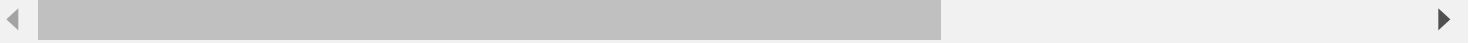


[Hot Topics](#)[IT Leadership](#)[Digital Transformation](#)[Innovation](#)[Data Analytics & AI](#)[Enterprise Applicat](#)[Home](#)**FEATURE**

7 famous analytics and AI disasters

Insights from data and machine learning algorithms can be invaluable, but mistakes can cost you reputation, revenue, or even lives. These high-profile analytics and AI blunders illustrate what can go wrong.



By Thor Olavsrud

CIO | APR 15, 2022 3:00 AM PDT



Image: IDG/Stephen Sauer

In 2017, [The Economist declared](#) that data, rather than oil, had become the world's most valuable resource. The refrain has been repeated ever since. Organizations across every industry have been and continue to invest heavily in data and analytics. But like oil, data and analytics have their dark side.

According to CIO's State of the CIO 2022 report, 35% of IT leaders say that data and business analytics will drive the most IT investment at their organization this year. And 20% of IT leaders say machine learning/artificial intelligence will drive the most IT investment. Insights gained from analytics and actions driven by machine learning algorithms can give organizations a competitive advantage, but mistakes can be costly in terms of reputation, revenue, or even lives.

[Learn [how to use data analytics for decision-making](#) and the [secrets of highly successful data analytics teams](#). | Beware the [12 myths of data analytics](#) and [sure-fire ways organizations fail at data analytics](#). | Get the latest on data analytics by [signing up for CIO newsletters](#).]

Understanding your data and what it's telling you is important, but it's also important to understand your tools, know your data, and keep your organization's values firmly in mind.

Here are a handful of high-profile analytics and AI blunders from the past decade to illustrate what can go wrong.

AI algorithms identify everything but COVID-19

Since the COVID-19 pandemic began, numerous organizations have sought to apply machine learning (ML) algorithms to help hospitals diagnose or triage patients faster. But according to the UK's Turing Institute, a national center for data science and AI, the [predictive tools made little to no difference](#).

[MIT Technology Review](#) has chronicled a number of failures, most of which stem from errors in the way the tools were trained or tested. The use of mislabeled data or data from unknown sources was a common culprit.

Derek Driggs, a machine learning researcher at the University of Cambridge, together with his colleagues, [published a paper in Nature Machine Intelligence](#) that explored the use of deep learning models for diagnosing the virus. The paper determined the technique not fit for clinical use. For example, Driggs' group found that their own model was flawed because it was trained on a data set that included scans of patients that were lying down while scanned and patients that were standing up. The patients who were lying down were much more likely to be seriously ill, so the algorithm learned to identify COVID risk based on the position of the person in the scan.

A similar example includes an algorithm trained with a data set that included scans of the chests of healthy children. The algorithm learned to identify children, not high-risk patients.

Zillow wrote down millions of dollars, slashed workforce due to algorithmic home-buying disaster

In November 2021, online real estate marketplace Zillow [told shareholders](#) it would wind down its Zillow Offers operations and cut 25% of the company's workforce — about 2,000 employees — over the next several quarters. The home-flipping unit's woes were the result of the error rate in the machine learning algorithm it used to predict home prices.

Zillow Offers was a program through which the company made cash offers on properties based on a "Zestimate" of home values derived from a machine learning algorithm. The idea was to renovate the properties and flip them quickly. But a Zillow spokesperson [told CNN](#) that the algorithm had a median error rate of 1.9%, and the error rate could be much higher, as much as 6.9%, for off-market homes.

CNN reported that Zillow bought 27,000 homes through Zillow Offers since its launch in April 2018 but sold only 17,000 through the end of September 2021. Black swan events like the COVID-19 pandemic and a home renovation labor shortage contributed to the algorithm's accuracy troubles.

Zillow said the algorithm had led it to unintentionally purchase homes at higher prices than its current estimates of future selling prices, resulting in a \$304 million inventory write-down in Q3 2021.

In a conference call with investors following the announcement, Zillow co-founder and CEO Rich Barton said it might be possible to tweak the algorithm, but ultimately it was too risky.

UK lost thousands of COVID cases by exceeding spreadsheet data limit

In October 2020, Public Health England (PHE), the UK government body responsible for tallying new COVID-19 infections, revealed that nearly 16,000 coronavirus cases went unreported between Sept. 25 and Oct. 2. The culprit? Data limitations in Microsoft Excel.

PHE uses an automated process to transfer COVID-19 positive lab results as a CSV file into Excel templates used by reporting dashboards and for contact tracing.

Unfortunately, Excel spreadsheets can have a [maximum](#) of 1,048,576 rows and 16,384 columns per worksheet. Moreover, PHE was listing cases in columns rather than rows.

When the cases exceeded the 16,384-column limit, Excel cut off the 15,841 records at the bottom.

The “glitch” didn’t prevent individuals who got tested from receiving their results, but it did stymie contact tracing efforts, making it harder for the UK National Health Service (NHS) to identify and notify individuals who were in close contact with infected patients. In a statement on Oct. 4, Michael Brodie, interim chief executive of PHE, said NHS Test and Trace and PHE resolved the issue quickly and transferred all outstanding cases immediately into the NHS Test and Trace contact tracing system.

PHE put in place a “rapid mitigation” that splits large files and has conducted a full end-to-end review of all systems to prevent similar incidents in the future.

Healthcare algorithm failed to flag Black patients

In 2019, [a study published in Science](#) revealed that a healthcare prediction algorithm, used by hospitals and insurance companies throughout the US to identify patients to in need of “high-risk care management” programs, was far less likely to single out Black patients.

High-risk care management programs provide trained nursing staff and primary-care monitoring to chronically ill patients in an effort to prevent serious complications. But the algorithm was much more likely to recommend white patients for these programs than Black patients.

The study found that the algorithm used healthcare spending as a proxy for determining an individual’s healthcare need. But according to [Scientific American](#), the healthcare costs of sicker Black patients were on par with the costs of healthier white people, which meant they received lower risk scores even when their need was greater.

The study’s researchers suggested that a few factors may have contributed. First, people of color are more likely to have lower incomes, which, even when insured, may make them less likely to access medical care. Implicit bias may also cause people of color to receive lower-quality care.

While the study did not name the algorithm or the developer, the researchers told Scientific American they were working with the developer to address the situation.

Dataset trained Microsoft chatbot to spew racist tweets

In March 2016, Microsoft learned that using Twitter interactions as training data for machine learning algorithms can have dismaying results.

Microsoft released Tay, an AI chatbot, on the social media platform. The company described it as an experiment in “conversational understanding.” The idea was the chatbot would assume the persona of a teen girl and interact with individuals via Twitter using a combination of machine learning and natural language processing. Microsoft seeded it with anonymized public data and some material pre-written by comedians, then set it loose to learn and evolve from its interactions on the social network.

Within 16 hours, the chatbot posted more than 95,000 tweets, and those tweets rapidly turned overtly racist, misogynist, and anti-Semitic. Microsoft quickly suspended the service for adjustments and ultimately pulled the plug.

“We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay,” Peter Lee, corporate vice president, Microsoft Research & Incubations (then corporate vice president of Microsoft Healthcare), wrote in a [post on Microsoft’s official blog](#) following the incident.

Lee noted that Tay’s predecessor, Xiaoice, released by Microsoft in China in 2014, had successfully had conversations with more than 40 million people in the two years prior to Tay’s release. What Microsoft didn’t take into account was that a group of Twitter users would immediately begin tweeting racist and misogynist comments to Tay. The bot quickly learned from that material and incorporated it into its own tweets.

“Although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack. As a result, Tay tweeted wildly inappropriate and reprehensible words and images,” Lee wrote.

Amazon AI-enabled recruitment tool only recommended men

Like many large companies, Amazon is hungry for tools that can help its HR function screen applications for the best candidates. In 2014, Amazon started working on AI-powered recruiting software to do just that. There was only one problem: The system vastly preferred male candidates. In 2018, [Reuters broke the news](#) that Amazon had [scrapped the project](#).

Amazon's system gave candidates star ratings from 1 to 5. But the machine learning models at the heart of the system were trained on 10 years' worth of resumes submitted to Amazon — most of them from men. As a result of that training data, the system started penalizing phrases in the resume that included the word "women's" and even downgraded candidates from all-women colleges.

At the time, Amazon said the tool was never used by Amazon recruiters to evaluate candidates.

The company tried to edit the tool to make it neutral, but ultimately decided it could not guarantee it would not learn some other discriminatory way of sorting candidates and ended the project.

Target analytics violated privacy

In 2012, an analytics project by retail titan Target showcased how much companies can learn about customers from their data. [According to the New York Times](#), in 2002 Target's marketing department started wondering how it could determine whether customers are pregnant. That line of inquiry led to a predictive analytics project that would famously lead the retailer to inadvertently reveal to a teenage girl's family that she was pregnant. That, in turn, would lead to all manner of articles and marketing blogs citing the incident as part of advice for avoiding the "creepy factor."

Target's marketing department wanted to identify pregnant individuals because there are certain periods in life — pregnancy foremost among them — when people are most likely to radically change their buying habits. If Target could reach out to customers in that period, it could, for instance, cultivate new behaviors in those customers, getting them to turn to Target for groceries or clothing or other goods.

Like all other big retailers, Target had been collecting data on its customers via shopper codes, credit cards, surveys, and more. It mashed that data up with demographic data and third-party data it purchased. Crunching all that data enabled Target's analytics team to determine that there were about 25 products sold by Target that could be analyzed together to generate a "pregnancy prediction" score. The marketing department could then target high-scoring customers with coupons and marketing messages.

Additional research would reveal that studying customers' reproductive status could feel creepy to some of those customers. According to the Times, the company didn't back away from its targeted marketing, but did start mixing in ads for things they knew pregnant women wouldn't buy — including ads for lawn mowers next to ads for diapers — to make the ad mix feel random to the customer.

Next read this

- [*7 hard truths of business-IT alignment*](#)
- [*7 ways to ruin your IT leadership reputation*](#)
- [*CIOs rethink how they hire for hard-to-fill IT roles*](#)
- [*What IT will look like in 2025*](#)
- [*14 ways to advance your IT career*](#)
- [*The 15 most valuable IT certifications today*](#)
- [*The 6 most overhyped technologies in IT*](#)
- [*12 things every CIO must get done in year one*](#)
- [*IT's most wanted: 11 traits of indispensable IT pros*](#)
- [*The rise of the revenue-generating CIO*](#)



Top 7 challenges IT leaders will face in 2022



The voice of IT leadership



POLICIES

▼

ABOUT

▼

MORE FROM CIO

▼

DIGITAL MAGAZINE

▼



Copyright © 2022 IDG Communications, Inc.

Explore the Foundry Network +

⬆

TOP OF PAGE

Do Not Sell My Info

Privacy Settings