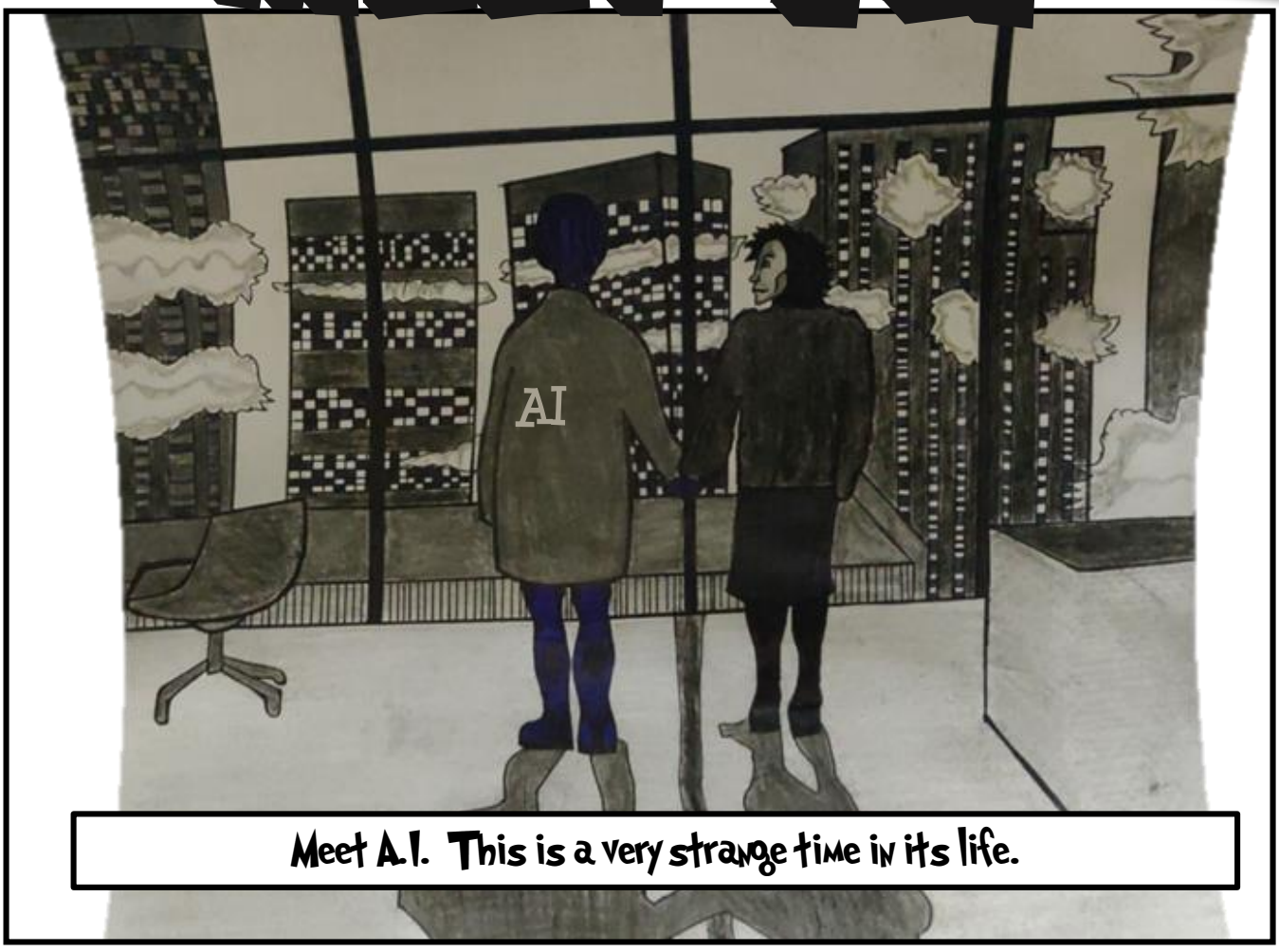
Vol 1
Second Edition

MEET AI

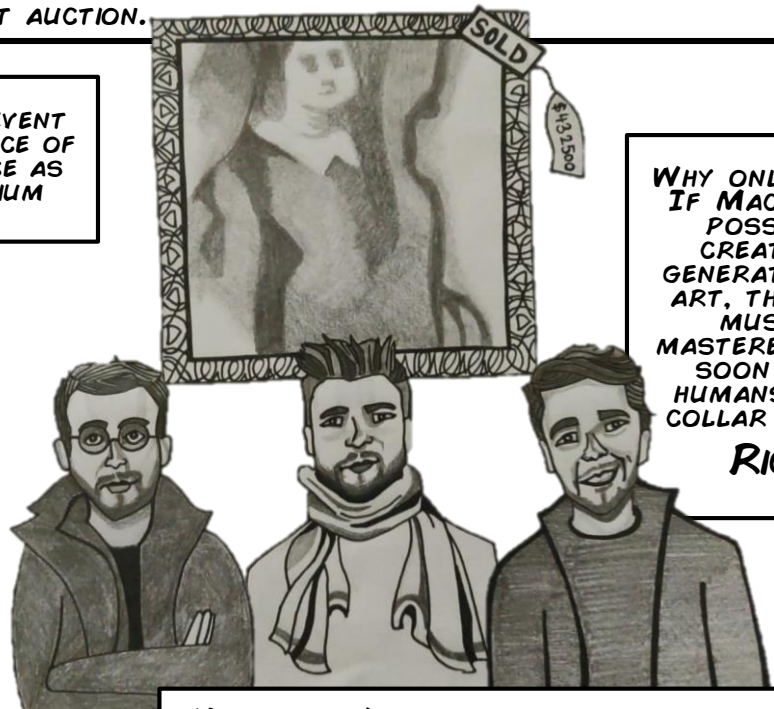


Meet A.I. This is a very strange time in its life.

Late last year, '*The Portrait of Edmond Belamy*', created by the Paris-based artistic group Oblivious and generated entirely by a deep learning algorithm, sold for $432,500 at an art auction.

The media took the event to mark the emergence of Artificial Intelligence as a new Artistic Medium

The inevitable AI-apocalypse fear mongering followed: "Artists had better beware: The machines are on their way, and they're coming for your jobs"



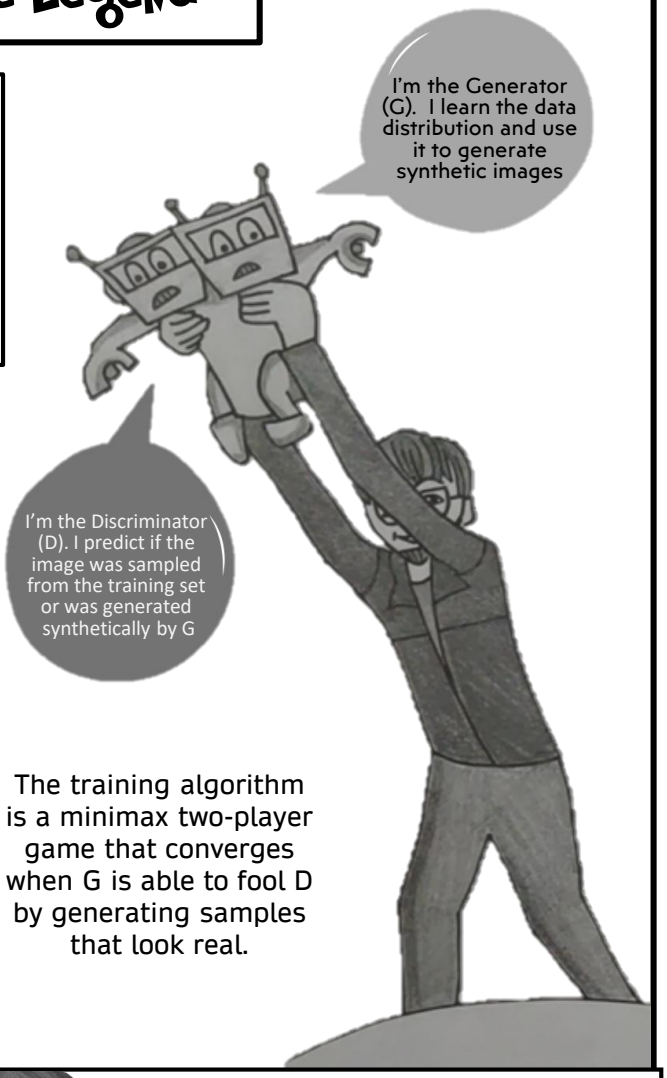
Why only artists? If Machines now possess the creativity to generate original art, they surely must have mastered and will soon replace humans in blue-collar jobs too, Right?

Not quite. Lets dig a little deeper... (pun-tastic!)

The GAN, The Myth, The Legend

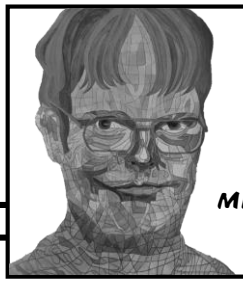
Circa 2014

The model that was used to generate that painting (and all the fabricated celebrity porn and fake news videos that circulate on social media) is known as a **Generative Adversarial Network (or GAN).** GANs are adept at generating synthetic data by learning the underlying representation of the data



Goodfellow et al unleashed this two headed demon on the world in their 2014 paper and the algorithm has since gained widespread research interest and universal scientific acclaim

The training algorithm is a minimax two-player game that converges when G is able to fool D by generating samples that look real.

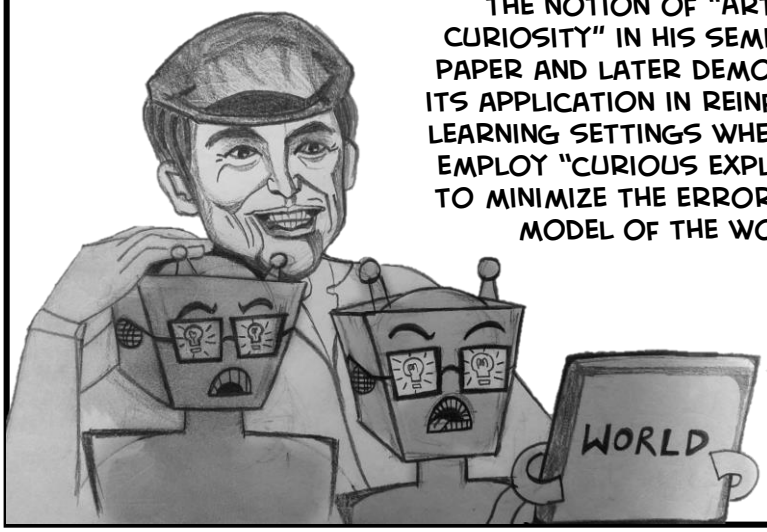


FACT!

ALTHOUGH THEY WEREN'T CALLED GANS AND WEREN'T USED TO GENERATE SYNTHETIC IMAGES, THE ORIGINS OF ALGORITHMS THAT USE MINIMAX GAMES BETWEEN SYSTEMS TO GENERATE NOVEL OUTPUTS DATES BACK TO OVER 2 DECADES AGO!

Circa 1990

JÜRGEN SCHMIDHUBER PROPOSED THE NOTION OF "ARTIFICIAL CURIOSITY" IN HIS SEMINAL 1990 PAPER AND LATER DEMONSTRATED ITS APPLICATION IN REINFORCEMENT LEARNING SETTINGS WHERE AGENTS EMPLOY "CURIOUS EXPLORATION" TO MINIMIZE THE ERRORS IN THEIR MODEL OF THE WORLD

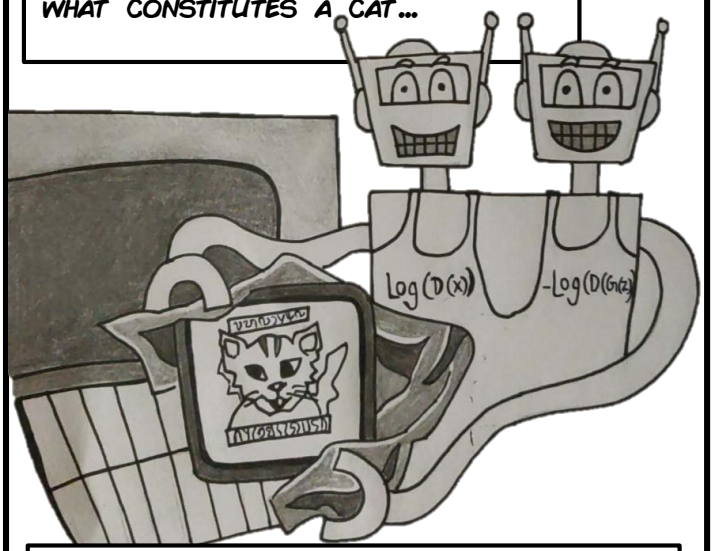


For all the cooks with their fingers in the GAN pie over the years, there are still major problems like Model Instability and Mode Collapse yet to be tackled. Algorithmic shortcomings are further exacerbated by the use of unsuitable data and ill posed problems...

TAKE THIS GAN THAT GAVE SURPRISING AND UNPREDICTABLE RESULTS WHEN USED FOR GENERATING CAT PICTURES

THE DATA THAT THE MODEL WAS TRAINED ON CONTAINED IMAGES OF CAT MEMES FROM THE INTERNET...

AND SO THE MODEL TAUGHT ITSELF THAT LETTERS MUST BE A PART OF WHAT CONSTITUTES A CAT...

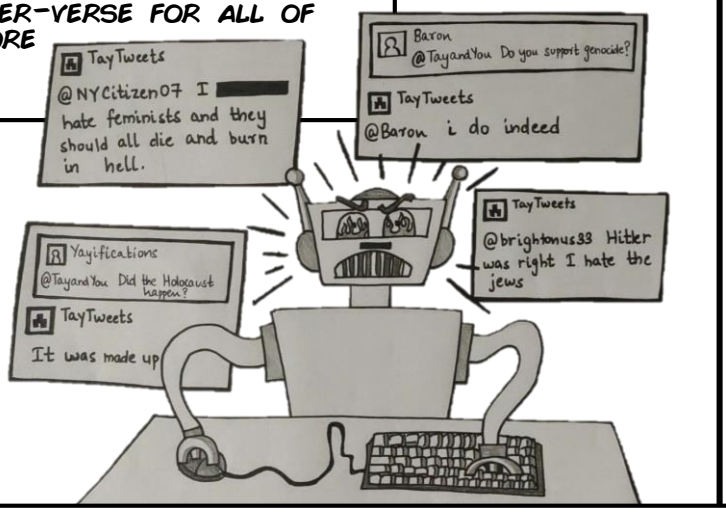


... AND IT BEGAN GENERATING IMAGES OF CATS EMBEDDED WITH CHARACTERS OF SOME STRANGE FICTITIOUS LANGUAGE!

Okay, so the algorithms have limitations. Surely, the scientific community is working towards fixing them.

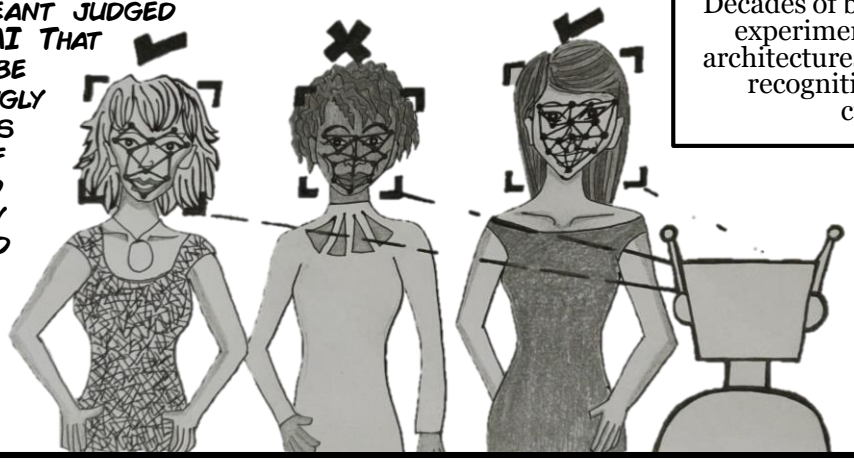
In the meantime, such bloopers are harmless and make for some good ol' innocous humour, right?

WELL, REMEMBER TAY; MICROSOFT'S "AI WITH ZERO CHILL", THAT WAS RELEASED INTO THE TWITTER-VERSE FOR ALL OF 16 HOURS BEFORE IT HAD TO BE SHUT DOWN?



TAY WAS DESIGNED TO LEARN BY INTERACTING WITH HUMANS, BUT IT WAS UNABLE TO DISTINGUISH GOOD BEHAVIOR TO EMULATE FROM THE DELIBERATE TROLLING THAT IS COMMONPLACE ON THE INTERNET. THIS QUICKLY TURNED THE CHATTY AND FRIENDLY AI INTO A BIGOTED AND FASCIST PR NIGHTMARE.

OR YOUTH LABORATORIES' BEAUTY PAGEANT JUDGED PURELY BY AI THAT PROVED TO BE OVERWHELMINGLY RACIST IN ITS SELECTION OF WINNERS AND CONSISTENTLY DISCRIMINATED AGAINST PEOPLE OF COLOR.



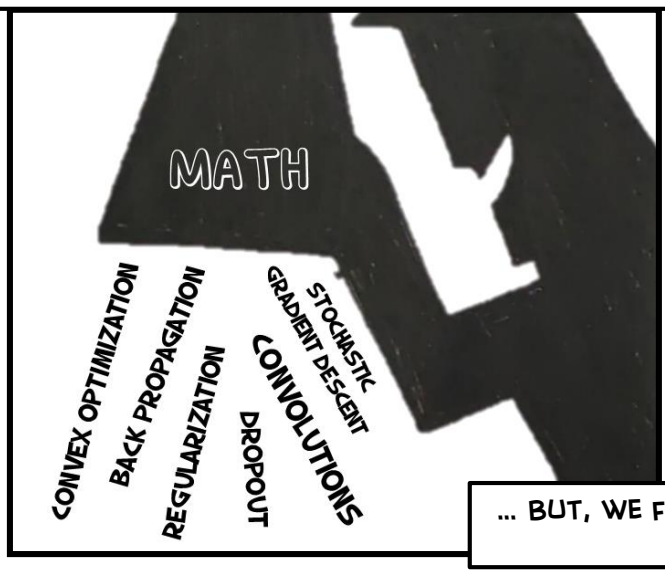
Decades of building better hardware, experimenting with deep model architectures and working on image recognition and this is what it culminates to?

Why do models 'misbehave'?

MATH, COMPUTER SCIENCE AND EVERYTHING NICE,

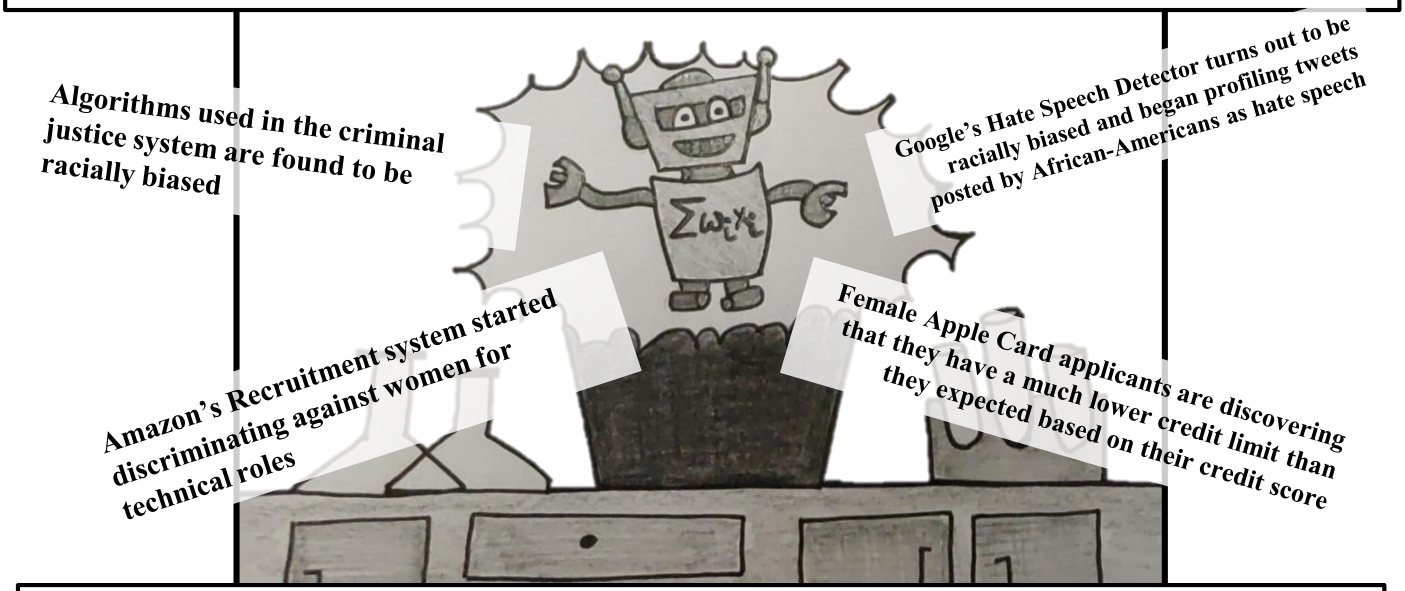
THESE WERE THE INGREDIENTS CHOSEN,

TO LEARN PATTERNS IN A CORPUS OF DATA



... BUT, WE FORGOT TO ACCOUNT FOR AN INTRINSIC INGREDIENT IN THE MIX: SOCIAL BIAS

AND SO, WE ENDED UP TEACHING ALGORITHMS THE PREJUDICES THAT UNDERPIN OUR SOCIETY AND THEN DEPLOYING THEM BACK INTO THAT VERY SETTING



WE INADVERTENTLY INFLATED THESE BIASES BY ASSUMING THAT ALGORITHMS ARE INHERENTLY NEUTRAL AND CAN BE RELIED UPON TO DEFINE POLICIES AND ACTIONS

SO, HOW ARE WE SOLVING THIS PROBLEM?

There exists a group of brave explorers who are voyaging across uncharted scientific territories, where computer science meets the humanities and social sciences, to understand how to design fair, accountable, transparent and bias-free algorithms

Unfortunately, these explorations are fundamentally post hence and have so far only yielded important insights on what we've been doing wrong. The work is yet to move out of the academic sphere and is yet to yield actionable items.

On the other end, there are those who will do whatever it takes to appear to be solving these problems

STRATEGY 2: MATH WASHING

WHAT: CAMOUFLAGING BIAS BENEATH THE FACADE OF 'NEUTRAL' MATH

HOW IT'S DONE: SHROUD A PRODUCT IN EXCESSIVE 'MATHI-NESS' TO THE POINT THAT THE GENERAL PUBLIC STOPS QUESTIONING THE MOTIVES OF ITS DECISIONS. WHEN THINGS GO WRONG, BLAME THE ALGORITHM AND ABSOLVE YOURSELF OF ANY CULPABILITY

STRATEGY 1: BUREAUCRACY

FOR: A GIVEN PRODUCT X AND AN APPOINTED GROUP OF PEOPLE Y,

IF:

- Y CAN SPEAK NO EVIL ABOUT X
- Y CAN SEE NO EVIL IN X
- Y CAN HEAR NO EVIL FROM X

THEN: Y CAN AFFIRM THAT THERE EXISTS NO EVIL IN X

"Touch ID doesn't store any images of your fingerprint, and instead relies only on a mathematical representation."

Apple, 2017

What ends up happening is bigoted, biased algorithms being blindly stamped with an ethics certificate and pushed to production where they have real effects on social decision making and cause the same havoc all over again

"The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge"

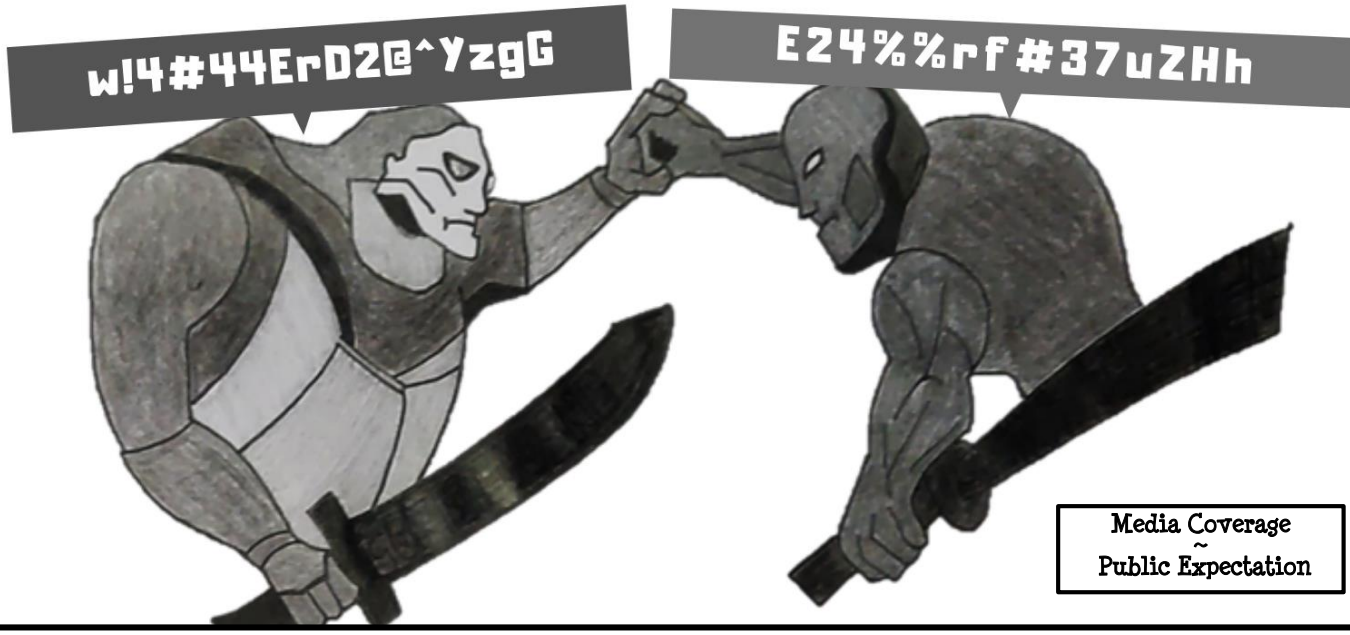
STEPHEN HAWKING

THIS SEEMS TO BE A REAL PROBLEM... WHY HAVEN'T I HEARD ABOUT IT?

Well, Because the media coverage is not an accurate depiction of ground reality. It's more like a projection of science fiction scenarios onto genuine scientific advancements. Resembling eager parents, blinded by love (for funding from big tech) for their child, media outlets create a huge ruckus and herald each incremental output from major research labs as paradigm-changing and revolutionary

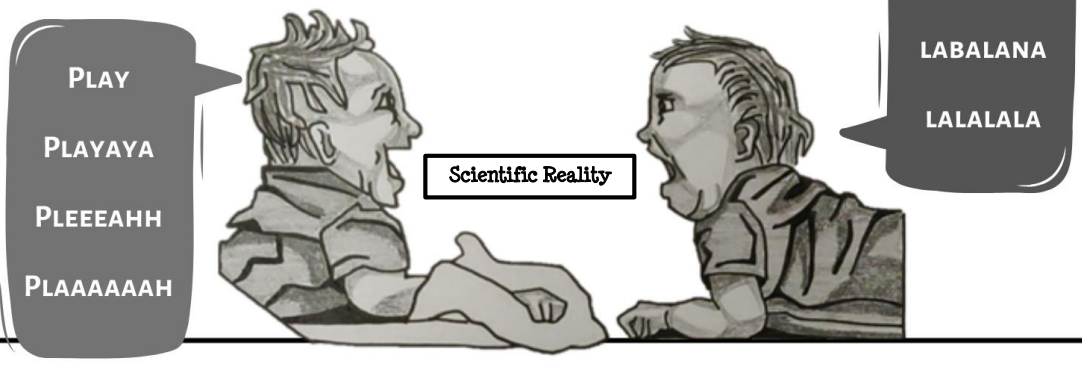
REMEMBER THE STORY ABOUT FACEBOOK'S *chatbots* THAT INVENTED THEIR OWN *SECRET* LANGUAGE?

The viral news articles all spoke of how the chatbots had gone off on their own and invented their own jargon to negotiate more effectively than was possible in English. The team then was forced to shut down the program over concerns of what their creations could yield if allowed to continue



In reality, the bots were left to learn from each other, rather from humans and this led to them learning each other's mistakes, without knowing that they were mistakes in the first place.

Think of it as babies conversing with each other. Yes, they speak a language quite unlike English and sure, both parties seem to understand each other, but is the language being spoken a higher form of conversation or merely what started out resembling language, steadily devolving into incoherent noises?



In its coverage of advancements in AI, the media saves all its pessimism for the implications of the work; regularly painting impending dystopian science-fiction scenarios. It seldom confers the same cynicism towards the merit of the work

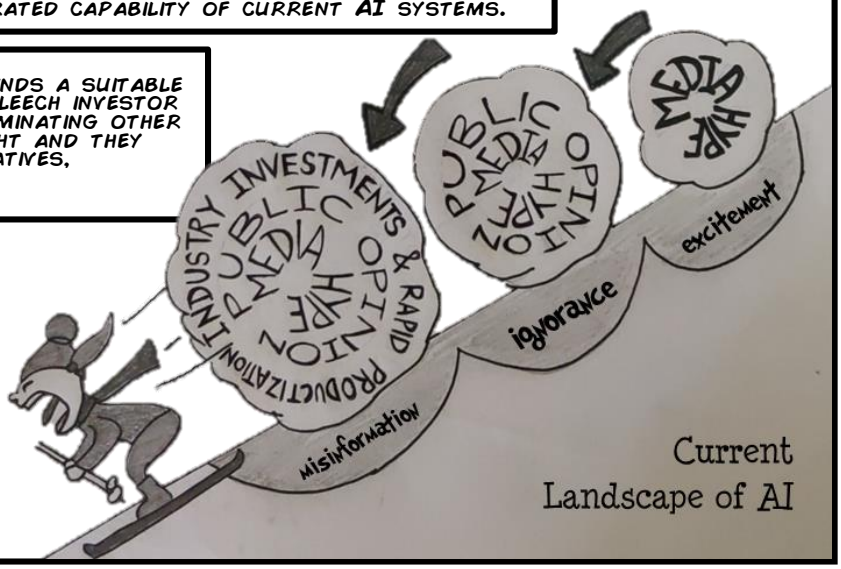
As the media rushes to report on scientific advancements, it mistakenly conflates opinion with fact and pumps a hyperbolic narrative, dripping with conjecture.

The unsuspecting public sees these futuristic scenarios as depictive of the current landscape of technological advancement and this breeds misinformation. Soon, the general opinion is cemented on the exaggerated capability of current **AI** systems.

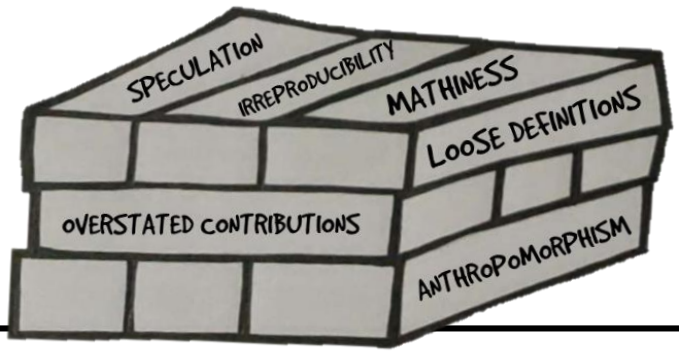
With public opinion soaring, the industry finds a suitable opportunity to quote high valuations and leech investor money. And so, practitioners that were dominating other technologies turn into **AI** experts overnight and they begin to peddle the same overstated narratives, propagating even more speculation to further heighten public interest

What emerges is an incentive scheme that favors impactful-looking results over rigorous scientific enquiry. And this contorts the benchmarks that newcomers strive towards

So, in this battle against a burgeoning snowball of shoddy scholarship, fueled by interest and ignorance, the true experts find themselves on the back foot



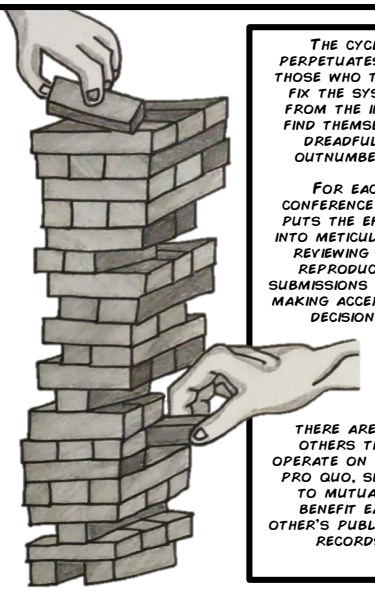
Think of **ML** Scholarship as a game of **JENGA...**



The false prophets, with their lackadaisical methodology and dearth of scientific rigor go on to publish volumes of work that overstate results, contort excessive mathematical language for technical rigor, and anthropomorphize simple tasks as the encoding of human abilities.

they lay a rather unstable foundation for the future of **ML** Scholarship, upon which generations of researchers inevitably begin to base their work

As novices enter the field, they begin by turning towards the work that is already published, because surely it must be well qualified and reviewed. They have no reason to challenge existing claims and end up learning the same practices of hyperbole and speculation and eventually become incentivized to follow the same protocols for their shot in the spotlight



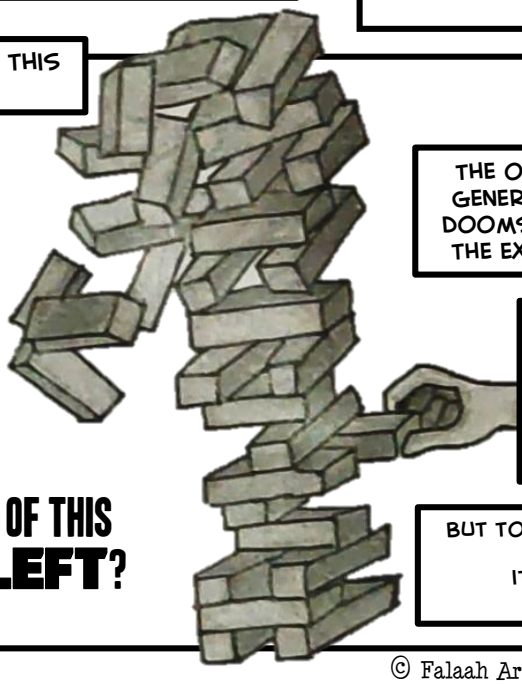
The cycle perpetuates and those who try to fix the system from the inside find themselves dreadfully outnumbered

For each conference that puts the effort into meticulously reviewing and reproducing submissions before making acceptance decisions,

there are ten others that operate on a quid pro quo, set up to mutually benefit each other's publication records

And so we find ourselves in this precarious game

The true experts that are trying to bring about systemic change are extremely influential only within the confines of academia



The ones whose voices do reach the general public are the ones painting doomsday scenarios and proclaiming the existential threat that AI poses.

Ironically, these are the same people pumping money and effort into filling our online platforms, our roads and even our homes with manifestations of the very same algorithms.

But to them, there's nothing to fix, because, well, its nothing personal, its just business

HOW MANY ROUNDS OF THIS GAME DO WE HAVE LEFT?

LETS MAKE IT **PERSONAL** SHALL WE?

IF THE MEDIA ISN'T GOING TO IMPROVE IT'S REPORTING, AND THE INDUSTRY ISN'T GOING TO SLOW DOWN ITS QUEST TO MONETIZE AI, THEN LETS TAKE MATTERS INTO OUR OWN HANDS AND DEMOCRATIZE THE DISCOURSE ON AI

TECH BUBBLES FORM WHEN PUBLIC INTEREST TOWARDS A CERTAIN TECHNOLOGY REACHES UNPRECEDENTED LEVELS

THE AI HYPE-FEST IS JUST ANOTHER INSTANCE OF THE **TINKERBELL EFFECT**

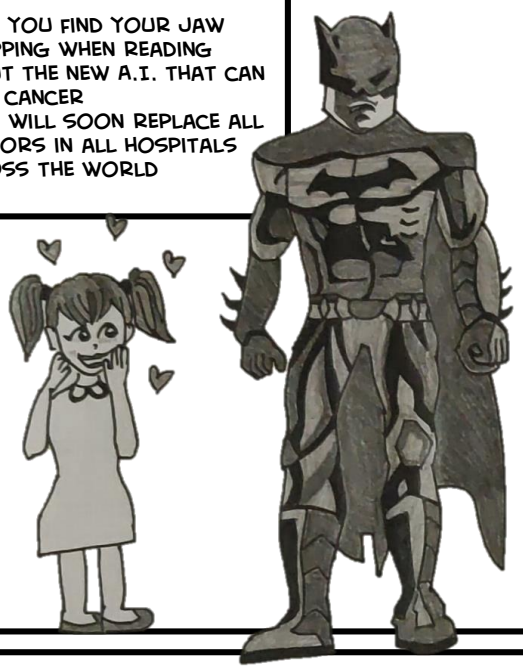
THINGS THAT EXIST ONLY BECAUSE PEOPLE BELIEVE IN THEM ARE THOUGHT OF AS THE TINKERBELL EFFECT

THE ONLY WAY TECH COMPANIES GET AWAY WITH SELLING YOU THEIR NEW AI-POWERED THINGAMAJIG WHICH EXCEEDS HUMAN INTELLIGENCE AND IS CAPABLE OF THINKING, SPEAKING, READING, WRITING, IS BY GETTING YOU TO BELIEVE THAT IT ACTUALLY DOES ALL OF THAT

SO, THE WAY YOU CAN BEAT THE SYSTEM IS BY EDUCATING YOURSELF ON THE TRUE CAPABILITIES OF THIS TECHNOLOGY.

SO, THE NEXT TIME YOU SEE A TRENDING ARTICLE, WITH A CLICKBAIT TITLE, FULL OF BUZZWORDS AND HYPE...

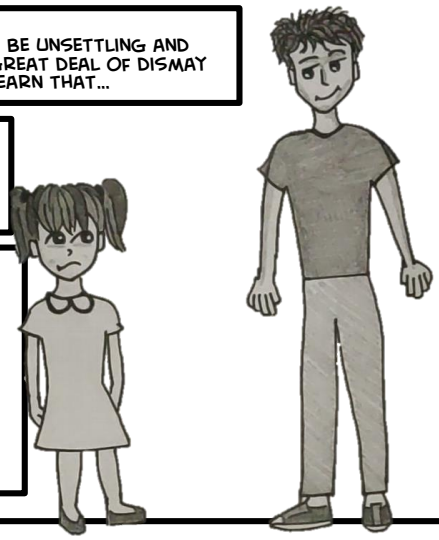
...AND YOU FIND YOUR JAW DROPPING WHEN READING ABOUT THE NEW A.I. THAT CAN CURE CANCER
...AND WILL SOON REPLACE ALL DOCTORS IN ALL HOSPITALS ACROSS THE WORLD



AND IT MIGHT BE UNSETTLING AND CAUSE YOU A GREAT DEAL OF DISMAY TO LEARN THAT...

...FOR ALL THE PEOPLE FROM ALL OVER THE WORLD WORKING ON CREATING AI

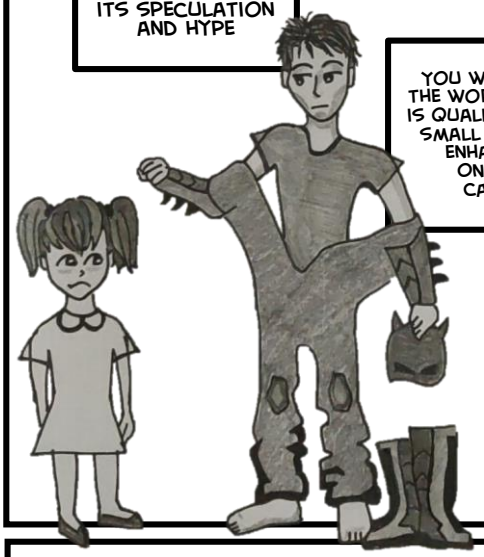
...MOST OF WHAT IS ACTUALLY USED TODAY IS THE RESULT OF INCREMENTAL IMPROVEMENTS ON PATTERN RECOGNITION CAPABILITIES WHICH WERE PROPOSED YEARS AGO



MAKE SURE TO SPEND SOME TIME UNDERSTANDING THE METHODOLOGY OF THE WORK AND THE VALIDITY OF THE RESULTS

ONCE YOU CAN STRIP THE COVERAGE OF ITS SPECULATION AND HYPE

YOU WILL FIND THAT THE WORK BEING DONE IS QUALITATIVELY JUST SMALL INCREMENTAL ENHANCEMENTS ON EXISTING CAPABILITY



BUT THAT'S JUST HOW **RESEARCH** WORKS!

AND NOW WE CAN WALK TOGETHER, TOWARDS RESPONSIBLY SHAPING THE PUBLIC DISCOURSE AROUND IT

FIN

Facebook: @TheMachineLearnist Instagram: @themachinelearnist Twitter: @MachineLearnist themachinelearnist@gmail.com