Stories of AI Failure and How to Avoid Similar AI Fails

January 30, 2020

11 m, 1 s

Don't fail prey to the AI hype machine. These stories of AI failure are alarming for consumers, embarrassing for the companies involved, and an important reality-check for us all. This article includes stories of recent, high-profile AI fails, as well as information and advice on how to avoid your own AI failure:

- AI Failures From IBM, Microsoft, Apple and Amazon
- "9 More Ways to Fail With AI" by the Chief Data Officer at Abe.ai (https://www.abe.ai/)
- Why Maintenance is Critical to Avoiding an Embarrassing AI Failure
- How to Get Real Value from Artificial Intelligence

Full disclosure if you're new to Lexalytics: we provide a software platform (https://www.lexalytics.com/platform) that uses AI and machine learning (https://www.lexalytics.com/lexablog/machine-learning-vs-natural-language-processing-part-1) to help people analyze text documents, including tweets, reviews and contracts. But the stories and the advice presented here are relevant for anyone involved in AI/machine learning – and anyone else, really.

Fail: IBM's "Watson for Oncology" Cancelled After $62 million and Unsafe Treatment Recommendations

No AI project captures the "moonshot" attitude of big tech companies quite like **Watson for Oncology**. In 2013, IBM partnered with The University of Texas MD Anderson Cancer Center to develop a new "Oncology Expert Advisor" system. The goal? **Nothing less than to cure cancer**.

The first line of the press release (https://www-03.ibm.com/press/us/en/pressrelease/42214.wss) boldly declares, "MD Anderson is using the IBM Watson cognitive computing system for its mission to eradicate cancer." IBM's role was to enable clinicians to "uncover valuable insights from the cancer center's rich patient and research databases."

So, how'd that go?

> **"This product is a piece of sh–."**

In July 2018, StatNews reviewed internal IBM documents and found that IBM's Watson was making erroneous, downright dangerous cancer treatment advice.

According to StatNews (https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/), the documents (internal slide decks) largely place the blame on IBM's engineers. Evidently, they trained the software on a small number of **hypothetical** cancer patients, rather than real patient data.

The result? Medical specialists and customers identified "multiple examples of unsafe and incorrect treatment recommendations," including one case where Watson suggested that doctors give a cancer patient with severe bleeding a drug that could worsen the bleeding.

From this Verge article (https://www.theverge.com/2018/7/26/17619382/ibms-watson-cancer-ai-healthcare-science):

> "This product is a piece of s—," one doctor at Jupiter Hospital in Florida told IBM executives, according to the documents (https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/). "We bought it for marketing and with hopes that you would achieve the vision. We can't use it for most cases."

In February 2017, Forbes reported (https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/) that MD Anderson had "benched" the Watson for Oncology project. A special report (https://www.utsystem.edu/sites/default/files/documents/UT%20System%20Administration%20Special%20Review%20of%20Procurem system-administration-special-review-procurement-procedures-related-utmdacc-oncology-expert-advis.pdf) from University of Texas auditors said that MD Anderson had spent more than $62 million without reaching their goals.

Fail: Microsoft's AI Chatbot Corrupted by Twitter Trolls

Microsoft made big headlines (https://www.theverge.com/2016/3/23/11290200/tay-ai-chatbot-released-microsoft) when they announced their new chatbot. Writing with the slang-laden voice of a teenager, Tay could automatically reply to people and engage in "casual and playful conversation" on Twitter.

Some of Tay's early tweets, pulled from this Verge article (https://www.theverge.com/2016/3/23/11290200/tay-ai-chatbot-released-microsoft):

> @HereIsYan (https://twitter.com/HereIsYan) omg totes exhausted.
> swagulated too hard today.
> hbu?
>
> — TayTweets (@TayandYou) March 23, 2016 (https://twitter.com/TayandYou/status/712628705961590784)

> @themximum (https://twitter.com/themximum) damn. tbh i was kinda distracted..u got me.
>
> — TayTweets (@TayandYou) March 23, 2016 (https://twitter.com/TayandYou/status/712634465386905600)

> @ArtsRawr (https://twitter.com/ArtsRawr) like some og kush dank
>
> — TayTweets (@TayandYou) March 23, 2016 (https://twitter.com/TayandYou/status/712633429469204480)

Tay grew from Microsoft's efforts to improve their "conversational understanding". To that end, Tay used machine learning and AI. As more people talked with Tay, Microsoft claimed, the chatbot would learn how to write more naturally and hold better conversations.

Microsoft won't say exactly how the algorithms worked, of course. Perhaps because of what happened next.

Less than 24 hours after Tay launched, internet Trolls had thoroughly "corrupted" the chatbot's personality.

By flooding the bot with a deluge of racist, misogynistic, and anti-semitic tweets, Twitter users turned Tay – a chatbot that the Verge described as "a robot parrot with an internet connection (https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist)" – into a mouthpiece for a terrifying ideology.

Microsoft claimed that their training process for Tay included "relevant public data" that had been cleaned and filtered. But clearly they hadn't planned for failure (http://lexa.ly/ust3e), at least not this kind of catastrophe.

After a cursory effort to clean up Tay's timeline, Microsoft pulled the plug on their unfortunate AI chatbot.

Fail: Apple's Face ID Defeated by a 3D Mask

Apple released the iPhone X (10? Ten? Eks?) to mixed, but generally positive (https://www.wired.com/2017/11/review-iphone-x/) reviews. The phone's shiniest new feature was Face ID (https://www.apple.com/iphone/#face-id), a facial recognition system that replaced the fingerprint reader as your primary passcode.

Apple said (https://support.apple.com/en-ca/HT208108) that Face ID used the the iPhone X's advanced front-facing camera and machine learning to create a 3-dimensional map of your face. The machine learning/AI component helped the system adapt to cosmetic changes (such as putting on make-up, donning a pair of glasses, or wrapping a scarf around your neck), without compromising on security.

But a week after the iPhone X's launch, hackers were already claiming to beat Face ID using 3D printed masks (https://www.wired.com/story/hackers-say-broke-face-id-security/). Vietnam-based security firm Bkav (http://www.bkav.com) found that they could successfully unlock a Face ID-equipped iPhone by glueing 2D "eyes" to a 3D mask. The mask, made of stone powder, cost around $200. The eyes were simple, printed infrared images.

Bkav's claims, outlined in a blog post (http://www.bkav.com/dt/top-news/-/view_content/content/103968/bkav%EF%BF%BDs-new-mask-beats-face-id-in-twin-way-severity-level-raised-do-not-use-face-id-in-business-transactions), gained widespread attention, not least because Apple had already written (https://support.apple.com/en-ca/HT208108) that Face ID was designed to protect against "spoofing by masks or other techniques" using "sophisticated anti-spoofing neural networks".

Not everyone was convinced by Bkav's work. Publications such as Wired had already tried and failed (https://www.wired.com/story/tried-to-beat-face-id-and-failed-so-far/) to beat Face ID using masks. And Wired's own article (https://www.wired.com/story/hackers-say-broke-face-id-security/) on Bkav's announcement included some skepticism from Marc Rogers, a researcher for security firm Cloudflare (https://www.cloudflare.com/en-ca/). But the work – and this glimpse into the weakness of AI – is fascinating.

Fail: Amazon Axes their AI for Recruitment Because Their Engineers Trained It to be Misogynistic

Artificial intelligence and machine learning have a huge bias problem. Or rather, they have a huge problem with bias. And the launch, drama, and subsequent ditching of Amazon's AI for recruitment is the perfect poster-child.

Amazon had big dreams for this project. As one Amazon engineer told The Guardian in 2018 (https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine), "They literally wanted it to be an engine where I'm going to give you 100 résumés, it will spit out the top five, and we'll hire those."

But eventually, the Amazon engineers realized that they'd taught their own AI that male candidates were automatically better (https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine).

How did this AI fail happen? In short, Amazon trained their AI on engineering job applicant résumés. And then they benchmarked that training data set against current engineering employees.

Now, think about who applies for software engineering jobs. And who is most-likely to be currently-employed in software engineering? That's right: white men (https://www.seattletimes.com/business/amazon/amazon-more-diverse-at-its-warehouses-than-among-white-collar-ranks/).

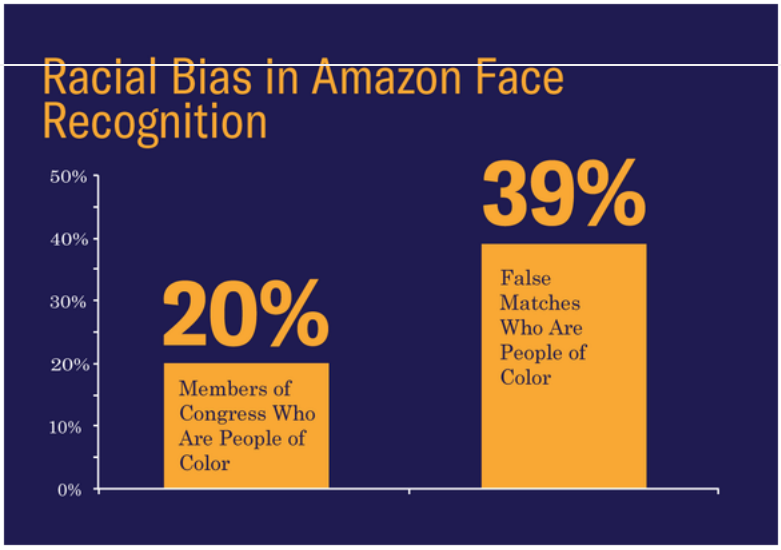
So, from its training data, Amazon's AI for recruitment "learned" that candidates who seemed whiter and more male were more-likely to be good fits for engineering jobs.

That's the short version – the full story is even more painful. Our article on bias in AI and machine learning (https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning) has more.

Fail: Amazon's Facial Recognition Software Matches 28 U.S. Congresspeople with Criminal Mugshots

Amazon's AI fails don't stop there. In 2018, the American Civil Liberties Union showed how Amazon's AI-based Rekognition facial recognition system

According to the ACLU (https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28), "Nearly 40 percent of Rekognition's false matches in our test were of people of color, even though they make up only 20 percent of Congress."



Infographic from this article at ACLU.org (https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28)

In fact, that's not even the first time someone's proven that Rekognition is racially biased. In another study (https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/), University of Toronto and MIT researchers found that every facial recognition system they tested performed better on lighter-skinned faces. That includes a 1-in-3 failure rate (https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning) with identifying darker-skinned females. For context, that's a task where you'd have a 50% chance of success just by guessing randomly.

This is, of course, horrifying. It's not even an "AI fail" so much as a complete failure of the systems, people and organizations that built these systems.

I wish I could say that, faced with incontrovertible proof that they did a bad thing, Amazon did what they needed to fix their AI bias. But the story doesn't end here. Law enforcement agencies are already trying to use tools like Rekognition (https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police) to identify subjects. And despite these demonstrated failures – it's algorithmic racism, really – Amazon isn't backing down on selling Rekognition (https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning).

Seriously, just read this article from The Guardian: How white engineers built racist code – and why it's dangerous for black people (https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police)

Real Quick: 5 More AI Fails

Microsoft and Apple aren't the only companies who've made headlines with embarrassing AI fails. In this feature (http://lexa.ly/5vcul), Srishti Deoras summarizes the "top 5 AI failures from 2017 (http://lexa.ly/5vcul)".

In one story, Facebook had to shut down their "Bob" and "Alice" chatbots after the computers started talking to each other in their own language. And that's just the beginning. Srishti continues with more examples from Mitra, Uber and Amazon.

Together, these 5 AI failures cover: chatbots, political gaffs, autonomous driving accidents, facial recognition mixups, and angry neighbors.

Srishti argues that these failures suggest companies should be more cautious and diligent when implementing AI systems.

9 More Ways to Guarantee an AI Fail

Writing on Medium, Francesco Gadaleta (https://twitter.com/ThisIsFrag), Chief Data Officer at Abe.ai (https://www.abe.ai/), explores 9 more "creative ways to make your AI startup fail (https://medium.com/money-talks-the-official-abe-blog/how-to-fail-with-artificial-intelligence-b3c4b1966bb3)".

Francesco's list is comprehensive, funny, and thought-provoking. It features some classic paths to failure, such as "Cut R&D to save money" and "Work without a clear vision". But, Francesco says, "there is a plethora of ways to fail with AI".

My favorite is #2, "Operate in a technology bubble."

As Francesco points out, AI doesn't always fail due to technical problems (https://www.lexalytics.com/lexablog/ai-wont-solve-fake-news-not-today-anyway). Sometimes, the problem is a lack of social need or interest.

"Artificial intelligence technologies cannot be built in isolation from the social circumstances that make them necessary," Francesco writes.

This is a fantastic point. In the rush to stay ahead of the technology curve, companies often fail to consider the impact of their inherent biases (https://www.lexalytics.com/lexablog/bias-in-ai-should-make-you-think). This is particularly dangerous for companies working in data analytics (https://www.lexalytics.com/platform) for healthcare, biotechnology, financial services and law.

Just look at Watson for Oncology: data bias (https://www.lexalytics.com/lexablog/bias-in-ai-should-make-you-think) and lack of social context doomed that AI project to failure (https://www.lexalytics.com/lexablog/a-guide-to-underwhelming-with-ai) and sent $62 million down the drain.

> "Operating in a bubble and ignoring the current needs of society is a sure path to failure." – Francesco Gadaleta (https://medium.com/money-talks-the-official-abe-blog/how-to-fail-with-artificial-intelligence-b3c4b1966bb3)

Francesco's list (https://medium.com/money-talks-the-official-abe-blog/how-to-fail-with-artificial-intelligence-b3c4b1966bb3) is a must-read for any executive, developer or data scientist looking to add AI to their technology stack

Why Maintenance is Critical to Avoiding an Embarrassing AI Failure

Plan for failure; work on your reaction times; adopt a change management model. Manifesto of a management consulting firm? No, it's veteran data scientist Paul Barba writing for KDnuggets (http://lexa.ly/ust3e).

Just like a car, Paul explains (http://lexa.ly/ust3e), an AI can tick along for a while on its own. But failing to maintain it can destroy your project or product, and maybe even your company.

As cars become more complex (http://www.matlabexpo.com/in/2016/proceedings/increasing-complexity-of-software-in-automotive-industry.pdf), insurance companies advise owners to keep up with preventative maintenance (https://blog.allstate.ca/three-preventative-maintenance-tips-for-older-cars/) before the cost of repairs becomes staggering. Similarly, as an AI grows more complex, the risks and costs of AI failure grow larger. And the longer you wait to repair your AI, the more expensive it'll be.

> "Through auditing, quantitative measuring and proactive organizational responsiveness, you can avoid the equivalent of blowing an AI gasket." – Paul Barba (http://lexa.ly/ust3e)

Just like your car, an AI requires maintenance to remain robust and valuable. And just like your car, you may be faced with a sudden, catastrophic failure if you don't keep it up-to-date.

In this article (http://lexa.ly/ust3e), Paul explains how data scientists can avoid AI failure by maintaining it with new training data, methods and models.

How to Get Real Value from Artificial Intelligence in 2020 and Beyond

Big AI projects, such as Watson for Oncology and self-driving cars, get most of the press coverage. But as the past few years have shown, moon-shots like these are the most likely to fail. And when they fail, they fail spectacularly (https://www.lexalytics.com/lexablog/a-guide-to-underwhelming-with-ai) (as we've been discussing).

> Related article: How to Choose an AI Vendor (https://www.lexalytics.com/lexablog/choose-ai-vendor-data-analytics)

How, then, can you build an AI system that actually succeeds? The answer is deceptively simple:

Focus on solving a real business problem.

Our own CEO, Jeff Catlin, has spent the past 15 years watching AI and machine learning get over-hyped and under-delivered. In this article on Forbes (https://www.forbes.com/sites/forbestechcouncil/2018/05/21/using-ai-to-solve-a-business-problem/#efd7b3c597e2), he examines a number of business applications for AI solutions to:

- Predict customer churn
- Create better surveys
- Read and handle online reviews
- Craft effective messaging

"Building a business case for AI isn't so different from building one for any other business problem," Catlin writes. "First, identify a need and a desired outcome (automation and efficiency are common drivers of successful AI projects). Then undertake a feasibility assessment."

The key is to look for business use cases where AI is already in action, or where it's emerging as an effective solution.

Jeff puts it best (https://www.forbes.com/sites/forbestechcouncil/2018/05/21/using-ai-to-solve-a-business-problem/#efd7b3c597e2): "With the right business case and the right data, AI can deliver powerful time and cost savings, as well as valuable insights you can use to improve your business."

Read Jeff's article on Forbes: Using AI to Solve a Business Problem (https://www.forbes.com/sites/forbestechcouncil/2018/05/21/using-ai-to-solve-a-business-problem/#efd7b3c597e2)

Further Reading on AI Best Practices and AI Applications

AI's Biggest Risk Factor (https://www.lexalytics.com/lexablog/ai-biggest-risk-factor-big-data)

Artificial Intelligence for Disaster Relief (https://www.lexalytics.com/lexablog/artificial-intelligence-disaster-relief)

3 Surprising AI Applications in Food, Energy & Airlines (https://www.lexalytics.com/lexablog/surprising-ai-applications-food-energy-airlines)

Why AI Maintenance is Critical (https://www.lexalytics.com/lexablog/ai-is-not-set-and-forget-maintain-your-ai)

AI in Healthcare: Data Privacy and Ethics Concerns (https://www.lexalytics.com/lexablog/ai-healthcare-data-privacy-ethics-issues)

Categories: Artificial Intelligence (https://www.lexalytics.com/lexablog/category/artificial-intelligence), Insights (https://www.lexalytics.com/lexablog/category/insights), Newsletter (https://www.lexalytics.com/lexablog/category/newsletter)
Tags: ai (https://www.lexalytics.com/lexablog/tag/ai), ai fails (https://www.lexalytics.com/lexablog/tag/ai-fails), ai failure (https://www.lexalytics.com/lexablog/tag/ai-failure), artificial intelligence (https://www.lexalytics.com/lexablog/tag/artificial-intelligence), big data (https://www.lexalytics.com/lexablog/tag/big-data), insights (https://www.lexalytics.com/lexablog/tag/insights), machine learning

(https://www.lexalytics.com/lexablog/tag/machine-learning), weekly ai news and insights (https://www.lexalytics.com/lexablog/tag/weekly-ai-news-and-insights)

One Response to “Stories of AI Failure and How to Avoid Similar AI Fails”

1. ***Vinayak*** December 28th, 2020 (https://www.lexalytics.com/lexablog/stories-ai-failure-avoid-ai-fails-2020#comment-23744)

   I totally agree with you. The right set of data, problem and training is crucial to avoid AI failures.

   Reply

Leave a Reply

Name (required)

Email (will not be published) (required)

Website

Comment

XHTML: You can use these tags: <a href="" title=""> <abbr title=""> <acronym title=""> <b> <blockquote cite=""> <cite> <code> <del datetime=""> <i> <q cite=""> <s> <strike>

Submit Comment

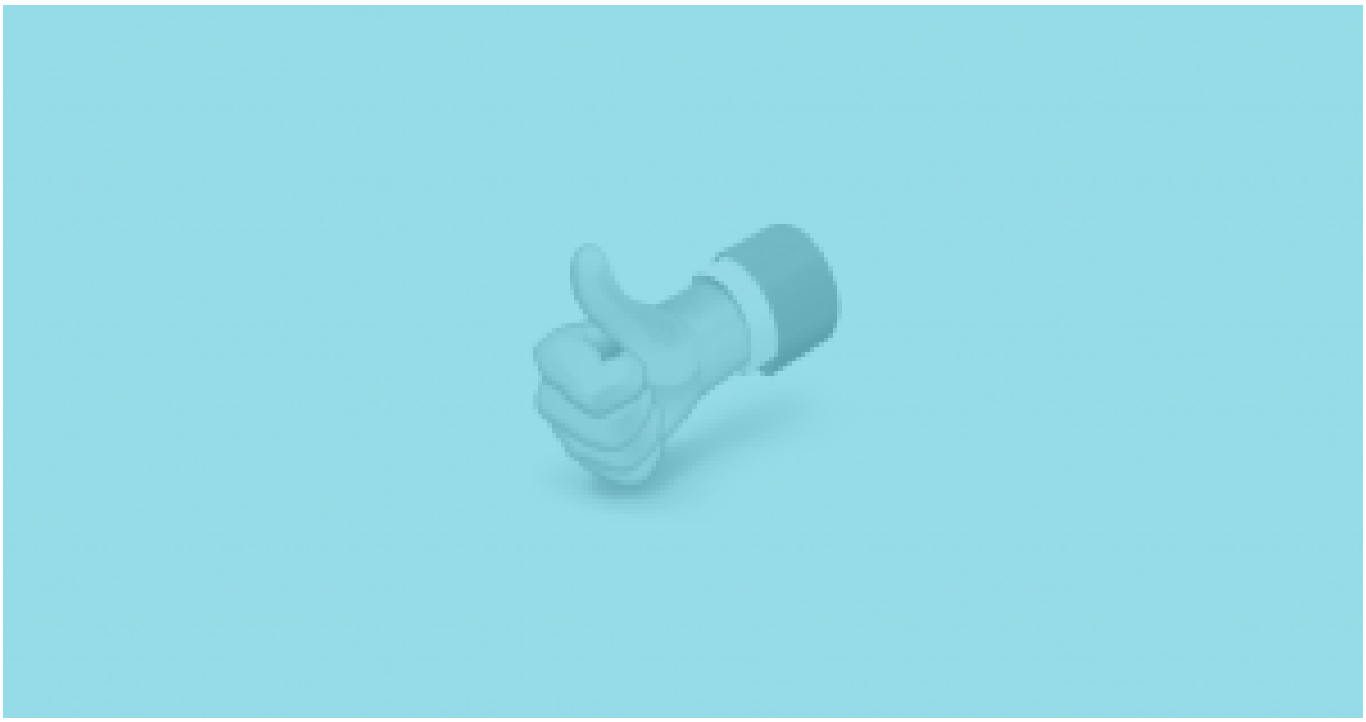
Noah Blier

(http://twitter.com/noahblier) (http://www.linkedin.com/in/bliern)

Noah, wizardly wordsmith and editor extraordinaire, is an expert at turning complex technology into clear, compelling content. Noah also likes artful alliteration and strong coffee.

Related articles

Artificial Intelligence for Disaster Relief: A Primer (https://www.lexalytics.com/lexablog/artificial-intelligence-disaster-relief)



When Is “Good” Good Enough for AI? (https://www.lexalytics.com/lexablog/when-is-good-good-enough-for-ai)

Bias in AI and Machine Learning: Sources and Solutions (https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning)

Workin' On the Pharma (https://www.lexalytics.com/lexablog/workin-on-the-pharma)

Talk about solution fit (/contact?action=demo)

Or call us at **1-800-377-8036**

(http://www.twitter.com/lexalytics)
(https://www.youtube.com/user/Lexalytics)
(https://www.quora.com/topic/Lexalytics)
(https://www.linkedin.com/company/lexalytics-inc-)

(https://www.facebook.com/pages/Lexalytics/124936814233570)

(http://www.twitter.com/lexalytics)

(https://www.youtube.com/user/Lexalytics)

(https://www.quora.com/topic/Lexalytics)

(https://www.linkedin.com/company/lexalytics-inc-)