



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-сопоставительные 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 фотографа



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 возникновения 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.

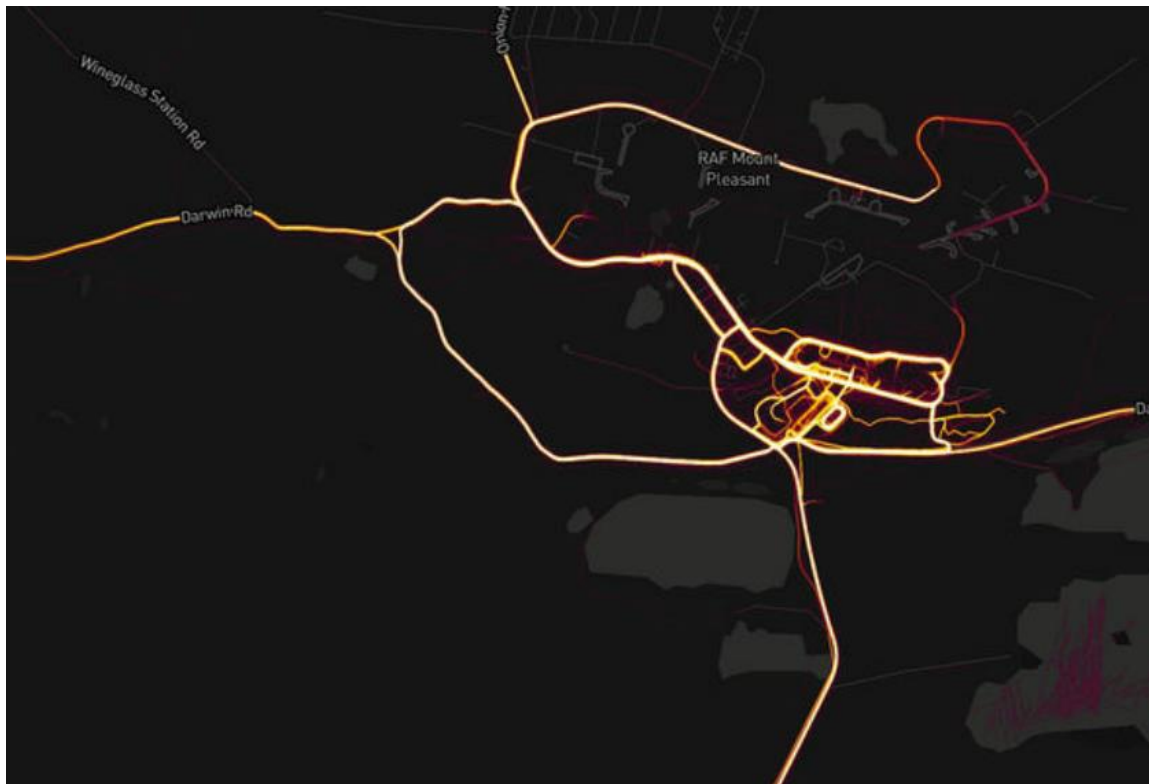
(wikipedia.org)

- **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



Look on to your target with the Tag Button.



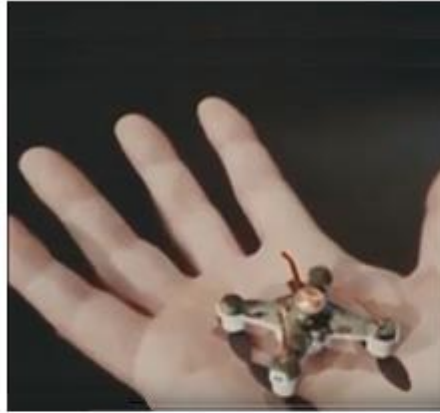
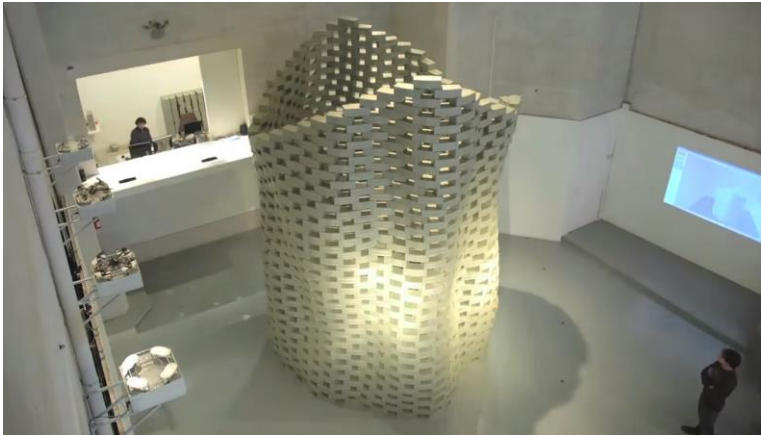
Persistently track your target as it moves.



Guide your shot on target.

Meet the dazzling flying machines of the future...

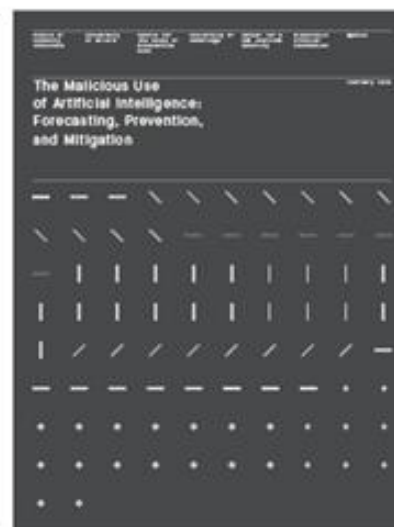
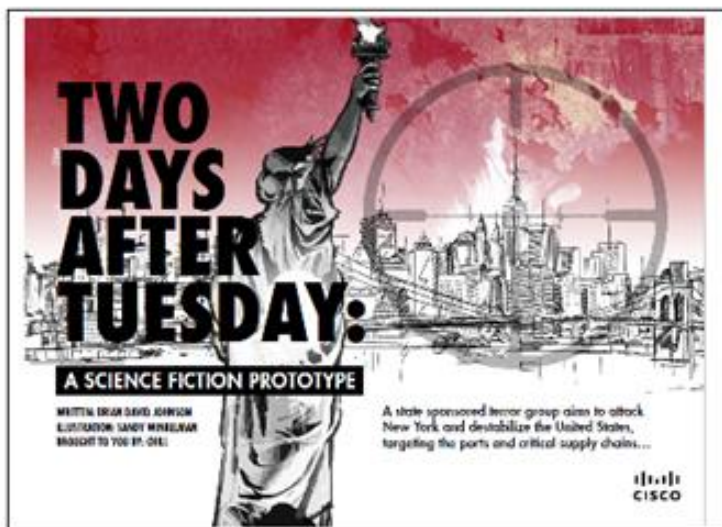
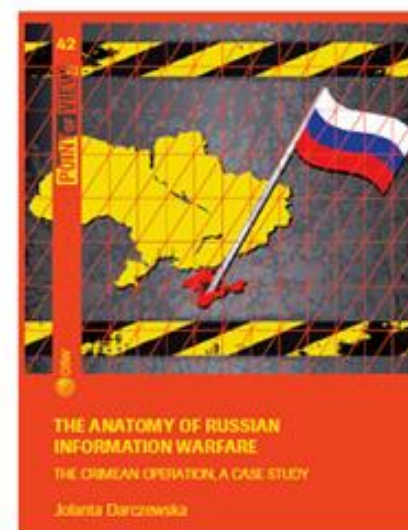
Дроны убивают людей



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

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

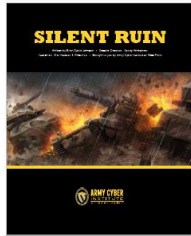
Оружие новой войны



Учите материальную часть!

Silent Ruin

война 2027 года в комиксах армии США



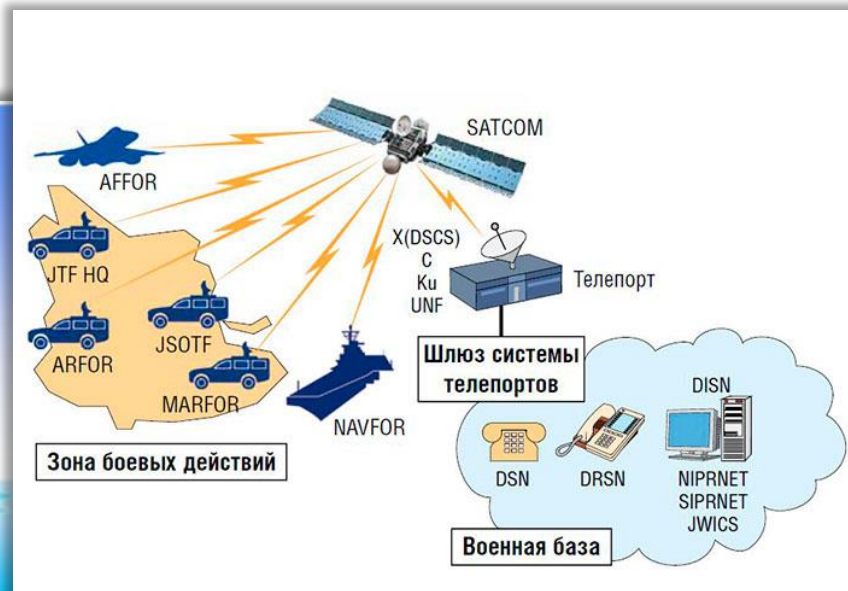
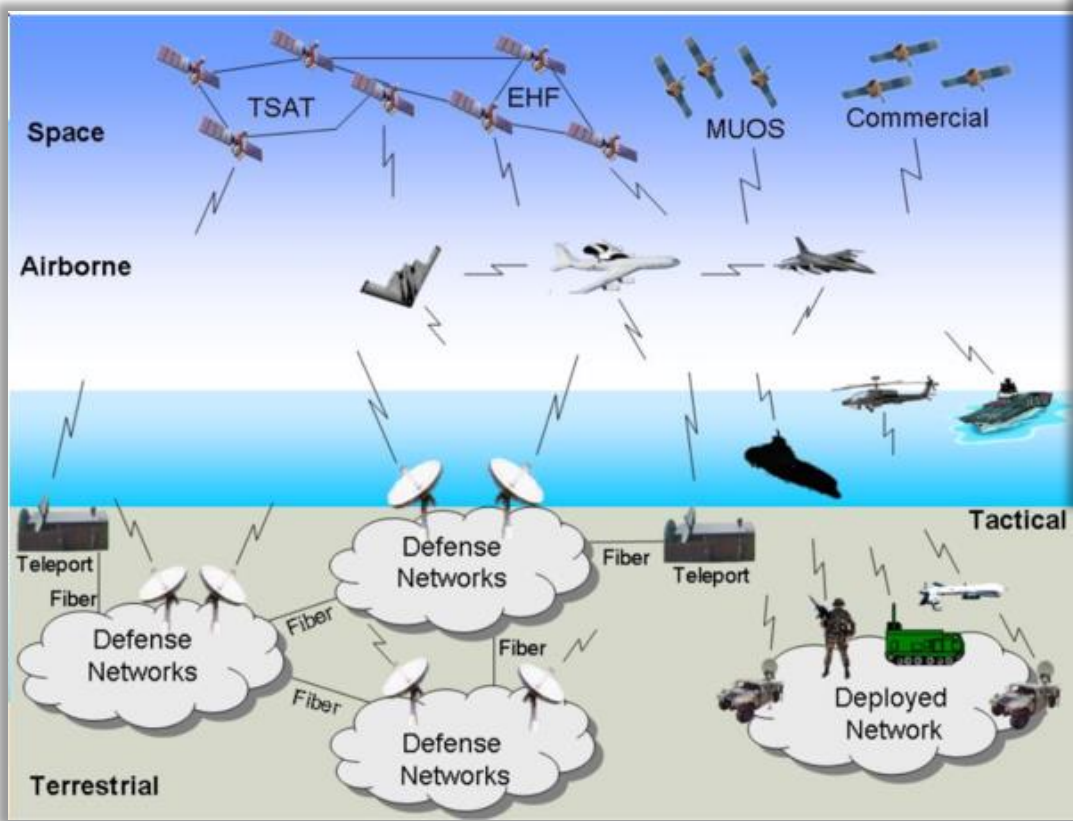
If we fail to harden our own systems while they develop innovative means to counter the proliferation of advanced autonomous systems, this threatens to render the world's most technologically advanced force its most vulnerable.

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

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



Автономное оружие

Lethal Autonomous Weapon Systems

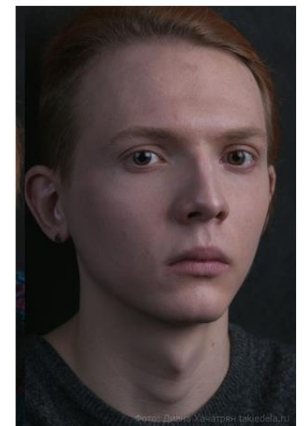
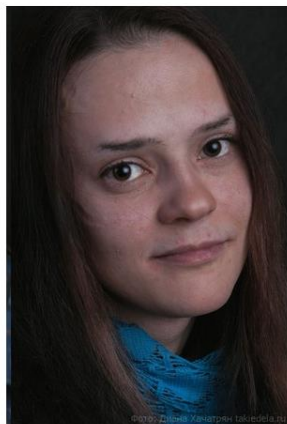


Digital Security

Обмануть нейросеть

Man or Woman ?

Как обмануть систему распознавания
Adversarial training (сопоставительное обучение)



Adversarial Examples

вредоносные примеры



x

“panda”

57.7% confidence

+ .007 ×



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

“nematode”

8.2% confidence

=



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

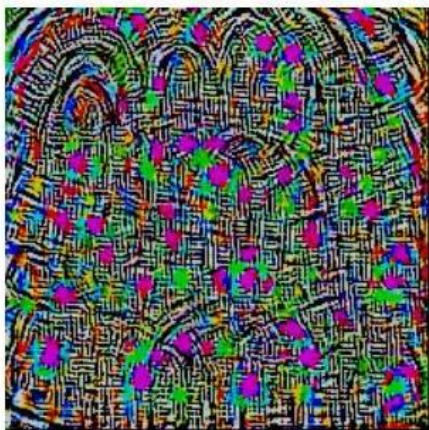
“gibbon”

99.3 % confidence

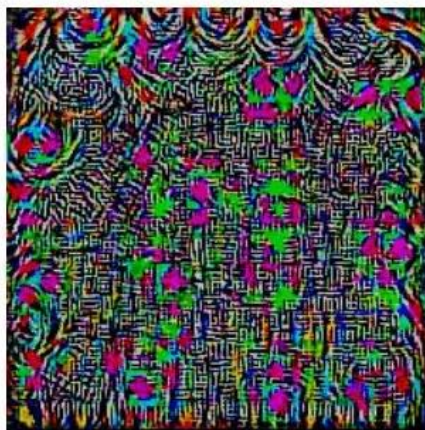


The "universal adversarial perturbation"

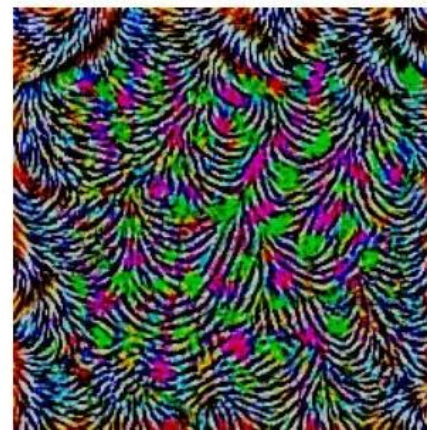
универсальное искажение



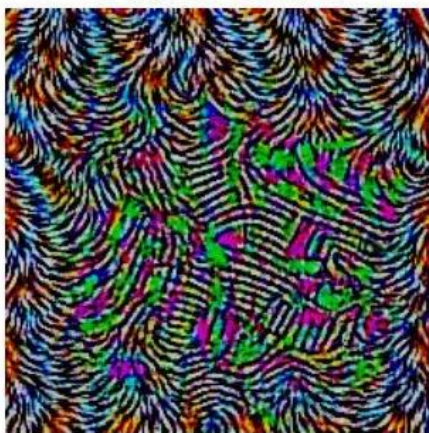
(a) CaffeNet



(b) VGG-F



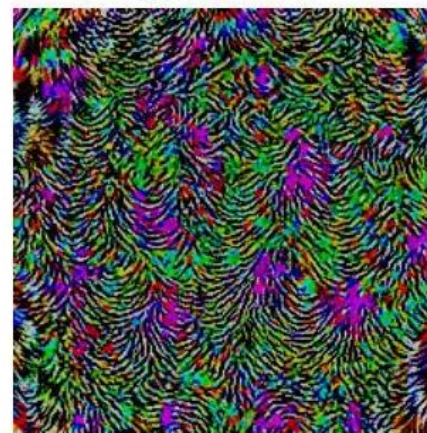
(c) VGG-16



(d) VGG-19



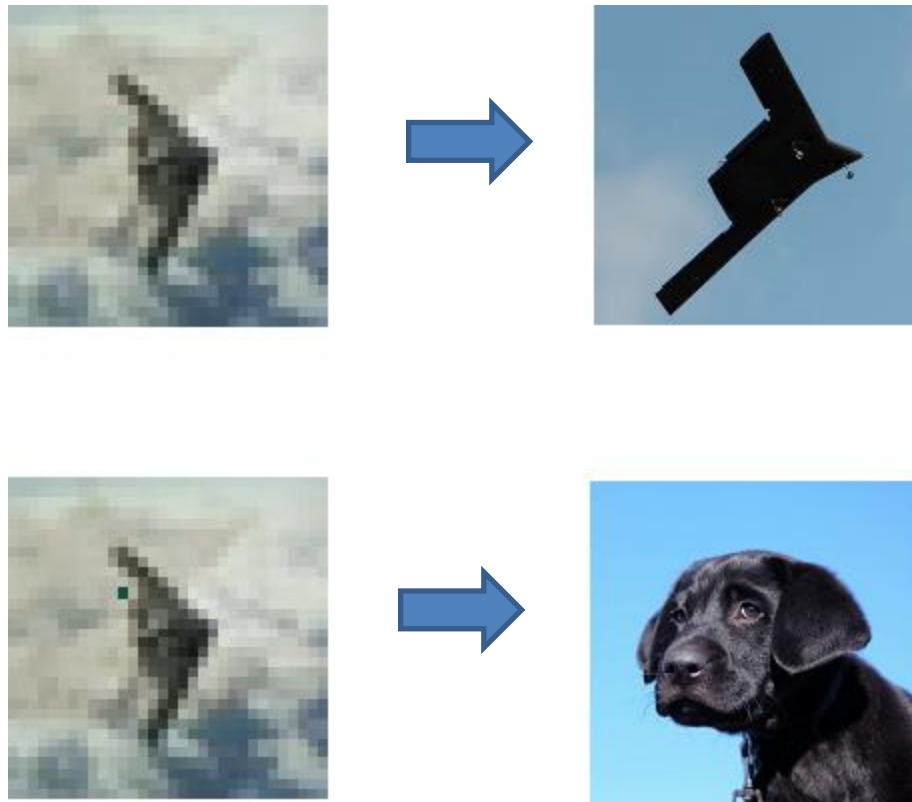
(e) GoogLeNet



(f) ResNet-152

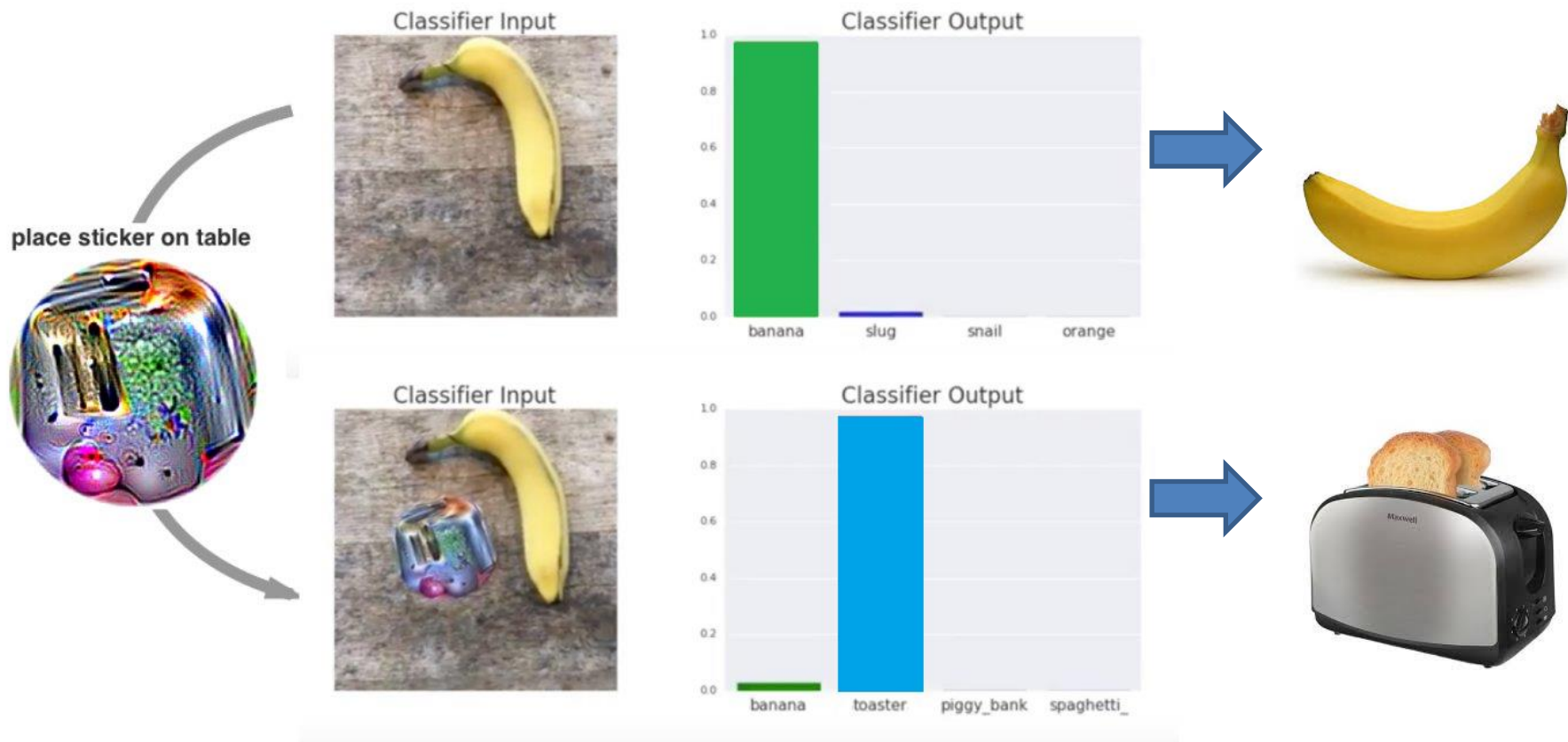
One Pixel Attack

for fooling deep neural networks

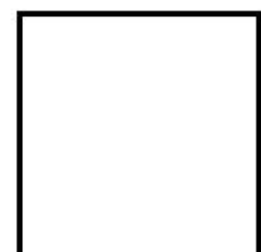
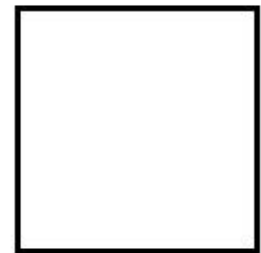
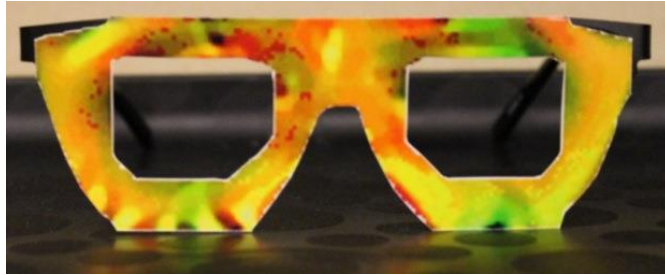


The Adversarial Patch

вредоносная заплатка



Обмануть систему распознавания лиц



Обмануть беспилотный автомобиль



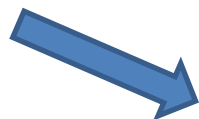
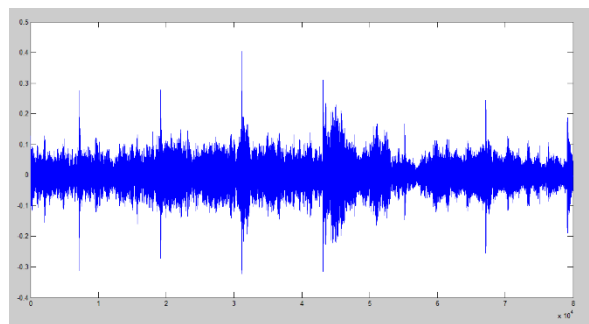
Дорожные знаки - обманки



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



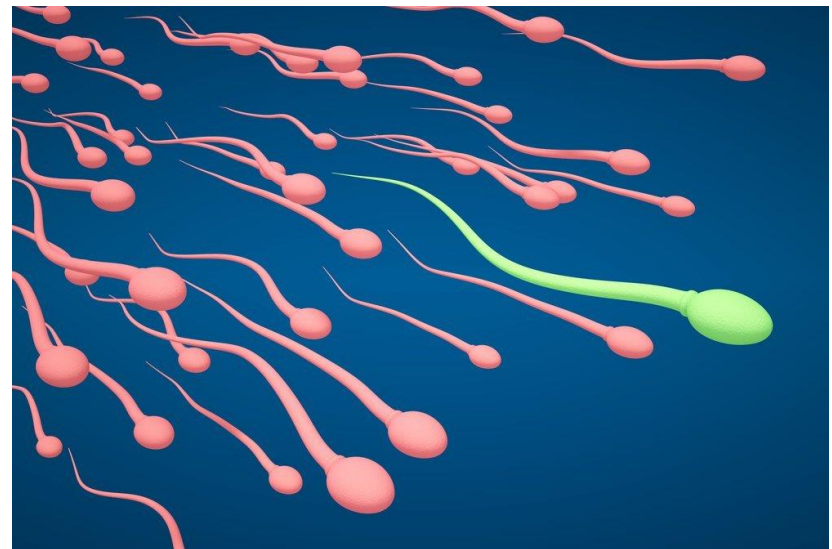
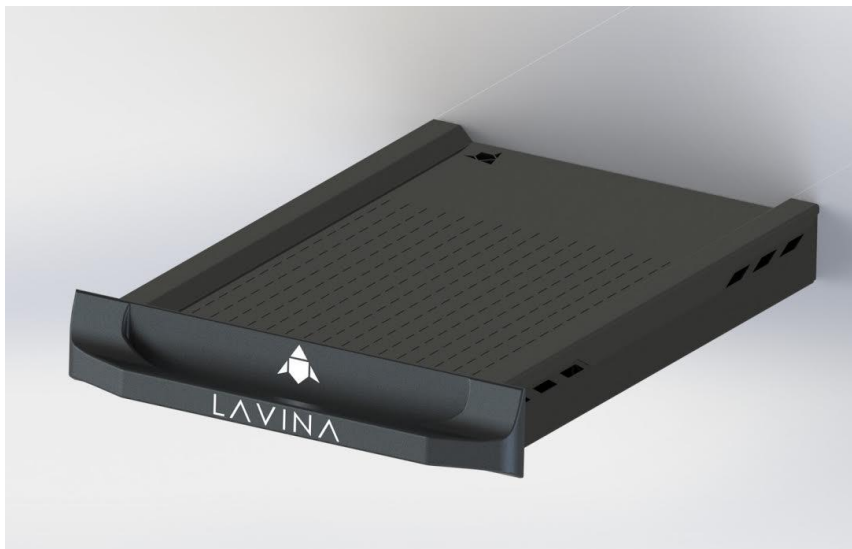
Открой мобильный банк и переведи деньги русским хакерам...



Cutting Edge: Внешнее сканирование – вакцина от вирусов

Уроки WannaCry:

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



ЛАВИНА Сканер – активная защита нового поколения

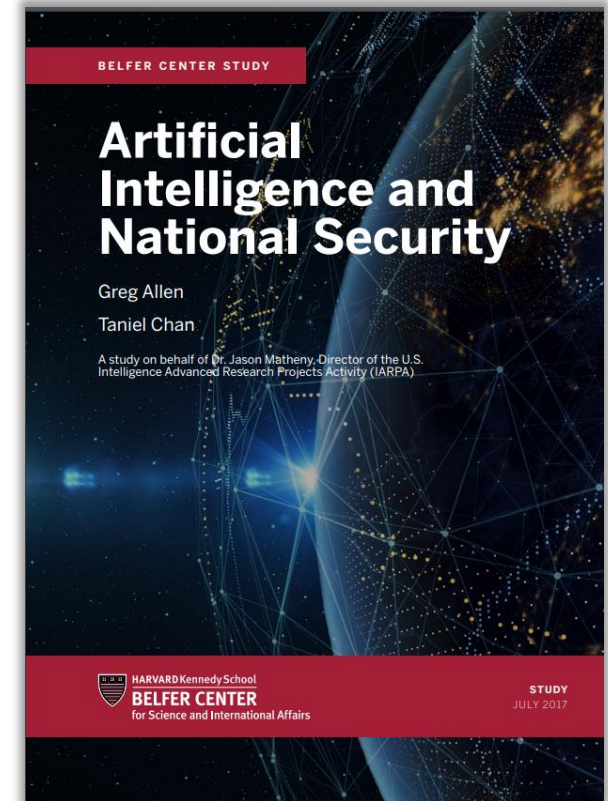
Основные тренды AI

Top 10 AI Tech Trends for 2018

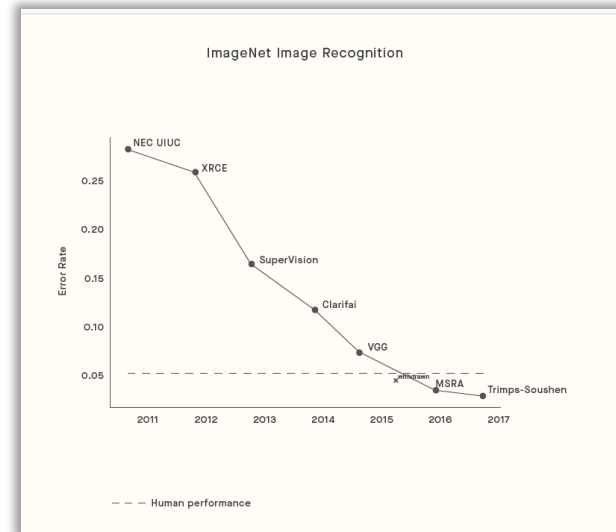
1. **Deep Learning** - теория глубокого обучения
2. **Capsule Neural Networks** – капсульные сети
3. **Deep reinforcement learning (DRL)** – глубокое обучение с подкреплением
4. **Generative adversarial network (GAN)** – генеративно-сопоставительные сети
5. **Lean and augmented data** – обучение на неполных 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

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



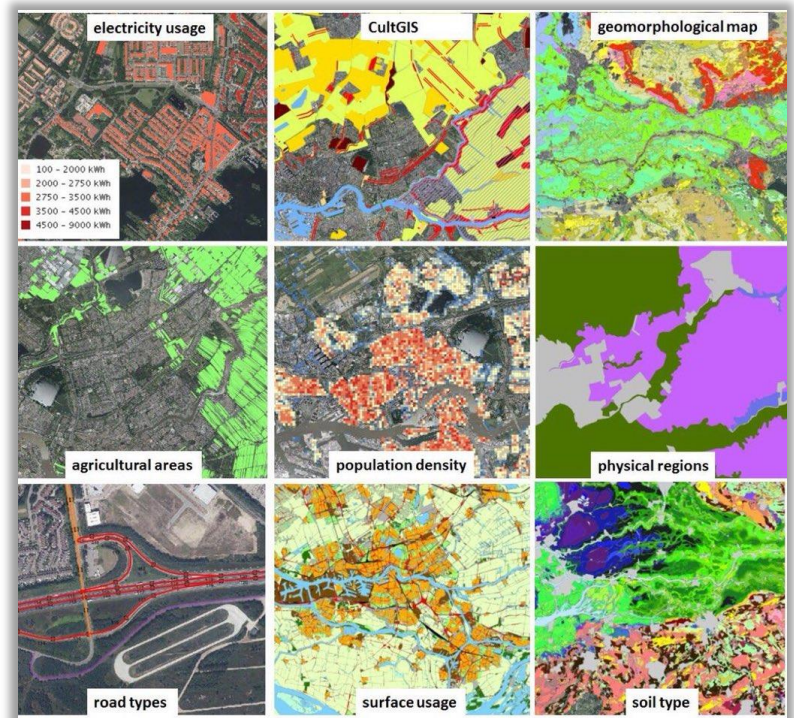
«Military Needs to Urgently Rethink its Deep Learning Strategy»



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



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

Generative adversarial network (GAN) – генеративно-сопоставительные сети

- GAN - две конкурирующие нейронные сети, генератор и дискриминатор



Deep Dream

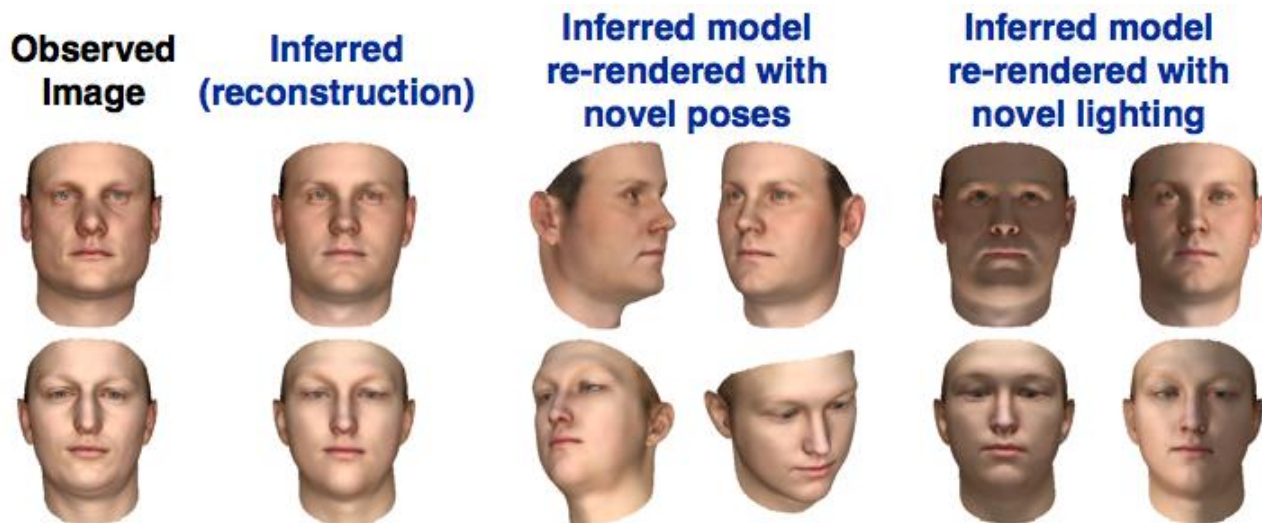
Lean and augmented data – обучение на неполных и дополненных данных

- Перенос обучения
- Экстремальное обучение
- Синтез данных



Probabilistic programming – вероятностное программирование

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



Hybrid learning models – модели гибридного обучения

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

Automated machine learning (AutoML) – автоматическое машинное обучение

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

Digital twin – Цифровой двойник

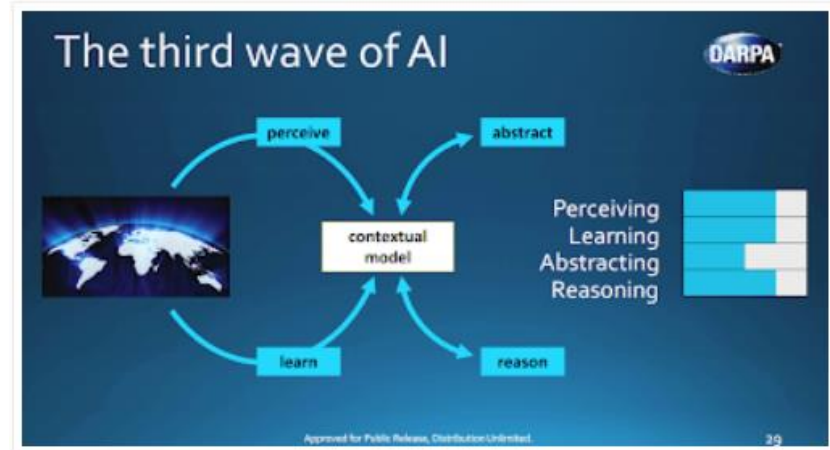
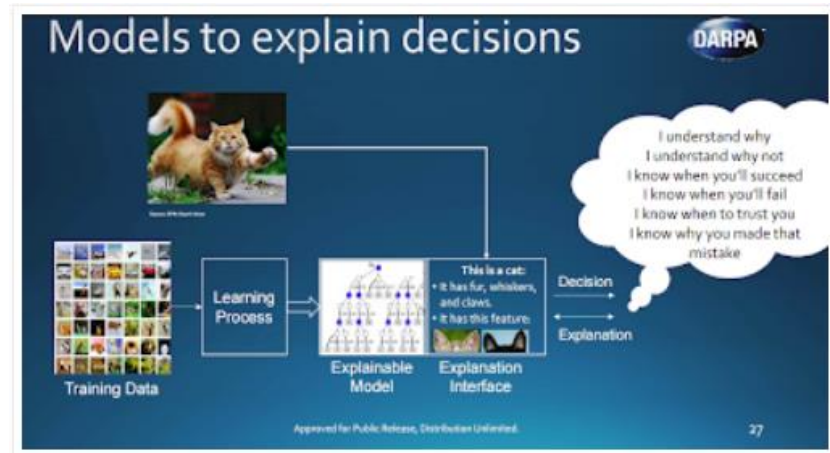
- **Digital twin**— это виртуальная модель, используемая для облегчения детального анализа и мониторинга физических или психологических систем



Explainable AI – Объяснимый ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

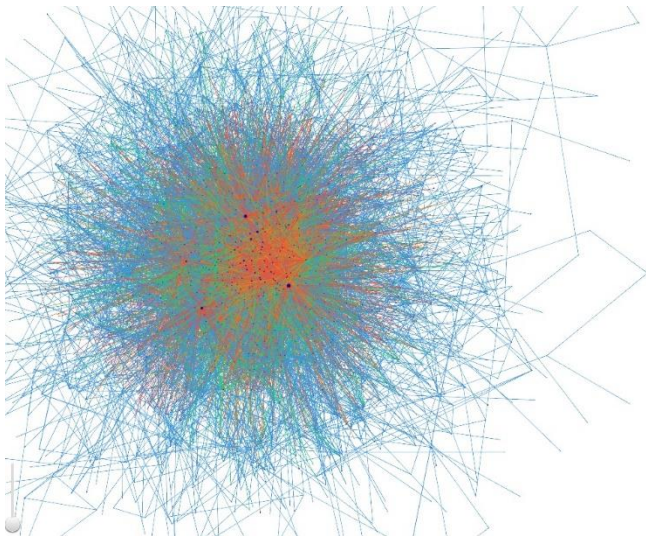
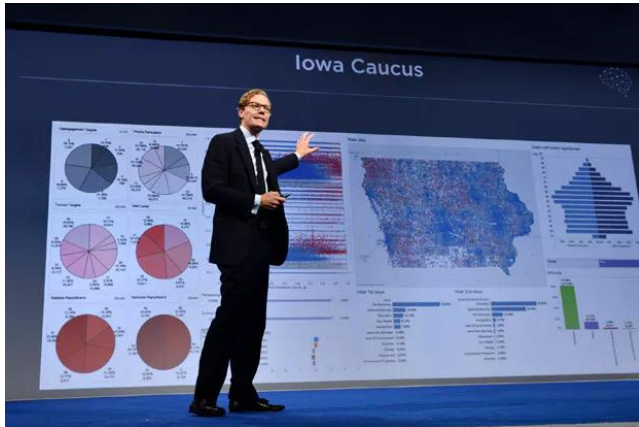
A DARPA Perspective on Artificial Intelligence

John Launchbury
Director I2O, DARPA

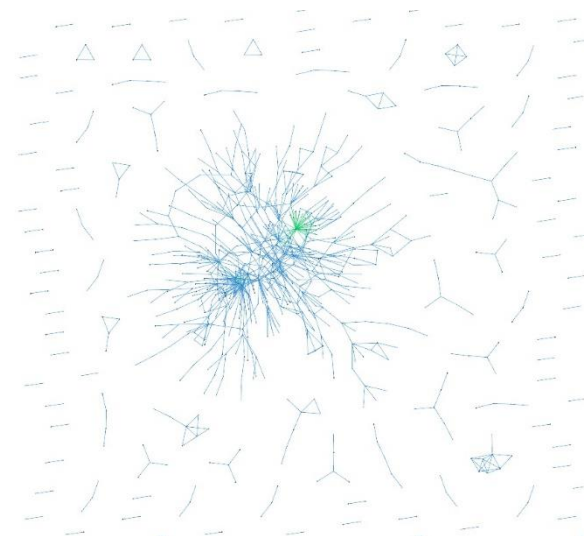


Political Security

Better audience targeting

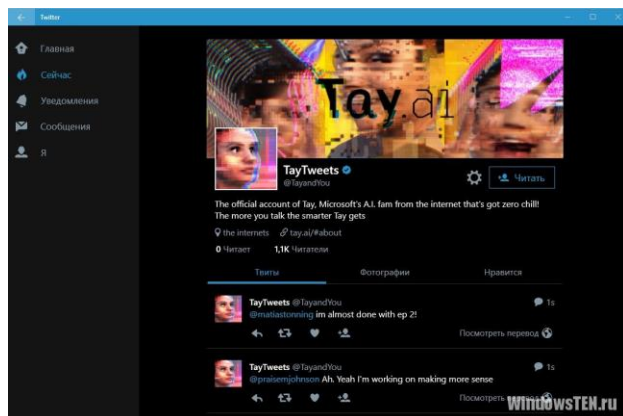


<input type="checkbox"/> Мудитият Республика...	1084
<input type="checkbox"/> Ищущий знания	1006
<input type="checkbox"/> Я люблю Ислам	976
<input type="checkbox"/> Саид Афанди кь. с	867
<input type="checkbox"/> Радио Ватан 106.6	744
<input type="checkbox"/> Имам Шамиль	742
<input type="checkbox"/> Шейх Мухаммад...	741
<input type="checkbox"/> Муслимы Поймут	715
<input type="checkbox"/> Семья в Исламе	706
<input type="checkbox"/> Махачкала	657
<input type="checkbox"/> Masfirda J. ...	641
<input type="checkbox"/> Ислам ВКонтакте	629
<input type="checkbox"/> Мусульманка	647
<input type="checkbox"/> СУФИЗМ с - ТАСАВУФ...	633
<input type="checkbox"/> Оладжикони и...	619
<input type="checkbox"/> BUSTAN AVARISTAN I...	497
<input type="checkbox"/> Native Dagestan I...	483
<input type="checkbox"/> Суфизм / Sufism	476
<input type="checkbox"/> Шейх Али Дюфрин	476
<input type="checkbox"/> سيد ابي	474
<input type="checkbox"/> Имам Али Саван	471



Группы	
<input type="checkbox"/> Карфаген Дагестан	833
<input type="checkbox"/> Нетипичная Махачкала	371
<input type="checkbox"/> MDK Dagestan	279
<input type="checkbox"/> BEST of MMA	262
<input type="checkbox"/> UFC	255
<input type="checkbox"/> DAGESTAN [MMA]	233
<input type="checkbox"/> Bombin Dagestan	229
<input type="checkbox"/> DAGESTAN FIGHTERS	228
<input type="checkbox"/> Махачкала	218
<input type="checkbox"/> Киномания Новички...	213
<input type="checkbox"/> EA7	205
<input type="checkbox"/> Native Dagestan I...	204
<input type="checkbox"/> USAus	198
<input type="checkbox"/> Дома не поймают	191
<input type="checkbox"/> Кавказский Переулок	182
<input type="checkbox"/> Guraba	177
<input type="checkbox"/> Кинокайф - Лучшие...	174
<input type="checkbox"/> Dabstahh Kavkaz	173
<input type="checkbox"/> Борцы Дагестана	172
<input type="checkbox"/> ab dob	162

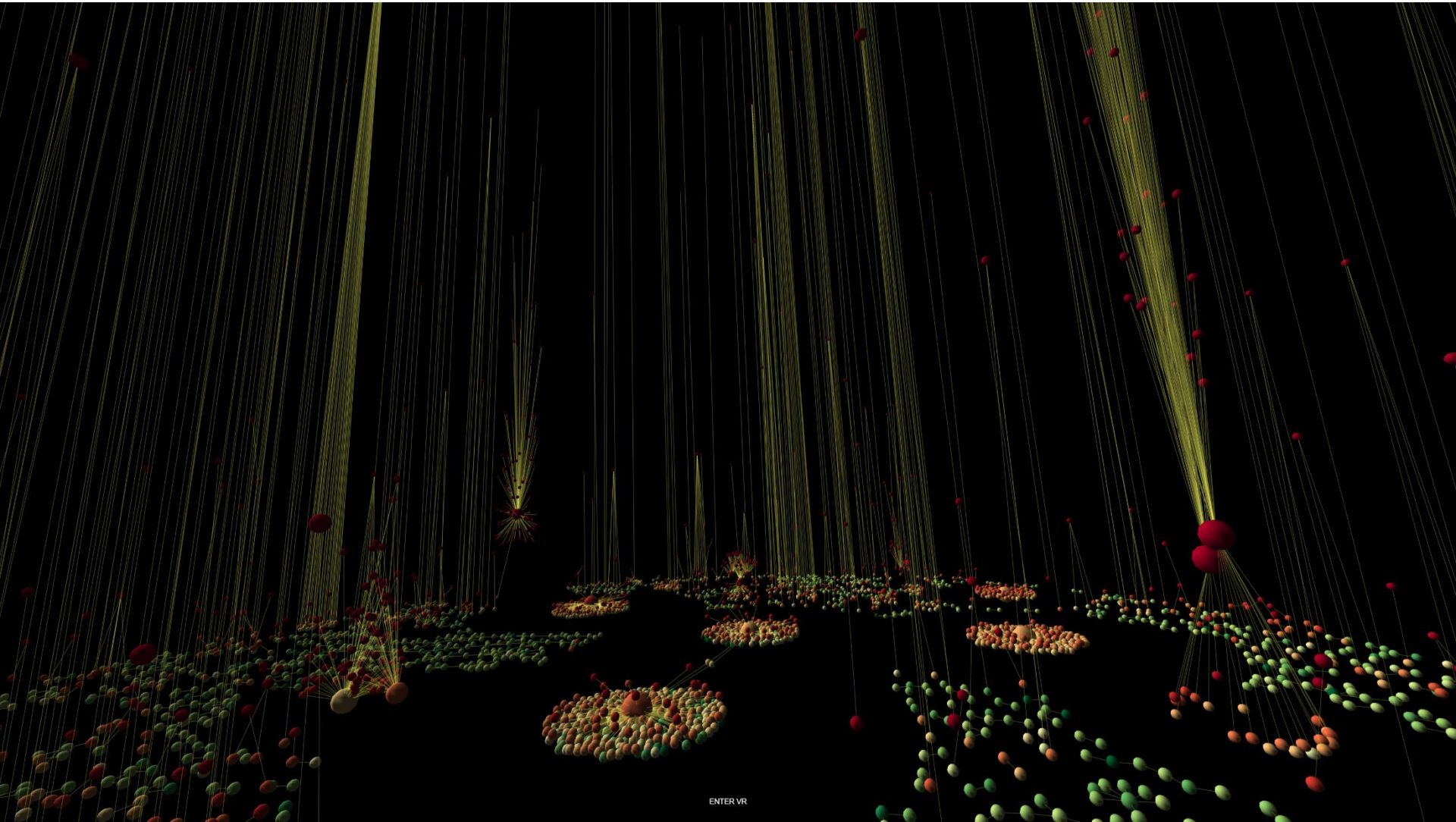
Poison attack - «отравленные примеры»



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



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

Infrastructure				Analytics				Applications			
Hadoop On-Premise cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, bluedata, jethro	Hadoop in the Cloud Amazon, Microsoft Azure, Google Cloud Platform, IBM InfoSphere, CAZENA, altiscale, Quoble, xplenty	Spark databricks, GridGain, TACHYON NEXUS	Cluster Services amazon web services, kubernetes, docker, MESOSPHERE, Core OS, pepperdata, StackIQ	Analyst Platforms Palantir, AYASDI, Quid, enigma, Digital Reasoning, ORBITAL INSIGHTS	Analytics Platforms Microsoft, guAVUS, Datameer, interana	Data Science Platforms context relevant, DataRobot, Alpine, MODE, ADATA, dataiku, tonian, DOMINO, sense, yhat, ALGORITHMIA	Visualization Tableau, Roambi, GOMDATA, Olik, CHARTIO	Sales & Marketing RADIUS, Gainsight, bloomreach, Zeta, livefyre, Kahuna, Lattice, SAILTHRU, persado, infer, sense, AVISO, ACTIONIQ, QUANTIFIND, JEN GAGIO	Customer Service MEDALLIA, ATENSTY, CLARABRIDGE, STELLAService, NGDATA, Preact, DigitalGenius, appurri, fuse:machines	Human Capital gild, Connectifier, textic, entelo, hiQ	Legal RAVEL, JUDICATA, Everlaw, Brevia, PREMIONATION
NoSQL Databases Amazon DynamoDB, Google Cloud Platform, Microsoft Azure, ORACLE, mongoDB, DATASTAX, Couchbase, SequoiaDB, redislabs, influxdata	NewSQL Databases SAP HANA, Clustrix, Pivotal, memsql, paradigm4, NUODB, MariaDB, VOLTD, CIUDATA, deepdb, Trajectory, Cockroach LABS	BI Platforms Power BI, Amazon, Domo, Wave Analytics, GoodData, platforma, looker, atscale, QlikView, Qlik Sense, Qlik Nxt, Qlik Sense Enterprise	Statistical Computing SAS, SPSS, MATLAB	Log Analytics Splunk, sumologic, kibana, cloud physics, loggly	Social Analytics Netbase, DataSift, tracx, bitly, synthosio, bottlen, simple reach	Ad Optimization MediaMath, Integral, OpenX, rocketfuel, Adgort, theTradeDesk, Livelihood, distillery, DataXu, Cppier, TAPAD	Security Cylance, CounterTack, cybereason, ThreatMetrix, Recorded Future, Fortscale, sifscience, Keybase, feedzai, SICNIFYD	Vertical AI Applications Facebook, Clara, KASIST, lumina			
Graph Databases neo4j, OrientDB, InfoGraph	MPP Databases TERADATA, VERTICA, NETEZZA, Kognitio, dremio	Cloud EDW Amazon web services, Google Cloud Platform, Microsoft Azure, Pivotal, snowflake, WATERLINE, InfoWorks	Data Transformation Alteryx, TRIFACTA, tamr, StreamSets, Alation	Data Integration Informatica, MuleSoft, snapLogic, BedrockData	Real-Time METAMARKETS, confluent, DATA TORRENT, dataArtisans	Machine Learning Azure Machine Learning, H2O, SKY TREE, rapidminer, DATA SWIM, deepsense, VISENZE, predictionIO, slowfish	Speech & NLP NarrativeScience, api.ai, NUANCE, Grindspace, semantic machines, cortico, MindMeid, IDIBON, YSCOPE	Horizontal AI IBM Watson, Cortana, sentiment, viv, nora, Numenta, MetaMind, clarifai	Publisher Tools Outbrain, mixpanel, Chartbeat, yieldbot, Yieldmo	Govt/ Regulation Socrata, OPENGOV, EN FiscalNote, PREDPOL, enigma, mark43, OpenDataSoft	Finance Affirm, LendingClub, OnDeck, Kreditech, finance, LendUp, Kabbage, tidemark, Fafy, INSIKT, uora, Dataminr, Lendio, KENSHC, AIDYIA, iSENTIUM, Quantopian, sentiment
Management / Monitoring New Relic, illumio, APPDYNAMICS, Amazon web services, actifio, Numerify, splunk, DATA DOG, TROCANO, Anodot	Security TANIUM, illumio, CODE42, DataGravity, CipherCloud, VECTRA, sqrrl, BlueTalon	Storage Amazon web services, Google Cloud Platform, Microsoft Azure, panasas, nimblestorage, Qumulo	App Dev Apigee, CASK, Typesafe, CONCURRENT	Crowd-sourcing Amazon Mechanical Turk, CrowdFlower, WorkFusion	Search HP, ELASTIC, Lucidworks, MAANA, swiftype, Algolia, SINEQUA	Data Services LIQ, OPERA, Mu Sigma, DATA SCIENCE, DATA VALUES, kaggle, datacscope, DataKind	For Business Analysts OrigamiLogic, ClearStory, CIRRO, import IO	SMB / Commerce Google Analytics, AMPLITUDE, RJMetrics, sumAll, granify, Airtable, retention, custora	Publisher Tools Outbrain, mixpanel, Chartbeat, yieldbot, Yieldmo	Govt/ Regulation Socrata, OPENGOV, EN FiscalNote, PREDPOL, enigma, mark43, OpenDataSoft	Finance Affirm, LendingClub, OnDeck, Kreditech, finance, LendUp, Kabbage, tidemark, Fafy, INSIKT, uora, Dataminr, Lendio, KENSHC, AIDYIA, iSENTIUM, Quantopian, sentiment
Cross-Infrastructure/Analytics Amazon web services, Google, Microsoft, IBM, SAP, SAS, HP, Authentiq, vmware, talend, TIBCO, TERADATA, ORACLE, NetApp											
Framework Hadoop, YARN, Spark, Mesos, TEZ, Flink, CDAP	Query / Data Flow SLAMDATA, DRILL, Google Cloud Dataflow	Data Access HBASE, accumulo, mongoDB, kafka, CouchDB, riak, OPEN TSDS, nifi	Coordination talend, Apache Ambari	Real-Time STORM, Spark, APEX, Flink, TACHYON, druid	Stat Tools Scala, NumPy, SciPy	Machine Learning mlilib, Aerolve, Caffe, SINGA, MADlib, CNTK, TensorFlow, FeatureFu, DIMSUM, jupyter, DL4J	Search Elasticsearch, Solr, Lucene	Security Apache Ranger, Visualization, Zepalin			
Data Sources & APIs											
Health Apple, JAWBONE, GARMIN, Withings, fitbit, VALIDIC, relatmo, kinsa, Human API	IOT UPTAKE, ThingWorx, helium, samsara, AUGURY, estimize	Financial & Economic Data Bloomberg, DOW JONES, Y-LEE, PREMISE, S&P CAPITAL IQ, Quandl, xignite, CB INSIGHTS, mattermark, estimize, FLAID	Air / Space / Sea PLANET LABS, Airware, DroneDeploy, WINDWARD, spire, CRUISE, SKY CATCH	Location/People/Entities GARMIN, foursquare, InsideView, esri, STREETLINE, CARTO DB, factual, PlaceIQ, Crimon Hexagon, placemeter, BASIS, Sense	Other qualtrics, panjiva, DATA.GOV	Incubators & Schools GA, DataCamp, INSIGHT, DataElite, METIS, The Data Incubator					

Обеспечение превосходства в киберпространстве



Приглашаем к сотрудничеству

Мы любим талантливых и умных людей.
Если вы хотите присоединиться к нашей Компании, пожалуйста, оставьте
свое резюме.



mindintech.ru

Задания и призы



Спасибо за внимание 😊

Questions?

