработки и применения систем ИИ и РТ.
  - Запрет на причинение вреда человеку по инициативе систем ИИ и РТ. По общему правилу, следует ограничивать разработку, оборот и применение систем ИИ и РТ, способных по своей инициативе целенаправленно причинять вред человеку.
  - Подконтрольность человеку в той мере, в которой это возможно с учётом требуемой степени автономности систем ИИ и РТ и иных обстоятельств.
  - Проектируемое соответствие закону, в том числе – требованиям безопасности: применение систем ИИ не должно заведомо для разработчика приводить к нарушению правовых норм.

Замена людей в некоторых областях

- Технологии со временем трансформируют рынок труда, заменив людей в некоторых областях
- Одни профессии уйдут, другие появятся



?

ВОПРОСЫ