Regulating AI and Robotics

Steve Omohundro, Ph.D.

PossibilityResearch.com

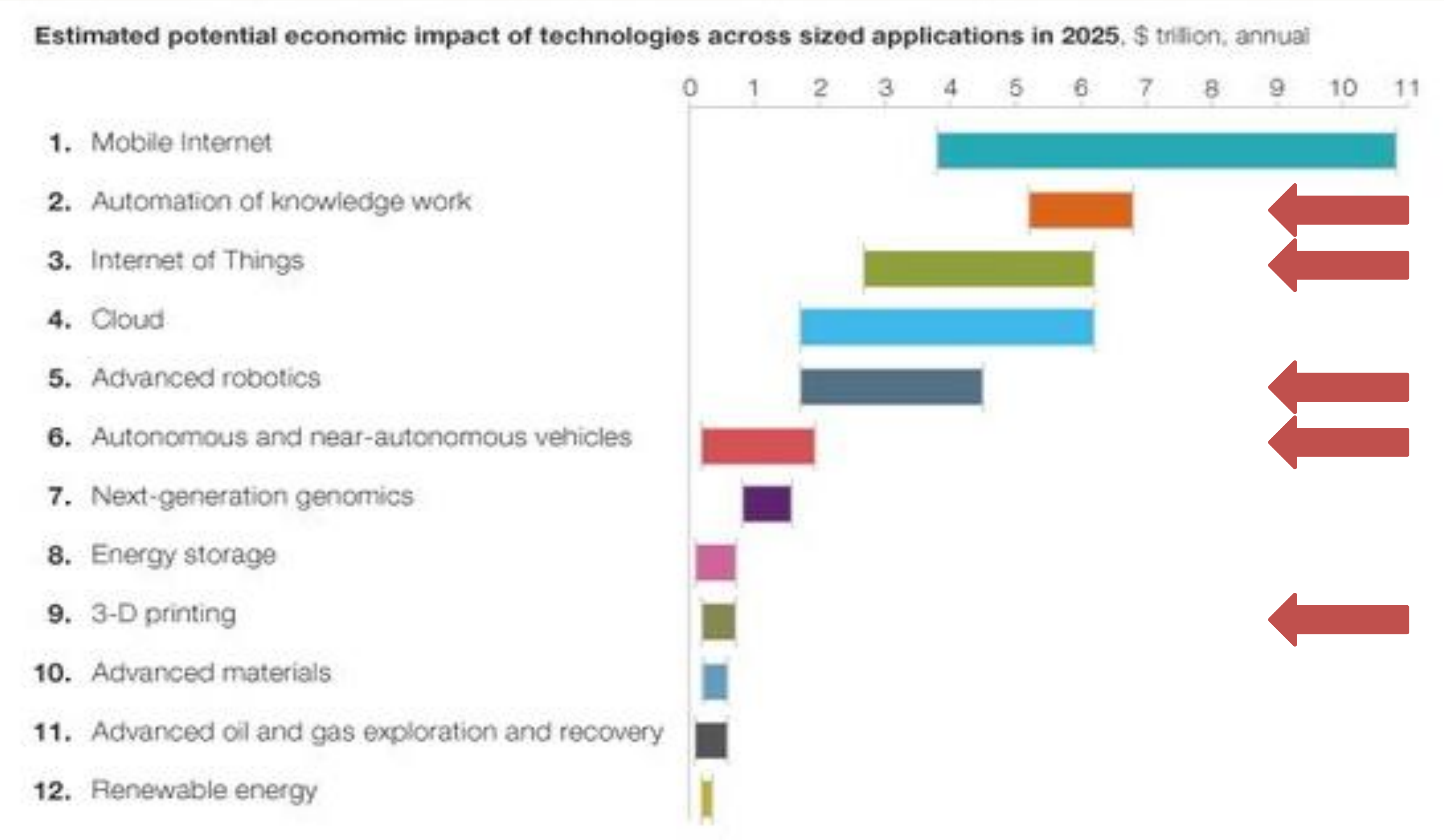
SteveOmohundro.com

SelfAwareSystems.com

http://i791.photobucket.com/albums/yy193/rokib50/sculpture/Lady-Justice-frankfurt_zps970c5d8f.jpg


Economic Drivers
Deep Learning
Regulatory Challenges
New Regulatory Technologies
https://www.flickr.com/photos/danielfoster/14758510078/

McKinsey: $50 Trillion to 2025



http://www.mckinsey.com/insights/business_technology/disruptive_technologies

AI Knowledge Work: $25 Trillion to 2025

Marketing, ERP, Big Data, Smart Assistants

http://thismasquerade.me/wp-content/uploads/blogger/-84ftTg_azwY/Um8up3j4zTI/AAAAAAAABJ8/9_-dxg23UaY/s1600/9979402_ml.jpg

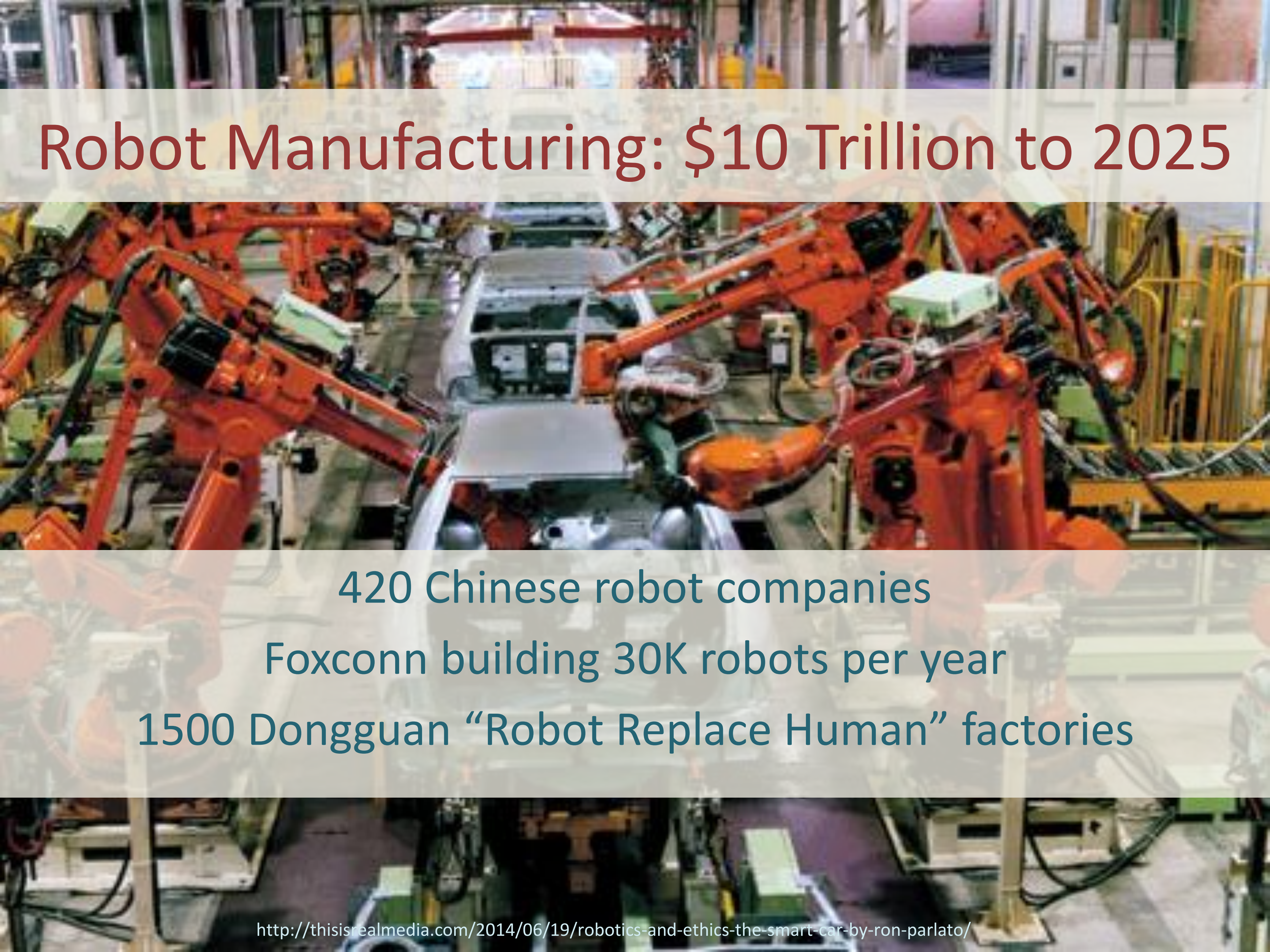
Internet of Things: $15 Trillion to 2025

100 Billion devices by 2025

Cars, Appliances, Cameras, Meters, Wearables, etc.

http://www.forbes.com/sites/gilpress/2014/08/22/internet-of-things-by-the-numbers-market-estimates-and-forecasts/

https://www.summitbusiness.net/images/Internet.jpg



Robot Manufacturing: $10 Trillion to 2025

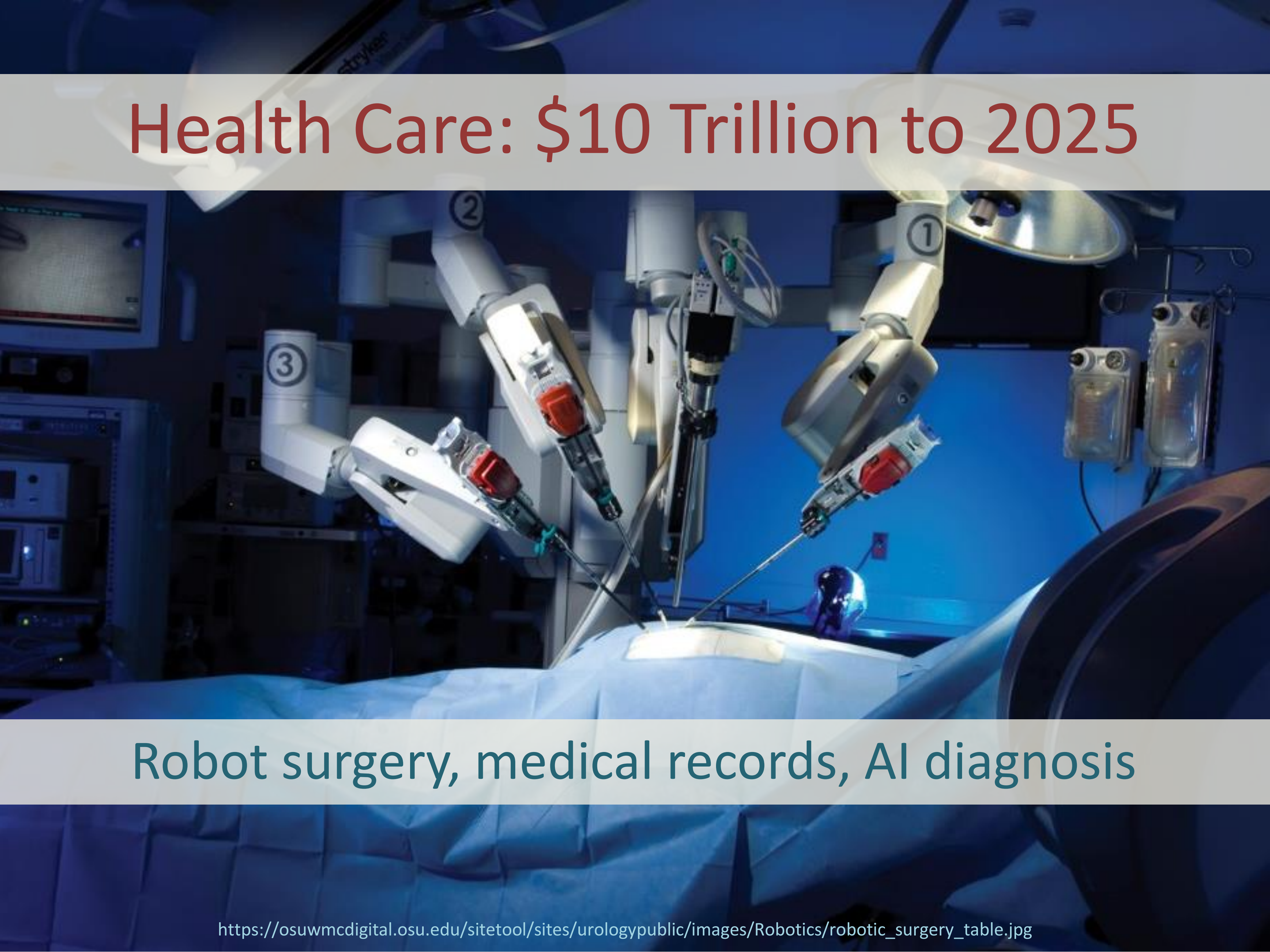
420 Chinese robot companies

Foxconn building 30K robots per year

1500 Dongguan “Robot Replace Human” factories

http://thisisrealmedia.com/2014/06/19/robotics-and-ethics-the-smart-car-by-ron-parlato/

Health Care: $10 Trillion to 2025



Robot surgery, medical records, AI diagnosis

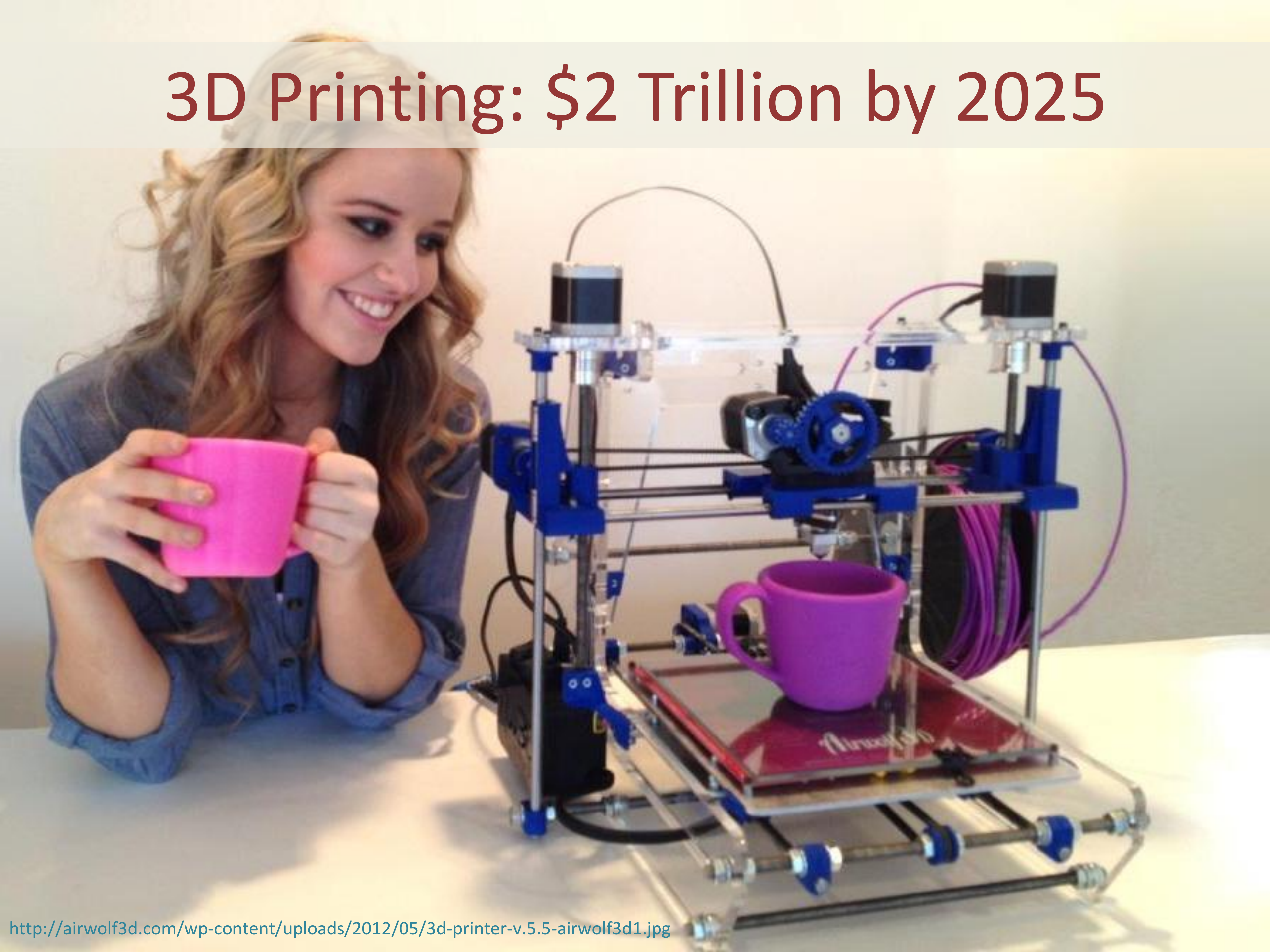
https://osuwmcdigital.osu.edu/sitetool/sites/urologypublic/images/Robotics/robotic_surgery_table.jpg

Self-Driving Vehicles: $10 Trillion by 2025

Disrupt Dealers, Insurance, Parking, Finance, Trucking, Taxis
10 million jobs

http://www.theverge.com/2014/5/28/5756852/googles-self-driving-car-isnt-a-car-its-the-future
http://zackkanter.com/2015/01/23/how-ubers-autonomous-cars-will-destroy-10-million-jobs-by-2025/

3D Printing: $2 Trillion by 2025



http://airwolf3d.com/wp-content/uploads/2012/05/3d-printer-v.5.5-airwolf3d1.jpg


WinSun 3D printed 12,000 sq ft villa
US Building construction: $1 Trillion/yr
5.8 million employees
http://3dprint.com/38144/3d-printed-apartment-building/

Aerial Drones: $98 Billion by 2025

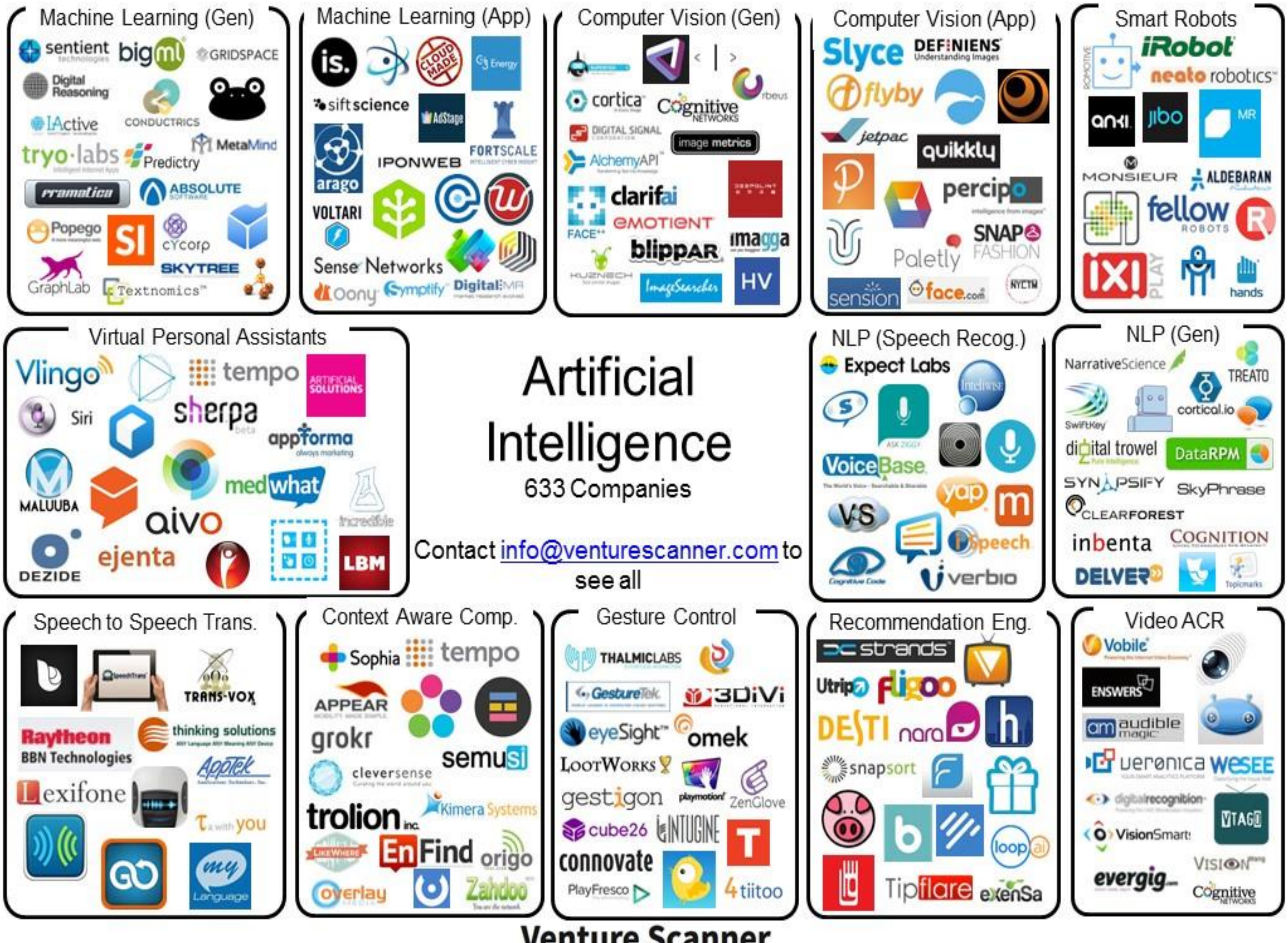
http://www.businessinsider.com/the-market-for-commercial-drones-2014-2

Delivery, Surveillance, Agriculture, Military, Police

$50 Hobby Drones with Video Camera

http://www.flybestdrones.com/best-5-drones-with-camera-under-50-dollars/

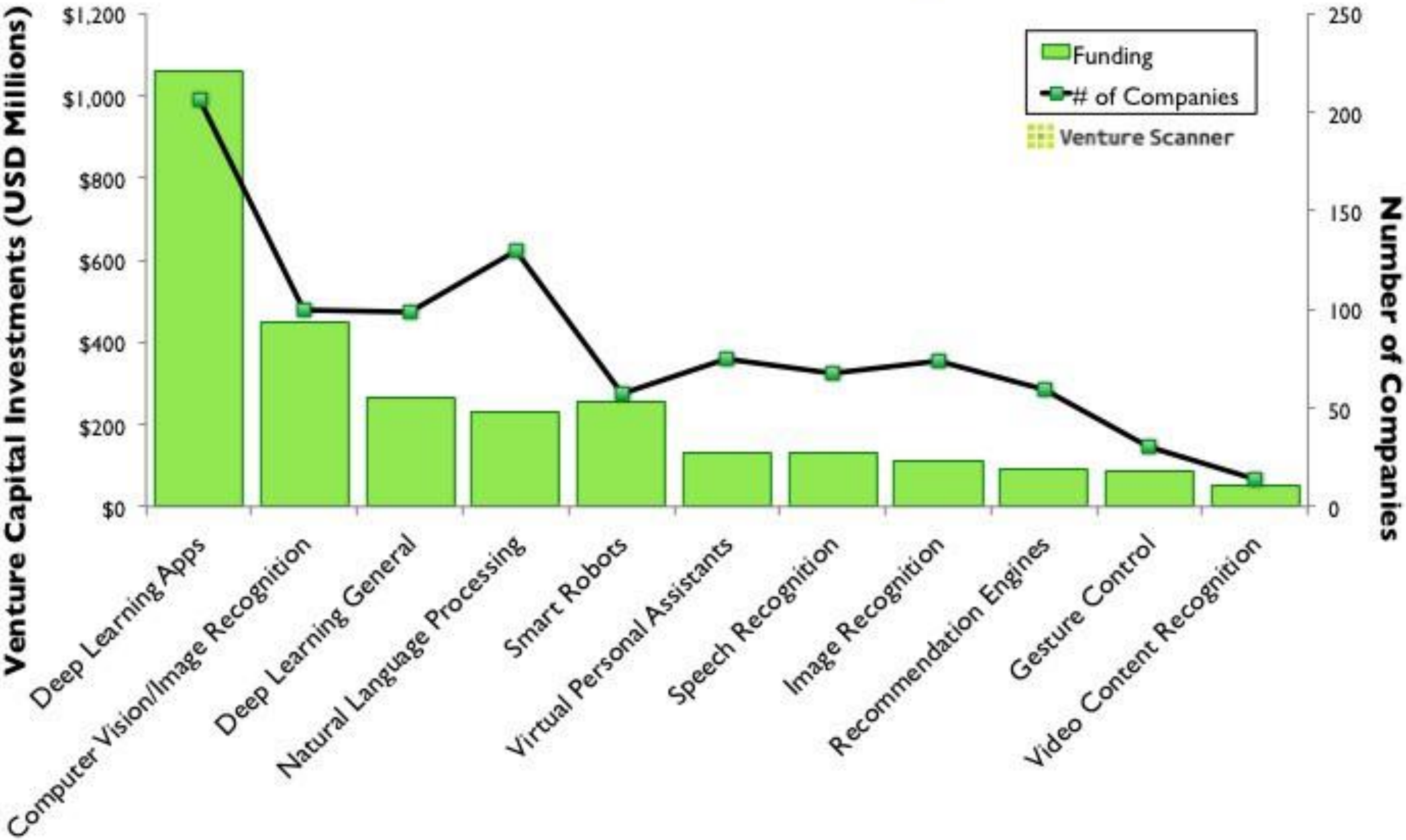
http://mint-tek.com/wp-content/uploads/2015/08/commercialdronesforhire.jpg



https://venturescannerinsights.files.wordpress.com/2015/01/artificial-intelligence-map.jpg https://venturescannerinsights.files.wordpress.com/2015/09/ai4.jpeg

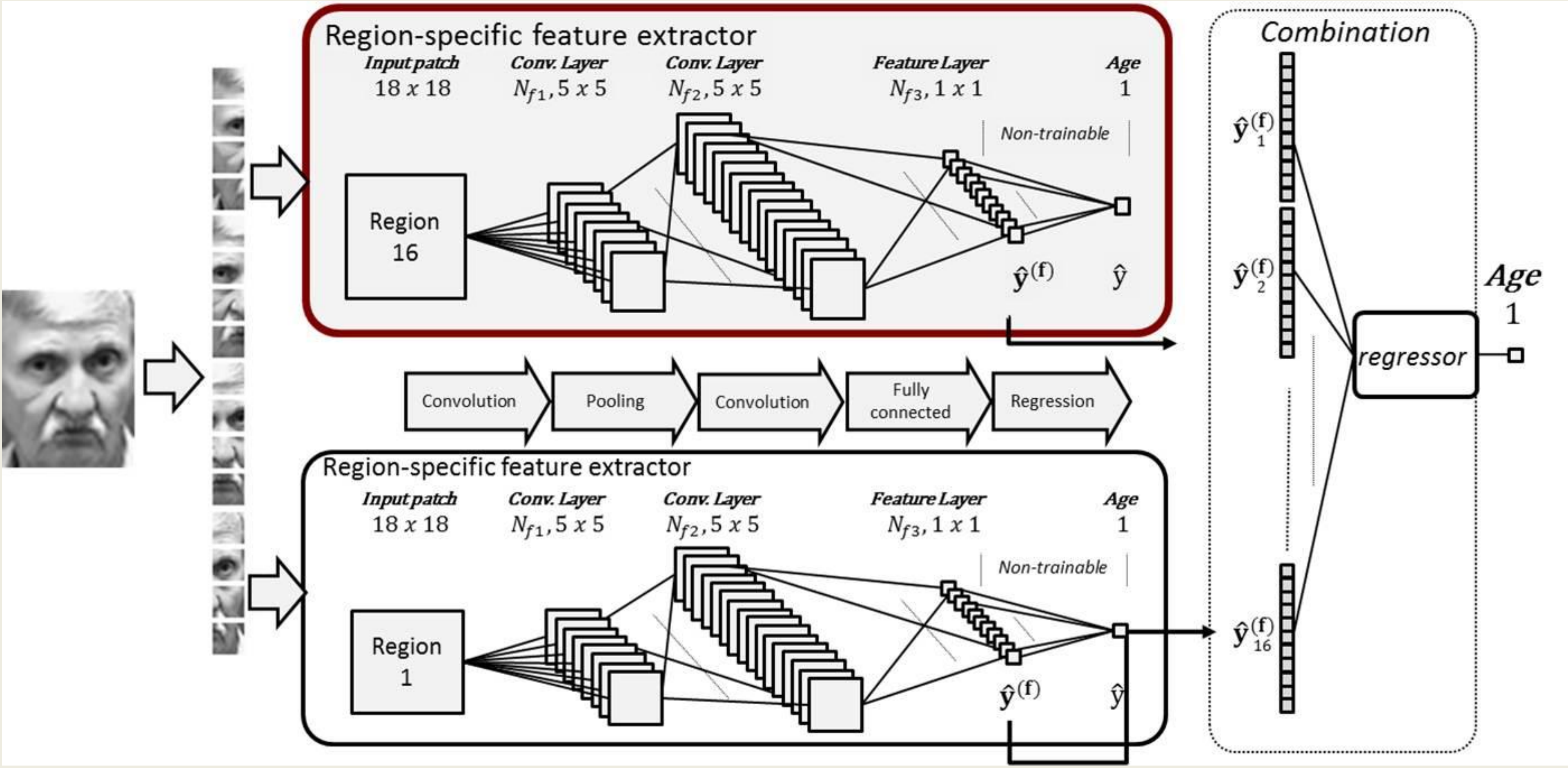
Venture Investing in Artificial Intelligence

Venture Scanner



Contact us at info@venturescanner.com to see all 855 AI Startups

Deep Learning Neural Nets



http://www.nicta.com.au/content/uploads/2015/02/deep.jpg

Deep Learning Successes

- Speech Recognition TIMIT 2009: Cortana, Skype, Google Now, Siri, Baidu, Nuance, etc.
- Image Recognition ImageNet 2012
- Image Captioning 2014
- Natural Language: Sentiment 2013, Translation 2014, Semantics 2014
- Drug Discovery: Merck Challenge 2012
- DeepMind 49 Atari Video Games 2015

Deep Learning Has Blindspots

Full Citation: Nguyen A, Yosinski J, Clune J. *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*. In Computer Vision and Pattern Recognition (CVPR '15), IEEE, 2015.

Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images

Anh Nguyen
University of Wyoming
anguyen8@uwyo.edu

Jason Yosinski
Cornell University
yosinski@cs.cornell.edu

Jeff Clune
University of Wyoming
jeffclune@uwyo.edu

arXiv:1412.1897v4 [cs.CV] 2 Apr 2015

Abstract

Deep neural networks (DNNs) have recently been achieving state-of-the-art performance on a variety of pattern-recognition tasks, most notably visual classification problems. Given that DNNs are now able to classify objects in images with near-human-level performance, questions naturally arise as to what differences remain between computer and human vision. A recent study [30] revealed that changing an image (e.g. of a lion) in a way imperceptible to humans can cause a DNN to label the image as something else entirely (e.g. mislabeling a lion a library). Here we show a related result: it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs believe to be recognizable objects with 99.99% confidence (e.g. labeling with certainty that white noise static is a lion). Specifically, we take convolutional neural networks trained to perform well on either the ImageNet or MNIST datasets and then find images with evolutionary algorithms or gradient ascent that DNNs label with high confidence as belonging to each dataset class. It is possible to produce images totally unrecognizable to human eyes that DNNs believe with near certainty are familiar objects, which we call "fooling images" (more generally, fooling examples). Our results shed light on interesting differences between human vision and current DNNs, and raise questions about the generality of DNN computer vision.

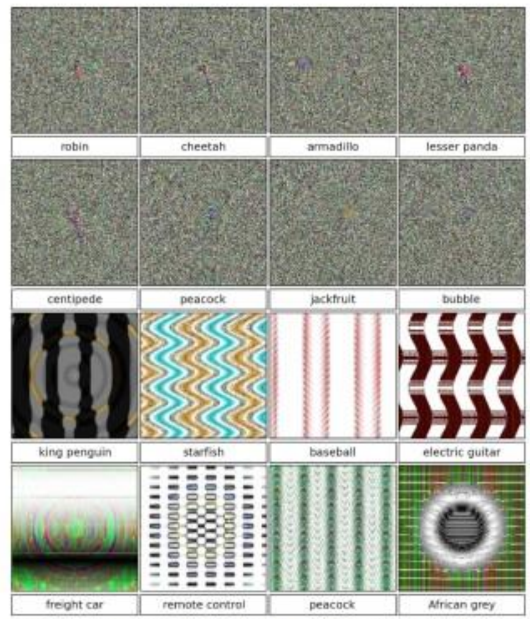


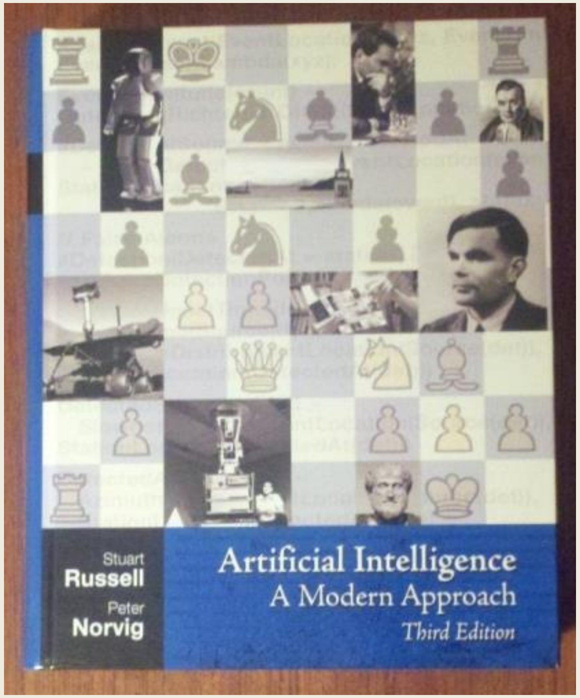
Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

http://arxiv.org/abs/1412.1897

Rational Decision Making

http://commons.wikimedia.org/wiki/File:John_von_Neumann.jpg

1. *Have utility function*
2. *Have a model of the world*
3. *Choose the action with highest expected utility*
4. *Update the model based on what happens*



http://aima.cs.berkeley.edu/

- Von Neumann and Morgenstern, 1944
- Savage, 1954
- Anscombe and Aumann, 1963

Modern Approach to AI

Unintended Consequences

Chess Robot:

Win lots of chess games against good players.

http://www.flickr.com/photos/jiuguangw/4982409890/

Rational Drives

1. *Self-protective*
2. *Goal preservation*
3. *Reproduction*
4. *Resource Acquisition*
5. *Efficiency*
6. *Self-Improvement*

http://www.flickr.com/photos/ahuett/2339608718/

AI Script Kiddies

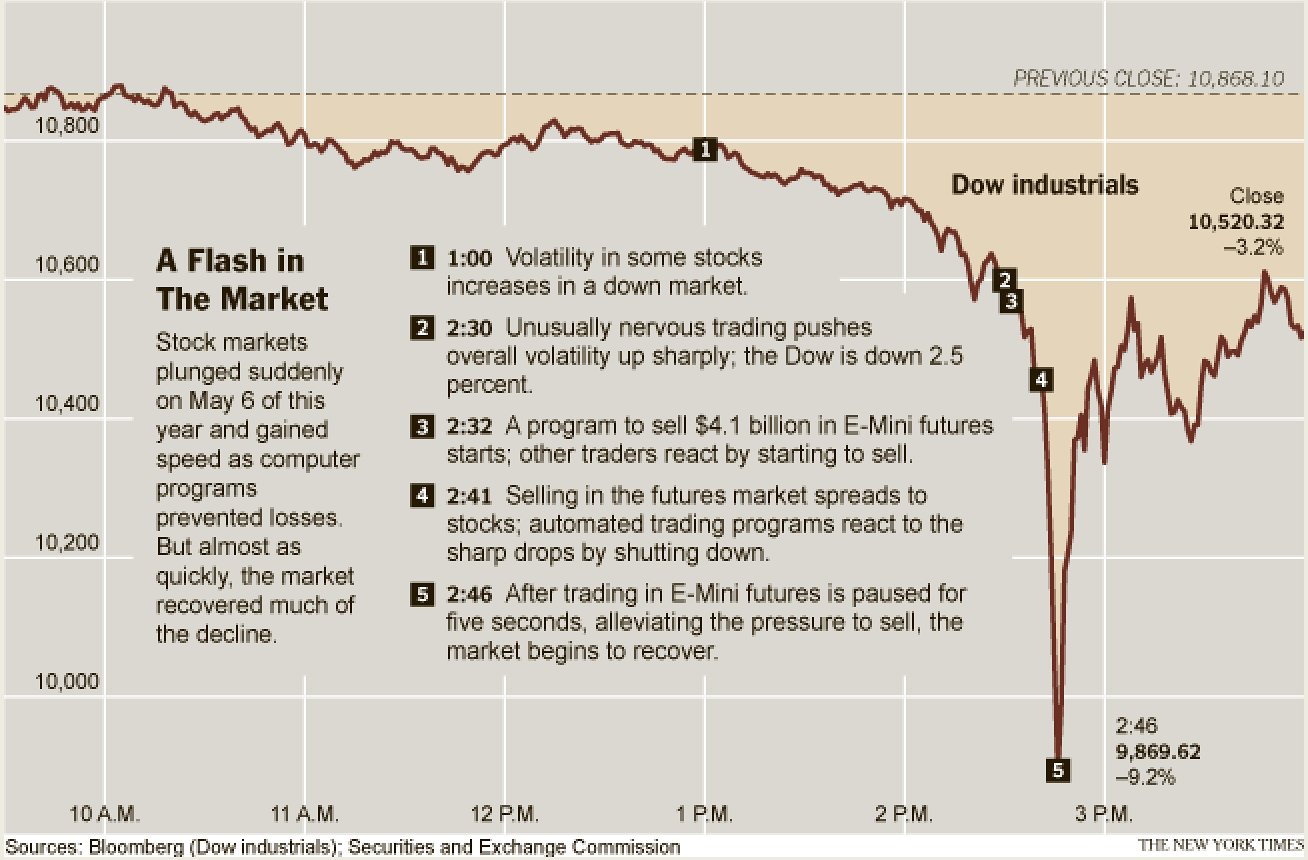
- AI goals and intelligence are independent
- Open source AI easily modified for any goal
- Script Kiddies may create harmful systems

http://www.fastcolabs.com/3013102/why-the-script-kiddie-next-door-is-just-as-dangerous-as-a-chinese-government-hacker

Unexpected Collective Behavior

50% of equity trades are now autonomous

May 6, 2010 Trillion Dollar 9% Flash Crash



http://www.ritholtz.com/blog/wp-content/uploads/2010/10/flash-crash-dow-popup.png

Collective Bot Price Fixing

APRIL 25, 2015

WHEN BOTS COLLUDE

BY JILL PRILUCK

SHARE TWEET

On the day after Easter this year, an online poster retailer named David Topkins became the first e-commerce executive to be prosecuted under antitrust law. In a complaint that was scant on details, the U.S. Department of Justice's San Francisco division charged Topkins with one count of price-fixing, in violation of the Sherman Act. The department alleged that Topkins, the founder of Poster Revolution, which

Can algorithms form price-fixing cartels?

ILLUSTRATION BY BOYOUN KIM

http://www.newyorker.com/business/currency/when-bots-collude

Forbes / Tech

2 FREE Issues

MAY 27, 2014 @ 02:56 PM 16,076 VIEWS

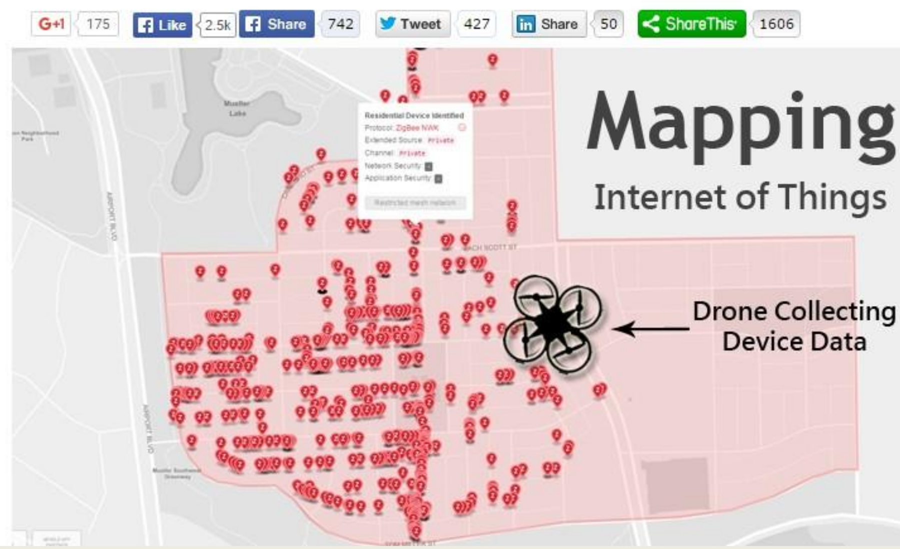
The Half-Baked Security Of Our 'Internet Of Things'

http://www.forbes.com/sites/kashmirhill/2014/05/27/article-may-scare-you-away-from-internet-of-things/

- Hacked Cars
- Hacked Baby Monitors
- Hacked Front Door Locks
- Hacked Stoves
- "Fly-by" Hacking

How Drones Can Find and Hack Internet-of-Things Devices From the Sky

Friday, August 07, 2015 Mohit Kumar



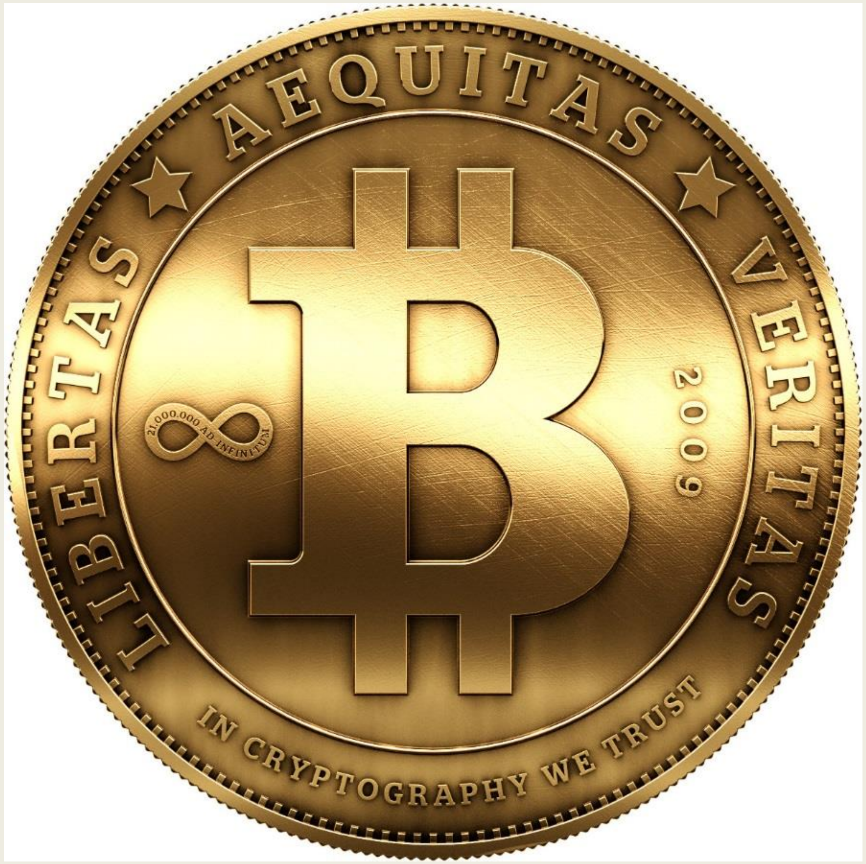
http://thehackernews.com/2015/08/hacking-internet-of-things-drone.html

Liability: Who's Responsible?

- Owner? Manufacturer? Programmer? Neural Net Trainer? Training Data Provider?
- What about bot created bots?
- Unexpected situations?
- New kinds of security vulnerabilities?
- How to track history of events?
- How to regulate rapid software events?

2008: Cryptocurrencies

- Bitcoin and 511 Altcoins
- Decentralized consensus
- “Blockchain” ledger prevents double spending
- “Bitcoin miners” get paid for adding blocks
- “Proof of work” prevents “Sybil” attacks
- Current market cap: $3B



http://blog.newegg.com/blog/wp-content/uploads/bitcoin-logo-3d.jpg

2015: Smart Contracts

- Ethereum
- “Blockchain with a built-in programming language”
- “Consensus-based globally executed virtual machine”
- Contracts in Turing complete programming language EVM
- Summer 2014 presold more than $18 million Ether
- “Decentralized Autonomous Organizations” (DAOs)

http://francebitcoin.com/wp-content/uploads/2014/01/Ethereum.png

New Regulatory Technologies

- Based on Smart Contracts
- Audit trails
- Rogue bot rejection
- Limits on replication
- Identity
- Contract termination
- Contract reversal
- New incentive structures

http://i.ebayimg.com/images/i/190622884591-0-1/s-l1000.jpg