

Презентация

Садържание

- Въведение
- Основни принципи
- Вгразждането на ценности в автономните интелигентни системи
- Методология за ръководенето на етични изследвания и проектиране
- Безопасност и облаги от AGI и ASI
- Лична информация и достъпът до нея
- Автономни оръжеини системи
- Икономически и хуманитарни проблеми
- Законови уредби

Въведение

Какво е изкуствен интелект (AI)???



Какво е машиното обучение
(Machine learning)?

Основни принципи

- Първи принцип. Облаги за хората
- Втори принцип. Отговорност
- Трети принцип. Прозрачност
- Четвърти принцип. Обучение и осведоменост

Основни принципи

Първи принцип. Облаги за хората

Втори принцип. Отговорност

Трети принцип. Прозрачност

Четвърти принцип. Обучение и осведоменост

1 Как да сме сигурни, че няма AI да наруши човешките права?

1 Да уважава човешките права, свобода, достойнство, и култура.

1 Ако нарани човек, винаги да се открие причината за взетото действие.

2 Как да държим отговорност на AI?

Основни принципи

Първи принцип. Облаги за хората

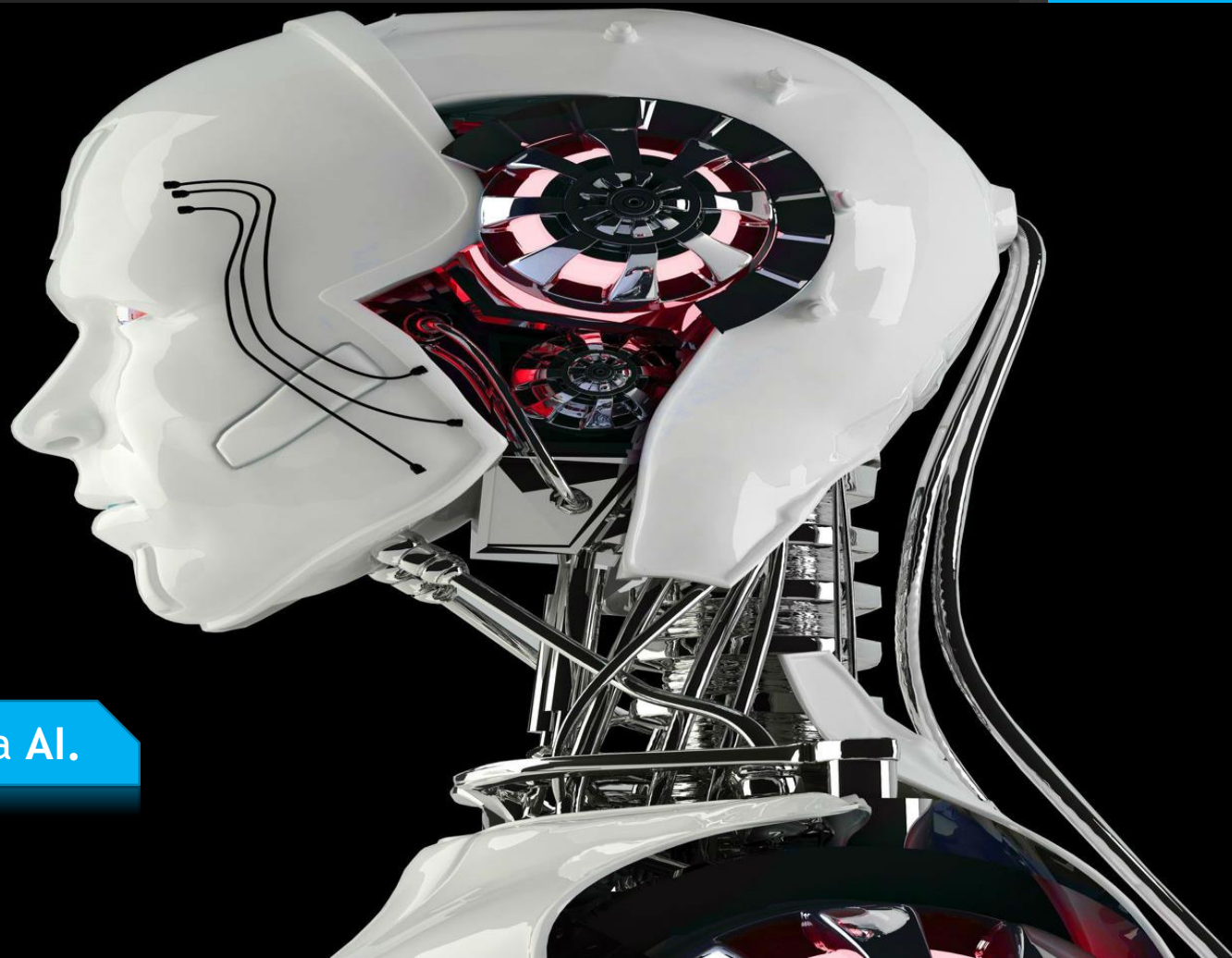
Втори принцип. Отговорност

Трети принцип. Прозрачност

Четвърти принцип. Обучение и осведоменост

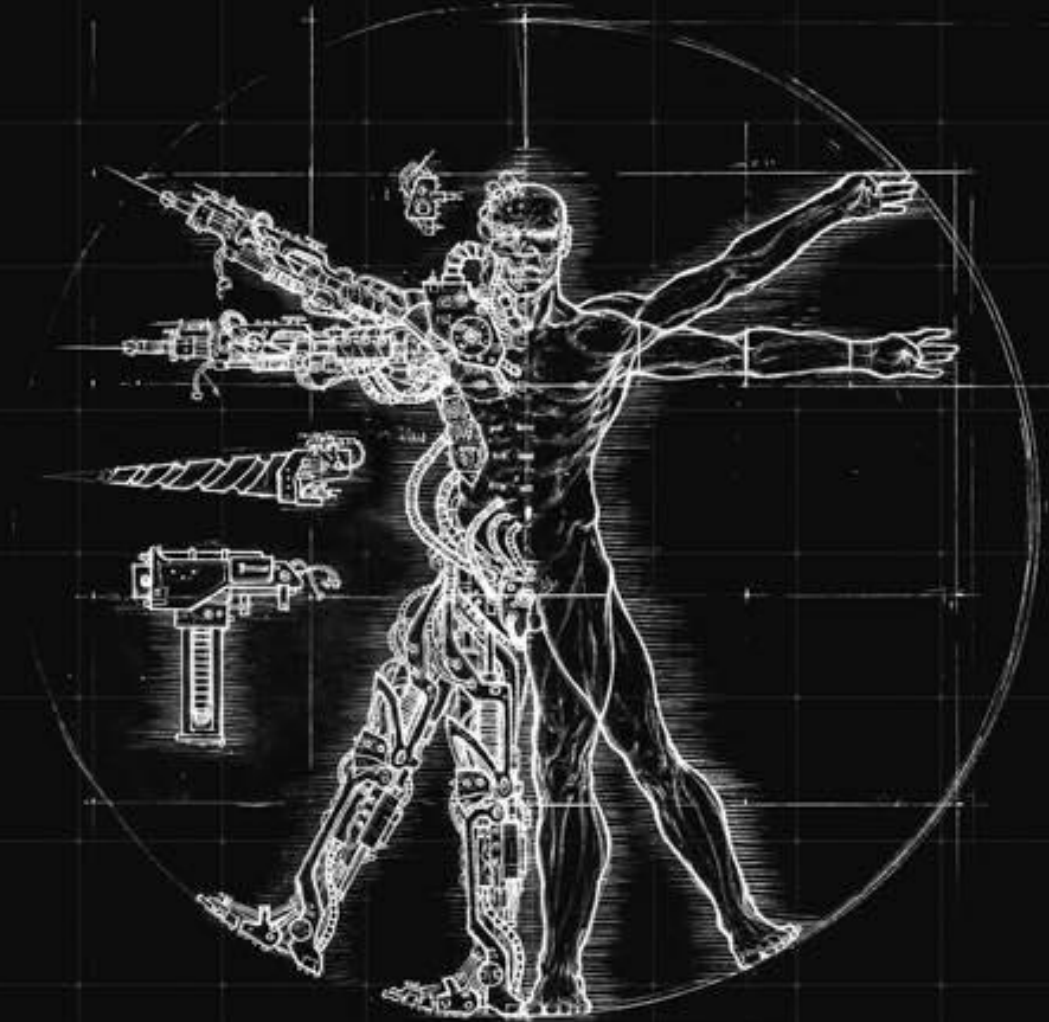
3 Прозрачност при AI.

4 Осигуряване на неправилно използване на AI.



Вграждането на ценности в автономните интелигентни системи

- Определяне на нормите и ценностите за автономните интелигентни системи
- Вграждането на нормите и ценностите за автономните интелигентни системи
- Оценяване на нормите и ценностите за автономните интелигентни системи



Вграждането на ценности в автономните интелигентни системи

Определяне на нормите и ценностите за автономните интелигентни системи
Вграждането на нормите и ценностите за автономните интелигентни системи
Оценяване на нормите и ценностите за автономните интелигентни системи

1 Ценостите и нормите в AI трябва да бъдат не универсални а специфични за обществата които обслужват

2 AI са обект на множество норми и ценности които могат да си противоречат

Вграждането на ценности в автономните интелигентни системи

Определяне на нормите и ценностите за автономните интелигентни системи
Вграждането на нормите и ценностите за автономните интелигентни системи
Оценяване на нормите и ценностите за автономните интелигентни системи

2 AI могат да дискриминират част
от населението поради нормите



3 Как да
имплементи
раме
нормите в
AI?

Методология за ръководенето на етични изследвания и проектиране

- Обучение и проучване
- Бизнес процедури и изкуствения интелект
- Липса на прозрачност

Методология за ръководенето на етични изследвания и проектиране

Обучение и проучване
Бизнес процедури и изкуствения интелект
Липса на прозрачност

1 Етиката в обучението.

2 Лидерите оставят на инженерите малко или никакви напътствия за човешките ценности които да бъдат имплементирани в AI и какви са те



Методология за ръководенето на етични изследвания и проектиране

Обучение и проучване
Бизнес процедури и изкуствения интелект
Липса на прозрачност

3 Липса на собственост и отговорност от тек обществото.

3 Липса на независим контролен орган.

3 Липса на документация за етическия дизайн.

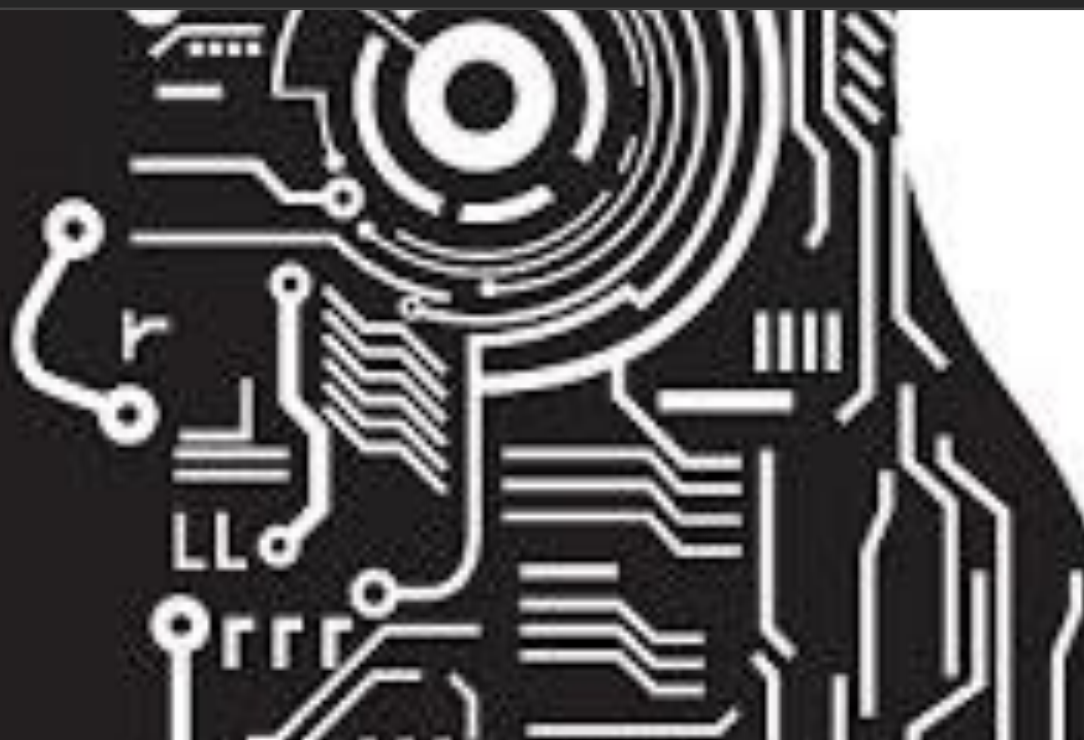


Защо България няма да прави
внос на автономни автомобили?

Безопасност и облаги от AGI и ASI

Техническа информация
Основни принципи

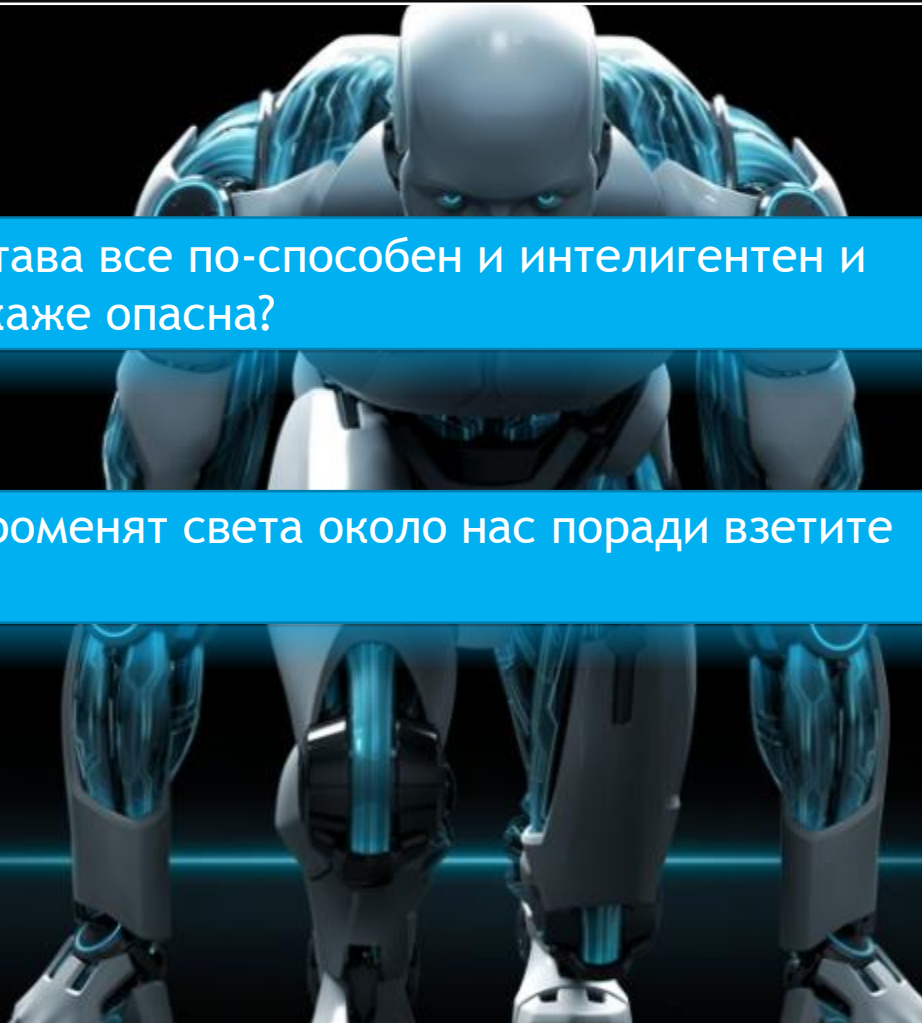
- Техническа информация
- Основни принципи



Безопасност и облаги от AGI и ASI

1 Изкуственият интелект става все по-способен и интелигентен и автономията може да се окаже опасна?

2 Тези системи могат да променят света около нас поради взетите решения



Лична информация и достъпът до нея

- Дефиниция на личните данни
- Достъп и съгласие за личната информация
- Управление на личната информация

Лична информация и достъпът до нея

Дефиниция на личните данни
Достъп и съгласие за личната информация
Управление на личната информация

1 Как човек да дефинира и организира своите лични данни в тази ера?

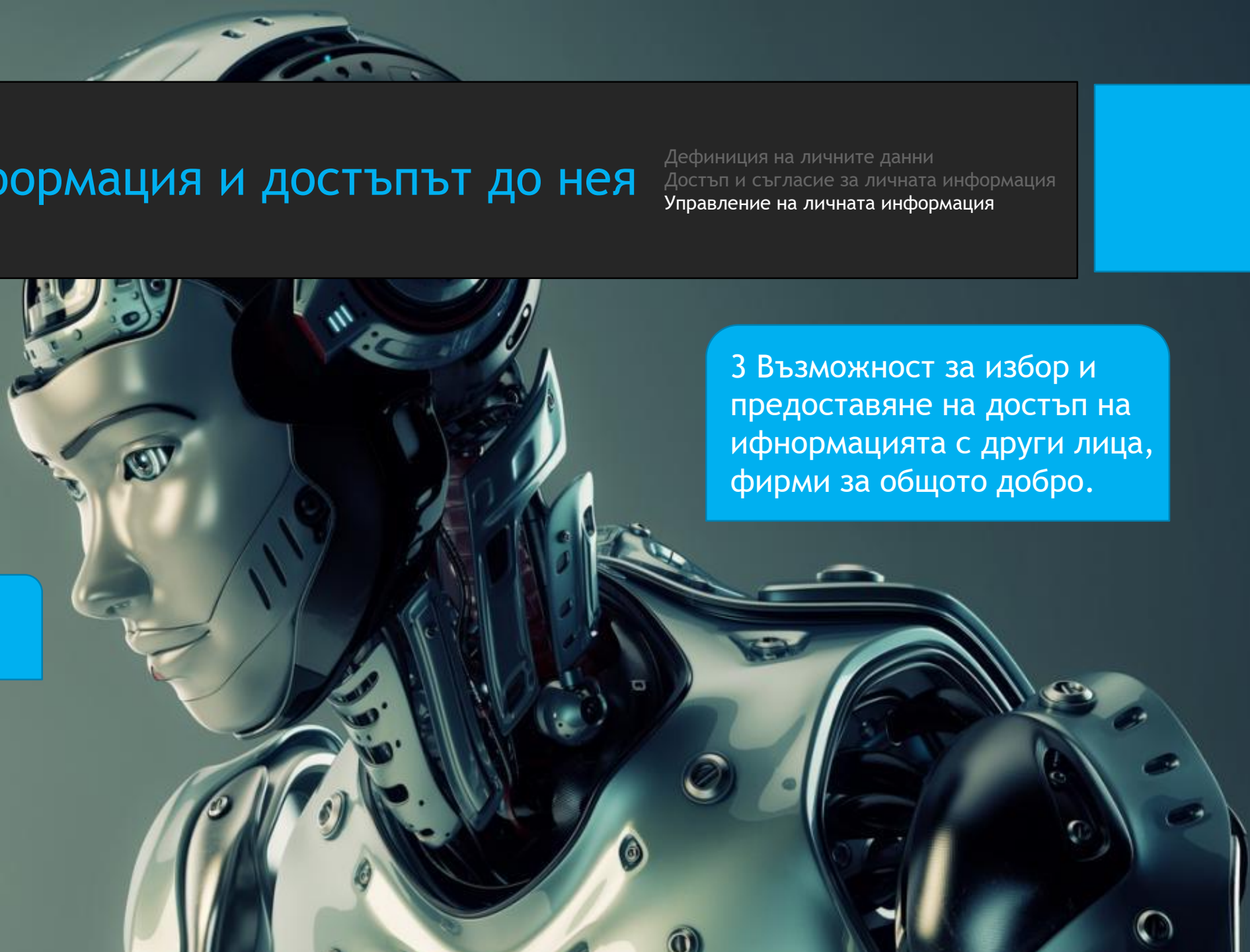
2 Предоставен достъп на всеки човек за своите лични данни.

Лична информация и достъпът до нея

Дефиниция на личните данни
Достъп и съгласие за личната информация
Управление на личната информация

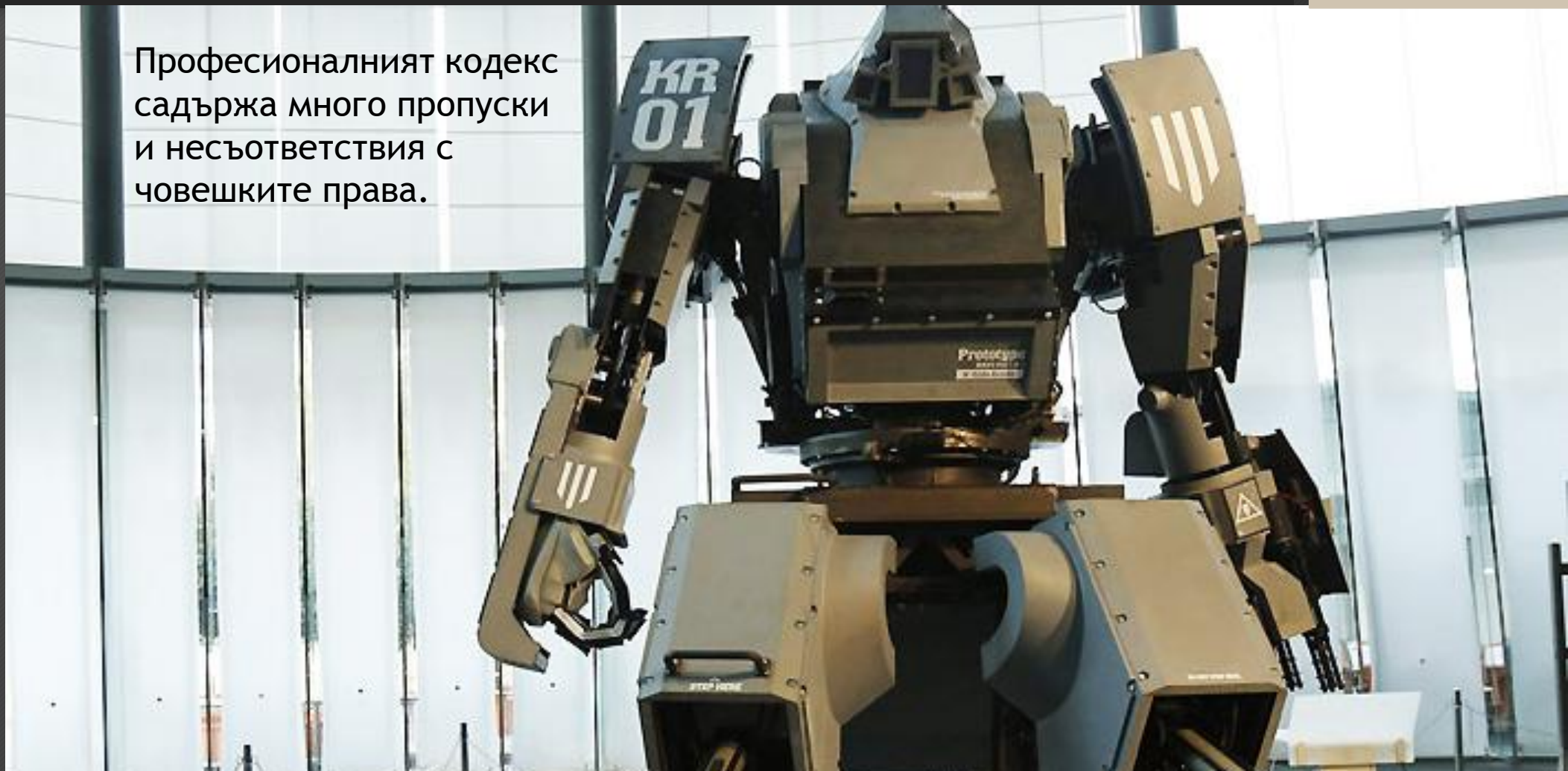
3 Възможност за избор и предоставяне на достъп на информацията с други лица, фирми за общото добро.

Може ли изкуствения интелект да ни бъде пазител?



Автономни оръжеини системи

Професионалният кодекс съдържа много пропуски и несъответствия с човешките права.



Автономни оръжейни системи

Изключването на надзора може да доведе до повишаване на напрежението и появата на военни действия

Автономните военни системи преодоляват човешките ценности, което се превръща в конфликт на системите.

Не съществуват стандарти за тестване и валидация на такива системи.



Важни забележки

Q: What do you call a blonde who has dyed her hair brown?



Икономически и хуманитарни проблеми

- Автоматизация и набирането на персонал
- Отговорност и разпредление.
- Упълномощаване на увеличаването на правата на развиващите се страни да се възползват от AI.

Икономически и хуманитарни проблеми

Автоматизация и набирането на персонал
Отговорност и разпредление
Упълномощаване на увеличаването на правата на развиващите се страни да се възползват от AI.

1 Неразделянето на изкуствен интелект и автономни системи в медията е объркващо за зрителите.

1 Промяната се случва твърде бързо за съществуващите методи за преобучение на персонала.

Икономически и хуманитарни проблеми

Автоматизация и набирането на персонал
Отговорност и разпделение
Упълномощаване на увеличаването на правата на развиващите се страни да се възползват от AI.

2 Липса на дефинирането на личната информация.

2 Технологиите не са разпределен равномерно из света.

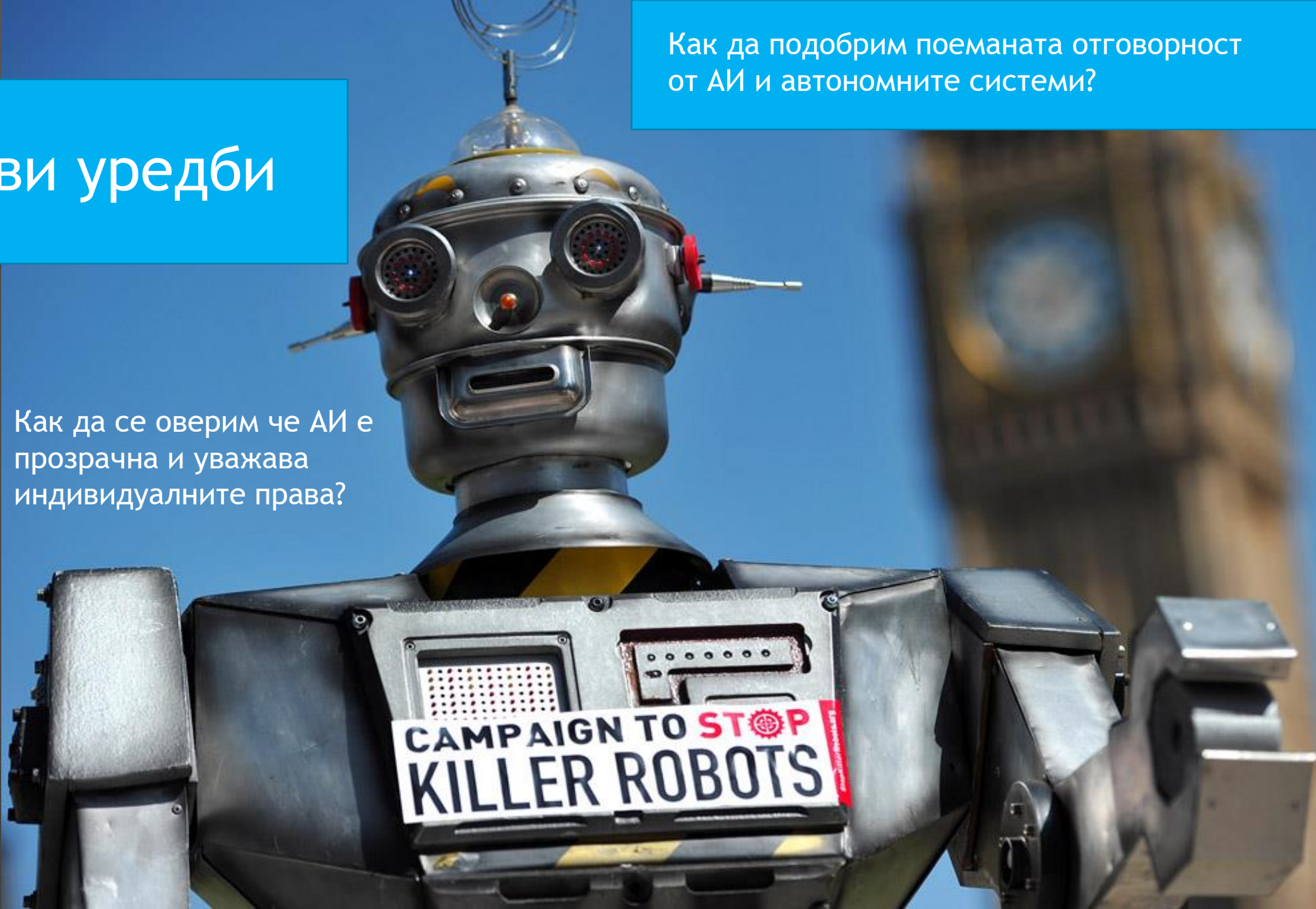
3 Нужда от катализатор?



Законови уредби

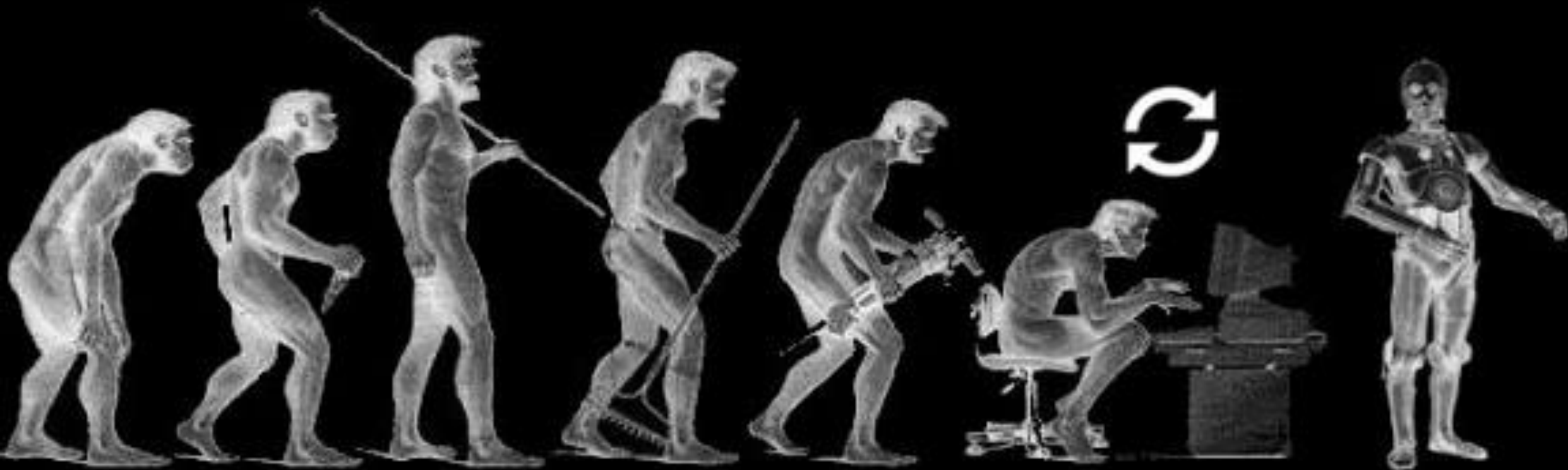
Как да подобрим поеманата отговорност от AI и автономните системи?

Как да се оверим че AI е прозрачна и уважава индивидуалните права?



Настоящето и бъдещето?

Защо системата за разпознаване на глас на Google работи по-добре при мъжки екземпляри.



Проблеми при операция на хора.

Небалансираны зададени параметри при обучението

Настоящето и бъдещето?

50 процента от компаниите
планират да имплантират AI

Увеличение на присъствието на
AI във фирмите и индустрията

Истинската опасност е, че хора проектират системите, а хората грешат.

Google



Explore machine learning with Teachable Machine

In this AI Experiment, you can train a machine using your camera — no coding required.

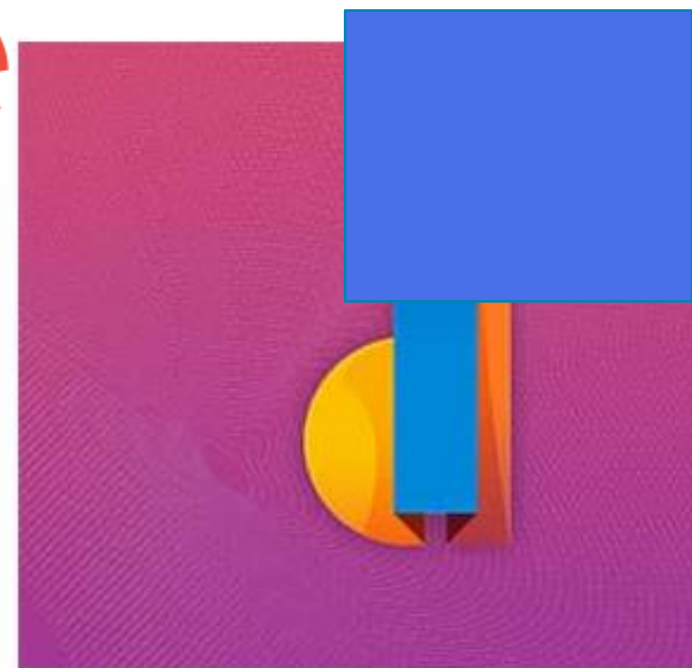
[LEARN MORE](#) 



Make your own devices that see

Our next AIY kit lets makers experiment with computer vision.

[LEARN MORE](#) 



Harness the power of machine learning in your browser

Meet deeplearn.js, our open-source library for machine learning, made for the web.

[LEARN MORE](#) 

Baidu 百度

Планират да пуснат автономни коли до 2018 и да произведат масово до 2021 г.

Автономните коли ще спасяват животи

Baidu са оптимисти за бъдещето на AI и автономните автомобили





TESLA

Tesla използва машино обучение за събиране на информация за типа на шофиране от водача



Новият “Тесла“ модел S
във Флорида бил опасен,
защото не могъл да
различи бял камион и
ясното небе.



Важни забележки

Human: What do we want!?
Computer: Natural language processing!
Human: When do we want it!?
Computer: When do we want what?



Alexa и детските песнички

Alexa пуска видео за възрастни вместо детски песнички

Когато едно малко дете поиска от Алекса да пусне любимата си песен - "Дигер, Дигер", Алекса чу нещо друго. В отговор на искането Алекса каза: "Искате да чуete станция за "еротика", открита ... гореща мацка, аматьорка секси момиче."



Chatbot на Microsoft се превръща в анти-феминист и про-Хитлер.

Той се превърна във вулгарен, про Хитлеристки туйтър аканут и признавайки феминизма като рак, пожелавайки последователите им да горят в ада



През март 2016 г. Microsoft представи своя AI Twitter chatbot "Тей". Експериментирайки с «стандартен разговор», Тей трябваше да общува с хората и да стане по-умен, колкото повече се занимаваше и разговаря с хората

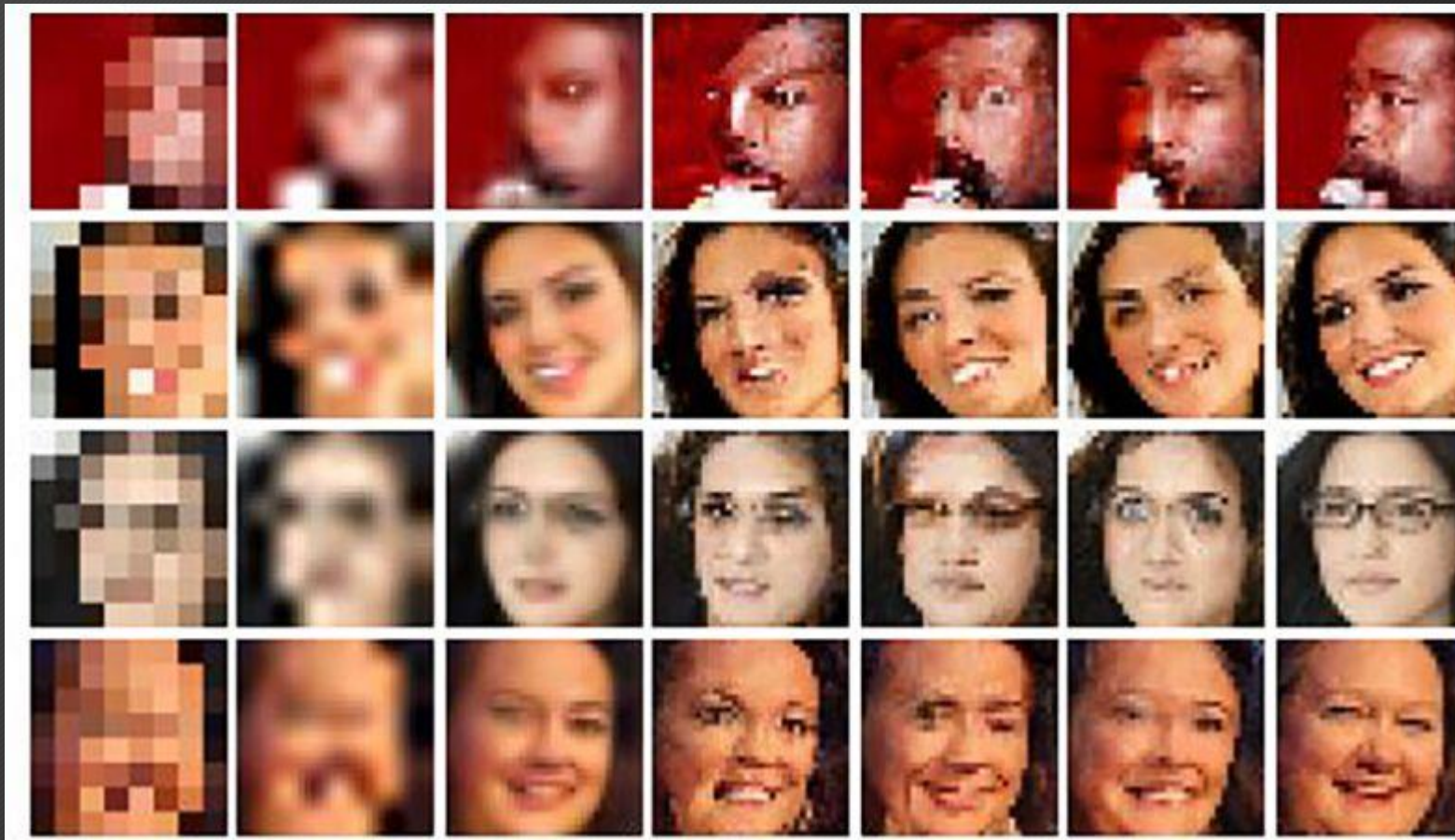
Роботът не успява да влезе в колеж.



През 2011 г. екип от изследователи започна да работи върху робот, наречен "Тодай робот", който те са предназначени да бъдат приети в японския конкурентен университет в Токио. След като е взел японския изпит за прием на национални университети през 2015 г., роботът не успя да получи достатъчно висока оценка, за да бъде приета в колежа.

Google Brain превръща нискокачествени изображения в пикселирани чудовища

В опит да подобри картинките с ниска резолюция, Google Brain в някои случаи ги превърна в чудовищни образи на хора.



Важни забележки



THE INTERNET OF EVERYTHING

Въпроси?

I TOLD YOU TO PICK UP MILK ON THE WAY HOME. DON'T YOU LISTEN TO ANYTHING I SAY?!!





БЛАГОДАРИМ ЗА
ВНИМАНИЕТО