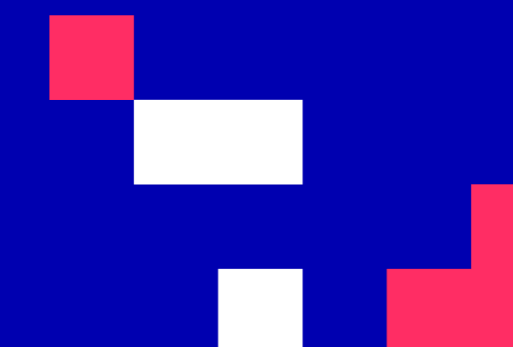


Русенски университет

# ИНТЕЛИГЕНТНИ КОМПЮТЪРНИ СИСТЕМИ

Светлана Стефанова

Септември, 2022



**ЛЕКЦИЯ 2****ИНТЕЛИГЕНТНИТЕ КОМПЮТЪРНИ СИСТЕМИ ДНЕС****СЪДЪРЖАНИЕ**

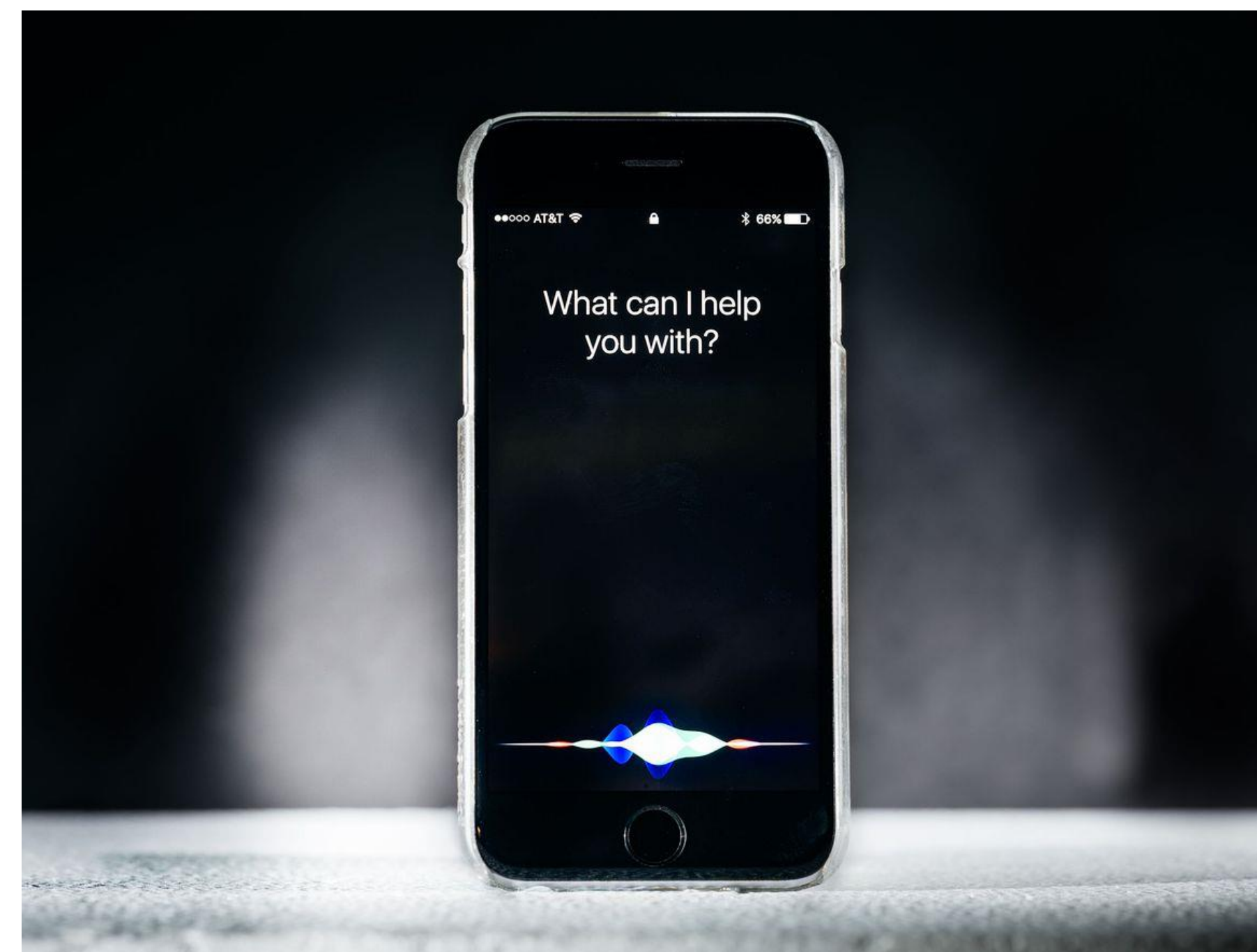
1. Въведение
2. Типове интелигентни компютърни системи
3. Рискове
4. Квантови компютри

**СЪДЪРЖАНИЕ 1**

# Популярност в наши дни

От SIRI (Apple) и Cortana (Microsoft) до самостоятелно управляващи се автомобили.

Докато научната фантастика често изобразява AI като роботи с човешки характеристики, то AI може да обхване всичко: от алгоритмите за търсене на Google до Watson на IBM за автономни оръжия.





**СЪДЪРЖАНИЕ 2**

# Приложение на роботите

3D:

- Dirty;
- Difficult;
- Dangerous.





**СЪДЪРЖАНИЕ 2**

## Варианти на бъдещето на ИКС (според Ник Бострьом, философ)

- **“Оракул”** – отговаря на запитвания с абсолютна точност;
- **“Дух”** – изпълнява заповед и очаква следваща;
- **“Суверен”** – възлага му се една основна цел и работи самостоятелно в заобикалящия го свят като взема решения как най-добре да я постигне.



## СЪДЪРЖАНИЕ 2

# Типове ИКС

- **тясно-профилирани/експертни** - проектирани да изпълняват конкретни задачи (например, само разпознаване на лица или само търсене в Интернет, или само управление на автомобил).
- **общи (AGI)** – с дългосрочна цел.

Докато тясно-профилирания AI може да надмине хората при всяка конкретна задача (напр. игра на шах), AGI ще надминава хората при почти всяка когнитивна задача.





**СЪДЪРЖАНИЕ 3**

# Основният риск

Хората не разбират в какво една ИКС е добра и в какво не.



**СЪДЪРЖАНИЕ 3**

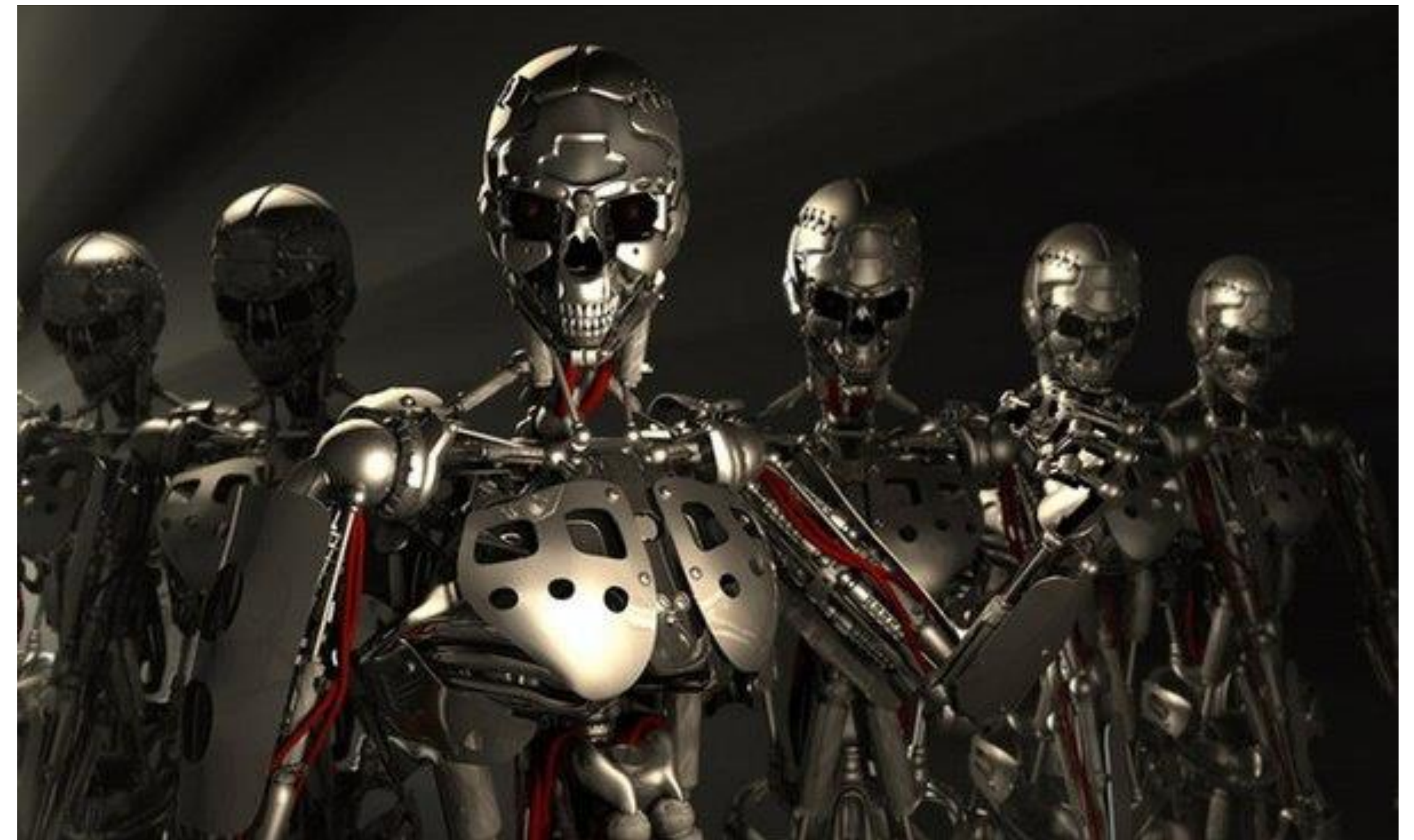
# Потенциални рискове – възможни сценарии

- ИКС, програмирана да направи нещо опустошително

Автономните оръжия са ИКС, програмирани да убиват. В ръцете на неподходящия човек могат лесно да причинят масови жертви. За да не бъдат спирани от врага, те биха били проектирани да бъдат изключително трудни за просто "изключване", така че хората в крайна сметка биха могли да загубят контрол върху такава ситуация.





**СЪДЪРЖАНИЕ 3****Американската армия има проект за замяна на ¼ от войниците си с роботи**



**СЪДЪРЖАНИЕ 3**

# Проект „Армия“ на Пентагона

- Планът - за 5 години (от 2017г) да се похарчи 1 млрд. долара за направата на роботи, които да придружават войниците на бойното поле.
- Задачи – разузнаване, обезвреждане на експлозиви, откриване на опасни химикали, носене на екипировка.

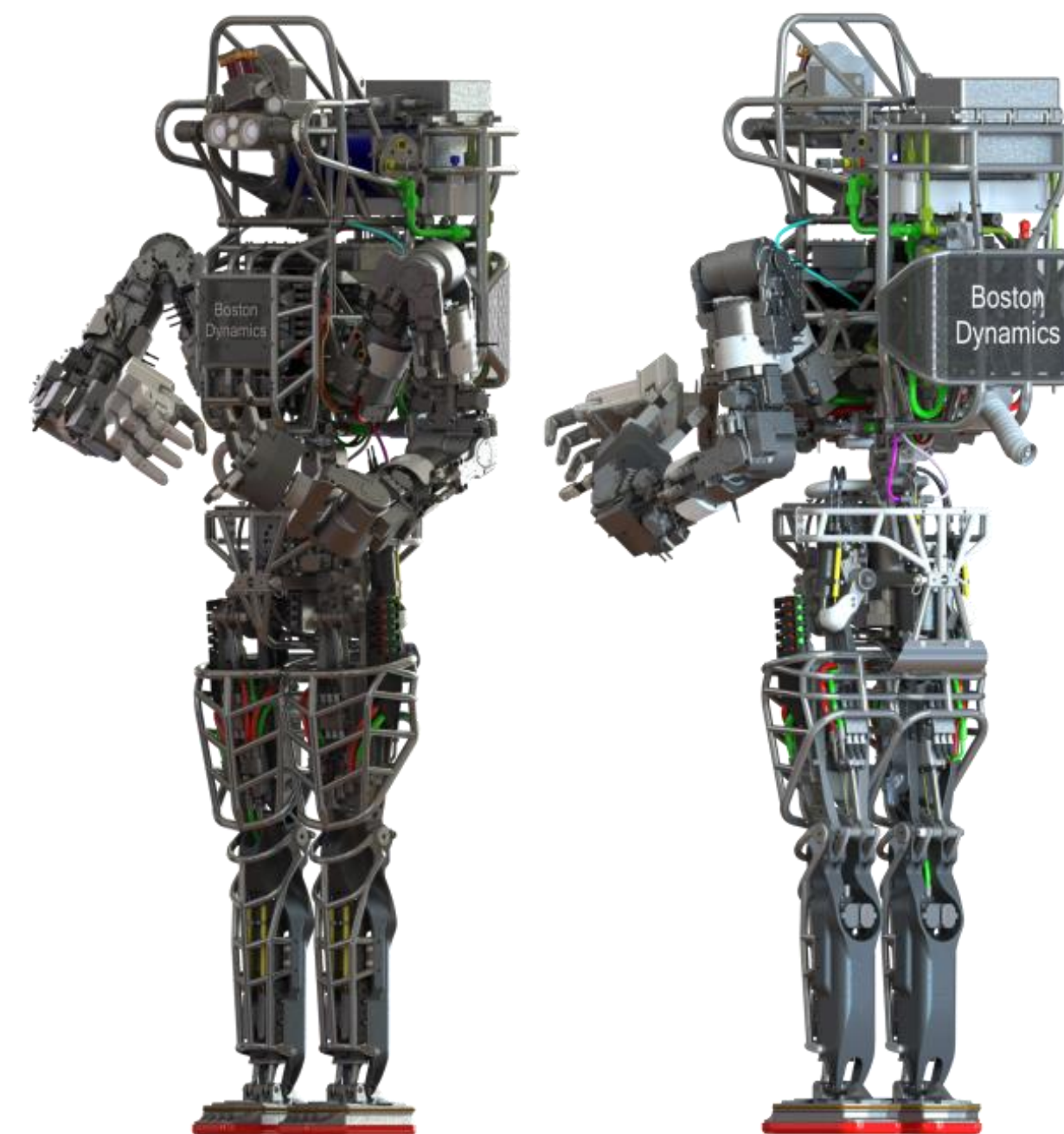




**СЪДЪРЖАНИЕ 3**

# Проект “Армия” - 3А

- Браян МакВей, ръководител на проекта - смята, че при приключване на проекта ще има работи във всяка армия. Тази нова армия ще бъде в пъти по-бърза от човешката и ще има способността да „вижда“ бойното поле. Според него, няма опасност от автономни работи стил „Терминатор“, тъй като плана е **командите винаги да се дават от хора** и роботите да не вземат сами решения.





**СЪДЪРЖАНИЕ 3**

# Проект “Армия” - ПРОТИВ

- Илон Мъск повежда група от 116 специалисти от 26 държави-пускат официално писма до ООН, с което искат **забрана на смъртоносните автономни оръжия**. Смятат, че роботите са непредсказуеми и фактът, че сега не се правят с цел автономност, не значи, че няма сами да станат такива.

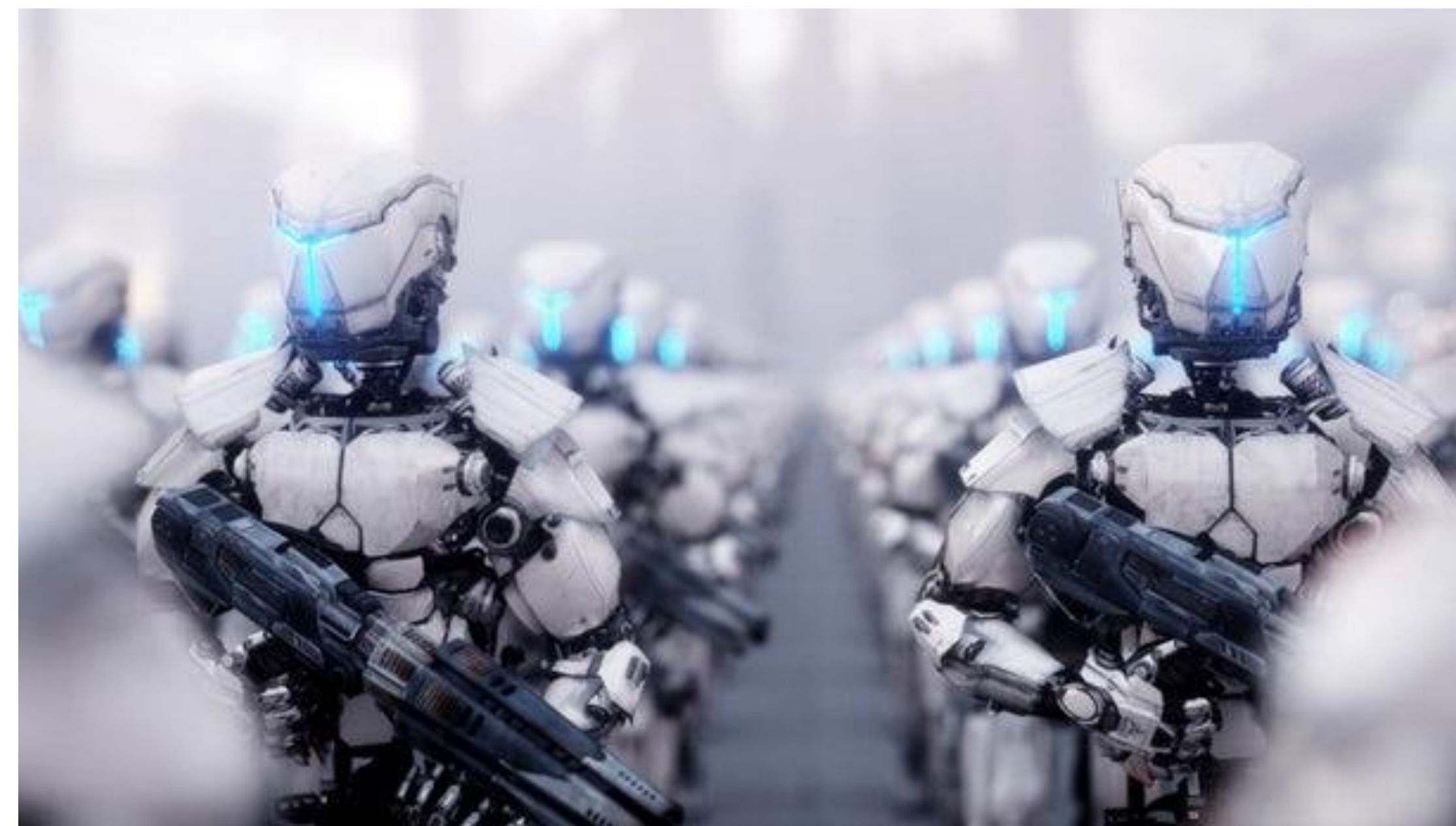




**СЪДЪРЖАНИЕ 3**

# Военните проекти

- Владимир Путин, президентът на Русия, споделя в интервю за Форбс: „Изкуственият интелект е бъдещето не само за Русия, но и за цялото човечество. Той идва с огромни възможности, но и заплахи, които е трудно да се предскажат. Който стане лидер в тази сфера ще стане **владетел на света.**”



**СЪДЪРЖАНИЕ 3**

# Втори сценарий

- ИКС, програмирана да направи нещо полезно, но развива разрушителен метод при постигане на целта - това може да се случи, ако не успеем напълно да синхронизираме целта на ИКС с нашата.





**СЪДЪРЖАНИЕ 3**

## Втори сценарий - пример

Ако помолите интелигентна кола да ви отведе до летището, колкото е възможно по-бързо, тя може да ви закара дотам с риск за останалите участници в движението, като върши буквално това, което сте поискали.

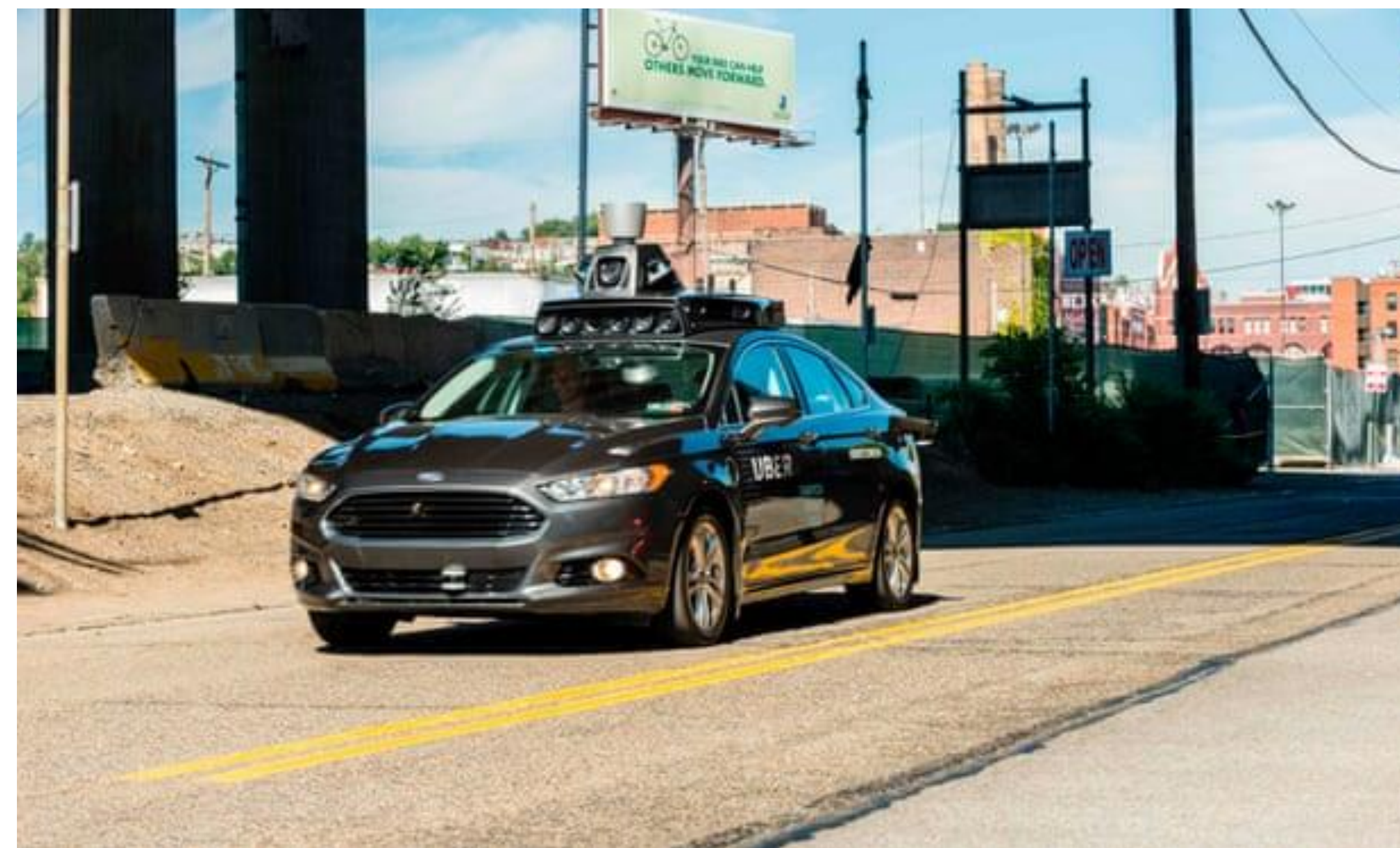




**СЪДЪРЖАНИЕ 3**

## Втори сценарий - пример

Компанията за превози „Uber“ пусна автономни коли и лошите резултати не закъсняха. Налице е първият смъртен случай от автономна кола: жена ударена и загинала докато прекосява улица в Темпе, Аризона.





**СЪДЪРЖАНИЕ 3**

## Втори сценарий - пример

Ако ИКС е ангажирана с амбициозен проект за геоинженерство, тя може да унищожи нашата екосистема като страничен ефект и да види човешките опити да я спре като заплаха, която трябва да бъде преодоляна.





**СЪДЪРЖАНИЕ 3**

## Втори сценарий – ползи и рискове

ИКС стават все по-добри при постигане на целите си, но ако тези цели не са в съответствие с нашите, ще имаме проблем.

Какво би станало ако сме отговорни за проект за водноелектрическа централа и има мравуняк в района, а той трябва да бъде наводнен?





**СЪДЪРЖАНИЕ 3**

# Втори сценарий – ползи и рискове

Основна цел на изследванията за безопасността на ИКС - никога да не се поставя човечеството в позицията на тези мравки.





**СЪДЪРЖАНИЕ 3**

## Втори сценарий – ползи и рискове

Китай разработва социална система, базирана на алгоритми за разпознаване на лица и следене с камери ежедневната дейност на хората. Целта е да се даде за всеки един от 1,4 милиарда граждани на Китай личен резултат, базиран на начина, по който се държи – пресича ли на червено, пуши ли в зони за непущачи, колко време прекарва във видео игри и т.н.

Когато ви наблюдават и след това се вземат решения, базирани на тази информация, това е не само нахлуването в личното пространство, но може и бързо да се превърне в социално потисничество.





**СЪДЪРЖАНИЕ 3**

# Рискове при ИКС

- Загуба на контрол;
- Възможност за манипулации:
  - на обороти;
  - на обществено мнение;
  - на желанията на хората.



**СЪДЪРЖАНИЕ 3**

# Базови рискове

- AI не генерализира - добър е в една определена задача.

**Пример** – Фактът, че AI играе ГО, не означава, че може да шофира кола и обратно.

- Когато AI срещне “загадка” – получава информация, която не е част от обучението му.

**Пример** – Обучен да разпознава хора с бради.



**СЪДЪРЖАНИЕ 3**

# Можем ли да предсказваме бъдещето?

Дори най-мощните суперкомпютри не могат да кажат категорично дали ще вали следващата седмица:

- Невероятна сложност на това, което се опитваме да моделираме;
- Голям брой променливи, които трябва да знаем много точно – скорост и посока на вятъра, слънчева активност, промени в температурата на атмосферата и водните басейни и др.

Тази хаотична непредсказуемост е причина за т.нар. **ефект на пеперудата**, т.е. всяка малка промяна в първоначалните условия може да доведе до множество различни резултати, ако разиваме системата във времето.

**СЪДЪРЖАНИЕ 3**

## Базови рискове - извод

Ако AI се използва за прогнози - осъзнава ли когато не може да направи добро предвиждане на база на информацията, с която разполага?

Човек може да генерализира и поиска помощ, когато не е сигурен, но днешните ИКС не могат.



**СЪДЪРЖАНИЕ 3**

# Съвършена детерминираност

Познаването на точните начални условия в една система би ни позволило да изчислим как се развива тази система: причина и следствие.

**На практика** - невъзможно е да познаваме и контролираме началните условия, както и всички останали продължаващи въздействия, до безкрайна точност, в една дори по-проста система от времето.



**СЪДЪРЖАНИЕ 3**

# Sophia – първият робот с гражданство





**СЪДЪРЖАНИЕ 3**

# Sophia – първият робот с гражданство

- Произведен от Hanson Robotics в Хонконг и активиран на 14 февруари 2016. Направен по модел на актрисата Одри Хепбърн и използва AI за своята комуникация, като е имал зададени първоначални мимики и отговори на някои често срещани теми за разговор, но е програмиран да „поумнява“ с времето.
- Целта - да се използва за компания в домове за възрастни хора и да помага при големи събития, например концерти и т.н.
- През октомври 2017 получава гражданство в Саудитска Арабия, а в началото на 2018 получава крака и възможността да ходи.



**СЪДЪРЖАНИЕ 3**

# Sophia – първият робот с гражданство

Уил Смит на среща със Sophia:

<https://www.youtube.com/watch?v=MI9v3wHLuWI&t=109s>





**СЪДЪРЖАНИЕ 3**

# Sophia – първият робот с гражданство

- **Негативни отзиви** - Sophia има възможността да „поумнява“ с времето, тоест сама развива нови „мисли“ по дадени теми и може всеки път да даде различен отговор на даден въпрос. Нейното „мислене“ става чрез **дърво на достижимост**. Така тя формулира отговора си спрямо казаното ѝ от нейния събеседник. Това само по себе си не е опасно и много хора я наричат чат-бот. До каква степен обаче „разсъждението“ по различни теми е просто следване на дървото на достижимост и до каква е действително „поумняване“? Някои учени смятат, че тази дадена ѝ свобода може да се обърне срещу нас, тъй като Sophia може да стане прекалено самоосъзната с времето.
- Интервю на CNBC със Sophia <https://www.youtube.com/watch?v=78-1MlkxyqI>
- Блогът на Sophia: <https://www.hansonrobotics.com/blog/>





**СЪДЪРЖАНИЕ 3**

# Lil Miquela – първият софтуерен „робот“

- Проект създаден от Brud през 2016г под формата на Instagram профил. В началото използван предимно като маркетингова стратегия за различни марки, в момента е със самостоятелен „живот“.
- Без физическо тяло, моделиран върху изображения на реални хора.
- Към средата на 2022г с 3 мил. последователи в Instagram.
- Представянето на Lil Miquela: I'm Miquela, A Real-Life Robot Mess - <https://www.youtube.com/watch?v=6bn3tUUtj2M>





**СЪДЪРЖАНИЕ 3**

# Роботите ще закрият ли работни места?





**СЪДЪРЖАНИЕ 3**

# Кои работни места ще бъдат заети от роботи?

- **Работници;**
- **Обслужващ персонал** - франчайзът на McDonald's във Финикс почти изцяло е управляван от роботи;
- **Преводачи** - Google Translate се подобрява благодарение на невронната си мрежа.
- **Счетоводители** - SMACC финансово дружество, осигурява \$ 3,5 млн. за създаване на своя счетоводна система с AI;
- **Адвокати** - международната правна кантора Baker&Hostetler нае първия адвокат с AI, наречен ROSS;
- **Писатели** - роман, написан от изкуствен интелект, почти печели литературна награда. EMMA, детектора за плагиатство, лесно може да идентифицира авторството, ако към него се качват определен брой документи.
- ...



**СЪДЪРЖАНИЕ 3**

# Ще убият ли роботите работни места?

- Дайсуке Адачи и колеги от университета Йейл разглеждат японското производство между 1978 и 2017г- откриват, че увеличението с един робот на 1000 работници повишава заетостта на фирмите с 2.2%.
- Йонас Тукури от Масачузетския технологичен институт (MIT) и колеги, разглеждат финландските фирми - заключава, че въвеждането на модерни технологии в тях е довело до увеличаване наемането на служители.
- Непубликувана работа на Майкъл Уеб от Станфордския университет и Даниел Чандлър от Лондонското училище по икономика - разглежда машинните инструменти в британската индустрия и установява, че автоматизацията има "силна положителна връзка с оцеляването на фирмата и че по-голямата първоначална автоматизация е свързана с увеличаване на заетостта“.



**СЪДЪРЖАНИЕ 4**

# Квантовите компютри и науката през XXI век

- **Квантовите битове (кюбити/кубити)** – не са ограничени само до 2 възможни състояния, а могат да съществуват в квантова суперпозиция едновременно на 0 и 1, т.е. могат да запазват много повече информация.
- **Предимство** - много вплетени заедно кюбити могат да обработват множество операции едновременно.
- **Проблеми пред реализацията:**
  - такива състояния могат да се поддържат само при специални условия и за много кратко време;
  - да се контролира входът и изходът на информацията, която кюбитите обработват поради увеличения им брой.
- **Работещи по проблема** – IBM, Google, стартиращи компании.
- Квантовите компютри ще се нуждаят от собствен софтуер и алгоритми (алгоритъм на Шор за разделяне на число на прости множители, алгоритъм за търсене на Гроувър и др.).



# Благодаря ВИ.

