

Great Power, Great Responsibility: The 2018 Big Data & AI Landscape

Published on June 27, 2018



Matt Turck | + Follow
Managing Director at FirstMark
14 articles

 3,472 88 754

It's been an exciting, but complex year in the data world.

Messaging



Just as [last year](#), the data tech ecosystem has continued to “fire on all cylinders”. If nothing else, data is probably even more front and center in 2018, in both business and personal conversations. Some of the reasons, however, have changed.

On the one hand, data technologies (Big Data, data science, machine learning, AI) continue their march forward, becoming ever more efficient, and also more widely adopted in businesses around the world. It is no accident that one of the key themes in the corporate world in 2018 so far has been “[digital transformation](#)”. The term may feel quaint to some (“isn’t that what’s been happening for the last 25 years?”), but it reflects that many of the more traditional industries and companies are now fully engaged into their journey to become truly data-driven.

On the other hand, a much broader cross-section of the public has become aware of the pitfalls of data. Whether it is through the very public debate over the [risks of AI](#), the Cambridge Analytica scandal, the massive Equifax data breach, GDPR-related privacy discussions or reports of growing [government surveillance in China](#), the data world has started revealing some [darker, scarier undertones](#).

Both are the flipside of the same phenomenon, which has been brewing for many years but is now in full display: just about everything (whether personal or professional) is rapidly getting digitized, and data technologies are becoming more adept than ever at processing and analyzing this massive data exhaust, increasingly in real time. From this can result **both magic and abuse**. The debate on how to combine this great power with a necessary sense of responsibility has become essential.

Let’s highlight some of the key trends and events of 2018.

Infrastructure & Analytics

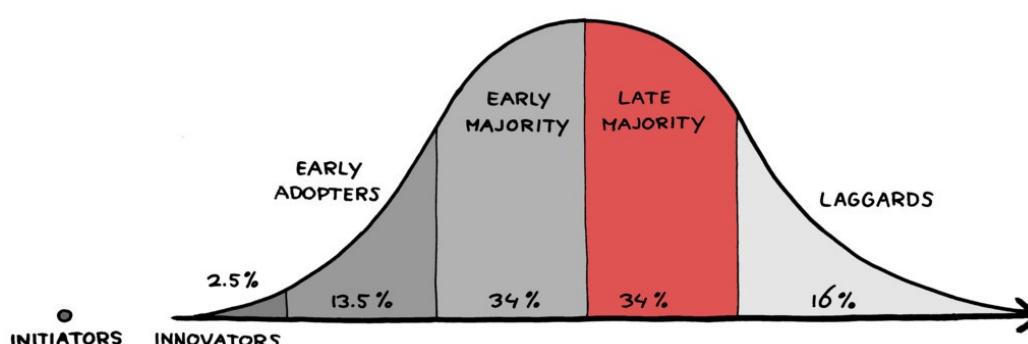
From an industry standpoint, the data ecosystem remains as exciting and vibrant as ever, with a rich tapestry of innovative startups, mature “scale-ups”, and many aggressive public technology vendors. Most importantly, many customers large and small are [deploying t](#)

Messaging



technologies in production at scale, and reaping undeniable value from their efforts.

As the cycle of replacing older IT technologies with more modern data products continues, it seems that the Big Data market (infrastructure, analytics) is cycling through the early majority of buyers and transitioning into the late majority of the traditional adoption curve.



In addition, the data world continues its inexorable evolution towards the cloud. It is actually staggering to see how fast the large public cloud providers (AWS, Azure, Google Cloud Platform, IBM) are growing, considering they already each generate billions of dollars of revenues every quarter. The trend raises ongoing concerns around **vendor lock-in**, and this may open up opportunities for startups offering **multi-cloud** solutions. However, to date even companies adopting multi-cloud strategies tend to still rely on one vendor as their primary provider.

As they keep growing, large cloud providers increasingly compete with each other by offering a wide array of Big Data, data engineering and machine learning tools through their platforms (e.g., Amazon Neptune, Google AutoML, etc.) - and often with aggressive pricing, to attract more developers, as their true business model is data storage. As the scope and sophistication of such tools keep growing, this has a big impact on the data technology landscape, making it arguably **harder for startups to compete**, at least for broad, horizontal opportunities. A bit more every year, the list of product announcements at big annual cloud vendor conferences (see AWS [re:Invent](#), for example) sends shockwaves in the startup industry, as they put cloud vendors in direct competition with dozens of VC-backed star

in one fell swoop. It will be interesting to see how public markets react to the upcoming Elastic IPO, an open-source software company that saw Amazon launch a direct competitor, Elasticsearch, three years ago.

Plenty of opportunities for startups remain, however, as long as they are sufficiently differentiated. Many in the space are scaling fast, and there are a number of particularly interesting, fast-growing segments in the infrastructure and analytics part of the ecosystem, including **streaming/real-time**, **data governance**, and **data fabrics/virtualization**. The explosion of interest in AI has also led to great opportunities (and a lot of funding) in **AI chips**, **GPU databases**, **AI devops** tools, and platforms enabling the **deployment of data science and machine learning in the enterprise**.

Machine Learning & AI

It's certainly been a **wild year in the world of AI research**, with anything from the prowess of AlphaZero to the staggering pace of release of new advances - new forms of Generative Adversarial Networks, Vicarious' new Recursive Cortical Networks, Geoff Hinton's new Capsule Networks. AI conferences like NIPS have grown to attract 8,000 people and thousands of academic papers are being submitted every day.

At the same time, the pursuit of **AGI remains elusive**, perhaps thankfully so. Much of the current wave of excitement (and fear) about AI results from the impressive performance of deep learning since 2012 but, in the AI research community, there's a **growing sense of "what now?"** as some question the foundations of deep learning (backpropagation) and others look to move past what they consider "brute force" approaches (lots of data, lots of computing power), perhaps in favor of more neuroscience-based approaches.

Far from fearing robot world domination, many in the AI research community are concerned that continued over-hyping of the field may eventually disappoint and lead to another AI nuclear winter.

Outside of AI research, however, we are just at the beginning of a wave of deployment:

Messaging



application of deep learning in the real world across a variety of problems involving speech recognition, image classification, object recognition and language, in different industries. If the infrastructure and analytics part of the ecosystem is getting to the late majority, we're still very much in early adopter territory for enterprise and vertical AI applications.

The **Cambrian explosion of deep-learning based startups** that started a year or two ago has mostly continued unabated, even though the AI startup market is (arguably) showing signs of finally cooling down. Expectations, round sizes and valuations remain high, but we are certainly past the phase where big Internet companies would snap up very early AI startups at high prices just for the talent. The air is also clearing up a bit and revealing "real" AI startups, versus a number of other companies that were leveraging the hype. Some of the AI startups that were founded in the 2014-2016 time frame are starting to hit early scale, and many are offering increasingly interesting products **across industries and verticals** including health, finance, "industry 4.0" and back office automation. Deep learning will continue bringing a lot of value in real world applications for years to come, and vertical-focused AI startups have many great opportunities ahead of them.

This continued explosion is very much a global phenomenon, with Canada, France, Germany, the U.K. and Israel being particularly active. However, **China seems to be playing at a completely different level in AI**, with reports of government-led pooling of data at mind-boggling scale (across Internet companies and municipalities), rapid advances in areas such as facial recognition and AI chips, and gigantic rounds of financing for its startups: according to CB Insights, China accounted for only 9% of global AI deal share but nearly 48% of global AI funding in 2017, up from 11% in 2016 (see some examples below).

In the same vein, **issues of data privacy** (and ownership and security) are emerging as a major concern around the world. In the early days of the Internet, data privacy was about protecting what we did online, a comparatively small portion of our activities. Correspondingly, only a small (albeit vocal and passionate) minority of people truly cared. As just about every aspect of our personal and professional lives is now connected to the Internet through an ever increasing array of connected devices, the stal

Messaging



are changing. With its ability to spot anomalies in massive data sets, predict outcomes and recognize faces, **AI is compounding the data privacy problem.**

A separate but related concern is that a lot of this data is owned by large Internet companies (GAFA). Some, like Facebook, have proven to be a less than perfect steward of it.

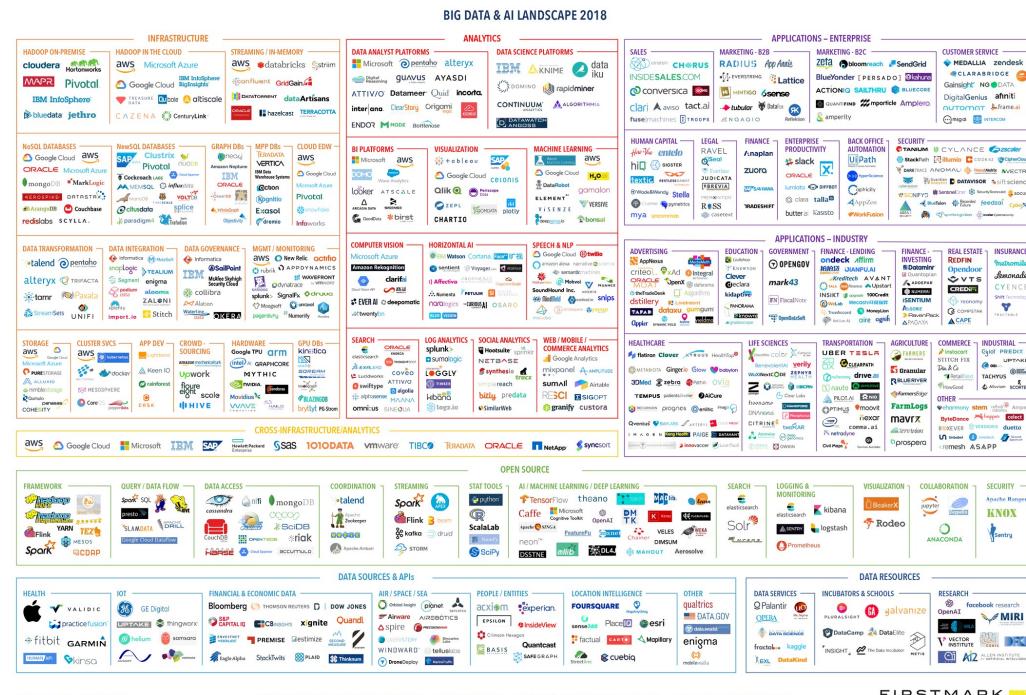
Nonetheless, this data provides them an unfair advantage in the race to produce ever more powerful AI.

Against those issues, an emerging theme is to think of the **blockchain as a possible foil against the risks of AI**, as well as a way for others, outside of GAFA, to produce great AI. Crypto economics are viewed as a way to incentivize individuals to provide their personal data and for machine learning engineers to build models by processing this data anonymously. It all remains very experimental, but some early marketplaces and networks are emerging

The 2018 Landscape

Without further ado, here's our 2018 landscape.

Quite semantic note: buzz terms come and go. Fewer people speak about “Big Data”, many more about “AI”, often to described the same reality. Consequently, we have slightly rebranded our 2018 landscape: it is now called the “Big Data & AI” landscape!



To see the landscape at full size, [click here](#).

(The image is high-res and should lend itself to zooming in well, including on mobile)

To download the full-size image, [click here](#).

To view a full list of companies in spreadsheet format, [click here](#).

This year, my FirstMark colleague [Demi Obayomi](#) provided immense help with the landscape.

We've detailed some of our methodology in the notes to this post. Thoughts and suggestions welcome – please use the comment section to this post.

Messaging



Who's in, who's out

On the exit front, the last year (since our 2017 landscape) has seen solid, but not extraordinarily strong.

A few key companies appearing on the landscape went public, in particular Cloudera, MongoDB Pivotal and Zuora. Others are preparing to go out at the time of this writing, such as Elastic.

Some notable acquisitions also occurred, including in particular Mulesoft (acquired by Salesforce post-IPO, for \$6.5B), Flatiron Health (acquired by Roched for \$2.1B), Appnexus (acquired by AT&T for \$1.6B), Syncsort and Vision Solutions (acquired for \$1.2B by Centerbridge Partners), Moat (acquired by Oracle for \$850M), Integral Ad Science (acquired by Vista Equity Partners for \$850M), eVestment (acquired by NASDAQ for \$705M) and Kensho (acquired by S&P Global for \$550M). It is worth noting that, other than Mulesoft, all those companies are headquartered on the East Coast (New York, Boston and Atlanta).

Many other companies were also acquired for smaller amounts: Gigya (SAP), Blue River Technology (Deere & Co), CoreOS (Red Hat), Guavus (Thales), Lattice Data (Apple), Socrata (Tyler Technologies) and PracticeFusion (AllScripts).

On the investment front, this was a year of big financing rounds for some Big Data and AI startups, particularly in China, with a number of oversized investments including Bytedance (\$3B in total across 2 rounds in 2017), NIO (\$1.6B across two rounds in 2017), and SenseTime (\$850M across two around in 2017 and 2018).

Major rounds of US companies appearing on the landscape include Snowflake Computing (\$263M Series A - see our recent [fireside chat](#) at [Data Driven NYC](#)), Cohesity (\$250M Series D), Dataminr (\$221M Series E), Affirm (\$200M Series E), Rubrik (\$180M Series D), Qualtrics (\$180M Series C - see an older but still relevant [fireside chat](#) at Data Driven NYC), Tanium (\$180M private equity round), ThoughtSpot (\$145M Series D) and Cov

Messaging



(\$100M private equity round) and C3IoT (\$100M Series F).

NOTES:

1) As every year, we couldn't possibly fit all companies we wanted on the chart. While the general philosophy of the chart is to be as inclusive as possible, we ended up having to be somewhat selective. Our methodology is certainly imperfect, but in a nutshell, here are the main criteria:

- Everything being equal, we gave priority to companies that have reached some level of market significance. This is a reasonably easy exercise for large tech companies. For growing startups, considering the limited amounts of data available, we often used venture capital financings as a proxy for underlying market traction (again, probably imperfect). So everything else being equal, we tend to feature startups that have raised larger amounts, typically Series A and beyond.
- Occasionally, we made editorial decisions to include earlier stage startups when we thought they were particularly interesting.
- On the application front, we gave priority to companies that explicitly leverage Big Data, machine learning and AI as a key component or differentiator of their offering. As discussed in the piece, it is a tricky exercise at a time when companies are increasingly crafting their marketing around an AI message, but we did our best.
- This year as in previous years, we removed a number of companies. One key reason for removal is that the company was acquired, and not run by the acquirer as an independent company.. In some select cases, we left the acquired company as is in the chart when we felt that the brand would be preserved as a reasonably separate offering from that of the

Messaging



acquiring company.

- 2) As always, it is inevitable that we inadvertently missed some great companies in the process of putting this chart together. Did we miss yours? Feel free to add thoughts and suggestions in the comments.
- 3) The chart is in png format, which should preserve overall quality when zooming, including on mobile.
- 4) As we get a lot of requests every year: feel free to use the chart in books, conferences, presentations, etc – two obvious asks: (i) do not alter/edit the chart and (ii) please provide clear attribution (Matt Turck, Demi Obayomi and FirstMark Capital).
- 5) Disclaimer: I'm an investor through FirstMark in a number of companies mentioned on this Big Data Landscape, specifically: ActionIQ, Cockroach Labs, Dataiku, Frame.ai, Helium, HyperScience, Kinsa, Timber, Sense360 and x.ai. Other FirstMark portfolio companies mentioned on this chart include Bluecore, Engagio, HowGood, Payoff, Knewton, Insikt, Optimus Ride, and Tubular. I'm a small personal shareholder in Datadog.

Report this

3,472 Likes



88 Comments

Show previous comments

Aaron Carl Labiaga
Intern at Starcom

5h ...

This would be helpful!

Like Reply

Messaging



David Earles, PMP

Delivering business strategy through Program management practices

2h ...

Raytchle Reiss Pg Dip DigM - interesting article

Like Reply



Add a comment...



Matt Turck

Managing Director at FirstMark

+ Follow

More from Matt Turck [See all 14 articles](#)



Beyond IoT: Building Decentralized, Intelligent Infrastructure

Matt Turck on LinkedIn

Frontier AI: How far are we from artificial “general” intelligence, really?

Matt Turck on LinkedIn

Growing Pains: The 2018 Internet of Things Landscape

Matt Turck on LinkedIn

Firing on All Cylind Data Landscape

Matt Turck on LinkedIn

Messaging

