Matt Turck

A Turbulent Year: The 2019 Data & AI Landscape

Twitter   Linkedin   Facebook

It has been another intense year in the world of data, full of excitement but also complexity.

As more of the world gets online, the **"datafication" of everything** continues to accelerate. This mega-trend keeps gathering steam, powered by the intersection of separate advances in infrastructure, cloud computing, artificial intelligence, open source and the overall digitalization of our economies and lives.

A few years ago, the discussion around "Big Data" was mostly a technical one, centered around the emergence of a new generation of tools to collect, process and analyze massive amounts of data. Many of those technologies are now well understood, and deployed at scale. In addition, over the last couple of years in particular, we've started adding layers of intelligence through data science, machine learning and AI into many applications, which are now increasingly running in production in all sorts of consumer and B2B products.

As those technologies continue to both improve and spread beyond the initial group of early adopters (FAANG and startups) into the broader economy and

world, the discussion is shifting from the purely technical into a necessary conversation around impact on our economies, societies and lives.

We're just starting to **truly get a sense of the nature of the disruption ahead**. In a world where data-driven automation becomes the rule (automated products, automated cars, automated enterprises), what is the new nature of work? How do we handle the social impact? How do we think about privacy, security, freedom?

Meanwhile, the underlying technologies continue to evolve at a rapid pace, with an ever vibrant ecosystem of startups, products and projects, heralding perhaps even more profound changes ahead. In that ecosystem, the year was characterized by the early innings of a long expected consolidation, and perhaps a passing of the guard from one era to another as early technologies are starting to give way to the next generation.

To try and make sense of it all, this is our sixth landscape and "state of the union" of the data and AI ecosystem. For anyone interested in tracking the evolution, here are the prior versions: 2012, 2014, 2016, 2017 and 2018.

Worth noting: as the term "Big Data" has now entered the museum of once-hot buzzwords, this year the chart will just be the "Data & AI Landscape".

Also, to make the reading more digestible, we'll break down the post into two parts:

Part I (this post) will include a few introductory thoughts on the rapidly evolving context around data privacy and regulation, which will have a profound impact on what can/cannot be done with data technologies; it will also include the landscape itself.

Part II will include a roundup of key trends on data infrastructure, analytics

and ML/AI.

Data, AI and society: The tide is shifting

In 2018, we noted how the data world had started to reveal some darker, scarier undertones, in the wake of the Cambridge Analytica scandal in particular.

This trend continued to develop in 2019. There were more data breaches, more privacy scandals. More stories of surveillance state in China (including this report on a Muslim town in Northwest China). More freaky examples of AI deepfakes, for which we are very unprepared.

As a result, the tide has started to shift in earnest.

Certainly, the debate around the dangers of AI, with all its sci-fi connotations, had captured imaginations already, and this year has seen more initiatives around thinking through those issues, such as the launch of Fei Fei Li's Institute for Human-Centered Artificial Intelligence.

But up until recently, questions around data ownership, privacy and security were met, for almost everyone but a vocal minority, with a resounding yawn.

Perhaps more than ever, **privacy issues jumped to the forefront of public debate** in 2019 and are now front, left and center. The fact that many of those issues were related to Facebook, a service known to billions, probably played an important role in sensitizing a much broader group of people around the world to the severity of the issues.

The data privacy landscape is also shifting, as governments are increasingly getting involved.

Regulation is certainly spreading in full force:

- GDPR, the European data protection and privacy regulation, came into effect in May 2018, and since then a few high profile fines have been announced including a €50 million fine issued to Google in January 2019 by the French data protection regulator and a £500,000 fine issued to Facebook in October 2018 by the UK's Information Commissioner's Office.
- The California Consumer Privacy Act (CCPA) will become effective on January 1, 2020.
- New York's privacy bill is "even bolder" than California's.
- San Francisco just voted to ban the use of facial recognition by city agencies.
- Illinois moved against video bots for hiring interviews.

Yet harsher government actions could take place. For starters, Facebook is likely to be fined up to $5B by the FTC over privacy issues. Perhaps most importantly, there have been increasing calls to **break up** the largest Internet franchises — too much power, too much data and not enough privacy. The clearest target has been Facebook (see this well- publicized opinion piece by one of its founders, Chris Hughes), but the discussion has included others as well (a proposal from presidential candidate Elizabeth Warren targets Google and Amazon).

Big Tech was already under pressure from within their own midst. Employees at Google, Amazon and Microsoft protested against the commercialization of their face recognition technology. Google relented. Amazon did not – some activist shareholders and employees tried to put a ban into effect, but were defeated.

For the FAANGs, privacy has become a new battleground, forcing their leaders to take much more of a public stance on the issue:

- Tim Cook, CEO of Apple, [warned us](#) about the "weaponization of data" which is leading us into a "**data industrial complex**."
- Sundar Pichai, CEO of Google, took a [public stand](#) on the issue in the NY Times.
- Mark Zuckerberg, CEO of Facebook, [vowed](#) to turn Facebook into a privacy-focused messaging and social networking platform.

To which extent such statements should be taken for face value, of course, is anyone's guess, and probably depends on the specific company and leader.

In Facebook's case, the launch of Libra, a global cryptocurrency, could arguably be considered as a way to continue making money in a "post-data", privacy-first world where the company would be less reliant on a pure advertising model based on user data – or as a way to collect even more personal data.

The debate around the impact of data and AI on privacy and society is obviously hugely important, and it is fundamentally healthy that it has become much more central over the last year or so.

Yet it is a complex discussion, which involves many nuances.

Our relationship to privacy continues to be a complicated one, full of mixed signals. People say [they care about privacy](#), but continue to purchase all sorts of connected devices that have uncertain privacy protection. They say they are outraged by Facebook's privacy breaches, yet Facebook continues to add users and beat estimates (both in [Q4 2018](#) and Q1 2019).

In the same vein, **how we decide to handle AI involves many trade-offs.** As all technologies, **AI is intrinsically neutral**, and whether it creates good or bad for society is ultimately a human decision. Take face recognition for example: it can be a tool for state surveillance, but it can also [help locate](#)
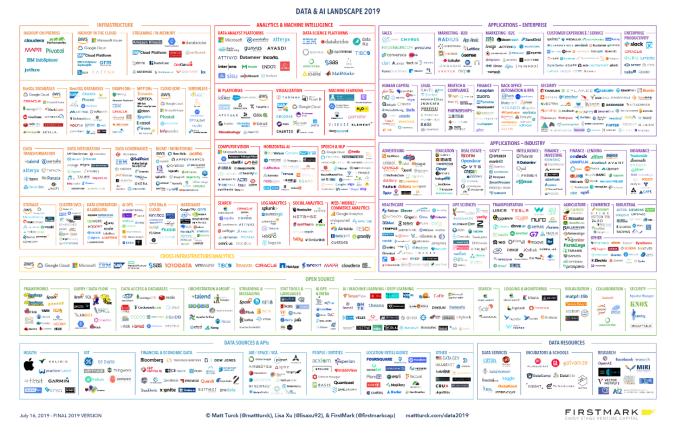
[victims of sex trafficking](#) Deciding how to regulate or curb AI, to the extent such a thing is even possible, would involve all sorts of second order consequences that are hard to predict. For example, if you regulate AI in the West, do you end up losing long term competitive advantage against China, which has a different set of rules (leaving aside any discussion on values)?

Data technologies: A vibrant but evolving landscape

While it is impossible in 2019 to ignore the broader questions of privacy, security and regulation around data and AI, the ecosystem of data technologies and products is **as exciting (and full!) as ever.**

The ecosystem is also evolving into some interesting ways, as some pioneering technologies such as Hadoop may be on their way out, replaced by cloud computing and Kubernetes, and entire segments, such as Business Intelligence, seem to be rapidly **consolidating**.

We'll dig into those various trends in some detail, but first, here's our 2019 Data & AI Landscape:

DATA & AI LANDSCAPE 2019

July 16, 2019 - FINAL 2019 VERSION   © Matt Turck (@mattturck), Lisa Xu (@lisaxu92), & FirstMark (@firstmarkcap)   matturck.com/data2019

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

Some key resources:

- View in **full size**: click here
- Underlying list: despite how busy the landscape is, we cannot possibly fit in every interesting company on the chart itself. As a result, we have a whole **spreadsheet** that not only lists all the companies in the landscape, but also hundreds more – click here.

A few additional comments:

- Yes, **you can zoom**! The image and all logos are very high-res, so you can navigate the landscape in detail by zooming. Works very well on mobile, too!
- This year, my FirstMark colleague **Lisa Xu** provided immense help with the landscape.
- We've detailed some of our methodology in the notes at the end of this post.
- Thoughts and suggestions welcome – please **use the comment section** to

this post. We'll probably publish two or three revisions of the chart until it's fully final.

Who's in, who's out?

The last year (since our 2018 landscape) has been active from an exit perspective.

Several companies on the landscape went public. Crowdstrike (NASDAQ:CRWD) and Elastic (NYSE:ESTC) reached big valuations at IPO time – $7B and $5B, respectively. Other IPOs included PagerDuty ($1.8B), Anaplan ($1.8B), and Domo ($500M).

Some very large acquisitions occurred in the last year, including Qualtrics (acquired by SAP for $8B), Medidata (acquired post-IPO by Dassault for $5.8B), Hortonworks ($5.2B merger with Cloudera), Imperva (acquired by Thoma Bravo for $2.1B), AppNexus (acquired by AT&T for up to $2B), Cylance (acquired by BlackBerry for $1.4B), Datorama (acquired by Salesforce for $800M), Treasure Data (acquired by Arm for $600M), Attunity (acquired post-IPO by Qlik for $560M), Dynamic Yield (acquired by McDonald's for $300M), and Figure Eight (acquired by Appen for $300M).

Notably, there has been a **wave of consolidation in business intelligence** in just the last quarter: Tableau (acquired by Salesforce for $15.7B), Looker (acquired by Google for $2.6B), Periscope Data (acquired by Sisense for $100M), ClearStory Data (acquired by Alteryx for $20M), and Zoomdata (acquired by Logi Analytics).

Many other companies on the 2018 landscape were acquired for smaller amounts: Alooma (Google), Bonsai (Microsoft), Euclid Analytics (WeWork), Sailthru (Campaign Monitor), Data Artisans (Alibaba), GRIDSMART (Cubic), Drawbridge (LinkedIn), Citus Data (Microsoft), Quandl (NASDAQ), Connotate

(import.io), Datafox (Oracle), Market Track (Vista Equity Partners), Lattice Engines (Dun & Bradstreet), Blue Yonder (JDA Software), SimpleReach (Nativo).

Also worth noting, the AI acqui-hire by large Internet companies, a fixture of 2016-2017, is not completely dead: Twitter acquired Fabula AI to strengthen its machine learning expertise, for example.

On the investment front, Big Data and AI startups continued to see big financing rounds. Investments in China were not quite as oversized as last year, when there were multiple companies that raised over a billion dollars. Chinese companies that raised large rounds this year included facial recognition company Face++ ($750M Series D), AI chip maker Horizon Robotics ($600M Series B), fleet management company G7 ($320M Series F), online tutoring platform Yuanfudao ($300M Series F).

In the US, huge investments went into autonomous vehicle companies, including Cruise ($1.9B across 2 rounds in 2018 and 2019), Nuro ($940M Series B), and Aurora ($600M Series B). RPA companies also saw massive rounds: UiPath ($800M across 2 rounds in 2018 and 2019) and Automation Anywhere ($550M across 2 rounds in 2018).

Other major rounds of US companies on the landscape include Verily Life Sciences ($1B private equity round), Cambridge Mobile Telematics ($500M), Clover Health ($500M Series E), Veeam Software ($500M), Snowflake Computing ($450M Series F), Compass ($400M Series F), Zymergen ($400M Series C), Dataminr ($392M Series E), Lemonade ($400M Series D), Rubrik ($260M Series E), Databricks ($250M Series E), and MediaMath ($225M Series D).

End of Part I.

June 27, 2019 / AI, Big Data

---

31 thoughts on "A Turbulent Year: The 2019 Data & AI Landscape"

Sarah McKenna

June 27, 2019 at 8:40 pm

Excited to see a couple web data extraction companies on the list. One that should be included is Sequentum Inc. (infrastructure category). Backed by Worldquant Ventures, Sequentum is nearly a decade old and supports thousands of software license customers, including large enterprises and government agencies. Our CG Enterprise product is best of breed technology for reliable web data extraction at scale and with governance. Find out more at http://www.sequentum.com .

Reply

Rob Koyfman

June 28, 2019 at 3:46 am

Suggestion to add Koyfin to the financial data and economics section.

Reply

Matthias

June 28, 2019 at 10:47 am

Where is RStudio, data.table and the tidyverse? As well as caret and mlr?

Reply

Richard

June 28, 2019 at 11:55 am

Hi Matt – First off, great piece. Thank you! Recognize the potential list is long and not everyone fits. Whether it's the logo infographic or the longer excel list, Amenity Analytics belongs in NLP and Text Analytics for Finance, Media and Business. Hopefully we can prove that out as we grow. Happy to chat and share our world view at your convenience. Thanks again!

Reply

Abhishek

June 28, 2019 at 1:56 pm

Missing Imply.io in streaming or newSQL dbs. Also, Starburst Data in the MPP category..

Reply

mattturck 👤

July 1, 2019 at 10:04 am

Thank you. Yes, adding both, definitely an oversight as we both had them speak at Data Driven NYC recently!

Imply.io here: https://www.youtube.com/watch?v=uKlJSDtHj-c&t=818s
Starburst here: https://www.youtube.com/watch?v=Ol6Awd-sBXo

We'll probably add Starburst to Data Analyst platforms, as ultimately they're the end users of the product.

Reply

Alan

July 3, 2019 at 6:15 pm

Also consider adding Dropbox (storage) and Nearmap (spatial data analytics)

Reply

Drake Pruitt

June 28, 2019 at 6:22 pm

Matt, great roundup. In Data Science group, "Continuum" was renamed "Anaconda" in 2017. Cheers.

Reply

mattturck 👤

July 1, 2019 at 10:01 am

Great, thank you

Reply

mattturck 👤

July 4, 2019 at 10:40 am

Now fixed / updated

Reply

Andrew Bloom

June 28, 2019 at 8:13 pm

Thanks for such a comprehensive report Matt. Suggest you also take a look at the B2B marketing data space. There are some large players that you have missed. Infogroup is one of the largest data intelligence companies powering everything from search and navigation to the acquisition efforts of leading brands. Check out this report

https://www.dnb.com/content/dam/english/economic-and-industry-insight/the-forrester-wave-b2b-marketing-data-providers-q3-2018.pdf

Reply

John Thuma

July 1, 2019 at 8:05 am

Where is Confluent?

Reply

mattturck 👤

July 1, 2019 at 10:00 am

Infrastructure / Streaming / In Memory

Reply

Steve Kurtz

July 2, 2019 at 10:00 am

Where is Narrative?

Reply

Ronald van Kleunen

July 1, 2019 at 1:14 pm

Dear Matt,
While is see a few companies in the spreadsheet that have been taken over
and being integrated into Cisco Systems (www.cisco.com /
https://en.wikipedia.org/wiki/Cisco_Systems).

But Cisco as a company itself is missing in the graphic/list

Hereby the links – Cisco Enterprise Networking and Intent-Based Networking using AI/ML

Cisco AI Network Analytics: Making Networks Smarter and Simpler to Manage
https://blogs.cisco.com/analytics-automation/cisco-ai-network-analytics-making-networks-smarter-simpler-and-more-secure

Improving Networks with Artificial Intelligence
https://blogs.cisco.com/enterprise/improving-networks-with-ai?ccid=cc000098&dtid=oblgzzz000659

Reply

Eric Prinsen

July 1, 2019 at 2:30 pm

Matt,

Can you list GSI Technology in the Infrastructure/Hardware category? GSI is already sampling the APU (in place parallel processor) for AI applications in Visual Search, Cheminformatics, Bioinformatics, Computer Vision and Big Data.

Reply

Eric Prinsen

July 1, 2019 at 2:40 pm

Some detail on GSI Technology's APU chip linked here (slides 9-16):

http://ir.gsitechnology.com/static-files/0e74b761-ea29-4411-8224-9fcad0994e81

Reply

Sunny Suh

July 1, 2019 at 5:07 pm

Qlik and Sisense are probably the most comparable, both should be in the bi platform. Curious why you updated and said alteryx in data analyst platforms instead of leaving it as clearstory data, is it because it's mostly integrated? For BI platforms, why call it just Microsoft instead of Microsoft Power BI? That is also more a visualization platform, similar to tableau. salesforce also has einstein, which is more of their BI tool, wave is more of a reporting tool. Let me know what you think! Would love to meet sometime as well.

Reply

---

mattturck 👤

July 4, 2019 at 10:46 am

Clearstory data — will probably not be maintained as a separate brand (links on CSD's homepage already point to Alteryx product pages) so we put Alteryx in the box instead

Other comments (Power BI, Einstein Analytics) — had received similar suggestion, updated in new version of landscape, thank you

Connecting – happy to, let's discuss separately

Reply

Corey Samuels

July 2, 2019 at 10:15 am

So much good info here, and I really appreciate the summary and highlights. Minor operational comment: There are a LOT of very useful links here. I found myself wishing each would open in a new tab so I could look but also continue reading your blog post without navigating back.

Reply

mattturck 👤

July 4, 2019 at 10:47 am

Good point, thank you. Looks like I can fix on a link by link basis, will try to give it a shot.

Reply

Dylan Steeg

July 2, 2019 at 10:21 am

Thank you Matt for continuing to update and publish the Data & AI Landscape. Please consider adding Aible to the Machine Learning section. Aible was recently recognized by Gartner and Forrester as an AutoML innovator. Please see details below.

i) Gartner's 2019 Cool Vendor in Analytics report:
https://enaible.aible.com/aible-gartner-cool-vendor

ii) Forrester New Wave™: Automation-Focused Machine Learning Solutions Q2 2019, saying,
"Unique among AutoML vendors, Aible gets that a model that maximizes accuracy almost never maximizes business impact."

Reply

Romina Aziz

July 2, 2019 at 11:30 am

What about Visier? It's not in the list and it's a really powerful HR analytics tool. https://www.visier.com/

Reply

Romina Aziz

July 2, 2019 at 11:31 am

Sorry my comment before got cut and I hit the ENTER button too fast. This is a great article and I enjoyed reading it thoroughly! Thanks for putting this together.

Reply

Matt Jones

July 3, 2019 at 2:22 am

Suggestion to add Geoblink to the LI space, they are disrupting the space and are near Series B phase, invested by top European VC firms

Reply

Kerim

July 4, 2019 at 5:03 am

Hi Matt,
once again thanks for the great work and effort you and the team have put into this.
Suggesting to add KNIME Analytics for Data Science Platforms. They have been Leader in Gartners „Data Science and ML Platforms" Quadrant for six years, now.

Reply

mattturck 👤

July 4, 2019 at 10:36 am

Thank you. KNIME is in the Data Science Platforms box already.

Reply

Kerim

July 4, 2019 at 11:11 am

Just recognized so as well, I was referring to the full res picture. It seems, that the link is not pointing to the latest version.

Reply

Phil Unger

July 4, 2019 at 8:13 am

Matt,

Thanks for putting this together! I did notice that TIBCO is noticably absent from your landscape diagram. Any particular reason why Spotfire, Statistica, etc were left off?

Reply

---

mattturck ▲

July 4, 2019 at 10:38 am

Thank you. TIBCO is "Data Science Platforms" and also in "Cross infrastructure / analytics", which is a category meant to include the largest companies on the landscape that straddle many boxes.

Reply

Andrew Thomas

July 4, 2019 at 5:50 pm

Suggestion to add
Immense Simulations Ltd (Immense.ai) Transport / Mobility Simulation as a Service (SaaS) company that has just completed Series A funding.