# DATA SCIENCE TRENDS 2022

# 1. Acknowledgments

2. Data Science Trends 2022

# 3. Trends According to 10 Data Science Experts

- Kirk Borne
- Iain Brown
- Kate Strachnyi
- David Regalado
- Jarrod Teo
- MH van Staden
- Besa H Bauta
- Smiriti Mishra
- Lisa-Christina Winter Done
- Ioannis Tsamardinos
- Aruna Pattam
- Michelle Dunivan

# 4. AI Time Journal Podcast

# Acknowledgments

Editing by TzuYi Juang Front Cover Image by Alexandra Jackson, Graphic Design by Madison Hovey, Review by Martin Russo, Jacks Kernohan

Special thanks go to:

Melissa Drew, Luis Villar, Kaitlyn Marer, Kate Strachnyi, David Regalado, Jarrod Teo, MH van Staden, Besa H, Bauta, Smiriti Mishra, Kirk Borne, Iain Brown, Lisa-Christina Winter Done, Ioannis Tsamardinos.

by Melissa Drew, Associate Editor at AI Time Journal

Data has become more critical than ever for an organization to remain competitive, to expand its products and service offerings, and adjust more quickly to changes in consumer demand in this global economy.

The value of Data Science has been around for quite some time. Personally, I have been advocating data as the key to better-informed decisions in the organization since 1997 and using data combined with cognitive technologies since 2004. However, it wasn't until the last few years the term Data Science solidified a more formal role in our economy with specific job descriptions, recruitment priorities, and most recently the development of educational certifications and university degrees.

Data Science is still rather fluid. In 2021, more terms and definitions became mainstream than what had seen in prior years with Data Fabric, Data Mesh, Lakehouse, AutoML, TinyML, NoCode ML, Model Drift, just to name a few. It is expected this trend will continue well into 2022.

#### Data Science Organization

Depending on the size and culture of the organization, a Chief Data Officer oversees the data science team, whereas a Chief Analytics Officers oversees the AI model engineers and developers. The primary role of the Data Organization to solve a specific business challenge by collecting hundreds of data sources both internally and externally to the organization. This data is cleaned, transformed, and enriched in preparation for model development. Examples of the more common Data Science roles are:

- (IT) Data Engineer / Data Architect: developing the cloud infrastructure to house the data
- (Business) Data Scientist: consolidating and curating data from multiple sources to cleanse, normalize, harmonize, and transform data.
- (Business) Data Analyst: use applications to create visualizations of the data, e.g., dashboard, charts, association graphs.

Note: Some organizations may not have distinct definitions between Data Scientist and Data Analyst while other organizations may require mathematics, SQL, and other programming backgrounds. Reading the entire job description is necessary when searching for data related positions.

#### Industry and Start Up

Data Science is also industry agnostic. The agriculture industry is using drones to determine optimal harvest times and earlier identification of crop disease. Computer Vision technology in the automotive industry will shift from exterior to Interior Sensing, which captures eye movements of the driver recognizing drowsiness. The Healthcare industry is exploring the use of data to hyper-personalize patient outcomes to improve diagnosis and personalize the patient health care experience.

The expanding use cases across all industries has increased the number of technology start-ups and small businesses using data to solve specific business challenge leading to another emerging trend, AI as a Service (AIaaS). This increase is new start-ups is so significant, it will force procurement and supply organizations to rethink their policies in managing and evaluating suppliers in 2022.

#### Regulation

While industries push forward to utilize AI across products and services, so does the need to develop federal governance and policies. As an example, the Federal Drug Administration is overhauling their guidelines and procedures used to evaluate Software as Medical Device (SaMD). If software is intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions and relies on AI/ML, the FDA considers it SaMD. Additionally, the FDA is considering not only evaluating the data and assumptions used in developing SaMD but is also proposing to interview the individuals responsible in developing the solution to validate the threshold bias.

#### Responsible AI

2021 highlighted some very public solutions that didn't quite accounts for common data bias before their products and services were sold publicly. This bias can be created based on the depth and breadth of data collected or not collected; if retrospective data versus prospective was available, social bias, and algorithm bias.

A few examples from 2021 where algorithms included a data bias was misaligned prioritization of patients for kidney transplants, potentially discrimination of a specific demographic for credit card qualifications, or proactively removing candidates when using an automatic resume screener. Additional topics expecting to trend in 2022 as a result are: Ethical AI, Trustworthy AI, and AI for Good.

The 2022 Data Science Trends ebook curates a collection of industry and consulting thought leaders who share their perspectives on the continuing and emerging trends, along with practical insights on the data scientist in the future.

# Trends According to 10 Data Science Experts

#### **Kirk Borne**

#### Chief Science Officer at DataPrime, Inc.

Some of the big data science trends in 2021 that will continue into 2022 include the increasingly incredible growth in the number of data science / AI / ML applications being deployed in many different domains -- for example: healthcare, medical, government, finance, retail, marketing, call centers, logistics/supply chain, and gaming.

The latter includes VR (Virtual Reality) and AR (Augmented Reality) implementations in the Metaverse. All these different domain applications have an "edge"



component. Edge intelligence essentially means dynamic actionable insights at the edge -- at the location and moment of data collection. Intelligent edge applications include dynamic insights-as-a-service from internet of things sensor deployments, which are governed through strategic enterprise observability strategies. These will soon dominate data science discussions and enterprise applications.

Organizations are demanding more productivity and value from their data, as well as from their AI, ML, Data Science assets and investments, and 2022 will see more of this. However, there are significant headwinds pushing back on these advancements caused by the huge data science talent gap. The growth in new data science training programs, bootcamps, and educational degree opportunities is rapid, but it is still falling behind the more rapid growth in demand. Consequently, we can expect organizations to opt for more low-code/no-code deployments (such as AutoML) to harness and maintain the growing data science momentum.

Associated with all these trends (growth in the demand for more value from data, bigger talent gap challenges, and AutoML model governance) is another big trend continuing into 2022, and that is the use of model monitoring services like MLOps and ModelOps, to monitor model drift, including data drift and concept drift. This became a serious concern during the pandemic but is now becoming a core data science requirement to monitor data science model ethics, explainability, robustness, risk, and trustworthiness.

# Iain Brown

#### Head of Data Science at SAS

The importance of data science in tackling today's biggest challenges has continued to gain momentum throughout 2021 and I foresee this carrying over to 2022. My top 3 picks for Data Science trends that will emerge next year are:

1. AI as a Service (AIaaS) – with the growing demand for data science skills still set to outpace supply, organisations will need to look for other means to meet their AI requirements. To support this, I foresee 2022 as the year AI



as a service taking center stage with well-established cloud-based offerings and AI vendors providing on demand services to answer future business challenges.

2. Rise of Natural Language Understanding (NLU) – with the recent renaissance in natural language processing (NLP), I predict the subset of this field, NLU, will have more focus in 2022. NLU deals with the AI-hard problem of machine reasoning and offers the promise of automation, translation, and Q&A capabilities.

3. Responsible AI (RAI) – rather than emerging next year, Responsible AI, which is already in the spotlight, will continue to gain momentum and clarity around how organisations adopt ethical principles by design. This will partly come through proactivity from developers as well as legal and regulatory standards.

# Kate Strachnyi

#### Founder of DataCated

We will likely see continued innovation within data platforms that make it simpler for the business to understand relevant data and reduce the time to insights.

Additionally, we will see the rise of the data mesh, data fabric, and Dataware; along with a move towards data centricity. Due to innovations within the metaverse, there will be an increase in the demand for data scientists who can handle the creation, handling, and assessment of large synthetic datasets.





## **David Regalado**

#### **Co-Founder at Data Engineering LATAM**

Definitely, Data Engineering would be the macro trend for 2022. More and more companies are realizing that they need usable data in order to obtain useful data. Making sense out of data it's not an easy task. It really is like bringing order out of chaos. You need to extract, cleanse, enrich, and transform this data. One thing is to do it in a spreadsheet and another completely different is to do it at scale (hopefully using the cloud). And don't even get me started on streaming data, that's another level!





# **Jarrod Teo**

#### **Chief Data Scientist of Direct Sourcing Solutions**

This year and last year have been all about retail and e-commerce businesses surviving Covid-19. My advice to clients over that time has been to allocate their marketing budgets to keep paying customers close.

If the world manages to tame the Omicron variant, that will all change in 2022. As we gain control over Covid-19, most companies will start to look at winning back the customers who have not been buying from them over the past two years. Expect the extensive use of data science to guide win-back marketing strategies designed to encourage customers to shop again. We might even see collaborations



with celebrities for online sales. If live events are permitted, how about fashion shows with invitations extended to premium customers singled out by an AI engine?

It's not far-fetched. DSS has Customer Segmenter, a real-time machine learning model that can easily identify high-value customers as well as those at risk of moving to competitors. It's a potent tool, especially when combined with DSS Next Product to Purchase to predict what next to sell those customers.

My clients are often surprised that they can earn up to \$5,000 per hour in sales. When you sell the right products to the right customers, you end up with happy customers who are loyal to your brand. That's the key to surviving tough times and also to the data science-based business decisions that help you win back customers in times of growth.

# MH van Staden

#### Data Analytics Leader of the Year 2020 | Speaker | Writer | Mentor | Coach

Despite the amazing feats achieved by cutting edge deep learning, in fields like image recognition, language processing, self-driving cars and gaming, application of these complex layered algorithms, in practice, in general, remains severely limited.

The superior accuracy of these 'black-box' systems is not sufficient, by itself, for wider adoption. A range of AI implementation ethical considerations, stakeholders, and affected parties, have the additional need for transparency.



Early forms of AI, developed hand in hand with clearly traceable and auditable, logical decisioning steps. With the focus on models capable of solving more complex challenges, explainability has taken a back seat.

This does not mean that some work has not been done, in advancing how we can better make sense of convoluted networks. In 2022 and beyond, I do however see interpretability increasingly being drawn into the forefront.

Within the spotlight, the way AI is explained, will increasingly be scrutinised from a variety of new, less technical perspectives, forcing refinements and new approaches, to making sense of what is going on underneath the hood.

Advances in unearthing the potential of human understandable and actionable AI insights, follows as a natural spin off benefit.

## Besa H Bauta

Chief Data and Analytics Officer, State of Texas, Department of Family and Protective Services, Office of Data and System Improvement

Hyper-parameterized models, Augmented Workforce, Low-code and No-code AI, Scalable AI, Real-time Data



	31 32 33 34 35 36 37 38 38 39	<pre>self.file = self.logdupes = self.logger = logsing if path:     self.file = cont(self)     self.file = cont(self)     self.file.cont(self)     self)     self.file.cont(self)     self     self.</pre>
	40 41 42 43 44 45 46 47 48	<pre>@classmethod def from_settings(cls.settings) debug = settings.gettings) return cls(job_dir(settings)) def request_seen(self.request) fp = self.request_interprints;</pre>
	49 50 51 52 53 54 55	<pre>if fp in setric add(fp)     self.fingerprints.add(fp)     if self.file:         self.file.write(fp + os.laesaad) def request_fingerprint(self. request)     return request_fingerprint(request)</pre>

#### Smiriti Mishra

#### Head of AI at Earthbanc | AI Speaker | Computational Neuroscience at KTH | Google Women Techmakers

I definitely see AIoT as a major trend for 2022. The converge of AI, IoT, 5G, cloud computing.

And of course AutoML and MLOps, I shared a document about it yesterday.

Also, more use of synthetic data, deepfakes etc.





# Lisa-Christina Winter Done

#### Co-Founder of Hakuna MaData & Two Kermis

2022 will be an incredible year for data science, but I believe the most impactful trend will be the considerable increase in public use of synthetic/generative technology (e.g. deepfakes) resulting both from the mass availability of usable implementations, and constant increase in quality, of, these techniques.





# **Ioannis Tsamardinos**

#### **CEO & Cofounder of JADBio AutoML**

To me, Deep Learning is continuously producing super exciting results. It is all over the news and the scientific literature. However, it mostly regards image, video, and text data. Other technologies that I am excited about that are particularly pertinent to molecular and clinical data are Automated Machine Learning and Causal Discovery.

Of course, my opinion is biased because I am involved in research in these fields. Automated Machine Learning has the potential to tremendously boost our productivity in



analyzing data. Causal Discovery seeks to go a step beyond standard Machine Learning and predictive modeling and identify causal relations and causal effects. It can answer what-if scenarios, e.g., what I should expect if I perform this treatment or intervention. Questions that are not possible to answer correctly with standard predictive models.

# Aruna Pattam

#### Head, AI & Data Science

Data science is not only changing how we do things, but also what. The last decade has seen its impact grow exponentially and it's continuing to be a driving force. Top 5 Emerging data science trends that I see for 2022 include:

1. AutoML: While Machine Learning (ML) is a hot topic, there is another trend that is becoming very popular is automated machine learning (autoML). AutoML is creating opportunities to democratize ML by allowing firms with limited data science expertise to develop and automate



analytical pipelines that have the ability to solve sophisticated business problems.

2. TinyML: This is a fast-growing field in data science. With the rise of Internet-connected devices, tiny machine learning algorithms are being developed that can analyze sensor data on these small electronics. These "TinyML" systems also operate at extremely low power for always-on use cases targeting battery operated gadgets such as wearables or IoT devices.

3. AI as a Service (AIaaS): AIaaS is an exciting new concept that will have profound effects on the future of business. It offers businesses humanlike capabilities that can be done by a computer system with low cost and flexibility by third party service providers. AIaaS comes with the ability for virtually anyone to access state-of-the art algorithms around the world without requiring an extensive investment in resources. This also provides companies with the ability to quickly scale their AI capabilities.

4. Generative AI & Synthetic Data: Generative AI had been making waves in the arts and entertainment industry. It's expected that this technology will soon become widespread in other industries. It has a huge potential in generating synthetic data such as synthetic faces for training the machine learning algorithms while avoiding privacy concerns.

5. Augmented analytics: Augmented analytics has the potential to change how decision-makers view their data by providing them with insights. Successful enterprises will be able automate this process, which can help reduce workload and give data science professionals time for higher value-added activities such as developing new algorithms or customized models that suit a company's unique needs.

Looking ahead to 2022, data science and machine learning are shaping up to be a more eventful year.

# **Michelle Dunivan**

#### Analytics Director at Superior Court of Arizona in Maricopa County

As no code and gig-economy analytics become more prevalent, small and medium sized businesses will have more access to the benefits of data science, and more potential to be data driven.

For those that did not make good use of data during the pandemic, they will exponentially fall behind. Not capitalizing on the biggest data-driven opportunity in our lifetimes will be compounded by the return to status quo,



which will affect recruitment and retention of data science and related talent. Especially with the limited talent pool in data science, there will be less and less incentive to remain in

Ethics will take a front seat this year, and emerge as a more explicit part of governance in all sizes and types of organizations, not just the most mature as it has been in the past. Resources like tactical guidelines and expectations will be made available by trade organizations and inter-industry collaborations to help define processes and considerations that will help all organizations to create a fair, equitable, profitable, and sustainable future for data science and AI.

# AI TIME JOURNAL

Listen to more fascinating AI & Data Science interviews on the AI TIME JOURNAL Podcast!

Available on Apple Podcasts, Spotify, and YouTube.

#### **Interview Series**

Impact of AI Data Science Leaders Data Science

#### **Interview Inquiries**

For Interview Inquiries, please visit: https://www.aitimejournal.com/interview-inquiries