



The latest news from Google AI

Google Research: Looking Back at 2019, and Forward to 2020 and Beyond

Thursday, January 9, 2020

Posted by Jeff Dean, Senior Fellow and SVP of Google Research and Health, on behalf of the entire Google Research community

The goal of [Google Research](#) is to work on long-term, ambitious problems, with an emphasis on solving ones that will dramatically help people throughout their daily lives. In pursuit of that goal in 2019, we made advances in a broad set of fundamental research areas, applied our research to new and emerging areas such as healthcare and robotics, open sourced a wide variety of code and continued collaborations with Google product teams to build tools and services that are dramatically more helpful for our users.

As we start 2020, it's useful to take a step back and assess the research work we've done over the past year, and also to look forward to what sorts of problems we want to tackle in the upcoming years. In that spirit, this blog post is a survey of some of the research-focused work done by Google researchers and engineers during 2019 (in the spirit of similar reviews for [2018](#), and more narrowly focused reviews of some work in [2017](#) and [2016](#)). For a more comprehensive look, please see our [research publications in 2019](#).

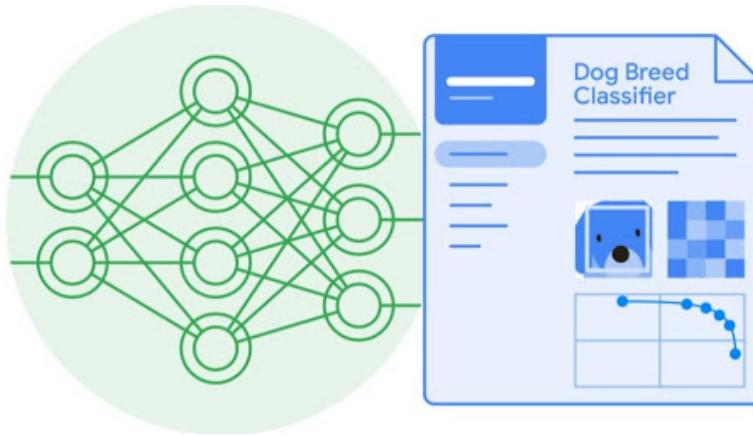
Ethical Use of AI

In 2018, we published a set of [AI Principles](#) that provide a framework by which we evaluate our own research and applications of technologies such as machine learning in our products. In June 2019, we published a [one-year update](#) about how these principles are being put into practice in many different aspects of our research and product development life cycles. Since many of the areas touched on by the principles are active areas of research in the broader AI and machine learning research community (such as bias, safety, fairness, accountability, transparency and privacy in machine learning systems), our goals are to apply the best currently-known techniques in these areas to our work, and also to do research to continue to advance the state of the art in these important areas.

For example, last year we:

- Published a [research paper about a new transparency tool](#), which enabled the

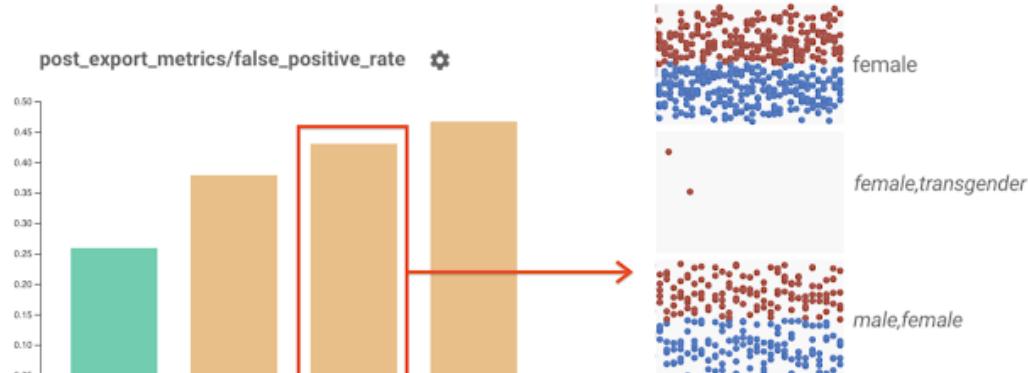
Launch of Model Cards for several of our Cloud AI products. You can see an



- Showed how [Activation Atlases](#) can help explore neural network behavior and can aid with interpretability of machine learning models.

Tour of an Activation Atlas

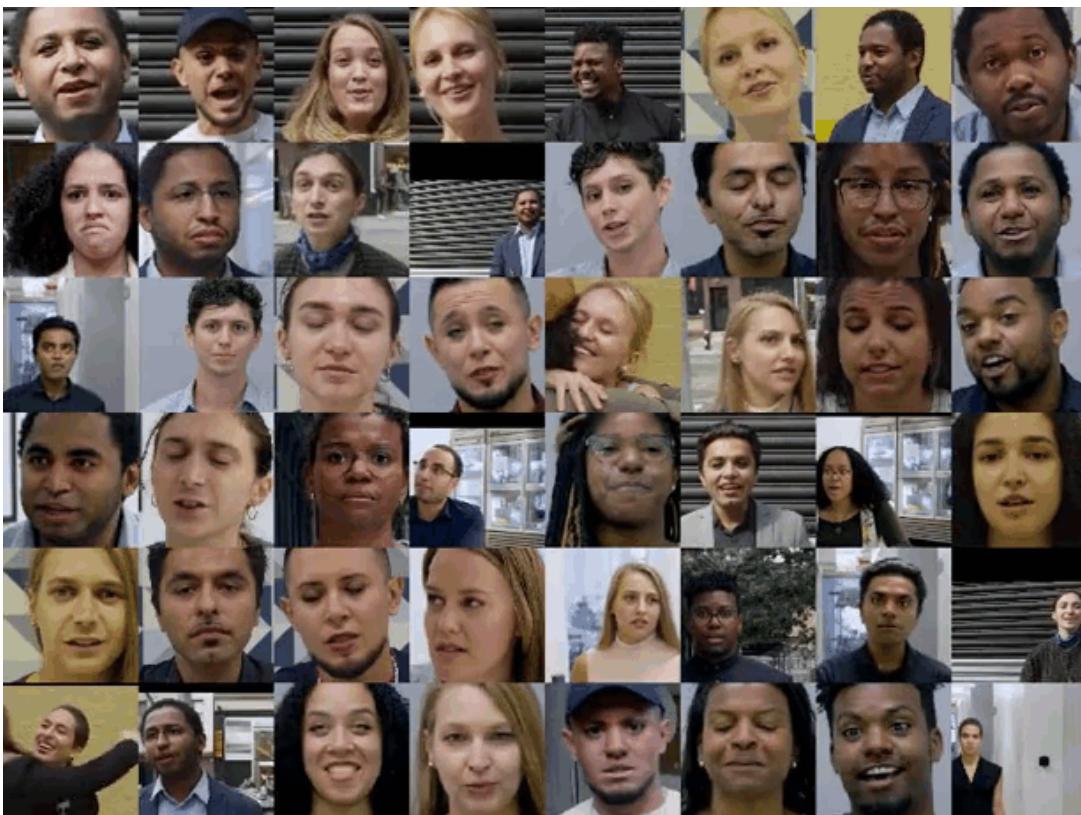
- Introduced [TensorFlow Privacy](#), an open-source library to enable training machine learning models with [differential privacy](#) guarantees.
- Released a beta version of [Fairness Indicators](#), to help ML practitioners identify unjust or unintended impacts of machine learning models.





Clicking on a slice in Fairness Indicators will load all the data points in that slice inside the What-If Tool widget. In this case, all data points with the "female" label are shown.

- Published a [KDD'19 paper](#) on how pairwise comparisons and regularization is incorporated into a large-scale production recommender system to improve ML Fairness.
- Published an [AIES'19 paper](#) about a case study on the application of fairness in machine learning research to a production classification system, and described our fairness metric, conditional equality, that takes into account distributional differences in implementing equality of opportunity.
- Published an [AIES'19 paper](#) about counterfactual fairness in text classification problems that asks the question: "How would the prediction change if the sensitive attribute referenced in the example were different?" and used this approach to improve our production systems that assess the toxicity of online content.
- Released a [new dataset to help with research to identify deepfakes](#).



A sample of videos from Google's contribution to the FaceForensics benchmark. To generate these, pairs of actors were selected randomly, and deep neural networks swapped the face of one actor onto the head of another.

AI for Social Good

There is enormous potential for machine learning to help with many important societal issues. We have been doing work in several such areas, as well as working to enable others to apply their creativity and skills to solving such problems. [Floods are the most common and the most](#)

deadly natural disaster on the planet, affecting approximately 250 million people each year. We have been using machine learning, computation and better sources of data to make significantly more accurate flood forecasts, and then to deliver actionable alerts to the phones of millions of people in the affected regions. We also hosted a workshop that brought together researchers with expertise in flood forecasting, hydrology and machine learning from Google and the broader research community to discuss ways to collaborate further on this important problem.



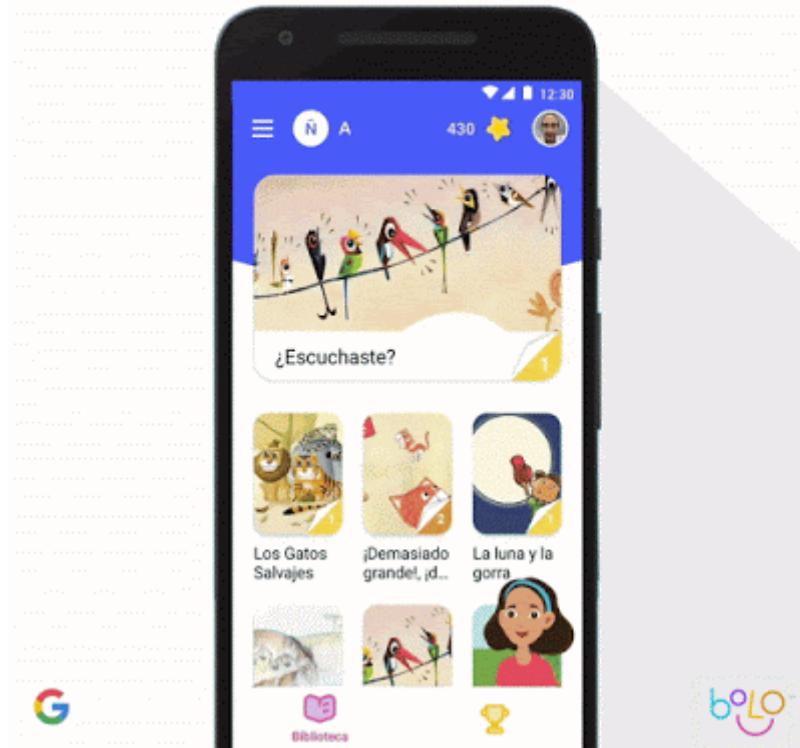
In addition to our flood forecasting efforts, we've been developing techniques to better understand the world's wildlife, collaborating with seven wildlife conservation organizations to use machine learning to help analyze [wildlife camera data](#) and collaborating with the U.S. NOAA to identify whale species and locations from sounds in underwater recordings. We've also created and released a set of tools for enabling new kinds of machine-learning-oriented biodiversity research. As part of helping to organize the [6th Fine-Grained Visual Categorization Workshop](#), Google researchers in our Accra, Ghana office collaborated with researchers at [Makerere University AI & Data Science research group](#) to create and run a Kaggle competition on the classification of [cassava plant diseases](#). As cassava is the second largest source of carbohydrates in Africa, plant health is an important food security issue, and it was great to see more than 100 participants across 87 teams participate in the contest.

In 2019 we updated [Google Earth Timelapse](#), enabling people to effectively and intuitively visualize how the planet has changed over the past 35 years. Further, we've been collaborating with academic researchers on new privacy-preserving ways to [aggregate data on human mobility](#), to give urban planners better information about how to design efficient environments with lower levels of carbon emissions.

Google Earth Timelapse: Urban Expansion A...



We've also applied machine learning to support childhood learning. According to the United Nations, [617 million children do not have basic literacy](#), a critical determinant of their quality of life. To help more children learn to read, our [Bolo app uses speech-recognition technology](#) that tutors students in real-time. And to increase access, the app works completely offline on low-cost phones. In India, Bolo has already [helped 800,000 children read stories](#) and speak half a billion words. Early results are encouraging; a [three-month pilot](#) among 200 villages in India showed an improvement in reading proficiency among 64% of pilot participants.



For older students, the [Socratic app](#) can [help high schoolers](#) with complex problems in math, physics and over 1,000 higher education topics. Based on a photo or verbal question, the app automatically identifies the question's underlying concepts and links to the most helpful online resources. Like the [Socratic method](#), the app doesn't directly answer questions, but instead leads students to discover the answer themselves. We're excited about the broad possibilities of improving educational outcomes around the world through things like Bolo and Socratic.

To expand the reach of our AI for Social Good efforts, in May we [announced](#) the grantees of our

AI Impact Challenge with \$25 million in grants from Google.org. The response was huge: we received over 2,600 thoughtful proposals from 119 countries. Twenty impressive organizations stood out for their potential to solve big social and environmental problems and were our initial set of grantees. A few examples of the work of these organizations:

- The [Fondation Médecins Sans Frontières \(MSF\)](#) is [creating a free smartphone application that uses image recognition tools to help clinical staff](#) in low-resource settings (currently being piloted in Jordan) to analyze anti-microbial images and advise on the appropriate antibiotics to use for a particular patient's infection.
- Over a billion people live in smallholder farm households. A single pest attack can [devastate their crop yields](#) and livelihoods. [Wadhwani AI](#) uses image classification models that can identify pests and provide timely advice on what pesticides to spray and when—ultimately improving crop yield.
- And deep in tropical rainforests, where illegal deforestation is a major driver of climate change, [Rainforest Connection](#) uses deep learning for bioacoustic monitoring and old cell phones to track rainforest health and detect threats.

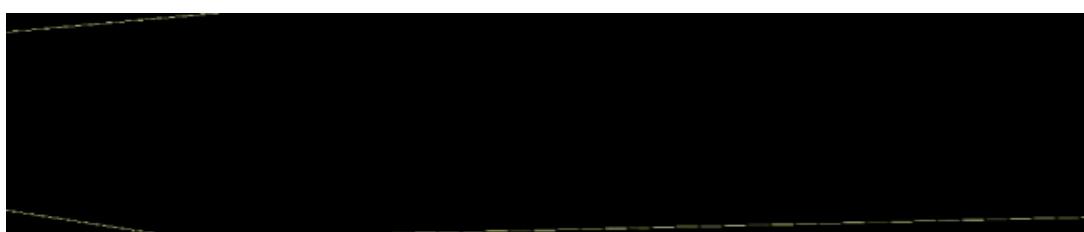


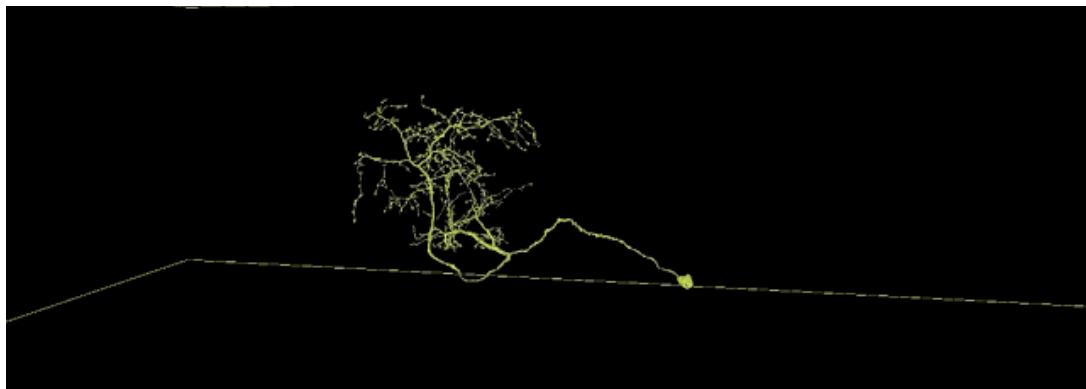
Our 20 AI Impact Challenge winners. You can learn more about the work of all the grantees [here](#).

Applications of AI to Other Fields

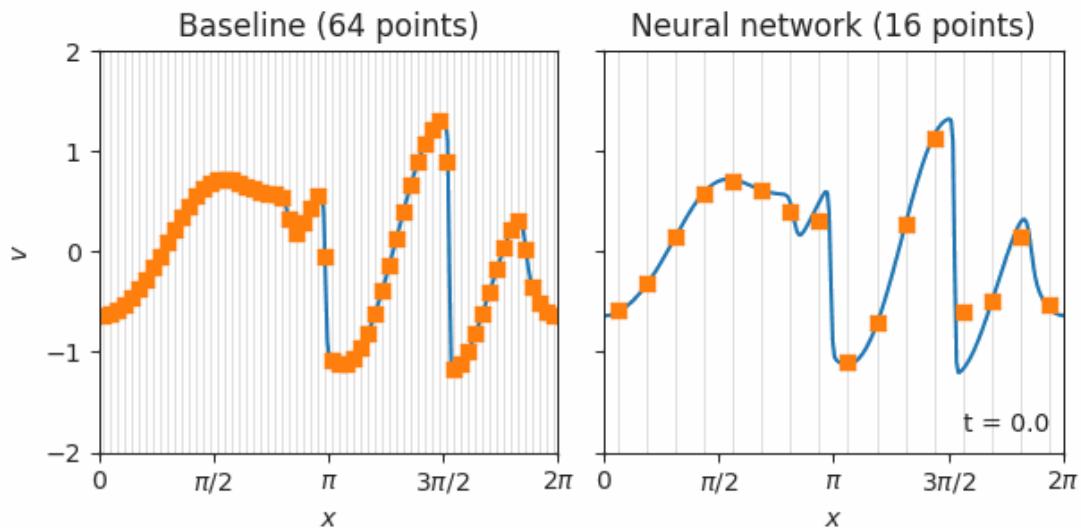
The application of computer science and machine learning to other scientific fields is an area that we are especially excited about and [have published a number of papers in](#), often in multi-organization collaborations. Some highlights from this year include:

- In [An Interactive, Automated 3D Reconstruction of a Fly Brain](#), we reported on a collaborative effort that achieved a milestone of mapping the structure of an entire fly brain, using machine learning models that were able to painstakingly trace each individual neuron.



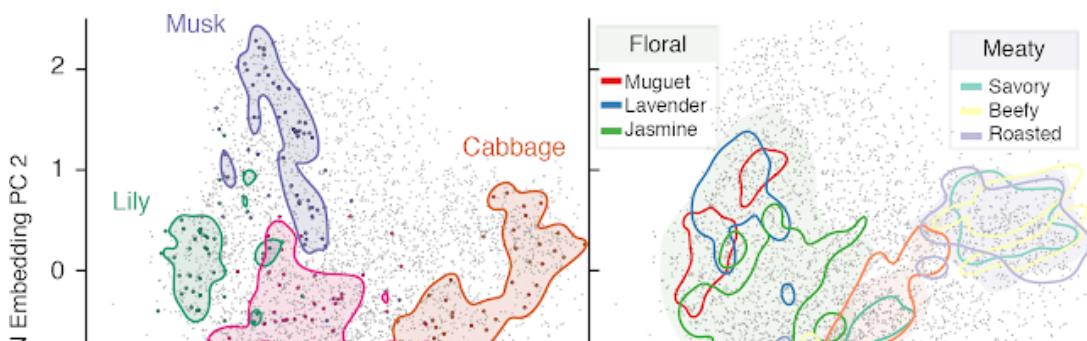


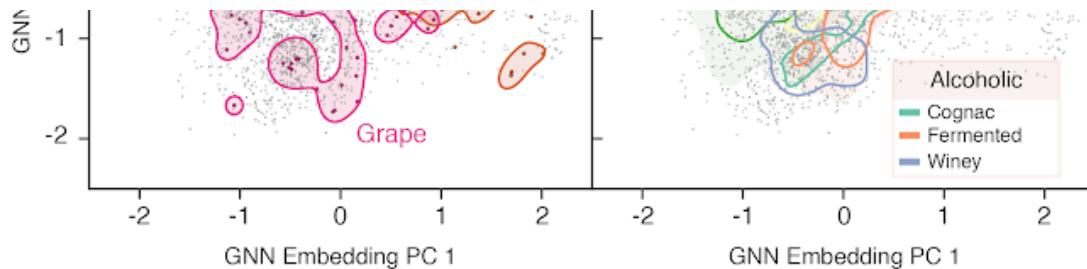
- In [Learning Better Simulation Methods for Partial Differential Equations](#) (PDEs), we showed how machine learning can be used to accelerate PDE computations, which are at the heart of many fundamental computational problems in climate science, fluid dynamics, electromagnetism, heat conduction and general relativity.



Simulations of [Burgers' equation](#), a model for shock waves in fluids, solved with either a standard finite volume method (left) or our neural network based method (right). The orange squares represent simulations with each method on low resolution grids. These points are fed back into the model at each time step, which then predicts how they should change. Blue lines show the exact simulations used for training. The neural network solution is much better, even on a 4x coarser grid, as indicated by the orange squares smoothly tracing the blue line.

- We gave machine learning models better scents of the world with [Learning to Smell: Using Deep Learning to Predict the Olfactory Properties of Molecules](#). We showed how to leverage graph neural networks (GNNs) to directly predict the odor descriptors for individual molecules, without using any handcrafted rules.





2D snapshot of our embedding space with some example odors highlighted. **Left:** Each odor is clustered in its own space. **Right:** The hierarchical nature of the odor descriptor. Shaded and contoured areas are computed with a [kernel-density estimate](#) of the embeddings.

- In work that combines chemistry and reinforcement learning techniques, we presented a [framework for molecule optimization](#).
- Machine learning can also help us in our artistic and creative endeavors. Artists have found ways to [collaborate with AI and AR](#) and create interesting new forms, from [dancing with a machine](#) to reimagine choreography, to [creating new melodies](#) with machine learning tools. ML can be used by novices, too. To honor the birthday of J.S. Bach, we featured a [ML-powered Doodle](#): just create your melody, and the ML tool can create accompanying harmonizations in Bach's style.

Assistive Technology

On a more personal scale, ML can help us in our daily lives. It's easy to take for granted our ability to see a beautiful image, to hear a favorite song, or to speak with a loved one. Yet over one billion people aren't able to access the world in these ways. ML technology can help by turning these signals—vision, hearing, speech—into other signals that can be well-managed by people with accessibility needs, enabling better access to the world around them. A few examples of our assistive technology:

- [Lookout](#) helps people who are blind or have low vision identify information about their surroundings. It draws upon similar underlying technology as Google Lens, which lets you search and take action on the objects around you, simply by pointing your phone.
- [Live Transcribe](#) has the potential to give people who are deaf or hard of hearing greater independence in their everyday interactions. You can get real-time transcriptions of conversations that the user is engaged in, even if the speech is in another language.
- [Project Euphoria](#) performs personalized speech-to-text transcription. For people with ALS and other conditions that produce slurred or non-standard speech, [this research](#) improves automatic speech recognition (ASR) over other state-of-the-art ASR models.
- Like Project Euphoria, [Parrottron](#) uses end-to-end neural networks to help improve communication, but the research focuses on automatic speech-to-speech conversion rather than transcription, presenting a speech interface that may be easier for some to access.
- Millions of images online don't have any text description. [Get Image Descriptions from Google](#) helps blind or low vision users understand unlabelled images. When a screen reader encounters an image or graphic without a description, Chrome can now create one automatically.

¹ We developed tools that can read visual text in audio form in [Image for Google](#). On

- We developed tools that can read visual text in audio form in [Lens for Google Go](#), greatly helping users who are not fully literate navigate the word-rich world around them.



Making Your Phone More Intelligent

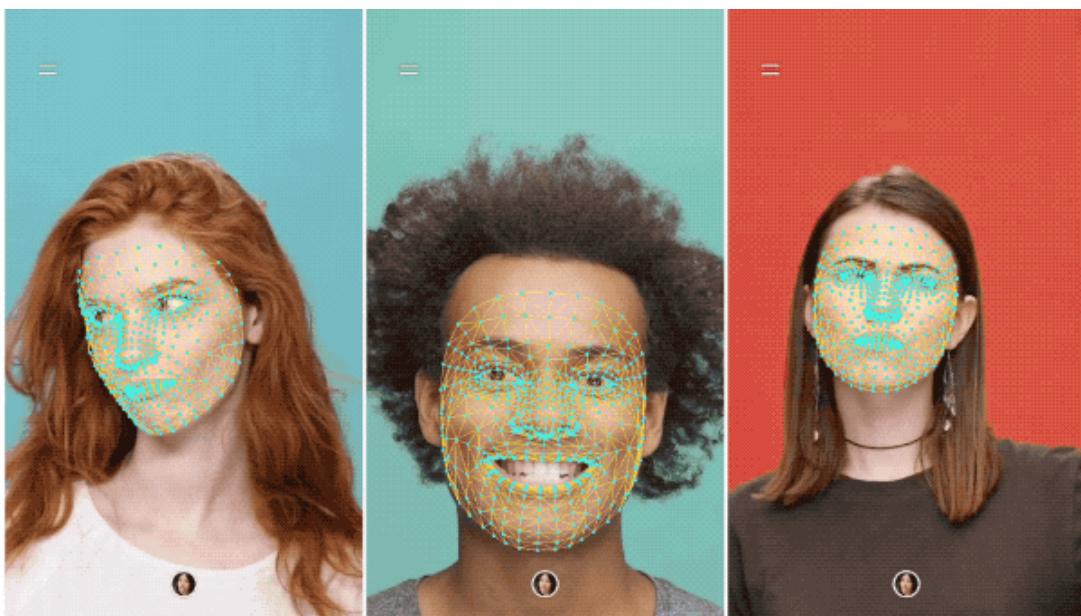
Much of our work serves to enable intelligent, personal devices by giving mobile phones new capabilities through the use of on-device machine learning. By making powerful models that can run on-device, we can ensure that these phone features are highly responsive and always available even in airplane mode or otherwise off the network. We've made progress in getting highly accurate [speech recognition models](#), [vision models](#) and [handwriting recognition models](#) all running on-device, paving the way for powerful new features. Some of this year's highlights include:

- The launch of [on-device captioning with Live Caption](#), giving always-available transcription of any video playing on your device.

- The creation of a powerful new [transcribing Recorder app](#), which can help index audio information and make it easily retrievable.
- Improvements to [Google Translate's camera translation](#), so that you can point at text in an unfamiliar language and get it instantly translated in context.



- Release of the [Augmented Faces API](#) in ARCore, enabling new [real-time AR self-expression tools](#).



- A demonstration of [on-device, real-time hand tracking](#), enabling new ways for users to interact with and control their devices with their hands.

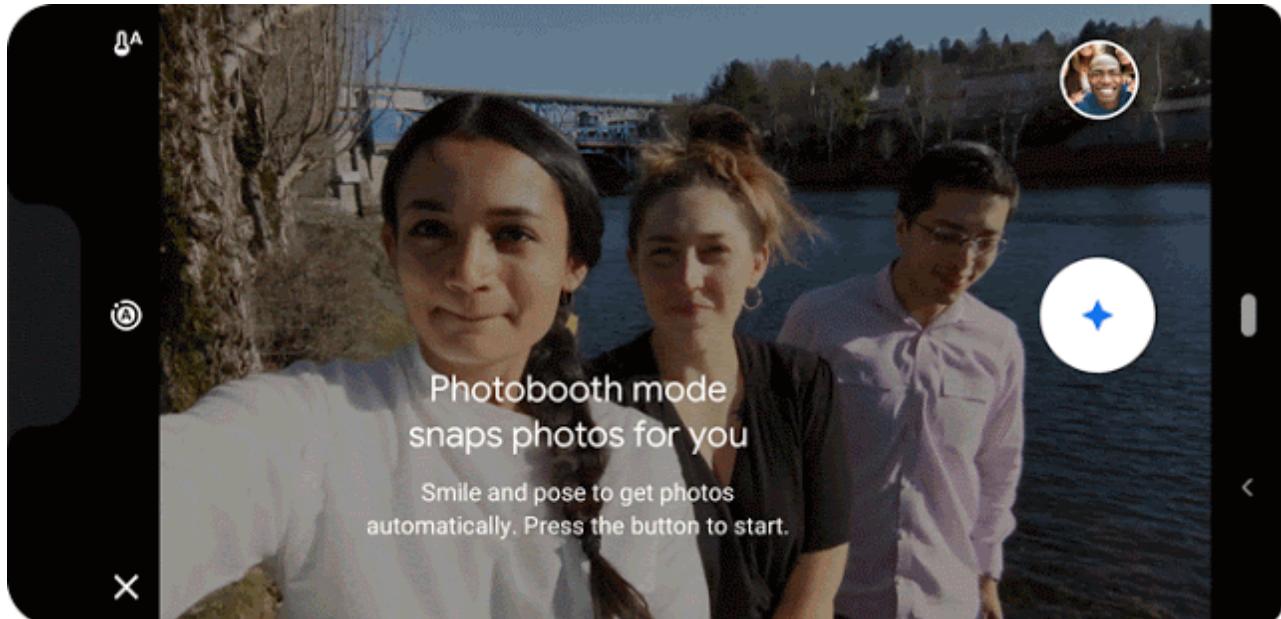


- Improved, RNN-based on-device handwriting recognition for on-screen mobile keyboards.
- The release of a new global localization approach using your smart phone's camera to help more accurately orient you and help you find your way in the world.



Federated learning (check out the [online comic description!](#)) is a [powerful machine learning approach invented by Google researchers in 2015](#), whereby many clients (such as mobile devices or whole organizations) collaboratively train a model, while keeping the training data decentralized. This enables approaches that have superior privacy properties in [large-scale learning systems](#). We are using federated learning in more and more of our products and features, while also working to advance the state of the art in many research problems in this space. In 2019, Google researchers collaborated with authors from 24 (!) academic institutions to produce a [survey article on Federated Learning](#), highlighting advances over the past few years as well describing a number of open research problems in the field.

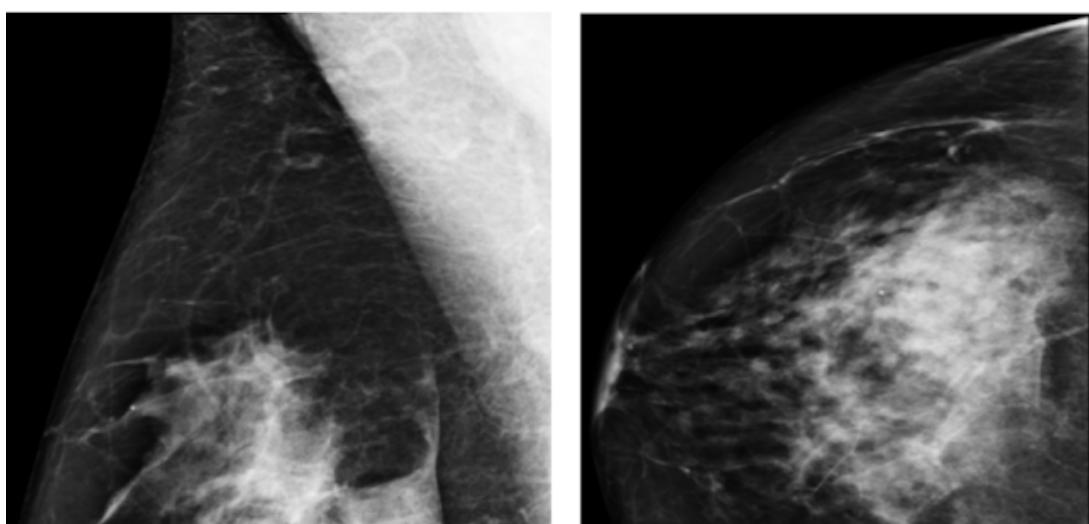
The field of computational photography has led to great advances in the image quality of phone cameras over the past few years, and this year was no exception. This year, we made it easier to [take great selfies](#), to take [professional-looking shallow depth of field images and portraits](#) and to use the [Night Sight feature](#) on Pixel Phones to take some stunning astrophotography [pictures](#). More technical details about this work can be found in papers on [multi-frame super resolution](#) and [mobile photography in very low-light conditions](#). All of this work helps enable you to take great pictures to remember life's magical moments as they happen.

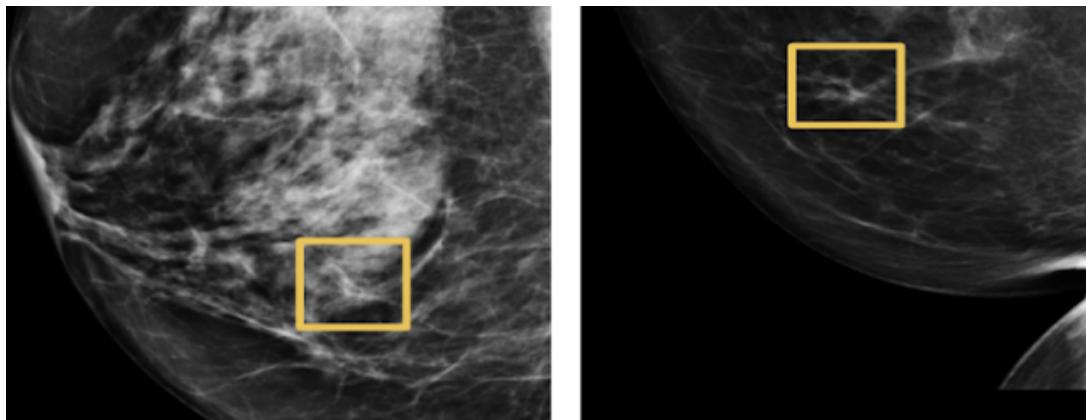


Health

In late 2018, we combined the Google Research health team, Deepmind Health and a team from Google's Hardware division focused on health-related applications to form [Google Health](#). In 2019 we continued the research we've been pursuing in this space, [publishing research papers](#) and [building tools](#) in collaboration with a variety of healthcare partners. Here are a few of the highlights from 2019:

- We showed that a [deep learning model for mammography](#) can assist physicians in spotting breast cancer, a condition that affects 1 in 8 women in the US during their lifetimes, with greater accuracy than experts, reducing both false positives and false negatives. The model trained on de-identified data from a UK hospital had similar gains in accuracy when used to evaluate patients in a completely different healthcare system in the U.S.





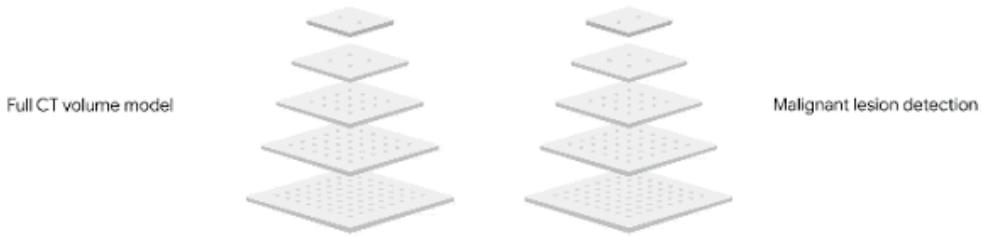
Example of a difficult-to-detect cancer case correctly identified by machine learning.

- We showed that a [deep learning model for differential diagnoses of skin diseases](#) can give results that are significantly more accurate than primary care physicians and on par with or perhaps slightly better than dermatologists.
- Working alongside experts from the US Department of Veterans Affairs (VA), DeepMind Health colleagues who are now part of Google Health showed that a [machine learning model can predict the onset of acute kidney injury](#) (AKI), one of the leading causes of avoidable patient harm, up to two days before it happens. In the future, this could give doctors a 48-hour head start in treating this serious condition.
- We [expanded the application of deep learning to electronic health records](#) with several partner organizations. You can read more about this work in our [2018 blog post](#).

Google Health - Tools to help healthcare providers deliver...



- We showed a [promising step forward for predicting lung cancer](#), where a deep learning model for examining the results of a single CT scan study performed on par or better than trained radiologists at early detection of lung cancer. Early detection of lung cancer dramatically improves survival rates.



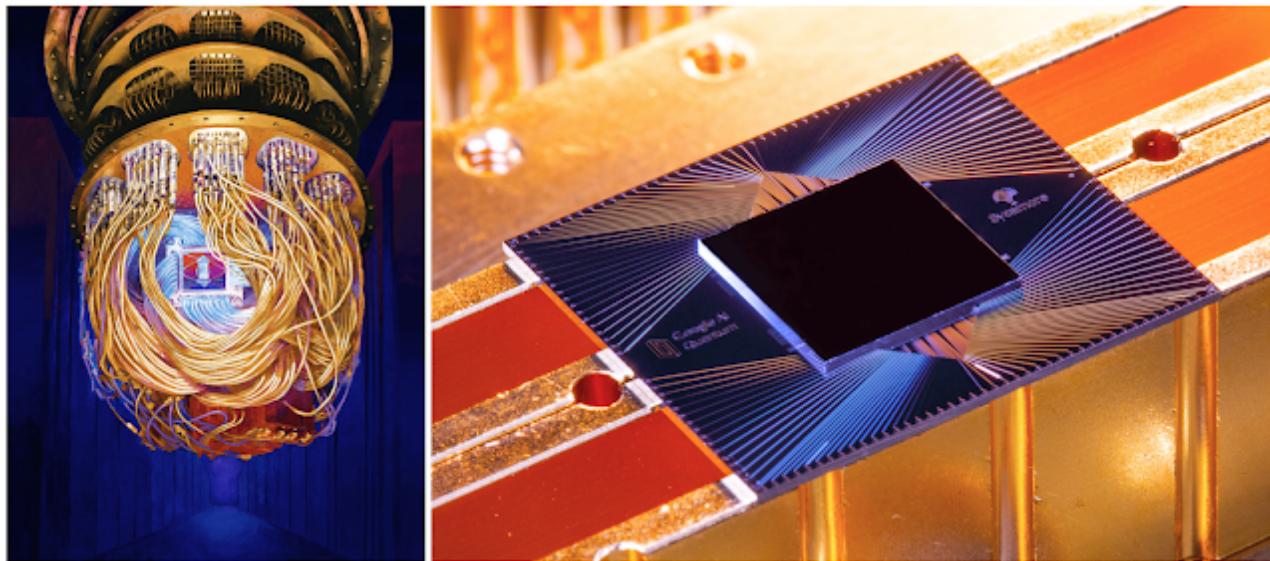
- We continued to [expand and evaluate our deployment of machine learning tools for detection and prevention of eye disease](#), in collaboration with [Verily](#) and our healthcare partners in India and Thailand.
- We published a research paper on [an augmented reality microscope for cancer diagnosis](#), whereby a pathologist can get real-time feedback about what parts of a slide are most interesting while examining tissue through a microscope. You can also read more about it in our 2018 blog post [here](#).
- We built a [human-centric, similar-image search tool](#) for pathologists to help them make more effective diagnoses, by allowing examination of similar cases.

SMILY : Similar image search for pathologists



Quantum Computing

In 2019, our quantum computing team demonstrated for the first time a [computational task](#) that can be executed exponentially faster on a quantum processor than on the world's fastest classical computer – just 200 seconds compared to 10,000 years.



Left: Artist's rendition of the Sycamore processor mounted in the cryostat. ([Full Res Version](#); Forest Stearns, Google AI Quantum Artist in Residence) Right: Photograph of the Sycamore processor. ([Full Res Version](#); Erik Lucero, Research Scientist and Lead Production Quantum Hardware)

Using quantum computers may make important problems in domains like materials science, quantum chemistry ([early example](#)) and large-scale optimization tractable, but in order to make this a reality, we'll have to continue to push the field forward. We are now focusing on implementing quantum error correction so that we will be able to run computations for longer. We are also working on making quantum algorithms [easier to express](#), the hardware [easier to control](#) and we have found ways to use [classical machine learning techniques like deep reinforcement learning](#) to build more reliable quantum processors. The achievements this year are encouraging and are early steps along the way to making practical quantum computing a reality for a wider variety of problems.

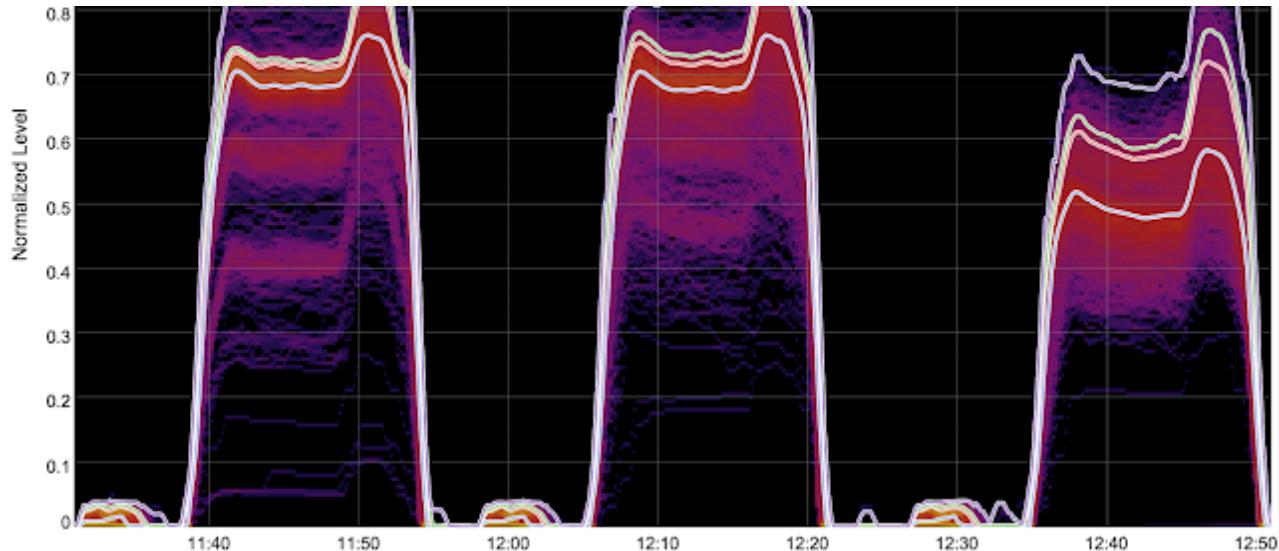
You can also read Sundar's thoughts on [what our quantum computing milestone means](#).

General Algorithms and Theory

In the general areas of [algorithms and theory](#), we continued our research from algorithmic foundations to applications, and also did work in [graph mining](#) and [market algorithms](#). A blog post [summarizing some of our work in graph learning algorithms](#) gives more details about that work.

We published a paper at [VLDB'19](#) titled "[Cache-aware load balancing of data center applications](#)," although an alternative title could be "Increase the serving capacity of your data center by 40% with this one cool trick!". The paper describes how we used [balanced partitioning of graphs](#) to specialize the caches in our web search backend serving system, thereby increasing the query throughput of our flash drives by 48%, and helping to enable a 40% increase in the throughput of the entire search backend.





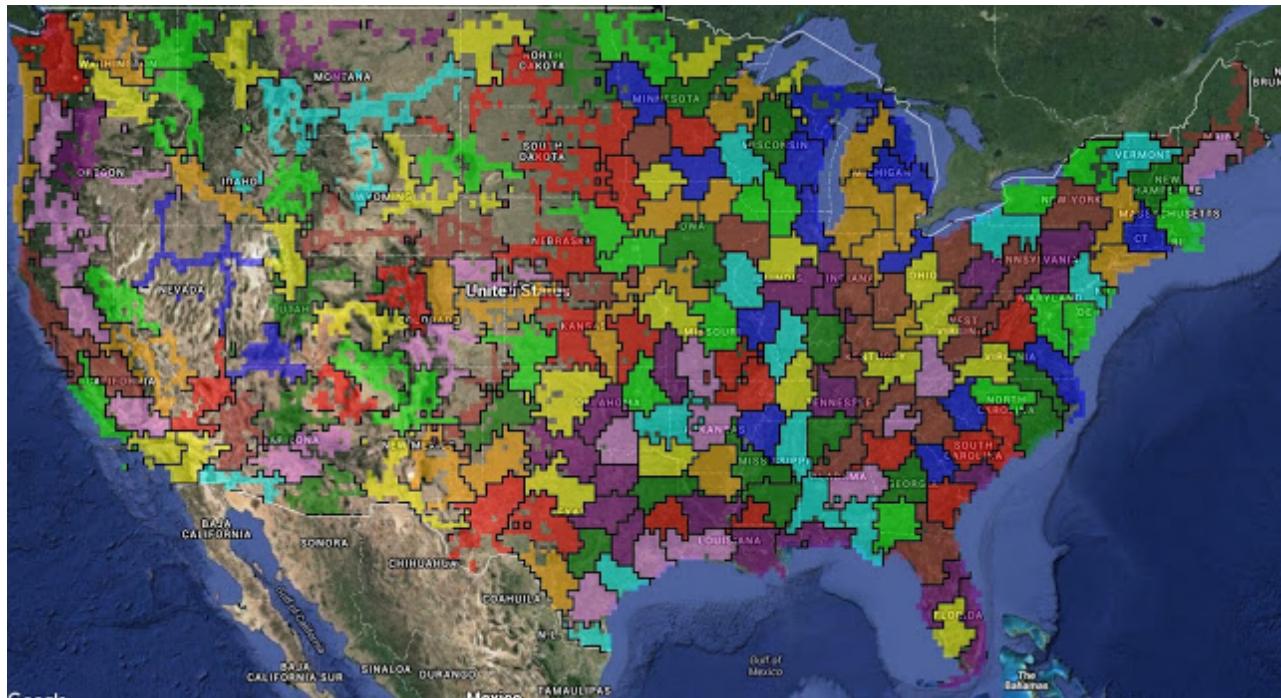
Heatmap of flash IO requests (resulting from cache misses) across web search serving leaves. The three humps represent random leaf selection, load balancing, and cache-aware load balancing (left to right). Lines indicate the 50th, 90th, 95th and 99.9th percentiles. From VLDB'19 paper, "Cache-aware load balancing of data center applications."

In an ICLR'2019 paper titled "[A new dog learns old tricks: RL finds classic optimization algorithms](#)," we discovered a new connection between algorithms and machine learning, showing how Reinforcement Learning can effectively find optimal (worst-case, uniform) algorithms for several classic online optimization combinatorial problems such as online matching and allocation.

Our work in scalable algorithms spans both parallel, online and distributed algorithms for big data sets. In a recent [FOCS'19 paper](#), we provided a near-optimal massively parallel computation algorithm for connected components. Another set of our papers improved parallel algorithms for matching (in [theory](#) and [practice](#)) and for [density clustering](#). And a third line of work concerned adaptively optimizing submodular functions in the black-box model, which has several applications in feature selection and [vocabulary compression](#). In a [SODA'19 paper](#), we presented a submodular maximization algorithm that is nearly optimal in three aspects: approximation factor, round complexity, and query complexity. Also, in another [FOCS 2019 paper](#), we provide the first online multiplicative approximation algorithm for PCA and Column Subset selection.

In other work, we introduce the semi-online model of computation that postulates that the unknown future has a predictable part and an adversarial part. For classical combinatorial problems such as [bipartite matching](#) ([ITCS'19](#)) and [caching](#) ([SODA'20](#)), we obtained semi-online algorithms to provide guarantees that smoothly interpolate between the best possible online and offline algorithms.

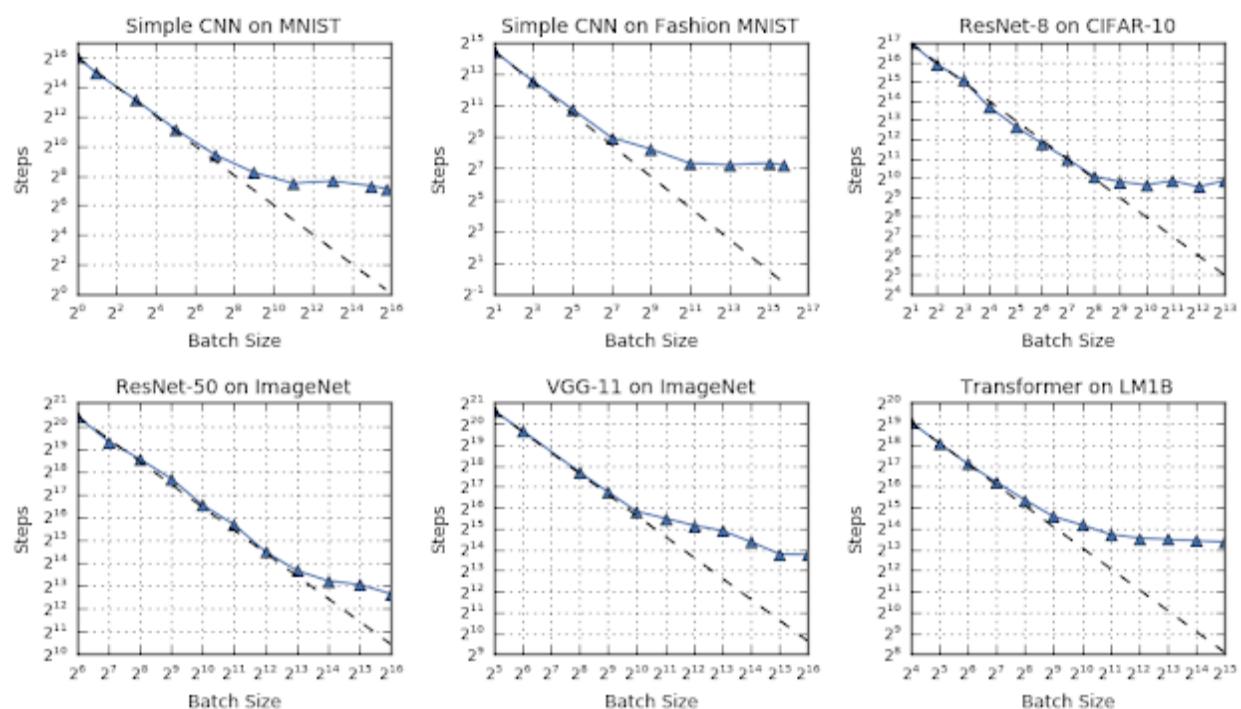
Our recent research in the area of [market algorithms](#) includes new understanding of the interaction between learning and markets, and [innovations in experimental design](#). For example, this [NeurIPS'19 oral paper](#) reveals the surprising competitive advantage that a strategic agent has when competing with a learning agent in a general repeated 2-player game. Recent focus on advertising automation has produced increased interest in automated bidding and understanding response behavior of advertisers. In a pair of [WINE 2019](#) papers, we study optimal strategy to maximize conversions on behalf of advertisers and further learn advertiser response behavior for any changes in the auction. Finally, we studied experimental design in the presence of interference where the treatment of one group may affect the outcomes of others. In a [KDD'19 paper](#) and a [NeurIPS'19 paper](#), we show how to define units or clusters of [units to limit interference while maintaining experimental power](#).



The clustering algorithm from the KDD'19 paper “[Randomized Experimental Design via Geographic Clustering](#)” applied to user queries from the United States. The algorithm automatically identifies metropolitan areas, correctly predicting, for example, that the Bay Area includes San Francisco, Berkeley, and Palo Alto, but not Sacramento.

Machine Learning Algorithms

In 2019, we conducted research in many different areas of machine learning algorithms and approaches. One major focus was in understanding the properties of training dynamics in neural networks. In the blog post [Measuring the Limits of Data Parallel Training for Neural Networks](#) highlighting this [paper](#), Google researchers presented a careful set of experimental results showing when scaling the amount of data parallelism (by making larger batches) is effective for allowing the model to converge faster (using data parallelism).



For all workloads we tested, we observed a universal relationship between batch size and training speed with three

distinct regimes: perfect scaling with small batch sizes (following the dashed line), eventually seeing diminishing returns as batch sizes grow (diverging from the dashed line), and maximal data parallelism at the largest batch sizes (where the trend plateaus). The transition points between the regimes vary dramatically between different workloads.

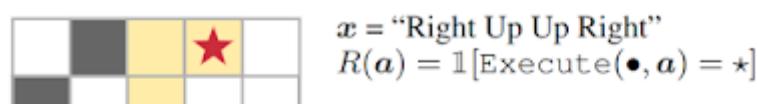
Model parallelism, in contrast to data parallelism, where a model is spread out across multiple computational devices, can be an effective way of scaling models. [GPipe](#) is a library that enables model parallelism to be more effective, in an approach similar to that used by pipelined CPU processors: when one part of the whole model is working on some of the data, other parts can be working on their part of the computation on different data. The results of this pipeline approach can be combined together to simulate a larger effective batch size.

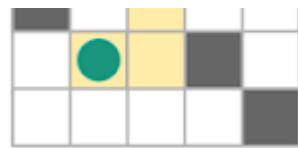
Machine learning models are effective when they're able to take raw input data and learn "disentangled" higher-level representations that separate different kinds of examples by properties that we want the model to be able to distinguish (cat vs. truck vs. wildebeest, cancerous tissue vs. normal tissue, etc.). Much of the focus on advancing machine learning algorithms is to encourage the learning of better representations that generalize better to new examples, problems or domains. This year, we looked at this problem in a number of different contexts:

- In [Evaluating the Unsupervised Learning of Disentangled Representations](#), we examined what properties affect the representations that are learned from unsupervised data, in order to better understand what makes for good representations and effective learning.
- In [Predicting the Generalization Gap in Deep Neural Networks](#), we showed that it is possible to predict the generalization gap (the gap between a model's performance on data from the training distribution versus data drawn from a different distribution) using statistics of the margin distribution, helping us better understand which models generalize most effectively. We also did some research on [Improving Out-of-Distribution Detection in Machine Learning Models](#), to better understand when a model is starting to encounter kinds of data it has never seen before. We also looked at [Off-Policy Classification](#) in the context of reinforcement learning, to better understand which models are likely to generalize the best.



- In [Learning to Generalize from Sparse and Underspecified Rewards](#), we also examined ways of specifying reward functions for reinforcement learning that enable learning systems to more directly learn from true objectives and be less distracted with longer, less-desirable sequences of actions that happen to achieve desired goals by accident.





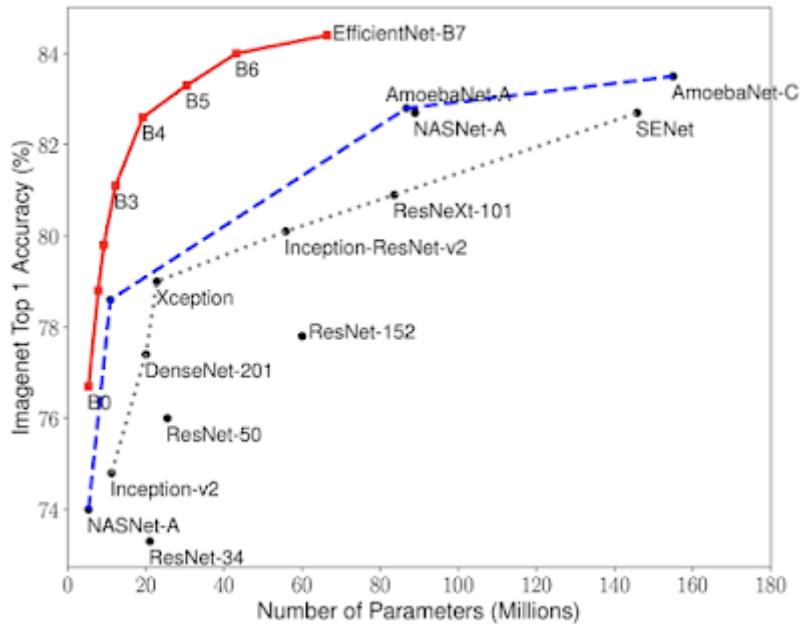
$$\begin{aligned} \mathbf{a}_1 &= (\rightarrow, \uparrow, \uparrow, \rightarrow) \\ \mathbf{a}_2 &= (\leftarrow, \rightarrow, \rightarrow, \uparrow, \uparrow, \rightarrow) \\ \mathbf{a}_3 &= (\uparrow, \rightarrow, \rightarrow, \uparrow) \\ R(\mathbf{a}_1) &= R(\mathbf{a}_2) = R(\mathbf{a}_3) = 1 \end{aligned}$$

In this instruction-following task, the action trajectories \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 reach the goal, but the sequences \mathbf{a}_2 and \mathbf{a}_3 do not follow the instructions. This illustrates the issue of underspecified rewards.

AutoML

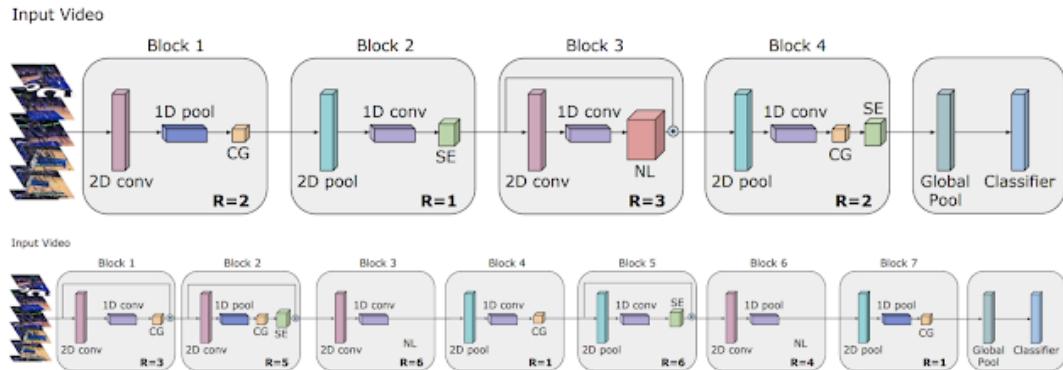
We continued our work on AutoML this year, an approach whereby algorithms that learn how to learn can automate many aspects of machine learning and often can achieve substantially better results than the best human machine learning experts for certain kinds of machine learning meta-decisions. In particular:

- In [EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling](#), we showed how to use neural architecture search techniques to achieve substantially better results on computer vision problems, including a new state-of-the-art result of 84.4% top-1 accuracy on ImageNet while having 8X fewer parameters than the previous best model.



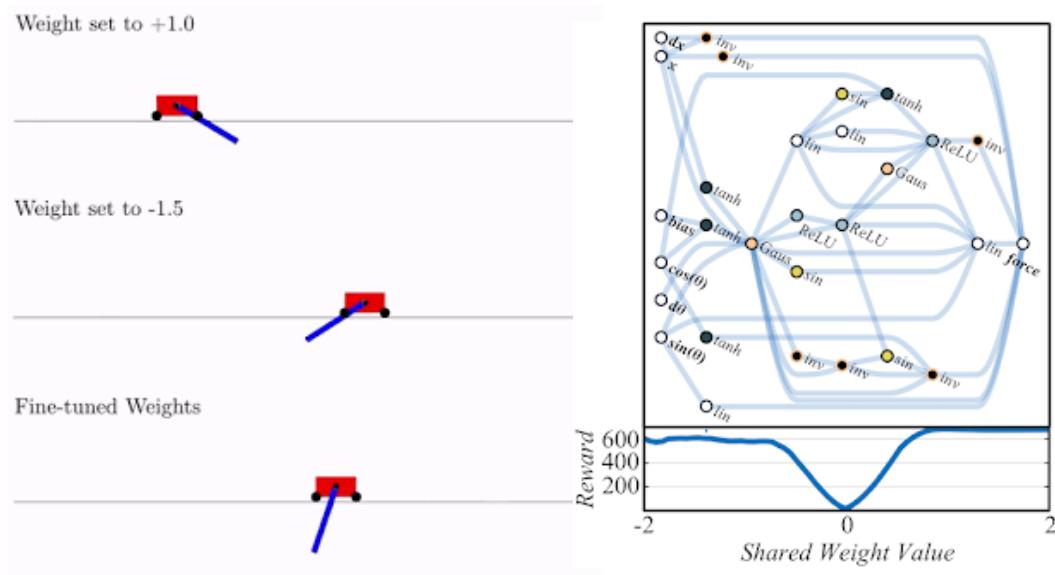
Model Size vs. Accuracy Comparison. EfficientNet-B0 is the baseline network developed by [AutoML MNAS](#), while Efficient-B1 to B7 are obtained by scaling up the baseline network. In particular, our EfficientNet-B7 achieves new state-of-the-art 84.4% top-1 / 97.1% top-5 accuracy, while being 8.4x smaller than the best existing CNN.

- In [EfficientNet-EdgeTPU: Creating Accelerator-Optimized Neural Networks with AutoML](#), we showed how a neural architecture search approach can find efficient models that are tailored to particular hardware accelerators, resulting in high accuracy, low-computational models for running on mobile devices.
- In [Video Architecture Search](#), we describe how we extended our AutoML work to the domain of video models, finding architectures that achieve state-of-the-art results, and also lightweight architectures that match the performance of hand-crafted models while using 50x less computation.



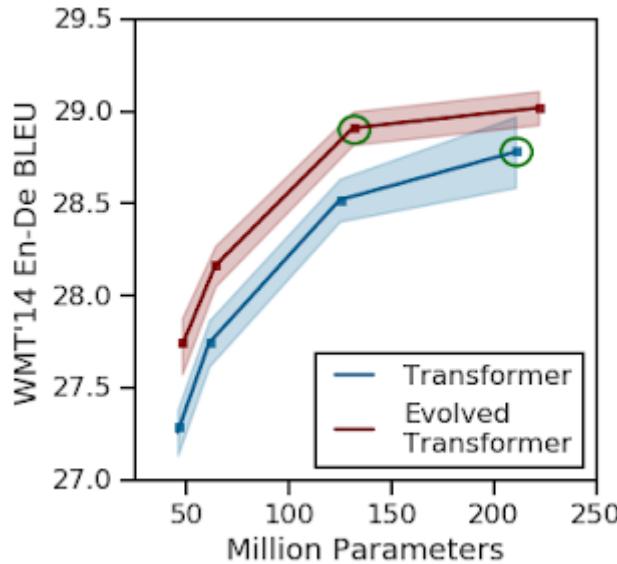
TinyVideoNet (TVN) architectures evolved to maximize the recognition performance while keeping its computation time within the desired limit. For instance, TVN-1 (top) runs at 37 ms on a CPU and 10ms on a GPU. TVN-2 (bottom) runs at 65ms on a CPU and 13ms on a GPU.

- We developed AutoML techniques for tabular data, unlocking an important domain where many companies and organizations have interesting data in relational databases, and often want to develop machine learning models on this data. We collaborated to release this technology as a new Google Cloud [AutoML Tables](#) product, and also discussed how well this system did in a new Kaggle competition in [An End-to-End AutoML Solution for Tabular Data at KaggleDays](#) (spoiler: AutoML Tables finished second out of 74 teams of expert data scientists).
- In [Exploring Weight Agnostic Neural Networks](#), we showed how it is possible to find interesting neural network architectures without any training steps to update the weights of the evaluated models. This can make architecture search much more computationally efficient.



A weight-agnostic neural network performing a Cartpole Swing-up task at various different weight parameters, and also using fine-tuned weight parameters.

- [Applying AutoML to Transformer Architectures](#) explored finding architectures for natural language processing tasks that significantly outperform vanilla Transformer models at substantially reduced computational costs.



Comparison between the Evolved Transformer and the original Transformer on WMT'14 En-De at varying sizes. The biggest gains in performance occur at smaller sizes, while ET also shows strength at larger sizes, outperforming the largest Transformer with 37.6% less parameters (models to compare are circled in green). See Table 3 in [our paper](#) for the exact numbers.

- In [SpecAugment: A New Data Augmentation Method for Automatic Speech Recognition](#), we showed that the approach of automatically learning data augmentation methods can be extended to speech recognition models, with the learned augmentation approaches achieving significantly higher accuracy with less data than existing human ML-expert driven data augmentation approaches.
- We launched our first speech application for [keyword spotting and spoken language identification using AutoML](#). In our experiments we found better models (both more efficient and better performance) than the human designed models that have been in this setting for some time.

Natural Language Understanding

The past few years have seen remarkable advances in models for natural language understanding, translation, natural dialog, speech recognition and related tasks. This year, one theme in our work was advancing the state of the art by combining modalities or tasks, to train more powerful and capable models. A few examples:

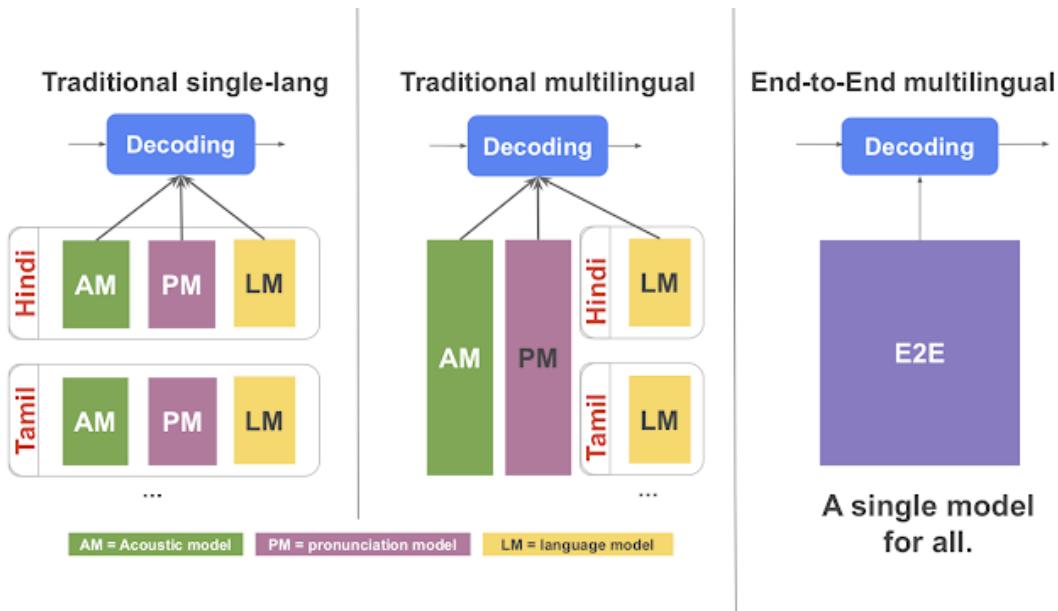
- In [Exploring Massively Multilingual, Massive Neural Machine Translation](#), we showed significant gains in translation quality by training a single model to translate between 100 languages, rather than having 100 separate models.



Left: Language pairs with larger amounts of training data generally have higher translation quality.

Right: Multilingual training, where we train a single model for all language pairs rather than separate models for each language pair, results in substantial improvements in **BLEU score** (a measure of translation quality) for language pairs without much data.

- In [Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model](#), we showed how combining speech recognition and language models together and training the system on many languages, can significantly improve speech recognition accuracy.



Left: A traditional monolingual speech recognizer comprised of Acoustic, Pronunciation and Language Models for each language. Middle: A traditional multilingual speech recognizer where the Acoustic and Pronunciation model is multilingual, while the Language model is language-specific. Right: An E2E multilingual speech recognizer where the Acoustic, Pronunciation and Language Model is combined into a single multilingual model.

- In [Translatotron: An End-to-End Speech-to-Speech Translation Model](#), we showed that it is possible to train a joint model to accomplish the (normally separate) tasks of speech recognition, translation and text-to-speech generation with nice benefits, like preserving the sound of the speaker's voice in the generated translated audio, as well as a simpler overall learning system.
- In [Multilingual Universal Sentence Encoder for Semantic Retrieval](#), we showed how to combine many different objectives to yield models that are significantly better at semantic retrieval (versus simpler word matching techniques). For example, in [Google Talk to Books](#), the query "[What fragrance brings back memories?](#)" yields the result, "*And for me, the smell of jasmine along with the pan bagnat, it brings back my entire carefree childhood.*"
- In [Robust Neural Machine Translation](#), we showed how to use an adversarial training procedure to significantly improve the quality and robustness of language translations.

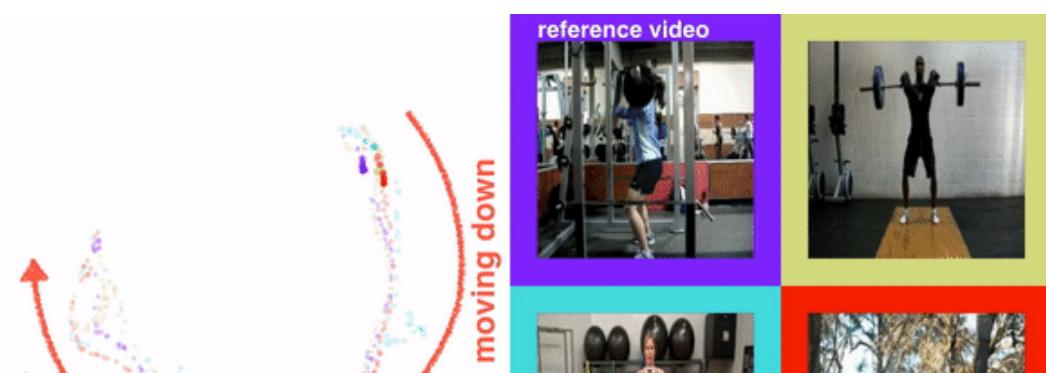
Left: The Transformer model is applied to an input sentence (*lower left*) and, in conjunction with the target output sentence (*above right*) and target input sentence (*middle right*; beginning with the placeholder "<sos>"), the translation loss is calculated. The AdvGen function then takes the source sentence, word selection distribution, word candidates and the translation loss as inputs to construct an adversarial source example. **Right:** In the defense stage, the adversarial source example serves as input to the Transformer model and the translation loss is calculated. AdvGen then uses the same method as above to generate an adversarial target example from the target input.

As our language understanding capabilities have improved, based on fundamental research advances like [seq2seq](#), [Transformer](#), [BERT](#), [Transformer-XL](#) and [ALBERT](#) models, we have seen increased use of these sorts of models in many of our core products and features like [Google Translate](#), Gmail's [Smart Compose](#), and [Google Search](#). This year, the [launch of BERT in our core search and ranking algorithms](#) led to the biggest improvement in search quality in the last five years (and one of the biggest ever), through better understanding of the subtle meanings of query and document words and phrases.

Machine Perception

Models for better understanding of still images have made remarkable progress in the last decade. Among the next major frontiers are models and approaches for understanding the dynamic world in fine-grained detail. This includes deeper and more nuanced understanding of images and video, as well as live and situated perception: understanding the audiovisual world at interactive rates and with a shared spatial grounding with the user. This year, we explored many aspects of advances in this area, including:

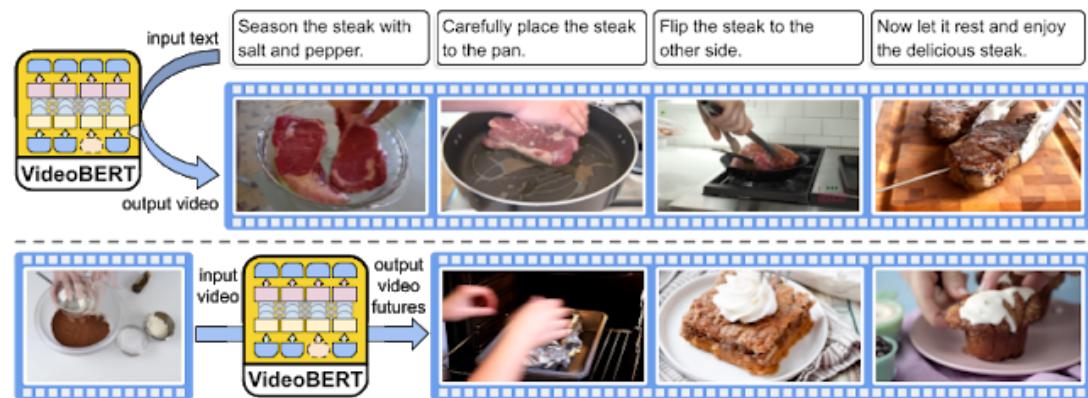
- Finer-grained visual understanding in Lens, enabling even more powerful [visual search](#).
- Helpful smart camera features such as Quick Gestures, Face Match and smart video call framing on the [Nest Hub Max](#).
- Technology for live and spatially-aware perception for helpfully augmenting the world around us through [Lens](#).
- Better models for [depth prediction from videos](#).
- Better representations for [fine-grained temporal understanding of videos using temporal cycle-consistency learning](#).





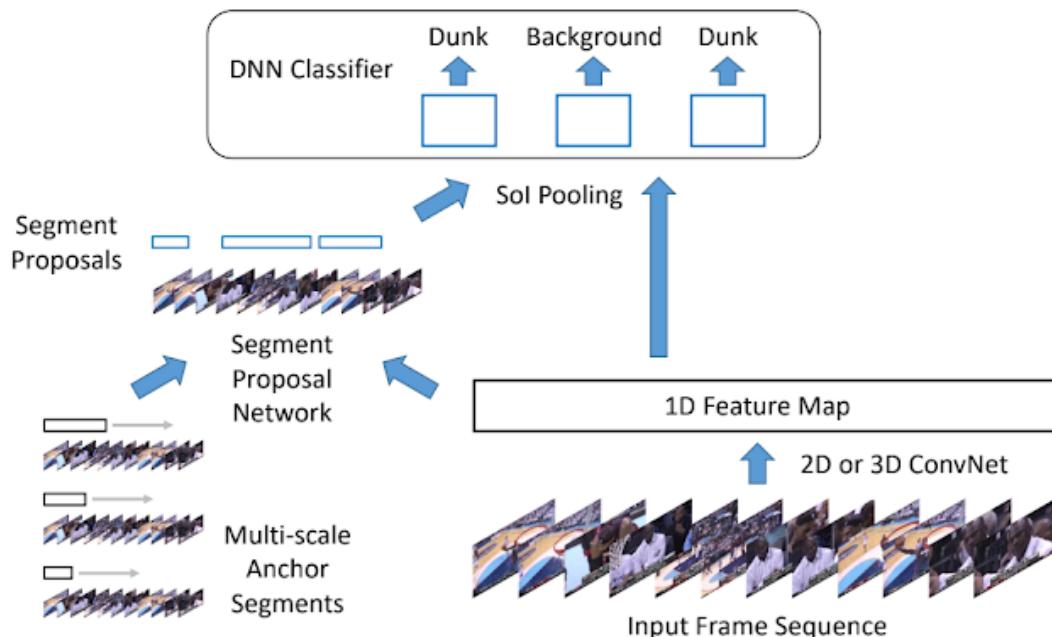
Right: Input videos of people performing a squat exercise. The video on the top left is the reference. The other videos show nearest neighbor frames (in the TCC embedding space) from other videos of people doing squats. Left: The corresponding frame embeddings move as the action is performed.

- Learning representations across text, speech and video that are temporally consistent from unlabeled videos.



Qualitative results from VideoBERT, pretrained on cooking videos. **Top:** Given some recipe text, we generate a sequence of visual tokens. **Bottom:** Given a visual token, we show the top three future tokens forecast by VideoBERT at different time scales. In this case, the model predicts that a bowl of flour and cocoa powder may be baked in an oven and may become a brownie or cupcake. We visualize the visual tokens using the images from the training set closest to the tokens in feature space.

- Being able to predict future visual inputs from observations of the past.
- Models that can better understand action sequences in videos, enabling you to better recall special video moments like “blowing out candles” or “sliding down a slide” in Google Photos.



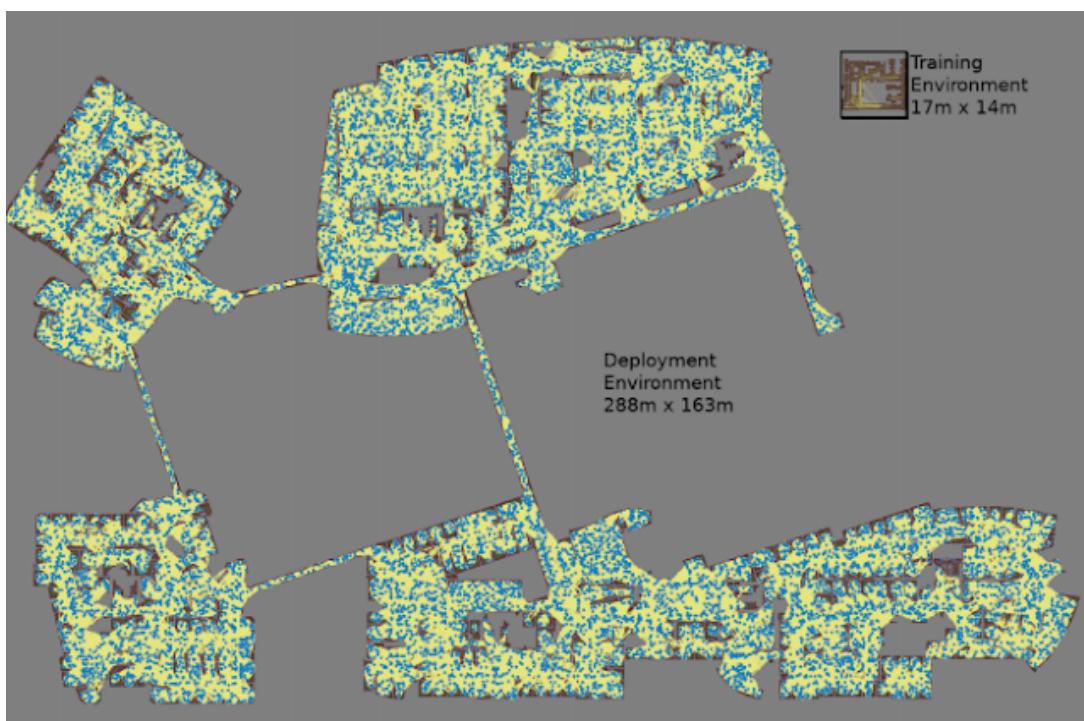
Architecture for temporal action localization.

We're quite excited about the prospects of continued improvements in the understanding of the sensory world around us.

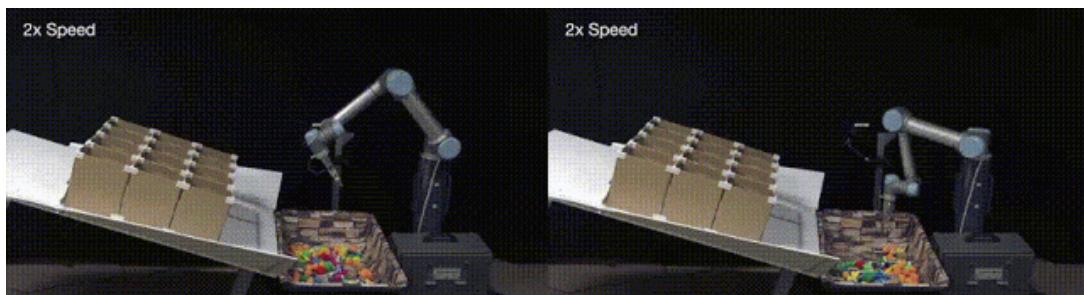
Robotics

The application of machine learning to robotic control is a [significant research area](#) for us. We believe this is a vital tool for enabling robots to operate effectively in complex, real-world environments like everyday homes and businesses. Some of the work we did this year includes:

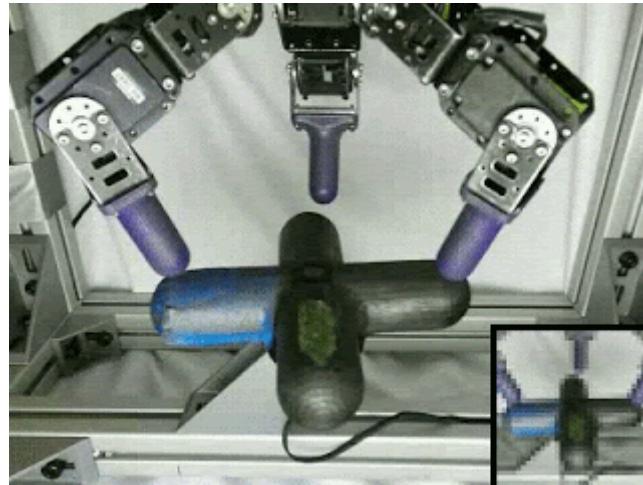
- In [Long-Range Robotic Navigation via Automated Reinforcement Learning](#), we showed how to combine reinforcement learning with long-range planning to enable robots to more effectively navigate complex environments (like our Google office buildings).



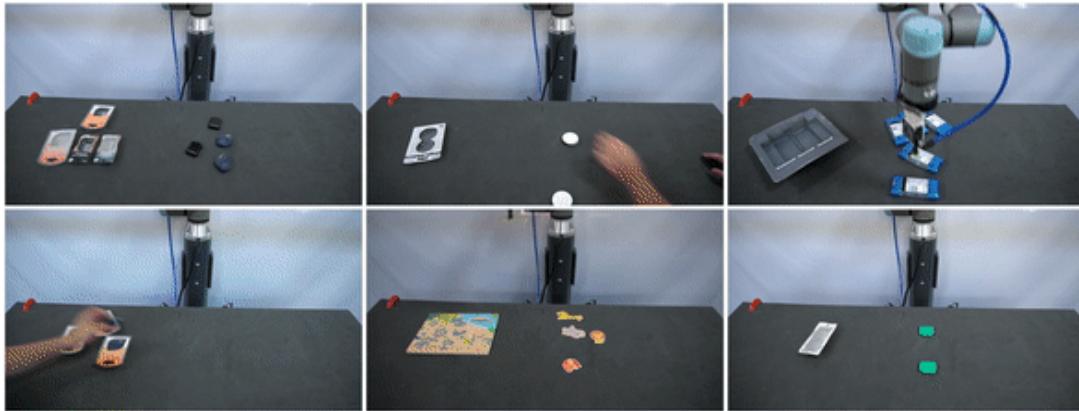
- In [PlaNet: A Deep Planning Network for Reinforcement Learning](#), we showed how to effectively learn a world model purely from the pixels of images, and how to leverage this model of how the world behaves in order to accomplish tasks with many fewer learning episodes.
- In [Unifying Physics and Deep Learning with TossingBot](#), we showed how robots can learn “intuitive” physics from experimentation in an environment, rather than being pre-programmed with physics models about the environment in which they are operating.



- In [Soft Actor-Critic: Deep Reinforcement Learning for Robotics](#), we showed that training a reinforcement learning algorithm to both maximize the expected reward (which is the standard RL objective) and to maximize the policy's entropy (so that learning favors policies that are more random), can help robots learn faster and be more robust to changes in their environment.



- In [Learning to Assemble and to Generalize from Self-Supervised Disassembly](#), we showed how robots can learn to assemble by first learning to disassemble things in a self-supervised way. Kids learn from taking things apart, and it appears that robots can as well!



- We introduced [ROBEL: Robotics Benchmarks for Learning with Low-Cost Robots](#), an open-source platform of cost-effective robots and curated benchmarks designed to facilitate research and development on physical robotics hardware in the real world.

Helping Advance the Broader Developer and Researcher Community

Open source is about more than code: it's about the community of contributors. It's been an exciting year to be part of the open source community. We [launched TensorFlow 2.0](#)—the biggest TensorFlow release to date—which makes building ML systems and applications easier than ever. We added support for fast [mobile GPU inference](#) to [TensorFlow Lite](#). We also [launched Teachable Machine 2.0](#), a fast, easy web-based tool which can train a machine learning model with the click of a button, no coding required. We announced [MLIR](#), open source machine learning compiler infrastructure that addresses the complexity of growing software and hardware fragmentation and makes it easier to build AI applications.

We saw the first year of [JAX](#), a new system for high-performance machine learning research. At [NeurIPS 2019](#), Googlers and the broader open-source community presented work using JAX ranging from [neural tangent kernels](#) to [Bayesian inference](#) to [molecular dynamics](#), and we launched [a preview of JAX on Cloud TPUs](#).

We open-sourced [MediaPipe](#), a framework for building perceptual and multimodal applied ML pipelines, and [XNNPACK](#), a library of efficient floating-point neural network inference operators. As of the end of 2019, we had enabled more than 1,500 researchers around the world to access [Cloud TPUs](#) for free via the [TensorFlow Research Cloud](#). Our [Intro To TensorFlow](#) at Coursera crossed 100,000 students. And we engaged with thousands of users while [taking TensorFlow on the road](#) to 11 different countries, hosted our first ever [TensorFlow World](#) and more.

With the help of TensorFlow, one college student [discovered two new planets](#) and built a method to help others find more. A data scientist originally from Nigeria trained a GAN to generate images [reminiscent of African masks](#). A developer in Uganda used TensorFlow to [create the Farmers Companion](#), an app that local farmers can use to fight a crop-destroying caterpillar. In snowy Iowa, researchers and state officials used TensorFlow to [determine safe road conditions](#) based on traffic behavior, visuals and other data. In sunny California, college students used TensorFlow to [identify pot holes and dangerous road cracks](#) in Los Angeles. And in France, a coder used TensorFlow to build a simple algorithm that [learns how to add color to black-and-white photos](#).

Open Datasets

Open datasets with clear and measurable goals are often very helpful in driving forward the field of machine learning. To help the research community find interesting datasets, we continue to index a wide variety of open datasets sourced from many different organizations with [Google Dataset Search](#). We also think it's important to create new datasets for the community to explore and to develop new techniques, and to ensure we [share open data responsibly](#). This year, we additionally released a number of open datasets across many different areas:

- [Open Images V5](#): An update to the popular [Open Images](#) dataset that includes segmentation masks for 2.8 million objects in 350 categories (so that it now has ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, and visual relationships).
- [Natural questions](#): the first dataset to use naturally occurring queries and find answers by reading an entire page, rather than extracting answers from a short paragraph.
- [Data for deepfake detection](#): we contributed a large dataset of visual deepfakes to the [FaceForensics benchmark](#) (mentioned above).
- [Google Research Football](#): a novel reinforcement learning environment where agents aim to master the world's most popular sport—football (or, if you're American, soccer). It's important for reinforcement learning agents to have GOOOAAALLLSS!
- [Google-Landmarks-v2](#): over 5 million images (2x that of the first release) of more than 200 thousand different landmarks.
- [YouTube-8M Segments](#): A large-scale classification and temporal localization dataset that includes human-verified labels at the 5-second segment level of YouTube-8M videos.
- [Atomic Visual Actions \(AVA\) Spoken Activity](#): A multimodal audio+visual video dataset for perception of conversations. In addition, academic challenges were

run for AVA action recognition and AVA: Spoken Activity

- [PAWS and PAWS-X](#): To help with paraphrase identification, both datasets contain well-formed sentence pairs with high lexical overlap, in which around half of pairs are paraphrase and half are not.
- [Natural language dialog datasets](#): CCPE and Taskmaster-1 both use a Wizard-of-Oz platform that pairs two people who engage in spoken conversations, to mimic a human-level conversation with a digital assistant.
- [The Visual Task Adaptation Benchmark](#): VTAB follows similar guidelines to ImageNet and GLUE but is based on one principle—a better representation is one that yields better performance on unseen tasks, with limited in-domain data.
- [Schema-Guided Dialogue Dataset](#): the largest publicly available corpus of task-oriented dialogues, with over 18,000 dialogues spanning 17 domains.

Research Community Interaction

Finally, we've been busy within the broader academic and research community. In 2019 Google researchers presented hundreds of papers, participated in numerous conferences and received many awards and other accolades. We had a strong presence at:

- [CVPR](#): ~250 Googlers presented 40+ papers, talks, posters, workshops and more.
- [ICML](#): ~200 Googlers presented 100+ papers, talks, posters, workshops and more.
- [ICLR](#): ~200 Googlers presented 60+ papers, talks, posters, workshops and more.
- [ACL](#): ~100 Googlers presented 40+ papers, workshops and tutorials.
- [Interspeech](#): Over 100 Googlers presented 30+ papers.
- [ICCV](#): ~200 Googlers presented 40+ papers, and several Googlers also won three prestigious ICCV awards.
- [NeurIPS](#): ~500 Googlers co-authored more than 120 accepted papers and engaged in various workshops and more.

We also brought together hundreds of Google researchers and faculty from across the globe to 15 separate research workshops hosted at Google locations. These workshops were on topics ranging from improving flood forecasting globally, to how to use machine learning to build systems that can better serve people with disabilities, to accelerating the development of algorithms, applications and tools for noisy-intermediate scale quantum (NISQ) processors.

Supporting academia and research communities outside of Google, we supported over 50 PhD students globally through our annual PhD Fellowship Program, we funded 158 projects as part of our [Google Faculty Research Awards 2018](#), and we held our third cohort of the [Google AI Residency Program](#). We also [mentored AI-focused startups](#).

New Places, New Faces

We've made lots of headway in 2019, but there's so much more we can do. To continue growing our impact around the world, we opened a [Research office in Bangalore](#), and we're expanding in other offices. If you're excited about working on these sorts of problems, [we're hiring!](#)

Looking Forward to 2020 and Beyond

The past decade has seen remarkable advances in the fields of machine learning and computer science, where we now have given computers the ability to see, hear and understand language better than ever before (see [a nice overview of important advances of the last decade](#)). In our pockets, we now have sophisticated computing devices that can use these capabilities to better

help us accomplish a multitude of tasks in our daily lives. We have substantially redesigned our computing platforms around these machine learning approaches by developing specialized hardware, giving us the ability to tackle ever larger problems. This has changed how we think about computing devices both in data centers (such as the inference-focused [TPUv1](#) and the training-and-inference focused [TPUv2](#) and [TPUv3](#)), as well as in low-power mobile environments (such as [Edge TPUs](#)). The [deep learning revolution will continue to reshape how we think about computing and computers](#).

At the same time, there are a huge number of unanswered questions and unsolved problems. Some directions and questions that we are excited about tackling in 2020 and beyond are:

- ***How can we build machine learning systems that can handle millions of tasks, and that can learn to successfully accomplish new tasks automatically?*** Currently, we're mostly training separate machine models for each new task, starting from scratch, or at best, from a model trained on one or a few highly related tasks. As such, the models we train are really good at one or a few things, but not good at anything else. However, what we truly want are models that are good at leveraging their expertise at doing many things, so that they are able to learn to do a new thing with relatively little training data and computation. This is a true grand challenge which will require expertise and advances in many areas spanning solid-state circuit design, computer architecture, ML-focused compilers, distributed systems, machine learning algorithms and domain experts across many other fields in order to build systems that can generalize to solve new tasks independently across a full range of application areas.
- ***How can we advance the state-of-the-art in important areas of artificial intelligence research like avoiding bias, increasing interpretability & understandability, improving privacy and ensuring safety?*** Advances in these areas are going to be critical as we use machine learning in more and more ways in society.
- ***How can we apply computation and machine learning to make advances in important new areas of science?*** There are important advances to be had by collaborating with experts in other fields in areas like climate science, healthcare, bioinformatics and many other areas.
- ***How can we ensure that the ideas and directions pursued by the machine learning and computer science research communities are put forth and explored by a diverse group of researchers?*** The work that the computer science and machine learning research communities are pursuing has broad implications for billions of people, and we want the set of researchers doing this work to represent the experiences, perspectives, concerns and creative enthusiasm of all the people of the world. How can we best support new researchers from diverse backgrounds entering the field?

Overall, 2019 was a very exciting year for research at Google and in the broader research community. We're excited about tackling the research challenges ahead of us in 2020 and beyond, and we look forward to sharing our progress with you!



[Google](#) · [Privacy](#) · [Terms](#)