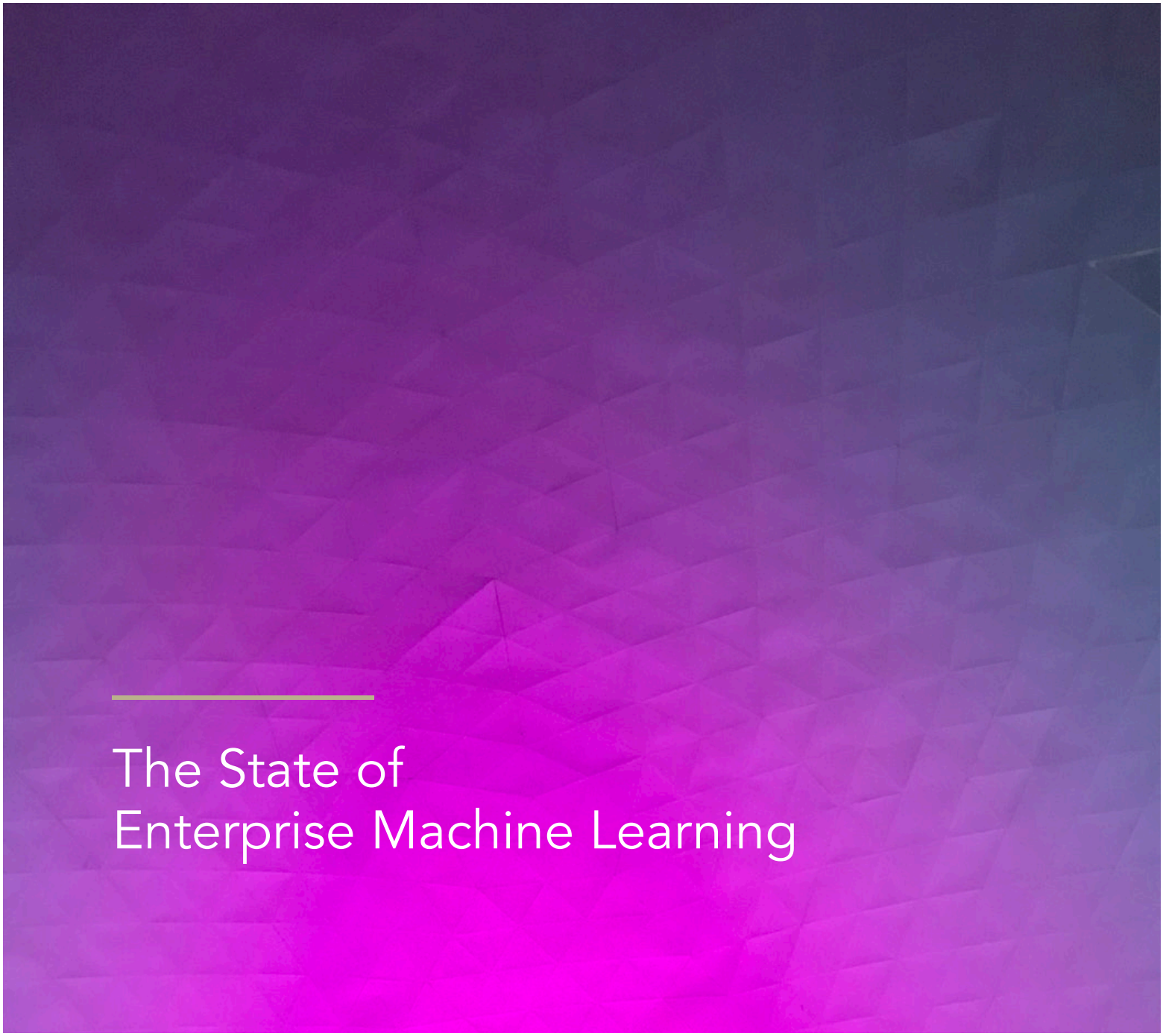
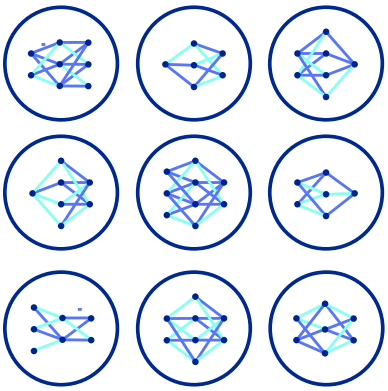
research

October 2018 | Algorithmia Research



The State of Enterprise Machine Learning



This report is the product of Algorithmia Research as part of our efforts to help everyone deploy Machine Learning models. More information can be found here.

Table of Contents

Overview of the Study & Results ··· 4

The Path to Machine Learning ··· 8

Survey Demographics ··· 10

Enterprise Machine Learning Use Cases ··· 12

Challenges Facing Enterprise Machine Learning ··· 13

Big Companies & Custom AI Layers ··· 14

Conclusion ··· 15

Overview

Rapid improvement in compute, data availability, and storage capabilities have precipitated a race to leverage Machine Learning as a competitive advantage. Many leading organizations were caught unprepared by the disruption of the Internet. Artificial Intelligence has the capacity to change business models on an even deeper level. We have seen most organizations are aware of the opportunities and risks which is leading them to proactively face the challenges their business models will likely face.

Artificial Intelligence will augment, automate, or replace most human and business processes.

In October 2018, we set out to understand how companies are reacting to Machine Learning, its adoption in the marketplace, how businesses adapt, and how the industry is evolving. We developed a survey with over 500 decision makers at companies with various employee sizes and industries to explore this topic.

Among surveyed organizations, we found that Machine Learning and Artificial Intelligence are a significant focus of organizations:

1. Data Scientists are facing many roadblocks.
2. Companies are increasing their investment.
3. Large enterprises are taking the lead in this initiative.
4. Machine learning leadership has no central location within an organization to date.
5. There are a broad number of use cases.

This report will go into detail on each of these challenges.

To note, organizations that make a concerted effort to focus on Machine Learning and Artificial Intelligence across their customer lifecycle are more successful. They have a higher propensity for brand loyalty, lower costs of operations, and many other benefits.

Key Finding: Data Scientists are facing many roadblocks

Most Data Science and Machine Learning teams are not able to focus on adding value. They spend the majority of their time on infrastructure, deployment, and data engineering. This leaves less than 25% of their time for training and iterating models.

Only 8% of respondents consider themselves “sophisticated” in their machine learning programs.

Qualitatively, many Data Scientists are fighting existing systems and processes without clear understanding from management.

Key Finding: Companies are quickly increasing their investment in Machine Learning

80% of respondents say their organization's investment in AI/ML has grown by at least 25% in the past 12 months, and this number climbs to 92% in organizations with greater than 10,000 employees.

Key Finding: Big companies are getting a lead

Employees within larger organizations feel significantly more satisfied with their progress than in smaller organizations—they are roughly 300% more likely to consider their model deployment "sophisticated" and 80% more likely to be "satisfied" or "very satisfied" with their progress when compared to people in companies of 500 employees or less.

Budgets are growing faster for organizations that consider themselves sophisticated. 51% of these companies have increased their AI/ML budgets by at least 25% this year.

Key Finding: Companies have not decided where Machine Learning leadership will come from

Overall 37% of respondents say their Machine Learning efforts are being directed primarily by management, while 55% say their efforts are emerging from engineers or other technical teams.

Key Finding: Companies are trying a broad number of use cases

Among enterprises (>10K) the biggest use case is increasing customer loyalty (59%), followed by increasing customer satisfaction (51%), and interacting with customers (48%).

In general, larger and more sophisticated companies noted more use-cases overall than smaller and less mature companies: as you get better at ML, you get smarter about where to apply it and gain clarity.

For larger organizations, cost savings is increasingly important: 43% of 1001-2500 companies put it as a use case, as well as 41% of 2501-10K, and 48% of greater than 10K companies.

The goal of this document is to give you a baseline understanding of the current maturity of the competitive landscape. The data from this survey shows that companies are rapidly maturing and running into common challenges. We hope this helps you navigate our quickly evolving field.

The Path to Machine Learning

Over the last ten years a series of trends has created the technical and operational ingredients for success. It began with a push for Big Data.

The Big Data wave encouraged companies to catalog, store, and clean their data sets. The promise was that a company's data is their biggest asset. Companies like Netflix and Facebook offered proof that leveraging massive data sets made decision making faster and more effective than traditional methods.

The next trend was Data Science. Using data to make decisions is not new. Actuaries and analysts have been using advanced statistics to develop Business Intelligence for a long time, but the term Data Science came to prominence when it became clear that leveraging data effectively is a huge advantage.

The world was introduced to a popular application of Data Science by author Michael Lewis in his 2003 book, *Moneyball*. The book detailed a strategy used by the Oakland Athletics to utilize data to make pragmatic decisions that went against the traditional wisdom of baseball clubs. They were able to outcompete their rivals on a shoestring budget. This was followed by the 2008 data driven political campaign by the Barack Obama campaign. Both pop culture events were very public, in some of the most competitive traditional fields, and showed the decisive victories that can come when you mix lots of data and smart people.

Big Data and *Data Science* provided the infrastructure, data, and talent to allow Machine Learning to become the next trend. Machine Learning frameworks have quickly evolved, giving algorithms access to solving a large set of problems that, until now were much more suited for humans.

Today Machine Learning is powering the legal discovery process for top law firms, is effectively reducing fraudulent credit card orders, and is even checking the grammar of this sentence as I write it. Machine Learning in production is quickly becoming a main objective for organizations across the world.

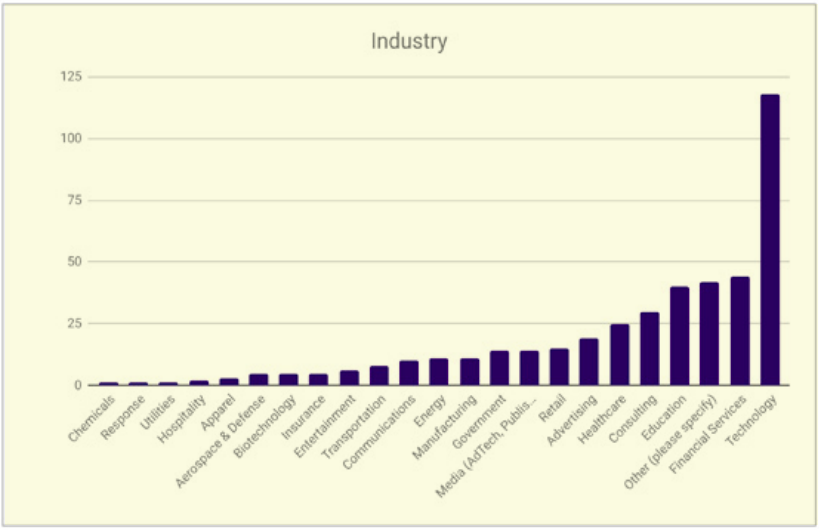
Needed: Machine Learning Infrastructure and Best Practices

In qualitative answers in the survey we see a common theme—productionizing and managing Machine Learning Models is a challenge most companies are currently dealing with. Data Scientists are often spread through an organization, so they do not have the infrastructure or support to put their own work into production, and there are not yet best practices.

Machine Learning is benefiting from massive investments of time, money, and focus—however, the results and maturity of these investments vary greatly. This report gives the first known look into the actual business progress being made across industries and company sizes. This report should help you understand the landscape of Machine Learning progress within companies.

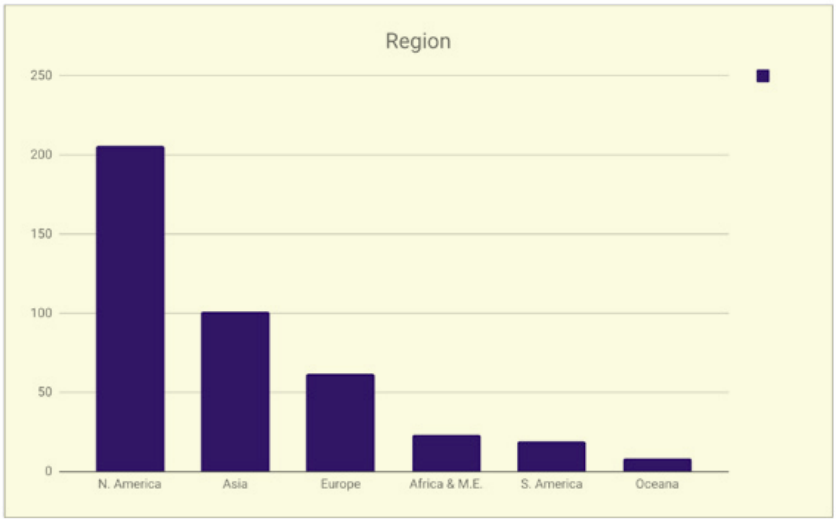
Demographics

Data Science touches many parts of an organization—our survey was able to get input from a wide variety of roles and levels of seniority. When notable, we call out the differences between roles and levels of seniority. This survey was conducted over the summer of 2018 and responses were received from 523 professionals working in Data Science and Machine Learning.



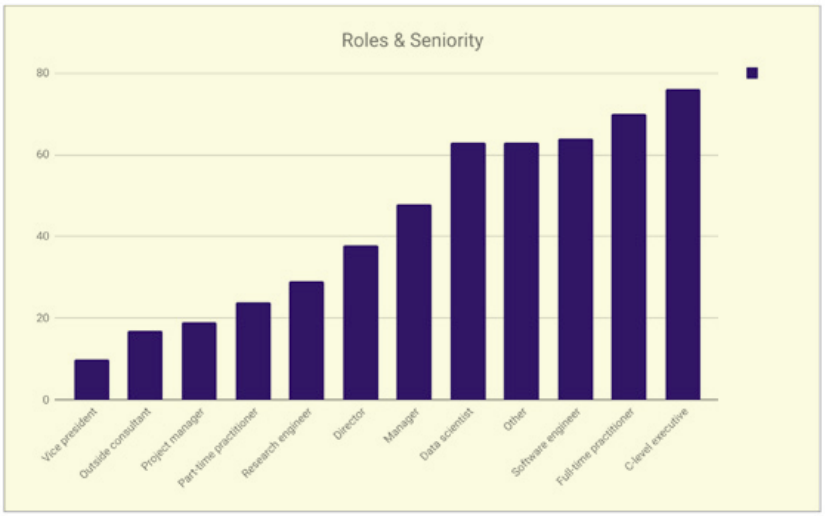
Geographic Location of Organization

Our survey received input from professionals around the globe with a majority coming from North America.



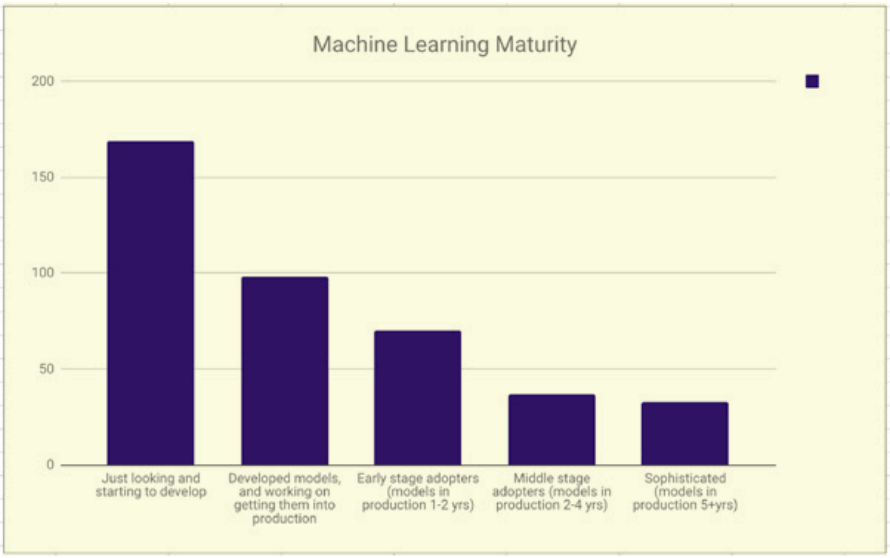
Roles & Seniority

Our survey was able to target a broad swath of roles and seniority levels.



Maturity

Roughly 40% of organizations are just beginning their Machine Learning efforts, and less than 10% consider themselves “sophisticated.” This leaves about half of organizations in the middle of their roadmap.



Of those productionizing models, we see about half are in the first two years of their efforts.

Machine Learning Use Cases

The most common use cases are focused on better understanding and engaging customers.

Big focus on the customer:

Generating customer insights and intelligence (#1), improving the customer experience (#2), interacting with customers (#5), increasing customer satisfaction (#6), and retaining customers (#7).

Among enterprises (>10K), the biggest use case is increasing customer loyalty (59%), followed by increasing customer satisfaction (51%), and interacting with customers (48%).

Larger organizations are putting significant efforts into using Data Science to identify areas of cost savings.

For larger organizations, cost savings is increasingly important: 43% of 1001-2500 companies put it as a use case, as well as 41% of 2501-10K and 48% of greater than 10K companies.

The focus on reducing costs is higher among sophisticated adopters

Larger and more sophisticated companies are investing more, and across a broader number of use cases.

In general, larger and more sophisticated companies filled in more use cases overall than smaller and less mature companies: as you get better at ML, you get smarter about where to apply it and gain clarity.

Challenges

- The biggest challenge faced by organizations is in productionizing Machine Learning models.
- 38% report difficulty in deploying models to the needed scale. Anecdotally, in the survey, we see many reasons for this including: DevOps or IT teams not having the resources, Data Scientists being expected to build the infrastructure to put their models into production, and a lack of existing infrastructure within the organization to support the unique needs of running ML models at scale.
- 30% report challenges in supporting different programming languages and training frameworks.
- Machine Learning models are created using a number of different programming languages and training frameworks. This adds an additional level of complexity because several models, written in distinct languages and frameworks, often must be pipelined together for a given task. Large companies like Facebook and Uber have solved some of these issues by building large internal tools like FBLearner and Michelangelo, respectively.
- 30% report challenges in model management tasks such as versioning and reproducibility.

Big Companies and Custom AI Layers

The biggest tech companies have a large headstart over the competition. They pioneered data collection, data engineering, and Machine Learning frameworks. Most recently we see a new category of infrastructure being built in house by the big companies. Here are some examples:

- Facebook has FBLearner
- Google has TFX
- Netflix has Notebook Data Platform
- Uber has Michelangelo
- Twitter has Cortex
- Airbnb has BigHead

We call this category of infrastructure as the "AI Layer." This infrastructure manages compute loads, automates deployment for ML models, and gives tools for managing ML portfolios across the organization.

Conclusion

In 2018, companies are in the 'pioneering' stage of Machine Learning. Meaning best practices are still being sorted out, Machine Learning is mostly looked at as an R&D effort, and the infrastructure required for productionizing models is not yet accessible.

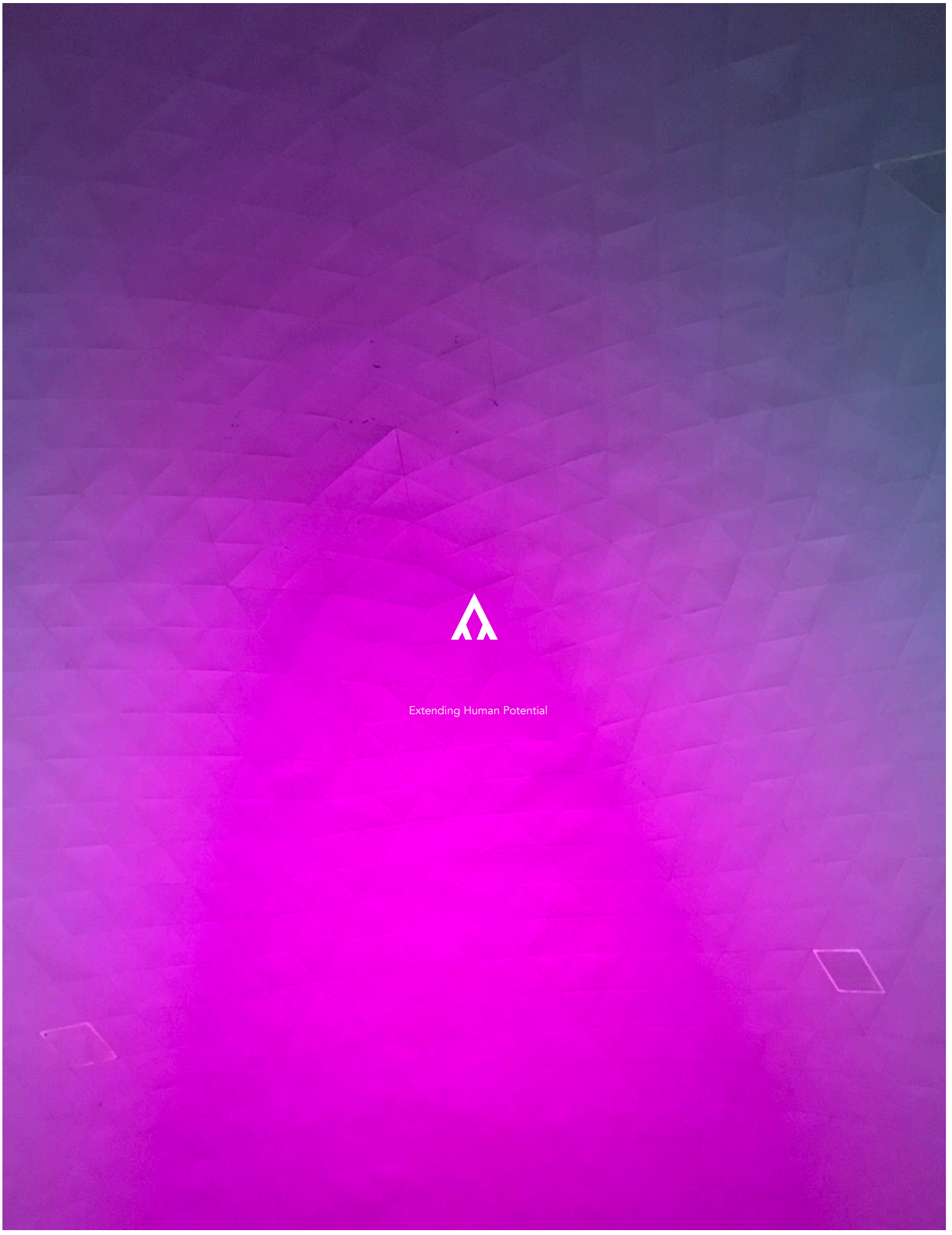
Big companies have an advantage because they have access to more data, can continue to invest in large R&D efforts, and have many problems that Machine Learning can solve cost effectively.

Throughout each step of maturity on the path to Enterprise Machine Learning the big tech companies have built custom infrastructure that then proliferates to the entire market. It began with Google's custom file structure maturing into Hadoop, and has provided the market with countless data cleaning and model training solutions.

We expect 2019 and 2020 to see large leaps in productionized Machine Learning as Data Scientists can easily deploy and manage their models.

About Algorithmia

Algorithmia is the leader in Machine Learning deployment. The Algorithmia AI Layer deploys and manages models in pioneering Fortune 100 companies, as well as US Intelligence Agencies and the United Nations. Algorithmia.com has nearly 80,000 engineers and Data Scientists deploying models on the AI Layer.


Extending Human Potential