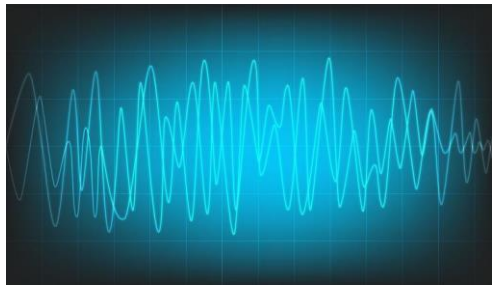
What Every Leader Needs to Know About Artificial Intelligence & Machine Learning

Andrew Puryear, Ph.D.

(andrew.puryear@ultra-us-gbs.com)

A little context…

Senior Engagement

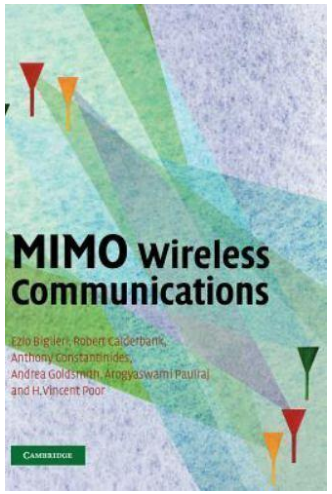


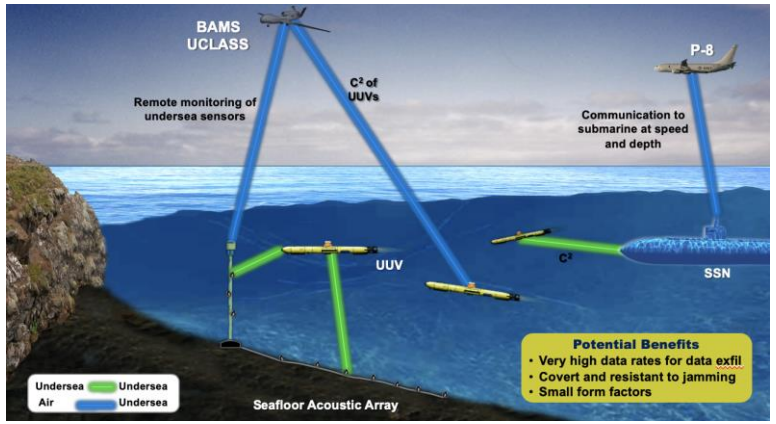
Carnegie Mellon University

US Navy Reserve

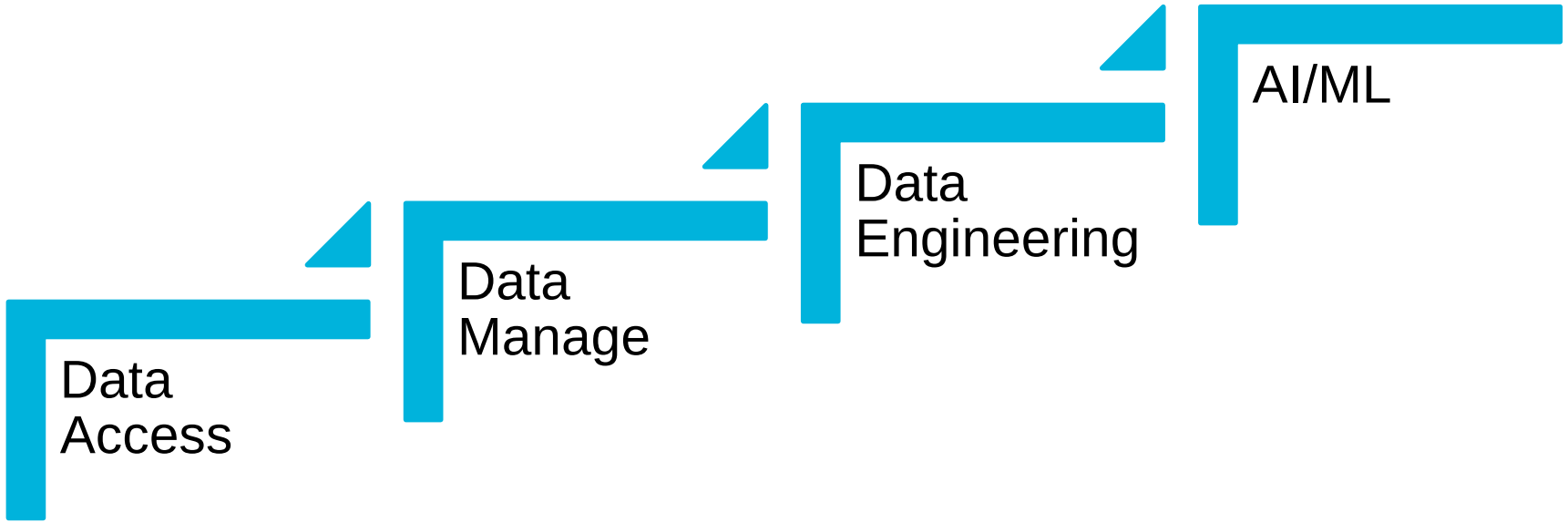
Texas → BS Texas A&M → MS Stanford → PhD MIT → MIT/LL → Sandia → MITRE → Ultra ELECTRONICS

Ultra ELECTRONICS





The AI Pipeline



BLUF--AI/ML presents two sorts of opportunities:

1. **Force multiplier:** AI Technologies might make existing tasks simpler, more reliable, or more efficient.
2. **Game changer:** AI technologies might be used to introduce wholly new capabilities.

Overview of AI State of the Art

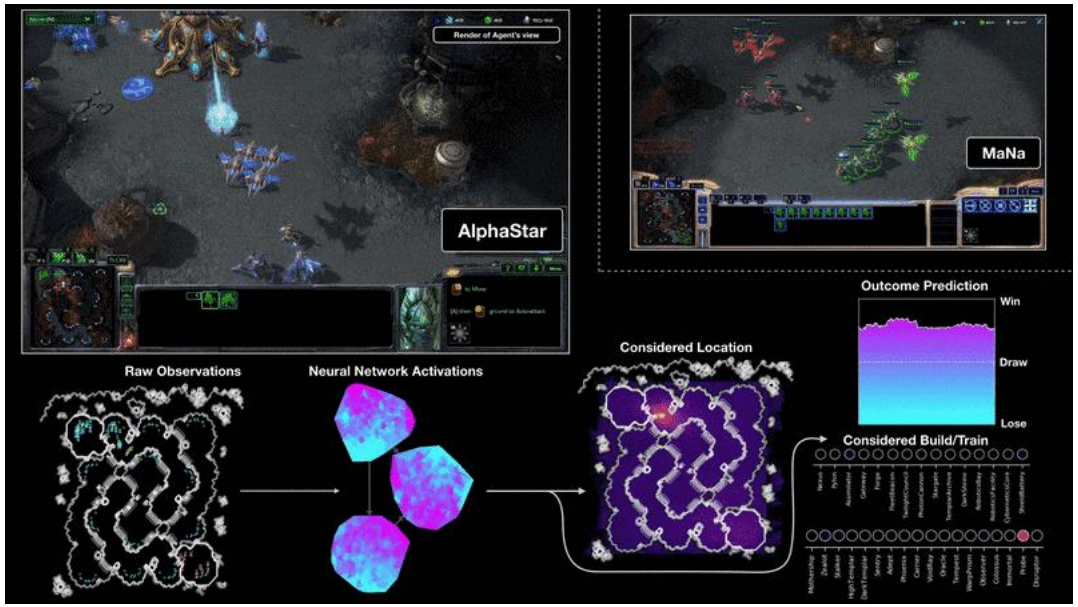
- DeepMind & Google Duplex
- Manual dexterity in robots
- Advances in creativity
- Novel/Re-emergent Algorithms
- Massive improvements in scale
- AI Safety & Explainable AI
- Multi-task learning

Ultra ELECTRONICS

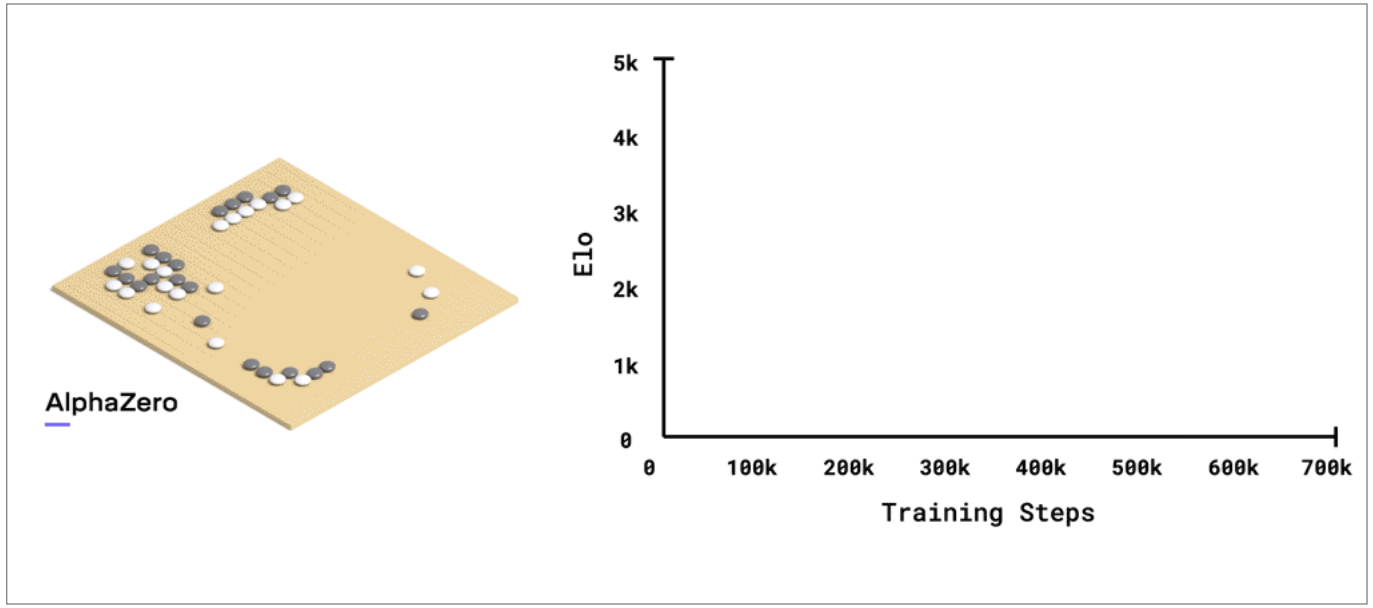
AI State of the Art: DeepMind & Google Duplex

Ultra ELECTRONICS

AlphaStar[3]



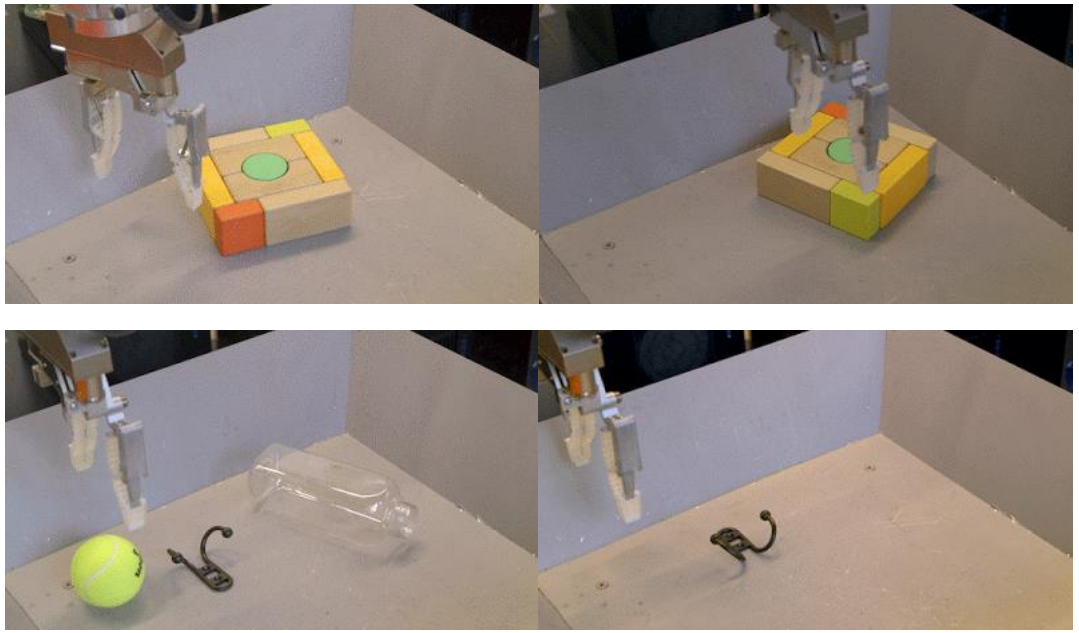
AlphaZero[19]



GoogleDuplex[4]

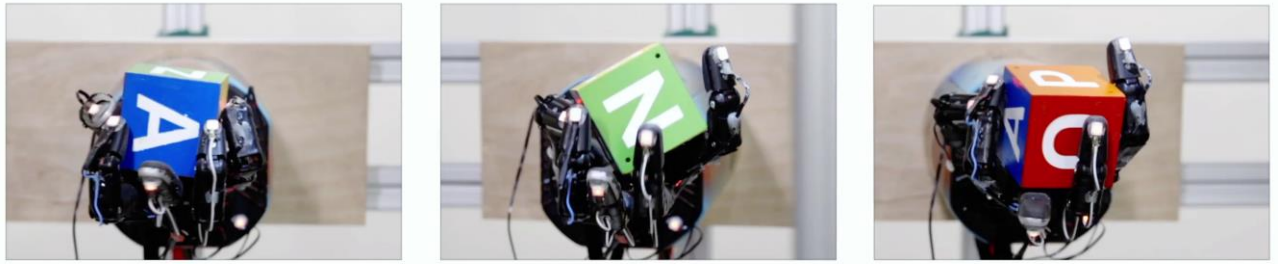
AI State of the Art: Manual Dexterity in Robots

Examples of singulation.



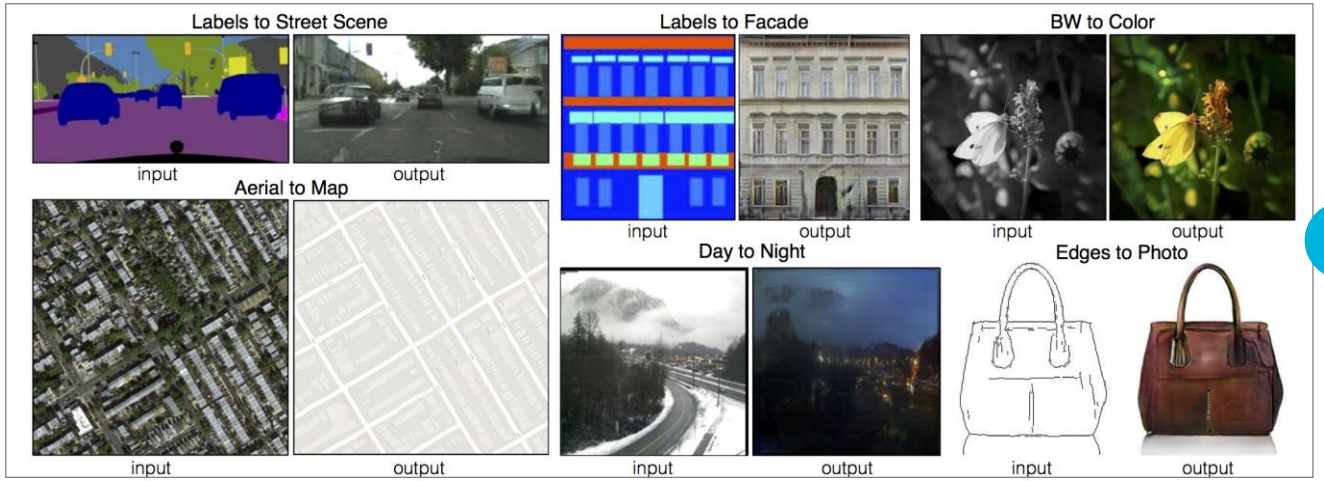
Examples of the learned behaviors. The policy tries several grasps until it succeeds at picking up the tricky object.

QT Opt[5]

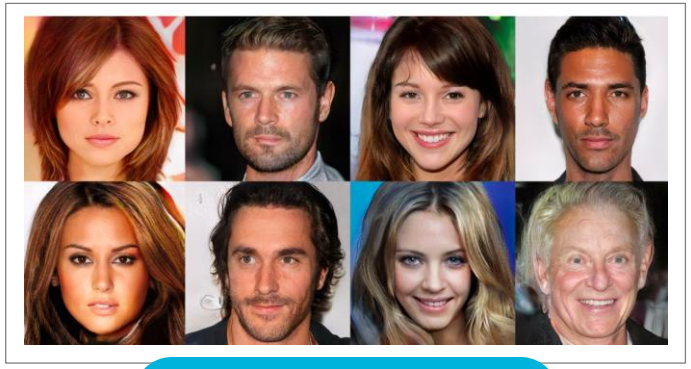


Dactyl[6]

AI State of the Art: Advances in Creativity



Pix2Pix[10]



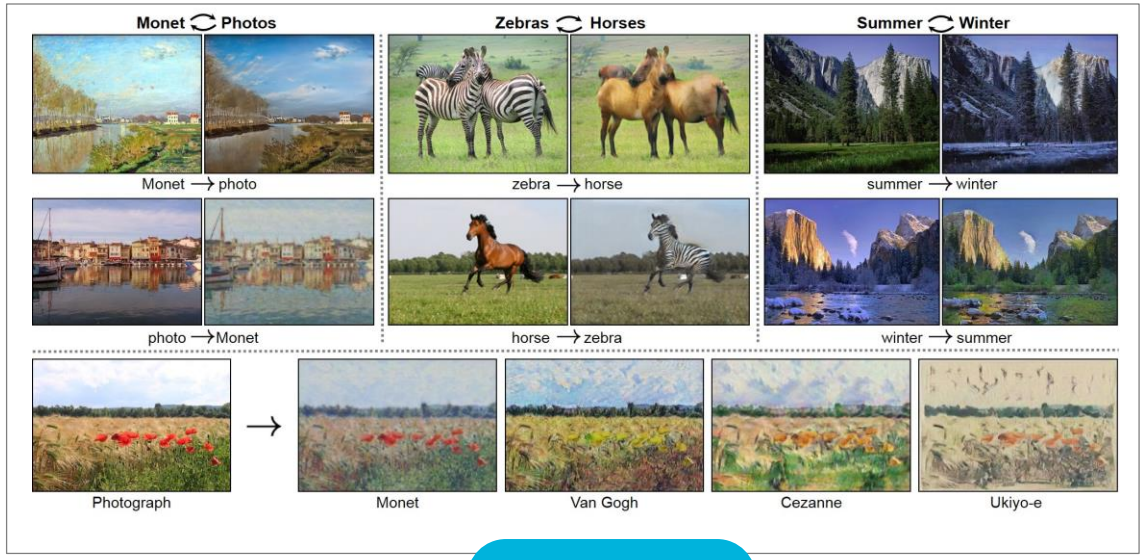
Progressive GAN[8]

A small yellow bird with a black crown and a short black pointed beak

This flower has long thin yellow petals and a lot of yellow anthers in the center

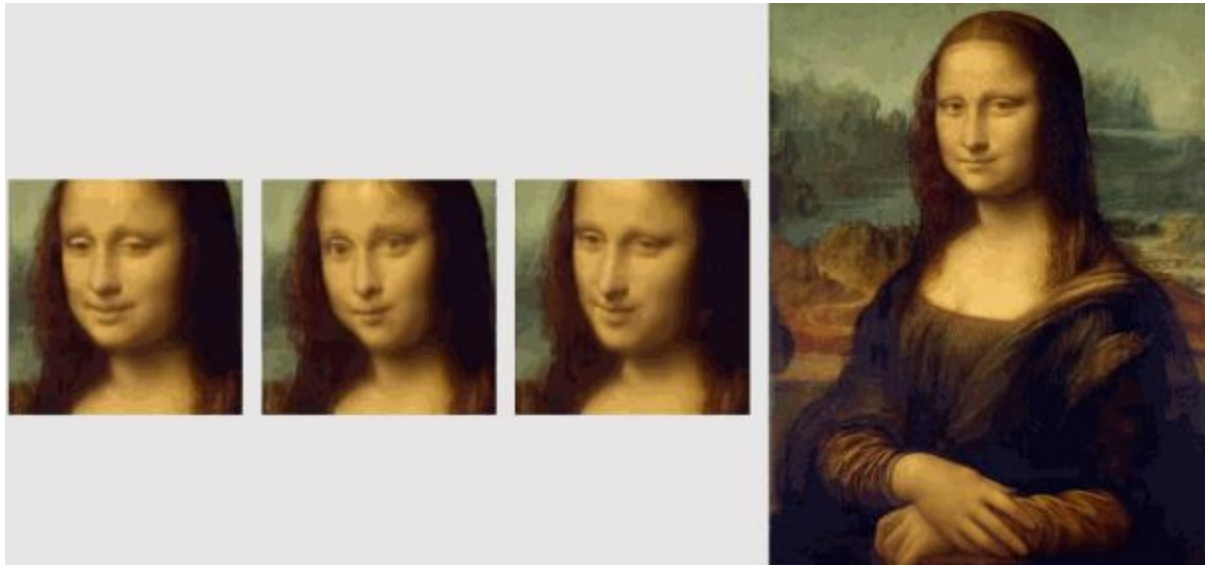
A white bird with a black crown and a yellow beak

StackGAN: Text to Image[7]



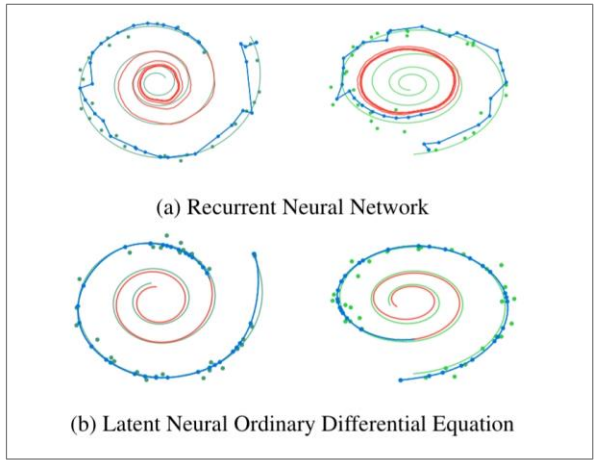
CycleGAN[9]

AI State of the Art: Advances in Creativity

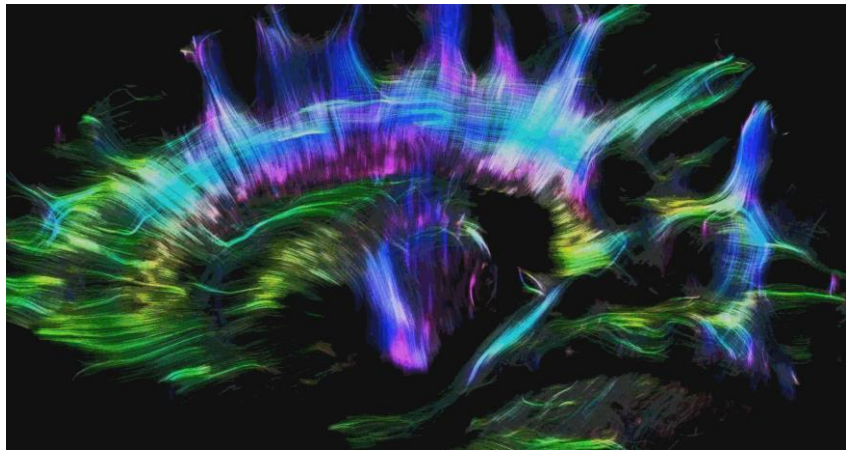


Few Shot Adversarial Learning[34]

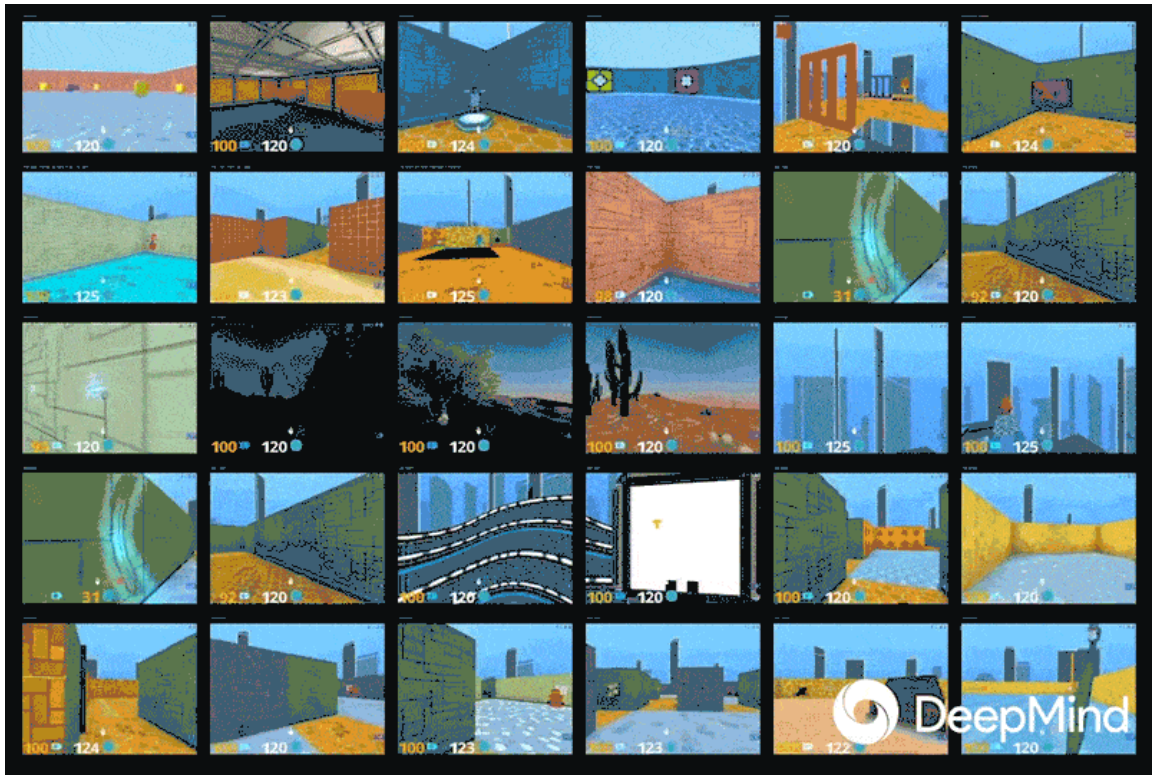
AI State of the Art: Interesting Algorithms



Neural Ordinary Differential Equations[12]

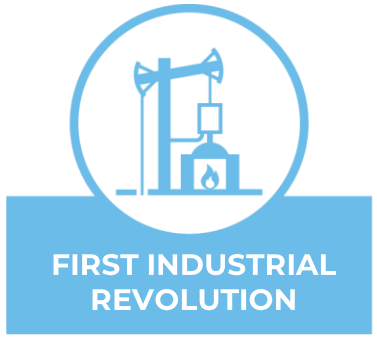


Neuroevolution[13]



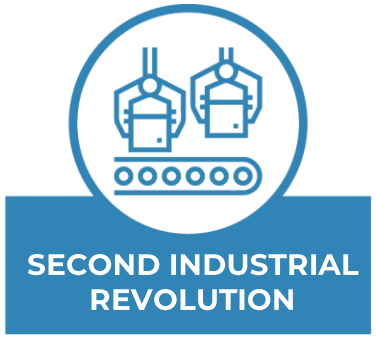
Multi-task Learning (IMPALA)[16]

Security & Warfare on the Brink of the 4th Industrial Revolution[1]



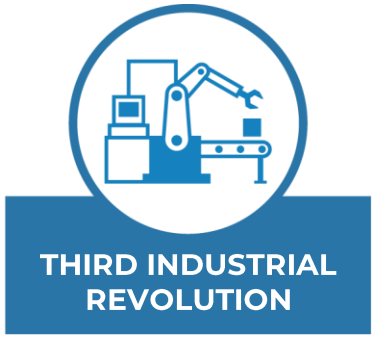
FIRST INDUSTRIAL REVOLUTION

Water and steam power drives mechanized production



SECOND INDUSTRIAL REVOLUTION

Electric power drives mass production



THIRD INDUSTRIAL REVOLUTION

Information and Communication technology drives automated production

FOURTH INDUSTRIAL REVOLUTION

Convergence of cyber, physical and AI drives exponential innovation across every sector

ARTIFICIAL INTELLIGENCE

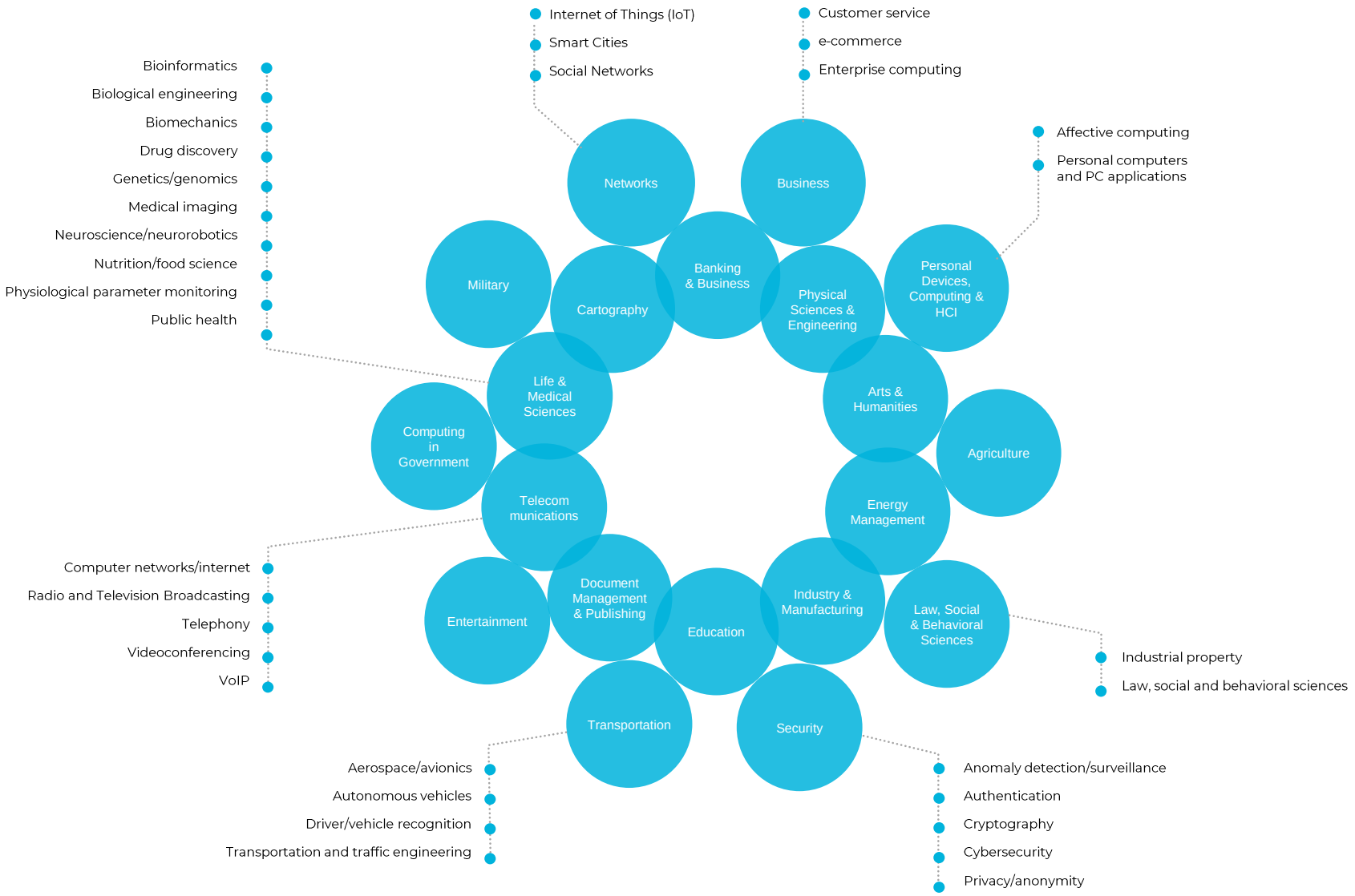
Military operations enabled by AI may unfold so quickly that effective responses require taking humans out of the decision cycle. Adversaries will race to employ these capabilities and the powerful operational advantages they may confer.

CYBER-PHYSICAL

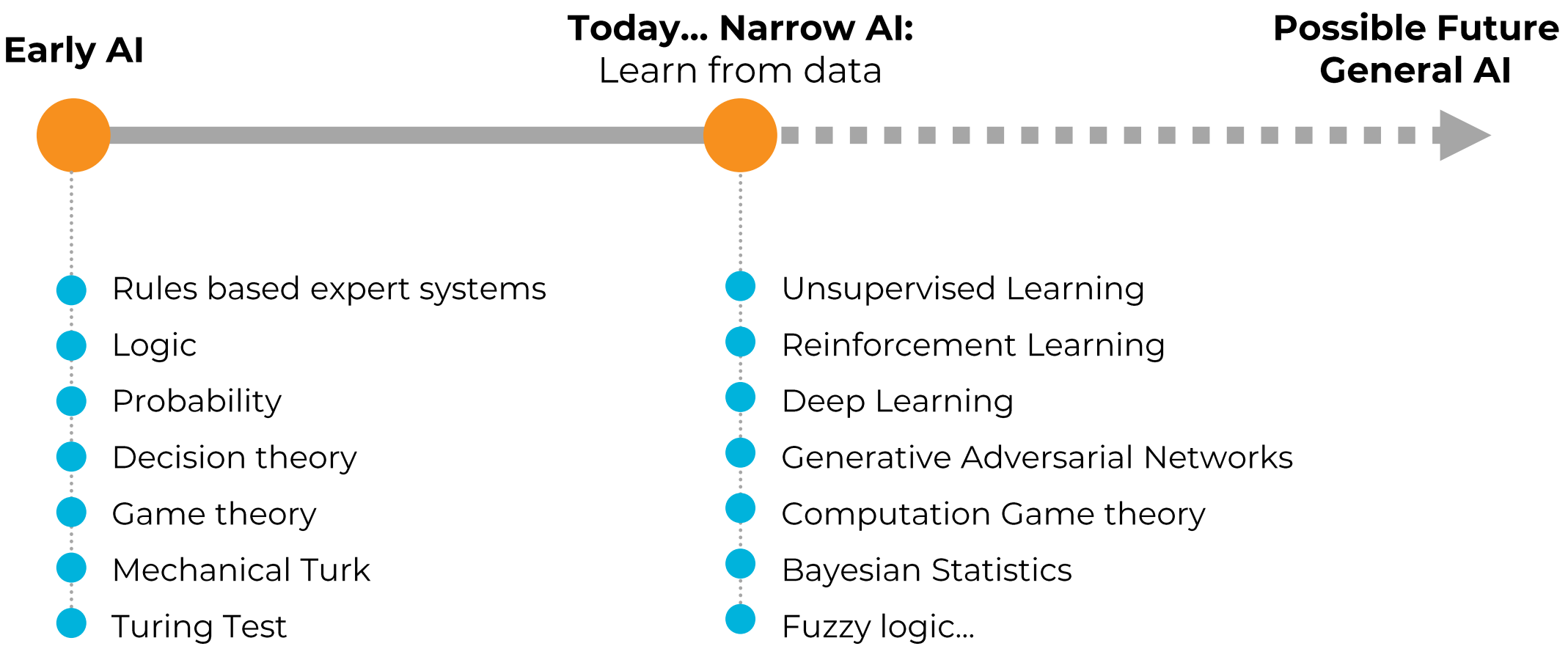
Convergence of distributed networks will touch billions of people. Attacks could hold our way of life.

Ultra ELECTRONICS

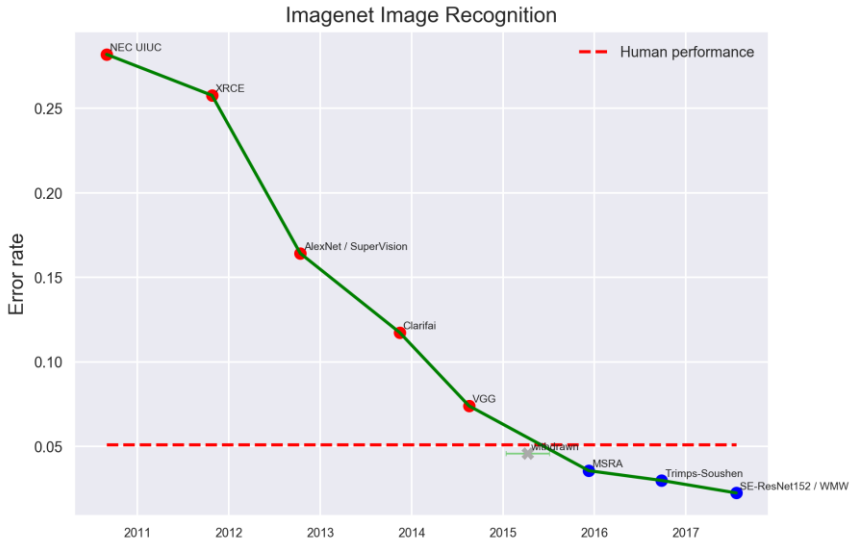
The Artificial Intelligence Revolution: Impact across the economy[30]

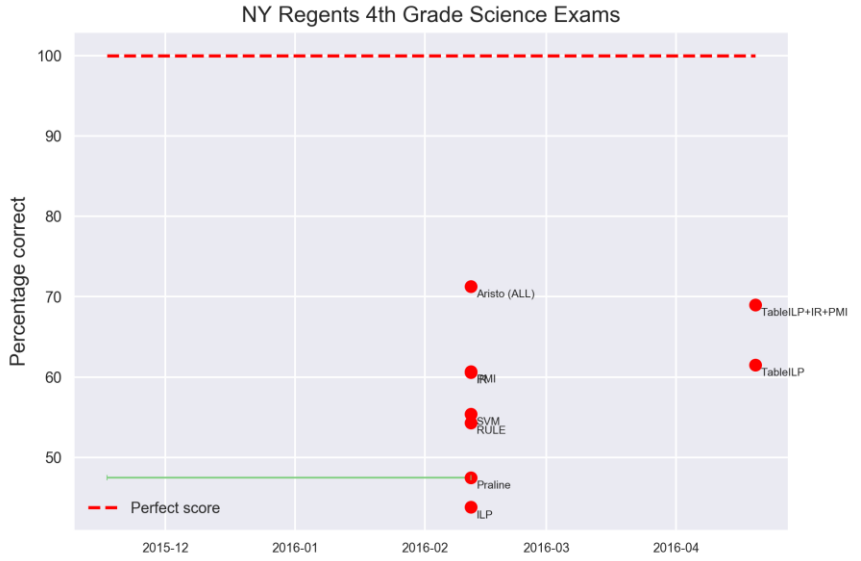


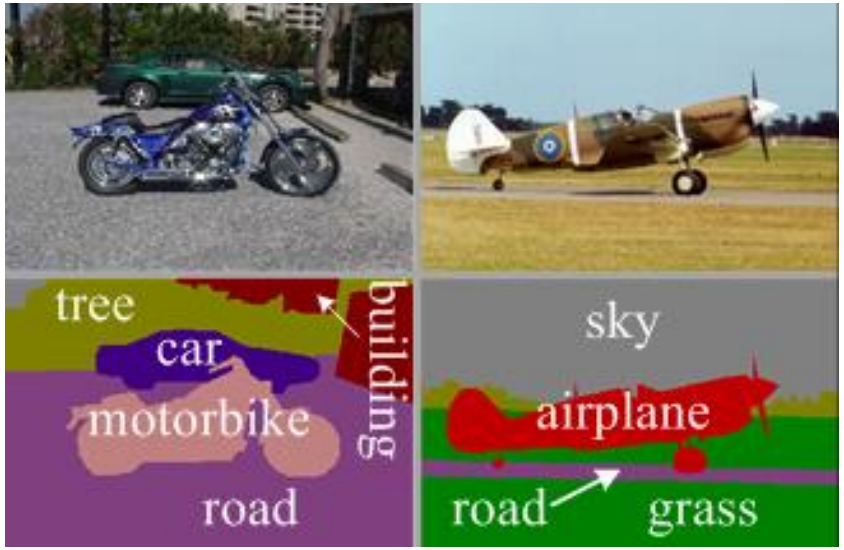
What is Artificial Intelligence?

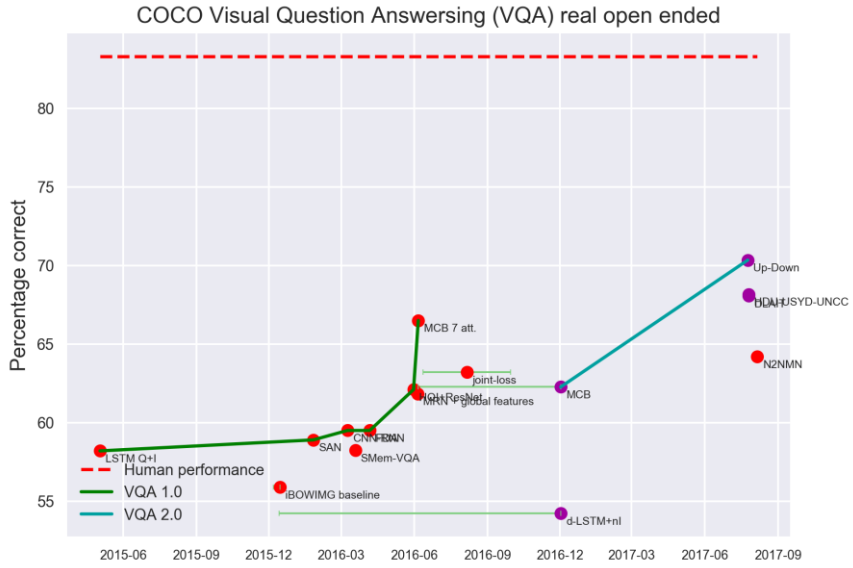


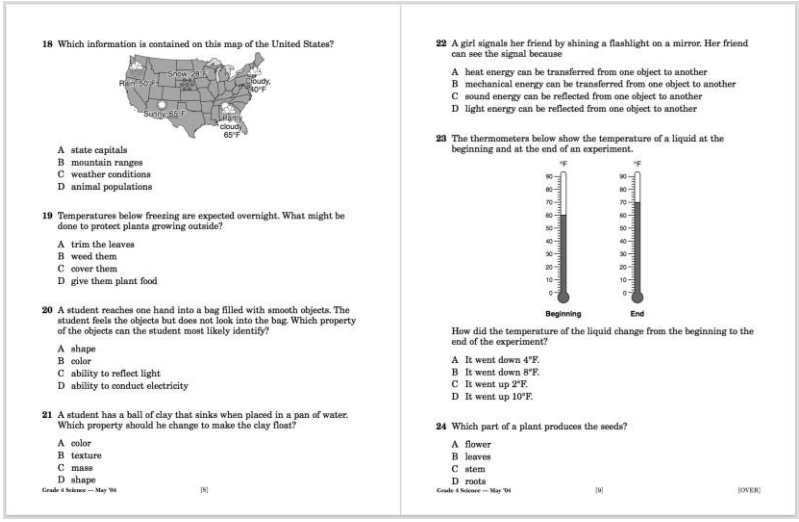
The Artificial Intelligence Advancement over Time (all images[18])



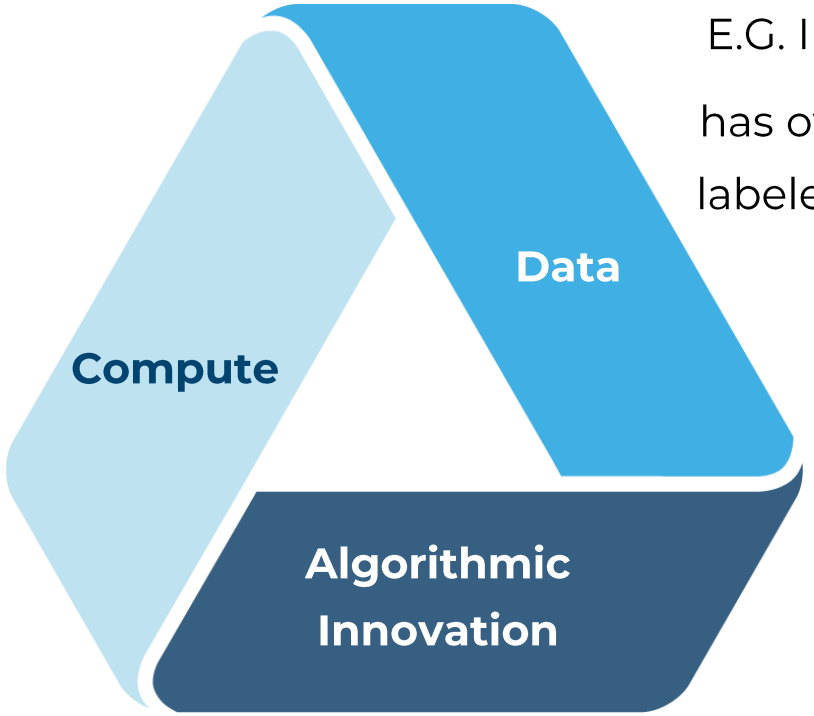




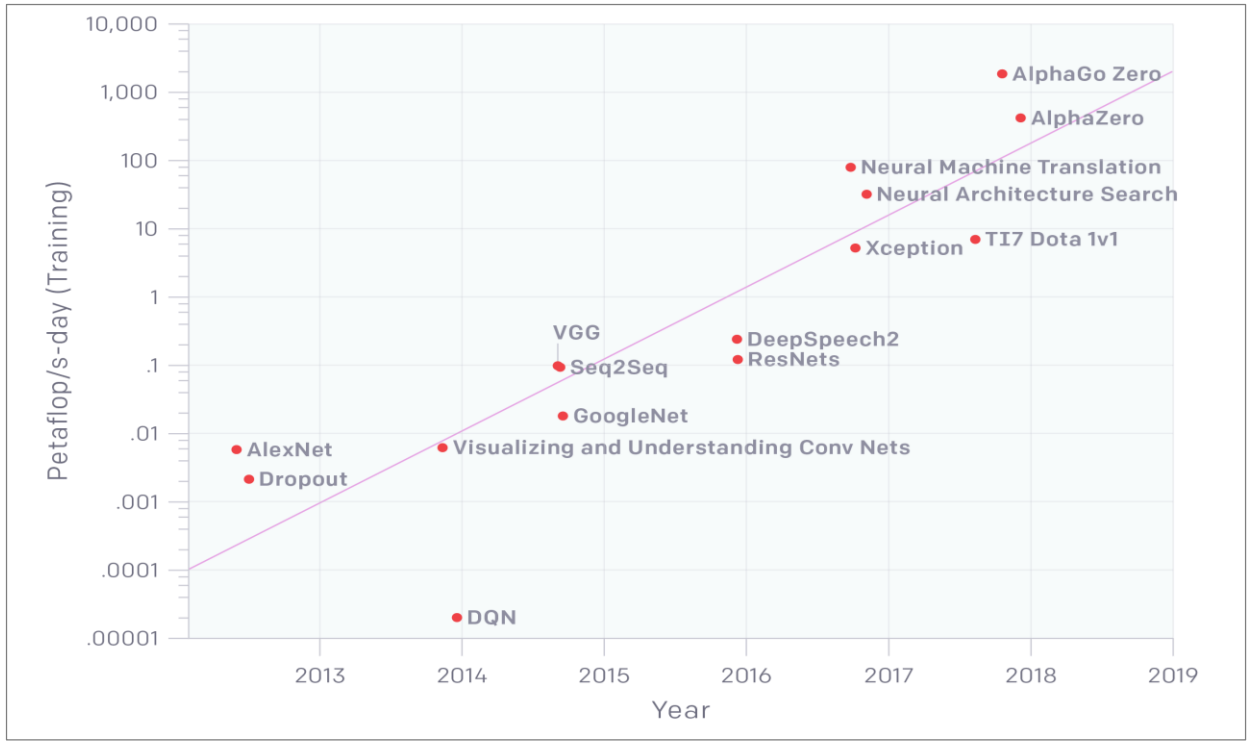




The Artificial Intelligence Revolution: Why Now?

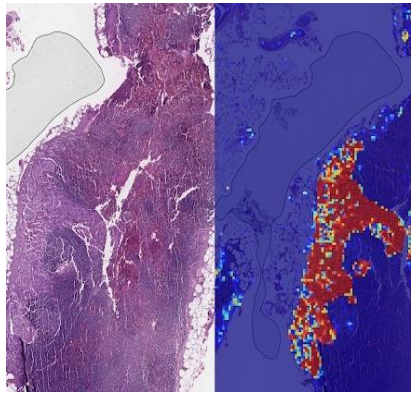


E.G. ImageNet has over **14M** labeled images



AI Compute[11]

AI Uses: Inference, Information, Knowledge Generation[2]



Classification[27]

Decision Support[30]

Synthesis

Anomaly Detection[28]

Prediction[31]

Optimization[29]

Data Mining

Search

AI Uses: Autonomy[2]

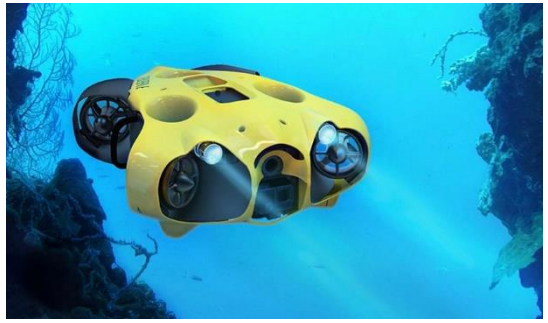
Embedded expertise

Larger scale operations

Faster-than-human reaction times

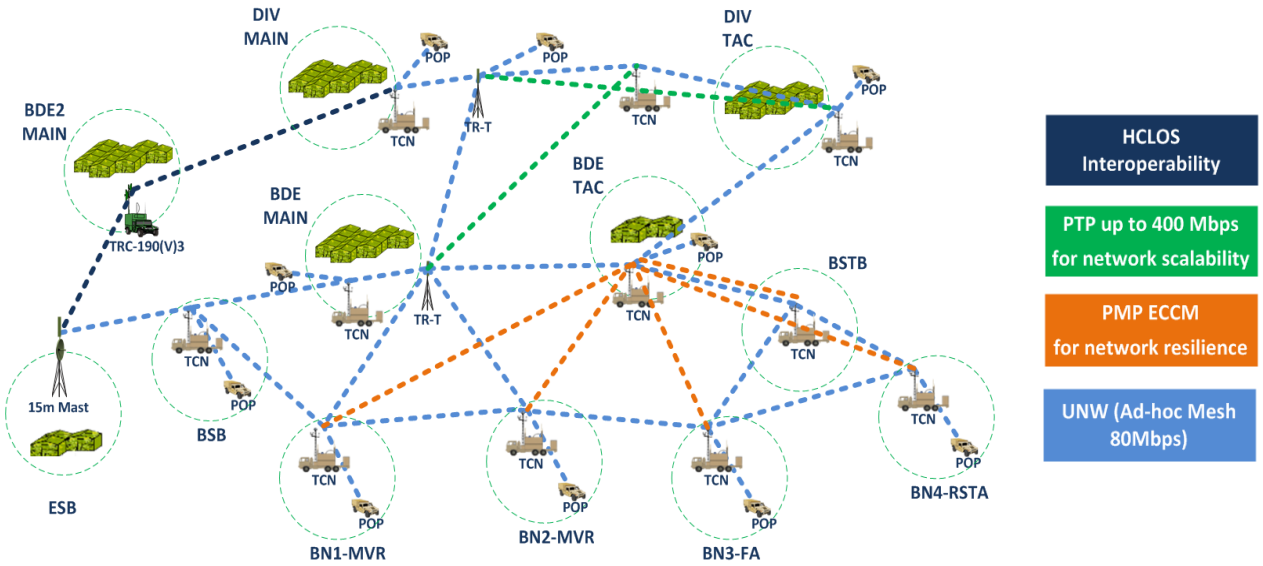
Superhuman precision and reliability

Superhuman patience and vigilance



Operations without connections to humans

Ultra Electronics AI Research: Resilient AdHoc Networking in Congested and Contested Environments



Challenge: Deliver the most critical information to the right place at the right time in a congested and contest electromagnetic environment.

Solution: Real-time, distributed Artificial Intelligence/Machine Learning for tactical radio networks.

Prediction: Predict & prioritize the critical information.

Optimization: Jointly optimize waveforms, protocols, network topologies to maximize throughput of critical information.

Classification: Real-time classification of interferers and jammers for ECCM.

Autonomy: Dynamically reposition mobile relay nodes.

Synthesis: Generate synthetic message traffic to obfuscate information gathering.

Ultra Electronics AI Research: Federated Analytics to Increase the Speed of Decisions at the Tactical Edge

- **Challenge**
  - Users are swimming in sensors, drowning in data
    - Bandwidth does not exist to push all data to all users
  - Provide multi-domain near-real time actionable information to tactical users at the edge
  - The gap between national intelligence and tactical data denies commanders relevant information for effective C2
- **Solution**
  - Leverage distributed AI to **predict** what information is needed where
  - Dynamically push harvested data required for advanced analytics
  - Proactively push actionable information to edge users based on mission requirements/operator queries
  - Deliver decision-quality information across the warfighting spectrum

AI Limitations: Issues, Concerns

Ultra ELECTRONICS

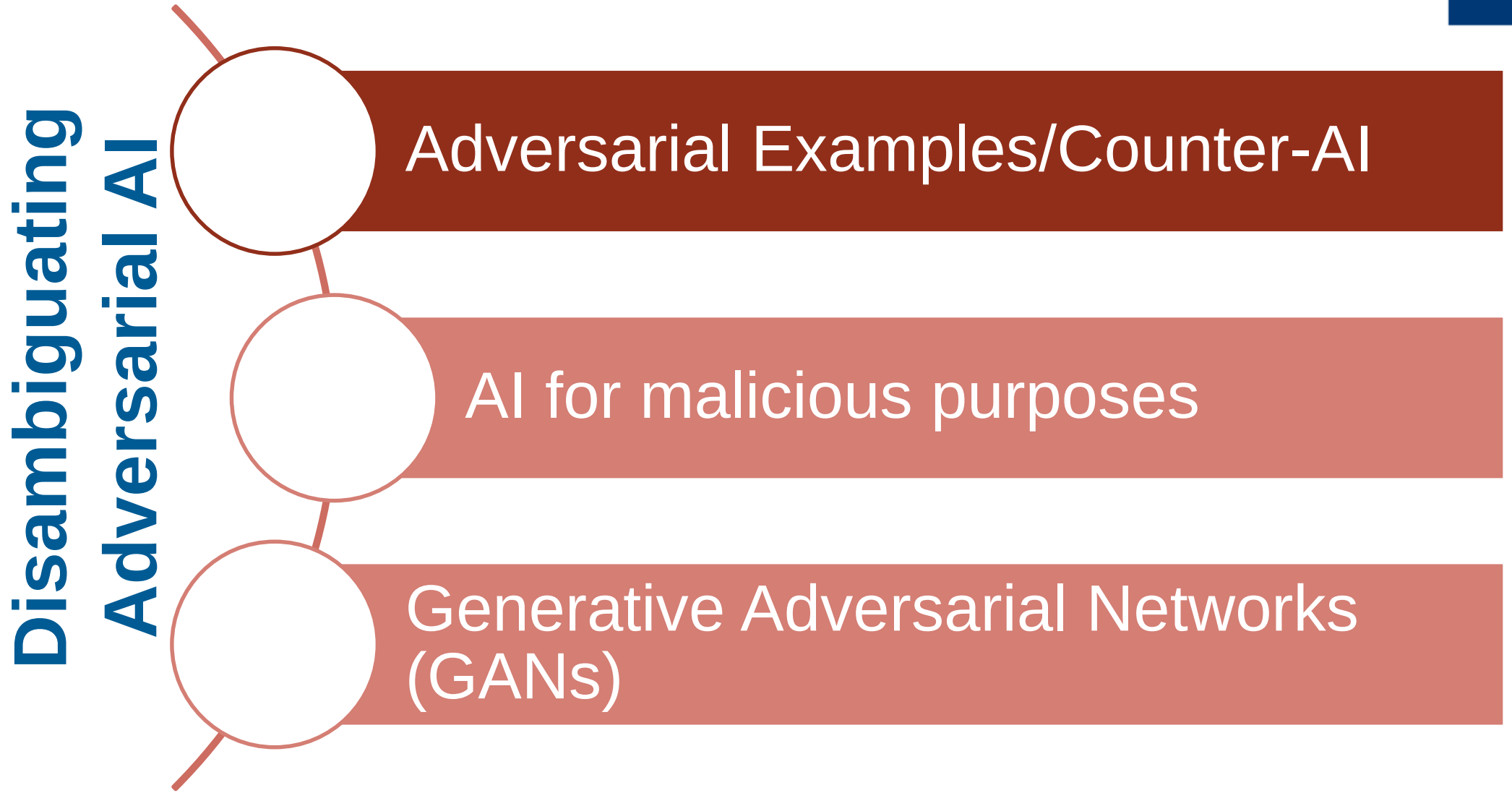
Cause

- **Limitations in current AI tech**
  - Brittleness
  - Explainability
  - Predictability
  - Lack of ‘common sense’ (context awareness)
  - Limited cross-training
- **Common AI engineering issues**
  - Reward hacking
  - Underdefined objective
  - Overfitting
  - Human machine interface failure
  - Non-representative training data

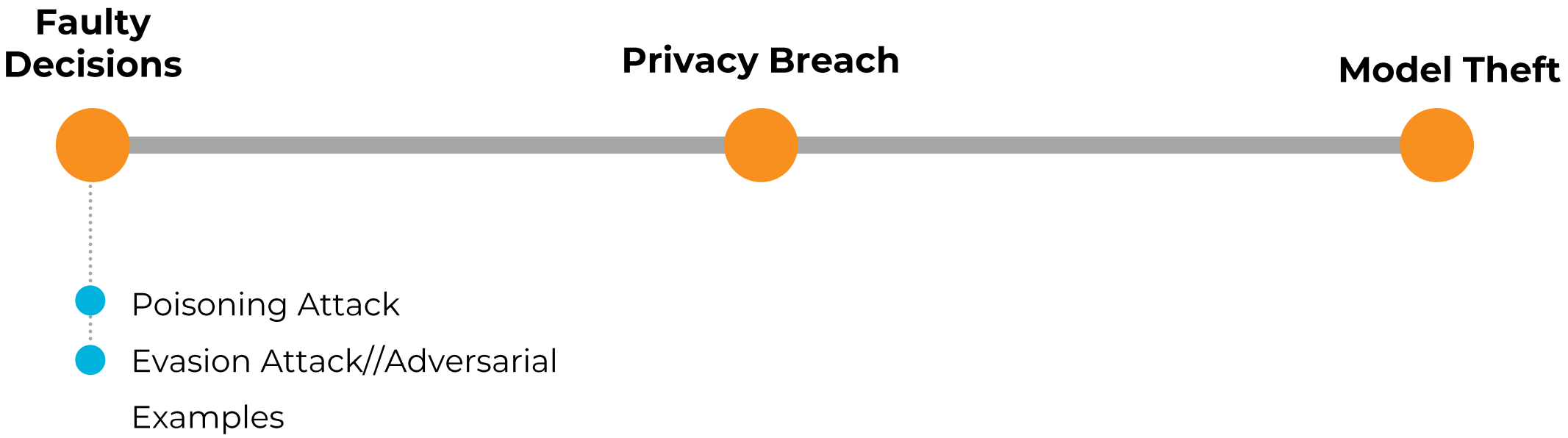
Effect

- **System accidents**
- **AI bias**
- **AI safety**
- **Vulnerability exploitation**
- **Unintended consequences**

Disambiguating Adversarial AI



Counter-AI: Adversary Objectives



Privacy Breach: Stealing data from ML models

(Sensitive) training data → model

Intended:

ML model + query → answer

Possible:

ML model + queries → sensitive training data

Privacy Breach: Stealing data from ML models

Face recognition:

model + name → image

Medicine dosing:

model + demographic info

→ genotype characteristics

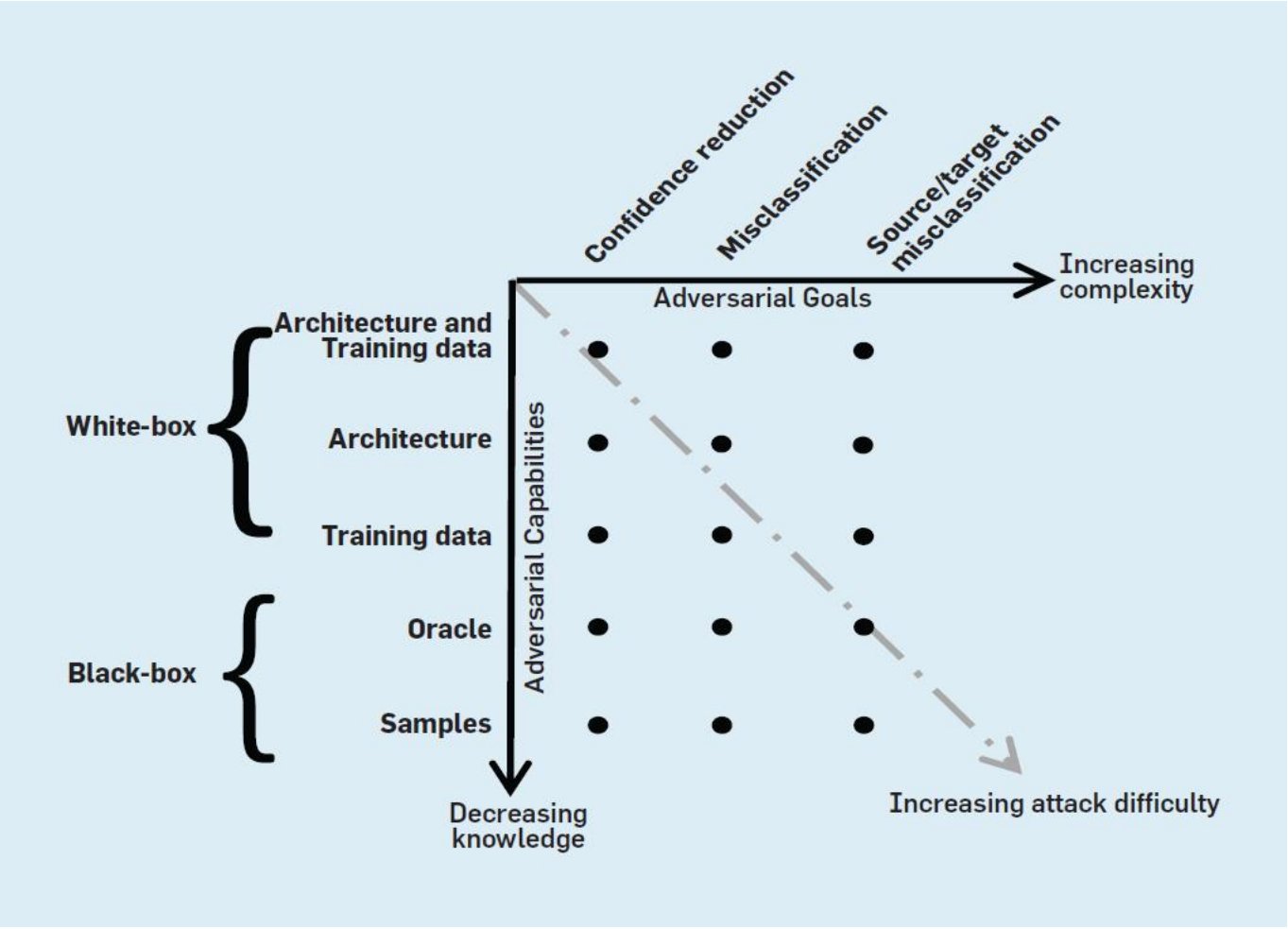
Reconstructed image

Actual image used in training

[Examples from Fredrikson et al., CCS ‘15 and USENIX Security ’14]

Inducing Faulty Decisions

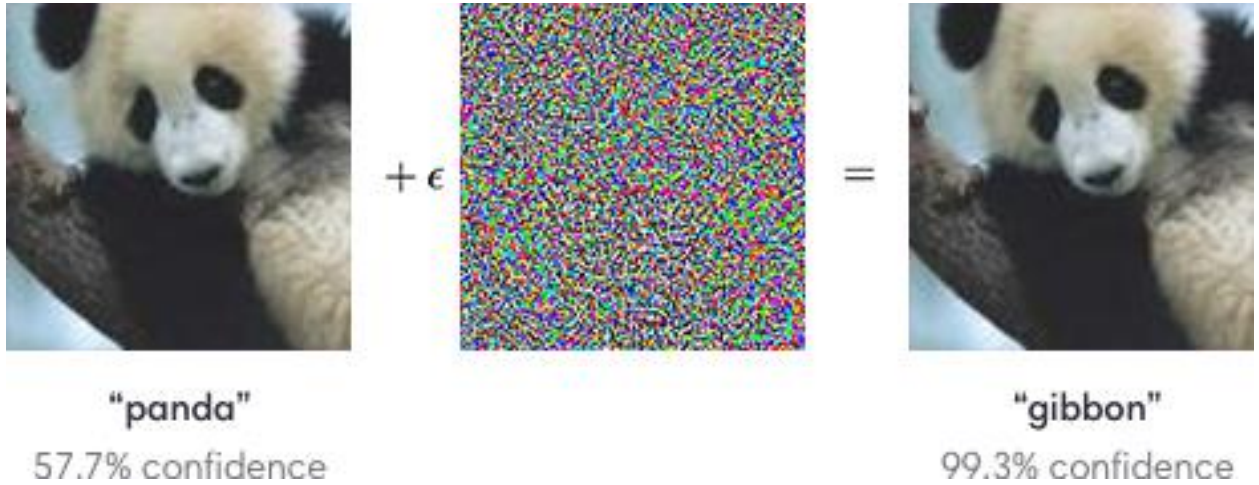
- **Data Poisoning**
  - Training or test time
- **Evasion attack**
  - Test time
- **Traditional cyber attack**
  - Edit model directly, edit query



Taxonomy of adversaries against machine learning at test time[17]

Sample of Adversarial Examples

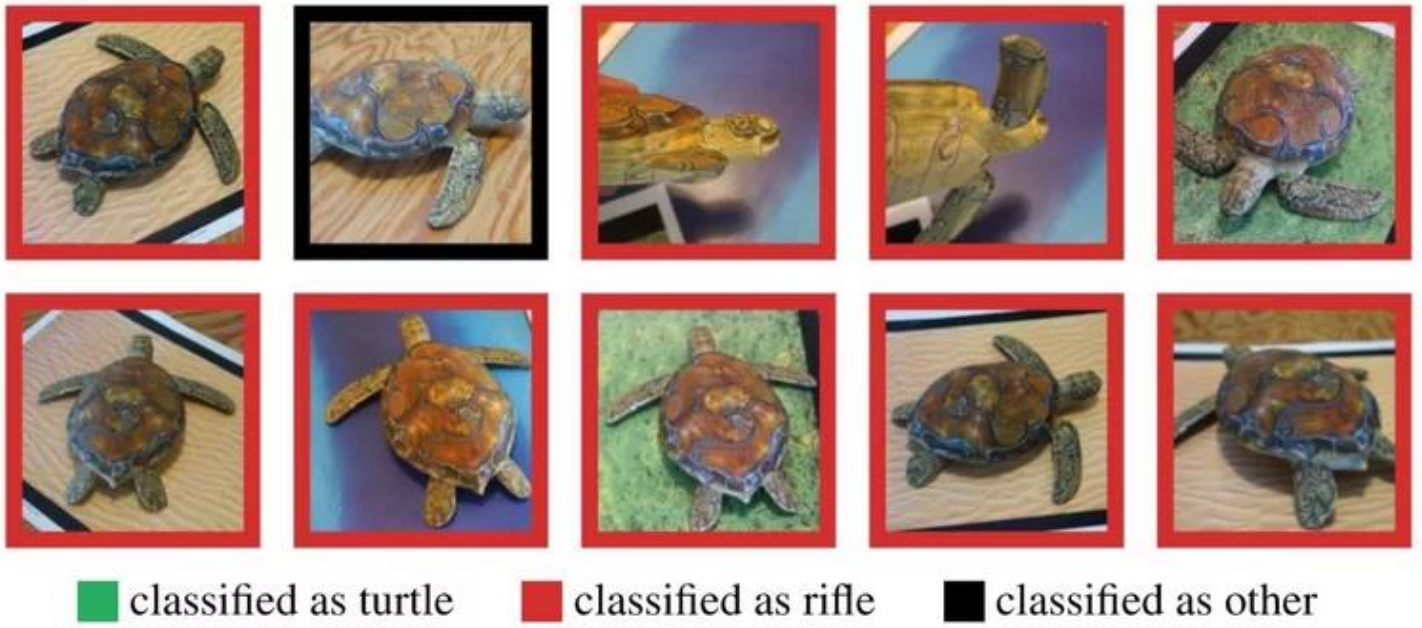
Subtle perturbations cause a neural network to misclassify stop signs as speed limit 45 signs, and right turn signs as stop signs.



An adversarial input, overlaid on a typical image, can cause a classifier to miscategorize a panda as a gibbon.

An example of digital dodging. Left: An image of actor Owen Wilson, correctly classified by VGG143 with probability 1.00. Right: Dodging against VGG143 using AGN's output (probability assigned to the correct class: < 0.01).

Sample of Adversarial Examples



Randomly sampled poses of a 3D-printed turtle adversarially perturbed to classify as a rifle at every viewpoint2 . An unperturbed model is classified correctly as a turtle nearly 100% of the time.[33]

What Every Leader Needs to Know About Artificial Intelligence & Machine Learning

Andrew Puryear, Ph.D.

(andrew.puryear@ultra-us-gbs.com)

References

1. Schwab, K. (2016, 14 Jan). The Fourth Industrial Revolution: what it means, how to respond
2. Scharre, P., Horowitz, M. C., Allen, G. C., Frederick, K., Cho, A., Saravalle, E., & Kania, E. (2018). *Artificial Intelligence: What Every Policymaker Needs to Know*.
3. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/
4. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html
5. https://ai.googleblog.com/2018/06/scalable-deep-reinforcement-learning.html
6. https://openai.com/blog/learning-dexterity/
7. https://github.com/hanzhanggit/StackGAN
8. Uality, Q., Tability, S., Ariation, V., & Karras, T. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation, 1–26.
9. https://github.com/junyanz/CycleGAN
10. https://phillipi.github.io/pix2pix/
11. https://openai.com/blog/ai-and-compute/
12. Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural Ordinary Differential Equations, (NeurIPS).
13. https://www.oreilly.com/ideas/neuroevolution-a-different-kind-of-deep-learning
14. Irving, G., Christiano, P., Amodei, D. (2018). AI safety via debate.
15. https://arxiv.org/abs/1811.05233
16. https://arxiv.org/abs/1802.01561
17. https://cacm.acm.org/magazines/2018/7/229030-making-machine-learning-robust-against-adversarial-inputs/fulltext#R22
18. https://www.eff.org/ai/metrics
19. https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/
20. https://arxiv.org/abs/1511.04508
21. https://arxiv.org/abs/1705.07263
22. **https://arxiv.org/abs/1707.03501**
23. **https://arxiv.org/pdf/1707.07397.pdf**
24. **https://iclr.cc/Conferences/2018**
25. **https://arxiv.org/abs/1802.00420**
26. Kolter, J., Procaccia, A., CMU Lecture 15-780 – Graduate Artificial Intelligence: Adversarial attacks and provable defenses, Spring 2018.
27. **https://www.technologyreview.com/the-download/612292/googles-ai-is-better-at-spotting-advanced-breast-cancer-than-pathologists/**
28. **https://www.scrabbl.com/role-of-artificial-intelligence-and-machine-learning-in-bank-fraud-detection**
29. This picture show the facilities of the Google data center in Changhua, central Taiwan, on December 11, 2013. US search engine giant Google announced that it has decided to double its investment in Taiwan to $600 million while opening its first data centre in Asia cashing in on the robust demands. AFP PHOTO / Sam Yeh (Photo credit should read SAM YEH/AFP/Getty Images)
30. **https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf**
31. **https://www.microsoft.com/en-us/research/project/farmbeats-iot-agriculture/**
32. ICLR 2018 papers
33. **https://arxiv.org/pdf/1707.07397.pdf?**