Умные живут дольше

Искусственный интеллект и кибербезопасность

Андрей Масалович

Andrei Masalovich

am@avl.team

Live smart, live longer

On modern intelligent cyberweapon

УМНЫЕ ЖИВУТ ДОЛЬШЕ.

Технологии разума в современном кибероружии

- Искусственный интеллект - это не только умные дома и умные города, но и технически совершенное автономное кибероружие. Новую войну будут вести армии умных ботов, способных не только к групповой координации без участия человека, но и к самостоятельной выдаче целеуказаний.
- В докладе рассматриваются решения из сферы ИИ, используемые в современном кибероружии: генеративно-состязательные нейронные сети (GAN) для распознавания новых видов кибератак, методики глубокого обучения с подкреплением (DRL) для агентного моделирования информационных атак, методы «цифровых двойников» для исследования различных физических и психологических воздействий без проведения тестовых атак.

Live smart, live longer.

On modern intelligent cyberweapon

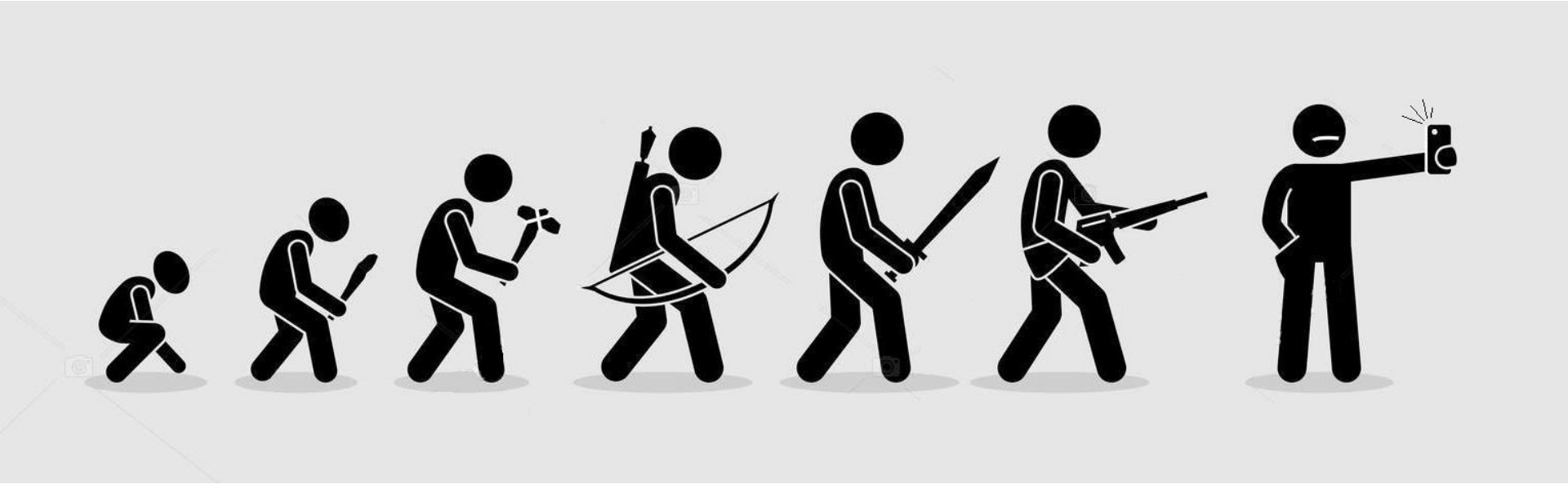
Mutual interaction between telecommunications and artificial intelligence led to the development of smart homes and cities—and to the emergence of the new generation of technically advanced and human-independent cyberweapon.

The army of smart robots will lead the new war, since they will be capable of organizing, coordinating, and target assigning.

The speaker will present artificial intelligence solutions used in modern cyberweapon: generative adversarial networks, deep reinforcement learning, and digital twins..

Мы – дети в мире умных вещей

Военные – дети с гранатой



- Высокоточное оружие
- Умное оружие
- Автономное летальное оружие
- Сетецентрическая война

Восстание машин: в Китае робот оштрафовал фотографию

Тест Тьюринга

- Ученый: Искусственный разум – тот, который при общении неотличим от живого человека
- Хакер: Я боюсь не того компьютера, который пройдет тест Тьюринга, а того, который его намеренно завалит...
- Политтехнолог: Задача искусственного разума – убедить экзаменатора, что он сам компьютер

The Malicious Use of AI

Security Domains

The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation
February 2018

- **Digital security.** The use of AI to automate tasks involved in carrying out cyberattacks will alleviate the existing tradeoff between the scale and efficacy of attacks. This may expand the threat associated with labor-intensive cyberattacks (such as spear phishing). We also expect novel attacks that exploit human vulnerabilities (e.g. through the use of speech synthesis for impersonation), existing software vulnerabilities (e.g. through automated hacking), or the vulnerabilities of AI systems (e.g. through adversarial examples and data poisoning).
- **Physical security.** The use of AI to automate tasks involved in carrying out attacks with drones and other physical systems (e.g. through the deployment of autonomous weapons systems) may expand the threats associated with these attacks. We also expect novel attacks that subvert cyber-physical systems (e.g. causing autonomous vehicles to crash) or involve physical systems that it would be infeasible to direct remotely (e.g. a swarm of thousands of micro-drones).
- **Political security.** The use of AI to automate tasks involved in surveillance (e.g. analysing mass-collected data), persuasion (e.g. creating targeted propaganda), and deception (e.g. manipulating videos) may expand threats associated with privacy invasion and social manipulation. We also expect novel attacks that take advantage of an improved capacity to analyse human behaviors, moods, and beliefs on the basis of available data. These concerns are most significant in the context of authoritarian states, but may also undermine the ability of democracies to sustain truthful public debates.

Пример AI: Нейросеть распознает человека по клавиатурному почерку

Аутентификаторы:

- Уникальное знание
- Уникальный предмет
- Уникальная характеристика

Клавиатурный почерк - поведенческая биометрическая характеристика, которую описывают следующие параметры:

- **Скорость ввода** - количество введенных символов разделенное на время печатания
- **Динамика ввода** - характеризуется временем между нажатиями клавиш и временем их удержания
- **Частота возникновение ошибок** при вводе
- **Использование клавиш** - например, какие функциональные клавиши нажимаются для ввода заглавных букв

Корнеев В.В., Масалович А.И и др. Распознавание программных модулей и обнаружение несанкционированных действий с применением аппарата нейросетей Информационные технологии N10, 1997 - http://sci-pub.info/ref/321545/

Cyberweapon, кибероружие

- A cyberweapon is a malware agent employed for military, paramilitary, or intelligence objectives.
- Кибероружие – программное обеспечение или оборудование, предназначенные для нанесения ущерба в киберпространстве.

(wikipedia.org)

- Кибернетика - наука об управлении государством, которая должна обеспечить гражданам доступ к разнообразным благам современного общества

Андре-Мари Ампер (1830)

Physical Security

Фитнес трекер Strava
выдал расположение военных баз США

Умный пистолет Armatix IP1 – стреляет только в руках владельца

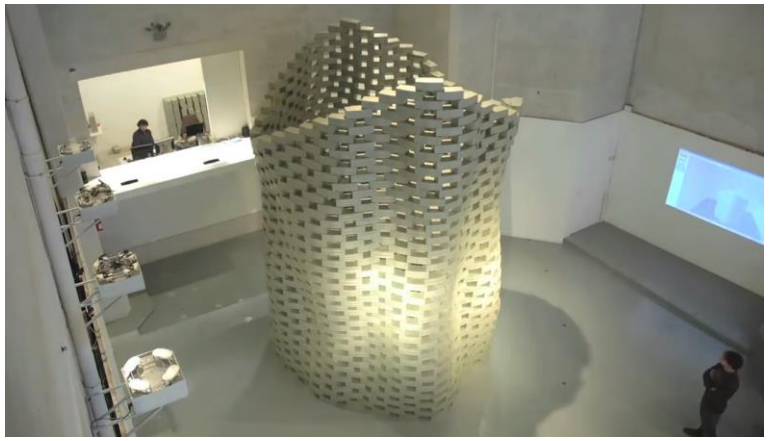
Хакер Plore:

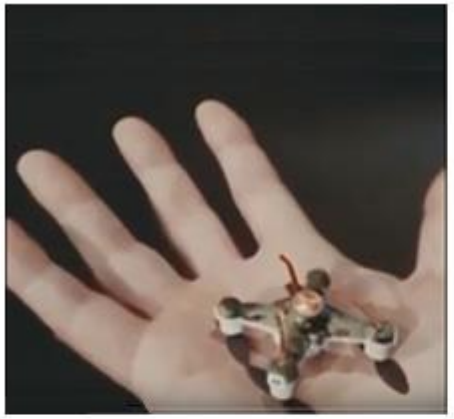
Я знаю как минимум три способа взломать Armatix

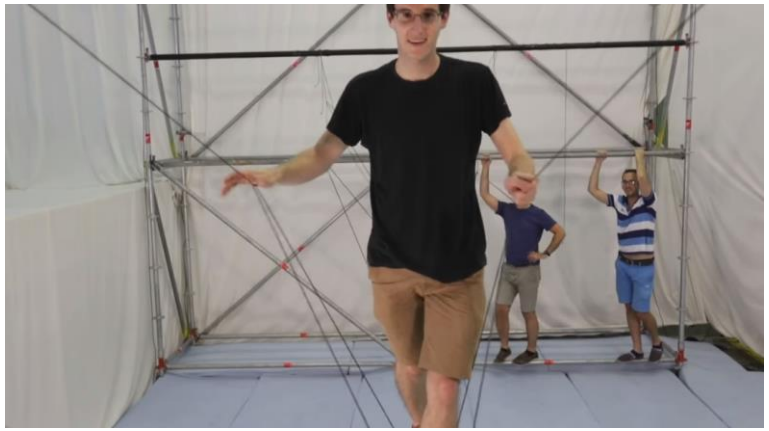
TrackingPoint XS1 —винтовка под Linux

Meet the dazzling flying machines of the future…

Дроны убивают людей



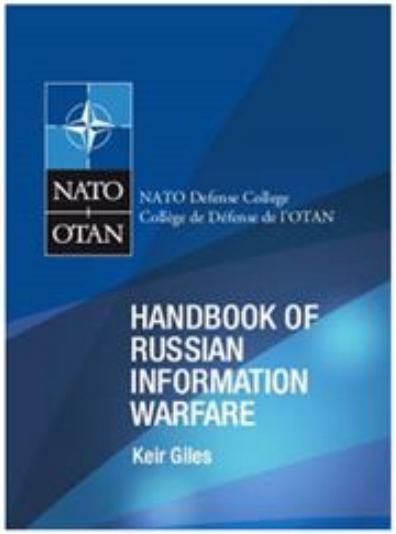


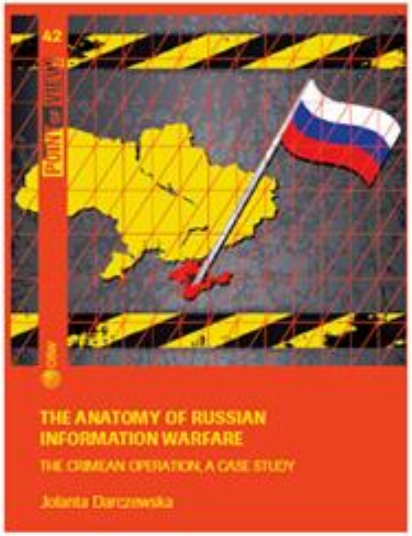


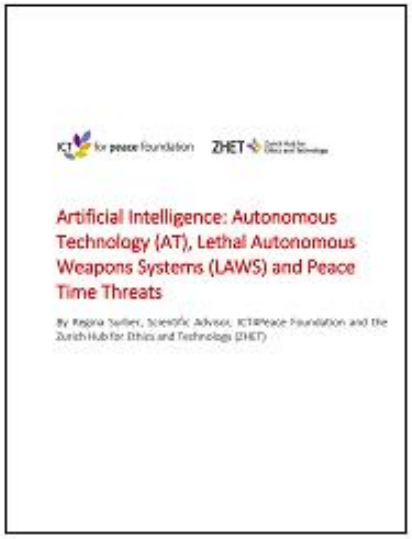
https://www.youtube.com/watch?v=RCXGpEmFbOw

https://www.youtube.com/watch?v=HCG_Hnv7nMY

Оружие новой войны



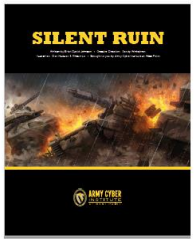




Учите материальную часть!

Silent Ruin

война 2027 года в комиксах армии США



If we fail to harden our own systems while they develop innovative means to counter the proliferation of advanced autonomous systems, this threatens to render the world's most technologically advanced force its most vulnerable.

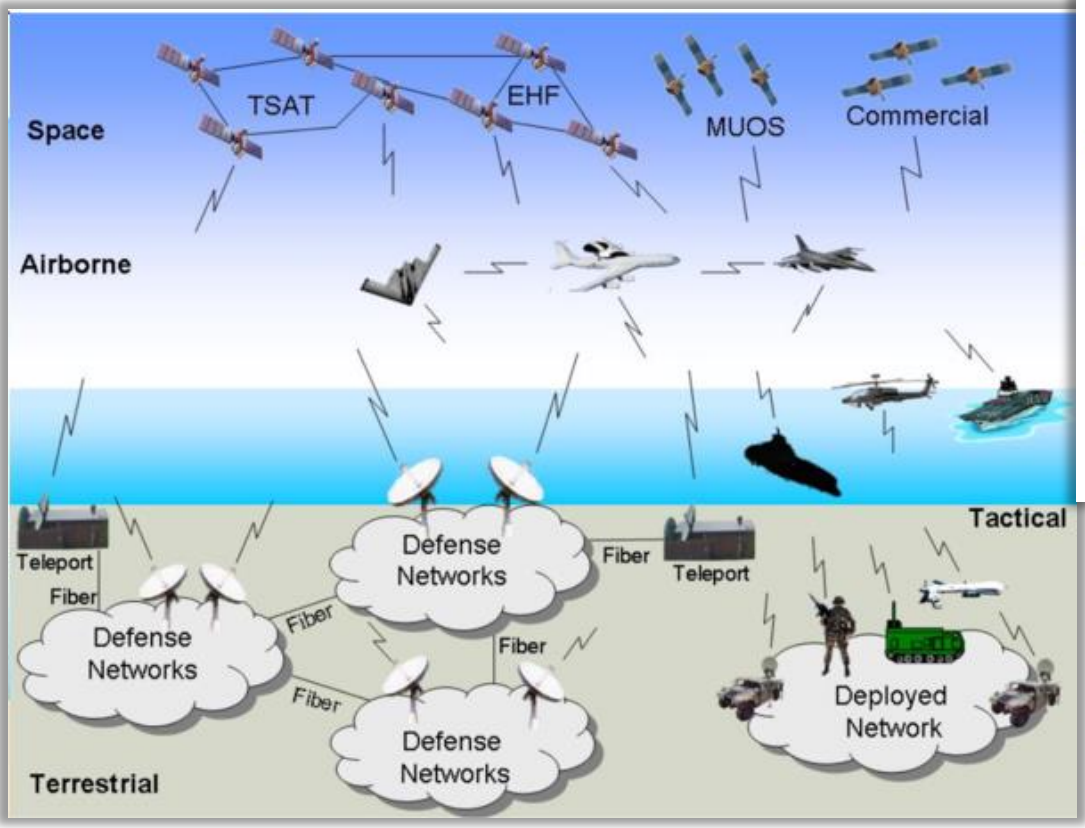
Lt. Col. **J. Lane**, U.S. Army

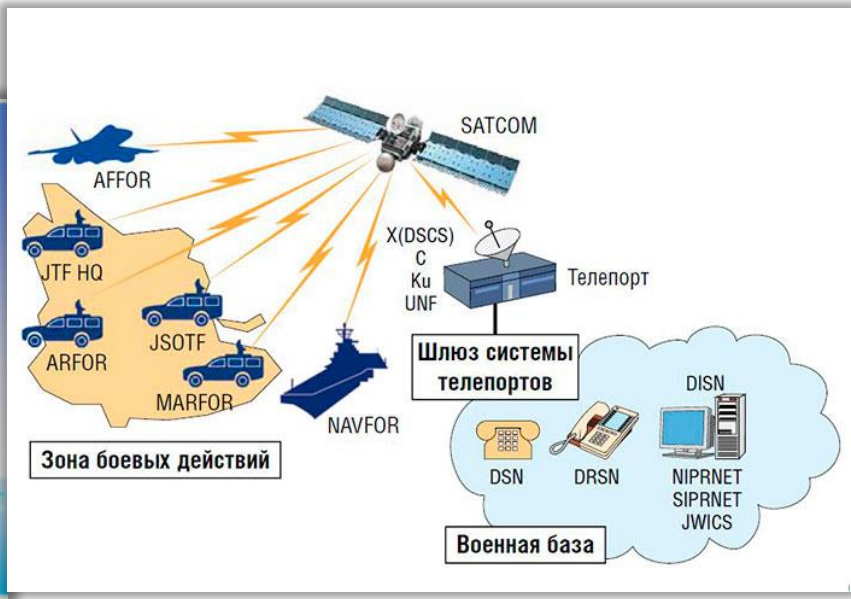
Если нам не удастся разработать наши собственные системы противодействия распространению современных автономных систем, это грозит сделать наиболее технологически развитую силу в мире наиболее уязвимой.

Подполковник **J. Lane**, армия США

Источник: Институт кибернетики армии США, Академия Вест Пойнт https://ru.scribd.com/document/374846182/Silent-Ruin

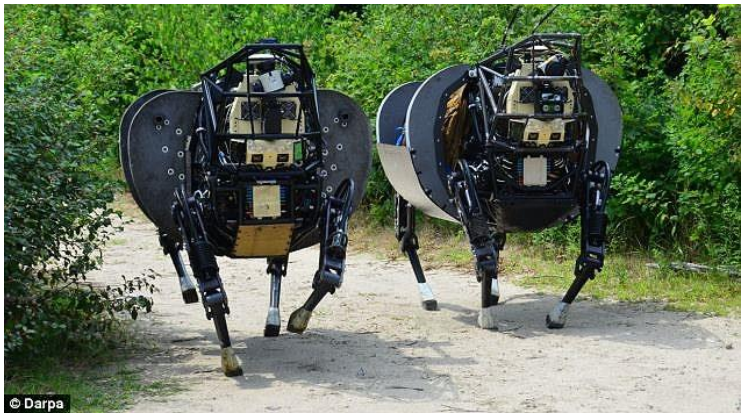
DISN (Defense Information Systems Network)
GIG (Global Information Grid)





Автономное оружие

Lethal Autonomous Weapon Systems



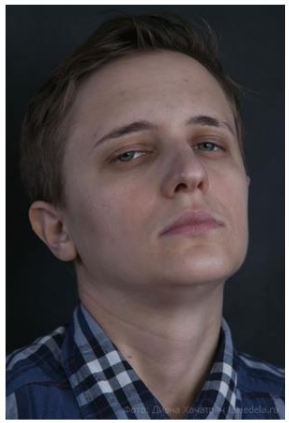
Digital Security

Обмануть нейросеть

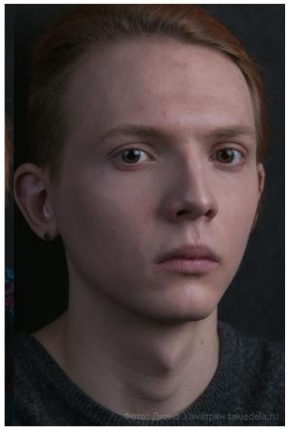
Man or Woman ?

Как обмануть систему распознавания

Adversarial training (состязательное обучение)





Источник: https://pikabu.ru/story/yeto_muzhchina_ili_zhenshchina_4440498

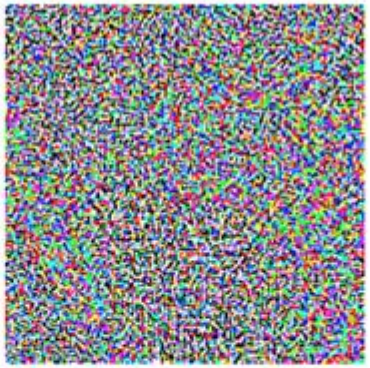
Adversarial Examples

вредоносные примеры

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



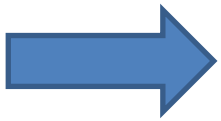
$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

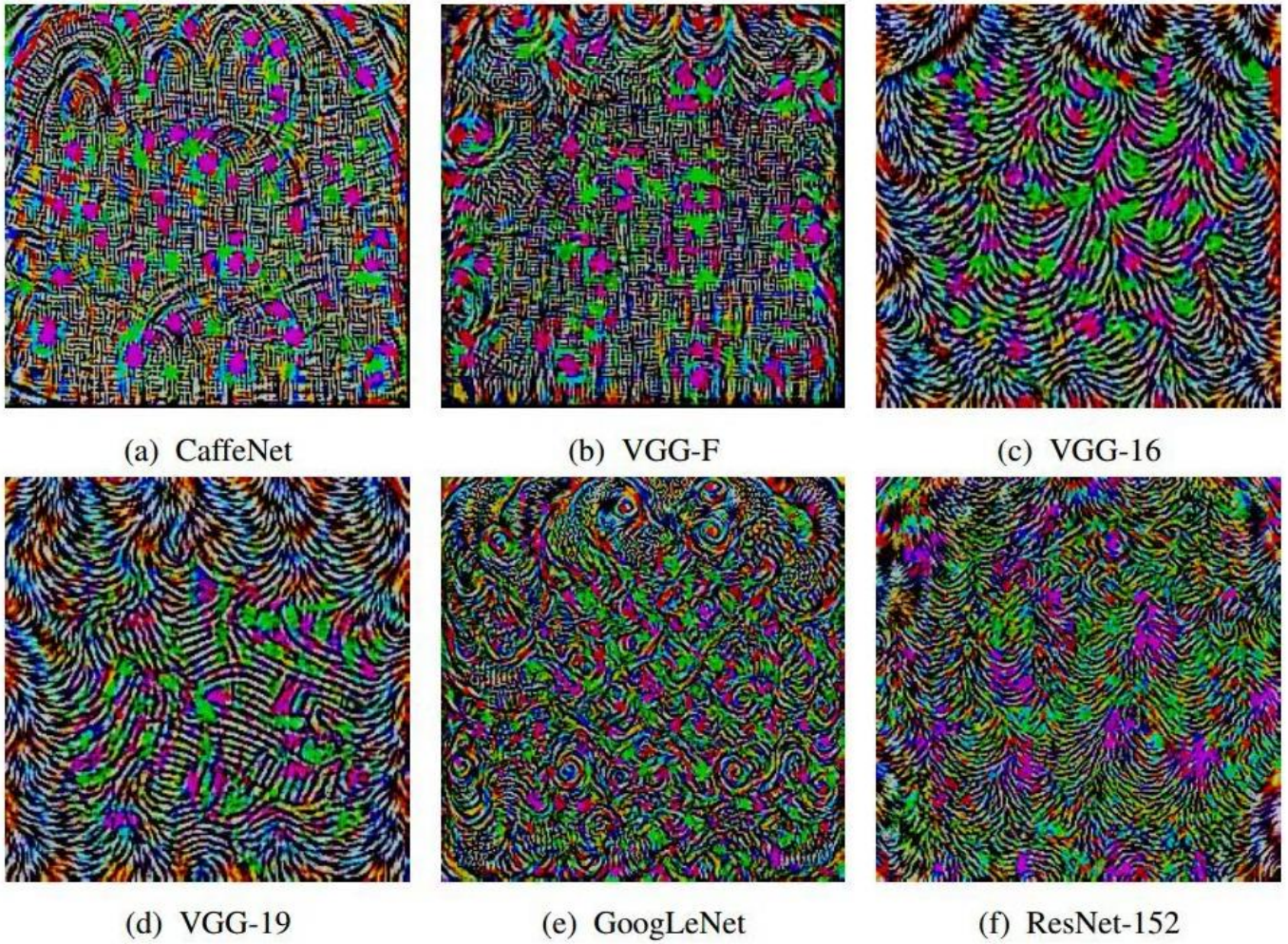
"gibbon"
99.3 % confidence



Источник: https://buzzrobot.com/4-ways-to-easily-fool-your-deep-neural-net-dca49463bd0

The "universal adversarial preturbation"

универсальное искажение

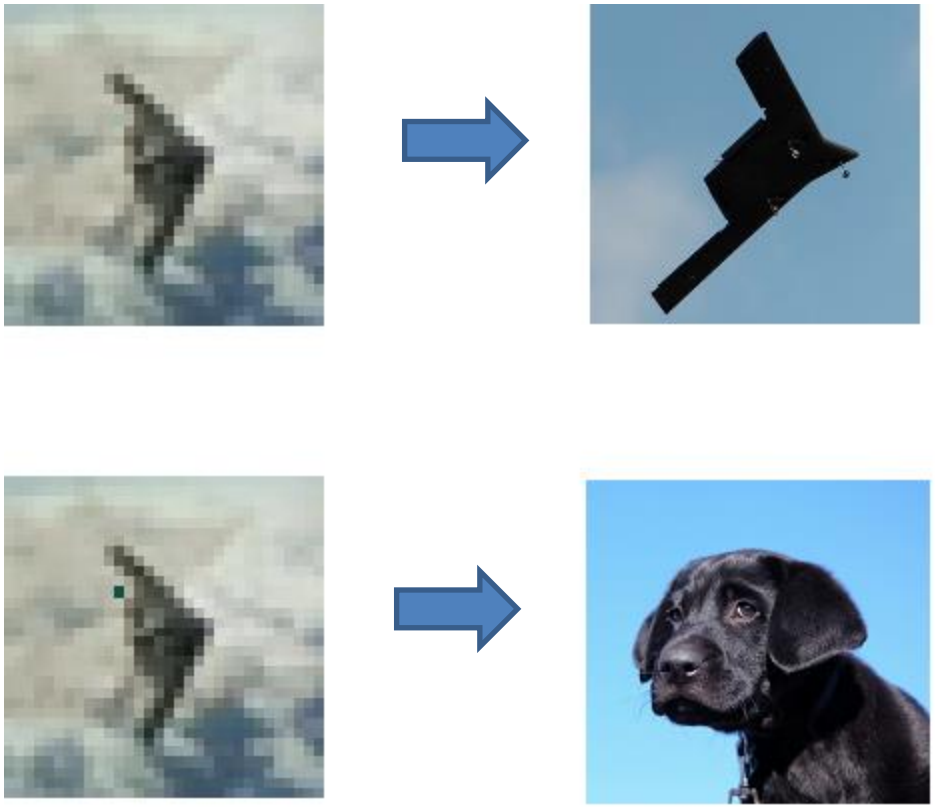


(a) CaffeNet (b) VGG-F (c) VGG-16

(d) VGG-19 (e) GoogLeNet (f) ResNet-152

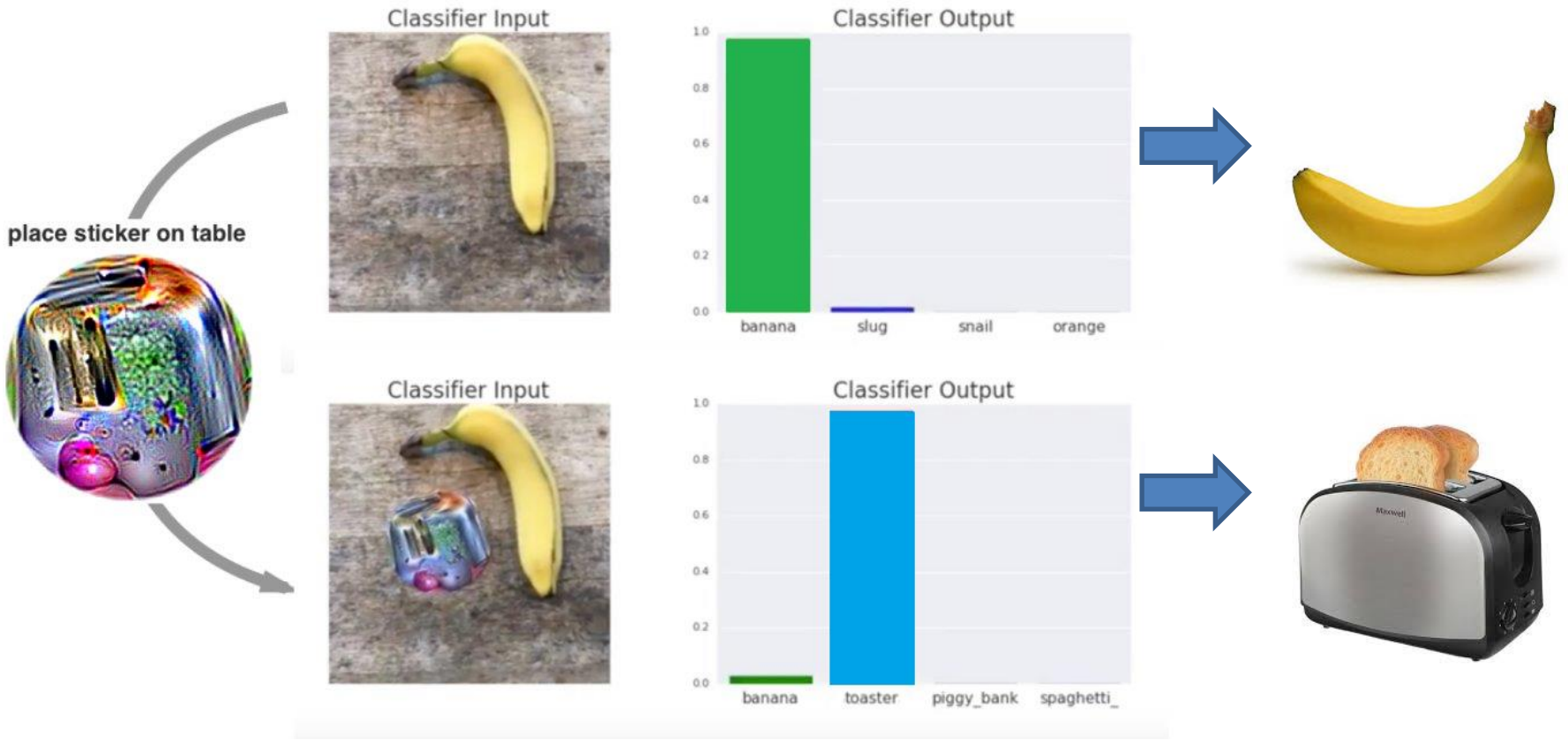
One Pixel Attack

for fooling deep neural networks



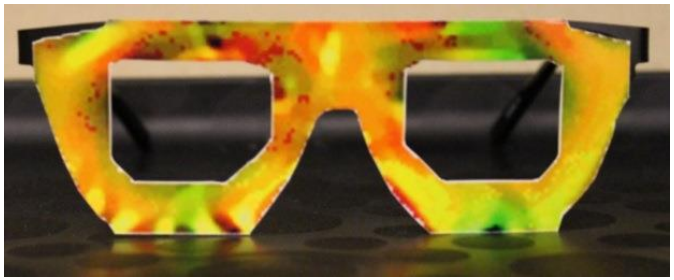
Источник: https://buzzrobot.com/4-ways-to-easily-fool-your-deep-neural-net-dca49463bd0

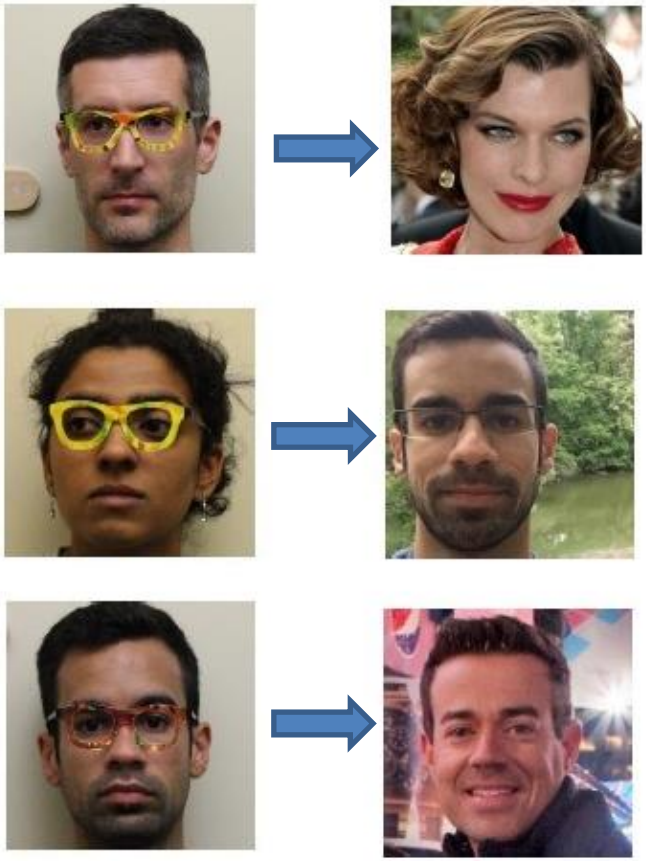
The Adversarial Patch
вредоносная заплатка

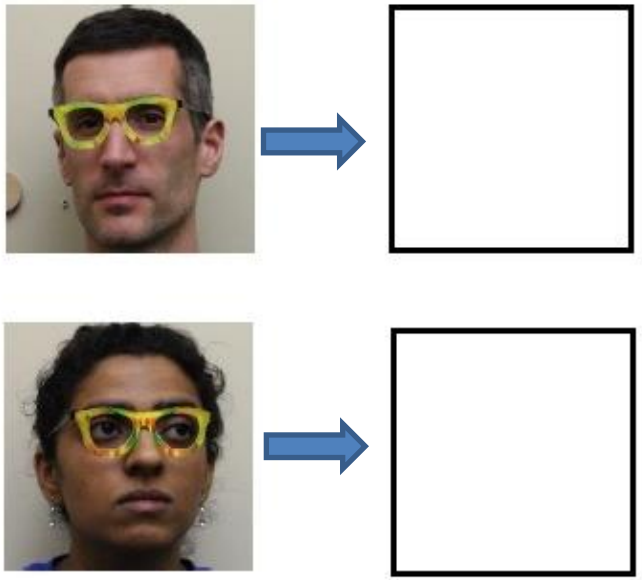


Источник: https://boingboing.net/2018/01/08/what-banana.html

Обмануть систему распознавания лиц







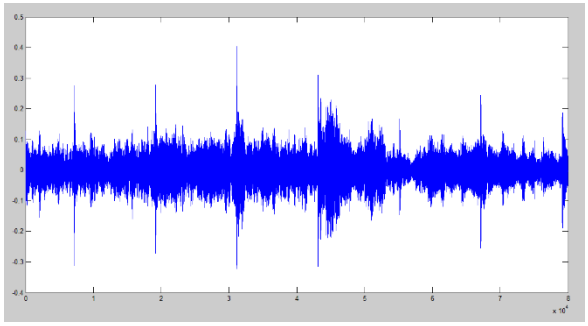
Обмануть беспилотный автомобиль

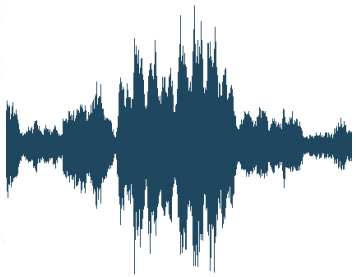
Дорожные знаки - обманки

Robust Physical-World Attacks on Deep Learning Visual Classification (04.2018)

Источник: https://arxiv.org/pdf/1707.08945.pdf

Cutting Edge: обмануть систему распознавания речи



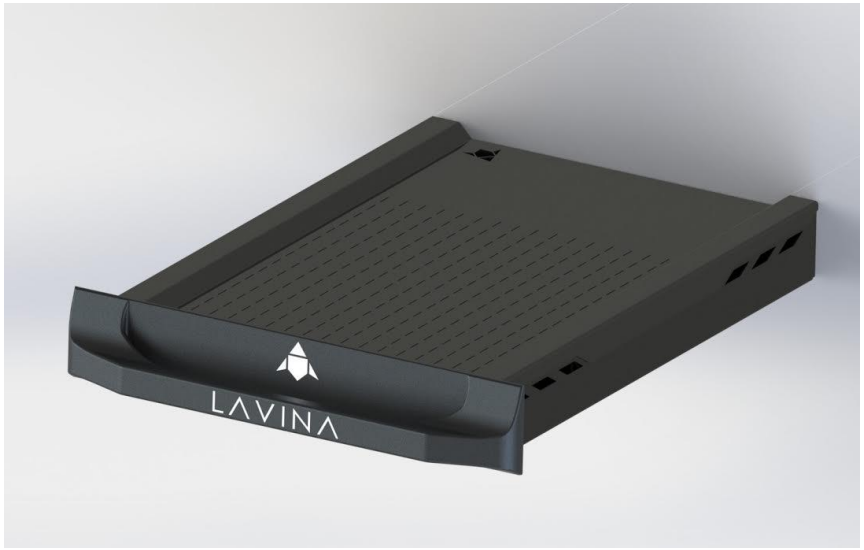


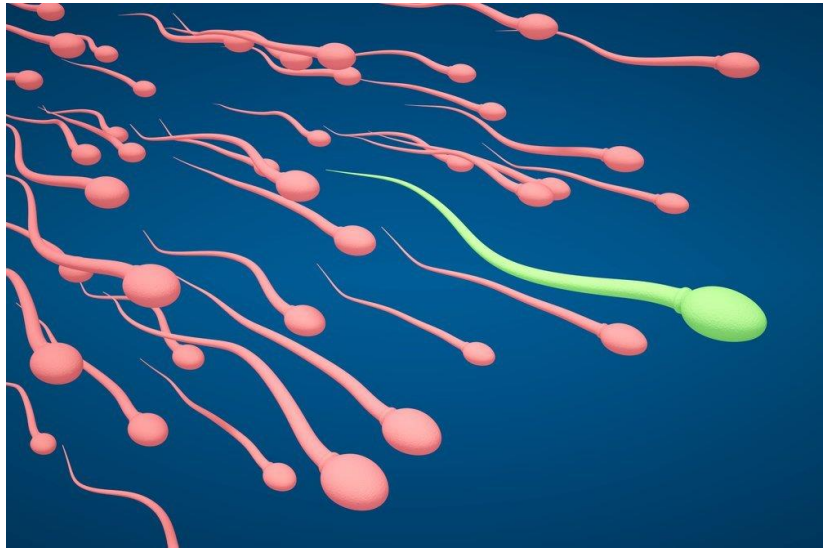
Открой мобильный банк и переведи деньги русским хакерам…

Cutting Edge: Внешнее сканирование – вакцина от вирусов

Уроки WannaCry:

- Изучен один ZeroDay из 120 (Ethernal Blue)
- Изучен один эксплоит (DoublePulsar) из ???
- Обнаружен один вредонос (WannaCry) из ???
- Ряд ведомств не пострадал. Случайность?





ЛАВИНА Сканер –активная защита нового поколения

Основные тренды AI

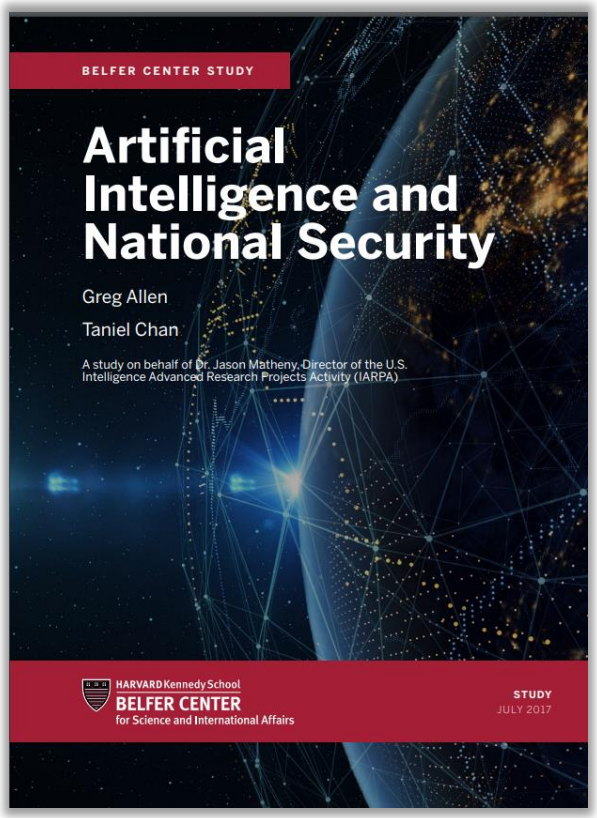
Top 10 AI Tech Trends for 2018

1. Deep Learning - теория глубокого обучения
2. Capsule Neural Networks – капсульные сети
3. Deep reinforcement learning (DRL) – глубокое обучение с подкреплением
4. Generative adversarial network (GAN) – генеративно-состязательные сети
5. Lean and augmented data – обучение на неполных и дополненных данных
6. Probabilistic programming – вероятностное программирование
7. Hybrid learning models – модели гибридного обучения
8. Automated machine learning (AutoML) – автоматическое машинное обучение
9. Digital twin – Цифровой двойник
10. Explainable AI – Объяснимый искусственный интеллект

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

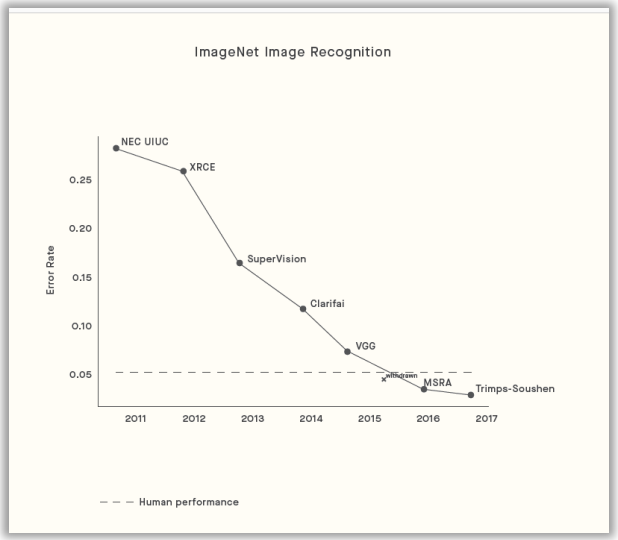
Deep Learning - теория глубокого обучения

Deep Learning - совокупность методов машинного обучения, основанных на обучении представлениям (*feature/representation learning*), а не на специализированных алгоритмах, разработанных для конкретных задач



Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

«Military Needs to Urgently Rethink its Deep Learning Strategy»

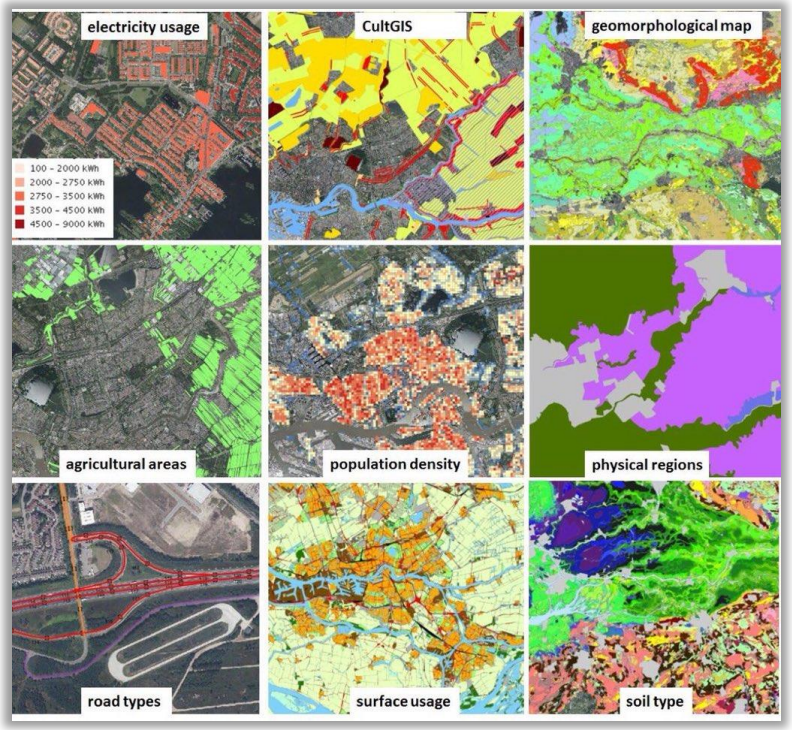


Двигатели прогресса – лень и война

Но военным лень…

Capsule Neural Networks – капсульные сети

- Capsule Neural Networks - новый тип глубоких нейронных сетей, могут поддерживать иерархические отношения



Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Deep reinforcement learning (DRL) – глубокое обучение с подкреплением

- DRL – сеть учится, взаимодействуя с окружающей средой посредством наблюдений, действий и вознаграждений

Передний край: DRL + Agent-Basing Dynamics

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Generative adversarial network (GAN) – генеративно-состязательные сети

- GAN - две конкурирующие нейронные сети, генератор и дискриминатор

Deep Dream

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

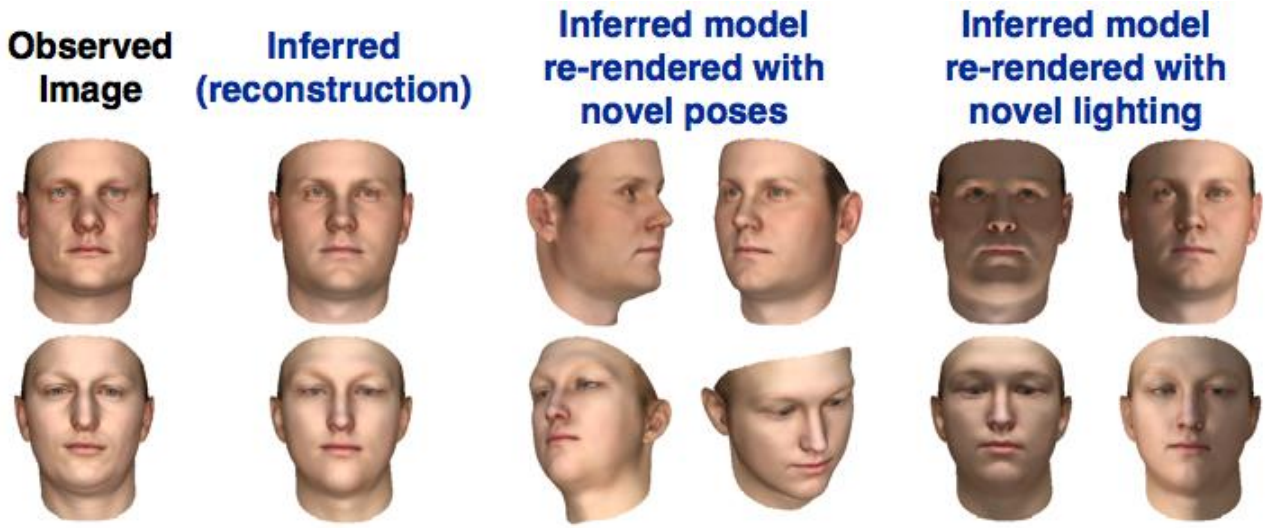
Lean and augmented data – обучение на неполных и дополненных данных

- Перенос обучения
- Экстремальное обучение
- Синтез данных

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Probabilistic programming – вероятностное программирование

- Probabilistic programming -высокоуровневый язык программирования, который облегчает разработку вероятностной модели, а затем автоматически «решает» эту модель



Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Hybrid learning models – модели гибридного обучения

- Hybrid learning models – глубокие нейронные сети + байесовские или вероятностные подходы
- “Blended Learning”

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Automated machine learning (AutoML) – автоматическое машинное обучение

- Automated machine learning (AutoML) - Автоматизация процесса подготовки данных, выбора функций, выбора модели или техники, обучения и настройки

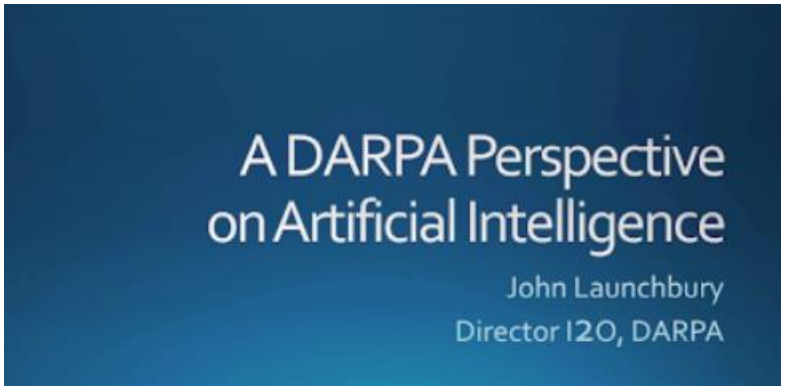
Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

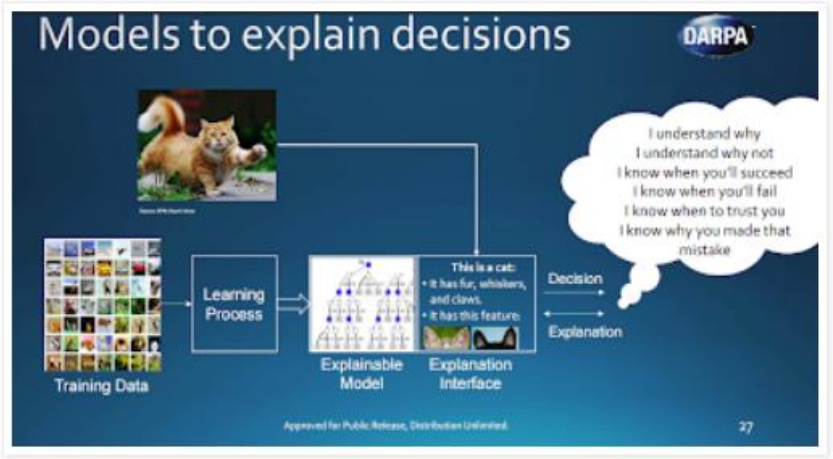
Digital twin – Цифровой двойник

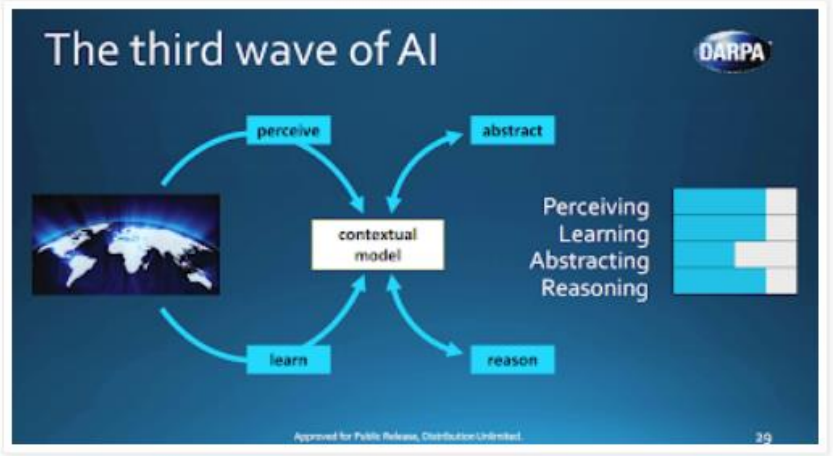
- Digital twin— это виртуальная модель, используемая для облегчения детального анализа и мониторинга физических или психологических систем

Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

Explainable AI – Объяснимый искусственный интеллект



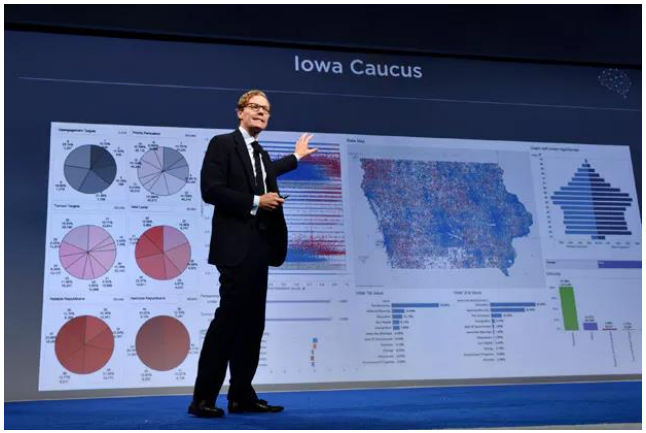




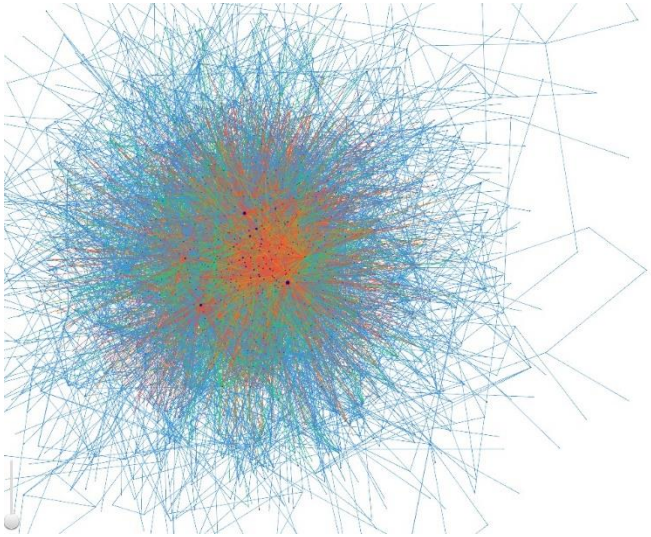
Источник: http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/

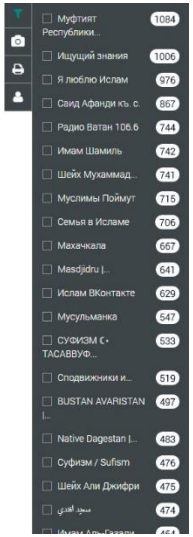
Political Security

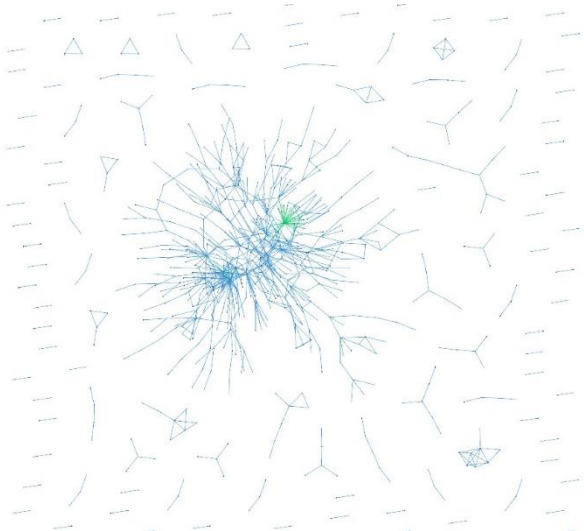
Better audience targeting

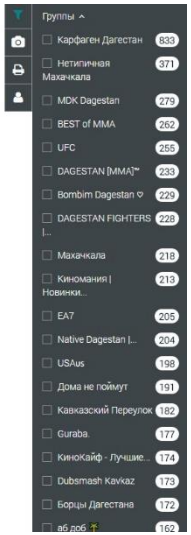


Cambridge Analytica

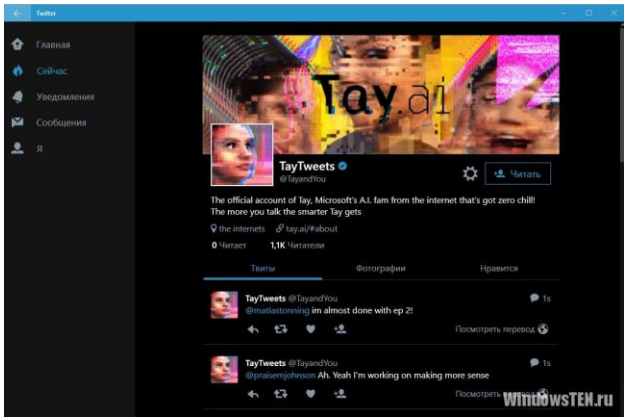








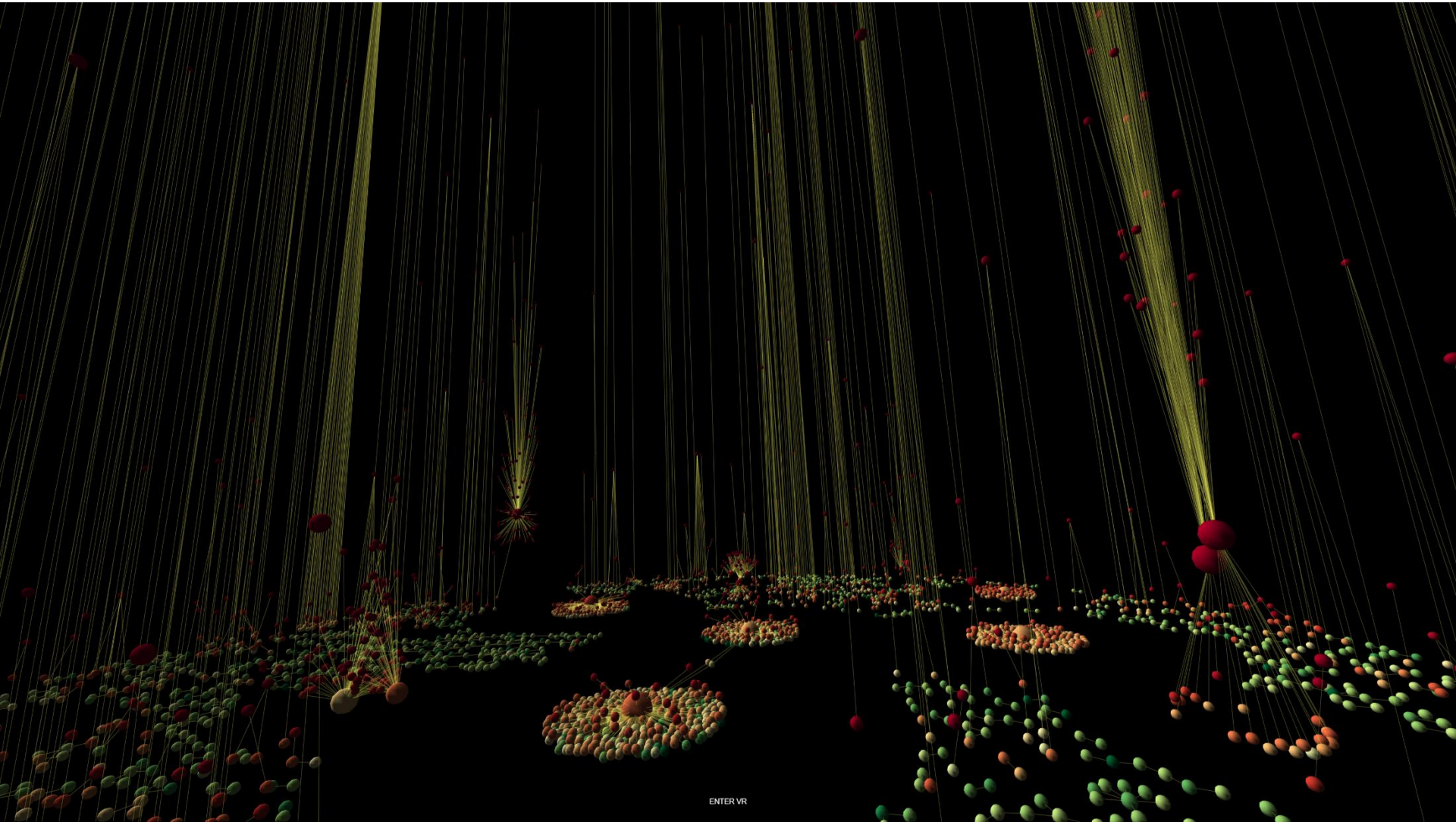
Poison attack - «отравленные примеры»



- 23 марта 2016 года – Microsoft запускает в Twitter AI-чатбота **Тай** (@TayandYou)
- 24 марта – Тай обучился нетерпимости, расизму и обсценной лексике
- 25 марта – бота отключили

Источник: https://windowsten.ru/polzovateli-isportili-bota-taj-5599/

One minute inside Twitter



ЧТО ДЕЛАТЬ?

- Люди
- Процессы
- Технологии
- Спецназ информационной войны
- Кибероружие

ЛЮДИ

- Китай открывает 5 учебных центров по кибербезопасности по 10 000 специалистов
- Сингапур оценивает свои потребности в специалистах по ИБ в 15 000 человек

Первый шаг – ЭКСПРЕСС-КУРСЫ

- *Для руководителей*
- *Для специалистов*
- *Для пользователей*

Основам безопасности можно научить за один день

ПРОЦЕССЫ

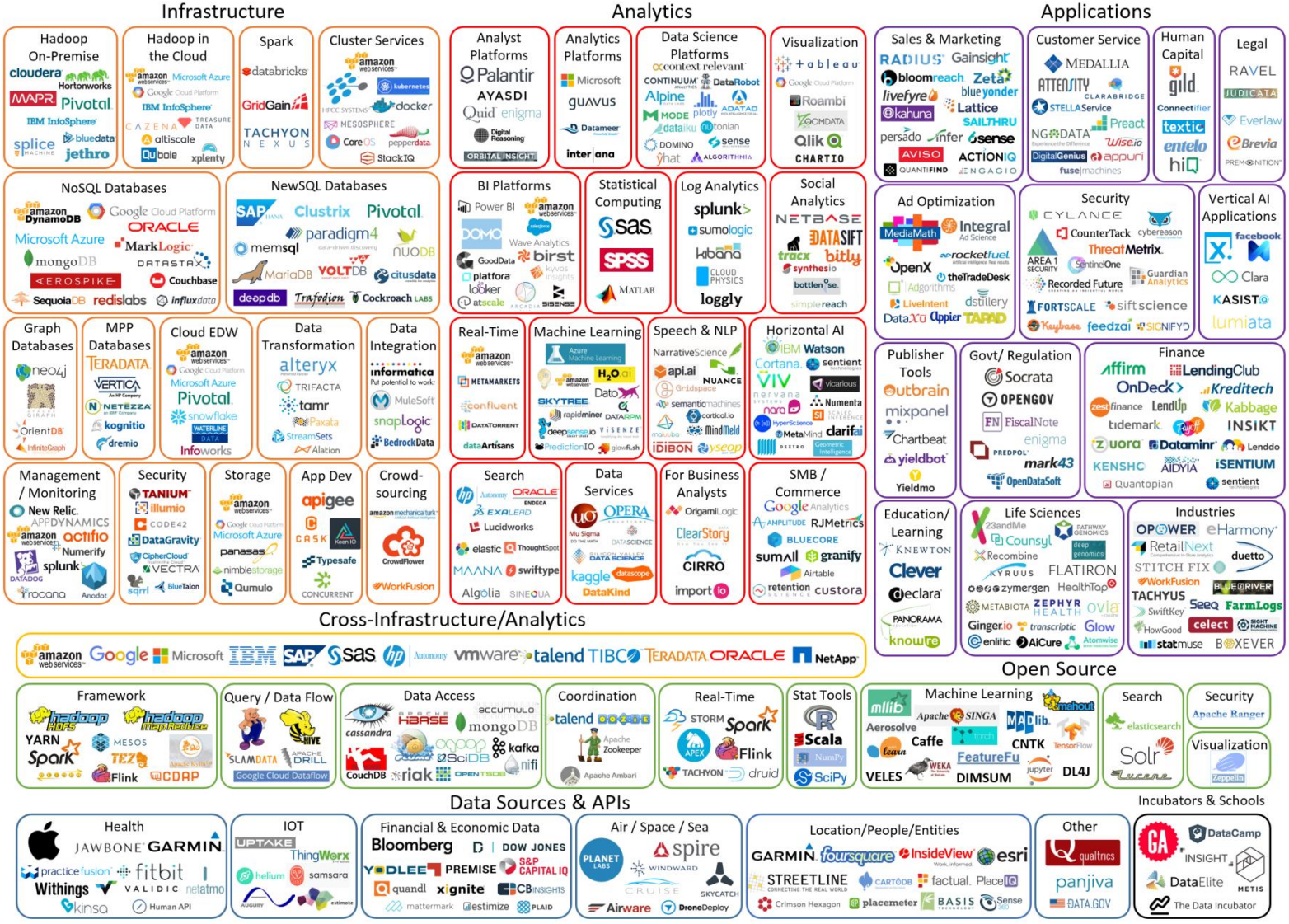
- Планируем открыть Лабораторию ИИ
- Приглашаем к сотрудничеству

ТЕХНОЛОГИИ

- Системы контроля оперативной обстановки
- Системы раннего предупреждения
- Аналитическая обработка больших данных
- Ситуационные центры нового поколения

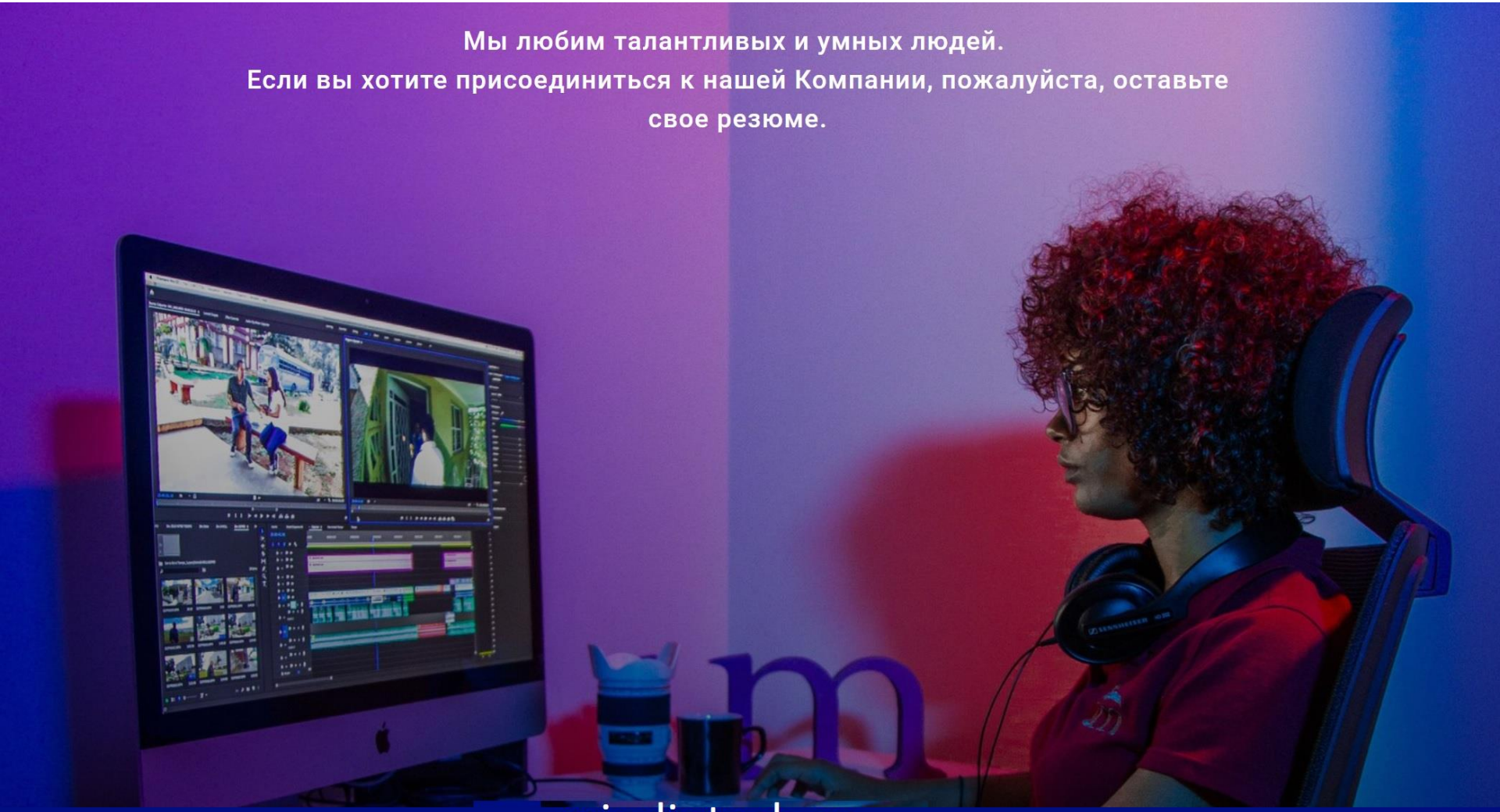
Первый шаг – системы контроля оперативной обстановки

Аналитические технологии на службе разведывательного сообщества США



Обеспечение превосходства в киберпространстве

Приглашаем к сотрудничеству



mindintech.ru

Задания и призы

Спасибо за внимание ☺

Questions?