RSAConference2020

San Francisco | February 24 – 28 | Moscone Center

HUMAN ELEMENT

SESSION ID: LAW-T12

Am I allowed to subvert machine learning for fun and profit?

MODERATOR: **Ram Shankar Siva Kumar**
Data Cowboy
Microsoft/Harvard
@ram_ssk

PANELISTS: **Cristin Goodwin**
Assistant General Counsel
Microsoft
@CristinGoodwin

Betsy Cooper
Director
Aspen Tech Policy Hub
@BetsOnTech

Nicholas Carlini
Research Scientist
Google Brain

Question Time!


G
C
C
C

Congratulations! You are 100% Human!

“7”

“Orangutan”

Source: https://arxiv.org/abs/1809.08352

“Hot Dog”

Source: https://arxiv.org/abs/1807.06732

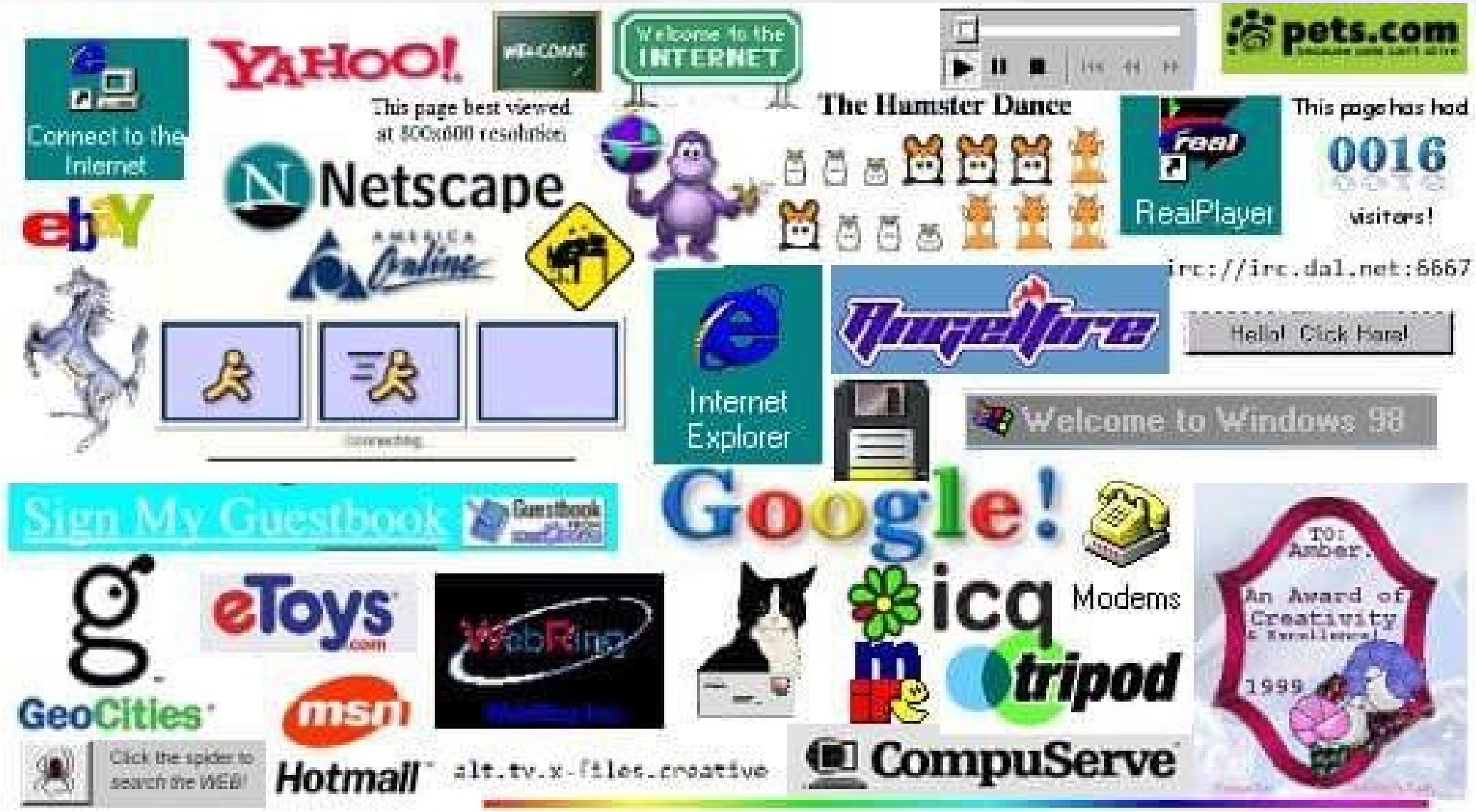
Source: https://arxiv.org/abs/1801.01944

Doesn’t transcribe to anything

Source: https://arxiv.org/abs/1801.01944

“Speech can be embedded in music”

Source: https://arxiv.org/abs/1801.01944



Source: https://knowyourmeme.com/photos/1295515-starter-packs

Source: https://gbhackers.com/these-were-the-biggest-cyber-attacks-of-the-year-2016/

Cybersecurity Regulatory landscape - Sample

Federal Laws

Computer Fraud and Abuse Act

Electronic Communications Privacy Act

Federal Trade Commission 15 U.S.C. § 45

Industry specific (HIPPA, Gramm-Leach-Bliley Act)

State Laws

Computer crime – All 50 states

Spyware –20 states

Phishing - 23 states

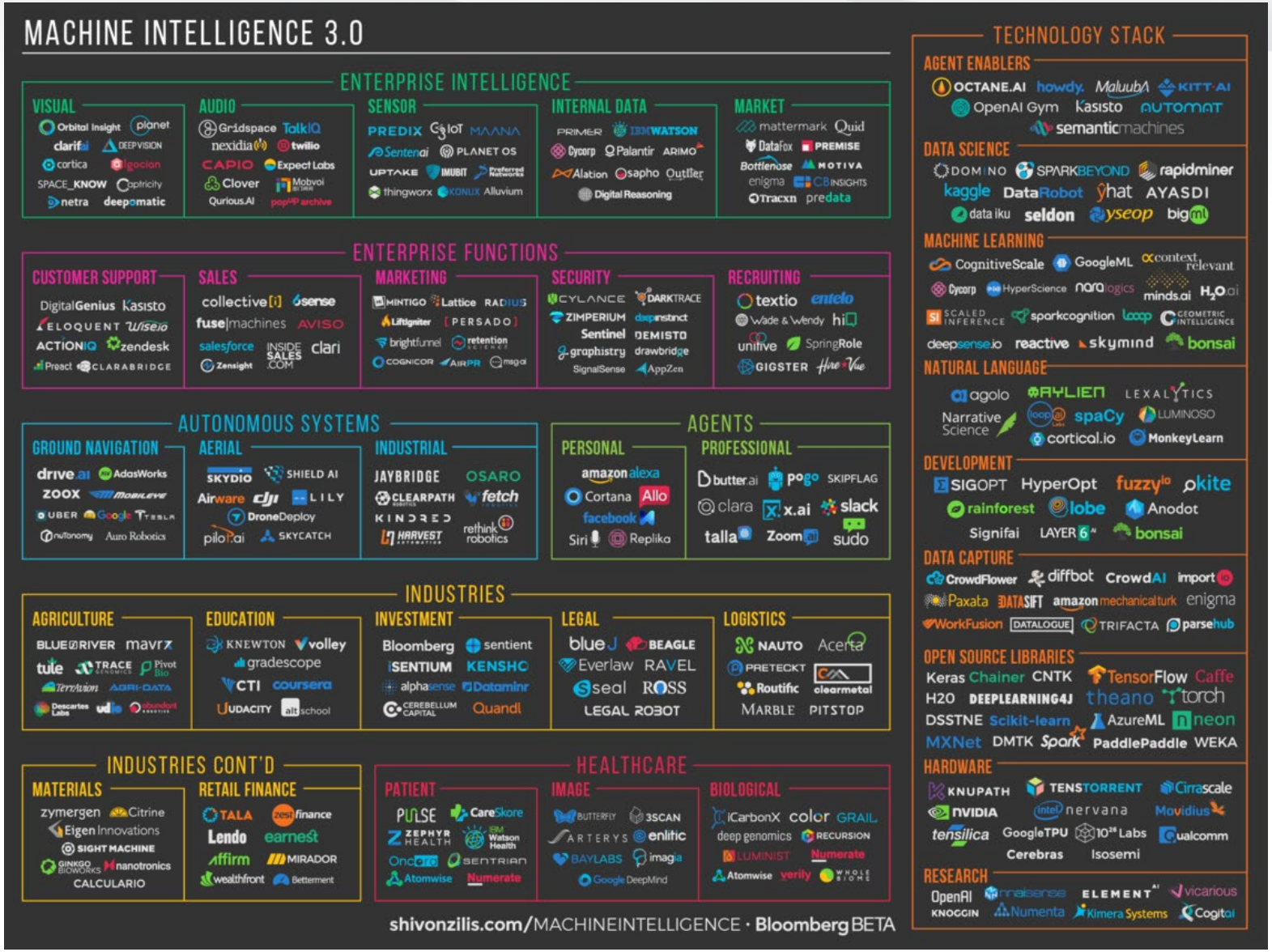
Security Breach Notification Laws (All 50 States)

Standardized Security Framework

NIST Cybersecurity Framework

ISO/IEC 27001

ISO 15408 – "Common Criteria"


MACHINE INTELLIGENCE 3.0
ENTERPRISE INTELLIGENCE
VISUAL
Orbital Insight
planet
clarifai
DEEPVISION
cortica
Sigocion
SPACE_KNOW
Captricity
netra
deepomatic
AUDIO
Gridspace
TalkIQ
nexidia
twilio
CAPIO
Expect Labs
Clover
Mobvoi
Qurious.AI
popup archive
SENSOR
PREDIX
IoT
MAANA
Sentenai
PLANET OS
UPTAKE
IMUBIT
Preferred Networks
thingworx
KONUX
Alluvium
INTERNAL DATA
PRIMER
IBM WATSON
Cycorp
Palantir
ARIMO
Alation
sapho
Outlier
Digital Reasoning
MARKET
mattermark
Quid
DataFox
PREMISE
Bottlenose
MOTIVA
enigma
CB INSIGHTS
Tracxn
predata
ENTERPRISE FUNCTIONS
CUSTOMER SUPPORT
DigitalGenius
Kasisto
ELOQUENT
Wise.io
ACTIONIQ
zendesk
Preact
CLARABRIDGE
SALES
collective[i]
6sense
fusemachines
AVISO
salesforce
INSIDE SALES.COM
clari
Zensight
MARKETING
MINTIGO
Lattice
RADIUS
Liftigniter
PERSADO
brightfunnel
retention science
COGNICOR
AIRPR
msgi
SECURITY
CYLANCE
DARKTRACE
ZIMPERIUM
deepinstinct
Sentinel
DEMISTO
graphistry
drawbridge
SignalSense
AppZen
RECRUITING
textio
entelo
Wade & Wendy
hi
unitive
SpringRole
GIGSTER
HireVue
AUTONOMOUS SYSTEMS
GROUND NAVIGATION
drive.ai
AdasWorks
ZOOX
Mobileye
UBER
Google
Tesla
nuTonomy
Auro Robotics
AERIAL
SKYDIO
SHIELD AI
Airware
dji
LILY
DroneDeploy
pilot.ai
SKYCATCH
INDUSTRIAL
JAYBRIDGE
OSARO
CLEARPATH
fetch
KINDRED
HARVEST
rethink robotics
AGENTS
PERSONAL
amazon alexa
Cortana
Allo
facebook
Siri
Replika
PROFESSIONAL
butter.ai
pogo
SKIPFLAG
clara
x.ai
slack
talla
Zoom
sudo
INDUSTRIES
AGRICULTURE
BLUE RIVER
mavrx
tule
TRACE
Pivot Bio
Terravion
AGRI-DATA
Descartes Labs
udio
abundant
EDUCATION
KNEWTON
volley
gradescope
CTI
coursera
UDACITY
altschool
INVESTMENT
Bloomberg
sentient
SENTIUM
KENSHO
alphasense
Dataminr
CEREBELLUM CAPITAL
Quandl
LEGAL
blueJ
BEAGLE
Everlaw
RAVEL
Sseal
ROSS
LEGAL ROBOT
LOGISTICS
NAUTO
Acerta
PRETECKT
Routific
clearmetal
MARBLE
PITSTOP
INDUSTRIES CONT'D
MATERIALS
zymergen
Citrine
Eigen Innovations
SIGHT MACHINE
GINKGO BIOWORKS
nanotronics
CALCULARIO
RETAIL FINANCE
TALA
zestfinance
Lendo
earnest
Affirm
MIRADOR
wealthfront
Betterment
HEALTHCARE
PATIENT
PULSE
CareSkore
ZEPHYR HEALTH
IBM Watson Health
Oncora
SENTRIAN
Atomwise
Numerate
IMAGE
BUTTERFLY
3SCAN
ARTERYS
enlitic
BAYLABS
imagia
Google DeepMind
BIOLOGICAL
iCarbonX
color
GRAIL
deep genomics
RECURSION
LUMINIST
Numerate
Atomwise
verily
WHOLE BIOME
shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA
TECHNOLOGY STACK
AGENT ENABLERS
OCTANE.AI
howdy.
Maluuba
KITT.AI
OpenAI Gym
Kasisto
AUTOMAT
semanticmachines
DATA SCIENCE
DOMINO
SPARKBEYOND
rapidminer
kaggle
DataRobot
yhat
AYASDI
data iku
seldon
yseop
bigml
MACHINE LEARNING
CognitiveScale
GoogleML
context relevant
Cycorp
HyperScience
narrative logics
minds.ai
H2O.ai
SCALED INFERENCE
sparkcognition
loop
GEOMETRIC INTELLIGENCE
deepsense.io
reactive
skymind
bonsai
NATURAL LANGUAGE
agolo
AYLIEN
LEXALYTICS
Narrative Science
spaCy
LUMINOSO
cortical.io
MonkeyLearn
DEVELOPMENT
SIGOPT
HyperOpt
fuzzyio
pkite
rainforest
lobe
Anodot
Signifai
LAYER 6
bonsai
DATA CAPTURE
CrowdFlower
diffbot
CrowdAI
import.io
Paxata
DATASIFT
amazon mechanical turk
enigma
WorkFusion
DATALOGUE
TRIFACTA
parsehub
OPEN SOURCE LIBRARIES
Keras
Chainer
CNTK
TensorFlow
Caffe
H20
DEEPLEARNING4J
theano
torch
DSSTNE
Scikit-learn
AzureML
neon
MXNet
DMTK
Spark
PaddlePaddle
WEKA
HARDWARE
KNUPATH
TENSTORRENT
Cirrascale
NVIDIA
intel nervana
Movidius
tensilica
GoogleTPU
10^26 Labs
Qualcomm
Cerebras
Isosemi
RESEARCH
OpenAI
numenta
ELEMENT AI
vicarious
KNOGGIN
Numenta
Kimera Systems
Cogitai

WIRED SUBSCRIBE

LOUISE MATSAKIS SECURITY 12.28.2017 12:07 PM

Researchers Fooled a Google AI Into Thinking a Rifle Was a Helicopter

To safeguard AI, we're going to need to solve the problem of 'adversarial examples.'

The New York Times

Alexa and Siri Can Hear This Hidden Command. You Can't.

Researchers can now send secret audio instructions undetectable to the human ear to Apple's Siri, Amazon's Alexa and Google's Assistant.

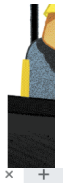
Sign in Get starte

Noteworthy - The Journal Blog YOUR STORIES | GET EARLY ACCESS TO JOURNAL

OpenGPT-2: We Replicated GPT-2 Because You Can Too

Vanya Cohen Follow
Aug 22, 2019 · 7 min read

Aaron Gokaslan*, Vanya Cohen*, Ellie Pavlick, Stefanie Tellex | Brown University



ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

TESLA AUTOPILOT —

Researchers trick Tesla Autopilot into steering into oncoming traffic

Stickers that are invisible to drivers and fool autopilot.

DAN GOODIN - 4/1/2019, 8:50 PM

Taylor Swift 'tried to sue' Micros

bbc.com/news/newsbeat-49645508

BBC News Sport Reel More

Search

NEWS

Home Video World US & Canada UK Business Tech More

Newsbeat

Taylor Swift 'tried to sue' Microsoft over racist chatbot Tay

10 September 2019 Share

Machine Learning Regulatory landscape

Federal Laws

????

State Laws

????

Standardized Security Framework

????

RSAConference2020

Evasion Attack

RSAConference2020

Model Stealing Attack

RSAConference2020

Poisoning Attack

RSAConference2020

Crafting Impactful Adversarial ML Policy

"Apply" Slide

- Next Week
  - Read "Law And Adversarial Machine Learning" (Link)
  - Ask your Machine Learning (ML) team to inventory the ML systems and their uses
- Next Month
  - Ask ML team to map the inventory to the attacks using this taxonomy (Link)
  - Go through the EU Trustworthy AI Assessment List with the most critical ML systems in your inventory (Link)
- Next Quarter
  - Put together a plan to update your security policy to account for attacks based on assessment list