strike

# Deep Dive: Threat modeling for Artificial Intelligence & Machine Learning

# This is NOT a Game.

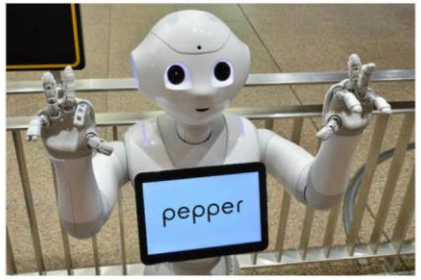
The Register — Biting the hand that feeds IT

Security

## Softbank's 'Pepper' robot is a security joke

Big-in-Japan 'bot offers root access through hard-coded password and worse bugs too

By Richard Chirgwin 29 May 2018 at 00:29 — 56 — SHARE




Softbank's popular anthropomorphic robot, Pepper, has myriad security holes according to research published by Scandinavian researchers earlier this month.

Amazon | Business | Technology

## Amazon's Alexa recorded and shared a conversation without consent, report says

A Portland woman said an Amazon device recorded a conversation in her home without her consent, and shipped the file to someone in her contact list, KIRO-TV reported.

By Matt Day, Seattle Times business reporter

Share story

The Register — Biting the hand that feeds IT

Security

## Bot-ched security: Chat system hacked to slurp hundreds of thousands of Delta Air Lines, Sears customers' bank cards

Hi! How may we pwn you today?

By Shaun Nichols in San Francisco 5 Apr 2018 at 19:41 — 22 — SHARE

Hackers are feared to have swiped sensitive personal information held by two of the best known companies in the US – after malware infected a customer support software maker.

Business > Technology

## Tesla Autopilot implicated in another emergency-vehicle crash


1 of 5

A Tesla sedan reportedly with the "Autopilot" driver-assistance system turned on crashed into a Laguna Beach Police Department SUV on Tuesday, May 29, 2018 in Laguna Beach, Calif. (courtesy of Laguna Beach Police Department)

REUTERS — World Business Markets Politics TV

MAY 2, 2018 / 11:16 AM / 2 MONTHS AGO

## Cambridge Analytica and British parent shut down after Facebook scandal

Reuters Staff — 4 MIN READ

venturebeat.com/2018/05/21/india-wants-to-use-ai-in-weapons-systems/

VB — NEWS — EVENTS — RESEARCH

AI

## India wants to use AI in weapons systems

KYLE WIGGERS @KYLE_L_WIGGERS MAY 21, 2018 11:25 AM

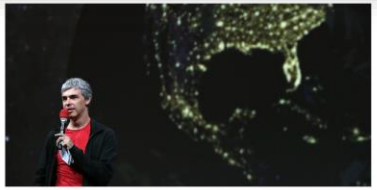
Image Credit: SNEHIT / Shutterstock

India will enlist the help of artificial intelligence to develop weapons, defense, and surveillance systems, government officials announced today.

"The world is moving towards an artificial intelligence-driven ecosystem," Dr. Ajay Kumar, secretary at the defense ministry, said in a statement. "India is also taking necessary steps to prepare our defense forces for the war of the future."

VB Recommendat

Upcoming Events



Google cofounder Larry Page Getty

- Google confirmed that it has a contract with the Department of Defense involving AI technology and drones, but has declined to go into detail.
- Google said the technology is being used for "non-offensive" uses.
- Google has long avoided being part of the military industrial complex, to the point where it seems to have been an unofficial company policy.
- The news that Google was working with the DoD reportedly upset many of the company's employees.

insurancenewsnet.com — Your industry. One source.

Now reading NEWSWIRES — TOPICS — INN EXCLUSIVES — NEWSWIRES — DOL RULE NEWS — PODCAST

newswires

April 26, 2018

## Artificial Intelligence: Commission Outlines a European Approach to Boost Investment and Set Ethical Guidelines

Targeted News Service (Press Releases)

BRUSSELS, Belgium, April 25 -- The European Commission issued the following news release:

Today the European Commission is presenting a series of measures to put artificial intelligence (AI) at the service of Europeans and boost Europe's competitiveness in this field.

The Commission is proposing a three-pronged approach to increase public and private investment in AI, prepare for socio-economic changes, and ensure an appropriate ethical and legal framework. This follows European leaders' call for a European approach on AI.

**L0phtCrack**
@L0phtCrackLLC

Investigation into the July data breach incident at Singapore's largest healthcare provider has revealed that local administrators made several critical mistakes that led to the breach, including the use of weak passwords and unpatched software.

Poor 'p@ssword' hygiene and unpatched systems led to cybersecurity breach...
securityboulevard.com

11:26 AM · Sep 24, 2018

# Which of you wants to be the Adminstrator?

# AI must be Secure and Protect Privacy

Can we protect customers and their AI solutions – services, tools, and infrastructure?

Can we help them protect their customer's and employee's personal data?

Can we protect data scientists, and their AI data?

Can we help them avoid costly mistakes and oversights?

Can we protect developers, and their AI applications?

Can we help them meet their regulatory obligations?

# Master Class of Manipulations

- Minimize human recognition of “evidence of manipulation” and maximize the negative impact on the classifier.
  - Injection of random noise at random levels & locations
  - Removal of color from images via sepia filters.
  - Modification of hue and saturation for random blocks/locations in an image.

Hackathon 2018: Hide in Plain Sight

# How easy was it?

- Subtle modification of input data can be used to force mis-classification without warning or failure indication of any kind.
- Trust is largely misplaced in uncurated, unsigned, public-access data sources.
- Machine learning algorithms lack the ability to monitor their own training progress or reject inputs which would compromise or degrade the effectiveness of future analysis.

# Noise – whole image and random partial

Original Classification:
**ox (score = 0.89517)**
water buffalo, water ox, Asiatic buffalo, Bubalus bubalis (score = 0.05105)
oxcart (score = 0.00753)
bison (score = 0.00357)
sorrel (score = 0.00052)

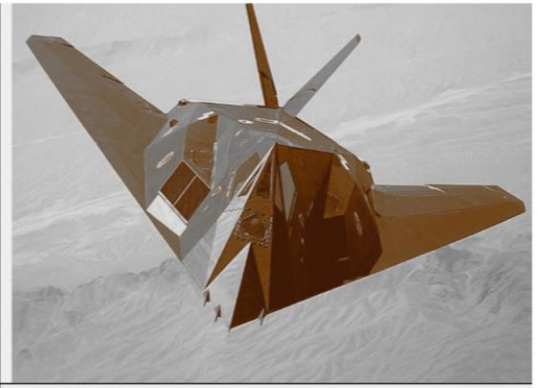
Manipulated Image (30% noise):
**Indian elephant, Elephas maximus (score = 0.19209)**
hog, pig, grunter, squealer, Sus scrofa (score = 0.15600)
ox (score = 0.04219)
African elephant, Loxodonta africana (score = 0.03333)
bison (score = 0.02342)

Original Classification:
**beagle (score = 0.78377)**
English foxhound (score = 0.04376)
basset, basset hound (score = 0.02375)
Walker hound, Walker foxhound (score = 0.01258)
bloodhound, sleuthhound (score = 0.00375)

Manipulated Image:
jaguar, panther, Panthera onca, Felis onca (score = 0.11067)
English foxhound (score = 0.02110)
bull mastiff (score = 0.02049)
Walker hound, Walker foxhound (score = 0.01989)
**beagle (score = 0.01631)**

# Effectiveness of subtle manipulations

Original Classification:
**warplane, military plane (score = 0.93798)**
missile (score = 0.01181)
projectile, missile (score = 0.01054)
wing (score = 0.00981)
space shuttle (score = 0.00462)



Sepia (manipulated) Image:
**sundial (score = 0.68659)**
warplane, military plane (score = 0.04613)
wing (score = 0.03205)
snowplow, snowplough (score = 0.02640)
plow, plough (score = 0.02189)


Original Classification:
**street sign (score = 0.96099)**
traffic light, traffic signal, stoplight (score = 0.00397)


Sepia (manipulated) Image:
**street sign (score = 0.67787)**
mailbox, letter box (score = 0.06781)

# Build upon the Details: Security Best Practices

## Build upon established security recommended practices:

- Use **Threat Modeling** and applying standard security controls.
- Extend your threat modeling to include specific use scenarios *outside* of the traditional technical use cases.
- Threat modeling for voice, video, and gesture-driven user experiences.
- SDL is necessary; however, *you will need to go beyond* these standards and controls to address new paradigms.

What is your AI's specific use scenario?

What are your assumptions about the AI interactions, the user's interactions, and an Attacker's interactions?

How do you Assure Provenance and Curate Lineage with Discretion?

How do you plan to operate, monitor and control your AI systems' plans and actions in response to observations and commands?

# Understand the business purposes & problems at hand.

Make sure AI solves the problems—
Do not build AI just for sake of building AI.

- Seek agreement with the customer on what the AI will predict.
- Discuss what risks the customer may have when the AI predicts wrong.
- Identify who, and what could potentially abuse the system, and why.

# Understand geopolitical & industry regulations

- Plan for regulatory requirements while designing an AI Solution
- Understand the impact of the AI uses, misuses, and actions to the human being.
- Consider and address requirements for regulatory compliance.

# Clear Business Purpose Aligned with Microsoft AI ethics principles

It is important that the AI Solution will be built upon an ethical foundation. The AI Solution must assist humanity and should be designed to address ethics principles.

## Our approach - Microsoft AI

https://www.microsoft.com/en-us/AI/our-approach-to-ai

**Fairness**
AI systems should treat all people fairly

**Inclusiveness**
AI systems should empower everyone and engage people

**Reliability & Safety**
AI systems should perform reliably and safely

**Transparency**
AI systems should be understandable

**Privacy & Security**
AI systems should be secure and respect privacy

**Accountability**
AI systems should have algorithmic accountability