



WHITE PAPER

Artificial Intelligence: 10 Predictions for 2019

Published 4Q 2018

Published in Partnership with:



ADITYA KAUL
Research Director

CLINT WHEELOCK
Managing Director

INTRODUCTION

The AI market is possibly one of the toughest to keep track of, as the pace of change is relentless. Making predictions is even harder for this market, as one tends to follow what is known as [Amara's Law](#), overestimating the potential of a technology in the short term and underestimating it in the longer term. Nevertheless, Tractica has identified 10 key predictions that cover various aspects of the ever-evolving AI market, based on our ongoing research and analysis including extensive primary research and interviews.

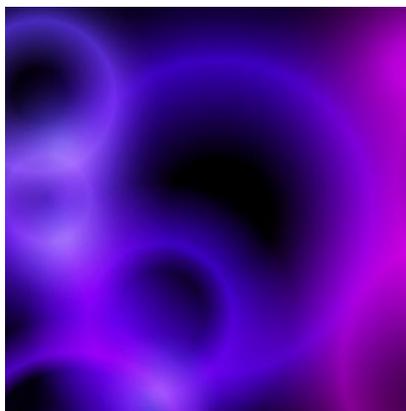


REINFORCEMENT LEARNING WILL SEE GREATER ADOPTION IN THE ENTERPRISE

Reinforcement learning (RL) has seen limited application within the enterprise so far, unlike its popularity in academic research circles. Within enterprise, RL has been applied in the manufacturing, energy, building automation, and automotive sectors, mostly related to optimizing and improving the performance of autonomous machines or control systems, which can be simulated in a RL-infused environment. Going into 2019, there are some reasons to be optimistic about RL and its application moving beyond academic papers and into enterprise environments.

One reason for this optimism is the announcement of [Horizon](#) by Facebook. Horizon is an open-source RL toolkit that extends the applicability of RL into a large-scale AI production environment where RL models can be initiated using offline data, after which they can be integrated into a live environment where models are trained in a constant feedback loop. Horizon has been used internally to improve several of its applications from managing the quality of 360° video to filtering suggestions for its Messenger application M.

Horizon has not just automated policy optimization at Facebook, but has also allowed it to use a better RL-based Deep Q-Network (DQN)-based model to improve performance of notifications without sacrificing quality. Horizon's ability to use RL to support "live AI models" is likely to be replicated across other platforms and is a great application of RL. Tractica expects to see more innovative uses of RL within the enterprise domain, which is part of a larger trend of starting to use unsupervised learning or semi-supervised learning.



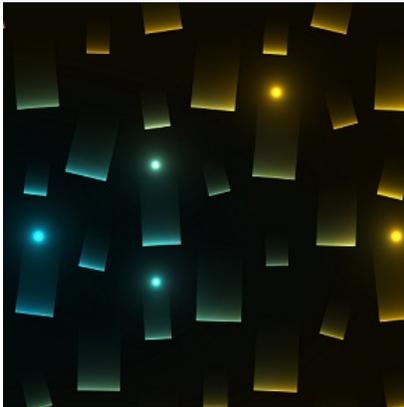
META LEARNING WILL SEE BABY STEPS IN THE ENTERPRISE DOMAIN

Meta learning is about AI learning to learn. In the context of deep learning or machine learning, it is about having meta AI or top-level AI improving the fundamental model, algorithm, architecture, or hyperparameters of targeted AI.

The field of [neuroevolution](#), which finds the most optimal neural network architecture using evolutionary algorithms, is one technique applied in meta learning. Rather than having human AI engineers architect the most optimal neural network, neuroevolution selects the

best network from a set of candidates' networks and breeds the best candidates to create a superset of ideal networks for a specific task. Google has already applied this to an [evolutionary image classifier algorithm](#) to create AmoebaNet, which is the state-of-the-art model for image classification. Other evolutionary meta learning models have emerged, such as [DARTS](#), [ENAS](#), and [NASNet](#).

While the output models are extremely fast and hardware efficient, the computational resources required for finding these new models are unscalable. For example, AmoebaNet took 3150 graphics processing unit (GPU) days and NASNet took 1800 GPU days. DARTS was much more efficient with 4 GPU days, while ENAS took only 16 GPU hours. So, there is a trend of finding better neural architectures with efficient hardware resources. These neural architecture search capabilities should start to make their way into enterprise AI toolkits and platforms in 2019.



ON-PREMISE DATA CENTER SPENDING FOR AI WILL INCREASE

The dominance of hyperscalers in AI has skewed the AI infrastructure market toward the cloud as internet giants like Google, Amazon, and Facebook continue to invest in building cloud infrastructure that supports “AI-first” business models both internally and externally. The rise of AI on the back of hyperscalers has impacted the server market, which has seen a dramatic rise of white box suppliers who serve hyperscalers, which has led to [traditional vendors like HPE, Dell, IBM, and Cisco feeling the pressure](#).

While hyperscaler spending on building AI infrastructure in the cloud should continue into 2019 and beyond, the enterprise market, especially large enterprises, is beginning to ramp up around on-premise data center spending. [Nvidia has mentioned an increase in on-premise deployments](#) of its GPU solutions with companies like SAP increasing their focus on AI as they drive enterprise implementations of AI into verticals like finance and healthcare where cloud-based implementations are the less-preferred option due to data security and privacy issues.

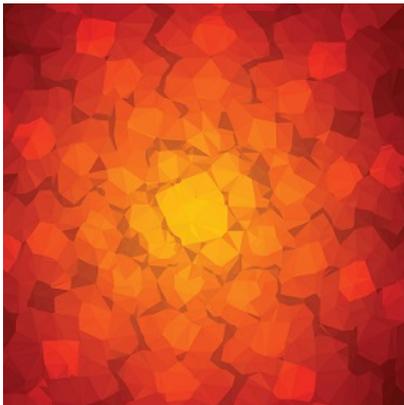
In 2019, we will see many more enterprises across multiple domains, both large and small, start to transition from proof of concept (PoC) to the live deployment phase. Therefore, expect to see more spending on collocated and on-premise servers and workstations that allow enterprises to control, build, and deploy their own their AI models. The proliferation of AI across different types and scales of enterprises tends to favor a much richer ecosystem of solutions, which includes on-premise and cloud solutions.



GPUS WILL SEE REAL COMPETITION EMERGE IN AI HARDWARE

The AI hardware market is heating up, especially for compute solutions, as GPUs start to face competition from offerings like Google's TPU, Intel Spring Crest (Nervana), Huawei's Ascend 910, Graphcore IPU, Amazon's Inferentia, and field programmable gate arrays (FPGAs). Many of these chips have yet to be released, but 2019 is the year when we will see Intel, Graphcore, Huawei, and Amazon all release their chips into the wider market. GPUs have been the workhorse for AI training, driving much of the 300,000X increase in compute capacity in the last 6 years. However, as GPUs see competition in training, Nvidia has plans to expand its capabilities into the inference market, where central processing units (CPUs) have dominated up to this point. Tractica expects to see increasingly hard-fought battles between GPUs and the rest of the chipsets, both in training and inference.

Over time, [Tractica expects custom chips or application-specific integrated circuits \(ASICs\)](#) to lead the deep learning chipset market, as AI processing becomes more decentralized and end users look for better price versus performance or power versus performance alternatives to GPUs. The competition will get real in 2019 for GPUs and companies like Nvidia.



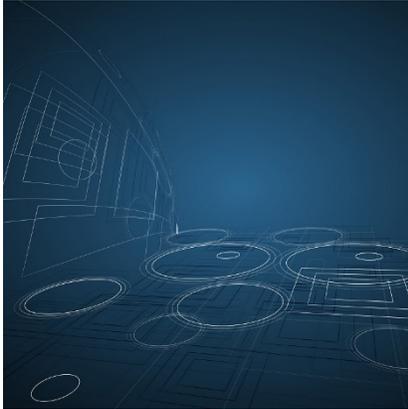
AI-ENABLED PHARMACEUTICALS WILL SEE MULTIPLE CANDIDATES REACHING EARLY TRIAL PHASE

Drug development in its current form is broken, especially when you pit it against the enormous challenge of finding cures to chronic conditions like cancer or other rare genetic conditions that impact millions of people across this planet. Pharmaceutical drug discovery today is a 10 to 12 year process, with the [average cost of a drug running as much as \\$2 to \\$3 billion](#). This is unsustainable, even if we were to limit it to one disease like cancer, with [barely 5% of the 500 cancer drug targets](#) having been met.

The use of AI in drug discovery is one of the most promising areas of AI application, as machine learning and deep learning techniques can take vast amounts of clinical data, image scans, and molecular data and then find correlations in multidimensional spaces, something that human brains cannot begin to comprehend. As with many innovations in AI, the majority of the innovation is coming from startups, with more than 100 startups focused on AI for drug discovery.

One of the top startups in this area is [Recursion Pharmaceuticals](#), who calls this new approach to drug discovery "new radical empiricism," which takes advantage of big data, large-scale automation, and AI to not just narrowly focus on one disease and one hypothesis, but to take a multifold approach to many experiments, many data sources, many diseases, and many treatments. Rather than follow a reductionist approach of uncovering a single drug for a disease and spending years looking for the answer, as was done in the past, companies like Recursion can now simply collect data, model data, analyze results, and repeat until they find the right solutions. This process is already yielding results, with Recursion announcing

its [first AI-enabled drug candidate going into Phase 1 clinical trials in 2018](#), and more than 30 candidates expected to come into clinical trial phase in the 2019 to 2020 timeframe. Recursion can run 250,000 experiments every week at the current scale and has hundreds of disease candidates that it is looking to target. Other startups like [Atomwise](#), [BenevolentAI](#), [Exscientia](#), and [Insilico Medicine](#), among many others, are all pushing for similar breakthroughs in 2019. At this pace, the first AI-enabled drug should hit the shelves within the 2020 to 2021 timeframe.



AUTONOMOUS RETAIL WILL REPLACE AUTONOMOUS VEHICLES IN THE NEW HYPE CYCLE

The launch of the Amazon Go store in 2016 kicked off the autonomous retail revolution. As Amazon plans to [launch more than 3,000 stores](#) across the United States and the rest of the world by 2021, the competition is already building up.

AI is at the heart of autonomous retail using cameras and sensors in combination with deep learning-driven software that recognize people, movement, pose, objects, and actions like picking up and keeping items on shelves. Without the advances in computer vision and deep learning, Amazon Go or any autonomous retail would not exist. The technology behind autonomous retail is commoditized enough to become a platform or a “full stack” solution that one could purchase “off the shelf.”

Autonomous retail startups like [Standard Cognition](#), [Trigo Vision](#), [AiFi](#), and [Zippin](#) are providing solutions for traditional retailers to deploy within their brick and mortar stores, while others like [Inokyo](#) and China’s [BingoBox](#) are going independent and competing on their own with companies like Amazon and traditional retailers. China is pushing ahead with lots of activity in autonomous retail with [JD.com](#) (7Fresh), [Alibaba](#) (Tao Cafe), [Tencent](#) (EasyGo), and [Huawei](#) all starting to push solutions in the area. BingoBox is known to have 300 stores already operating in China, far ahead of Amazon Go’s current rollout. In China, many of the autonomous store solutions use radio frequency identification (RFID) trackers, rather than AI-based computer vision. However, RFID trackers are known to be giving way to AI-based computer vision, as they offer improved product and customer tracking capabilities.

Autonomous retail or cashierless checkout technology will be in the news during 2019, as retailers and grocery chains start to experiment with the technology, launch dedicated cashierless sections within existing stores, and launch new autonomous retail outlets, as traditional retail starts to face up to competition from Amazon and the large internet companies in China. The venture capital (VC) investment in autonomous retail is extremely small compared to autonomous cars, however, 2019 should see much more investment pouring in with possible acquisitions even, as the technology moves onto the hype cycle.



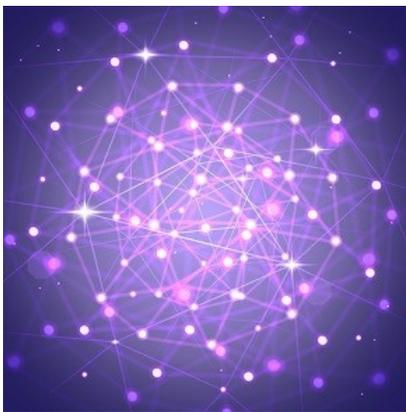
THE FIRST MAJOR AI-BASED CYBERATTACK WILL OCCUR

AI is the new mantra in cybersecurity, with deep learning and machine learning becoming commonplace across cybersecurity threat detection. While AI is being used as a defensive tool to identify and tag new and emerging threats that are rapidly changing, AI can also be used to come up with new threats and bypass security systems as an offensive tool. In 2018, we have already seen some sophisticated attacks like [Trickbot](#) and [Doppelgangers](#), where malware has stealth tactics, obfuscation techniques, and locking mechanisms that are currently hard coded but, in the future, could use AI and machine

learning to perform a “wait and watch” maneuver, staying dormant on a system and attacking at an appropriate time by learning the switches, ports, and channels from which to find back doors or simply transmit and receive data. Some of these techniques have already been seen in nation-state sponsored attacks like the infamous [Stuxnet](#) attack that crippled Iran’s nuclear facilities.

Phishing emails are one of the most common forms of transmitting malware into an enterprise, where an unsuspecting employee clicks on a phishing email that then allows malware to be downloaded onto the enterprise’s information technology (IT) system. Natural language processing (NLP) techniques are becoming much more sophisticated at natural language generation, with AI generating ultrarealistic email text with language that is indistinguishable from a genuine human. Why stop at emails when you can create [malware inducing chatbots](#) that can trick users into clicking. AI-enabled phishing and rogue chatbots are a genuine threat that could exacerbate matters in 2019 and beyond.

Some of these techniques and code are becoming generally available either commercially or through open-source websites like [arXiv](#) and [GitHub](#). These AI techniques could easily fall into the hands of criminal gangs or rogue nation-states and be used to create sophisticated cyberattacks. [AI-based adversarial attacks](#) on vision systems is one example of freely available code being misused. The more likely scenario is a nation-state developing AI-based threats and using it as a deterrent, similar to nuclear weapons. [Small proxy cyberwars](#) between nation-states are already a reality today, with disinformation campaigns being the most commonly used weapon of choice, which is another area where AI could worsen the situation. Nation-states are much more likely to possess and cultivate AI talent, or use proxies to create offensive cyberthreats and tactics, which could be used against another nation-state or even a large corporation. Unfortunately, the possibility of a major cyber-offense operation enabled by AI is only going to grow in 2019.

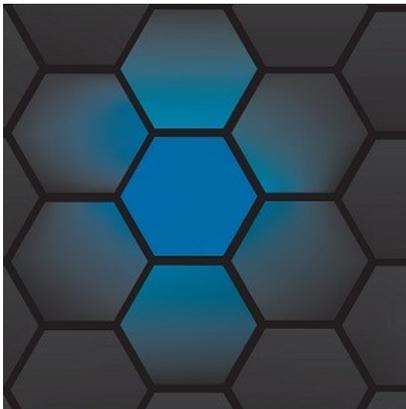


GOOGLE DUPLEX WILL BE AVAILABLE ON YOUR SMARTPHONE

The smartphone has been the ideal vehicle for consumers to experience the power of AI, including face unlocking, photo enhancements, image recognition, face recognition, personal voice assistants, and predictive typing, among other applications. As the smartphone gains much more horsepower in terms of running AI models on-device, we are likely to see many more use cases emerge. One that is likely to emerge in the near term and that could become more than just a gimmicky feature is that of “customized human-like voice assistants.” The [Google Duplex demo](#) showed how an AI system

could be easily mistaken for a human, with the right intonations, pauses, and hesitations. Google Duplex is meant to work as an assistant to make reservations on your behalf, and it has already been launched in beta to a select group of Pixel owners in select U.S. cities. In 2019, this is likely to expand, although expect Google to keep it a closed launch, rather than open to everyone. Also expect to see a few additional uses of Duplex on smartphones, such as answering phone calls on your behalf and having an “enhanced AI voicemail” service that has a short conversation with your friend to understand the reason for the call.

However, beyond Duplex, the technology of generating human-like voices is certainly gaining traction. Text-to-speech technology is becoming much more powerful with [Baidu](#) and [Google](#) both recently announcing new algorithms. However, the ability to train a voice assistant that mimics your voice is also becoming a reality. Baidu has already claimed to be able to [mimic a human voice with 1 minute of training](#), but 2019 could be the year when Google announces something similar, powered by Duplex technology. This would also open the market for synthesized human-like voices for audiobooks or voice assistants, which one could download, adjust, or even train on the smartphone. The human voice will no longer be human!



“AI-FIRST” BUSINESS MODEL WILL GIVE WAY TO “PRIVACY-FIRST AI”

Google’s CEO, Sundar Pichai, declared in 2017 that Google is now an “AI-first” company where AI is at the heart of the products and solutions, allowing them to drive AI from the bottom up, as they improve conversational, multi-device, contextual, learning, and adaptive capabilities into their products. Since then, we have seen AI-first becoming both a mantra and a disruptive challenge across boardrooms and PowerPoint presentations.

Fast forward to the end of 2018 when the mood has shifted considerably as hyperscaler companies like Facebook, Google, and Amazon are being questioned about their data practices in their pursuit of being AI-first. Both Facebook and Google have had to provide testimony to the U.S. Congress about how they collect, share, and use data to run their AI engines. Facebook particularly has a lot to answer in terms of how it handles user data, and why it has been [sharing data with partners without informing users](#). Apple’s CEO Tim Cook has called the AI-first models, which are largely ad-based, a “[‘data-industrial’ complex which threatens both democracy and privacy](#).” There are [growing indications](#) now that the U.S. government will follow the European Union (EU) in passing a “privacy protection” law along similar lines as the European General Data Protection Regulation (GDPR) ruling. Data monopolies are under threat and 2019 will see further erosion of the data hoarding power.

There is also a growing movement around Web 3.0 and the decentralized web, which is being built using Ethereum and blockchain technology. This is closely tied to the decentralized data exchanges and privacy-first movement, which is being enabled by organizations like [Ocean Protocol](#) and by powerful venture capital advocates like Chris Dixon in his essay “[Why Decentralization Matters](#).”

It is clear that the Web 2.0 phase is coming to an end, the centralized data monopoly architectures have run their course, and we are now in open territory. AI is clearly being increasingly embedded into the fabric of the internet, whether it is in the cloud or on the device. And so, the next phase of the internet will be driven by AI, but in a decentralized fashion. Blockchain (and Ethereum) is the one technology that can help build the foundation for this new era. The hype around cryptocurrency and Bitcoin should start dying out in 2019 and the real-world applications will start to see deployments. Projects like [SingularityNET](#), which bring together AI and the decentralized blockchain movement, should gain traction in 2019, although enterprise acceptance might take longer.



AI HYPER NATIONALISM WILL THREATEN GLOBAL COOPERATION

Ian Hogarth's essay on [AI Nationalism](#) really caught the attention of many people in 2018, as it describes how AI and machine learning will become the driving force of the new geopolitics and that we are entering a dangerous phase. On one hand, AI is seen as the spark that would propel countries like China into science and technology superpowers. On the other hand, as [Stephen Fry describes AI](#), it is like the fire that Greek gods discovered that could both destroy and heal, and that we should be careful in how we wield this "fire" power. The fire that destroys has ample fuel to fan it. Governments and policy

makers need to pay attention to and be wary of the threat of AI displacing jobs, AI decisions being biased against certain racial groups or sexes, AI creating further inequality, AI being used as a mass surveillance tool, AI creating further data monopolies, and AI leading to a new arms race, especially one that puts autonomous weapons into the hands of rogue nation-states.

In 2019, we will see increasing national protectionism with regard to AI and data storage but also technology transfer, mergers, and acquisitions. Countries like [India have already called on new restrictions](#) around hyperscaler companies and cross-border data flows. The [U.S. government has blocked mergers](#) like Broadcom taking over Qualcomm. Also, [Huawei's recent troubles with the United States and Canada](#) point toward emerging geopolitics that will be dictated by technology domination.

AI is currently being driven on a global level by three models:

- The United States is mostly being controlled by the hyperscaler data monopolies with the government in the backseat for now.
- Europe has more of a centralized push from government with a robust ecosystem of startups emerging, but the focus is very much ethical and privacy-centric AI.
- China is driven by a close partnership between government (central and regional) and the private sector, where the singular goal is to achieve global AI dominance, and privacy and ethics are secondary.

For now, the Chinese model seems to be winning and has the best odds of achieving the goal of artificial general intelligence (AGI), most likely in our lifetimes. There is also a fourth model that is emerging, as discussed in the previous section around a user-driven, ground-up, and decentralized AI. This model goes against AI nationalism and is more of a global movement that takes AI away from large corporations and governments. This fourth model will become prominent in 2019 and will go head to head with AI hyper nationalism.

ADDITIONAL READING

Tractica's [Artificial Intelligence Advisory Service](#) examines use cases and business models for the application of artificial intelligence technologies in enterprise, consumer, and government markets. Research focus includes analysis of technology trends that are driving the development of more robust cognitive technologies, the use cases and value propositions for artificial intelligence in specific industries, and the key industry players who are shaping the next stage of market evolution. Granular market sizing, segmentation, and forecasting models provide industry participants with an objective assessment of the business opportunity for artificial intelligence software, as well as the hardware and professional services that will enable the software deployments.

Artificial Intelligence Use Cases

258 Use Case Descriptions, Examples, and Market Sizing and Forecasts Across Enterprise, Consumer, and Government Markets

Published 4Q 2018

<https://www.tractica.com/research/artificial-intelligence-market-ecosystem/>

This Tractica report provides a comprehensive examination of more than 250 AI use cases across the consumer, enterprise, and government markets. The report defines, contextualizes, and offers real-world examples and revenue forecasts for each use case, organized by industry sector. It serves as a referential compendium to Tractica's ongoing market forecasting of the AI space, offering an overview and analysis for each use case included in the model. As applications are analyzed in depth, the report also identifies common themes across the broader AI market.

Deep Learning

Enterprise, Consumer, and Government Applications for Deep Learning Software, Hardware, and Services: Market Analysis and Forecasts for 125 Use Cases across 30 Industries

Published 4Q 2018

<https://www.tractica.com/research/deep-learning/>

This Tractica report examines the practical application of deep learning within consumer, enterprise, and government markets. The report provides strategic analysis and market forecasts for 125 deep learning use cases within 30 industry sectors, including an assessment of benefits and implementation considerations along with a quantification of the market opportunity for each use case. Market forecasts include deep learning software, hardware, and services revenue, segmented by world region, for the period from 2017 through 2025. The report also includes profiles of 28 key players within the evolving deep learning market ecosystem.

Artificial Intelligence Market Ecosystem

200 Key Players and Industry Participants

Published 4Q 2018

<https://www.tractica.com/research/artificial-intelligence-market-forecasts/>

This Tractica report provides an in-depth examination of the market ecosystem for AI. The ecosystem is subcategorized into multiple segments, which include hyperscalers, professional service providers, application-focused solution providers, AI development platforms, virtual digital assistant platforms, computer vision, natural language processing, deep learning, machine reasoning, strong AI, compute/networking/storage hardware, and chipsets. The report includes detailed profiles of 200 companies and 1,000 are included in the companion Excel databook.

Artificial Intelligence Market Forecasts

294 Consumer, Enterprise, and Government Use Cases for Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, Machine Reasoning, and Strong AI Across 30 Industry Sectors

Published 3Q 2018

<https://www.tractica.com/research/artificial-intelligence-market-forecasts/>

This Tractica report provides a quantitative assessment of the market opportunity for AI across the consumer, enterprise, and government sectors. The report includes market sizing, segmentation, and forecasts for 294 AI use cases distributed across 30 industries, and 165 unique use cases. The market forecasts span the period from 2017 through 2025 and include segmentation by the six fundamental AI technologies: machine learning, deep learning, computer vision, NLP, machine reasoning, and strong AI. Revenue forecasts are further segmented by software, hardware, and services, in addition to segmentation by world region.

Artificial Intelligence for Edge Devices

Edge-Based AI Chipsets and Accelerators for Mobile Phones, Smart Speakers, Head-Mounted Displays, Automotive, PCs/Tablets, Drones, Security Cameras, and Robots: Global Market Analysis and Forecasts

Published 3Q 2018

<https://www.tractica.com/research/artificial-intelligence-for-edge-devices/>

This Tractica report provides a quantitative and qualitative assessment of the market opportunity for AI edge processing across several consumer and enterprise device markets. The device categories include automotive, consumer and enterprise robots, drones, head-mounted displays, mobile phones, PCs/tablets, security cameras, and smart speakers. The report includes segmentation by processor type, power consumption, compute capacity, and training versus inference for each device category, with unit shipment and revenue forecasts for the period from 2017 to 2025.

 **The AI Summit**

AI is already here. It's being implemented in every industry by leading organizations looking to reach unprecedented efficiencies – human capacity truly reimagined.

The AI Summit is the world's first and largest conference & exhibition series to decipher the practical implications of AI for corporate organizations. We showcase the groundbreaking solutions that are transforming business productivity today and will transform organizations in the future.

The AI Summit – as the pioneering show in the AI space – is only open to qualified business leaders, not students or the general public; the quality of AI Summit attendees is simply unrivalled.

For more information, [visit the website](#).

2019 GLOBAL EVENT SCHEDULE

The AI Summit London 2019	June 12-13, 2019
The AI Summit APAC 2019	July 30-31, 2019
The AI Summit Singapore 2019	September 10-12, 2019
The AI Summit San Francisco 2019	September 25-26, 2019
The AI Summit Cape Town 2019	November 12-14, 2019
The AI Summit New York 2019	December 11-12, 2019

AI Summit attendees may receive a 20% discount off standard pricing of Tractica's research reports by contacting Tractica and mentioning the AI Summit. Please see the back page for contact information.

Published 4Q 2018

© 2018 Tractica

5541 Central Avenue, Suite 150

Boulder, CO 80301

Tel: +1.303.248.3000

Email: info@tractica.com

www.tractica.com

This publication is provided by Tractica. This publication may be used only as expressly permitted by license from Tractica and may not otherwise be reproduced, recorded, photocopied, distributed, displayed, modified, extracted, accessed or used without the express written permission of Tractica. Notwithstanding the foregoing, Tractica makes no claim to any Government data and other data obtained from public sources found in this publication (whether or not the owners of such data are noted in this publication). If you do not have a license from Tractica covering this publication, please refrain from accessing or using this publication. Please contact Tractica to obtain a license to this publication.