Why Your Business Needs an AI Platform

Gus Cavanaugh, Sales Engineer, Anaconda

AI is Transformational

Artificial Intelligence is undeniably transforming technology as we know it. Leading tech companies are investing heavily, with Google CEO Sundar Pichai going so far as to claim that AI is more important to humanity than electricity or fire [1]. So far, however, AI's promise has not translated into concrete strategies and applications that most businesses can use to actually generate value.

This paper discusses the economic impact of AI, how some companies are using AI to generate business value today, and how Anaconda Enterprise serves as an AI enablement platform for organizations seeking to harness the power of AI at scale.

Defining AI

While AI often conjures images of killer robots, the core technology behind today's AI advances is only deadly for those who hate math. The recent explosive growth of AI is largely due to a form of machine learning called *deep learning*. Advances in self-driving cars, smart assistants, and superhuman Go performance [2] are all due to improvements in deep learning models.

But what exactly is deep learning? Nailing down a definition often is a source of confusion in itself, because deep learning is basically a new term for an old concept: neural networks, which are collections of mathematical operations performed on arrays of numbers. But the models generated by deep learning techniques are significantly more complex or "deep" than traditional neural networks, and also involve much greater amounts of data.

Over the last few years, the combination of Big Data, open source advances in deep neural networks, and powerful hardware (notably GPUs) have led to breakthroughs in a number of areas, including image classification, speech transcription, and autonomous driving. These advances have brought about tremendous economic implications.

The Economic Impact of AI

Companies poised to harness AI will generate significant financial returns. McKinsey predicts deep learning will generate $3.5-$5.8 trillion in annual economic value in the near future. And while the top line numbers are large, the industry breakdown is especially illustrative [3].

McKinsey believes the biggest gains will come in retail and travel. But even in aerospace and defense (the industry McKinsey predicts will see the lowest impact), McKinsey still forecasts $45 billion in economic impact. For context, this is equivalent to the GDP of Lithuania [4].

That's not to say these gains will be evenly distributed internationally. In a 2018 PWC report, researchers forecast that the greatest gains in AI will come from China. US companies have an advantage today, but there is no guarantee that it will last [5].

Artificial intelligence (AI) has the potential to create value across sectors.
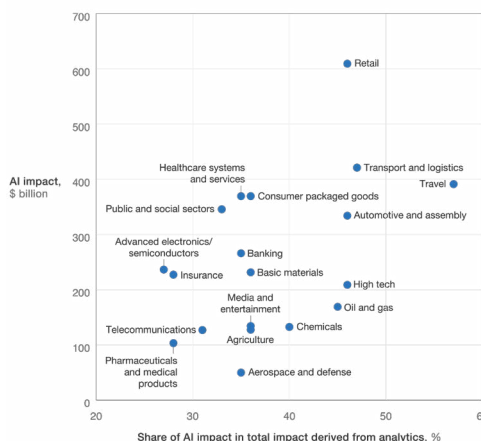
McKinsey&Company | **Source:** McKinsey Global Institute analysis

Furthermore, while AI represents a huge opportunity for your business, keep in mind that it also represents a huge opportunity for your competitors—including companies that might not be your competitors today but soon will be due to advances in AI. According to a 2017 BCG and MIT Sloan report, 75% of companies surveyed believe that AI will enable new competitors to enter their existing market [6].

The message is clear: companies must adopt AI now, or risk getting left behind.

How AI is Being Used

Deep learning has achieved amazing breakthroughs in image classification, speech recognition, handwriting transcription, and text-to-speech conversion, among other areas. That might sound a little abstract, so let's discuss some concrete examples where these advances are practically applied in industry today. In particular, we'll focus on image classification.

Hedge Funds

Hedge funds have invested heavily in machine learning and other advanced analytic techniques, as they are constantly searching for new sources of information to make better trading decisions. Powerful quantitative models serve as core pillars of the hedge business and it's no surprise that funds are early adopters of deep learning.
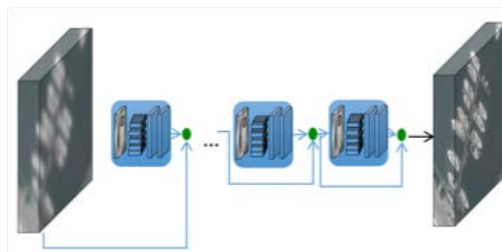

With so much money at stake, hedge funds increasingly are turning to "alternative data" to generate leading indicators of market trends. For example, they can input satellite images into GPU-accelerated neural networks that will estimate everything from the number of ships in a port to the amount of crops growing in a field. By using deep learning techniques to generate higher quality inputs, they can improve the outputs of their existing quantitative models.

Insurance

While the insurance industry has a long quantitative tradition, this heavily regulated, risk-averse field does not have the same association with cutting-edge data science and machine learning as financial trading. Nevertheless, insurance companies are still adopting deep learning techniques in innovative ways.

Let's take homeowner's insurance as an example. The condition of a home's roof is critical to accurately pricing coverage. Traditionally, many insurance companies rely on homeowner-reported roof age to assess roof condition, an approach that is obviously subject to error. But with deep learning, insurance companies can use photographs of a roof to create a deep learning model that will provide a much more accurate representation of the roof's quality. This enables insurance companies to reduce home insurance risk.
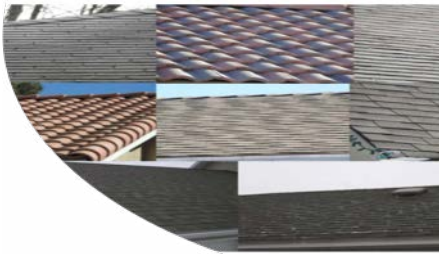
Insurance companies also are using image classification techniques to make the work of insurance adjusters faster and more accurate. Rather than relying on an insurance adjuster to read the odometer of a car that has been in an accident, a deep learning model can ingest a photograph of the odometer and determine the correct reading.

Again, it's important to note that these companies are not throwing away their existing models in favor of entirely new approaches. Instead, they are leveraging new AI techniques to improve the inputs to their models.

Consumer Products

Olay, the multi-billion dollar skin care brand, recently launched a "Skin Advisor" application using neural networks. Prospective customers submit a photograph, and the model returns a customized skin assessment and suggests corrective products for the customer to purchase.

While this might seem like a novelty product, it has had a significant financial impact[7]. Average basket sizes (the amount of goods a customer purchases at one time) are up 40% and conversion rates have doubled. By harnessing AI, Olay is able to offer a personalized shopping experience and customers are responding by buying more products.

## The Key to Harnessing AI in the Enterprise

Success with AI comes down to an organization's ability to build machine learning models at scale and rapidly deploy them.

The concept in this graphic is easy to read but challenging for most companies to implement. And this is a real problem. The majority of companies struggle to even deploy machine learning models, let alone rapidly iterate on this entire cycle.

Today, many companies are comfortable with the beginning stages of the data science process. Data scientists and data engineers are comfortable loading data, training models on a subset of data, scoring models, serializing models, and writing predictor functions. While these activities are by no means easy, they are not sufficient to generate business value. It is simply not good enough to build and train a model on your laptop or development environment and leave it there.

To generate real business impact, models must be deployed. And model deployment and model management are difficult tasks to
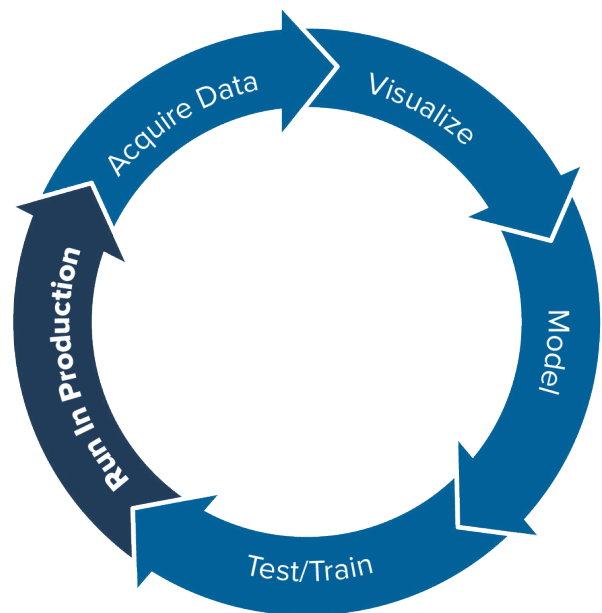

complete, let alone automate. Data scientists who attempt to deploy models find themselves wearing several hats: software developer, IT administrator, security officer. They struggle to wrap their model, serve their model, route traffic and handle load-balancing, and ensure all of this is done securely.

Even if successful, the difficulty of deploying their model discourages iteration. But it is imperative that companies automate the entire data science lifecycle, and crucial to this process are scale and speed.

Scale

The importance of scale in model training and production cannot be overstated. It's not enough to build and train models on a laptop with a subset of data. Data scientists must scale their model training in order to build powerful models. According to OpenAI, we've seen a 10X increase per year in the number of petaflops required to train world-beating AI[8].
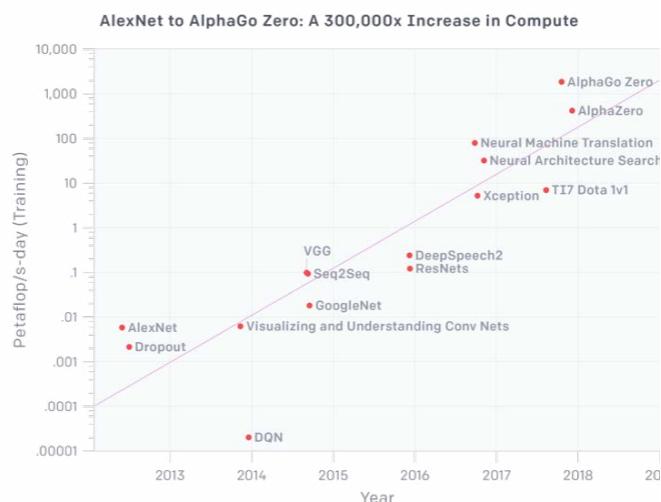
Source (OpenAI)

Whereas previously this level of compute would be the domain of super computers, thanks to GPUs and TPUs, this level of scale in model training is now economically feasible. But deploying these massive levels of compute requires significant supporting infrastructure. Thus today's challenge is not where to find the computing power, but how to manage an environment that supports it.

Speed

It's not just scale in model training that is required to have success with AI. The most accurate model in the world has zero value if it isn't deployed into production. And unfortunately, too many organizations strand their models in PowerPoint slides, PDFs, or on the notebooks of their data scientists. Organizations must rapidly deploy and manage models to provide value to the business.

In our experience, many organizations lack a clear path for data scientists to deploy models into production-like environments. Data scientists build models locally on their laptops but lack the supporting infrastructure to train them at scale or deploy them as REST APIs. Instead they pass their models to software engineering teams or other groups to be re-written and deployed. This is not only

CUSTOMER SPOTLIGHT

nationalgrid

Anaconda Enterprise customer National Grid is a British-based utility company. They chose Anaconda Enterprise to serve as the foundation of their transition to data-driven utility asset management. Previously they relied on a time-based maintenance schedule (think "change the oil every three thousand miles"), which left them blind to costly outages and unscheduled maintenance. With Anaconda Enterprise, they have built machine learning models to predict asset failure, reducing unnecessary maintenance and preventing costly outages.

Even better, once Anaconda Enterprise was installed, National Grid's data scientists found innovative ways to apply it that management did not forsee. For example, one data scientist noticed a group of engineers watching hours of video footage as part of that team's visual inspection of National Grid's powerlines. Specifically, the engineers were looking for the transformers that appear every few miles, and as such they spent most of their time fast-forwarding through the tape until a transformer appeared.

The data scientist took that video footage and used an open source package for object recognition called OpenCV, available within Anaconda Enterprise, to turn hours of mind-numbing footage into just the few critical minutes containing the transformers that the engineers needed to see. Needless to say, the engineers were delighted as this new approach saved them hours of time.

time-consuming but also error-prone, as the changing model dependencies alter the model's output.

There is a solution. Organizations must automate their AI/ML pipelines. Unfortunately, this is easier said than done.

Automating Operational AI is Hard

Leading tech companies like Google, Facebook, and Uber recognize the challenges of automating AI pipelines at scale. While there is tremendous open source innovation in machine learning libraries (Google's own TensorFlow serving as a great example), the open source community has not created tooling to support model development, training, and deployment. In a 2014 report[9], Google AI researchers reflected on the significant supporting infrastructure required to deploy production machine learning systems: "A mature system might end up being (at most) 5% machine learning code and (at least) 95% glue code." This glue code, or infrastructure code, represents a significant barrier to making AI operational.

Uber reported similar challenges[10]. Uber's data scientists could only train their models on their laptops, inhibiting their ability to scale their model training to the data volumes and compute cycles required to build powerful models. Further, Uber did not have a process to store model versions, leading data scientists to waste valuable time recreating the work of their colleagues.

Worst of all, they did not have a standard process for deploying their models into production, thereby severely limiting the potential business value from their work: "...there was no established path to deploying a model into production–in most cases, the relevant engineering team had to create a custom serving container specific to the project at hand."

In order to allow their data scientists to focus on building better models, Google, Uber, Facebook, Twitter, and other leading tech companies built their own AI platforms. These platforms automate the supporting infrastructure, the glue code, if you will, that is required to build, train, and deploy AI at scale.

Of course, we don't advise you to build your own platform. Google has TFX. Uber has Michelangelo. Twitter has Cortex. You have Anaconda Enterprise.

Anaconda Enterprise: An AI Enablement Platform for Teams at Scale

Anaconda Enterprise is an AI enablement platform for teams at scale. Organizations understand the business value that AI can bring and they desperately want to harness these exciting tools before their competitors do. While many companies experiment with AI and ML, few see transformative business value. Data science teams are stuck in silos. Projects are left orphaned in notebooks and inboxes. Business value is but a distant dream, pushed further away when the latest internal cutting-edge model lies dormant in a well-formatted PowerPoint slide.

It doesn't have to be this way. It is true that automating AI is hard. As discussed above, the best tech companies in the world have felt the

exact same pain. They understood the stakes involved and they made the investment of time and money to build their own platforms to automate AI.

But you don't have to do the same. You have Anaconda Enterprise. Anaconda Enterprise automates your organization's AI pipelines from laptops to training clusters to production clusters with ease. Anaconda Enterprise supports your organization no matter the size, scaling from a single user to thousands running on one laptop to thousands of machines. No headaches, no IT nightmares. Anaconda Enterprise automates the undifferentiated heavy-lifting, the 95% glue code that prevents organizations from rapidly training and deploying models at scale.

And that's only the beginning. With Anaconda Enterprise, your data scientists can focus on the areas where they can bring the biggest business impact. Your data scientists say goodbye to DevOps, software engineering, and IT. One platform provides them all of the tools they need to connect, share, and deploy data science projects.

From an IT perspective, Anaconda Enterprise provides automated AI pipelines with zero headaches. IT administrators rest easy knowing that AE's cloud-native architecture makes scaling simple. Security officers are confident that all data science assets—packages, projects, and deployments—are managed securely with appropriate access control configured automatically.

CUSTOMER SPOTLIGHT

Citi uses Anaconda Enterprise in their credit card business to make instant credit-line decisions, making spending convenient for customers. The data science team builds models collaboratively using Anaconda Enterprise while ensuring secure access to the Citi data lake. Citi's machine learning models are used in production to make fast credit decisions, providing improved support to credit card customers.

Anaconda Enterprise is also used for anti-money-laundering. Using Anaconda Enterprise, Citi builds models that identify and flag suspicious transaction patterns in real time. Anaconda Enterprise acts as a secure gateway to data lakes and secures the model development process, which protects Citi from fraud and also meets all regulatory requirements.

Anaconda Enterprise: Closer Look

Anaconda Enterprise is the only platform to combine core AI technologies, governance, and cloud-native architecture. Each piece—core AI, governance, and cloud native—are critical components to enabling organizations to automate AI at speed and scale.
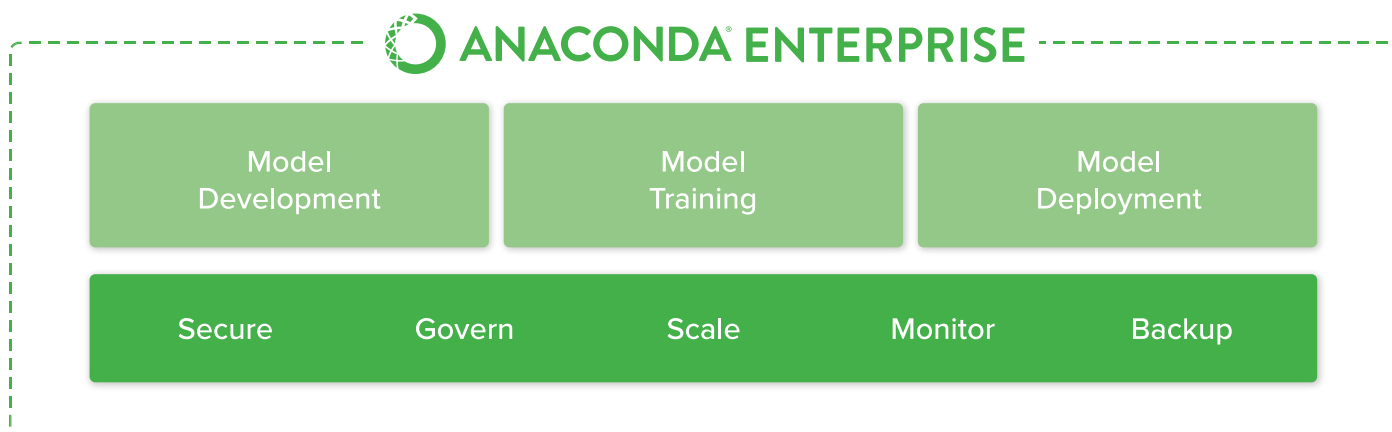
Core AI

The beating heart of Anaconda Enterprise is the collection of core AI/ML tools in Python and R that data scientists use to build models. These tools are provided via Anaconda Distribution, the most popular open source distribution of Python and R for data science. Anaconda Distribution brings powerful open source packages to 6M+ data scientists and developers around the world, allowing them to build and train models on their laptops with ease.

With Anaconda Enterprise, data scientists have access to 1,400+ packages for data science out of the box, hosted internally to ensure access and availability behind your company's firewall. When users login to Anaconda Enterprise, they can choose from example projects and preconfigured project templates that solve common problems and make getting started easy. Never waste time setting up ad-hoc environments. Anaconda Enterprise requires zero setup from users—just login and choose a notebook template, then start importing packages, exploring data, and building models.

Governance

While leading organizations want to harness the power of open source tools for AI/Ml, it is imperative they maintain control over their entire ML pipeline. Basic needs include the ability to track users, projects, and deployments, integrate with enterprise identity providers, and implement role-based access control. But leading organizations know that's only the beginning. IT organizations need full control over their open-source supply chain, including the ability to track entire project lineage from packages to source code to deployment logs.

Anaconda Enterprise enables governance and supports full reproducibility from laptop to server to cluster. This provides peace of of mind, as data scientists will know that they have full control of their model lineage, guaranteeing reproducibility on all model versions, and IT admins will know their data scientists are using approved packages, have access to the compute instances they need,

and are operating in a manner consistent with enterprise IT and security policies.
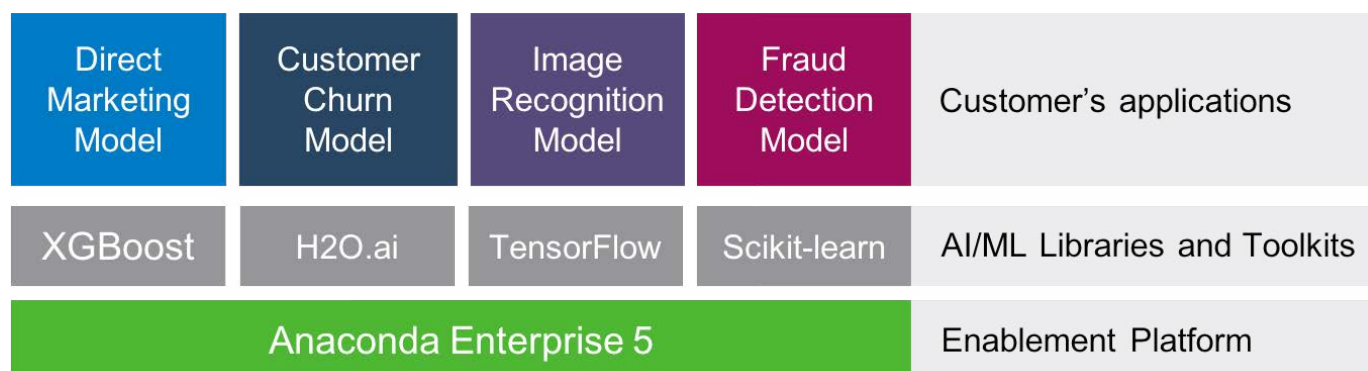
Cloud Native

For organizations interested in building industry-leading models, old-fashioned Java apps and drag and drop GUI-based tools will not suffice. Instead, organizations must adopt cloud native data science. Cloud native applications (which run both on-premises and in the cloud) follow modern, performant, API-oriented application development practices. Anaconda Enterprise supports cloud native data science applications and is itself a cloud native application. This is extremely important because, as we have established, success with AI comes from speed and scale in model training and deployment.

Embracing cloud native data science principles, all Anaconda Enterprise projects and deployments run inside of containers, automatically orchestrated and abstracted from the view of the data scientist. This provides data science users with flexibility and speed: the flexibility to choose the compute requirements necessary for model training or model deployment without requiring any expertise in containerization technology; and the speed that allows data scientists to deploy models as REST APIs with just one click, enabling them to quickly iterate through the machine learning lifecycle and thus provide more value to the business.

Anaconda Enterprise's cloud native approach also makes IT's life easy. With Anaconda Enterprise, IT admins can allow data scientists to self-serve compute needs while maintaining control. Even better, scaling Anaconda Enterprise is simple. IT admins need only run one command to add node(s) to their cluster while Anaconda Enterprise takes care of the rest.

What You Can Build with Anaconda Enterprise

Organizations use Anaconda Enterprise as a platform to provide the core AI, governance, and cloud native scale required to be successful with AI. This means that Anaconda Enterprise enables other AI/ML technologies and toolkits, rather than competes with them. You can easily plug third-party tools on top of Anaconda Enterprise to increase their effectiveness.

Direct Marketing Model	Customer Churn Model	Image Recognition Model	Fraud Detection Model	Customer's applications
XGBoost	H2O.ai	TensorFlow	Scikit-learn	AI/ML Libraries and Toolkits
Anaconda Enterprise 5				Enablement Platform

With Anaconda Enterprise, customers focus on the business problem they want to solve, confident that Anaconda Enterprise supports the tools they need for the job. Take fraud detection models, which are a popular application of AI today. Customers build their fraud detection models with Anaconda Enterprise, using pre-packaged open source toolkits like TensorFlow, Keras, and XGBoost, or third-party tools like H20, DataRobot, and Tableau. Anaconda Enterprise enables rapid data exploration, model development, model training, and model deployment at scale using Anaconda Enterprise's core cloud native infrastructure and whichever tools they prefer.

The Path Forward

Envision a world where data science teams can rapidly iterate on machine learning models, regularly pushing projects into production at scale with full reproducibility. How would that impact your business?

Anaconda Enterprise lets users deploy anything they can create with Anaconda Distribution into a production-like environment with just one click. With Anaconda Enterprise, your organization has a platform on which to build world-class machine learning and AI. Constructed on a foundation of open source, forged by real-world enterprise use, Anaconda Enterprise is the AI enablement platform for teams at scale. If you're interested in harnessing the power of AI, contact one of our ambassadors to schedule a demo.

Endnotes

[1] Clifford, C. (2018, February 1). "Google CEO: A.I. is more important than fire or electricity." *CNBC*. Retrieved from https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html

[2] Silver, D. et al. (2017, October 18). "Mastering the game of Go without human knowledge." *Nature*. Retrieved from https://www.nature.com/articles/nature24270

[3] McKinsey Global Institute. (April 2018). Notes from the AI Frontier: Insights from Hundreds of Use Cases. McKinsey & Company.

[4] The World Bank. Retrieved 14 August 2018 from https://data.worldbank.org/country/Lithuania

[5] PwC. (2017). Sizing the prize: What's the real value of AI for your business and how can you capitalise?

[6] Ransbotham, S. et al. (2017, September 6). "Reshaping Business with Artificial Intelligence: Closing the Gap Between Ambition and Action." MIT Sloan Management Review in collaboration with The Boston Consulting Group. Retrieved from https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/

[7] Marshall, M. (2018, July 19). "How Olay used AI to double its conversion rate." *VentureBeat*. Retrieved from https://venturebeat.com/2018/07/19/how-olay-used-ai-to-double-its-conversion-rate/

[8] Amodei, D. and Hernandez, D. (2018, May 16). "AI and Compute." OpenAI blog. Retrieved from https://blog.openai.com/ai-and-compute/

[9] Sculley, D. et al. (2014). "Machine Learning: The High Interest Credit Card of Technical Debt." Google AI. Retrieved from https://ai.google/research/pubs/pub43146

[10] Herman, J. and Del Balso, M. (2017, September 5). "Meet Michelangelo: Uber's Machine Learning Platform." UBER Engineering blog. Retrieved from https://eng.uber.com/michelangelo/

About Anaconda

With over 6 million users, Anaconda is the world's most popular Python data science platform. Anaconda, Inc. continues to lead open source projects like Anaconda, NumPy, and SciPy that form the foundation of modern data science. Anaconda's flagship product, Anaconda Enterprise, allows organizations to secure, govern, scale, and extend Anaconda to deliver actionable insights that drive businesses and industries forward. Visit www.anaconda.com.