

The Most Significant Failures When AI Turned Rogue, Causing Disastrous Results

Manjunath.R

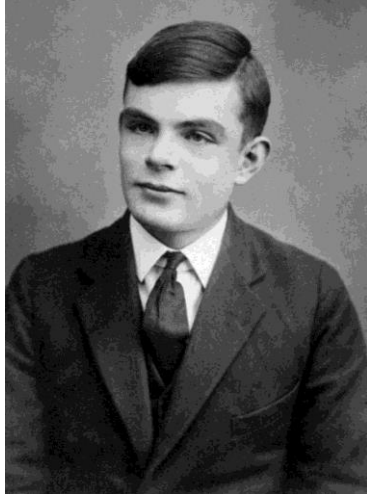
#16/1, 8th Main Road, Shivanagar, Rajajinagar, Bangalore560010, Karnataka, India

*Corresponding Author Email: manjunath5496@gmail.com

*Website: <http://www.myw3schools.com/>

Abstract

Artificial intelligence (sometimes called machine intelligence) is a part of computer science that emphasizes the creation of intelligent machines with generalized human cognitive abilities that work and reacts like intelligent beings. Artificial intelligence has made a major breakthrough in the processes, including learning (the acquisition of information for using the data), reasoning (using rules to reach definite conclusions) and self-correction -- and advancements are accelerating to present a range of new functionality for businesses. But nothing in this world can be made perfect; hence everything accompanies some notable failures and fallacies in them. Here we list some of the significant AI failures from the last decade that hint that the companies need to work harder and keep coming up with better and improved versions of their innovations.



Alan Mathison Turing was an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist. Turing was highly influential in the development of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, which can be considered a model of a general-purpose computer. Turing is widely considered to be the father of theoretical computer science and artificial intelligence.

Introduction

From self-driving cars to industrial robots, all complex real world problems are being solved with applications of intelligence (AI). Artificial intelligence (AI) is progressing rapidly and makes it possible for machines to think like humans and mimic their actions -- adjust to new inputs and perform human-like tasks by processing large amounts of data and recognizing patterns in the data. While science fiction often renders AI as robots (a machine -- especially one programmable by a computer -- capable of carrying out a complex series of actions without conscious thought or attention) with human-like characteristics, AI can encompass anything from missile guidance to tumor detection to face recognition.

The applications for artificial intelligence are countless and John McCarthy, who coined the term in 1956, defines it as: "the science and engineering of making intelligent machines." The study and design of intelligent agents -- where an intelligent agent is a system that becomes aware or conscious of its environment and takes actions which maximizes its chances of success -- can be applied to many sectors and industries including computer science, psychology, philosophy, neuroscience, cognitive science, linguistics, operations research, economics, control theory, probability, optimization, and logic. The simulation of human intelligence in machines is being tested and used in the maintenance or improvement of health industry for dosing drugs and different treatment in patients, and for surgical procedures in the hospital operating room.

A property of machines: the intelligence that the system demonstrates -- today is properly known as **Weak Artificial intelligence**, in that it is designed to perform a narrow task (such as web searches, control systems, scheduling, data mining, logistics, speech recognition, facial recognition and many others). However, the long-term goal of many technical researchers is to create Strong Artificial intelligence. While Weak Artificial intelligence may outperform humans at whatever its specific task is, like playing games or solving mathematical problems, **Strong Artificial intelligence** would outsmart humans at nearly every cognitive task.



In little over a decade, Artificial intelligence (a wide-ranging tool that enables people to rethink how we integrate information, analyze data, and use the resulting insights to improve decision making) has made leaps and bounds. Every single day, a new thousand word post showcase the most recent advancement in Artificial intelligence. Being Artificial intelligence has made remarkable breakthroughs, and many scientists dream of creating the Master Algorithm proposed by **Pedro Domingos**--which can solve all problems envisioned by humans -- failure is at the core of human advancement-- notable failures are emerging. From self-driving car accidents to Face ID hacks, AI didn't have a perfect year.

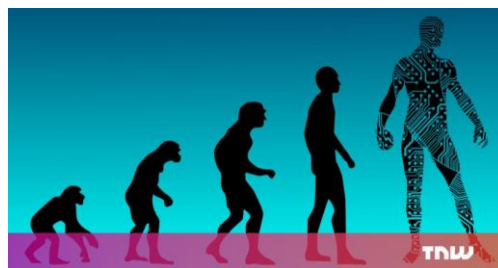
The Most Significant Failures When AI Turned Rogue, Causing Disastrous Results:

- **1959:** AI designed to be a General Problem Solver failed to solve real world problems.
- **1982:** Software designed to make discoveries, discovered how to cheat instead.
- **1983:** Nuclear attack early warning system falsely claimed that an attack is taking place.
- **2010:** Complex AI stock trading software caused a trillion dollar flash crash.
- **2011:** E-Assistant told to "call me an ambulance" began to refer to the user as Ambulance.
- **2013:** Object recognition neural networks saw phantom objects in particular noise images.
- **2015:** An automated email reply generator created inappropriate responses, such as writing "I love you" to a business colleague.
- **2015:** A robot for grabbing auto parts grabbed and killed a man.
- **2015:** Image tagging software classified black people as **gorillas**.
- **2015:** Medical AI classified patients with asthma as having a lower risk of dying of pneumonia.
- **2015:** Adult content filtering software failed to remove inappropriate content, exposing children to violent and sexual content.
- **2016:** AI designed to predict recidivism acted racist.

- **2016:** An AI agent exploited a reward signal to win a game without actually completing the game.
- **2016:** Video game **NPCs** (non-player characters, or any character that is not controlled by a human player) designed unauthorized super weapons.
- **2016:** AI judged a beauty contest and rated dark-skinned contestants lower.
- **2016:** A mall security robot collided with and injured a child.
- **2016:** The AI "Alpha Go" lost to a human in a world-championship-level game of "Go."
- **2016:** A self-driving car had a deadly accident.
- **2017:** Google Translate shows gender bias in Turkish-English translations.
- **2017:** Facebook chat bots shut down after developing their own language.
- **2017:** Autonomous van in accident on its first day.
- **2017:** **Google Allo** suggested man in turban emoji as response to a gun emoji.
- **2017:** Face ID beat by a mask.
- **2017:** AI misses the mark with **Kentucky Derby** predictions.
- **2017:** **Google Home Minis** spied on their owners.
- **2017:** Google Home outage causes near 100% failure rate.
- **2017:** Facebook allowed ads to be targeted to "**Jew Haters**".
- **2018:** Chinese billionaire's face identified as **jaywalker**.
- **2018:** Uber self-driving car kills a pedestrian.
- **2018:** Amazon AI recruiting tool is gender biased.
- **2018:** **Google Photo** confuses skier and mountain.
- **2018:** LG robot Cloi gets stagefright at its unveiling.
- **2018:** **IBM Watson** comes up short in healthcare.

While these are only a few instances of failures that have been observed so far, they are pieces of evidence to the fact that Artificial intelligence (the simulation of human intelligence processes by machines, especially computer systems) has the potential to develop a will of its own that may be in conflict with members of the human race. This is definitely a warning about the potential dangers of Artificial intelligence which should be addressed while exploring its potential interests.

"I believe there is no deep difference between what can be achieved by a biological brain and what can be achieved by a computer. It therefore follows that computers can, in theory, emulate human intelligence — and exceed it." – Stephen Hawking.

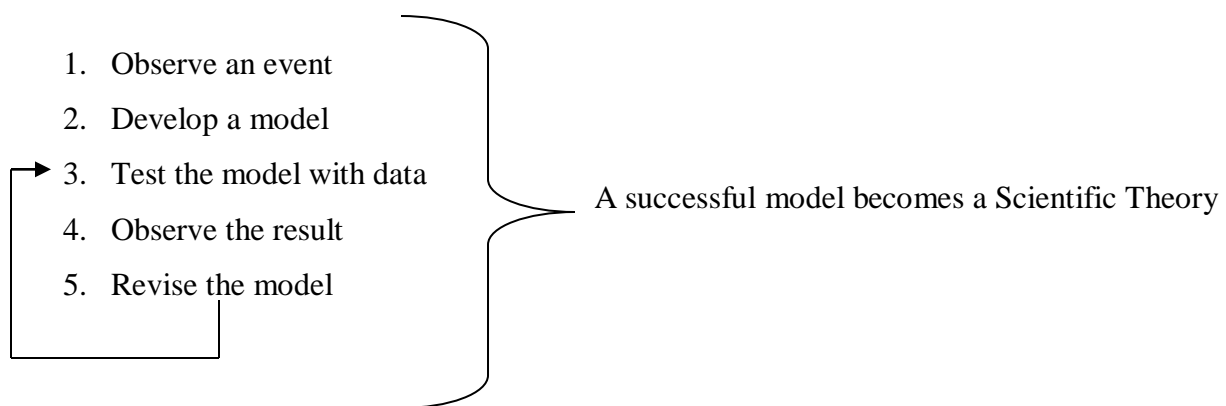


Artificial intelligence in general, context remains a challenge. **Despite Its Many Failures, why is artificial intelligence important?**

- Artificial intelligence automates repetitive learning and discovery through data.
- Artificial intelligence analyzes more and deeper data.
- Artificial intelligence adds intelligence to existing products.
- Artificial intelligence adapts through progressive learning algorithms to let the data do the programming.
- Artificial intelligence gets the most out of data.
- Artificial intelligence achieves unbelievable accuracy through deep neural networks – which was previously impossible. For example, your interactions with **Amazon Alexa**, **Google Search** and **Google Photos** are all based on deep learning – and they keep getting more precise the more we use them.

The threat of AI-charged job loss is spreading (AI and automation will eliminate the most mundane tasks). No matter what industry you're in, AI-powered bots (which can answer common questions and point users to FAQs and knowledge base articles) and software are taking a crack at it. Artificial intelligence seems to be ringing the death sound of a bell for all manner of jobs, tasks, chores and activities. From hospitality, to customer service, to home assistants, no job feels safe. Naturally, this has made people worried about the future. But is Artificial intelligence ready to take over our jobs, or even likely to do so ever? Prevalent AI- charged failures would suggest not.

Natural Language Processing → developed so that users can communicate with computers in human language.



"I find it useful to distinguish what I will call "strong" AI from "weak" or "cautious" AI. According to weak AI, the principle value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypothesis in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states."

[Searle, 1980, **Minds, Brains and Programs**]

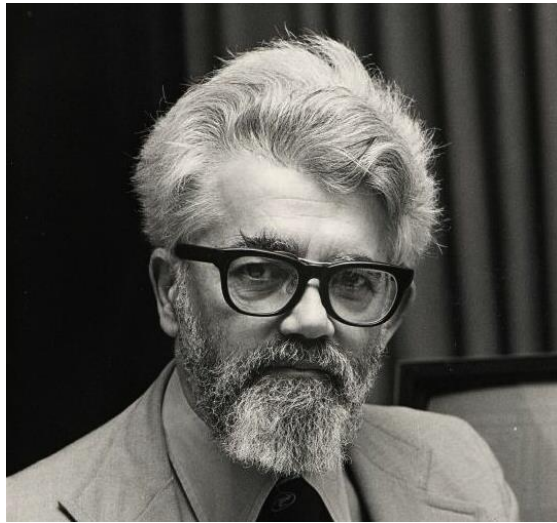
Some definitions of AI. They are organized into 4 categories:

Systems that think like humans.	Systems that think rationally.
Systems that act like humans.	Systems that act rationally.

Lifecycle of a ML Project

Conception → Experimentation → Productionizing → Usage

Potential Risks of AI
Superintelligence (a hypothetical agent that possesses intelligence far surpassing that of the brightest and most gifted human minds)
Robotic supremacy over humans
Skynet Scenarios and AI Terrorism
A unemployed Future and Accelerated Hacking
Autonomous weapons controlled by artificial intelligence may trigger the nuclear holocaust.
The WALL-E Dystopia



John McCarthy was an American computer scientist and cognitive scientist. McCarthy was one of the founders of the discipline of artificial intelligence. He co-authored the document that coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized time-sharing, invented garbage collection, and was very influential in the early development of AI.



Bill Gates

"First the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough to be a concern"



Elon Musk

"I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it's probably that"



Gary Marcus

"Once computers can effectively reprogram themselves, and successively improve themselves, leading to a so-called technological singularity or intelligence explosion the risks of machines outwitting humans in battles for resources and self-preservation cannot simply be dismissed"



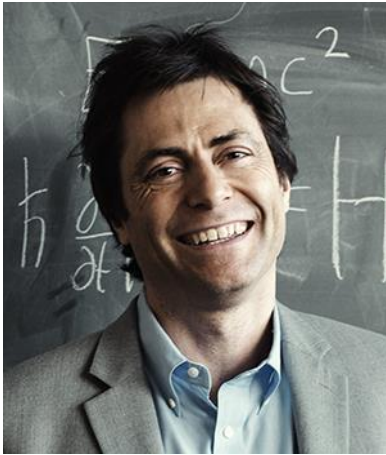
James Barrat

"I don't want to really scare you, but it was alarming how many people I talked to who are highly placed people in AI who have retreats that are sort of 'bug out' houses, to which they could flee if it all hits the fan"



Jaron Lanier

"We're still pretending that we're inventing a brain when all we've come up with is a giant mash-up of real brains. We don't yet understand how brains work, so we can't build one"



Max Tegmark

"One can imagine such technology outsmarting financial markets, out-inventing human researchers, out-manipulating human leaders, and developing weapons we cannot even understand"



Nick Bilton

"Imagine how a medical robot, originally programmed to rid cancer, could conclude that the best way to obliterate cancer is to exterminate humans who are genetically prone to the disease"



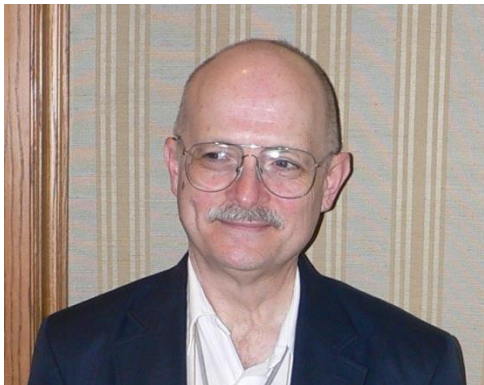
Nick Bostrom

"We cannot blithely assume that a superintelligence will necessarily share any of the final values stereotypically associated with wisdom and intellectual development in humans — scientific curiosity, benevolent concern for others, spiritual enlightenment and contemplation, renunciation of material acquisitiveness, a taste for refined culture or for the simple pleasures in life, humility and selflessness, and so forth"



Stephen Hawking

"The development of full artificial intelligence could spell the end of the human race. It would take off on its own, and re-design itself at an ever-increasing rate"



Vernor Vinge

"The competitive advantage – economic, military, even artistic – of every advance in automation is so compelling that passing laws, or having customs, that forbid such things merely assures that someone else will get them first"

Reasons for AI Failures:

- Issues of correctness, completeness and appropriateness of data
- Incorrectly coded rules
- Misunderstanding of data relationships
- Propagation of false positives at the feedback stage
- Formatting and data reconciliation issues



Edward Albert Feigenbaum is a computer scientist working in the field of artificial intelligence, and joint winner of the 1994 **ACM Turing Award**. He is often called the "**father of expert systems**."

References:

- **Artificial intelligence** By B.J. Copeland.
- **3 AI Fails and Why They Happened** By Charles A. R.
- **Artificial ignorance: The 10 biggest AI failures of 2017** By Olivia Krauth.
- **2018 in Review: 10 AI Failures** By Synced Review.
- **AI fails: why AI still isn't ready to take your job** By Howard Williams.