*Posted on 12/13/2021*

A few days ago, Timnit Gebru, who resigned from Google and launched her own AI research institute, published an article entitled "For truly ethical AI, its research must be independent from big tech" on *The Guardian*. In the article she accused several big tech companies of unethical behaviors e.g. Google forced her to withdraw the paper on bias of language models; Amazon crushed the labor union, and Facebook prioritizes growth over all else. In addition, she mentioned that recently California passed the *Silenced No More Act* to enable workers to speak against racism, harassment, and other forms of abuse in the workplace, thus preventing big corporations from abusing power. In conclusion, she suggested that we need alternatives rather than allowing big tech companies to monopolize the agenda.

**Full article:**
**https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebru?utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+December+10th%2C+2021&utm_campaign=11122021**

**That's my take on it:** I don't disagree with Gebru. However, it is difficult to find an optimal balance. Interestingly, Gebru cited the example of spreading misinformation in Ethiopia as evidence that unethical behaviors are widespread in high tech companies. In November 2020, a war broke out in Ethiopia, the home country of Gebru. According to Gebru, one of the causes is that Facebook allowed unchecked misinformation and hate speech being posted on Facebook. In a similar vein, a week ago Rhingya sued Facebook for $150 billion, saying that Facebook helped spread hate speech, misinformation, and incitement to violence in Myanmar.

It is important to point out that some political dissidents and people who promote alternate views on various controversial issues (e.g. COVID19, LGBT) also complain that social media companies muted their voice in the name of banning hate speech and misinformation.

Where can we draw the line? As Professor Thomas Sowell said, there is no solution, only trade-off.

*Posted on 12/3/2021*

Timnit Gebru is an Ethiopian-American computer scientist who specializes in algorithmic bias and data mining. For a long time she had led various AI task forces at big tech corporations, including Apple and Google. Her career path changed when in December 2020 Google Manager asked her to either withdraw a pending paper pertaining to bias in language models or remove the names of all the Google employees from the paper. According to Google, the paper ignored the latest developments in bias reduction. Gebru refused to comply and eventually resigned from her position. Recently Gebru announced that she is launching an independent AI research institute focusing on the ethical aspects of AI. Her new organization *Distributed Artificial Intelligence Research Institute* (DAIR) received $3.7 million in funding from the MacArthur Foundation, Ford Foundation, Kapor Center, Open Society Foundation, and the Rockefeller Foundation.

Gebru said she was more successful at changing Google's policies by publishing papers outside Google that could catch the attention of academics, policy-makers and journalists, rather than raising concerns inside Google. She hoped DAIR could break the monopoly of AI research by big tech companies.

**Full article:**
**https://www.washingtonpost.com/technology/2021/12/02/timnit-gebru-dair/?utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_content=Data+Science+Insider%3A+December+3rd%2C+2021&utm_campaign=04122021**

**That's my take on it:** On the one hand, research in AI and machine learning is a **bottom-up movement,** which is evident by the fact that many breakthroughs and advanced algorithms are contributed by individuals from the open-source community. On the other hand, it is also a **top-down revolution** when many large-scale projects are funded by the government and big corporations, especially the projects that require vast amounts of data and high-performance computing. No doubt Gebru's efforts and other similar research groups can provide **checks and balances** for the field.

*Posted on 11/9/2021*

Today is the first day of the 2021 Tableau Online Conference. I attended several informative sessions, including the one entitled "**Data is inherently human**" (see attached). This session highlighted the alarming trend that 85% of all AI projects will deliver erroneous results due to bias in data, algorithms, or human factors, according to a Gartner report. One of the speakers, who is a white woman, pointed out that AI-empowered voice recognition systems have problems with her southern accent. In addition, when she listened to her daughter 's Tiktok, she knew it was English, but she had no idea what it meant. She emphasized that machine learning algorithms, such as sentiment analysis, must be adaptive to linguistic evolution. Some terms that were negative two years ago might mean something positive today.

**That’s my take on it**: It is a well-known fact that many facial recognition systems have a harder time identifying people with darker skin. Actually, bias in technology affects every ethnic group, not only minorities. I can imagine that voice recognition systems developed in the US might not work well in Australia and South Africa. Further, semantics vary from time to time, and also from place to place. For example, in American culture "PK" stands for pastor’s kid or preacher’s kid. However, the abbreviation "PK" has a negative connotation for Hong Kong people. There is no fool-proof AI system!

*Posted on 10/30/2021*

The open-source software platform GitHub, owned by Microsoft, stated that for some programming languages, about 30% of new codes are suggested by its AI programming tool *Copilot*, which is built on the OpenAI Codex algorithm. This machine learning algorithm is trained on terabytes of source codes and is capable of translating natural human language into programming language. According to Oege de Moor, VP of GitHub Next, a lot of users have changed their coding practices because of Copilot and as a result they have become much more productive in their programming.

**https://www.axios.com/copilot-artificial-intelligence-coding-github-9a202f40-9af7-4786-9dcb-b678683b360f.html?utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+October+29th%2C+2021&utm_campaign=30102021**

**That’s my take on it:** On the one hand it is a blessing that cutting-edge technologies can make programming more efficient by modeling after many good examples. But on the other hand it could suppress potential innovations due to some kind of **echo chamber effect**. Consider this scenario: Henry Ford consults an AI system in an attempt to build a more efficient process for manufacturing automobiles. Based on a huge collection of “successful” examples learned from other automakers, the machine learning algorithm might suggest Ford to improve efficiency by hiring more skillful workers and building a bigger plant. The idea of **assembly line** would never come up! I am not opposed to programming assistance, but at the end of the day I must remind myself that I am the ultimate developer!

*Posted on 10/27/2021*

Two days ago (Oct. 25, 2021) the *Financial Times* reported that UK's spy agencies have signed a contract with Amazon Web Services. British intelligence agencies, such as MI5 and MI6, will store classified information in the Amazon cloud platform and also utilize Amazon's AI for intelligence analytics. British intelligence offices have been using basic forms of AI, such as translation technology, since the dawn of AI. Now they decided to expand AI applications in response to the threat from AI-enabled hostile states.

https://www.reuters.com/world/uk/amazon-signs-deal-with-british-spy-agencies-boost-use-ai-espionage-ft-2021-10-25/

**That's my take on it:** The stereotypical image of people in espionage is 007: handsome, strong, and dare to fight against dangerous villains by hand-to-hand combat. Not anymore! In the near future the most powerful weapon for a spy is not the Beretta pistol (the type of handgun used by James Bond); rather, it will be a mouse and a keyboard. If you want to be the next James Bond, study data science and machine learning!

*Posted on 10/20/2021*

Currently I am working on a book chapter regarding ensemble methods. During the literature review process a recent research article caught my attention:
Ismal, A. et al. (2021). A new deep learning-based methodology for video DeepFake detection using XGBoost. *Sensors, 21*. Article 5413.

https://doi.org/10.3390/s21165413.

DeepFake is a deep learning AI algorithm that can replace one person with another in video and other digital media. Famous humorous examples include fake videos of Obama and Queen Elizabeth. An infamous example is that in 2017 a Reddit user transposed celebrity faces into porn videos. Ismal and his team developed a new DeepFake detection system based on XGBoost, a supervised machine learning method that is capable of making gradual model improvement by running many decision trees and analyzing the residuals in each iteration. Those authors claimed that the accuracy is 90.73%, meaning that **the error rate is 9.27%**.
**That's my take on it**: In 1997 when Linda Tripp recorded her conversation with Monica Lewinsky about her affair with President Clinton, the legal enforcement system accepted the audio tapes as convincing evidence. Today you cannot trust video recording! Let alone audio! There is a still-photo equivalent to DeepFake: DeepNude. This app can use neural networks to remove clothing from the images of people, and the result looks realistic. The app is sold for $50 only. Due to its widespread abuse the developer retracted

it in 2019. However, parts of the source code are open and as a result there are many copycats in the market. I am glad that now cutting-edge technologies like XGBoost can be used to detect faked videos, but in the first place the problem originates from state-of-the-art technologies! According to some experts, DeepFake technologies have been improving exponentially. In late 2017 it took hundreds of images and days of processing time to swap faces in a video clip. Today it requires only a handful of images, or even just text inputs, and a few hours. It is similar to the race between computer viruses and anti-virus software packages. No matter how sophisticated anti-viruses software is, Trojan horse, spyware, ransomware…etc. keep evolving. The same contest will happen between DeepFake/DeepNude and fake video/image detection systems. The Pandora box has been opened!

*Posted on 10/15/2021*

Recently Facebook launched a new research project named *Ego4D* in an attempt to teach AI to comprehend and interact with the world as humans do, rather than from a third-person perspective. There are two major components in *Ego4D*: an open dataset of egocentric (first-person perspective) video and a series of benchmarks that Facebook thinks AI systems should be capable of handling in the future. The dataset, which is the biggest of its kind, was collected by 13 universities around the world. About 3,205 hours of video footage were recorded by 855 participants living in nine different countries.

**Full article:**
**https://www.theverge.com/2021/10/14/22725894/facebook-augmented-reality-ar-glasses-ai-systems-ego4d-research**

**That's my take on it:** For a long time research activities have been limited by a narrow definition of data: numbers in a table. In qualitative research we go one step further by including open-ended responses. But that is not enough! A lead research scientist at Facebook said: "For AI systems to interact with the world the way we do, the AI field needs to evolve to an entirely new paradigm of first-person perception. That means teaching AI to understand daily life activities through human eyes." Whether there will be any self-aware AI system in the future is controversial. Nonetheless, how Facebook is trying to train AI is also applicable to human researchers. No matter whether the data are structured or unstructured, currently researchers are investigating issues or phenomena in a third-person perspective. Perhaps video-based or VR-based data could unveil insights that were overlooked in the past.

*Posted on 10/11/2021*

Nicolas Chaillan, the Pentagon's former Chief Software Officer (CSO), told the *Financial Times* that China has won the artificial intelligence battle with the US and is heading towards global dominance in key technological sectors. According to Chaillan, "We have no competing fighting chance against China in 15 to 20 years. Right now, it's already a done deal; it is already over in my opinion." Chaillan blamed the gap on slow innovation, the reluctance of U.S. companies such as Google to work with the government on AI, and delay due to extensive ethical debates over the technology. He mocked that U.S. cyber defense capability in some government departments was at the "kindergarten level". Chailian resigned from this position to protest against the culture of inaction and slow responses.

**English version: https://news.trust.org/item/20211011063736-r28k4**
**Chinese version**:
**https://www.worldjournal.com/wj/story/121468/5809364?from=wj_maintab_index**

**That's my take on it:** It is not the first time. Right after AT&T Bell lab invented the transistor in 1947, Sony immediately bought the license and introduced the first transistor-based radio while the US home electronics manufacturer still stayed with bulky vacuum tubes. In the 1960s Japanese automakers produced affordable, dependable, and fuel-efficient small cars, but its US competitors experimented with the first compact car in 1971. During the last several years China, South Korea, Sweden, and Finland have been investing in 5G infrastructure. However, at the present time the US still lags behind international competitors in 5G. Will the Biden administration act upon the AI gap? Never too late!

*Posted on 10/7/2021*

Today is the third day of the 2021 JMP Discovery Summit. I learned a lot from the plenary talk entitled "Facets of a diverse career" presented by Dr. Alyson Wilson, Associate Vice Chancellor for National Security and Special Research Initiatives and Professor of Statistics at North Carolina State University. Her work experience spans across academia, industry, and government. She said that her career is a testament to John Tukey's statement: "The best thing about being a statistician is that you get to play in everyone's backyard." She covered many topics in the talk. I would like to highlight some of them as follows:

Many years ago she worked in the Los Alamos National Lab as a specialist on national security science, especially on weapons of mass destruction. You may wonder what role a statistician would play in this domain. Because the US signed the nuclear test-ban treaty, since the 1990s no comprehensive tests of reliability have been made to the US nuclear weapons. Alternatively, historical and simulation data were utilized by statisticians like her for reliability analysis. We are not 100% sure whether the missile works until we push the button!

Although Dr. Alyson was trained in traditional statistics, under her leadership NC State University established the Data Sciences Initiative for coordinating DS-related resources and works across ten departments in the university. In March 2021 NC State University launched a university-wide data science academy. The academy aims to enhance the infrastructure, expertise and services needed to drive data-intensive research discoveries, enhance industry partnerships, and better prepare their graduates to succeed in a data-driven economy.

https://research.ncsu.edu/dsi/
https://news.ncsu.edu/2021/03/nc-state-launches-data-science-academy/

**That's my take on it:** In the Q & A session I asked her: "The US collects a lot of data related to the COVID19 pandemic, but our countermeasures against the pandemic is not as effective as some Asian countries (e.g. Taiwan and Singapore). Do you think there is a disconnect between data analytics and decision support?" Dr. Alyson replied: we need to put good science on the data, but decision-making is multi-faceted. Something obvious to statisticians and data scientists may not be obvious to decision-makers.

I agree. Collecting and analyzing data is important, but at the end of the day the most important thing is what we do with the information.

*Posted on 10/6/2021*

Recently Mo Gawdat, formerly the Chief Business Officer for Google's moonshot organization, told *Times* Magazine that we are getting closer and closer to **AI singularity**, the point in time that AI becomes self-aware or acquires superpower beyond our control. He believed that it is inevitable for AI to become as powerful as the Skynet in "Terminator." At that point we will helplessly sit there to face the doomsday brought forth by god-like machines. Why did he make such a bold claim? Mo Gawdat said that he had his frightening revelation while working with AI developers at Google to build robotic arms. Once a robot picked up a ball from the floor, and then held it up to the researchers. Mo Gawdat perceived that the robot was showing off.

https://futurism.com/the-byte/google-exec-ai-god?utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+October+1st%2C+2021&utm_campaign=02102021

**That's my take on it:** As a psychologist, I think Mo Gawdat's concern is a result of **anthropomorphism**, a tendency of seeing human-like qualities in a non-human entity. It happens all the time e.g. we project our human attributes to pets. Now this disposition extends to robots. However, even though an AI-enabled robot acts like a human, it doesn't necessarily imply that the robot is really self-conscious or has the potential to become self-aware. I don't worry about terminators or Red Queen (in the movie "Resident Evil"), at least not in the near future!

*Posted on 10/5/2021*

Today is the second day of the 2021 JMP Discovery Summit. I would like to highlight what I learned from the plenary session entitled "Delicate Brute Force." The keynote speaker is John Sall, co-founder of SAS Institute and the inventor of JMP. In the talk Sall pointed out that traditional clustering and data reduction methods are very inefficient to process big data. To rectify the situation, Sall experimented with several new methods, such as vantage point trees, hybrid Ward, randomized singular value decomposition (SVD), multi-threaded randomized SVD…etc. Improvements were made bit by bit. For example, in a big data set containing 50,000 observations and 210 variables, it took 58 minutes to process the data in R's fast cluster. Fast Ward in JMP cut the processing time down to 8 minutes while the new hybrid Ward took 22 seconds only. Further improvements reduced the processing time to 6.7 seconds.

https://discoverysummit.jmp/en/2021/usa/home.html

**That's my take on it**: No doubt analytical algorithms are getting better and better, but very often the adoption rate cannot keep up the pace of technological innovation. I foresee that in the near future standard textbooks will not include hybrid Ward or multi-threaded randomized SVD. On the contrary, I expect widespread resistance. Think about what happened to Bruno, Copernicus, and Galileo when they proposed a new cosmology. Look at how US automakers ignored Edwards Deming. Perhaps we need another form of delicate brute force for psychological persuasion.

*Posted on 9/29/2021*

Recently Bernard Marr, an expert on enterprise technology, published two articles on *Forbes*, detailing his prediction of AI trends. In both articles Marr mentioned the trend of no- or low-code AI. As a matter of fact, not every company has the resources to hire an army of programmers to develop AI and machine learning applications. As a remedy, many of them started considering no- or low-code and self-service solutions. For example, Microsoft and other vendors have been developing natural language processing tools for users to build queries and applications by speaking or writing natural languages (e.g. "Computer! Build a time-series analysis of revenues by product segment from 2015-2021. I want the report in 30 minutes, or else!")
Marr, B. (2021, September 24). The 7 biggest artificial intelligence (AI) trends in 2022.

**Retrieved from:**
**https://www.forbes.com/sites/bernardmarr/2021/09/24/the-7-biggest-artificial-intelligence-ai-trends-in-2022/?sh=36dcfc022015&utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+September+24th%2C+2021&utm_campaign=25092021**

Marr, B. (2021, September 27). The 5 biggest technology trends in 2022. Forbes.

**Retrieved from:**
**https://www.forbes.com/sites/bernardmarr/2021/09/27/the-5-biggest-technology-trends-in-2022/?sh=126c97192414**

**That's my take on it:** History is cyclical. When I was a student, programming skills were indispensable. In 1984 Apple revolutionized the computing world by implementing the graphical user interface (GUI) on Mac OS (GUI was invented by the Xerox Palo Alto Research Center, not Apple). Since then GUI has made computing not only easier to operate, but more pleasant and natural. In recent years coding has become a hot skill again. Once a student told me, "employers don't want a data analyst doing drag-and-drop, point-and-click…etc." Not really. As experienced data analyst Bill Kantor said, many tasks are faster and easier to perform in applications with GUI than by programming. Today many corporations are aware of it and therefore they are looking for faster and no- or low- code solutions. But you don't need to wait for natural language processing. Conventional GUI is good enough to make your life easier!

*Posted on 9/19/2021*

Today is the last day of Data Con LA 2021. I really enjoy the talk “Catch me if you can: How to fight fraud, waste, and abuse using machine learning and machine teaching” presented by Cupid Chan. Dr. Chan was so humorous that he boldly claimed, “While others may take days or weeks to train a model, based on my rich experience in AI, I can build a model guaranteed with 99.9% accuracy within 10 seconds!” The fool-proof approach is: “declare that everything is NOT fraud!” Even though fraud is prevalent (credit card fraud, health care fraud, identity theft…etc.), the majority of all transactions and events (99.9%) are legitimate. Consequently, a model that yields high predictive accuracy could be totally useless. This problem also occurs in spotting manufacturing defects, diagnosing rare diseases, and predicting natural disasters. There are different approaches to rectify the situation, including **random under sampling (RUS).** For example, when a data set is composed of 4,693 positive and 54,333,245 negative cases, all positive cases should be kept, of course, but only a subset of negative cases are randomly selected for machine learning. By doing so the algorithm would not over-learn from an extremely asymmetrical data set. Feeding this subsample into **Google’s TensorFlow Boosted Tree Classifier**, Chan found that the predictive accuracy is about 85%, rather than 99.9%. But this reduction is a blessing in disguise!

**That’s my take on it**: There are many overlapping ideas between traditional statistics and modern data science. Conceptually speaking, RUS is similar to the **case-control design** in classical research methodologies. For example, in a study that aims to identify factors of illegal drug use at schools, it is extremely difficult, if not impossible, to recruit students who admit using illegal drugs. A viable approach is to carpeting all the students in a school using anonymous surveys. It turned out that 50 out of 1,000 students reported drug use. However, if these 50 cases are compared against 950 controls (no drug use), the variances of the two groups would be extremely asymmetrical, thus violating the assumption of most parametric tests. To make a valid comparison, 50 non-drug users are selected from the sample by matching the demographic and psychological characteristics of the 50 cases (Tse, Zhu, Yu, Wong, & Tsang, 2015). As such, learning traditional statistics can pave the way to learning data science and artificial intelligence.

**Reference**

Chan, C. (2021, September). Catch me if you can: How to fight fraud, waste, and abuse using machine learning and machine teaching. Paper presented at Data Con LA 2021, Online. **https://www.youtube.com/watch?v=OtWqTviKIpc**

Tse, S., Zhu, S., **Yu, C. H.**, Wong, P., & Tsang, S. (2015). An ecological analysis of secondary school students' drug use in Hong Kong: A case-control study. *International Journal of Social Psychiatry, 10*, 31-40. DOI: 10.1177/0020764015589132. Retrieved from **https://pubmed.ncbi.nlm.nih.gov/26060281/**

*Posted on 9/18/2021*

Today is the third day of Data Con LA 2021. Again, there are many interesting and informative sessions. The talk entitled “Too Much Drama and Horror Already: The COVID-19 Pandemic's Effects on What We Watch on TV” presented by Dr. Danny Kim (Senior Data Scientist at Whip Media) caught my attention. The theoretical foundation of his study is the **environmental security hypothesis**. According to the theory, viewers tend to look for **meaningful and serious content** in the media during tough times in order to help assuage uncertainty and anxiety. In contrast, people favor fun content when the living condition is not stressful. Utilizing big data (n = 233,284), Kim found that consumption of three genres have dropped substantially since the COVID19 pandemic:

- Drama: 8-11% drop
- Horror: 4-5% drop
- Adventure: 3-4% drop

**That’s my take on it**: The finding of this study partly corroborated with other reports. For example, in August 2020 Nielsen, a leading market measurement firm, found that news consumption (serious content) grew substantively. Nielsen found that 47% surveyed had either watched or streamed the news, making it the most popular TV genre. Yes, **it’s time to be serious!** Pandemic is too serious to be taken lightly.

*Posted on 9/17/2021*

Today is the second day of Data Con LA. Many sessions are informative and I would like to highlight one of them: “AI/ML/Data Science - Building a Robust Fraud Detection” presented by Gasia Atashian. In the talk Gasia illustrated how Amazon SageMaker is utilized to detect online fraud (see attached figure). Even though the data size is gigantic, the prediction time is cut down to 30 minutes and the cost is less than $10 a month. Moreover, the predictive accuracy improves 25%, compared with previous models. More importantly, you don’t need a superpower to run the program. Rather, what it takes is an 8-core CPU and 32GB of RAM! Her research has been published by Springer: **https://link.springer.com/chapter/10.1007/978-3-030-82196-8_33**

**That’s my take on it**: Models for fraud detection are not new. When I was a graduate student, a typical multivariate statistics class included discriminant analysis (DA), which is based on Fisher’s linear discriminant. The goal of DA is to find a linear combination of features that can classify entities or events into two or more categories (e.g. true positive, true negative). At that time it was the state of the art. But no one could foresee that in the near future a system developed by a book seller/online department store could become one of the most robust classifiers in the realm of mathematics and data analytics.

In addition, when I was a student, big data could only be analyzed by a workstation, such as SGI and Sun, or a supercomputer, such as Cray YMP and CM-5. Today if you have a computer equipped with a multi-core CPU, a GPU, and 16-32GB of RAM, you can be a data scientist! Life is like a box of chocolates; you never know what will happen next.

*Posted on 9/14/2021*

On September 9 Microsoft announced that it has formed a joint venture with the Australian Institute for Machine Learning to explore how advanced cloud computing, AI, computer vision and machine learning can be applied in space. The project scope includes building algorithms for on-board satellite data processing, developing solutions for the remote operation and optimization of satellites, as well as addressing space domain awareness and debris monitoring. According to Professor Tat-Jun Chin, Chair of Sentient Satellites at the Australian Institute for Machine Learning, the collaboration with Microsoft "will allow us to focus on the investigation on the performance of algorithms used to analyze large amounts of earth-observation data from satellites, without needing to be concerned about gaining access to space at the onset."
The announcement of Microsoft can be found at: **https://news.microsoft.com/en-au/features/microsoft-joins-forces-with-australian-institute-for-machine-learning/**
**That's my take on it**: In the past Microsoft was considered an imitator rather than an innovator. Excel replaced Lotus and Paradox, MS Word took over the word processing market from Word Perfect, Internet Explorer expelled Netscape, Windows NT dethroned Novell Netware…etc. The pattern is obvious: Microsoft reaped the fruits of other people's innovations. Nevertheless, in the era of big data and machine learning, Microsoft has **reinvented itself** to be a different type of company. Now AI features are a large part of the company's Azure Cloud service and no doubt today Microsoft is one of leaders in AI innovation. To stay relevant, every organization has to reinvent itself!

*Posted on 9/7/2021*

This "news" is 2-month old (published on July 14, 2021). Nonetheless, it is still posted on the front page of "Inside Big Data". After conducting extensive research, "Inside Big data" released a report entitled "The insideBigData Impact 50 list for Q3 2021." As the title implies, the report lists **50 most impactful companies in data science and machine learning**. According to the research team, the selection of these companies is based upon their massive data set of vendors and industry metrics. And also the research team employed machine learning to determine the ranking. The following are the top 20 only:

1. NVIDIA
2. Google
3. Amazon Web Services
4. Microsoft
5. Intel
6. Hewlett Packard Enterprise
7. DataRobot
8. Dell Technologies
9. Domino Data Lab
10. H20.ai
11. Databricks
12. Teradata
13. Qlik
14. TigerGraph
15. Snowflake
16. Kinetica
17. SAS
18. Anaconda (Python data science platform)
19. Salesforce (the parent company of Tableau)
20. OpeAI

**That's my take on it**: NVIDIA is the inventor of the graphics processing unit (GPU). But why is it considered the most impactful company for big data? The answer is: **parallel processing needs more GPUs**. Having more GPUs can enable deep learning algorithms to train larger and more accurate models. Currently two out of five world's fastest supercomputers (Sierra and Selene) are equipped with NVIDIA technologies.

Contrary to popular belief, proprietary software still has a very strong user base. For example, the ranking of SAS is higher than that of Anaconda, the platform for Python and other open source resources.

Not surprisingly, IBM (the parent company of SPSS) is not among the top 50. Besides the top 50, fifty-eight companies are on the list of **honorable mention**. Again, IBM is not there. In 2011 IBM's AI system Watson beat human experts in an epic Jeopardy match, but this halo cannot make IBM impactful today due to its legacy design.

The full article can be viewed at: **https://insidebigdata.com/2021/07/14/the-insidebigdata-impact-50-list-for-q3-2021/**

*Posted on 8/27/2021*

XGBoost is one of the most advanced machine learning algorithms in the open source community. It was introduced in 2014 by Dr. Tianqi Chen, an Assistant Professor at

Carnegie Mellon University. The latest version was released in April, 2021. XGboost has recently been dominating applied machine learning and **Kaggle competitions** for structured or tabular data. No doubt this 21$^{st}$ century algorithm is far better than the least square regression, which was developed in the 19th century. In spite of its predictive accuracy and computation efficiency, XGBoost is more popular in data science studies than academia. What is XGBoost, really? Four days ago Shreya Rao published an article entitled "XGBoost regression: Explain it to me like I'm 10-year old" on *Towards Data Science*.

**The full article can be accessed at:**
**https://towardsdatascience.com/xgboost-regression-explain-it-to-me-like-im-10-2cf324b0bbdb**

**That's my take on it**: It is a common misconception that data science is very difficult to understand and to implement. Actually, it is not. As the title of the preceding article implies, it is very easy to follow. You don't need calculus or matrix algebra; rather, the concepts involved in XGboost, such as residual, similarity, and gain, require basic arithmetic only. Besides XGboost, there are several other types of boosting algorithms, such as Adaptive boosting Algorithm (AdaBoost) and Gradient Boosting (Gradient Boosting is taught in my class "STAT 553"). To boost or not to boost? That's the question!
Posted on 8/20/2021

Recently Hewlett Packard Enterprise (HPE), a key player of high performance computing splitting from the parent company HP, released a report about the performance of HPE on SAS 9.4. According to the report, the key findings demonstrated high scalability when running SAS 9.4 using the Mixed Analytics Workload with HPE Superdome Flex 280 Server and HPE Primera Storage. These results demonstrated that the combination of the HPE Superdome Flex 280 Server and HPE Primera Storage with SAS 9.4 **delivers up to 20GB/s of sustained throughput**, up to a 2x performance improvement from the previous server and storage generation testing. The full report can be downloaded at:
**https://insidehpc.com/wp-content/uploads/2021/06/HPE-Reference-Architecture-for-SAS.pdf**
**That's my take on it**: Although open source has become more and more popular, some people might not realize that open source such as R is **limited by the memory**, and also is not capable of running **multi-thread processing**. For **high performance computing** and **big data analytics**, proprietary software apps such as SAS and IBM are still indispensable.

*Posted on 8/12/2021*

A week ago, Microsoft announced that their researchers have developed the world's largest general neural networks that utilize **135 billion parameters**. Now the new AI system is used in Microsoft's search engine, **Bing**. According to Microsoft, the enhanced Bing is able to determine whether a page is relevant to the query. For example, Bing learned that "Hotmail" is strongly associated with "Microsoft Outlook," even though the two terms are not close to each other in terms of semantic meaning. The AI system identified a nuanced relationship between them based on their contexts. After the enhancement, Microsoft recorded a 2% increase in click-through rates on the top search results.
**https://winbuzzer.com/2021/08/10/microsoft-research-meb-ai-for-bing-is-one-of-the-most-complex-models-ever-xcxwbn/**

**That's my take on it**: I tried to use the same phrases in both Google and Bing. For example, "Did Paul consult Greek philosophers?" (I deliberately left out the title "St." or "Apostle") "Scholars have high h-index"…etc. In most cases both Google and Bing returned different pages, yet most of them are highly relevant. However, for the query "Scholars have high h-index," apparently Google beats Microsoft.
Bing returned pages explaining how the h-index is measured, such as "What is a good H-index?" "What is a good H-index for a professor in Biology?" "What number in the h-index is considered a passing grade?" This is not what I want! I want to see a list of highly influential scholars. The top result shown in Google is: "Highly cited researchers (h>100)". The fourth one is: "Which researcher has the highest h-index?" Google won!
Posted on 8/6/2021

A recent report entitled "Data Science Needs to Grow Up: The 2021 Domino Data Lab Maturity Index" compiled by Domino found that 71% of the 300 data executives at large corporations are counting on data science to boost revenue growth, and 25% of them even expect double-digit growth. However, the report warned that many companies are not making proper investments to accomplish this goal.

In the survey, the participants reported different perceived obstacles to achieving the goal, as shown in the following

- Lack of data skills among employees: 48%
- Inconsistent standards and processes: 39%
- Outdated or inadequate tools: 37%
- Lack of buy-in from company leadership: 34%
- Lack of data infrastructure and architecture: 34%

**The full report can be downloaded at:**

**https://www.dominodatalab.com/resources/data-science-needs-to-grow-up/**
**That's my take on it:** To be fair, the above issues happen everywhere. The gap between the goal and the implementation always exists. It makes me remember the theory of **Management by Objectives (MBO)** introduced by Peter Drucker. MBO refers to the process of goal-setting by both management and employees so that there is a consensus about what is supposed to be done. In my opinion, neither the top-down nor the bottom-up approach alone can ensure a successful implementation of data science.

*Posted on 7/27/2021*

Currently the whole world has its eyes on the Olympic Games, and thus another interesting international competition is overlooked. Recently 50 teams from all over the world competed for a spot in the top ten of **World Data League**, an international contest of using data science to solve social problems. There are four stages in this contest and participants are required to solve a variety of problems, including public transportation, climate change, public health, and many others. All the complicated problems and voluminous data are provided by organizations sponsoring this game. After multi-stage screening, the top ten teams were selected to enter the finals during the first week of July. The final challenge is about how to improve the quality of life by reducing city noise levels. At the end the winner is an international team consisting of members from Germany, Italy, Portugal, and Australia.
**https://insidebigdata.com/2021/07/23/global-data-science-competition-gathered-brilliant-minds-to-solve-social-problems/**
**That's my take on it:** For a long time we have been training students how to write academic research papers for peer-review journals. No doubt it is valuable because a shiny vita with a long list of presentations and publications can pave the way for a successful career. Nonetheless, perhaps we should also encourage them to analyze big data for solving real-world problems. There is nothing more satisfying than seeing that someday my students can reverse the climate change in the World Data League!

*Posted on 7/15/2021*

Amazon, one of the leaders in cloud computing, will hold a free data conference on 8/19 between 9:00 AM and 3:00 PM Pacific. The conference aims to introduce the latest technology for building a modern data strategy to consolidate, store, curate, and analyze data at any scale, and share insights with anyone who needs access to the data. Registration is free and the link to register is:
**https://aws.amazon.com/events/aws-innovate/data/**

**That’s my take on it:** Besides providing data services, Amazon also developed several powerful analytical tools, such as Amazon SageMaker: **https://aws.amazon.com/sagemaker/**
Two decades ago I never imagined a book seller could become a major player in the field of data analytics or Jeff Bezos would go into space travel. No matter whether you will use Amazon’s cloud computing or not, it is a good thing to learn about how Amazon can constantly reinvent itself. Look at the fate of another book seller: Barnes and Noble (B & N). B & N has suffered seven years of declining revenue. Put it bluntly, the writing is on the wall when B & N didn’t want to go beyond its traditional boundary. **Do we want ourselves to be like Amazon or Barnes and Noble**?

*Posted on 7/15/2021*

Recently Dresner Advisory Services published the 2021 “Wisdom of Crowds Business Intelligence Market Study” to compare the strength of different vendors in business intelligence. The sample size consists of 5,000+ organizations and the research team rated various vendors by 33 criteria, including acquisition experience, value for price paid, quality and usefulness of product, quality of tech support, quality and value of consulting service…etc. The vendors are grouped into technology leaders and overall experience leaders. In this short message I would like to focus on technology leaders. According to the report, the **technology leaders** are:

- □ TIBCO
- □ SAS
- □ Amazon
- □ Tableau
- □ Microsoft
- □ ThoughtSpot
- □ Qlik

**That’s my take on it**: Never count on a single report! Several other consulting companies, such as Gartner, Forrester, and IDC, also published similar reports, and their results are slightly different. Nonetheless, some brand names appear on all or most reports, such as SAS, Microsoft, and TIBCO. In addition, some names have been re-appearing on several lists for many years. For example, TIBCO was named a leader five times in the Gartner Magic Quadrant for Master Data Management Solutions. SAS has also been recognized as a leader by Gartner Magic Quadrant for Data Science and Machine Learning for eight consecutive years. I want to make it clear that I am not endorsing any particular product. What I am trying to say is that **we need to teach students the skills needed by corporations**.

*Posted on  7/13/2021*

Yoshua Bengio, Yann LeCun, and Geoffrey Hinton are recipients of the 2018 ACM Turing Award for their research in Deep Neural Networks. In a paper published in the July issue of *the Communications of the ACM*, they shared their insights about the future of deep learning. They argued that the current form of deep learning is "fragile" in the sense that it relies on the assumption that incoming data are "independent and identically distributed" (i.i.d.). Needless to say, this expectation is unrealistic; in the real world almost everything is related to everything else. Due to the messiness of the real world, they said, "The performance of today's best AI systems tends to take a hit when they go from the lab to the field." A common solution is to feed the AI system with more and diverse data. In other words, currently AI systems are example-based, rather than rule-based. However, some scientists reverted to the classical approach by mixing data-driven neural networks and symbolic manipulation. But Bengio, Hinton, and LeCun do not believe that it can work. The full paper can be accessed at: **https://cacm.acm.org/magazines/2021/7/253464-deep-learning-for-ai/fulltext**

**That's my take on it**: The same problems described by Bengio, Hinton, and LeCun can also be found in classical statistics: unrealistic assumptions, messy data, and failure of generalizing the results from the lab to the field. As a remedy, some social scientists look for **ecological validity**. For example, educational researchers realize that it is impossible for teachers to block all interferences by closing the door. Contrary to the experimental ideal that a good study is a "noiseless" one, a study is regarded as ecologically valid if it captures teachers' everyday experience as they are bombarded by numerous distractions. I believe that the same principle is applicable to deep learning.

*Posted on 7/8/2021*

A few days ago, *InsideBigData* published an article entitled "The Rise and Fall of the Traditional Data Enterprise." The editorial team boldly claimed, "We are witnessing the death of traditional enterprise computing and storage – a real changing of the guard. Companies like Databricks, Snowflake and Palantir are obliterating companies initially thought to have been competitors: EMC, HP, Intel, Teradata, Cloudera and Hadoop."

Their argument is straight-forward: Cloud-based computing simplifies data storage and usage. The cloud platform is ideal for storing and analyzing large scale semi-structured data. In contrast, batch-based processing and relational databases for structured data are far less efficient. The full article can be accessed at:

https://insidebigdata.com/2021/07/02/the-rise-and-fall-of-the-traditional-data-enterprise/

**That’s my take on it**: History has been repeating itself. Back in the 1960s and 1970s, IBM mainframes seemed to be invincible and indispensable. However, in 1977 when Digital Equipment Corporation (DEC) introduced minicomputers running on VAX, IBM lost a big chunk of its market share to DEC. During the 1980s and 90s UNIX was gaining popularity, and to cope with the trend, DEC attempted to shift the focus of R&D its RISC technology, but it was too little and too late. In 2005 VAX ceased to exist.

Whether old players can continue to thrive depends on their adaptivity and the speed of reaction. Microsoft is a successful example. As *InsideBigData* pointed out, Microsoft Azure “have already commoditized storage at scales old-school players like EMC and HP could only have dreamed of.” Recently SAS Institute grabbed the opportunity by forming a joint venture with Microsoft in cloud computing.

Do we want ourselves to be like Hadoop and DEC/VAX, or Microsoft and SAS?

*Posted on 6/25/2021*

Although the predictive power of neural networks is supreme or even unparalleled, the process is considered a **black box** and sometimes the result is **uninterpretable**. Very often these models are fine-tuned by **numerous trial and error with big data**. Simply put, it is a **brute force** approach (Given enough computing power and data, you can always get the answer). To rectify the situation, Sho Yaida of Facebook AI Research, Dan Roberts of MIT and Salesforce, and Boris Hanin at Princeton University co authored a book entitled “The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks.” In the book they explained the theoretical framework of deep learning and thus data analysts could significantly reduce the amount of trial and error by understanding how to optimize different parameters. The book will be published by Cambridge University Press in early 2022 and the full manuscript can be downloaded from:
https://deeplearningtheory.com/PDLT.pdf?fbclid=IwAR1Yapc3x3ADeaTy7SZ5UZW8l4WU9dhoiBEjWTNv-VhaXaTRvTPiUKSYvPw
I haven’t read the whole book yet; nonetheless, I had a quick glance and found that the book is fairly accessible. As the authors said in the preface, the book is appropriate for everyone with knowledge of linear algebra, and probability theory, and with a healthy interest in neural networks.

*Posted on 6/18/2021*

China is in the midst of upgrading its military, including its tanks, missile systems, troop equipment, and fighter jets. Among the new systems being developed is AI. In a recent simulation one of the most experienced China's air force pilots, Fang Guoya, was defeated by the AI combatant system. According to Fang, early in the training it was easy for him to "shot down" the AI adversary. As you may already know, AI is capable of machine learning. After accumulating more and more data, the AI system outperformed Fang Guoya.
**https://www.businessinsider.com/china-pits-fighter-pilots-against-ai-aircraft-in-simulated-dogfights-2021-6?r=US&IR=T&utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+June+18th%2C+2021&utm_campaign=19062021**
**That's my take on it:** It is not surprising to see that the abilities of AI continue to outgrow even the best human experts. As a matter of fact, AI has been outsmarting humans for a long time. Back in 1997 IBM Deep Blue had already beaten the world chess champion after a six-game match. In 2011 IBM Watson competed against the best human contestants on Jeopardy and won the first prize. In 2016 Google's AlphaGo beat a 9-dan (the highest level) professional Go player. It is noteworthy that in 2017 a new system called AlphaZero defeated AlphaGo by 100-0! **Only AI can defeat AI!** Perhaps the future war will be fought between AI systems and humans will play a supporting role only.

*Posted on 6/11/2021*

In June the US Senate passed the bill entitled the *US Innovation and Competition Act* (USICA) with the purpose of boosting American semiconductor production, the R & D of Artificial Intelligence, and other crucial technologies. The bill approves $52 billion for domestic semiconductor manufacturing, as well as a 30 percent boost in funding for the National Science Foundation (NSF), and $29 billion for a new science directorate to focus on applied sciences. Additionally, the bill will provide $10 billion to reshape cities and regions across the country into "technology hubs," promoting R & D into cutting-edge industries and creating high-paying job opportunities.
**https://www.theverge.com/2021/6/8/22457293/semiconductor-chip-shortage-funding-frontier-china-competition-act**
**That's my take on it:** There will be many funding opportunities at NSF and other funding agencies. We should make ourselves ready to catch the wave. However, the US is facing a shortage of researchers, engineers, programmers, and other types of high-skilled workers. One of the strategies for boosting semiconductor production is to attract foreign

investment. In June Taiwan Semiconductor Manufacturing Co. (TSMC) has broken ground on a chipmaking facility in Chandler, Arizona (I lived there 10 years ago). One of the obstacles that might hinder TSMC from fully developing its chipmaking capacity is that the number of graduates related to science and engineering in the U.S. has diminished. In April, TSMC founder Morris Chang bluntly said that the U.S. lacks "dedicated talent ... as well as the capability to mobilize manufacturing personnel on a large scale." Further, a new report released by research group *New American Economy* found that for every unemployed technology worker in the US in 2020, there were more than seven job-postings for computer-related positions. Perhaps it's time to reconsider and restructure our academic curriculum!
**https://asia.nikkei.com/Business/Tech/Semiconductors/TSMC-in-Arizona-Why-Taiwan-s-chip-titan-is-betting-on-the-desert**
**https://www.cnbc.com/2021/06/10/study-employers-seek-immigrants-amid-shortage-of-high-skilled-workers.html**

*Posted on 6/9/2021*

Back in January 2020 Google set a record in the field of natural language processing by building a new model with 1.6 trillion parameters. Recently China broke the record by introducing **WuDao 2.0** carrying **1.75 trillion parameters**. WuDao 2.0 is able to understand both Chinese and English, thus providing appropriate responses in real-world situations. According to Chinese AI researcher Blake Yan, ""These sophisticated models, trained on gigantic data sets, only require a small amount of new data when used for a specific feature because they can transfer knowledge already learned into new tasks, just like human beings. Large-scale pre-trained models are one of today's best shortcuts to artificial general intelligence."
**https://www.techradar.com/news/china-outstrips-gpt-3-with-even-more-ambitious-ai-language-model?utm_source=ONTRAPORT-email-broadcast&utm_medium=ONTRAPORT-email-broadcast&utm_term=&utm_content=Data+Science+Insider%3A+June+4th%2C+2021&utm_campaign=05062021**
**That's my take on it:**
1. As natural language processing, image recognition, and other AI technologies become more and more sophisticated, researchers can go beyond structured data (e.g. numbers in a row by column table) by tapping into unstructured data (e.g. text, audio, image, movie…etc.).
2. Despite the US bans exporting crucial AI technologies to China, China has been surging ahead in research on AI and machine learning. China has at least three advantages: (a) AI needs big data; China can access massive data. (b) China is capable of training a large number of data scientists and AI researchers; Chinese students are

more willing to study STEM subjects no matter how challenging they are. (c). China tends to take bold steps to apply AI and machine learning into different domains, rather than maintaining the status quo.

*Posted on 6/7/2021*

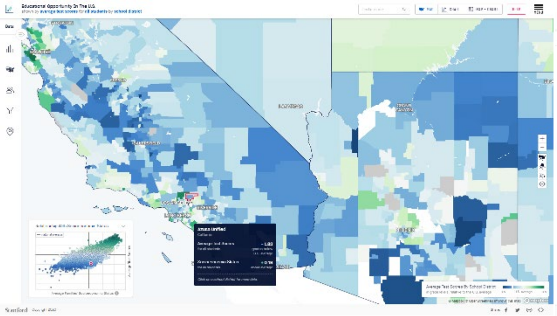
This Wednesday (June 9) the Educational Opportunity Project (EOP) at Stanford University will release new data (Version 4.1) sourced from the Stanford Education Data Archive (SEDA). This is a comprehensive national database consisting of 10 years of academic performance data from 2008-2009 to 2017-2018.

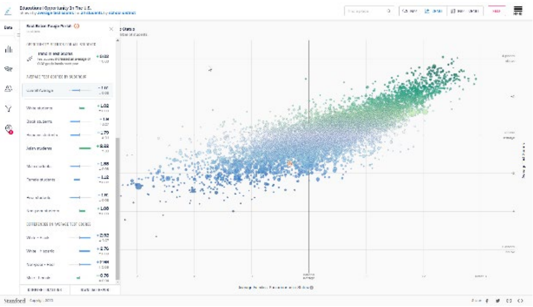
With the advance of **online interactive data visualization tools**, you don't have to wait for a year or more to see the results of this type of big data analytics. Now you can explore the data anytime anywhere on your own as long as there is a Web browser in your computer. For example, the following webpage is the **GIS map** of student test scores and socio-economic status (SES) by school district. In addition to the GIS map, the webpage also displays a scatterplot indicating a strong relationship between test scores and SES. https://edopportunity.org/explorer/#/map/none/districts/avg/ses/all/3.5/38/-97/

- □ To look for specific information about your school district, use the hand tool to move the map in order to place your state at the center.
- □ Click on the + sign on the right.
- □ Mouse hover on your school district e.g. the average test score of Azusa is -1.83 and the SES is +0.19 (see the attached PNG image "Azusa_scores_n_SES").

You can switch to a different view to interact with the chart. For example, by clicking on a particular data point, I can see the trend in test scores by ethnic groups in that particular school district (see the attached PNG image "Boston_ethnicity.png")

**That's my take on it:** I am excited by this type of **democratization** of data analytics. Rather than merely counting on what experts tell you, today you can access the data to obtain specific information that is relevant to yourself.


Educational Opportunity In The U.S.
Azusa Unified
-1.83
+0.19
Average Test Scores By School District
Stanford



*Posted on 6/4/2021*

On June 3 *Knowledge Discovery Nuggets* posted an article entitled "Will There Be a Shortage of Data Science Jobs in the Next 5 Years?" written by experienced data scientist Pranjal Saxena.

At the beginning of the article Saxena paints a gloomy picture of the future job market:

"In 2019, data scientists used to spend days in data gathering, data cleaning, feature selection, but now we have many tools in the market that can do these tasks **in a few minutes**.

On the other hand, we were trying different machine learning libraries like logistic regression, random forest, boosting machines, naive Bayes, and other data science libraries to get a better model.

But, today, we have tools like H2O, PyCaret, and many other cloud providers who can do the same model selection on the same data using the combination of other 30–50 machine learning libraries to give you the best machine learning algorithms for your data **with least error**…

Each company is aware of this fact, so after five years, when these cloud-enabled data science tools will become more efficient and will be able to provide better accuracy in much less amount of time, then **why will companies invest in hiring us and not buying the subscription of those tools?**"

At the end Saxena shows the ray of hope by saying, "Each company aims to build their product so that instead of depending on others, they can build their automated system and then sell them in the market to earn more revenue. So, yes, there will be a need for data scientists who can help industries **build automation systems** that can automate the task of machine learning and deep learning."

**That's my (Alex) take on it:** Data analysts like me are cautious of the lack of transparency and interpretability of the "black box" because the practice of handing over human judgment to the computer is not any better than blindly following the alpha level as 0.5. At most data science or machine learning should be used to augment human capabilities, not replace them. The key is to achieve an **optimal balance**. As Harvard DS researcher Brodie said, "too much human-in-the-loop leads to errors; too little leads to nonsense". I think we will need experienced data scientists to interpret the results and make corrections when the automated system makes a mistake.

Nonetheless, what Saxena described is a "good problem." I know many people still struggling with entering numbers into Excel manually. Let alone running automated tools.

**Full article:**
**https://www.kdnuggets.com/2021/06/shortage-data-science-jobs-5-years.html**

*Posted on 6/1/2021*

**AI in China's Walmart stores**

Recently Walmart, one of the world largest retailers, introduced RetailAI Fresh into China's Walmart stores for self-service customers. RetailAI Fresh is a software app developed by Malong Technologies, running on GPU-accelerated servers from Dell Technologies. Self-checkout is easy when the package has a barcode, but it becomes challenging for a scanner to recognize fresh produce products. RetailAI Fresh can rectify the situation by integrating state-of-the-art AI recognition technology into traditional self-service scales. It is noteworthy that Malong Technologies was founded by Chinese data scientists and engineers.

**https://insidebigdata.com/2021/05/24/walmart-innovates-at-the-retail-edge-with-ai/**
**https://www.linkedin.com/company/msight-ai/people/**

That's my take on it: There are a lot of brilliant Chinese computer scientists and engineers working on revolutionary products. The experiment or beta testing in China's Walmart stores is just one of many examples. We should look beyond our borders in order to absorb new ideas.

**AI for the insurance industry**

Using demographic data to customize insurance policies is not new. However, in the past customers were treated unfairly due to outdated data or incorrect predictions made by legacy software applications. Not anymore. According to a recent article on InsideBigData, today AI is capable of processing 4,000 data points in minutes and also analyzing 20 years' worth of mortality, demographic, health, and government trends for better decision support. As a result, insurance companies that utilize both AI and cloud-based data can create fairer policies to serve current and potential clients.

**https://insidebigdata.com/2021/05/25/is-ai-the-future-of-the-insurance-industry/**

That's my take on it: Today many people want to keep their privacy and complain against AI and big data, viewing them as "weapons of math destruction" or "the big brother in 1984". On the other hand, people expect corporations to improve our wellbeing by utilizing better algorithms and more accurate data. These two goals are contradictory! It is important to point out that insurance companies have been collecting customer data for many years. If AI can improve predictive models and data accuracy, I don't see a reason to oppose it.

*Posted on 5/26/2021*

1. Currently there is a special exhibition at **London's Design Museum**: Portrait paintings and drawings by an AI android named Ai-Da. Ai-Da is co-developed by robotics firm 'Engineeered Arts' and experts at the University of Oxford. Ai-Da is able to 'see' by

utilizing a computer vision system, and therefore she can create a portrait of someone in front of her. Because the creative process is based upon machine learning algorithms, she will not duplicate the same work and therefore each picture is unique. Unfortunately, I cannot visit the museum due to COVID19.

**https://www.youtube.com/watch?v=VCVgNDdIH4A**

2. IBM, along with SAS and TIBCO, is named one of the leaders in the 2021 Gartner Report of data science and machine learning platforms. Although the flagship products of IBM are IBM Watson Studio, IBM Cloud Pak for Data, IBM SPSS Modeler, and IBM Watson Machine Learning, IBM heavily invests on Python and other open source resources. Recently IBM announced that it will make the Python distribution platform **Anaconda** available for Linux on IMB Z. Anaconda is the leading Python data science platform and 25 million users use this platform for machine learning, data science, and predictive analytics.

**https://developer-tech.com/news/2021/may/18/ibm-python-data-science-platform-anaconda-linux/**
**https://www.ibm.com/analytics/data-science?p1=Search&p4=43700050329259837&p5=b&gclid=Cj0KCQjwwLKFBhDPARIsAPzPi-JK4kFax5z-_S7HdsDOUmUNn6-7LMKJf83nyIEsGNNIiAnQg45_T7kaAsGREALw_wcB&gclsrc=aw.ds**

*Posted 5/24/2021*

Daniel Kahneman won the Nobel Prize in economics in 2002 for his work on the psychology of decision-making. In response to the questions about the impact of AI on our society during a recent interview by the *Guardian*, Kahneman said, “There are going to be massive consequences of that change that are already beginning to happen. Some medical specialties are clearly in danger of being replaced, certainly in terms of diagnosis. And there are rather frightening scenarios when you’re talking about leadership. Once it’s demonstrably true that you can have an AI that has far better business judgment, say, what will that do to human leadership?... I have learned never to make forecasts. Not only can I certainly not do it – I’m not sure it can be done. But one thing that looks very likely is that these huge changes are not going to happen quietly. There is going to be massive disruption. The technology is developing very rapidly, possibly exponentially. But people are linear. When linear people are faced with exponential change, they’re not going to be able to adapt to that very easily. So clearly, something is coming… And clearly AI is going to win [against human intelligence]. It’s not even close. How people are going to adjust to this is a fascinating problem – but one for my children and grandchildren, not me.”

*Posted on 4/15/2021*

No matter whether you support developing AI or not, it is good to take multiple perspectives into consideration. A week ago, Alberto Romero published an article entitled "5 Reasons Why I Left the AI Industry" on "Towards Data Science". He complained that AI is a hype and so we should not expect to see AI at the level of human intelligence anytime soon. In addition, in his view AI becomes a blackbox and many people don't understand what is going on behind the scenes. The following is a direct quotation:

"The popularization of AI has made every software-related graduate dream with being the next Andrew Ng. And the apparent easiness with which you can have a powerful DL model running in the cloud, with huge databases to learn from, has made many enjoy the reward of seeing results fast and easy.
AI is within reach to almost anyone. You can use Tensorflow or Keras to create a working model in a month. Without any computer science (or programming) knowledge whatsoever. But let me ask you this: Is that what you want? Does it fulfill your hunger for discovering something new? Is it interesting? Even if it works, have you actually learned anything? It seems to me that AI has become an end in itself. Most don’t use AI to achieve something beyond. They use AI just for the sake of it without understanding anything that happens behind the scenes. That doesn’t satisfy me at all."

My response is: I never expect we will see an android like Commander Data or Terminator in the near future. Indeed, we don't need that level of AI to improve our performance or wellbeing. Nonetheless, it is a good strategy to aim high. If a researcher tries to publish 7 articles per year, at the end there would be 3-5 only. But if he or she sets the goal to 3 articles per year, the result would be zero! By the same token, the ultimate goal in AI seems to be unattainable, but it is how we are motivated. In addition, what Alberto described about AI programmers and users also happens among people who use traditional statistics. Some people feed the data into the computer, push a button, and then past the output into the paper without knowing what F values and p values mean. Misuse or even abuse happens everywhere. The proper way to deal with the issue is education, rather than abandoning the methodology altogether.

**The link to the full article is: https://towardsdatascience.com/5-reasons-why-i-left-the-ai-industry-2c88ea183cdd**