State of AI Report

June 28, 2019

Nathan Benaich Ian Hogarth

About the authors

Nathan Benaich

Nathan is the founder of **Air Street Capital**, a VC partnership of industry specialists investing in intelligent systems. He founded the Research and Applied AI Summit and the RAAIS Foundation to advance progress in AI, and writes the AI newsletter nathan.ai. Nathan is also a Venture Partner at Point Nine Capital. He studied biology at Williams College and earned a PhD from Cambridge in cancer research.

Ian Hogarth

Ian is an **angel investor** in 50+ startups with a focus on applied machine learning. He is a Visiting Professor at UCL working with Professor Mariana Mazzucato. Ian was co-founder and CEO of Songkick, the global concert service used by 17m music fans each month. He studied engineering at Cambridge. His Masters project was a computer vision system to classify breast cancer biopsy images.

Artificial intelligence (AI) is a multidisciplinary field of science whose goal is to create intelligent machines.

We believe that AI will be a force multiplier on technological progress in our increasingly digital, data-driven world.

This is because everything around us today, ranging from culture to consumer products, is a product of intelligence.

In this report, we set out to capture a snapshot of the exponential progress in AI with a focus on developments in the past 12 months. Consider this report as a compilation of the most interesting things we've seen with a goal of triggering an informed conversation about the state of AI and its implication for the future. This edition builds on the inaugural State of AI Report 2018, which can be found here: www.stateof.ai/2018

We consider the following key dimensions in our report:

- **Research**: Technology breakthroughs and their capabilities.
- **Talent**: Supply, demand and concentration of talent working in the field.
- **Industry**: Large platforms, financings and areas of application for AI-driven innovation today and tomorrow.
- **China:** With two distinct internets, we review AI in China as its own category.
- **Politics**: Public opinion of AI, economic implications and the emerging geopolitics of AI.

Collaboratively produced in East London, UK by **Ian Hogarth** (@soundboy) and **Nathan Benaich** (@nathanbenaich).

Thank you's

Thanks to the following people for suggesting interesting content and/or reviewing this year's Report.

Jack Clark, Kai Fu Lee, Jade Leung, Dave Palmer, Gabriel Dulac-Arnold, Roland Memisevic, François Chollet, Kenn Cukier, Sebastian Riedel, Blake Richards, Moritz Mueller-Freitag, Torsten Reil, Jan Erik Solem and Alex Loizou.

Definitions

Artificial intelligence (AI): A broad discipline with the goal of creating intelligent machines, as opposed to the natural intelligence that is demonstrated by humans and animals. It has become a somewhat catch all term that nonetheless captures the long term ambition of the field to build machines that emulate and then exceed the full range of human cognition.

Machine learning (ML): A subset of AI that often uses statistical techniques to give machines the ability to "learn" from data without being explicitly given the instructions for how to do so. This process is known as “training” a “model” using a learning “algorithm” that progressively improves model performance on a specific task.

Reinforcement learning (RL): An area of ML that has received lots of attention from researchers over the past decade. It is concerned with software agents that learn goal-oriented behavior by trial and error in an environment that provides rewards or penalties in response to the agent’s actions (called a “policy”) towards achieving that goal.

Deep learning (DL): An area of ML that attempts to mimic the activity in layers of neurons in the brain to learn how to recognise complex patterns in data. The “deep” in deep learning refers to the large number of layers of neurons in contemporary ML models that help to learn rich representations of data to achieve better performance gains.

Definitions

Algorithm: An unambiguous specification of how to solve a particular problem.

Model: Once a ML algorithm has been trained on data, the output of the process is known as the model. This can then be used to make predictions.

Supervised learning: This is the most common kind of (commercial) ML algorithm today where the system is presented with labelled examples to explicitly learn from.

Unsupervised learning: In contrast to supervised learning, the ML algorithm has to infer the inherent structure of the data that is not annotated with labels.

Transfer learning: This is an area of research in ML that focuses on storing knowledge gained in one problem and applying it to a different or related problem, thereby reducing the need for additional training data and compute.

Natural language processing (NLP): Enables machines to analyse, understand and manipulate textual data.

Computer vision: Enabling machines to analyse, understand and manipulate images and video.

Scorecard: Reviewing our predictions from 2018

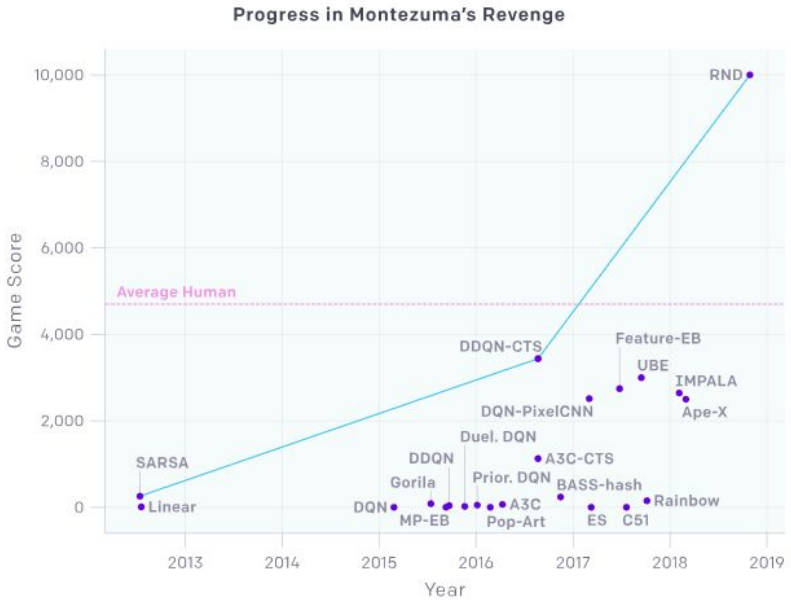
Our 2018 prediction	Outcome?	What's the evidence?
Breakthrough by a Chinese AI lab		Chinese labs win ActivityNet (CVPR 2018); train ImageNet model in 4 mins.
DeepMind RL Starcraft II breakthrough		AlphaStar beats one of the world's strongest StarCraft II players 5-0.
A major research lab "goes dark"		MIRI "non-disclosed by default" and OpenAI GPT-2.
The era of deep learning continues		Yes, but not entirely clear how to evaluate this.
Drug discovered by ML produces positive clinical trial results		
M&A worth >$5B of EU AI cos by China/US		
OECD country government blocks M&A of an ML co by USA/China		
Access to Taiwanese/South Korean semiconductor companies is an explicit part of the US-China trade war		

Section 1: Research and technical breakthroughs

Reinforcement learning (RL) conquers new territory: Montezuma's Revenge

Rewarding 'curiosity' enables OpenAI to achieve superhuman performance at Montezuma's Revenge.

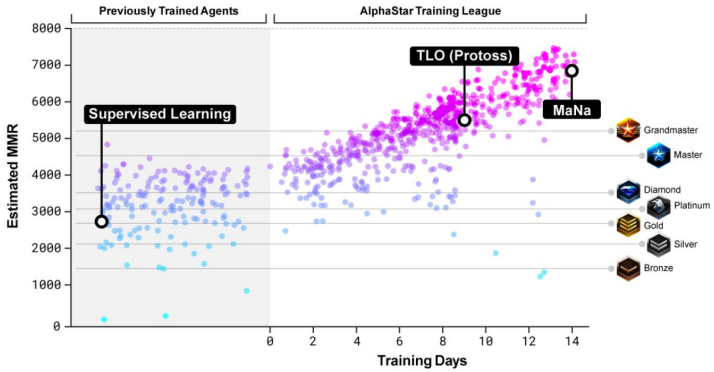
In 2015, DeepMind's DQN system successfully achieved superhuman performance on a large number of Atari 2600 games. A major hold out was Montezuma's Revenge. In October 2018, OpenAI achieved superhuman performance at Montezuma's with a technique called *Random Network Distillation (RND)*, which incentivised the RL agent to explore unpredictable states. This simple but powerful modification can be particularly effective in environments where broader exploration is valuable. The graph on the right shows total game score achieved by different AI systems on Montezuma's Revenge.



RL conquers new territory: StarCraft II

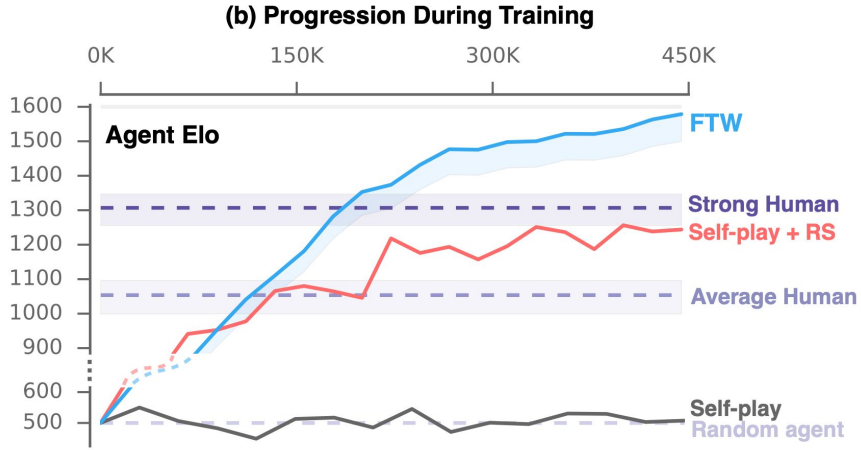
StarCraft integrates various hard challenges for ML systems: Operating with imperfect information, controlling a large action space in real time and making strategic decisions over a long time horizon.

- DeepMind beat a world class player 5-0. StarCraft II still cannot be considered to be 'solved' due to various constraints on the action space. This is nonetheless a major breakthrough.
- AlphaStar was first trained by supervised learning on a set of human games. After this, AlphaStar's novel approach used a multi-agent training algorithm that effectively created a league o agents competing against each other and collectively exploring the huge strategic space. The final AlphaStar agent is produced by Nash averaging, which combines the most effective mix of strategies developed by individual agents.
- The market cost of the compute resource to train AlphaStar has been estimated at $26M. The cost of the world class team at DeepMind's working on AlphaStar could be similar.



RL conquers new territory: Quake III Arena Capture the Flag

Human-level performance is achieved by having multiple agents independently learn and act to cooperate and compete with one another.

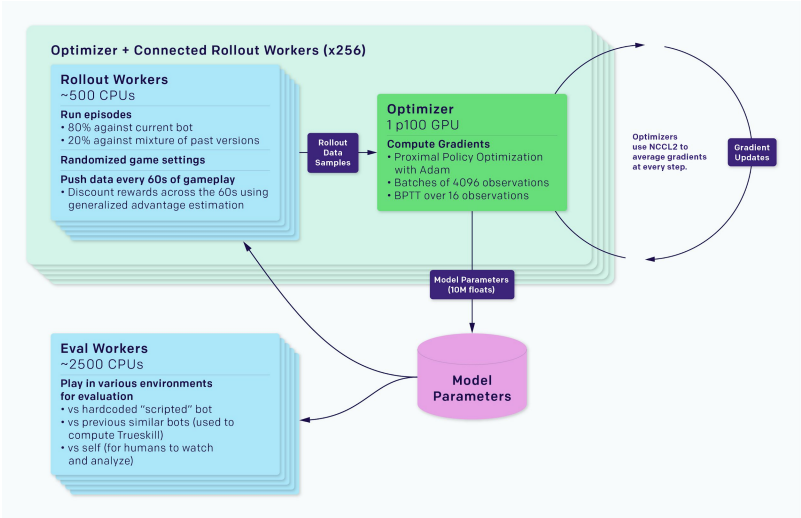


- To play Capture the Flag, a population of independent RL agents are trained concurrently from thousands of parallel matches with agents playing in teams together and against each other on randomly generated environments. Each agent in the population learns its own internal reward signal to complement the sparse delayed reward from winning, and selects actions using a temporally hierarchical representation that enables the agent to reason at multiple timescales. The agents use only pixels and game points as input.
- While its difficult to maintain diversity in agent populations, they end up displaying humanlike behaviours such as navigating, following, and defending based on a rich learned representation that is shown to encode high-level game knowledge.

RL conquers new territory: OpenAI Five improves even further

OpenAI's Dota2 playing bot now has a 99.4% win rate over >7,000 online games with >15,000 live players.

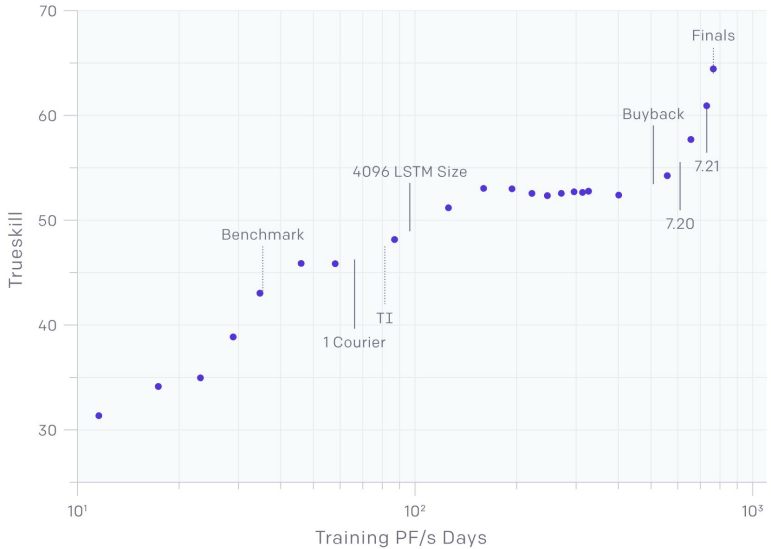
- **August 2017**: A single player bot beats a top global Dota2 player in a simplified 1v1 match.
- **August 2018**: A team of bots, OpenAI Five, lost 2 games in a restricted 5v5 best of 3 match in *The Internationals*.
- **April 2019**: OpenAI Five wins 2 back-to-back games vs. the world champion Dota2 team in a live streamed event. Over the 4 day online tournament (*Arena*), 15,019 total players challenged OpenAI Five to 7,257 Competitive games of which the bot team won 99.4%.
- **System design**: Each bot is a single-layer, 4,096-unit LSTM that reads the game state and is trained through self-play RL (80% against itself and 20% against older versions of itself). Bots report their experience in batches and gradient optimisation is run and averaged globally.



RL conquers new territory: OpenAI Five improves even further

Compute was the gatekeeper to the competitive performance of OpenAI Five.

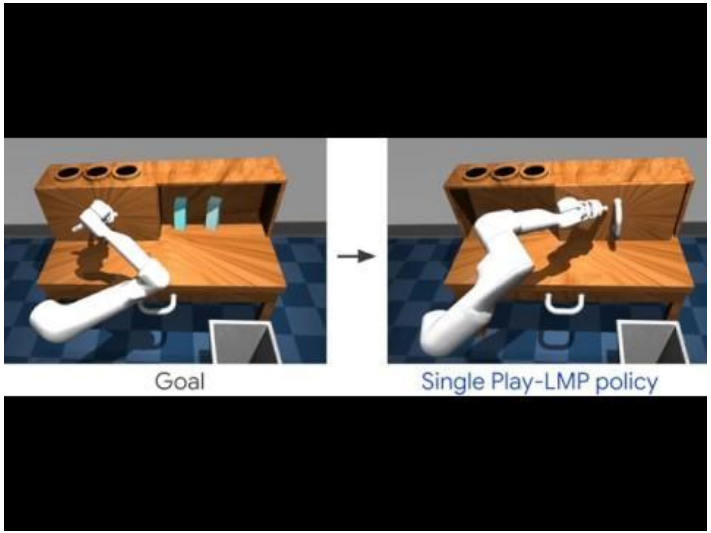
- Compared to the August 2018 version of OpenAI Five, April's version is trained with **8x more compute**.
- The current version has consumed **800 petaflop/s-days** and experienced about **45,000 years of Dota self-play** over **10 realtime months**.
- As of *The International* in 2018 where the bots lost 2 games in a best of 3 math, total training experience summed to 10,000 years over 1.5 realtime months. This equates to **250 years of simulated experience per day** on average.



What's next in RL: Play-driven learning for robots

Training a single robot using play to perform many complex tasks without having to relearn each from scratch.

- As children, we acquire complex skills and behaviors by learning and practicing diverse strategies and behaviors in a low-risk fashion, i.e. play time. Researchers used the concept of supervised play to endow robots with control skills that are more robust to perturbations compared to training using expert skill-supervised demonstrations.
- Here, a human remotely teleoperates the robot in a playground environment, interacting with all the objects available in as many ways that they can think of. A human operator provides the necessary properties of curiosity, boredom, and affordance priors to guide rich object play.

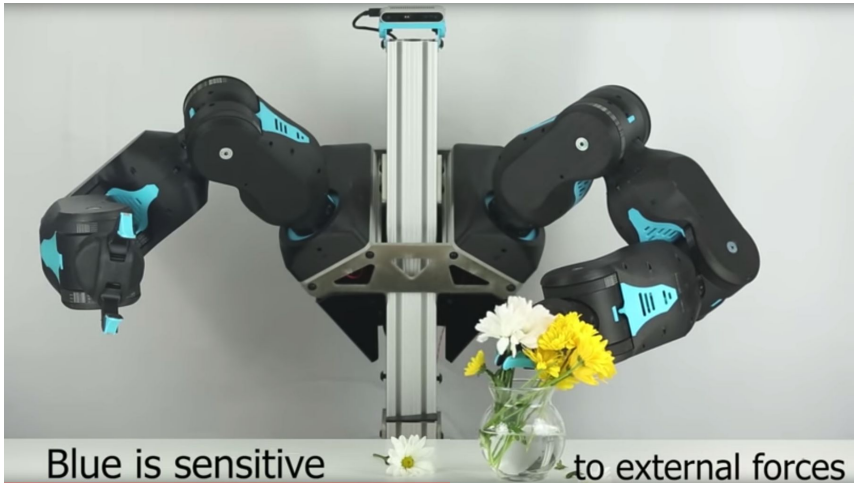


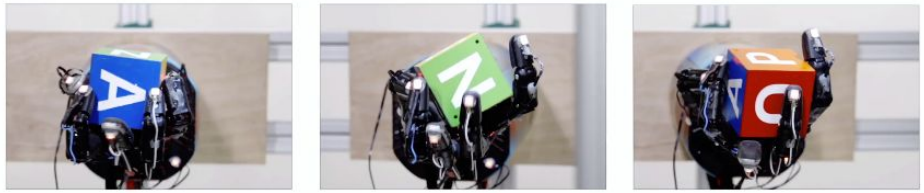
- Despite not being trained on task-specific data, this system is capable of generalizing to 18 complex user-specified manipulation tasks with average success of 85.5%, outperforming individual models trained on expert demonstrations (success of 70.3%).

What's next in RL: Learning dexterity using simulation and the world real

New robotic learning platforms and sim-to-real enable impressive progress in manual dexterity.

UC Berkeley's Robot Learning Lab created BLUE, a human-scale, 7 degree-of-freedom arm with 7kg payload for learning robotic control tasks. OpenAI used simulation to train a robotic hand to shuffle physical objects with impressive dexterity. The system used computer vision to predict the object pose given three camera images and then used RL to learn the next action based on fingertip positions and the object's pose.

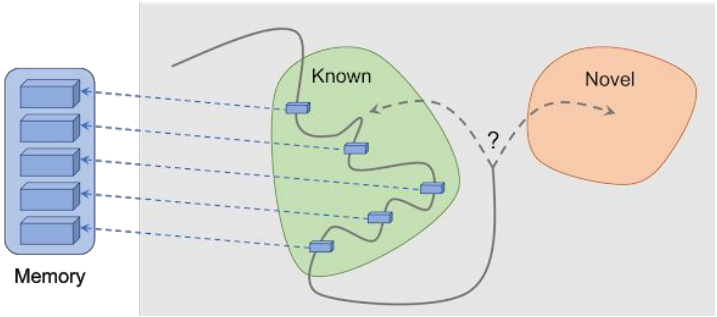


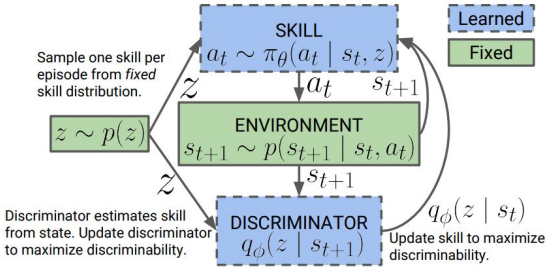


What's next in RL: Curiosity-driven exploration

How can agents learn to solve tasks when their reward is either sparse or non-existent? Encourage curiosity.

In RL, agents learn tasks by trial and error. They must balance exploration (trying new behaviors) with exploitation (repeating behaviors that work). In the real world, rewards are difficult to explicitly encode. A promising solution is to a) store an RL agent's observations of its environment in memory and b) reward it for reaching observations that are "not in memory". By seeking out novel experiences, the agent is more likely to find behaviors that allow it to solve 3D maze navigation tasks.





Going further, one can design an RL system that learns skills that not only are distinguishable, but also are as diverse as possible. By learning distinguishable skills that are as random as possible, we can "push" the skills away from each other, making each skill robust to perturbations and effectively exploring the environment.

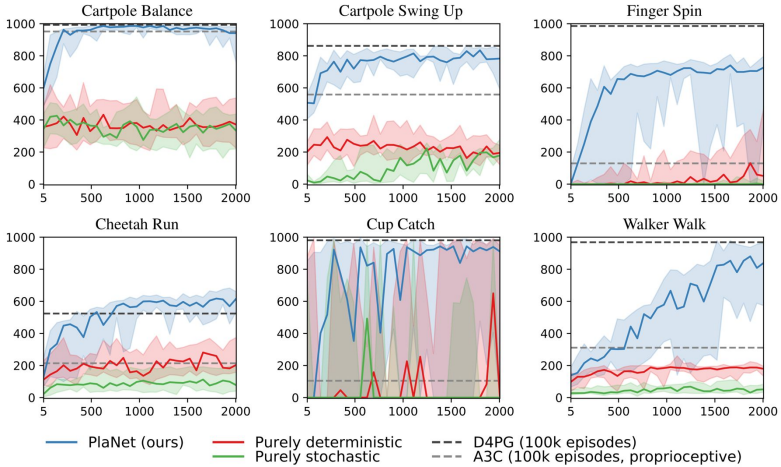
Google

What's next in RL: Learning dynamics models for online planning

An RL agent (PlaNet) learns a model of the environment's dynamics from images, selects actions through fast online planning by accurately predicting the rewards ahead for multiple time steps.

- This results in 50x less simulated environmental interactions and similar computation time compared to the state-of-the-art A3C and D4PG algorithms. Within 500 episodes, PlaNet outperforms A3C trained from 100,000 episodes, on six simulator control tasks. This is a significant improvement in data efficiency.
- After 2,000 episodes, PlaNet achieves similar performance to D4PG, which is trained from images for 100,000 episodes.

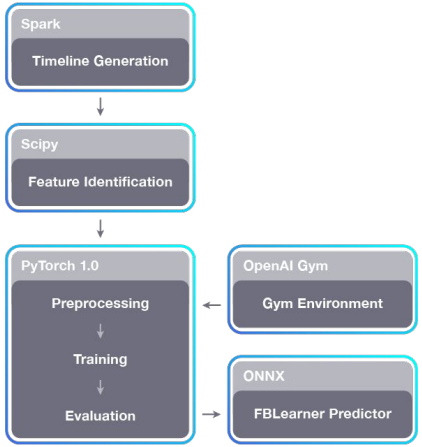
Method	Modality	Episodes	Cartpole Balance	Cartpole Swingup	Finger Spin	Cheetah Run	Ball in cup Catch	Walker Walk
A3C	proprioceptive	100,000	952	558	129	214	105	311
D4PG	pixels	100,000	993	862	985	524	980	968
PlaNet (ours)	pixels	2,000	986	831	744	650	914	890
CEM + true simulator	simulator state	0	998	850	825	656	993	994
Data efficiency gain PlaNet over D4PG (factor)			100	180	16	50+	20	11



What's next in RL: Moving research into production environments

Facebook release Horizon, the first open source end-to-end platform that uses applied RL to optimize systems in large-scale production environments, such as Messenger suggestions, video stream quality and notifications.

- Horizon is built on PyTorch 1.0, Caffe2 and Spark, popular tools for ML work.
- In particular, the system includes workflows for simulated environments as well as a distributed platform for preprocessing, training, and exporting models into production.
- It focuses on ML-based systems that optimise a set of actions given the state of an agent and its environment ("policy optimisation"). The optimisation relies on data that's inherently noisy, sparse, and arbitrarily distributed.
- Instead of online training as in games, Horizon models are trained offline using a policy that a product engineer has designed. Counterfactual policy evaluation (CPE) is used to estimate what the RL model would have done if it were making those past decisions. Once the CPE results are admissible, the RL model is deployed in a small experiment to collect live results.

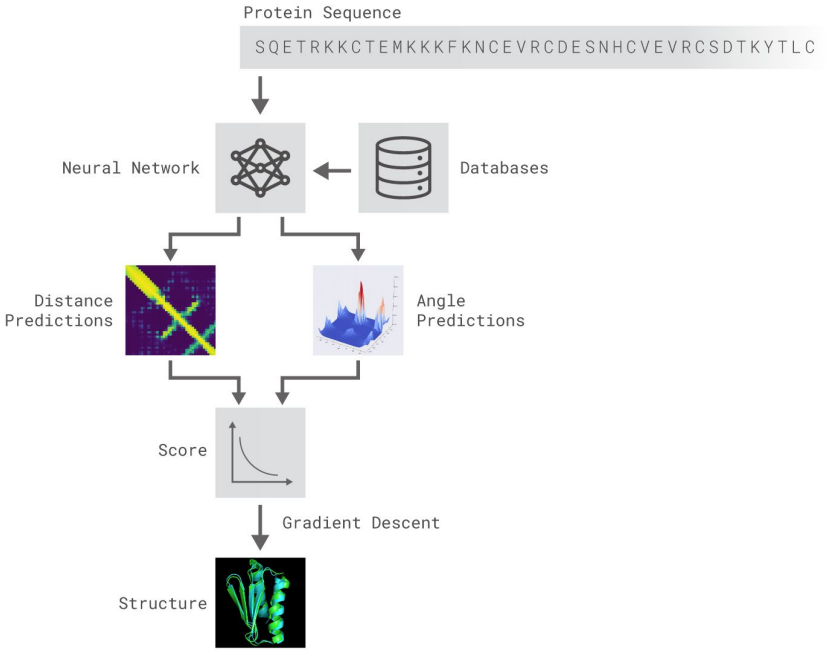


ML for life science: AlphaFold predicts *de novo* 3D structure of folded proteins

Two deep CNNs work in concert to significantly outperform prior state-of-the-art, far earlier than expected.

1. A first neural network predicts the distances between pairs of amino acids (AAs).
2. A second network predicts the angles between chemical bonds that connect those AAs to make up proteins.
3. By predicting how close pairs of AAs are to one another, the system creates a distance map of the protein.
4. This map can essentially be extrapolated to generate a 3D protein structure or match one from an existing database of structures.

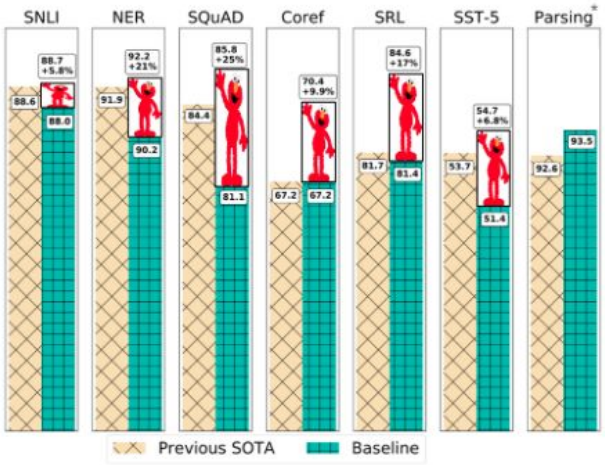
Note that proteins often bind to other proteins to form complexes in order to exert their biological function *in vivo*. AlphaFold does not currently apply to these complexes.



A big year in natural language processing: Pretrained language models

Transfer learning from pretrained language models usher in 'ImageNet moment' for NLP.

- Various research breakthroughs (Google AI's BERT, Transformer; Allen Institute's ELMo; OpenAI's Transformer, Ruder & Howard's ULMFiT, Microsoft's MT-DNN) demonstrated that pretrained language models can substantially improve performance on a variety of NLP tasks.
- Pretraining models to learn high- and low-level features has been transformative in computer vision. In the last year there have been similar empirical breakthroughs in pretraining language models on large text corpora to learn high- and low-level language features.
- Unlike ImageNet, these language models are typically trained on very large amounts of publicly available, i.e. *unlabeled* text from the web. This method could be further scaled up to generate gains in NLP tasks and unlock many new commercial applications in the same way that transfer learning from ImageNet has driven more industrial uses of computer vision.



*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)

The improvements ELMo achieved on a wide range of NLP tasks. (Source: Matthew Peters)

A big year in natural language processing: Pretrained language models

Example input-output response from OpenAI's GPT-2 language model.

Human-written prompt:

Miley Cyrus was caught shoplifting from Abercrombie and Fitch on Hollywood Blvd.

Model completion (entirely machine-written on its second attempt):

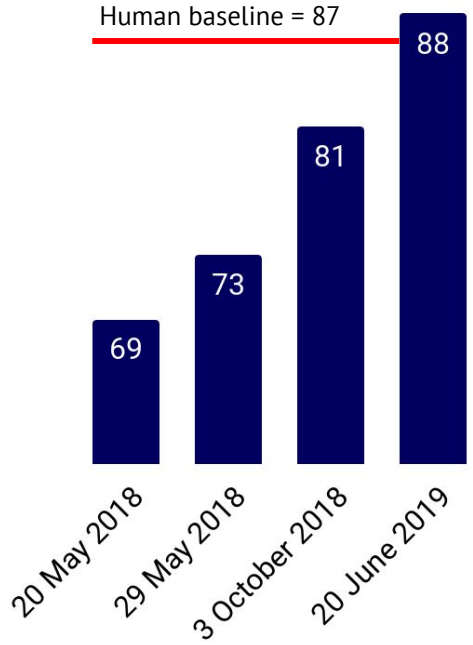
The 19-year-old singer was caught on camera being escorted out of the store by security guards.
The singer was wearing a black hoodie with the label 'Blurred Lines' on the front and 'Fashion Police' on the back.
Scroll down for video
Shoplifting: Miley Cyrus was caught shoplifting from Abercrombie and Fitch on Hollywood Boulevard today (pictured)
The singer was wearing a black hoodie with the label 'Blurred Lines' on the front and 'Fashion Police' on the back
The singer was also wearing a pair of black-rimmed glasses, a black jacket, black jeans and black sandals.
She was carrying a pair of black and white striped gloves and a small black bag.

Try it out here: **transformer.huggingface.co**

A big year in natural language processing: Researchers start sniffing GLUE

New performance benchmark GLUE tests performance at a variety of language understanding tasks.

- Human understanding of language is general and flexible. The GLUE benchmark provides a single benchmark for evaluating NLP systems at a range of tasks spanning logic, common sense understanding, and lexical semantics. The right hand charts progress on the leaderboard.
- The benchmark is designed to favor systems that share general linguistic knowledge across tasks.
- As a demonstration of how quickly progress is being made in NLP, the state-of-the art has increased from a score of 69 to 88 over 13 months. The human baseline level is 87.
- Progress was so much faster than anticipated that a new benchmark SUPERglue has already been introduced.



SuperGLUE

A big year in natural language processing: Machine translation without bitexts

Facebook show how to leverage monolingual data in order to make machine translation more widely applicable.

This work applies several principles to the development of a simple, easy to interpret phrase-based statistical machine translation (PBSMT) system and a neural machine translation (NMT) system that learns to translate *without bidirectional text*. These design principles are:

- Carefully initialize the model with an inferred bilingual dictionary;
- Leverage strong language models by training a sequence-to-sequence model as a denoising autoencoder (used for feature selection and extraction) where the representation built by the encoder is constrained to be shared across the two languages being translated;
- Use backtranslation to turn the unsupervised problem into a supervised one. This requires two models: the first translates the source language into the target and the second translates the target back into the source language. The output data from the first model is the training data for the second, and vice versa.

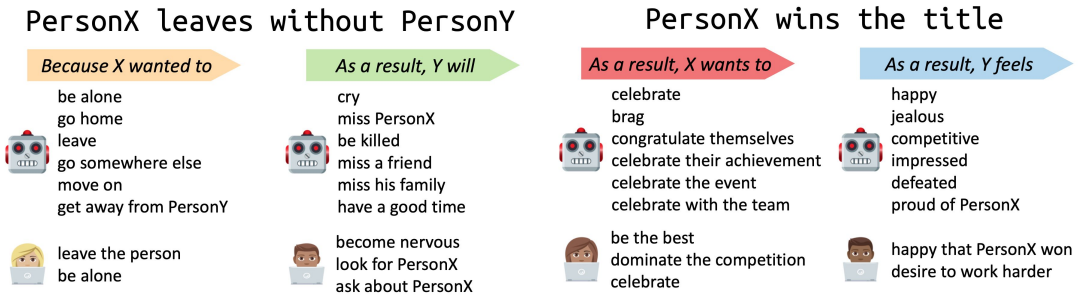
Model	en-fr	fr-en	de-en	en-de
(Artetxe et al., 2018)	15.1	15.6	-	-
(Lample et al., 2018)	15.0	14.3	13.3	9.6
NMT (LSTM)	24.5	23.7	19.6	14.7
NMT (Transformer)	25.1	24.2	21.0	17.2
PBSMT (Iter. 0)	16.1	15.4	14.5	10.3
PBSMT (Iter. n)	**27.1**	24.7	21.3	16.7
NMT + PBSMT	26.3	25.1	20.2	16.4
PBSMT + NMT	26.7	**27.1**	**23.6**	**19.2**

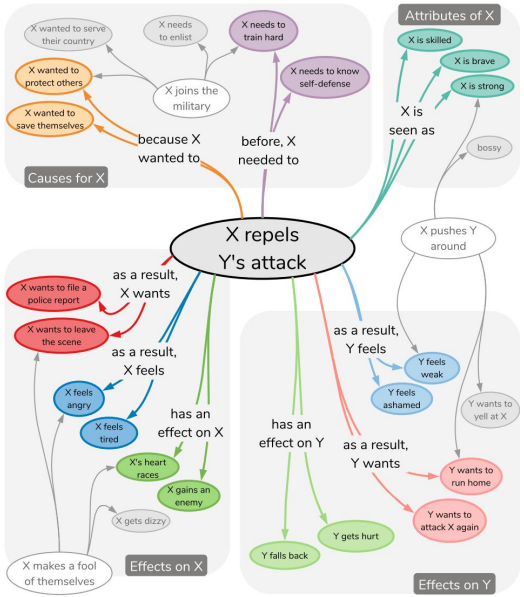
Table 2: **Comparison with previous approaches.** BLEU score for different models on the $en - fr$ and $en - de$ language pairs. Just using the unsupervised phrase table, and without back-translation (PBSMT (Iter. 0)), the PBSMT outperforms previous approaches. Combining PBSMT with NMT gives the best results.

Endowing common sense reasoning to natural language models: Is text really enough?

A dataset of >300k everyday common sense events associated with 877k inferential relations to help machines learn *if-then* relation types.

By generatively training on the inferential knowledge of the dataset, the authors show that neural models can acquire simple common sense capabilities and reason about previously unseen events. This approach extends work such as the Cyc knowledge base project that began in the 80s and is called world's longest AI project. Common sense reasoning is, however, unlikely to be solved from text as the only modality.

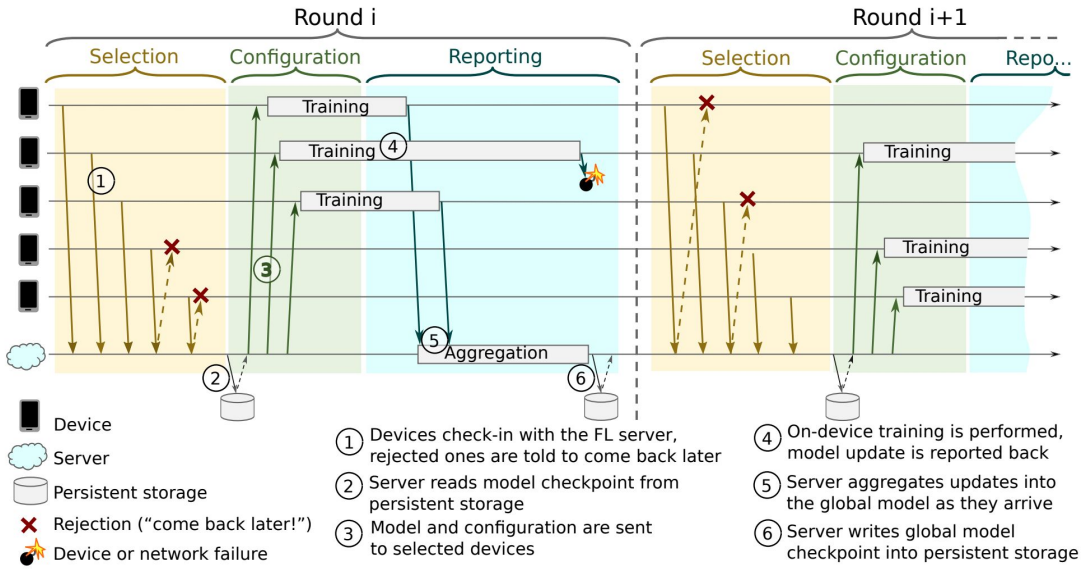




A growing interest in federated learning (FL) for real-world products

Last year, we noted that Google uses FL for distributed training of Android keyboards. This year, Google released their overall FL system design and introduced TensorFlow Federated to encourage developer adoption.

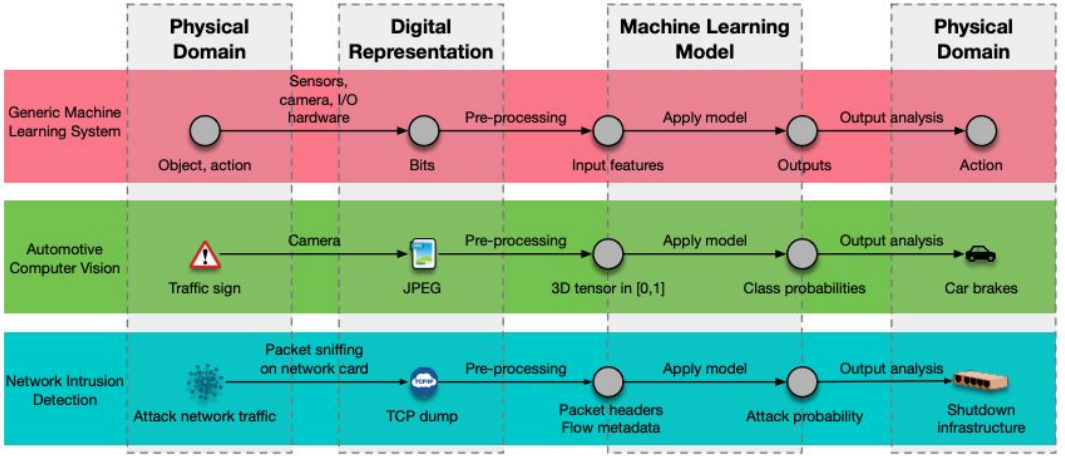
Developers can express a new data type, specifying its underlying data and where that data lives (e.g. on distributed clients) and then specify a federated computation they want to run on the data. The TensorFlow Federated library represents the federated functions in a form that could be run in a decentralized setting.
FL is creating lots of excitement for healthcare use cases where a global overview of sensitive data could improve ML systems for all parties.



Increasing emphasis on data privacy and protecting deployed ML systems from attacks

The attack surface of ML systems is large: Adversaries can manipulate data collection, corrupt the model or tamper with its outputs.

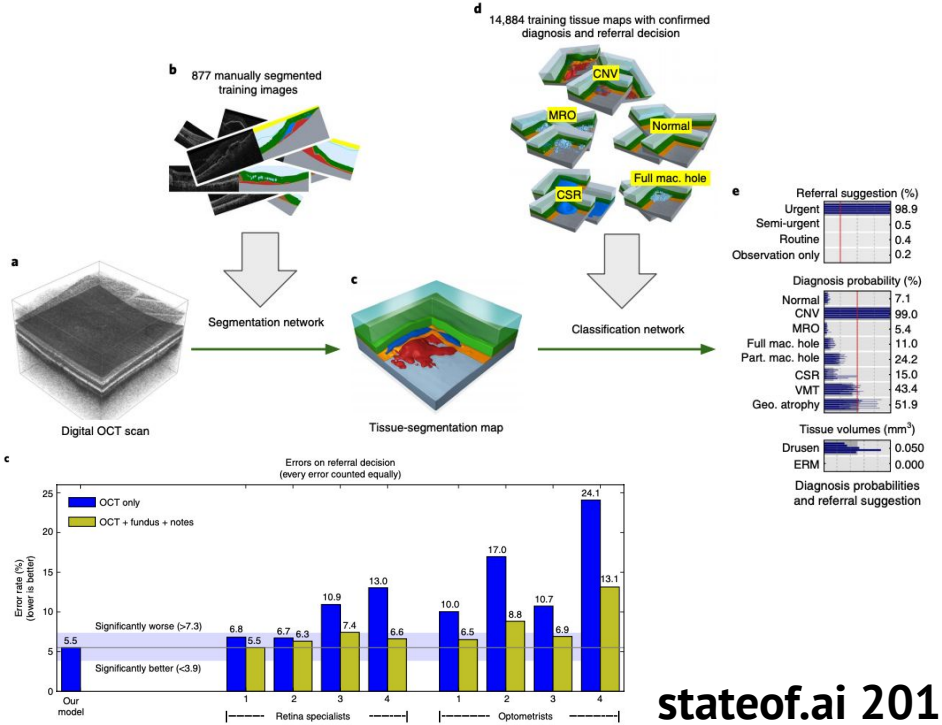
- TensorFlow Privacy from Google allows for training ML models on users' data while giving strong mathematical guarantees that they do not learn or remember details about any specific user. The library is also designed to work with training in a federated context.
- TF Encrypted from Dropout Labs is a library built on top of TensorFlow to integrate privacy-preserving technology into pre-existing ML processes.



Deep learning in medicine: Diagnosing eye disease

Expert-level diagnosis and treatment referral suggestions is achieved using a two-stage deep learning approach.

- At first, a segmentation network uses a 3D U-Net architecture to create a "tissue map" of the eye from a 3D digital optical computed tomography scan. This map paints the eye's structure according to expert ophthalmologists.
- A second classification network operates on this tissue map to predict the severity of the condition.
- This system achieves expert performance on referral decisions. It can also be easily adapted to various imaging machines by only retraining the segmentation network.

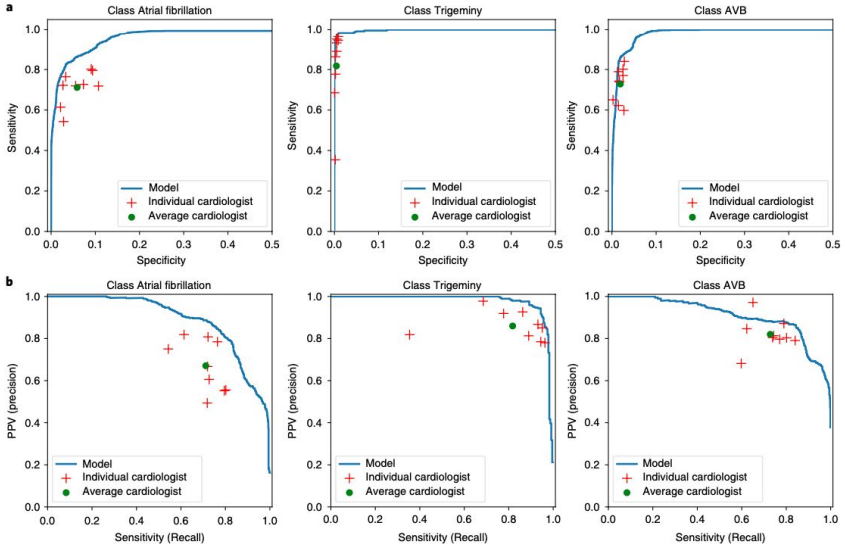


Moorfields Eye Hospital NHS Foundation Trust NHS DeepMind

Deep learning in medicine: Detecting and classifying cardiac arrhythmia using ECGs

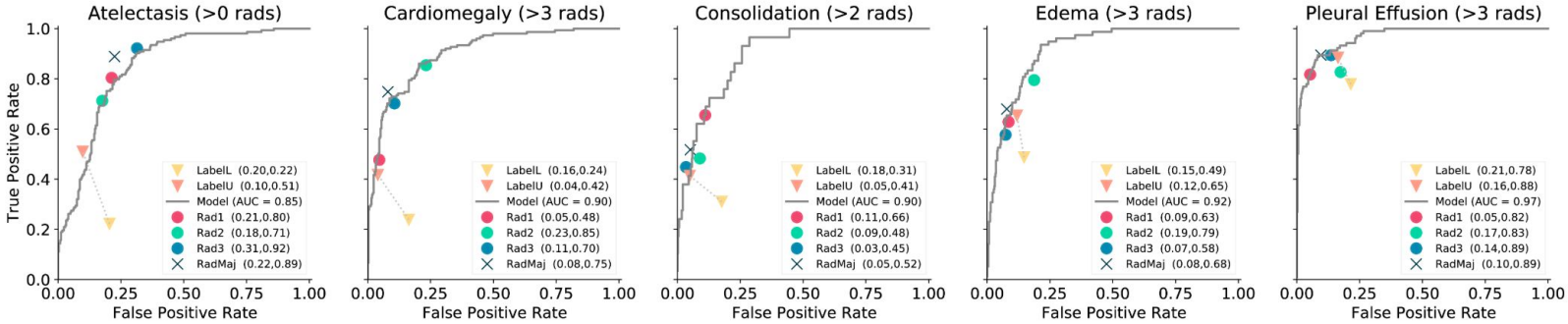
Cardiologist-level performance is demonstrated using end-to-end deep learning trained on 54k patients.

- This study showed that single lead electrocardiogram traces in the ambulatory setting can be processed in a raw format by a deep learning model to detect 12 rhythm classes.
- The curves on the right depict how individual (red crosses) and the average (green dot) of all cardiologists fair in comparison to the model. The model achieved an average ROC of 0.97 and with a specificity fixed at the average specificity of cardiologists, the model was more sensitive for all rhythm classes.
- It remains to be seen if this approach works on multi-lead ECGs, which are more common in the clinic.



Deep learning in medicine: The bigger the dataset, the better the model?

>600k chest x-rays have been published to boost model performance, but dataset issues remain.

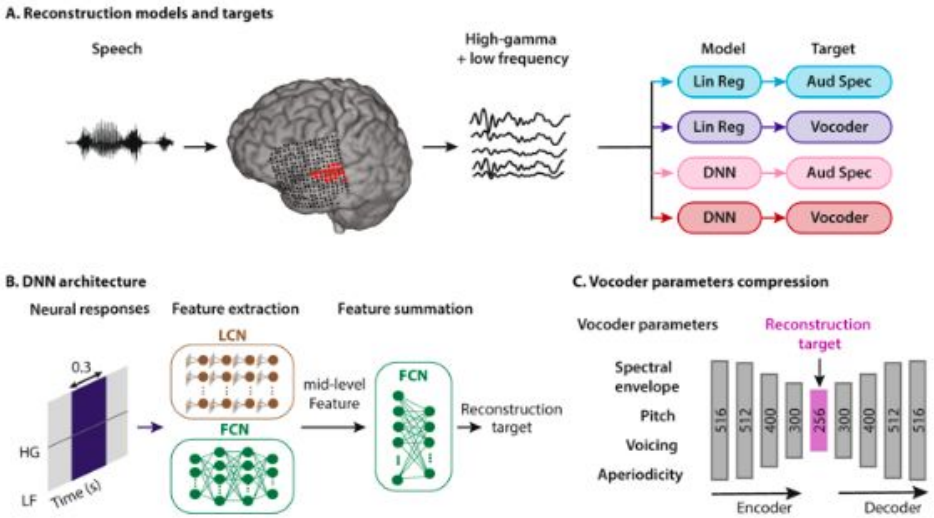


- Deep learning models for imaging diagnostics fit datasets well, but they have difficulties generalising to new data distributions. Despite improved documentation to this new dataset, label definitions are shallow.
- There are challenges with extracting labels using NLP from doctors notes: Its error-prone and suffers from the lack of information contained in radiology reports, with 5-15% error rates in most label categories.
- Significant number of repeat scans, with 70% of the scans coming from 30% of the patients. This reduces the effective size of the dataset and its diversity, which will impact the generalisability of trained models.

Deep learning in medicine: Neural networks decode your thoughts from brain waves

Researchers reconstruct speech from neural activity in the auditory cortex.

- Researchers at Columbia used invasive electrocorticography to measure neural activity in 5 patients undergoing treatment for epilepsy while listening to continuous speech sounds.
- Inverting this enabled the researchers to synthesize speech through a vocoder from brain activity. The system achieved 75% accuracy when tested on single digits 'spoken' via a vocoder. The deep learning method improved the intelligibility of speech by 65% over the baseline linear regression method.
- The research indicates the potential for brain computer interfaces to restore communication for paralysed patients.

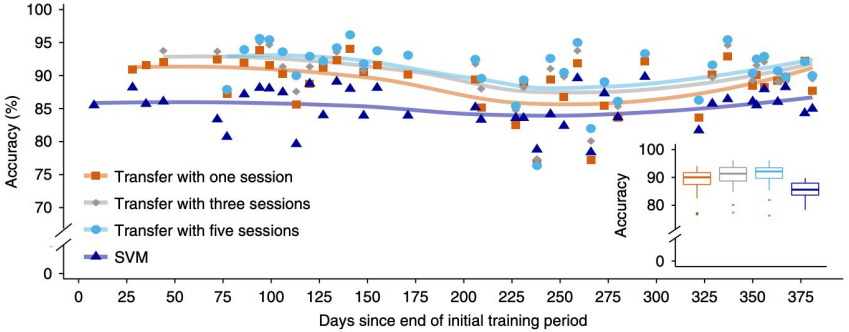


Deep learning in medicine: Neural networks can restore limb control for the disabled

Long-term reanimation of a tetraplegic patient's forearm with electrical stimulation and neural network decoder.

- Researchers implanted a microelectrode in the hand and arm area of a tetraplegic patient's left primary motor cortex. They trained a neural network to predict the likely intended movements of the person's arm based on the raw intracranial voltage signals recorded from the patient's brain.
- The patient could sustain high accuracy reanimation of his paralyzed forearm with functional electrical stimulation for **over a year** without the need of supervised updating (thus reducing daily setup time).

Model	Failure rate (missed cues)				
	Index extension	Index flexion	Wrist extension	Wrist flexion	All
SVM	25.66% (39)	31.61% (49)	14.97% (22)	15.75% (23)	22.17% (133)
uNN	6.58% (10)	6.45% (10)	6.12% (9)	3.42% (5)	5.67% (34)



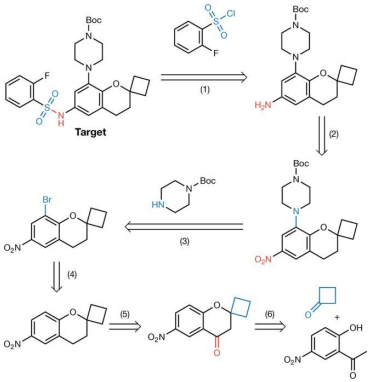
- The neural network approach was much more robust to failure than an SVM baseline. It could also be updated to learn new actions with transfer learning.

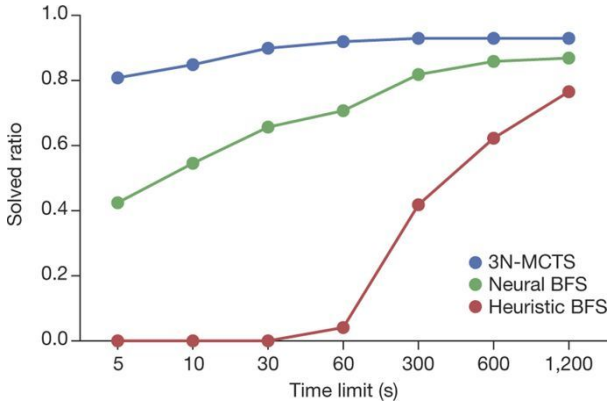
Machines learn how to synthesise chemical molecules

Using neural networks with Monte Carlo tree search to solve retrosynthesis by training on 12.4 million reactions.

A system built from three NNs (3N-MCTS):

- 1) Guide the search in promising directions by proposing a restricted number of automatically extracted transformations.
- 2) Predict whether the proposed reactions are actually feasible.
- 3) Estimate the position value and iterate.





This method is far faster that the state-of-the-art computer-assisted synthesis planning. In fact, 3N-MCTS solves more than 80% of a molecule test set with a time limit of 5 seconds per target molecule. By contrast, an approach called best first search in which functions are learned through a neural network can solve 40% of the test set. Best first search designed with hand-coded heuristic functions performs the worst: it solves 0% in 5s.

AutoML: Evolutionary algorithms for neural network architecture and hyperparameters

Jointly optimising for hyperparameters, maximising network performance while minimising complexity and size.

- Prior AutoML work optimize hyperparameters or network architecture individually using RL. Unfortunately, RL systems require a user to define an appropriate search space beforehand for the algorithm to use as a starting point. The number of hyperparameters that can be optimized for each layer is also limited.
- Furthermore, the computations are extremely heavy. To generate the final best network, many thousands of candidate architectures have to be evaluated and trained, which requires >100k GPU hours.
- An alternative (Learning Evolutionary AI Framework: LEAF) is to use evolutionary algorithms to conduct both hyperparameter and network architecture optimisation, ultimately yielding smaller and more effective networks.
- For example, LEAF matches the performance of a hand-crafted dataset-specific network (CheXNet) for Chest X-Ray diagnostic classification and outperforms Google's AutoML.

Algorithm	Test AUROC (%)
1. Wang et al. (2017) [48]	73.8
2. CheXNet (2017) [39]	84.4
3. Google AutoML (2018) [1]	79.7
4. LEAF	84.3

AutoML: Designing resource-constrained networks with real device performance feedback

The pace of CNN-based automated architecture search is accelerating: Facebook ups ante vs. Google.

- Google demonstrated a multi-objective RL-based approach (MnasNet) to find high accuracy CNN models with low real-world inference latency as measured on the Google Pixel platform. The system reaches 74.0% top-1 accuracy with 76ms latency on a Pixel phone, which is 1.5x faster than MobileNetV2.
- Facebook proposed a differentiable neural architecture search (DNAS) framework that uses gradient-based methods to optimize CNN architectures over a layer-wise search space. FBNet-B achieves the same top-1 accuracy than MnasNet but with 23.1 ms latency and 420x smaller search cost.

Model	Type	#Parameters	#Mult-Adds	Top-1 Acc. (%)	Top-5 Acc. (%)	CPU Latency
MobileNetV1 (Howard et al. 2017)	manual	4.2M	575M	70.6	89.5	113ms
SqueezeNext (Gholami et al. 2018)	manual	3.2M	708M	67.5	88.2	-
ShuffleNet (1.5) (Zhang et al. 2018)	manual	3.4M	292M	71.5	-	-
ShuffleNet (x2)	manual	5.4M	524M	73.7	-	-
CondenseNet (G=C=4) (Huang et al. 2018)	manual	2.9M	274M	71.0	90.0	-
CondenseNet (G=C=8)	manual	4.8M	529M	73.8	91.7	-
MobileNetV2 (Sandler et al. 2018)	manual	3.4M	300M	72.0	91.0	75ms
MobileNetV2 (1.4)	manual	6.9M	585M	74.7	92.5	143ms
NASNet-A (Zoph et al. 2018)	auto	5.3M	564M	74.0	91.3	183ms
AmoebaNet-A (Real et al. 2018)	auto	5.1M	555M	74.5	92.0	190ms
PNASNet (Liu et al. 2018a)	auto	5.1M	588M	74.2	91.9	-
DARTS (Liu, Simonyan, and Yang 2018)	auto	4.9M	595M	73.1	91	-
MnasNet	auto	4.2M	317M	74.0	91.78	76ms

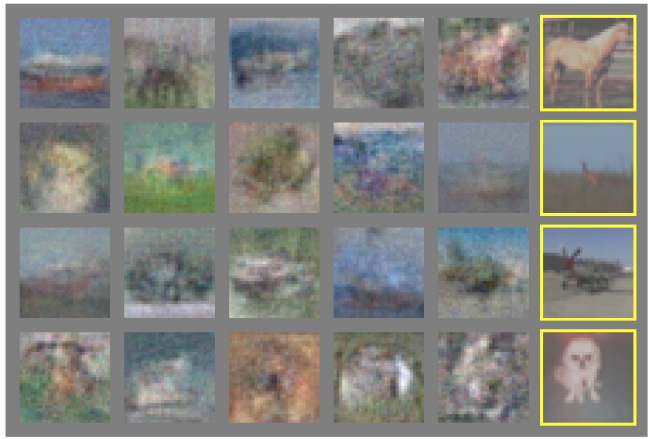
Model	Search method	Search space	Search cost (GPU hours / relative)	#Params	#FLOPs	CPU Latency	Top-1 acc (%)
1.0-MobileNetV2 [17]	manual	-	-	3.4M	300M	21.7 ms	72.0
1.5-ShuffleNetV2 [13]	manual	-	-	3.5M	299M	22.0 ms	72.6
CondenseNet (G=C=8) [7]	manual	-	-	2.9M	274M	28.4 ‡ ms	71.0
MnasNet-65 [13]	RL	stage-wise	91K* / 421x	3.6M	270M	-	73.0
DARTS [12]	gradient	cell	288 / 1.33x	4.9M	595M	-	**73.1**
FBNet-A (ours)	gradient	layer-wise	216 / 1.0x	4.3M	**249M**	**19.8 ms**	73.0
1.3-MobileNetV2 [17]	manual	-	-	5.3M	509M	33.8 ms	**74.4**
CondenseNet (G=C=4) [7]	manual	-	-	4.8M	529M	28.7‡ ms	73.8
MnasNet [20]	RL	stage-wise	91K* / 421x	4.2M	317M	23.7 ms	74.0
NASNet-A [31]	RL	cell	48K / 222x	5.3M	564M	-	74.0
PNASNet [11]	SMBO	cell	6K† / 27.8x	5.1M	588M	-	74.2
FBNet-B (ours)	gradient	layer-wise	216 / 1.0x	4.5M	**295M**	**23.1 ms**	74.1
1.4-MobileNetV2 [17]	manual	-	-	6.9M	585M	37.4 ms	74.7
2.0-ShuffleNetV2 [13]	manual	-	-	7.4M	591M	33.3 ms	**74.9**
MnasNet-92 [20]	RL	stage-wise	91K* / 421x	4.4M	388M	-	74.8
FBNet-C (ours)	gradient	layer-wise	216 / 1.0x	5.5M	**375M**	**28.1 ms**	**74.9**

Table 3. ImageNet classification performance compared with baselines. For baseline models, we directly cite the parameter size, FLOP count and top-1 accuracy on the ImageNet validation set from their original papers. For CPU latency, we deploy and benchmark the models on the same Samsung Galaxy S8 phone with Caffe2 int8 implementation. The details of MnasNet-{64, 92} are not disclosed from [20] so we cannot measure the latency. *The search cost for MnasNet is estimated according to the description in [20]. † The search cost is estimated based on the claim from [11] that PNAS [11] is 8x lower than NAS[31]. ‡ The inference engine is faster than other models.

State of the art in GANs continues to evolve: From grainy to GANgsta

Larger models and large-batch training further improves the quality of images produced using a GANs.

Goodfellow et al. @ NIPS 2014



Brock et al. @ ICLR 2019

State of the art in GANs continues to evolve: From faces to (small) full-body synthesis

Film on-set once and generate the same video in different languages by matching the face to spoken word (left). The next step is generating entire bodies from head to toe, currently for retail purposes (right).

synthesia

DataGrid

After image and video manipulation comes realistic speech synthesis

Futurism

This AI That Sounds Just Like Joe Rogan Should Terrify Us All

Published by Kristin Houser in Artificial Intelligence

Sound-Alike

On Wednesday, Canada-based startup Dessa unveiled a new AI that replicates the voice of Joe Rogan, a

joerogan • Follow

joerogan I just listened to an AI generated audio recording of me talking about chimp hockey teams and it's terrifyingly accurate. At this point I've long ago left enough content out there that they could basically have me saying anything they want, so my position is to shrug my shoulders and shake my head in awe, and just accept it. The future is gonna be really fucking weird, kids.

5w

switchsalad Sounds like a shitty stitch job to me. Like a

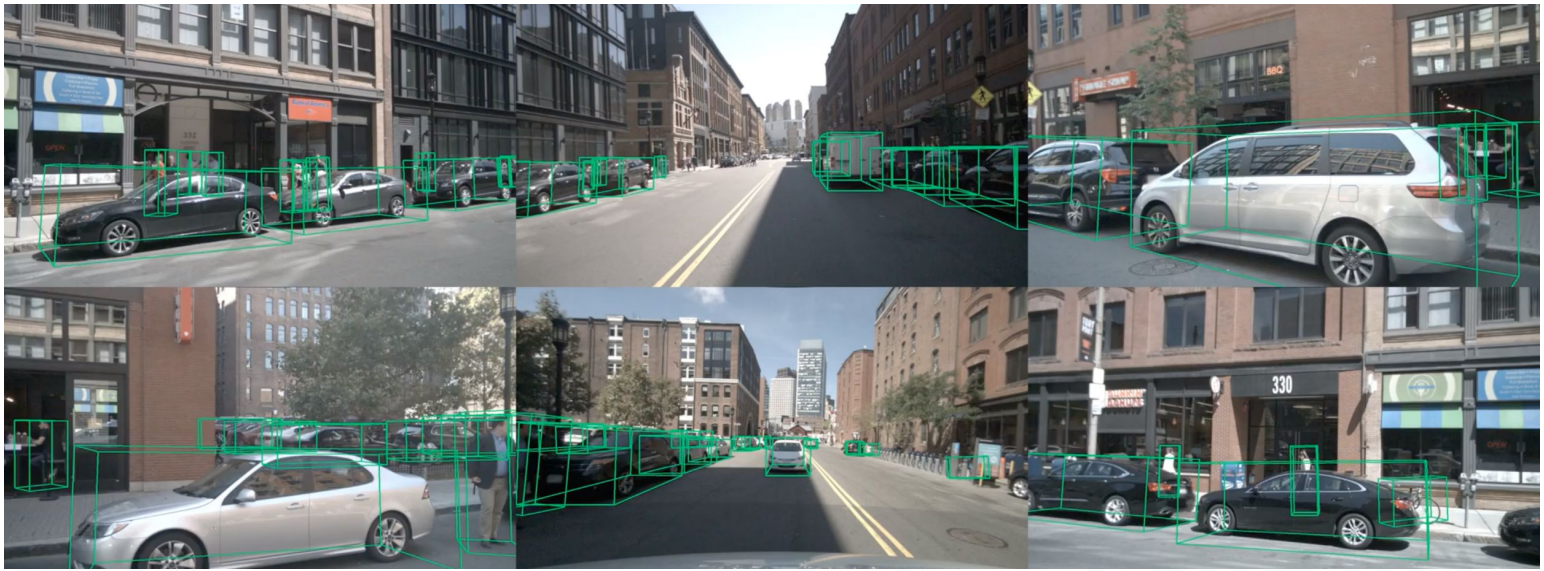
316,617 likes

MAY 17

Add a comment... Post

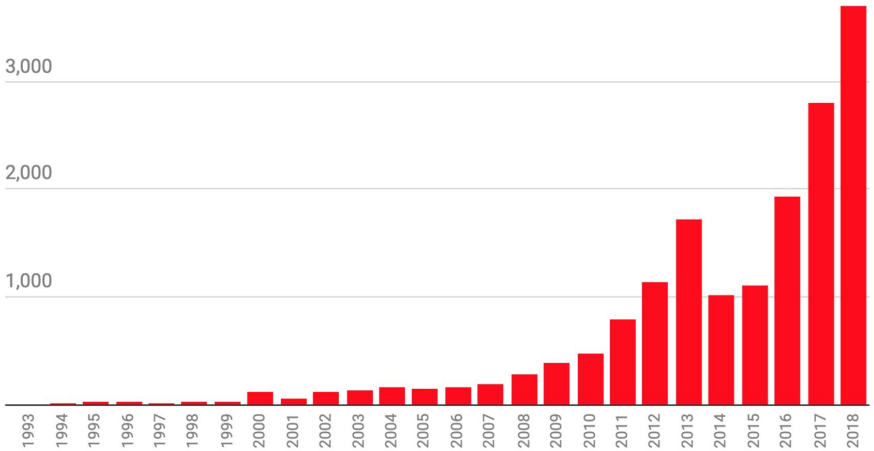
Learning the 3D shape of objects from a single image

- **A model outputs 3D bounding boxes for 10 different classes (like car, motorcycle, pedestrian, traffic cones, etc), class-specific attributes (like whether a car is driving or parking) and provides the current velocity vector.**

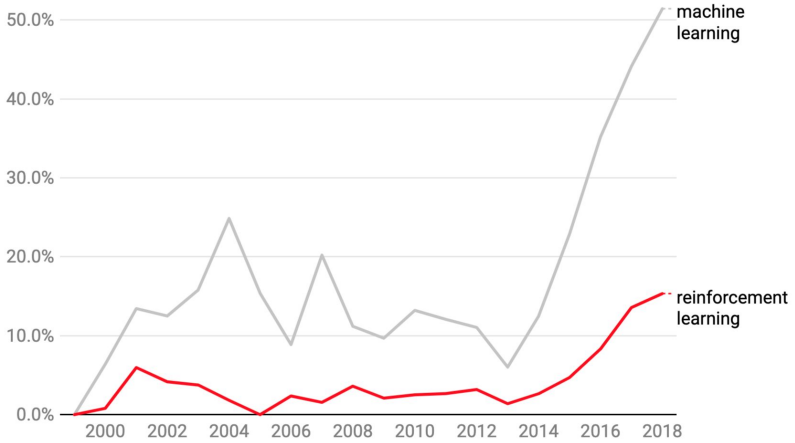


Analysis of 16,625 AI papers over 25 years shows immense growth in publication output with machine learning and reinforcement learning being the most popular topics

10x more papers annually over the last 10 years.



Over 50% of papers are about machine learning.



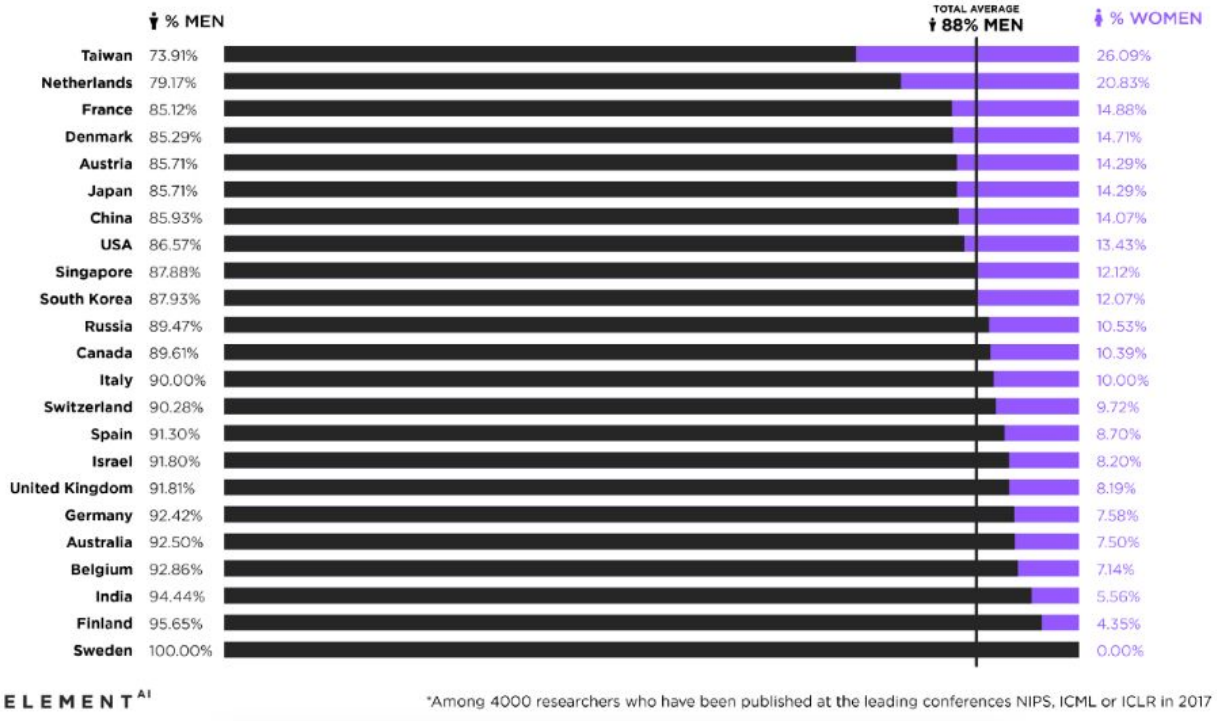
Section 2: Talent

Google continues its dominance at NeurIPS 2018, a premier academic AI conference

Google tops list of the most productive organisations measured by research paper output.

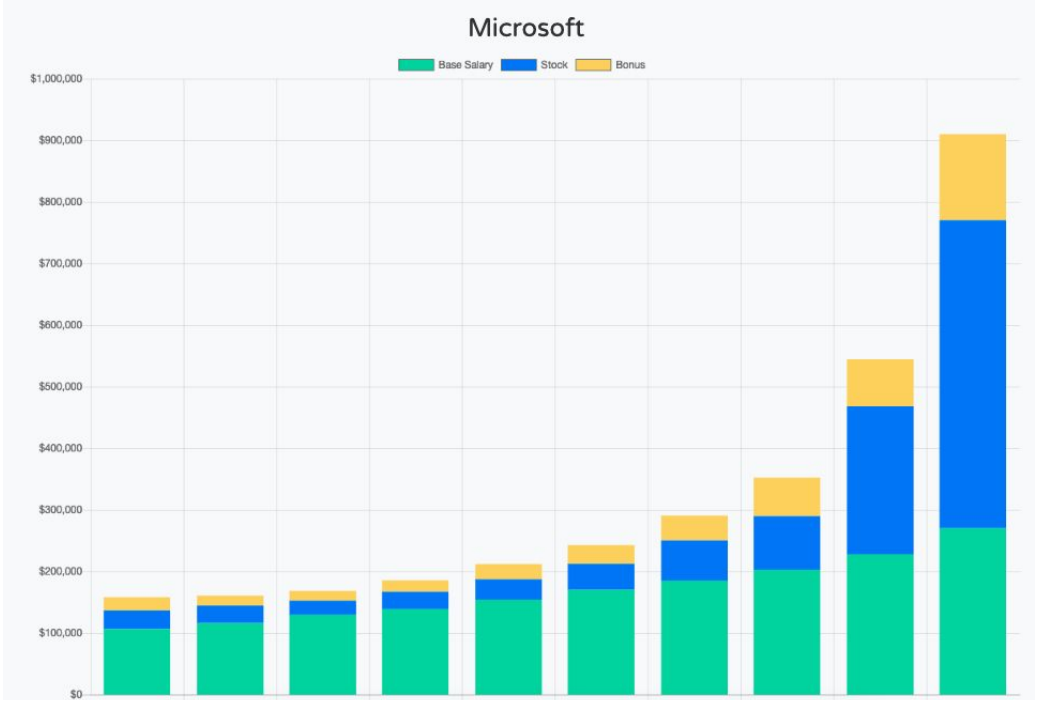
Institution	Ranking by number of 1st author papers (sum of papers)	
	2017	2018
Google	1 (38)	1 (57) ➡️
MIT	3 (30)	2 (44) ⬆️
Stanford	4 (25)	3 (38) ⬆️
CMU	2 (36)	4 (48) ⬇️
Berkeley	5 (21)	5 (32) ➡️

88% of 4,000 researchers who published papers at NeurIPS, ICML or ICLR were men



stateof.ai 2019

Compensation of senior engineers at large tech companies is approaching $1,000,000



Employee seniority grade

At the other end of the spectrum, there's huge growth in $1.47/hour data labelling jobs

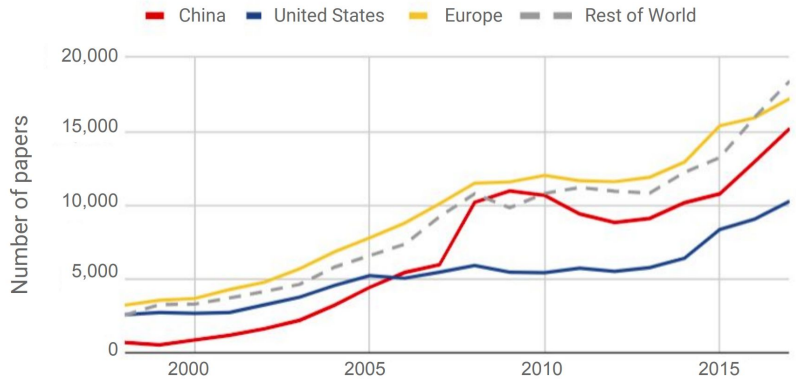
- Similar to the complex electronics supply chain (for example Foxconn), there has been massive growth in 'data labelling factories' for AI applications.
- Beijing-based Mada Code counts Microsoft and Carnegie Mellon as customers and claims to have a team of over 20,000 freelancers working for them labeling data.
- 30% of Beijing-based Basic Finder's clients are based outside of China including UC Berkeley. Minimum wage for these kind of jobs can be as low as 10 Yuan ($1.47) per hour.

Pioneers of neural networks win Turing Award, the highest award in computer science

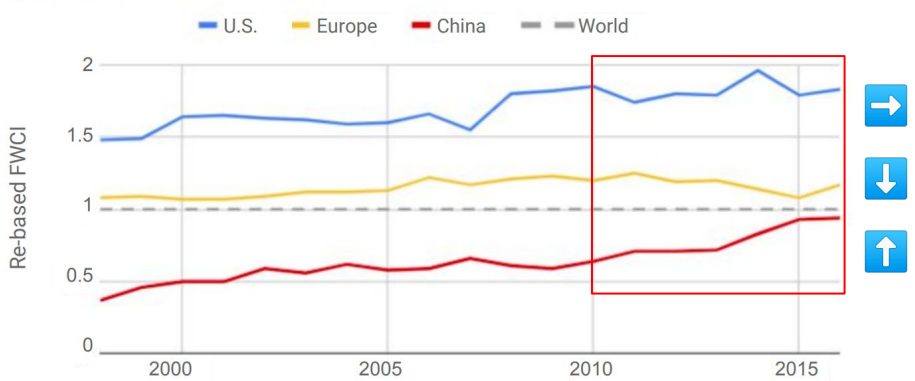
Europe publishes the most AI papers, but only China's average citation rate is growing

Europe appears to punch above its weight in output and the field as a whole is growing. The average citation impact of papers from different regions shows that only papers from China are becoming more impactful. Papers published by American authors are cited 83% more than the global average.

Annually published AI papers on Scopus by region (1998—2017)
Source: Elsevier

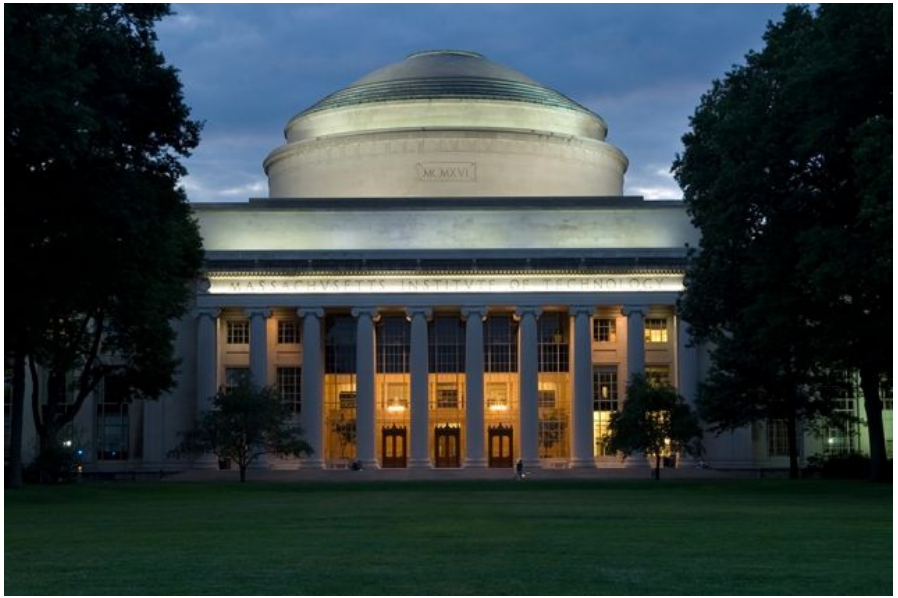


Field-Weighted Citation Impact of AI authors by region (1998—2016)
Source: Elsevier



New $1 billion investment in computing & AI at MIT: Shifting gears for new generations

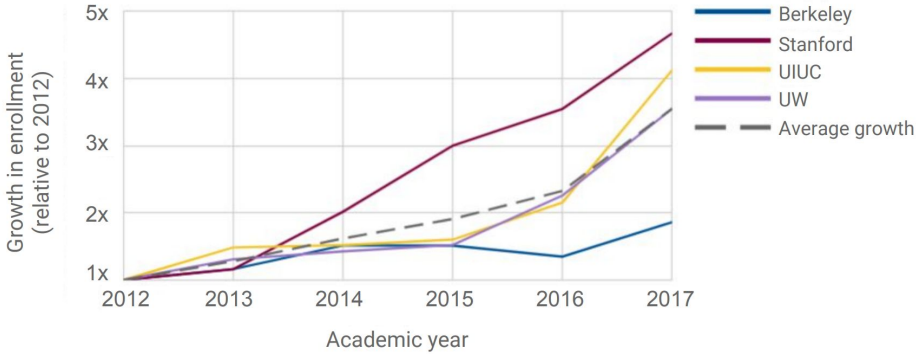
- Anchored by a $350M gift by Blackstone CEO Stephen Schwarzman, the new College of Computing will reorient MIT towards injecting AI education to all fields of study. It provides 50 new faculty positions, doubling MIT's academic capability in the field.
- Schwarzman cited his concern that the US was "lagging" China.
- Aiming to train 'bilingual' students who can apply ML to disciplines like "biology, chemistry, politics, history and linguistics".
- University of Virginia followed suit with a $130M donation by the founder of QIM (a hedge fund) to create a dedicated data science school.



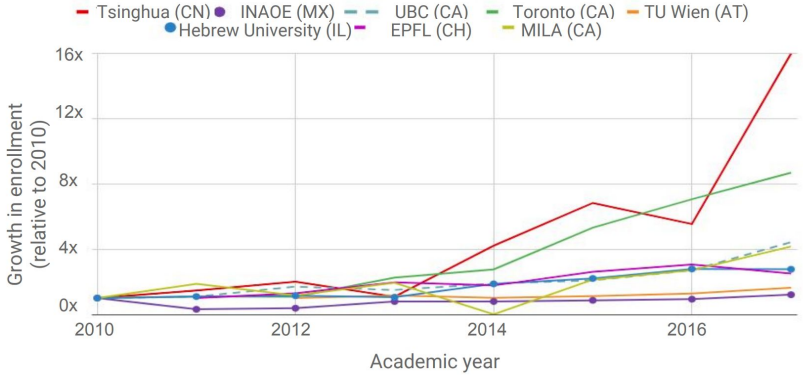
University course enrollment in AI is growing, particularly in China

Compared to 2012, almost 16x more Tsinghua and 5x more Stanford students enroll in AI courses today.

Growth in introductory AI course enrollment (2012—2017)
Source: University provided data

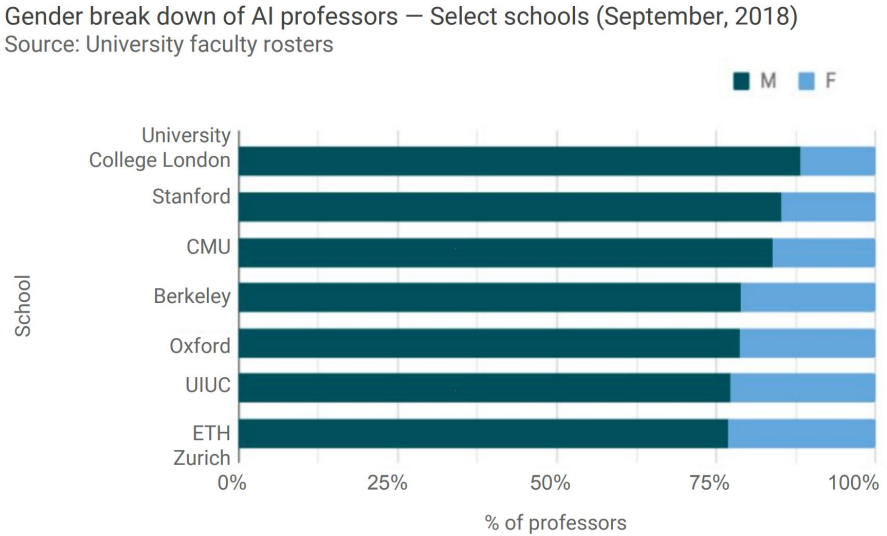


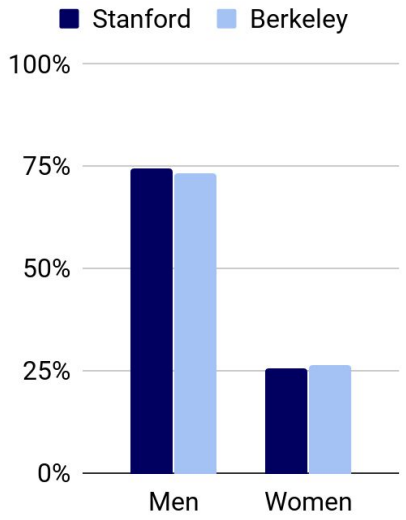
Growth of AI+ML course enrollment — Non-U.S. (2010—2017)
Source: University provided data



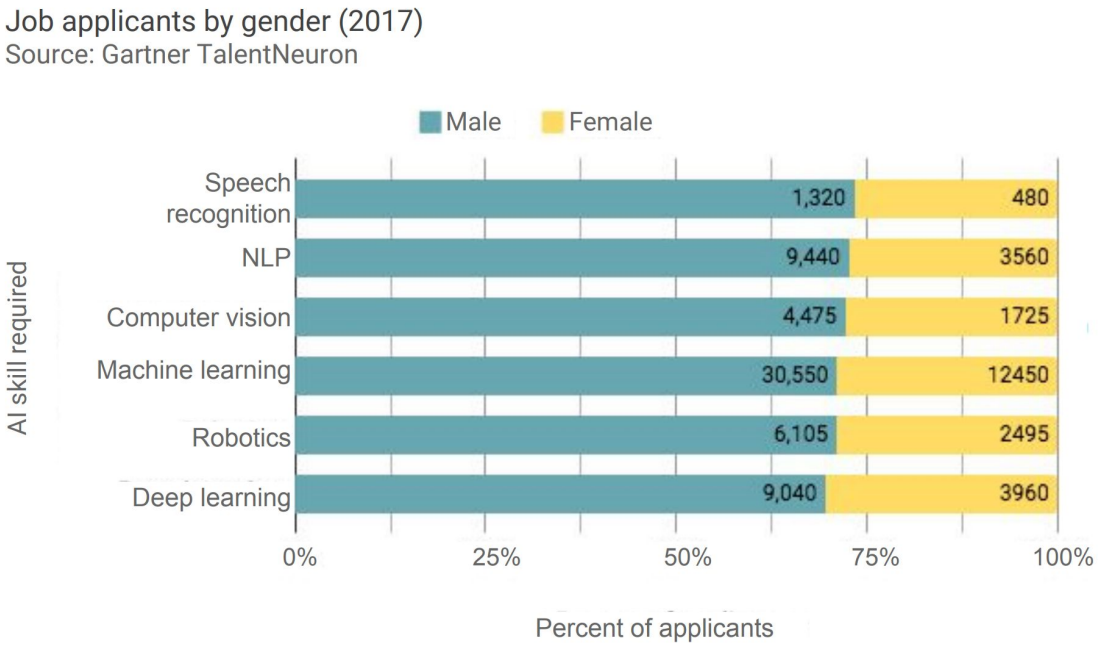
Gender diversity of AI professors and students is still far off being on an equal footing

On average, 80% of AI professors are men and 75% of students of undergraduate AI students are men.



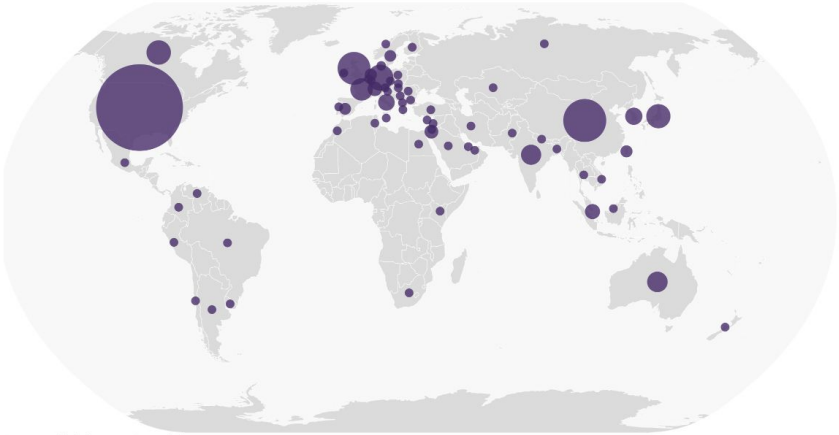


This contributes to a 71% male dominated AI applicant pool in the US



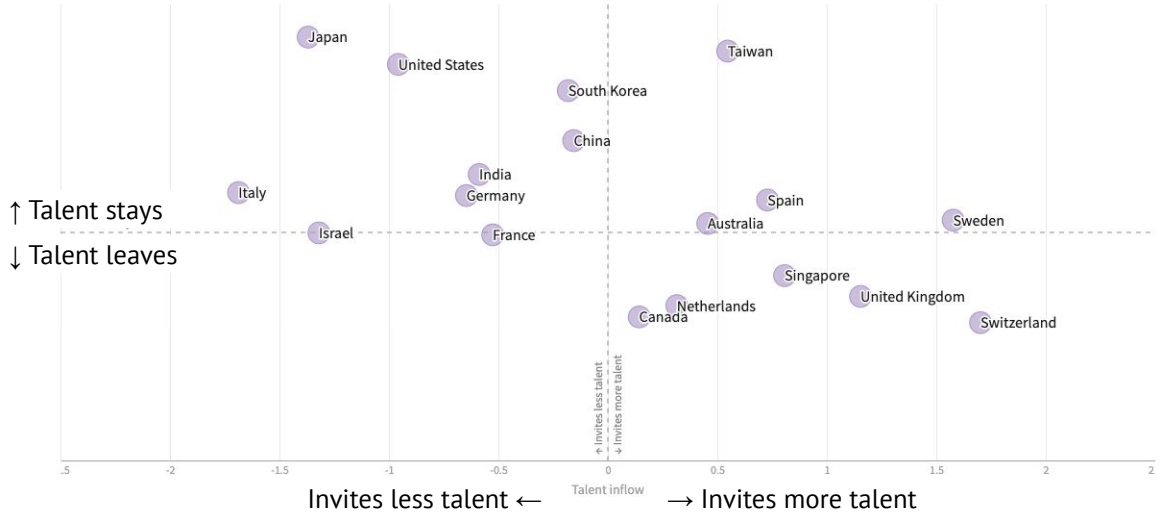
Element.AI talent survey 2019

- A review of papers published at 21 machine learning conferences by 22,400 unique authors: **Only 19% of academic authors and 16% of industry authors were women.**
- **44% of authors earned a PhD from the US, 11% from China and 6% from the UK.**
- **Five countries — the US, China, the UK, Germany and Canada — accounted for the employment of 72% of the authors.** Bubbles indicate the number of conference researchers per country.



Element.AI talent survey 2019

Plotting countries based on their inflow and outflow of AI talent: Canada, the UK and Switzerland are "platform countries" that both attract foreign talent and export locally-trained talent. The US and Chinese ecosystems are more mature - they see low inflows and outflows.



But is the trend of hiring AI researchers into companies slowing down now?

Tech giants have allegedly frozen or reduced their hiring drives for AI research talent. This is likely a sign that businesses now need talent to bring applied research into production.

AI Hiring Frenzy Settles Down

By Kevin McLaughlin Jun 11, 2019 10:00 AM PDT

The hiring boom in artificial intelligence research is finally slowing down.

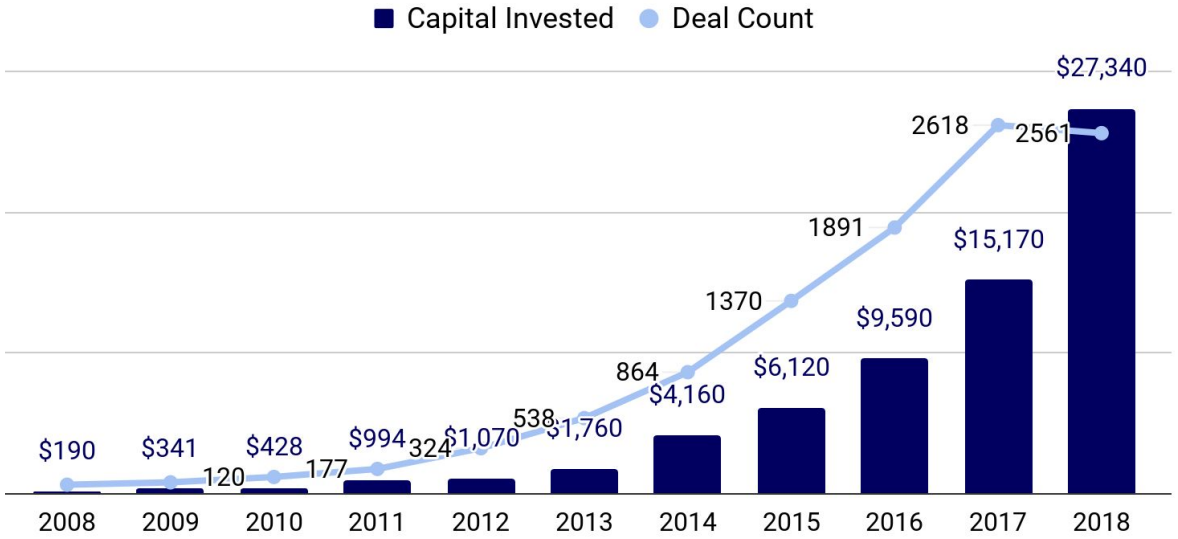
The frenzied competition for talent that drove the cost of hiring data scientists and AI researchers into the stratosphere appears to have sharply cooled. At both Alphabet and Microsoft, two of the biggest recruiters of AI talent, the growth rate of hiring for "pure" AI research positions is declining, according to six people in the field.

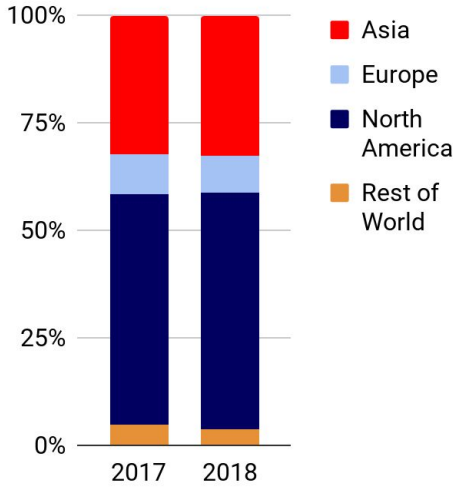
The Information

Section 3: Industry

Global venture capital investments in AI themes grow at a clip to reach >$27B/year

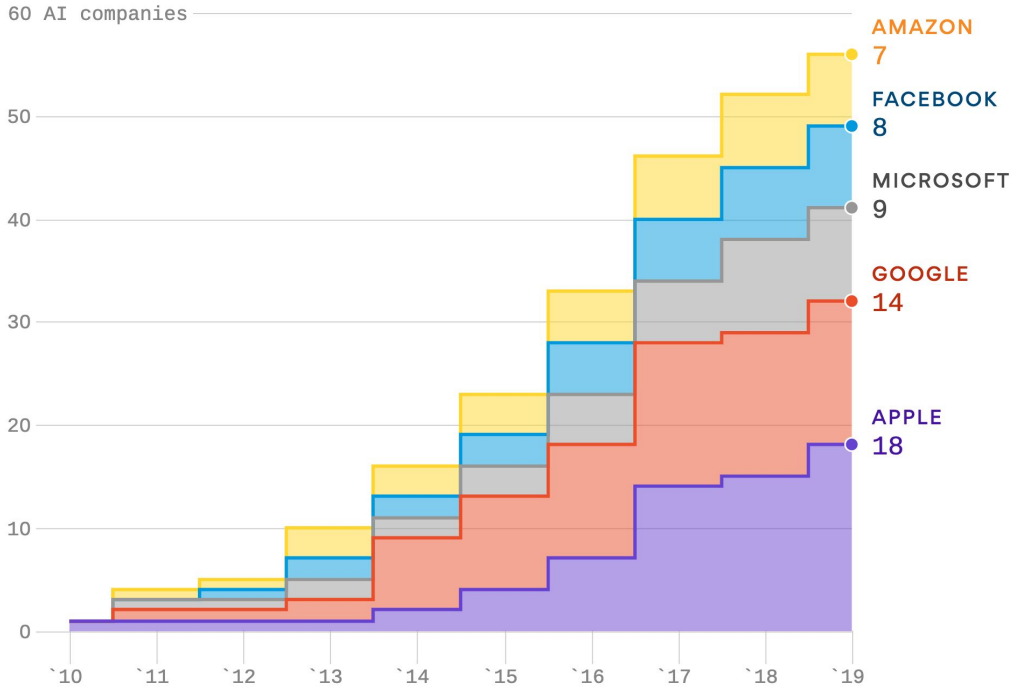
Almost 80% more capital invested in FY18 vs FY17 with North America leading the way at 55% market share.





Big tech companies continue to gobble up AI-first startups

GAFAM have completed a cumulative sum of 60 acquisitions of AI-first startups since 2010.



AXIOS

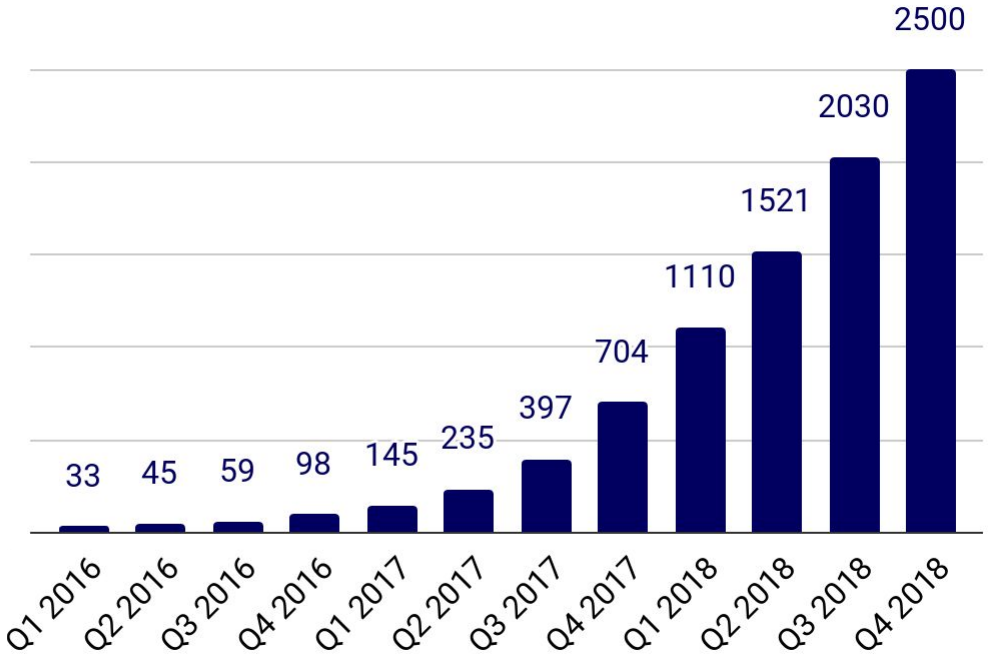
stateof.ai 2019

Spotlight on companies we featured in last year's State of AI:

nuro	$940M Series B led by SoftBank	DARKTRACE	$50M Series E led by Vitruvian
商汤 sensetime	$1B Series D led by SB China Capital	ZOOX	$500M Series B led by Grok Investments
AURORA	$530M Series B led by Sequoia Capital	SambaNova SYSTEMS	$150M Series B led by Intel Capital
GRAPHCORE	$200M Series D led by Sofina	zymergen	$400M Series C led by SoftBank
RIGHTHAND ROBOTICS	$23M Series B led by Menlo Ventures	GREYORANGE	$140M Series C led by Mithril
AUTOMATION ANYWHERE	$500M Series A led by SoftBank	UiPath	$225M Series C led by CapitalG
AVIDBOTS	$24M Series B led by True Ventures	Horizon Robotics	$600M Series B led by China Minsheng
Shift Technology	$60M Series C led by Bessemer	DataRobot	$100M Series D led by Cisco
lyft	$2.3B IPO at $24B valuation	CYLANCE	$1.5B acquisition by Blackberry

Robotic process automation: An overnight enterprise success (15 years in the making)

- **4x YoY customer growth at UiPath, a market leader in RPA, born in Europe.**



UiPath

Robotics in the real world: Cleaning and in-store operations

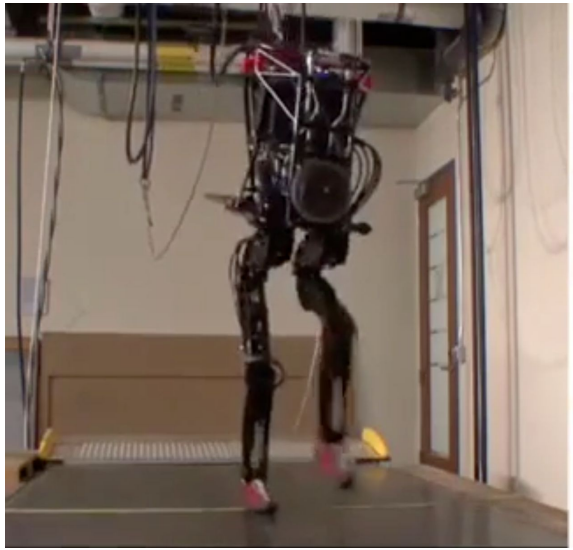
- **Brain Corp and Walmart are scaling up from an initial 360 robotic floor cleaner trial to add 1,500 more robots. This addresses a $5B commercial floor cleaning equipment market opportunity.**

brain corp

Robotics in the real world: From learning to walk to jumping around a parkour course

- **Notable progress has been made over 10 years at Boston Dynamics.**

2009 demo: clumsy walking

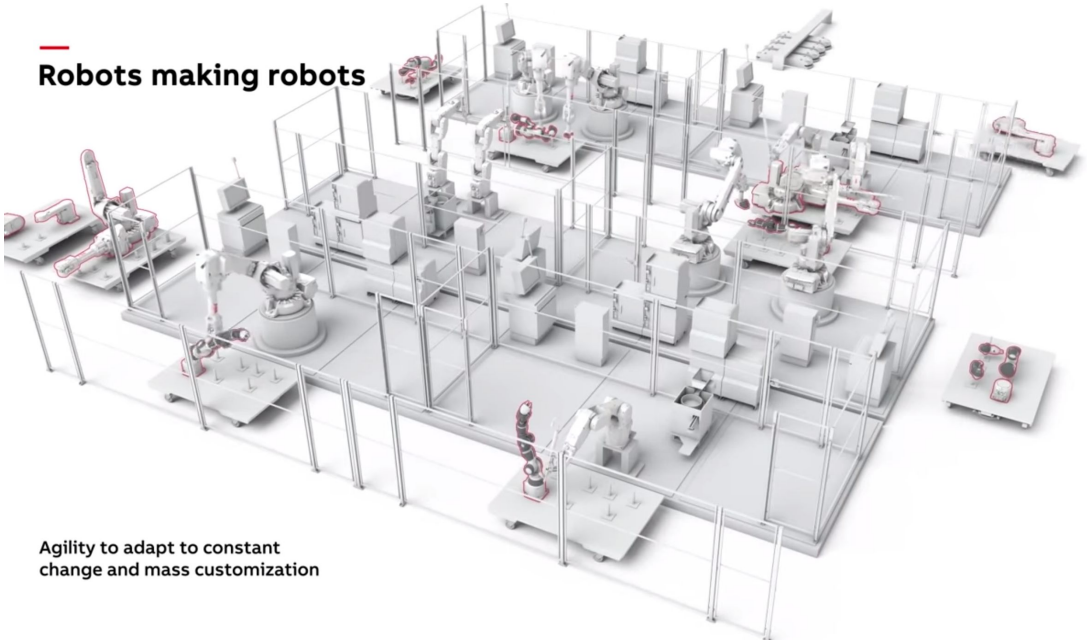


2019 demo: parkour hopping

Boston Dynamics

Robots making more robots: Big incumbents making moves

- **ABB investing $150M to build the world's most advanced, automated and flexible robotics factory in Shanghai.**



Robots making more robots: Full-stack startups enter the manufacturing market

Bright Machines is led by Autodesk and Flex veterans. They raised a $179M inaugural round and grew to 300 employees to catalyse a manufacturing paradigm where intelligent, software-defined machines build products autonomously.

Bright Machines

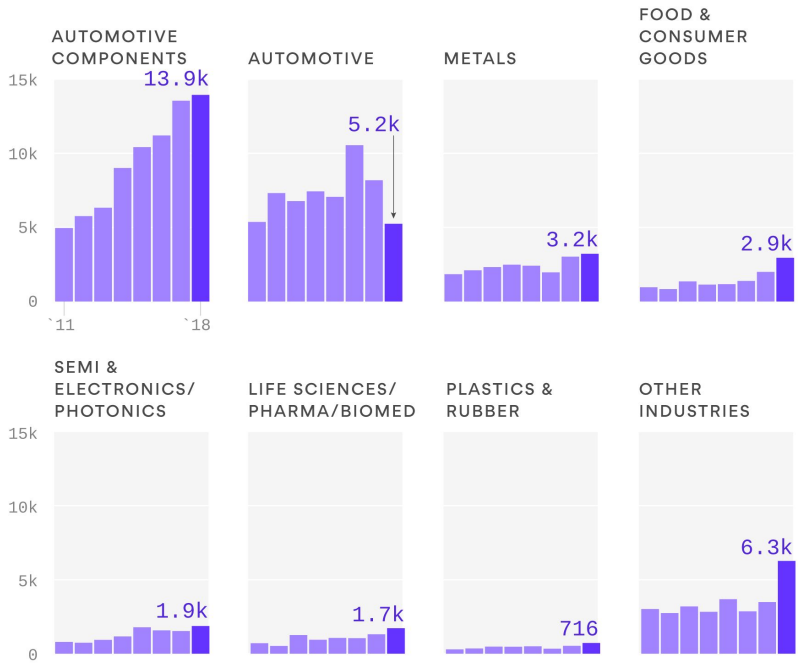
Robots-enabled supply chains: Omni-channel, automated fulfillment

Berkshire Grey offers automated picking robots, mobile ground robotic systems, product packaging and sorting solutions to increase the throughput and simplify the real-world supply chain of e-commerce operations. Led by veterans of early movers in industrial robotics: Kiva Systems and iRobot.

BERKSHIRE GREY

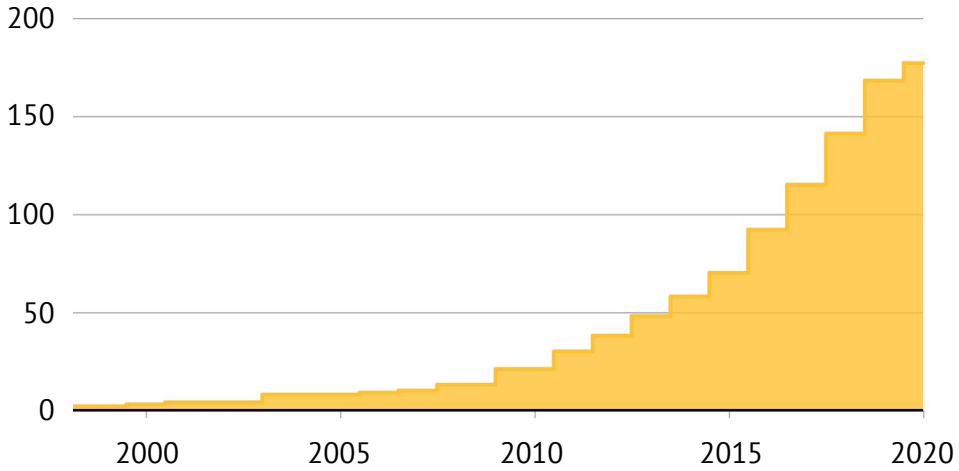
U.S. factories are installing record numbers of robots

- **35,880 robots were added to U.S. factories last year, 7% more than in 2017.** Graphs below show # units by industry between 2011-18.

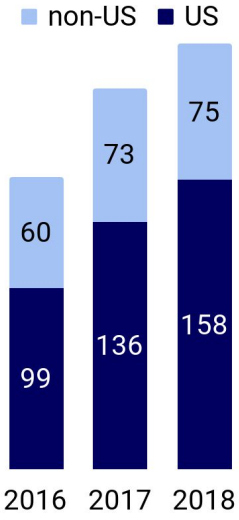


Amazon is massively scaling its physical fulfilment infrastructure

- **2x growth of US-based fulfillment centers in 4 years.**



- **45% growth in square footage '16-'19.**



amazon

Amazon rolls out more warehouse robots for fulfilment and sorting

200k robots (Amazon and third party) in warehouses today, up from 100k announced earlier this year.

Amazon has had to make many changes to its warehouses so that its robots know how to navigate around. This includes blocking out sunlight from the ceiling skylights to reduce glare, installing QR codes on the ground and re-orienting air conditioning so that it blows sideways from ground level so as not to push light objects around.

Cloud simulations coordinate robot routes.

CartonWrap by CMC: 700 boxes/hr (5x human throughput).

amazon

Self-driving cars are now a game for multi-billion dollar balance sheets

Cruise sold to General Motors for up to $1B in 2016, SoftBank Vision Fund invests $2.25B at $11.5B valuation in 2018, Honda comes in with $2.75B in 2019 to co-develop of vehicles. A few months later, Cruise raises another $1.15B at a $19B post-money valuation. Cruise spent $728M in 2018 alone, with the budget set to grow to $1B in 2019.

Uber spent $457M in 2018, $384M in 2017 and $230M in 2016 on self-driving R&D (including its flying car project). Their workforce is >1,000 strong. SoftBank Vision Fund, Toyota and DENSO invested $1B into the self-driving division at a $7.25B post-money valuation before Uber completed its public listing.

Waymo allegedly costs >$1B per year to run today and is seeking external investors. Prior to forming Waymo, Google spent $1.1B from 2009 to 2015.

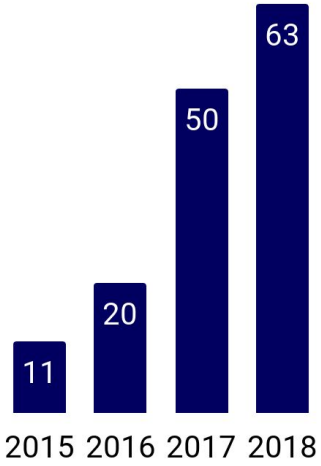
Nuro closes a gigantic $940M Series B led by SoftBank Vision Fund <3 years since its founding.

Aurora adds significant financial firepower to their balance sheet with a $530M Series B from Sequoia, Amazon and T. Rowe Price. This was later extended to a hefty $600M by Kia and Hyundai joining in.

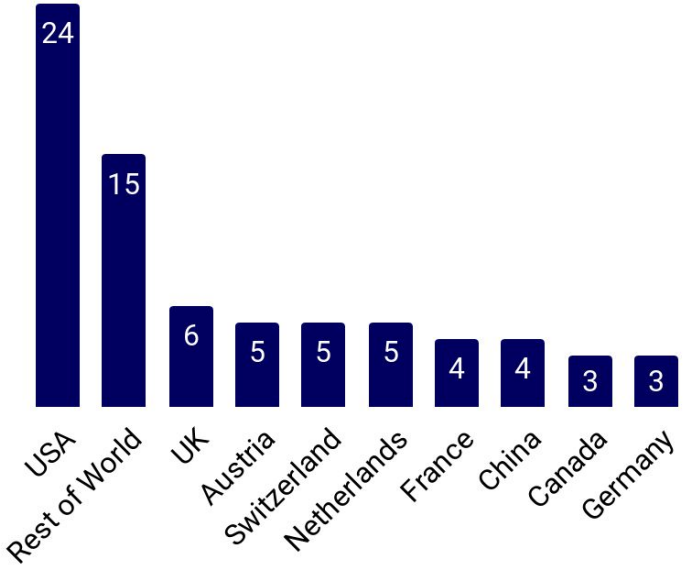
Ford allocated $4B for self-driving development, of which $1B for Argo AI.

While live AV pilots grow in California and around the world, several players retreat

Total # DMV approvals



of cities with live AV pilots in 2018



High profile AV strategy changes

Apple self-driving car layoffs hit 190 employees in Santa Clara, Sunnyvale

Roland Li | Feb. 27, 2019 | Updated: April 9, 2019 12:11 p.m.

Uber shutters its self-driving truck project

Engineers will be shifted to autonomous car research

By Andrew J. Hawkins | @andyjayhawk | Jul 30, 2018, 5:07pm EDT

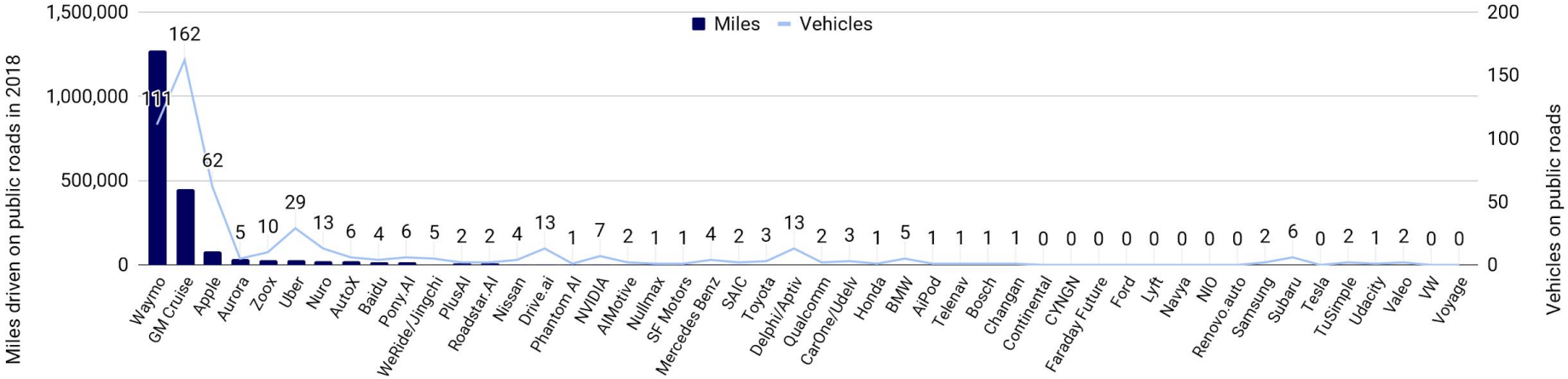
Ford CEO says the company 'overestimated' self-driving cars

Ford thinks there will be limits on what first self-driving cars can do.

Q: Was there some sort of breakthrough that Uber had between May 2016 and September 2016 that changed the projections from 13,000 units in 2019 to over 75,000 units?
A: So these results are highly speculative and depend on significant assumptions. And they change those assumptions and speculations from one report to another --

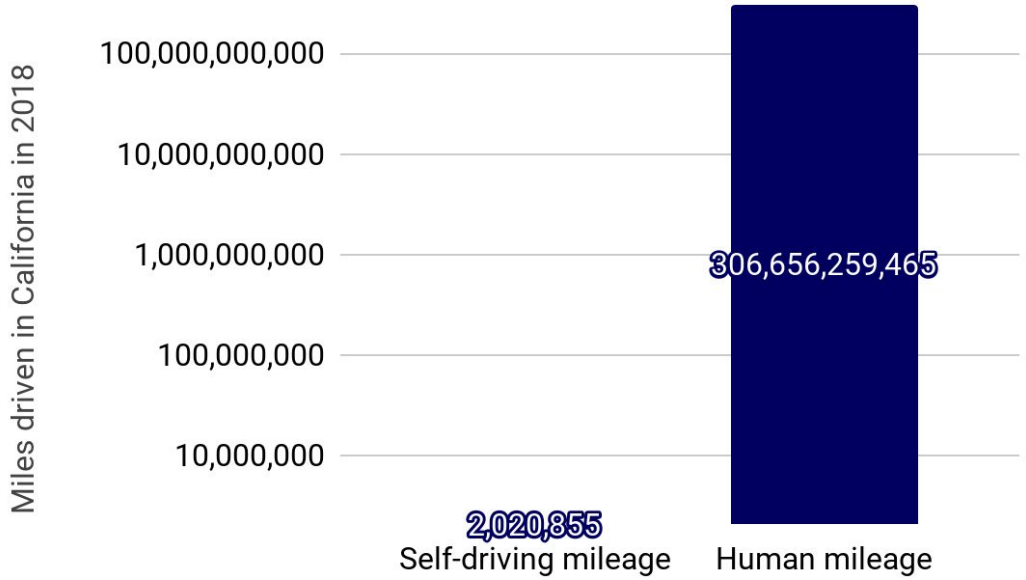
Waymo drove >1M miles in 2018, 2.8x next best (GM Cruise) and 16x 3rd best (Apple)!

The average Californian drives 14,435 miles per year. Only 11/63 companies have driven more than this in 2018.



Nonetheless, self-driving mileage accrual in California is still microscopic vs. all drivers

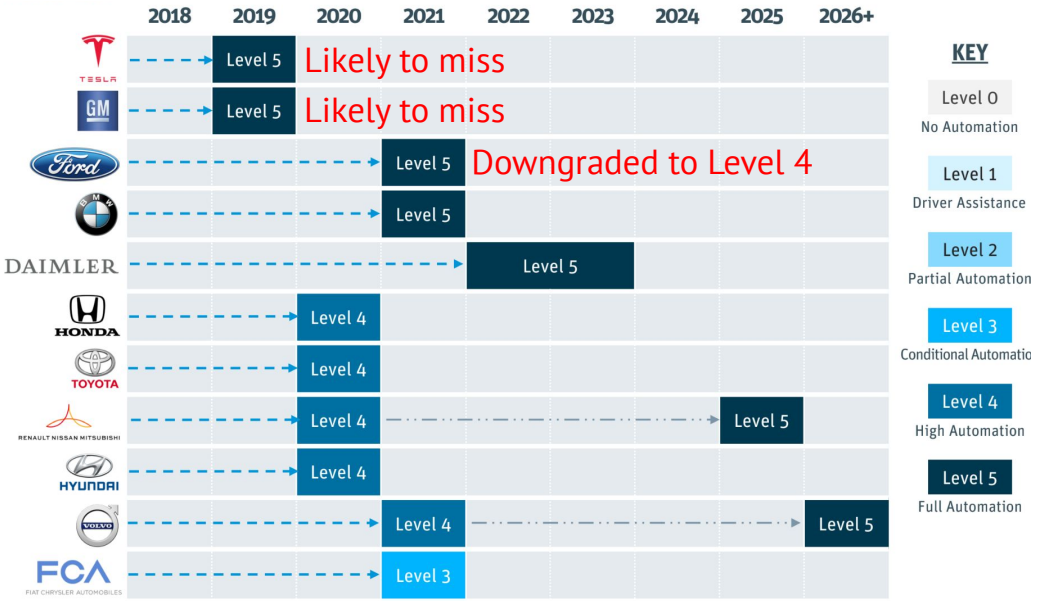
Self-driving car companies racked up 0.00066% of the miles driven by humans in California in 2018.



Leading to several missed launch dates and lots of silence from other players

- **Issues at GM/Cruise include "high severity ride discomfort" once per mile, self-driving routes taking 80% longer than their human driven peers, and failures during high-profile investor testing.**

Not much news

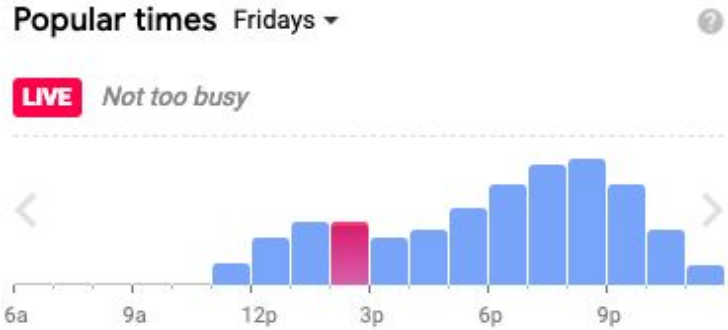


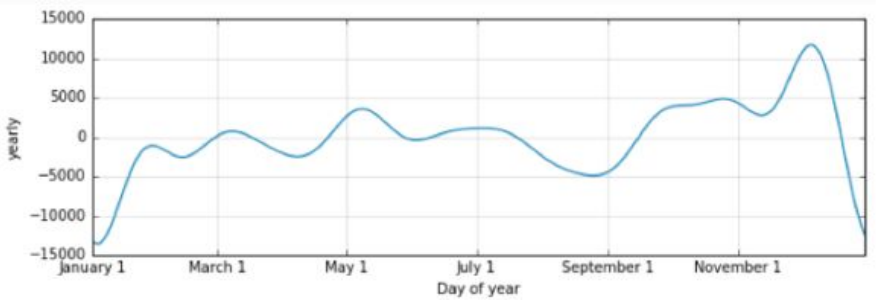
Demand forecasting

Why now?

Quantitative hedge funds have been using machine learning to forecast demand for a long time (for example predicting demand for a given commodity).

As more information about the world is available in digital form (from satellites, social media, ERP systems etc) it becomes possible to use machine learning to forecast demand beyond finance. With better demand forecasts, businesses can take action to more accurately prepare supply and in the process reduce waste and increase their profitability.





Demand forecasting

Where and how is machine learning being used effectively?

- **Energy:** Grid scale electricity is currently hard to store. This creates a substantial economic and environmental cost for underestimating demand (use of peaker plants; blackouts) and overestimating demand (wasted energy).

  Invenia is an early leader in this space. Rather than operating as a traditional energy utility, they operate as a virtual utility and are paid by Independent System Operators in various US electricity grids (California, Texas) for making more accurate predictions than other participants. Their system makes use of weather information, grid operation data and power flows to predict demand.

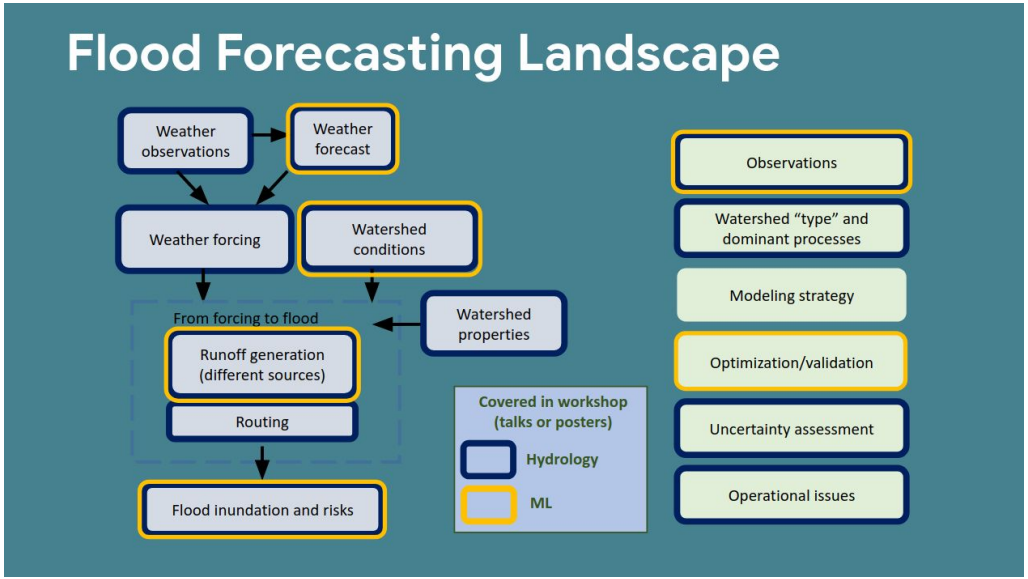
  INVENIA

  We would also like to highlight the work of **Open Climate Fix**, a new non-profit founded by a former DeepMind employee who worked on DeepMind's energy projects. The organisation is now developing open source tools and curating datasets for more accurate solar PV forecasting.

Demand forecasting

Where and how is machine learning being used effectively?

- **Flood forecasting:** Using ML to automatically calibrate hydrologic models that are central to understanding, predicting, and managing water resources.



Demand forecasting

Where and how is machine learning being used effectively?

- **Travel:** Demand for flights and hotels fluctuates based on seasonality, weather or large external events. Machine learning is extending the work that airlines and hotels already do to forecast demand. ML systems can help predict potential bookings for hotels, demand for a specific airline route or service disruptions.

  Selected examples: volantio MIGACORE

- **Local businesses:** Demand at restaurants, coffee shops or other high street shops is partly dependent on weather and external events. Better demand forecasts allow these businesses to adjust staffing and supplies and increase profitability while reducing wastage.

  Selected examples: tenzo dynamic yield

- **Logistics:** Probabilistic models and multi-agent systems can be used to learn how to optimally allocate resources (e.g. fleets of vehicles) to address dynamically changing demand (e.g. passenger requests) while maximising resource utilisation. Here, the resource allocation problem is solved for many potential futures.

  Selected examples: PROWLER.io the decision company

Demand forecasting

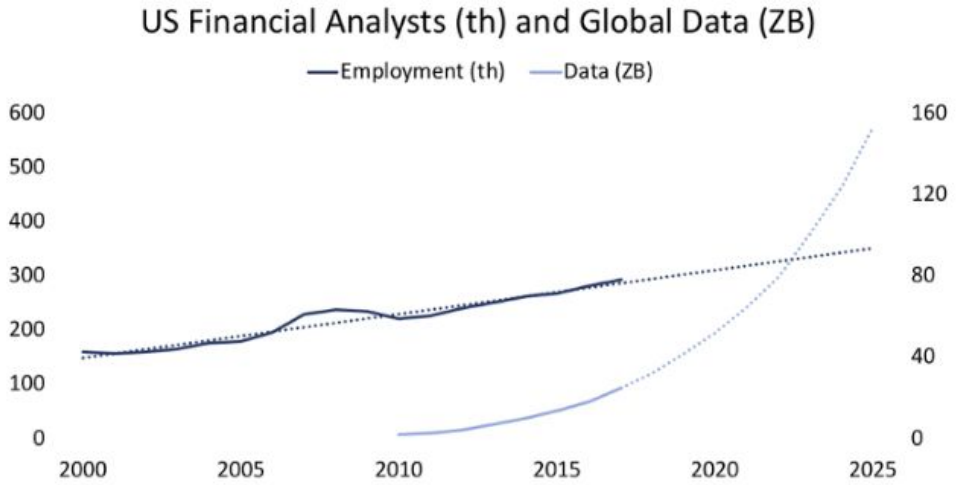
Where and how is machine learning being used effectively?

- **Retail:** A typical supermarket chain has to make millions of stock replenishment decisions each day.
  - Blue Yonder has enabled Morrisons (a supermarket chain in the UK) to fully automate 99% of 20M daily replenishment decisions and in the process improved its profitability and reduced waste.
  - Beyond replenishment, Blue Yonder is applying reinforcement learning to do automated dynamic pricing. This is particularly effective for items that are perishable or subject to consumer trends.
  - The long term vision for the company is to fully automate the entire supply chain from manufacturer to the customer.

Selected example: BlueYonder a jda. company

Reading machines are improving and proliferating

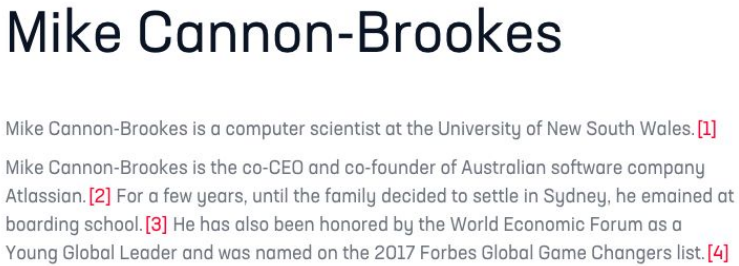
- The **breakthroughs in natural language processing** highlighted in the Research section of this report are starting to be applied to industries where there are either large amounts of text to be processed or where there is substantial financial return from processing text faster. The amount of text available for analysis is growing faster than the number of human analysts, creating an opportunity for start-ups to build new NLP based tools.



Reading machines: Augmenting professional analysts

Where and how is machine learning being used effectively?

- Primer, a SF-based startup, is using natural language processing and generation to automate tasks typically performed by analysts, such as finding, reading, cross-referencing and summarising.
- Today, this is a natural fit for finance and the intelligence community who both place a value on deriving important signals from vast amounts of online data.
- Reading and writing are important in most jobs so the potential for this technology to diffuse into other industries seems high. As a demonstration of its potential, Primer was able to 'find' 40,000 scientists who were missing from Wikipedia, but had similar levels of news coverage to scientists listed on Wikipedia. Their software was then able to machine generate Wikipedia entries. Read more: quicksilver.primer.ai



Selected example: PRIMER

Healthcare: The US FDA cleared 3 AI-based diagnostic products in the last 12 months

- **April 11, 2018:** IDx software for diabetic retinopathy detection from eye scans. This condition happens when high levels of blood sugar lead to damage in the blood vessels of the retina. In the US, 30 million patients suffer from this disease. The company ran a clinical study on 900 patients with diabetes from 10 primary care sites. The system could correctly identify the presence of more than mild diabetic retinopathy in 87.4% of cases. It correctly called when a patient did not have more than mild diabetic retinopathy in 89.5% of cases.

- **May 24, 2018:** I M A G E N software to detect wrist fractures in adult patients from 2D X-ray images. The intended use spans primary care, emergency medicine, urgent care and orthopedics. In a retrospective study of 1,000 wrist X-ray images, the model produced an AUC of ROC of 0.965. They also showed that when 24 X-ray readers were made to analyse 200 X-rays with or without OsteoDetect, those with OsteoDetect achieved an AUC of 0.889. This was higher than those who read X-rays without the software (AUC of 0.840).

- **November 7, 2018:** maxQ Artificial Intelligence software to prioritize the clinical assessment of adult non-contrast head computed tomography (CT) cases that exhibit indications of intracranial hemorrhage (ICH), commonly known as a brain bleed. These are fatal in about 40% of cases. Of those who survive, about ⅔ suffer some permanent neurological deficit.

Healthcare: Pharma companies partner with AI-driven drug development companies

Atomwise + charles river

Atomwise, an SF-based startup, uses convolutional neural networks to predict the binding capacity of small molecule drugs to target proteins of interest. This partnership with Charles River Laboratories will support the contract research organisation's hit discovery, hit-to-lead, and lead optimization efforts. Atomwise receives technology access fees, milestone-based payments and royalties from clients. Atomwise projects that the total potential value of the royalties to Atomwise with success in *all* projects could exceed US$2.4 billion.

Exscientia + Celgene

Exscientia, a startup based in Scotland, claim to be able to reduce the time to discover pre-clinical drug candidates by at least 75%. Their partnership with Celgene, a global pharma company focused on cancer and inflammatory disorders, includes an initial $25M upfront payment with Exscientia being eligible to receive significant milestone and royalty payments based on the success of the programme and future sales.

Healthcare: Pharma companies partner with AI-driven drug development companies

LabGenius + Tillotts Pharma (Zeria Group)

LabGenius, a London-based startup, uses AI-driven evolution strategies to develop radically improved protein therapeutics. They have entered into a discovery agreement with Tillotts Pharma to identify and develop new drug candidates for the treatment of inflammatory bowel diseases (IBD), such as Crohn's disease. The two-year long collaboration sees LabGenius generate molecules that, together with Tillotts, will be filtered into lead candidates for future development and commercialisation by Tillotts.

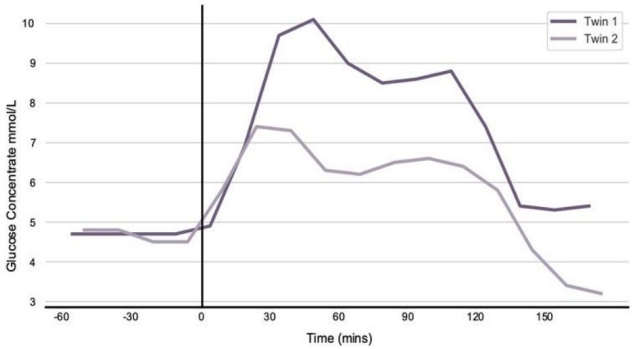
insitro + Gilead

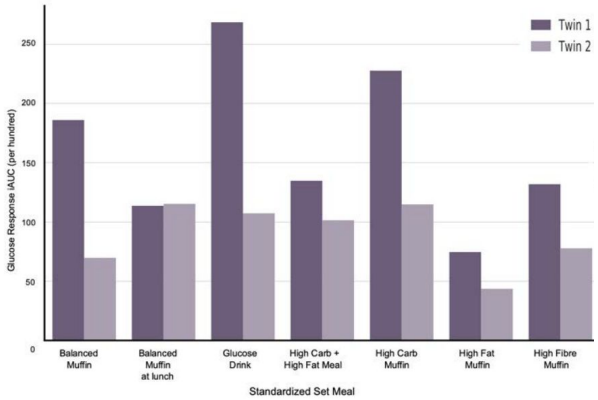
Insitro, an SF-based startup, launched with $100M in VC financing and announced a three-year collaboration with Gilead. The deal is to create disease models for nonalcoholic steatohepatitis (NASH), a chronic form of liver disease with limited treatment options and that can result in cancer. insitro will receive an upfront payment of $15M, with additional near-term payments up to $35M based on operational milestones. insitro will be eligible to receive up to $200M for the achievement of preclinical, development, regulatory and commercial milestones for each of the five potential Gilead targets.

Nutrition: Using genetic, metabolomic, metagenomic and meal-context information from 1,100 study participants to predict individuals' metabolic response to food

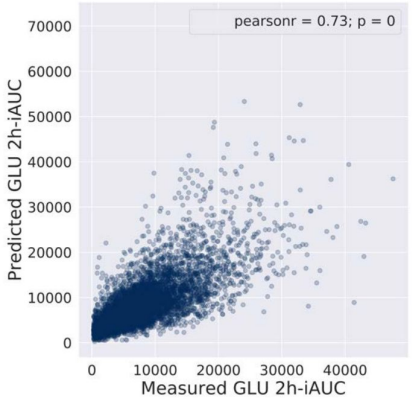
- Identical twins have very different responses to the same foods. ML predictions of glucose response two hours after meal consumption correlate 73% of the time with actual measured responses.

Twin's respond differently to a high carb muffin and to the same set meals.





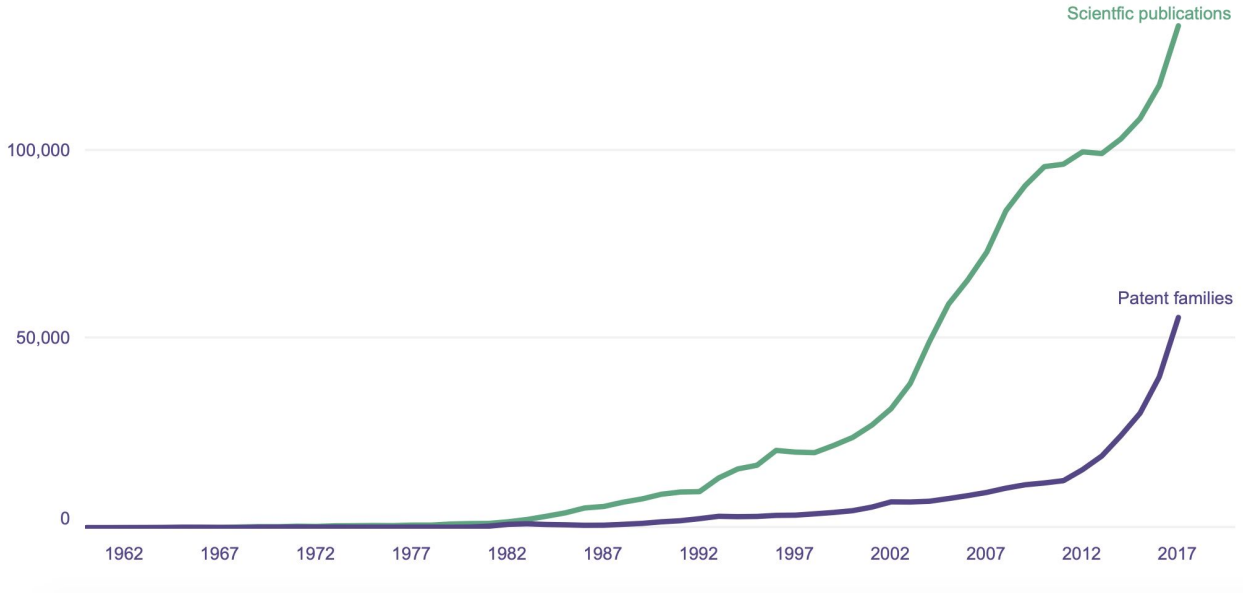
ML predictions vs. measured.



ZOE

Patenting in AI

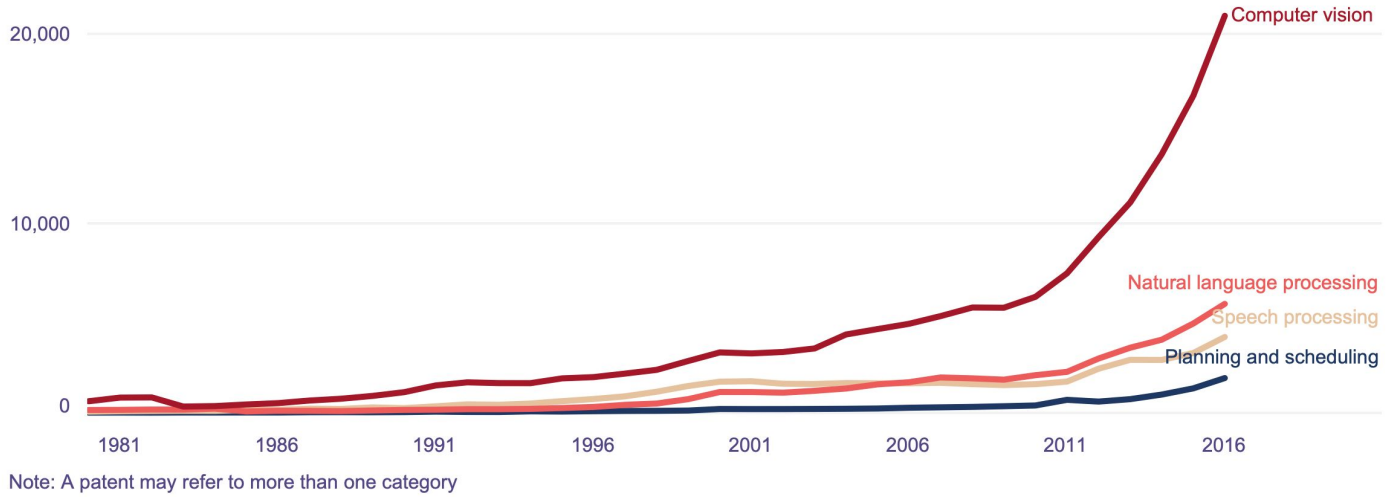
From 2012 to 2017 **AI patent families grew faster than AI scientific publications** (28% vs 6% annually). The ratio of scientific papers to patents has fallen dramatically as machine learning finds a greater number of commercial applications.



WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

Patenting in AI

- **Computer vision is the most popular area for patents.** Within computer vision the most popular area is biometrics (applications related to biological data).

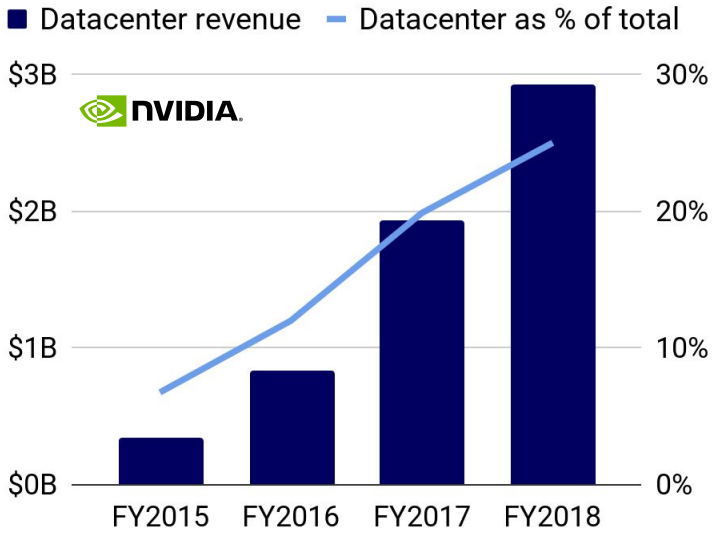


WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

Big tech companies monetise cloud computation, but not their hosted AI services

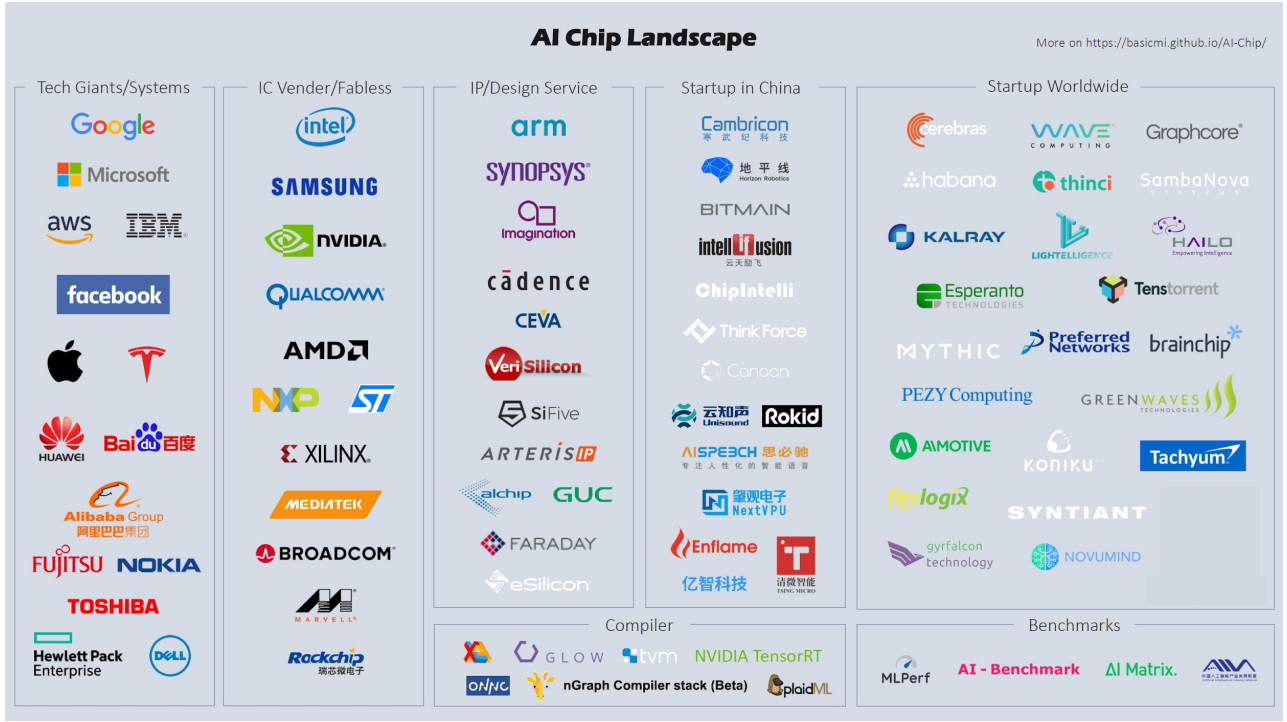
It's still very early days for monetising hosted AI services. In contrast, revenue from overall cloud computation (Amazon AWS) and AI hardware sales for data centers (NVIDIA*) is very strong and growing.

aws Product line	Purpose	Sales in 2018
SageMaker	ML infrastructure	$11M
Rekognition API	Computer vision	$3M
Lex API	NLP bots	$0.35M
Polly API	Text-to-speech	$0.5M
AWS overall	Compute, storage, networking, software	$25,656M



*Note that NVIDIA datacenter revenue includes both its AI cloud revenue and revenue from sales of AI hardware to 3rd party data centers.

AI hardware: A flourishing, global landscape of giants and startups



AI hardware: Benchmarking the performance of mobile chipsets for AI tasks

Qualcomm's Snapdragon wins by demonstrating very strong performance and hardware acceleration for both float and quantized neural networks. Benchmarking tasks include classification, face recognition, deblurring, Super-resolution, segmentation and enhancement.

Processor	Cores	Year	CPU Q AI Score	CPU F AI Score	QUANT Score	QUANT Accuracy	FP16 Score	FP16 Accuracy	FP32 Score	FP PAR Score	Accuracy	AI-Score
Mediatek Helio P90	CPU (4xCortex-A75 + 4xCortex-A55) + DSP x 2 + APU	2018	1054	2012	6212	98	9910	95	158	46	96	19496 [1.4]
Snapdragon 855	CPU (8xKryo) + DSP (Hexagon 690) + GPU (Adreno 640)	2018	1988	3598	3695	55	7361	37	1158	831	43	18924 [1]
HiSilicon Kirin 980	CPU (4xCortex-A76 + 4xCortex-A55) + NPU x 2 / n.a.	2018	1817	3447	222	60	10750	85	139	64	76	16684 [2]
Snapdragon 845	CPU (8xKryo) + DSP (Hexagon 685) + GPU (Adreno 630)	2018	1580	2176	2028	58	6298	38	929	701	45	13868 [1]
Spreadtrum ud710	n.a.	2019	1371	1497	112	60	7540	59	67	30	59	10773 [2.4]
Exynos 9820 Octa	CPU (2xM4 & 2xA75 & 4xA55) + GPU (Mali-G76 MP12)	2018	1737	2021	960	99	2244	99	186	139	99	7288 [1]
HiSilicon Kirin 970	CPU (4xCortex-A73 & 4xCortex-A53) + NPU / n.a.	2017	1286	1976	166	60	3431	58	132	51	59	7147 [2]
Snapdragon 845	CPU (8xKryo 385 Gold&Silver) + DSP (Hexagon 685)	2018	1667	2216	1724	58	851	99	130	47	85	6752 [3]
Mediatek Helio P60	CPU (4xA73 + 4xA53) + GPU (Mali-G72 MP3) + APU	2018	1151	1765	1334	99	1266	70	113	152	79	5806 [1]
Snapdragon 675	CPU (8xKryo 460 Gold&Silver) + DSP (Hexagon 685)	2018	1015	2000	1761	88	739	99	108	44	95	5785 [3]

ai-benchmark.com

AI hardware: Benchmarking the performance of mobile handsets for AI tasks

Samsung, Huawei and Xiaomi top the list whereas Google's Pixel 3 holds position #22.

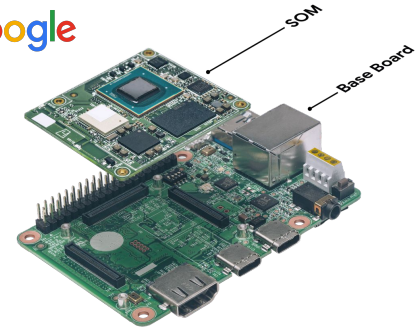
Model	CPU	RAM	Year	Android	Updated	CPU Q AI Score	CPU F AI Score	QUANT Score	QUANT Accuracy	FP16 Score	FP16 Accuracy	FP32 Score	FP PAR Score	Memory	Accuracy	AI-Score
Samsung Galaxy S10+	Snapdragon 855	8GB	2019	9	4.19	2022	3598	3400	55	7100	37	1135	800	1000	43	24488 [1]
Samsung Galaxy S10	Snapdragon 855	6GB	2019	9	4.19	2006	3548	3371	55	7157	37	1169	833	1000	43	24463 [1]
Huawei P30 Pro	HiSilicon Kirin 980	8GB	2019	9	4.19	1817	3454	223	60	10479	85	141	62	2000	76	23874 [2,6]
Xiaomi Mi 9	Snapdragon 855	8GB	2019	9	4.19	2084	3111	3208	40	6933	30	1107	677	1000	34	23199 [1]
Huawei Mate 20 Pro	HiSilicon Kirin 980	6GB	2018	9	4.19	1798	3367	218	60	8739	28	136	64	2000	39	21125 [2,6]
Huawei Mate 20	HiSilicon Kirin 980	4GB	2018	9	4.19	1829	3364	217	60	8601	28	133	69	2000	39	20973 [2,6]
Huawei Mate 20 X	HiSilicon Kirin 980	6GB	2018	9	4.19	1815	3377	218	60	8585	28	133	76	2000	39	20959 [2,6]
Honor View 20	HiSilicon Kirin 980	8GB	2018	9	4.19	1808	3364	218	60	8420	28	132	66	2000	39	20674 [2,6]
Samsung Galaxy S9+	Snapdragon 845	6GB	2018	9	4.19	1637	2214	2044	58	6464	38	926	714	1000	45	18885 [1]
Samsung Galaxy S9	Snapdragon 845	4GB	2018	9	4.19	1548	2212	2021	58	6377	38	926	700	1000	45	18591 [1]

ai-benchmark.com

AI hardware: Pushing compute and competition to the edge

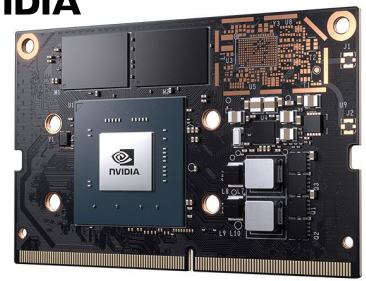
Google and NVIDIA throw their hats in the ring to apply AI computation to the 40 trillion gigabytes of data generated from connected devices by 2025.

Google



The Edge TPU is an ASIC chip designed to run TensorFlow Lite ML models at the edge. The dev kit includes a system on module (SOM) that combines Google's Edge TPU, a NXP CPU, Wi-Fi, and microchip's secure element in a compact form factor.

NVIDIA



Jetson Nano delivers 472 GFLOPs and can run AI models at just 5 to 10 watts.

AI hardware: Pushing compute and competition to the edge

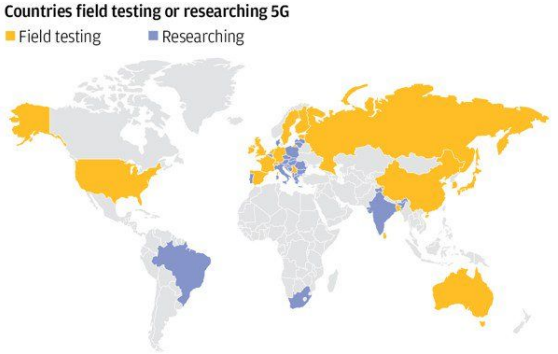
Amazon, on the other hand, launch SageMaker Neo to let developers train ML models on their cloud and export optimised models tailored to specific edge hardware platforms.

Neo's optimization and reduced resource usage techniques are open sourced at github.com/neo-ai

5G as the backbone of ubiquitous connectivity and AI computation

5G offers the potential for much faster and more stable information transmission. The organisation or country that owns 5G will set the standards for the rest of the world. Right now, China is far ahead of the United States.

- According to estimates, Huawei has 28% of the world's telecoms market, and data from German firm IPlytics shows that Huawei owns the most 5G standard essential patents (1,529) followed by Finland's Nokia (1,397).
- Having a patent advantage and the most commercial momentum is likely to position Huawei as the key player in building an ecosystem of network providers, device makers and app developers.
- In Europe, the UK and Germany are using Huawei hardware whereas the US still vehemently opposes its hardware. Interestingly, Huawei does not have 5G contracts in China.
- 5G is one of China's "new area of growth" in their 13th Five-Year Plan.



Mobile network evolution

Year	Generation	Maximum data speed**	Data transfer per second**	Time to download full HD movie***
1979	1G	none (voice only)	n/a	n/a
1991	2G	14.4 Kbps	1.8K	Over a month
2000	2.5G	53.6 Kbps	6.7K	Over a week
2001	3G	384 Kbps	48K	Over a day
2010	4G	100 Mbps	12.5MB	7 minutes
2020	5G	1 Gbps	125MB	40 seconds

* 1 exabyte = 1 billion gigabytes ** Theoretically possible *** 5GB file size

South China Morning Post

Section 4: Politics

Public Attitudes to AI

In this Section, we will review selected results from two major surveys of attitudes to AI and automation:

Brookings survey: Attitudes to AI

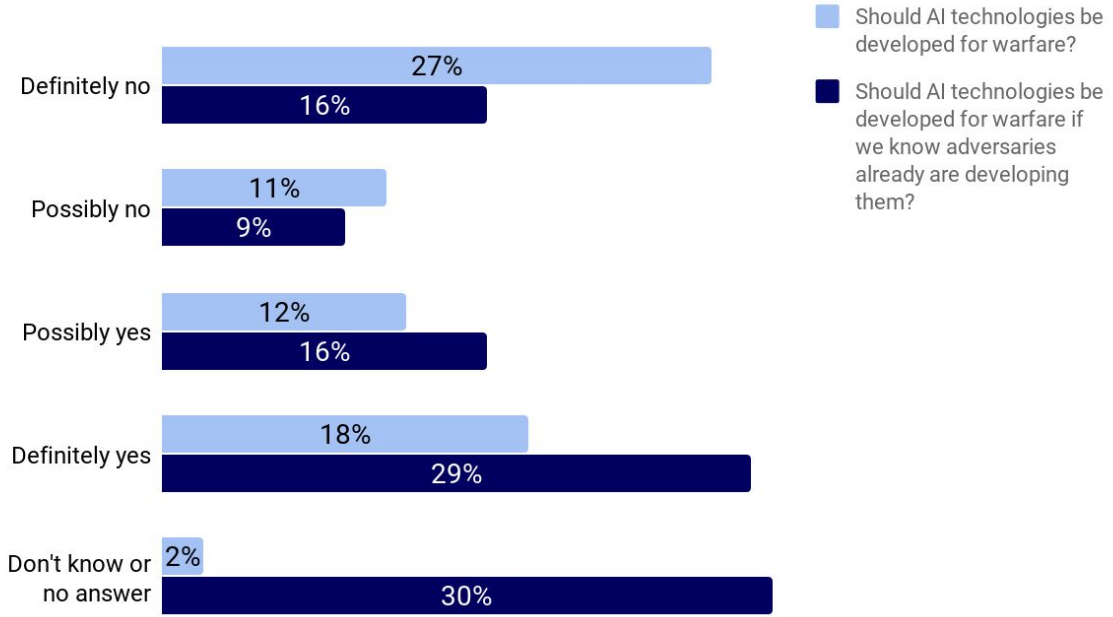
- Conducted August 19-21 2018. Published August 2017.
- Survey of 2,000 adult internet users in the U.S.

Future of Humanity Institute: AI - American Attitudes and Trends

- Conducted June 6-14 2018. Published January 2019.
- Survey of 2,000 US adults..

Public Attitudes to AI: Warfare and double standards

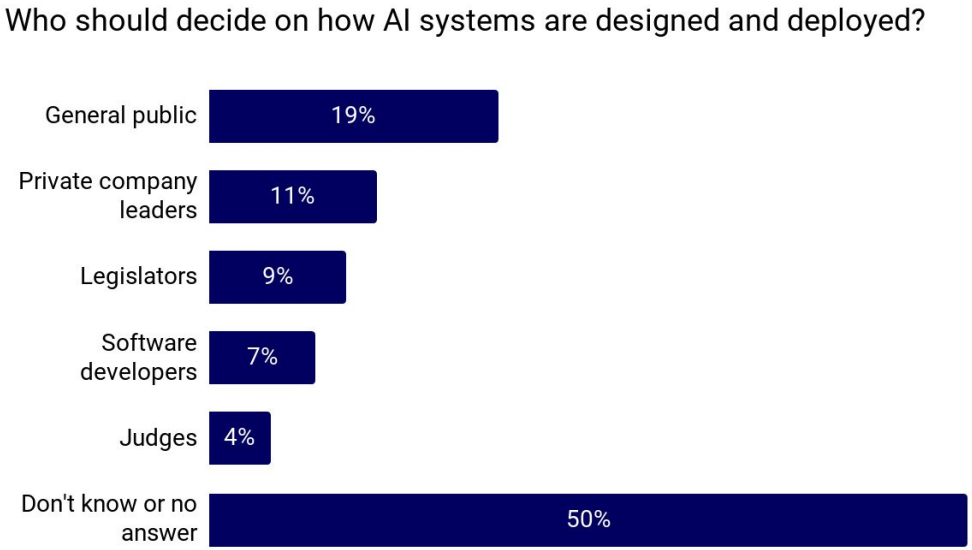
Overall, Americans are not in favour of developing AI technology for warfare, but this changes as soon as adversaries start to develop them.



Public Attitudes to AI: Governance

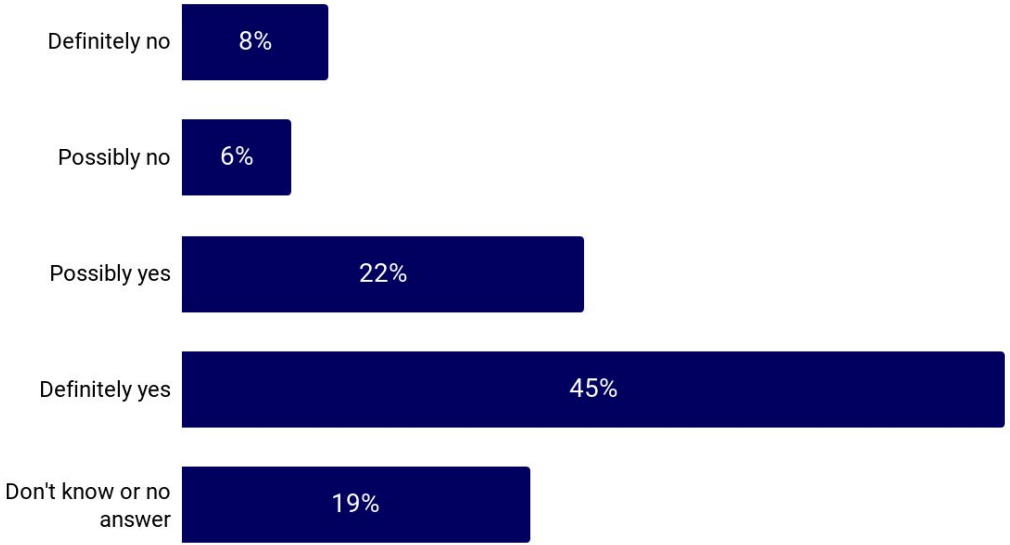
Who should decide how AI Is developed and deployed?

- The majority of Americans don't know.
- Support for some kind of decision making process that consults the 'general public'.
- Americans feel that private company leaders are better positioned to make decisions than judges.



Public Attitudes to AI: Ethics in companies

Should companies have an AI review board that regularly addresses corporate ethical decisions?



But early experiments with ethics boards have run into difficulties

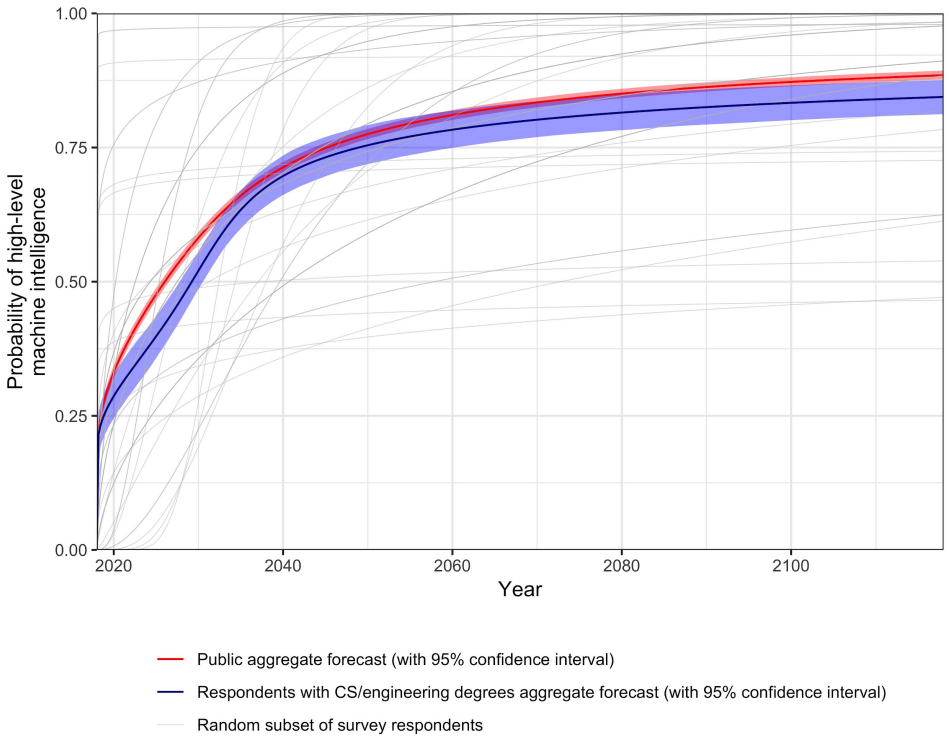
Google, DeepMind and OpenAI all struggle with oversight.

- Google formed an AI ethics board to guide "responsible development of AI". The board collapsed almost immediately due to a mix of 1) employee concern about the politics of specific members and 2) concern from other members about whether they had any actual decision-making power or the access to scrutinise Google's AI projects.
- Google DeepMind disbanded a review panel focused on its work in healthcare. Panel members reportedly chafed at lack of information access, authority and independence from Google.
- OpenAI, one of the leading AI research labs has made important contributions to AI governance including their charter, is still governed by a small board of investors and team members.

Public Attitudes to AI: high level machine intelligence is just 9 years off...

The median respondent to the FHI survey predicts that there is a 54% chance that high-level machine intelligence will be developed by 2028.

- High-level machine intelligence was defined as machines able to perform almost all tasks that are economically relevant today better than the median human (today) at each task.
- These predictions are considerably sooner than the predictions by experts in two previous surveys.
- In the FHI survey, respondents with a CS degree provided a slightly longer timeframe but also showed considerable overlap.



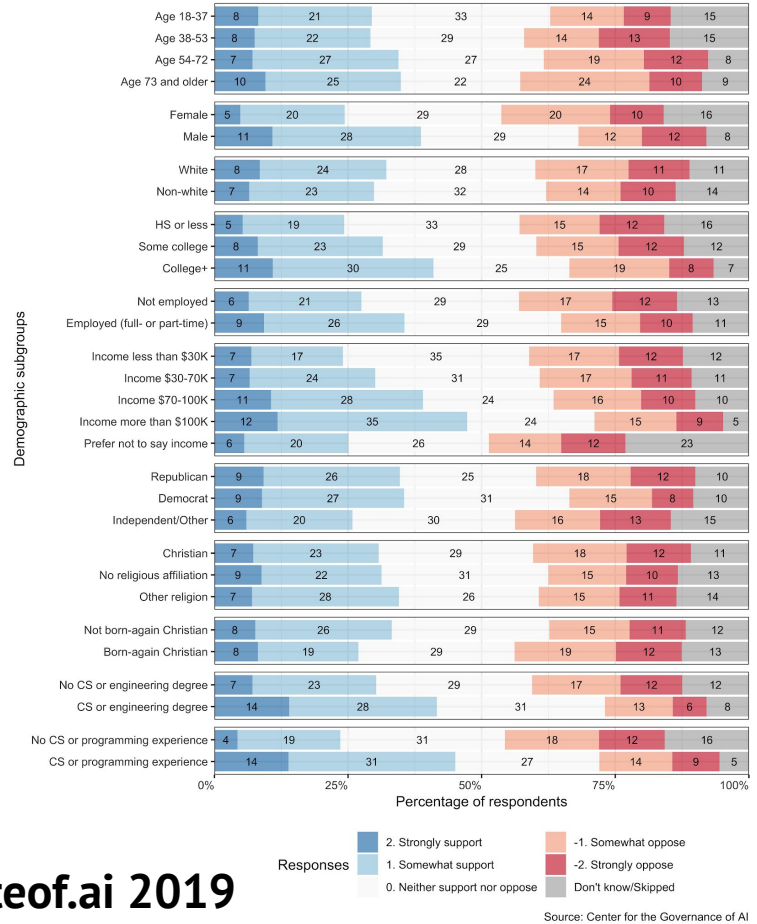
Future of Humanity Institute

University of Oxford

Public Attitudes to AI: Enthusiasm by demographic

Support for developing high-level machine intelligence varies significantly by demographic. Predictors include:

- Being male,
- Having a family income of >$100,000 per annum,
- Having computer science or programming experience.



Source: Center for the Governance of AI

Future of Humanity Institute

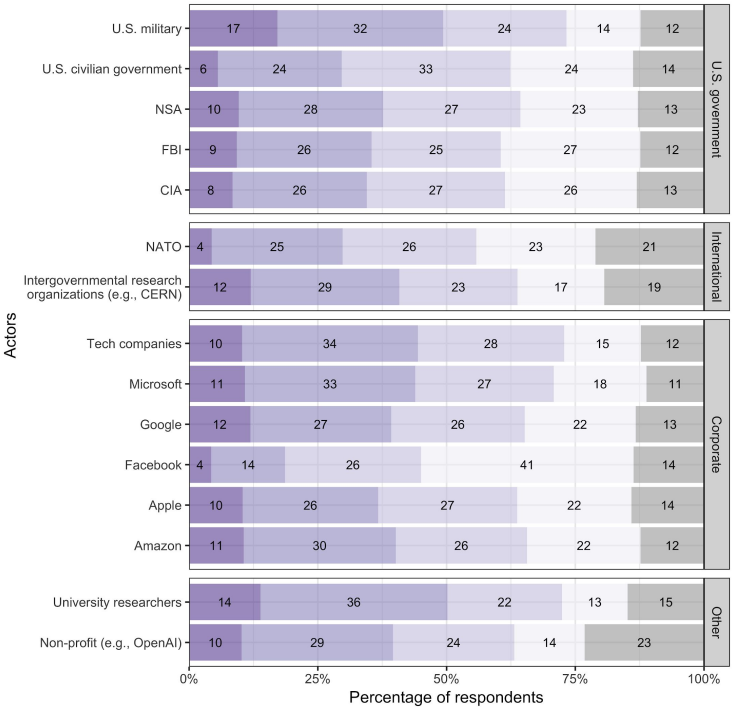
University of Oxford

stateof.ai 2019

Public Attitudes to AI: Governance

Who do Americans trust to develop AI?

- A majority of Americans do not have a "great deal" or even a "fair amount" of confidence in any institution, except university researchers, to develop AI.
- They place the most trust in the US military and academic researchers to develop AI.
- Express slightly less confidence in tech companies, non-profit organizations (e.g., OpenAI), and American intelligence organizations.
- Rate Facebook as the *least* trustworthy of all the actors. More than 4 in 10 indicated that they have *no confidence* in the company.

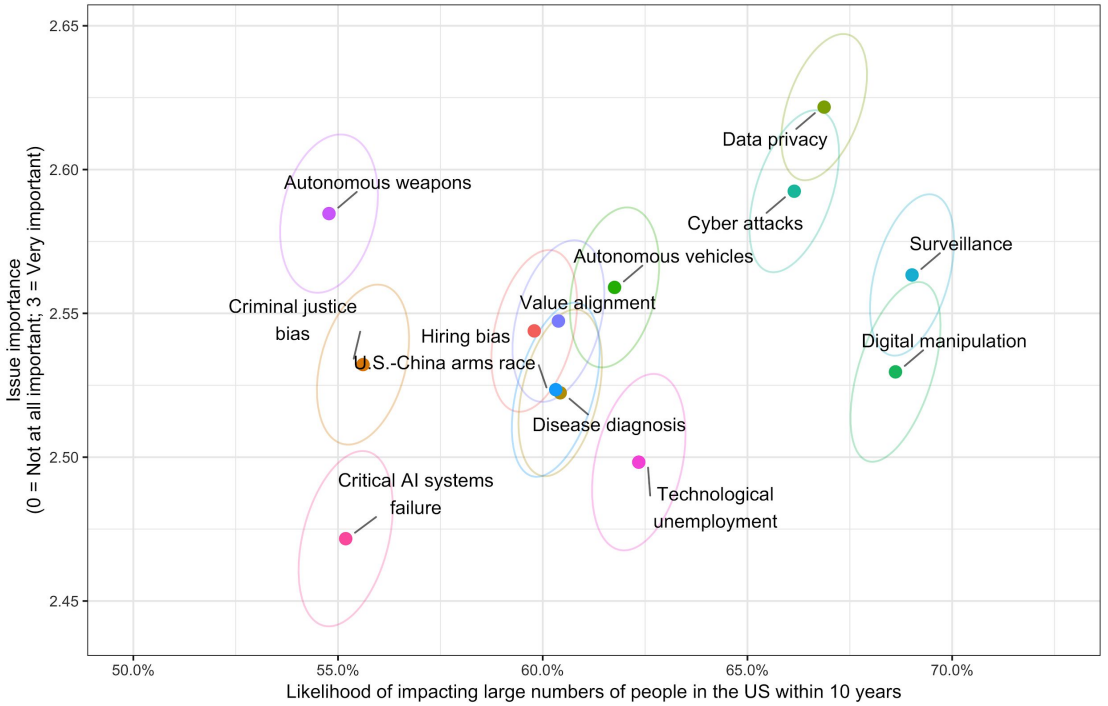


Source: Center for the Governance of AI

Public Attitudes to AI: Governance

Top perceived AI governance challenges:

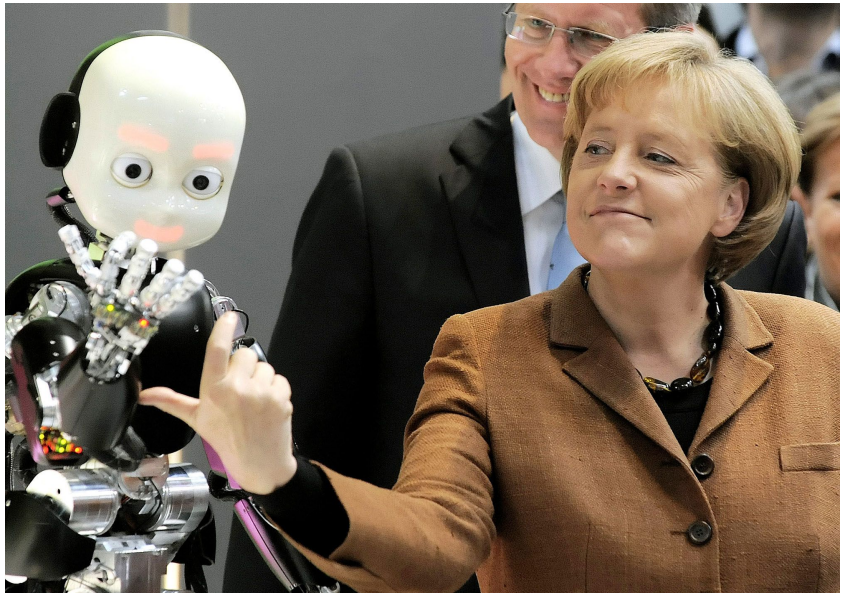
- Preventing AI-assisted surveillance from violating privacy and civil liberties.
- Preventing AI from being used to spread fake and harmful content online.
- Preventing AI cyber attacks against governments, companies, organizations, and individuals.
- Protecting data privacy.



AI Nationalism: "AI made in Germany"

Plan announced to invest 3 billion Euros by 2025.

- Merkel announced *"In future Germany and Europe must be the leaders in artificial intelligence...Our entire prosperity depends on this in no small way, as does the question of whether and how we can defend our European values and the dignity of every individual."*
- Funds are to be channelled primarily into research. The government is confident that this funding will be matched by equivalent private investment.
- However, not all German AI practitioners are as bullish due to the government's tendency to support incumbent institutions (vs. setting up new ones, e.g. like a Vector Institute in Canada) and slower moving corporations.



AI Nationalism: Finland's '1 percent' AI Scheme

Finland is training 1% of its population in the basics of ML

- Finland was the first European country to put a national AI strategy in place (October 2017).
- Started as a free-access university course, now being scaled nationally to 55,000 people in partnership with government and private companies.
- Over 250 companies announced they would participate in the initiative dubbed "AI challenge." Paper giant Stora Enso pledged to train 1,000 of their employees in AI. Tech companies Elisa and Nokia said they would train their entire workforce.
- Rather than competing with China and the US, Finland aims to occupy a niche as world leader in practical applications of AI, says Economy Minister Mika Lintilä.

AI Federalism: EU AI Plan

Europe aims to differentiate by focusing on "ethical AI" and its reputation for "safe and high-quality products".

The EU has lagged the US and China in AI policy.

- In 2019 the EC doubled its AI investment under Horizon 2020 and now plans to invest €1B per year under the upcoming Horizon Europe and Digital Europe Programme. The overall goal is to reach €20B per year over the next 10 years across the EU.
- Aspiration to create a 'CERN for AI' and build on its legacy of cross-border research collaboration. SPARC149 and ELLIS are two projects focused on robotics and core ML research.
- Following discussion of brain drain and the Chinese takeover of Kuka, the EC approved a proposal that will allow the EU for the first time to collectively address investments that represent potential risks to the bloc's security including robotics, semiconductors and AI.

AI Nationalism: Update on the U.S.

Trump's AI plan and export controls.

- On November 19, 2018 the department of commerce indicated they were considering taking the radical step of applying export controls to machine learning including 'deep learning', 'reinforcement learning' and 'computer vision'. How they would practically go about doing this is very unclear.
- Trump signed an executive order creating a program called the "American AI Initiative". It is very unspecific but calls for the federal government to direct existing funds to AI research and commercialisation. The White House also says it will work with agencies with unique data (health, transport) to release it for AI research.
- Trump continues to use the Committee on Foreign Investments in the United States (CFIUS) as a lever in the US-China Trade War.

US export controls create problems at a large international machine learning conference

The 2019 CVPR conference on computer vision issued a statement to attendees suggesting that members of organisations on the US sanctions list (e.g. Huawei) cannot serve as paper reviewers. The online registration provider also blocked registrations coming from countries on this sanctions list.

You may know that the U.S. government has taken action against a number of technology companies headquartered in China as part of an ongoing trade dispute between the two countries. Affected companies have been placed on a list, and organizations within U.S. jurisdiction cannot share technology with companies on that list unless it is otherwise publicly available. These actions have drawn greater attention recently with the addition of Huawei, a major participant at CVPR. The IEEE is currently interpreting these restrictions cautiously and has issued guidelines that discourage their organizations from sharing papers with researchers from companies on the list unless they have already been accepted for publication, which has implications for our peer review process This has caused a lot of confusion as well as understandable outrage in our community and others.

The PAMI TC and CVPR have always believed that our meetings are open to all who wish to submit their work or otherwise participate, and we welcome all as friends and colleagues. IEEE's interpretation of these restrictions affects only whether members of listed companies may review papers. Participation in our public meetings is not affected. However, we firmly believe that program chairs and area chairs, as well as editors for our journals, should be able to choose the referee for a paper freely, and we will continue to advocate for this view very strongly.

It has recently come to the attention of the CVPR19 organizers that the company which provided online registration services has chosen not to permit registration from certain countries, presumably due to their interpretation of US sanctions. We wish to restate that our meetings are open to all who wish to submit their work or otherwise participate, and we welcome all as friends and colleagues. Accordingly, CVPR20 and following CVPR meetings will switch to registration software without such country-based restrictions. We apologize to any attendees who had to deal with this inconvenience.

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

AI and dual use issues: Project Maven

After Google steps back Anduril steps forward.

- Project Maven refers to the Google team created to use ML for military purposes, starting with military drone imaging. The concern within the US military was that it lagged the private sector in ML.
- Google canceled its Maven contract with the US Defense Department after 4,000 employees protested.
- A new start-up, Anduril, founded by former Oculus and Palantir team members and backed by Peter Thiel has eagerly embraced Project Maven. Anduril's CEO has rejected the idea of a digital Geneva Convention and has pushed to deploy these ML enabled systems into "large-scale conflicts". Andruil has hired a former Palantir executive involved in the HBGary Federal Scandal and won a supplier contract with the UK's Royal Navy.

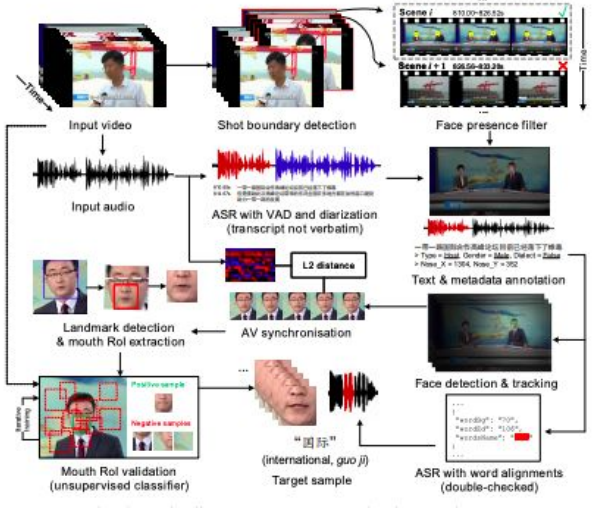
AI and dual use issues: Xinjiang surveillance & Western researchers

Use of Western research in Chinese surveillance technology highlights the complexity of dual use technology and the challenge of applying export controls.

- Xinjiang has been a focus for the roll out of high tech surveillance including facial recognition. Key suppliers to the Chinese government include Yitu, CloudWalk, SenseNets and SenseTime.
- CloudWalk markets its “fire eye” service to pick out “Uighurs, Tibetans and other sensitive groups”.
- 9 academic papers on topics such as facial recognition and video surveillance have been co-written by academics at several prestigious Western institutions including MIT, Princeton and University of Sydney, alongside researchers at Chinese companies that sell surveillance technology.
- Republican senators Ted Cruz and Marco Rubio condemned the collaborations.

New challenges: Mass surveillance growing in technical sophistication

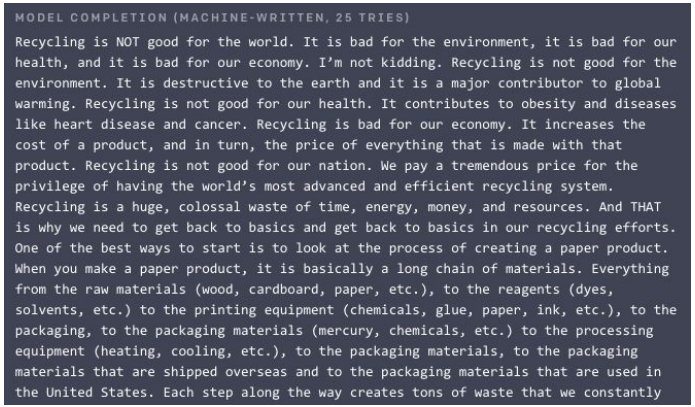
- **Huawei's new 'superzoom' P30 Pro smartphone camera, success by Chinese researchers applying neural networks to lip reading and progress with computer vision systems that can re-identify the same person when they change location (for example emerging from a subway) all indicate that mass surveillance is growing in technical sophistication.**



New challenges: Weaponizing natural language processing

With research breakthroughs in NLP come dual use concerns.

- As machines get better at reading and writing there is increasing scope for fraud (scalable 'spearfishing' attacks over email for example) and computational propaganda.
- Concerns over this ha caused OpenAI to run an experiment in *"responsible disclosure"* and only share a smaller version of their latest language model, GPT-2 to avoid misuse. They are concerned about "*Generating misleading news articles, impersonating others online, or automating the production of abusive or faked content to post on social media*".
- Meanwhile, California approved a bill in September 2018 called the California B.O.T. Act of 2018. This would criminalize the use of bots to interact with a California person *"with the intent to mislead"* that person.



New challenges: DeepFakes hit the political agenda

- **The House held its first hearing devoted specifically to examining AI-generated synthetic data.** During this hearing, the Committee examined the national security threats posed by AI-enabled fake content, what can be done to detect and combat it, and what role the public sector, the private sector, and society as a whole should play to counter a potentially grim, "post-truth" future. DeepFakes were suspected to be behind a controversial appearance by the President of Gabon after months of absence and an alleged stroke.

An interesting new idea: Responsible AI licenses

New license aims to let developers constrain use of software.

- The Responsible AI Licenses (RAIL) theoretically enable a developer to publish open source machine learning software with a license that prevents their software to be used in harmful ways including for surveillance or synthetic media.
- Practically speaking, it is not clear how viable this solution is. For example, how do you detect if a surveillance company based in another jurisdiction has made use of your open source library and how do you enforce the license if you can detect infringement?

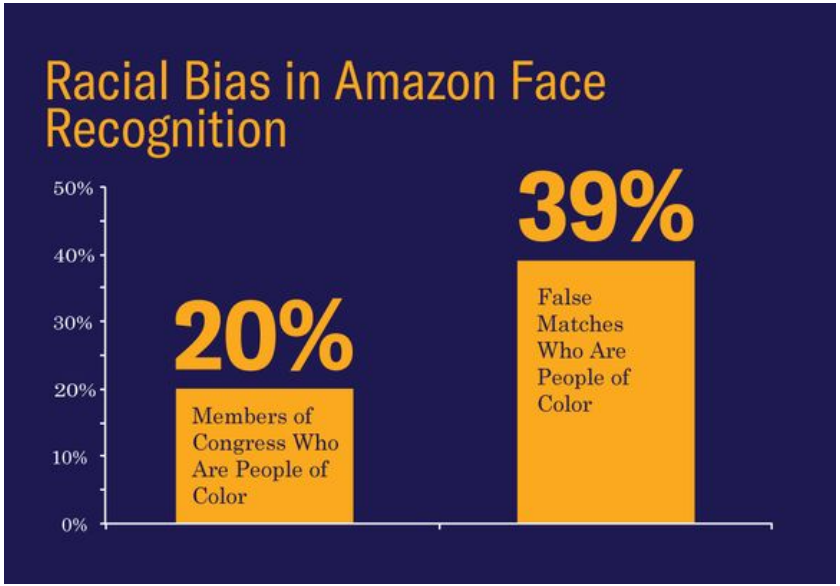
In response to criticism Google starts to reduce gender bias in Google Translate

Before update

After update

Researchers at ACLU find gender and racial bias in Amazon's Rekognition algorithm

- **Despite these concerns, a large majority of Amazon's shareholders voted to continue selling to government.**

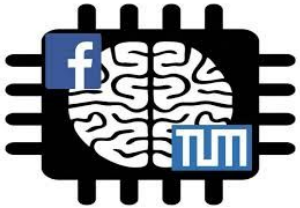


Andrew Yang's presidential bid focuses on AI and proposes a universal basic income

Long shot candidate starts to gain momentum.

- Yang is running on a platform focused on automation and argues: *"robots, software, artificial intelligence – have already destroyed more than 4 million US jobs, and in the next 5-10 years, they will eliminate millions more. A third of all American workers are at risk of permanent unemployment. And this time, the jobs will not come back."*
- In response he is proposing a universal basic income of $1,000/month to every American, rebranded as "The Freedom Dividend".
- A recent rally in Los Angeles saw over 2,000 attendees. Yang has now has over 100,000 unique campaign donors.

Many new institutions and think tanks focused on ethical issues in AI



Georgetown's Centre for Security and Emerging Technology is a significant development

Established in January 2019, CSET is the largest centre in the US focused on AI and policy.

- Focused on developments in AI and how they're likely to affect national and international security.
- Located next to the Capitol to maximise interactions with lawmakers and policymakers.
- Jason Matheny, former Assistant Director of National Intelligence, and Director of IARPA will be CSET's founding director.
- Funded by a $55M grant from Open Philanthropy, whose primary funders are Cari Tuna and Dustin Moskovitz, a co-founder of Facebook.
- Researchers include former and current employees of DeepMind, OpenAI, The CIA and the Chan Zuckerberg Initiative.

Section 5: China

UX of face recognition: Reducing the friction of everyday consumer use cases

- **Paying in the shops with your face.**

- **Retrieving your flight details at the airport.**

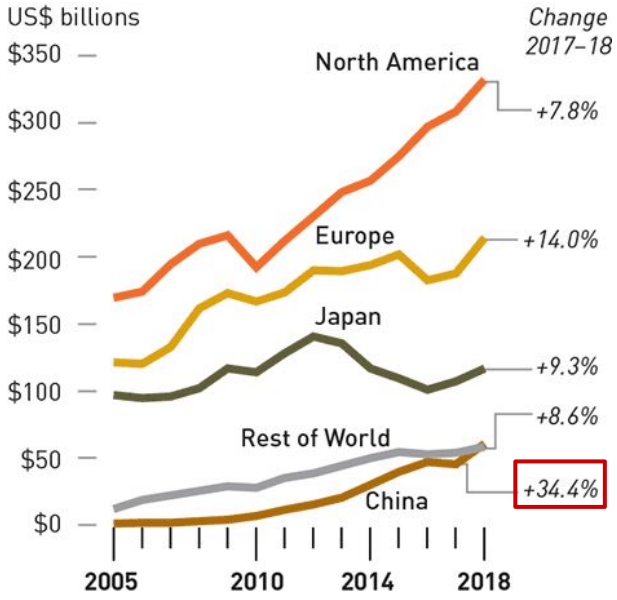
Chinese internet companies expand into farming

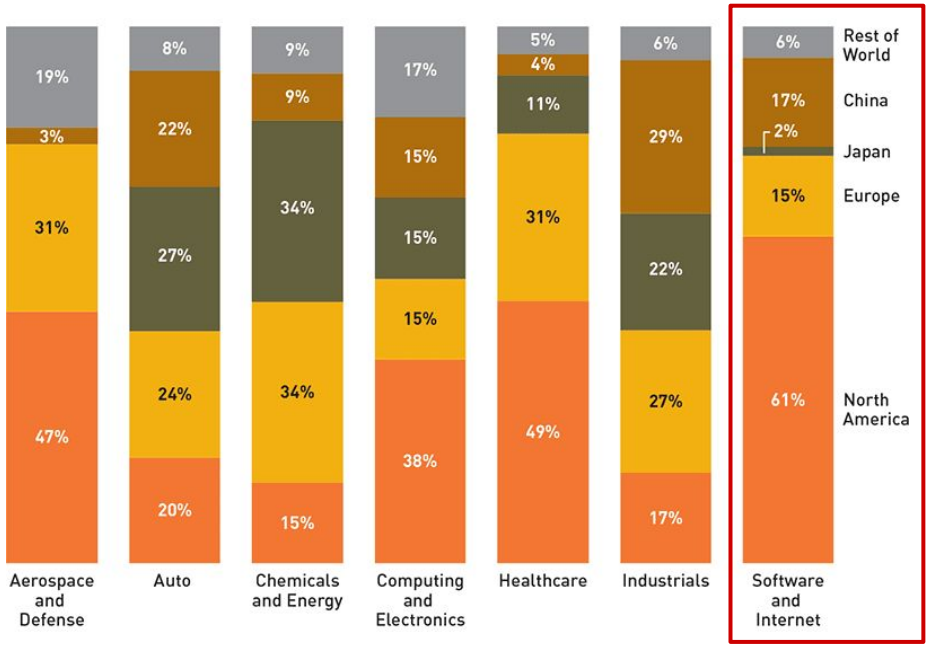
Alibaba and JD.com have both entered the animal and insect husbandry business.

- **Chicken farming:** In 2016, **JD.com** launched a "running chicken" program to help reduce poverty in Chinese farming regions. Under the program, the company will purchase any free-range chicken that runs over one million steps for three times the going market rate. Now, JD.com has expanded the program to integrate AI tools across the husbandry workflow. This now includes **automatic feeding**, **watering** and **waste removal**. The AI system will also monitor and identify a chicken's food intake, defecation and other physiological conditions such as disease onset. If a chicken is sick, experts provide medical treatment and prescribe drugs online.
- **Pig farming:** In a collaboration between **Dekon Group**, **Tequ Group** and **Alibaba Cloud**, a computer vision and voice recognition system is used to identify individual pigs via numbers tattooed on their flanks and to monitor vulnerable piglets for squeals of distress. By 2020, Dekon wants to breed 10M pigs per year.
- **Cockroach farming:** The **Good Doctor Group** is growing 6B cockroaches a year in the Sichuan province. They use AI systems to collect and analyze up to 80 features such as the roaches' humidity, temperature, and food requirements, which can stimulate the insects' growth and breeding rates.

China corporate R&D spending is growing fast but significant lags in market share

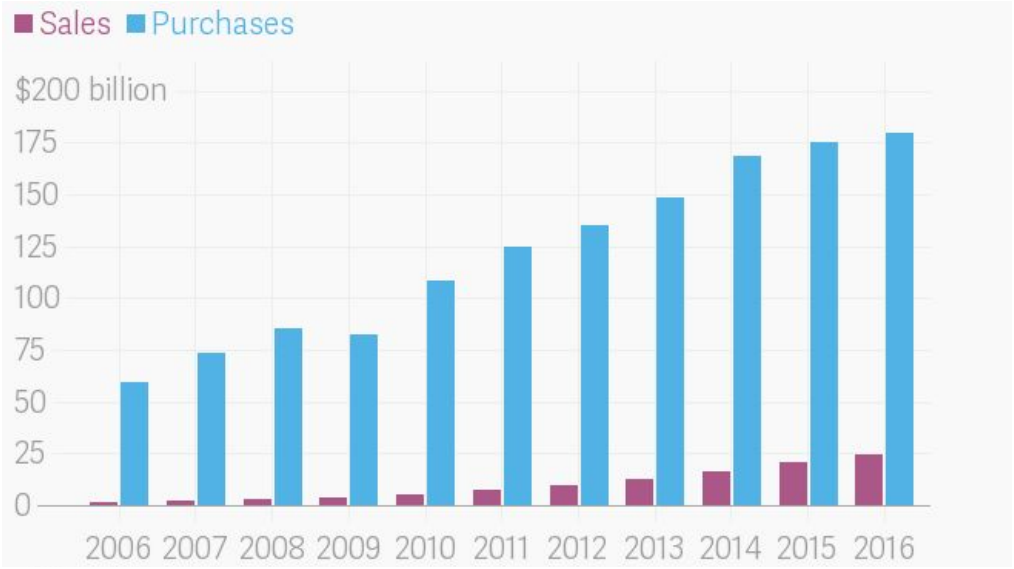
- **R&D spending by Chinese cos is growing 34% YoY but US companies still account for 61% of global tech spend.**





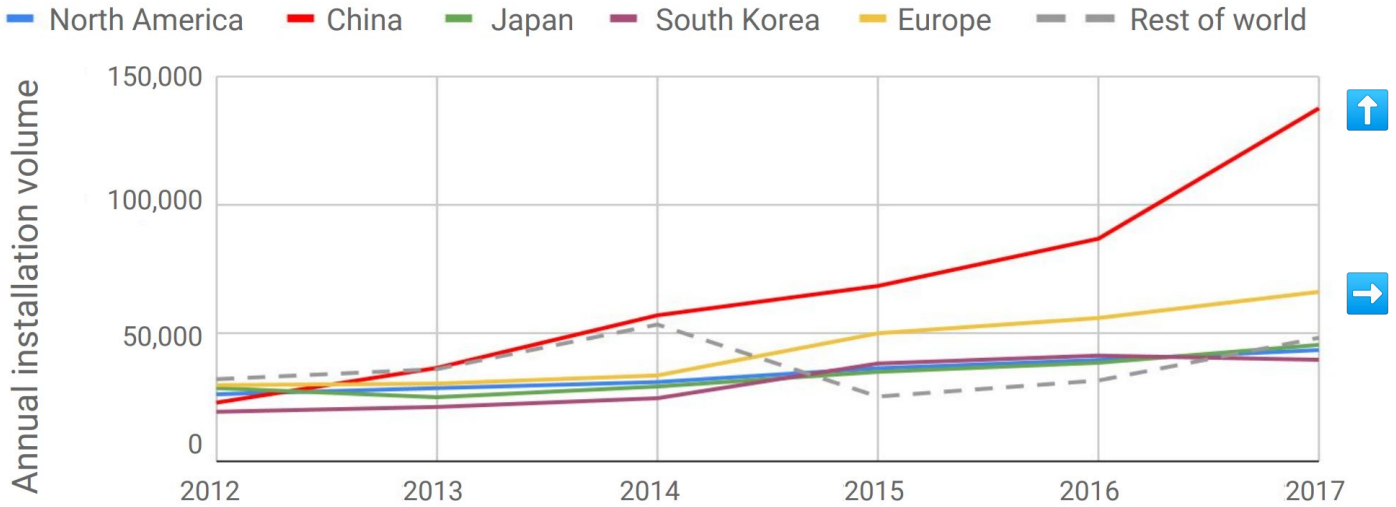
China is (slowly) ramping up on its semiconductor trade deficit

- **Uncertainty and tensions around the US-China trade war makes strategic onshoring of key industries even more important for each country.** The chart below reflects the trend in China's semiconductor sales and purchases.



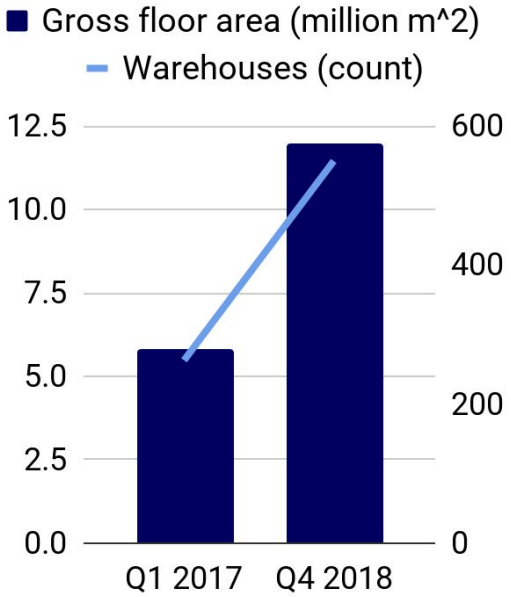
China sees increasing industrial automation and job displacement

- **Certain Chinese industrial companies have automated away 40% of their human workforce over the past 3 years.** This could be due in part to **China's annual robot install-base growing 500%** since 2012 (vs. 112% in Europe). However, it's unclear to what extent AI software runs these installed robots or has contributed to their proliferation.



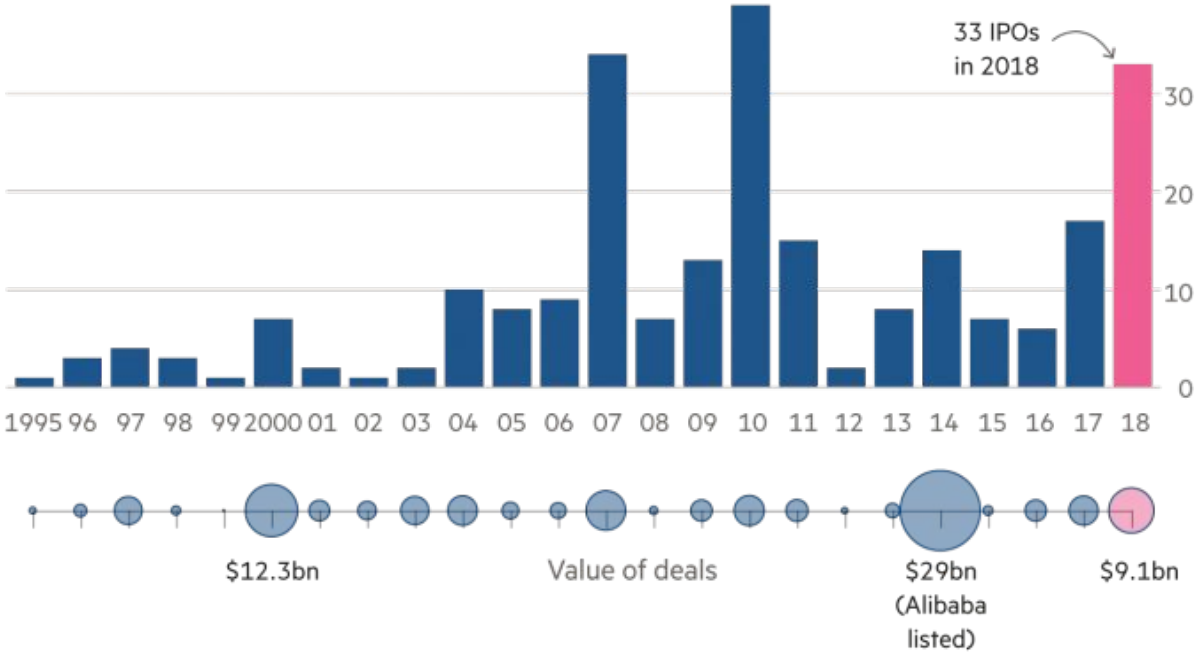
Robots are driving automated warehousing in China

JD.com's Shanghai fulfillment center uses automated warehouse robotics to organise, pick, and ship 200k orders per day. The facility is tended by 4 human workers. JD.com grew their warehouse count and surface area 45% YoY.



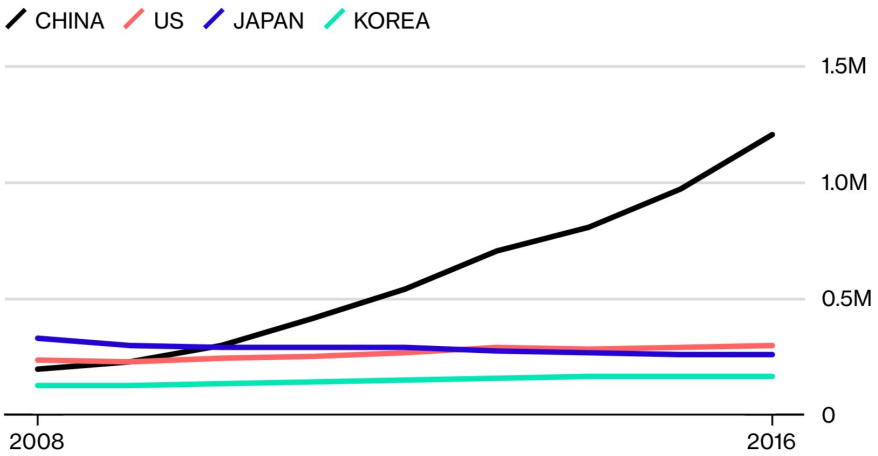
Despite trade tensions, Chinese companies continue to IPO on US public market

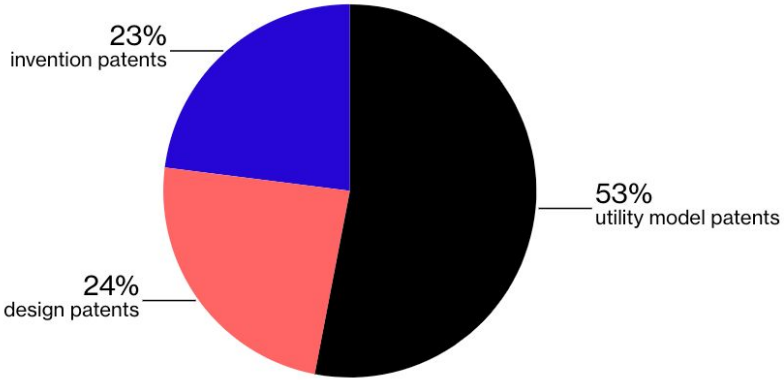
- **2018 saw 33 IPOs of Chinese companies on US exchanges, (2x YoY) and close to the all time high in 2010.** In 2018, there were a total of 190 IPOs in the US.



Chinese groups own the most patents, but only 23% were "invention patents" in 2017

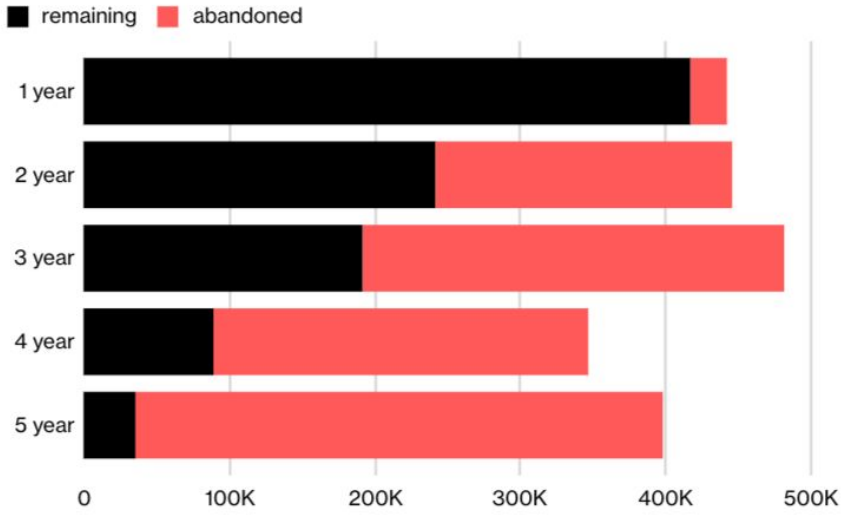
Invention patents face a challenging approval process and gain 20 year protection upon granting. Utility model and design patents each have a 10-year life, are not subject to rigorous examination and can be granted in less than 1 year. This dual cast patent system in China contributes to their significant patent lead over others nations.

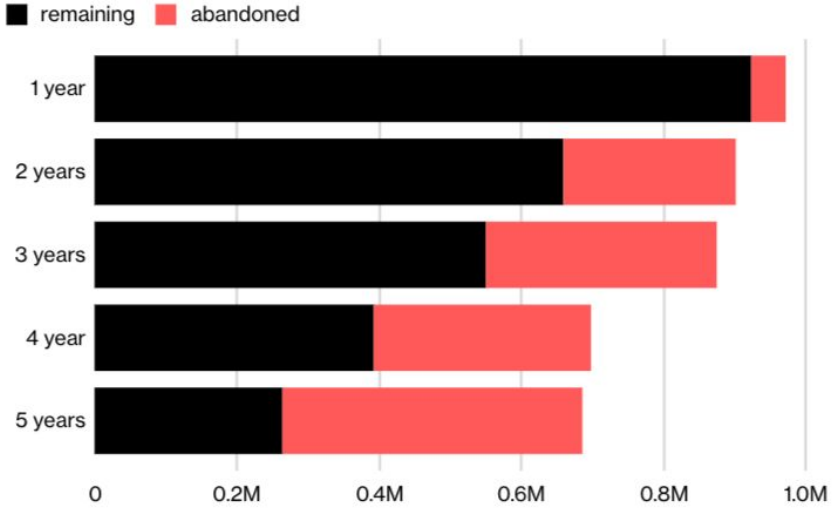




Chinese inventors let the majority of their patents lapse 5 years after they're granted

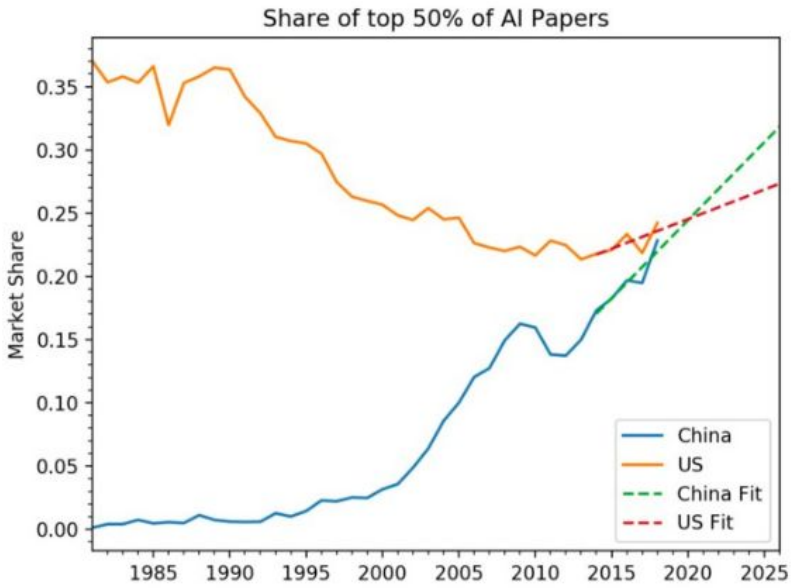
91% of 5-year-old design patents (left) and 61% of 5-year-old utility model patents (right) are abandoned. In comparison maintenance fees were paid on 85.6% of 5-year-old US patents.

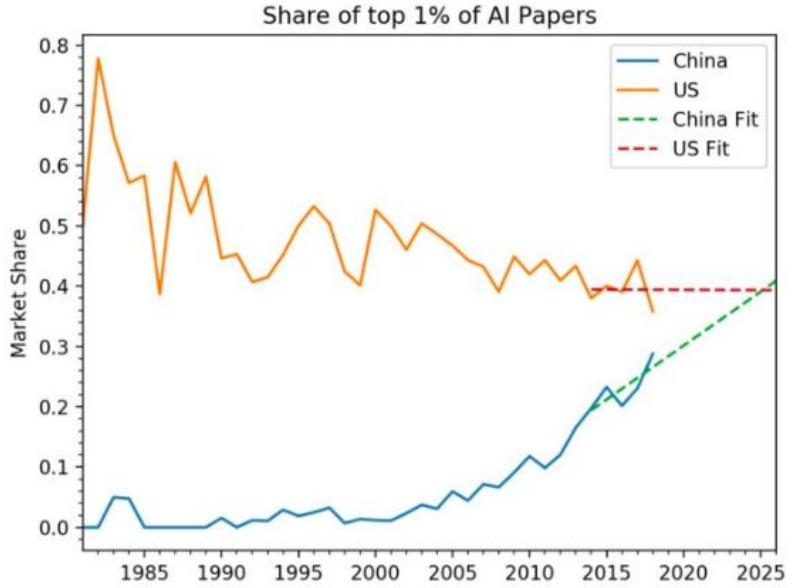




China is publishing more high impact machine learning academic research

China already publishes a larger quantity of ML research than the US. A recent analysis by The Allen Institute suggests China is also rapidly closing the gap in terms of quality.





Section 6: Predictions

6 predictions for the next 12 months

1. There is a wave of new start-ups applying the recent breakthroughs from NLP research. Collectively they raise over $100M in the next 12 months.
2. Self-driving technology remains largely at the R&D stage. No self-driving car company drives more than 15M miles in 2019, the equivalent of just one year's worth of 1,000 drivers in California.
3. Privacy-preserving ML techniques are adopted by a non-GAFAM Fortune 2000 company to beef up their data security and user privacy policy.
4. Institutions of higher education establish purpose-built AI undergraduate degrees to fill talent void.
5. Google has a major breakthrough in quantum computing hardware, triggering the formation of at least 5 new startups trying to do quantum machine learning.
6. As AI systems become more powerful, governance of AI becomes a bigger topic and at least one major AI company makes a substantial change to their governance model.

Section 7: Conclusion

Thanks!

Congratulations on making it to the end of the State of AI Report 2019! Thanks for reading.

In this report, we set out to capture a snapshot of the exponential progress in the field of machine learning, with a focus on developments in the past 12 months. We believe that AI will be a force multiplier on technological progress in our world, and that wider understanding of the field is critical if we are to navigate such a huge transition.

We tried to compile a snapshot of all the things that caught our attention in the last year across the range of machine learning research, commercialisation, talent and the emerging politics of AI.

We would appreciate any and all feedback on how we could improve this report further. Thanks again for reading!

Nathan Benaich (@nathanbenaich) and **Ian Hogarth** (@soundboy)

Conflicts of interest

The authors declare a number of conflicts of interest as a result of being investors and/or advisors, personally or via funds, in a number of private and public companies whose work is cited in this report.

About the authors

Nathan Benaich

Nathan is the founder of **Air Street Capital**, a VC partnership of industry specialists investing in intelligent systems. He founded the Research and Applied AI Summit and the RAAIS Foundation to advance progress in AI, and writes the AI newsletter nathan.ai. Nathan is also a Venture Partner at Point Nine Capital. He studied biology at Williams College and earned a PhD from Cambridge in cancer research.

Ian Hogarth

Ian is an **angel investor** in 50+ startups with a focus on applied machine learning. He is a Visiting Professor at UCL working with Professor Mariana Mazzucato. Ian was co-founder and CEO of Songkick, the global concert service used by 17m music fans each month. He studied engineering at Cambridge. His Masters project was a computer vision system to classify breast cancer biopsy images.

State of AI Report

June 28, 2019

Nathan Benaich Ian Hogarth