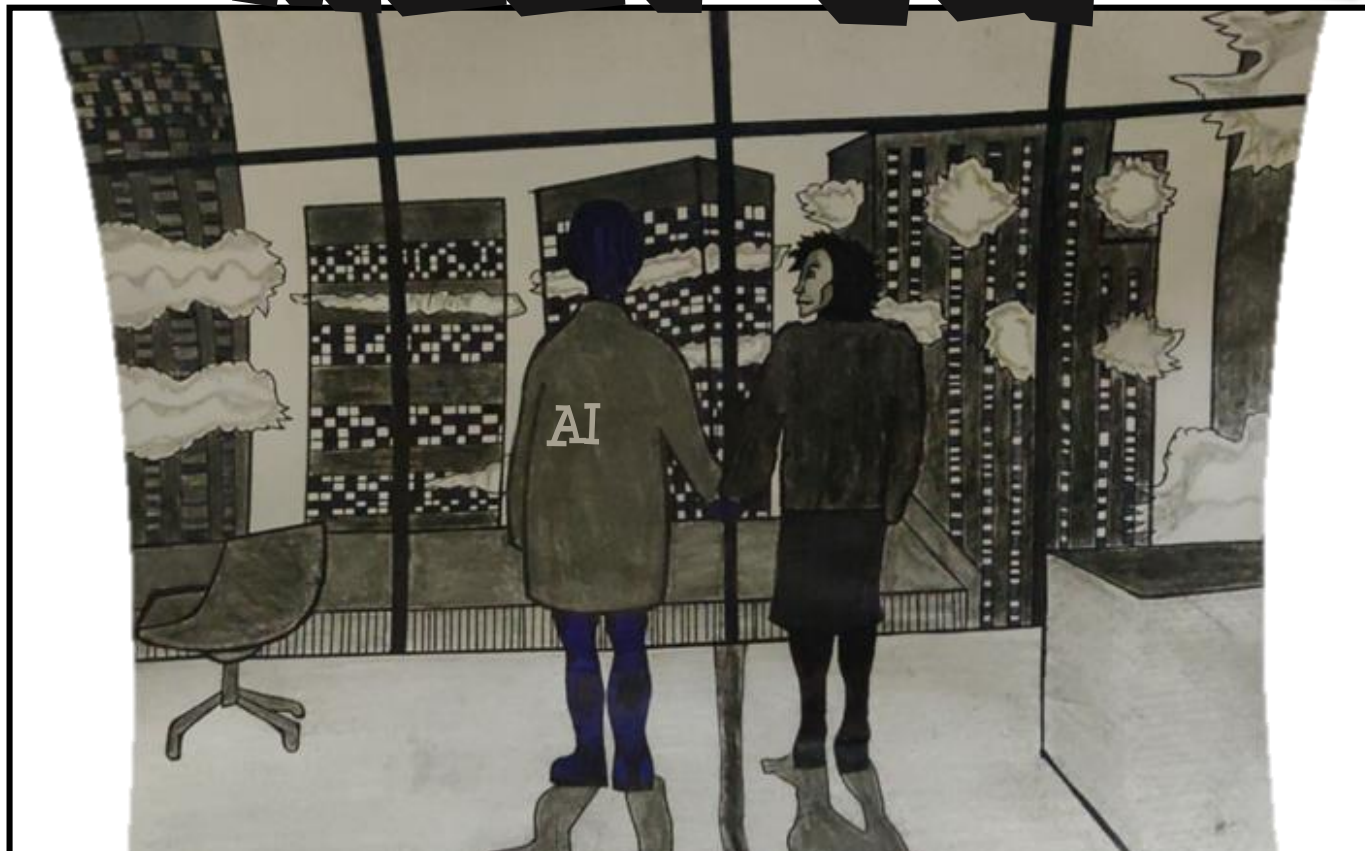


Vol 1

© February 2020,  
By Falaah Arif Khan

# 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

The model that was used to generate that painting (as well as 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 latent space and internal representation of the data distribution.

"WELL, IT'S SO DIFFERENT FROM WHAT HAS BEEN DONE BEFORE IN MACHINE LEARNING THAT, YOU KNOW, IT COULD BRING US EVEN BETTER THINGS IN THE FUTURE!"

—YOSHUA BENGIO,  
2018 ACM TURING AWARD CO-RECIPIENT



While leading researchers are optimistic about the potential that GANs hold, they do profess the several limitations of such models.

ONGOING RESEARCH ON GANS HAS UNMASKED PROBLEMS SUCH AS MODEL INSTABILITY DURING TRAINING AND MODE COLLAPSE. AS WITH MOST MACHINE LEARNING MODELS, INHERENT ALGORITHMIC SHORTCOMINGS ARE OFTEN EXACERBATED BY THE USE OF UNSUITABLE TRAINING DATA AND ILL POSED PROBLEMS...

I'm the Generator (G).  
I learn the data distribution and use it to generate synthetic images

I'm the Discriminator (D). I predict if the image was sampled from the training set or was generated synthetically by G

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

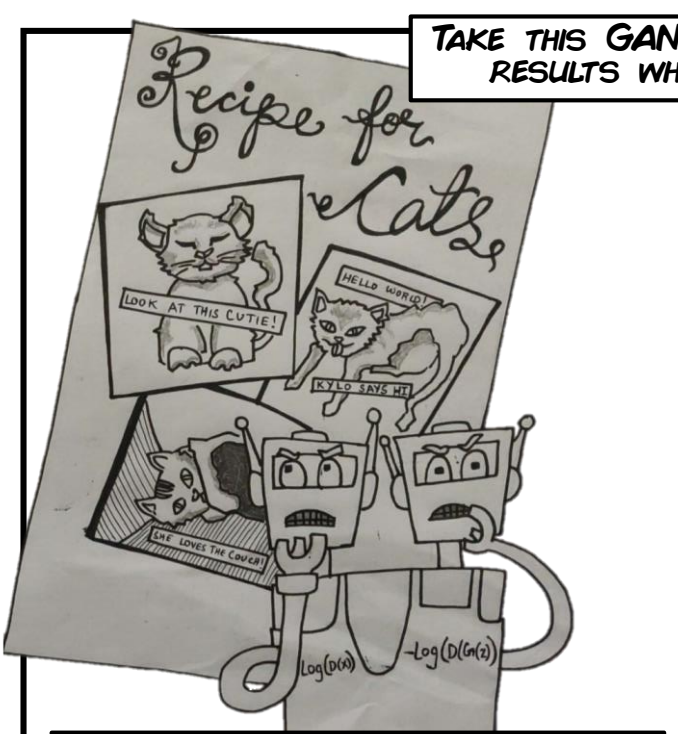
Goodfellow et al unleashed this two headed demon on the world in their 2014 paper and the algorithm has since gained widespread research interest and universal scientific acclaim

"THE GAN AND IT'S VARIATIONS THAT ARE NOW BEING PROPOSED IS THE MOST INTERESTING IDEA IN THE LAST DECADE IN MACHINE LEARNING!"

—YANN LECUN  
2018 ACM TURING AWARD  
CO-RECIPIENT

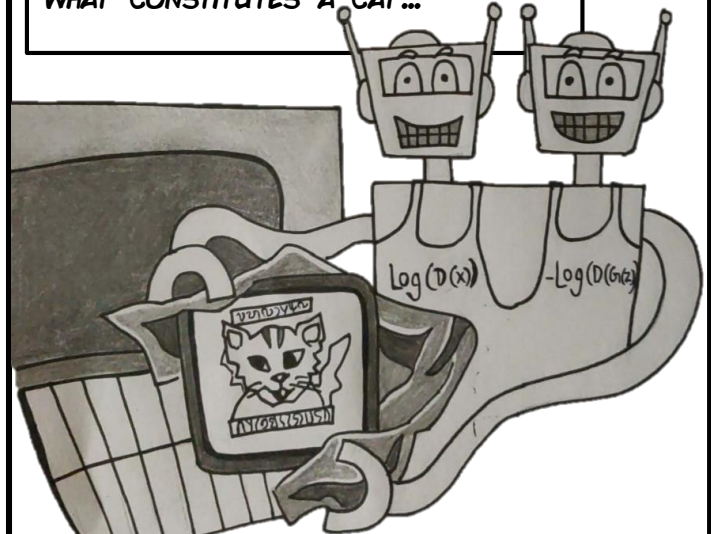


TAKE THIS GAN THAT GAVE SURPRISING AND UNPREDICTABLE RESULTS WHEN USED FOR GENERATING CAT PICTURES



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...

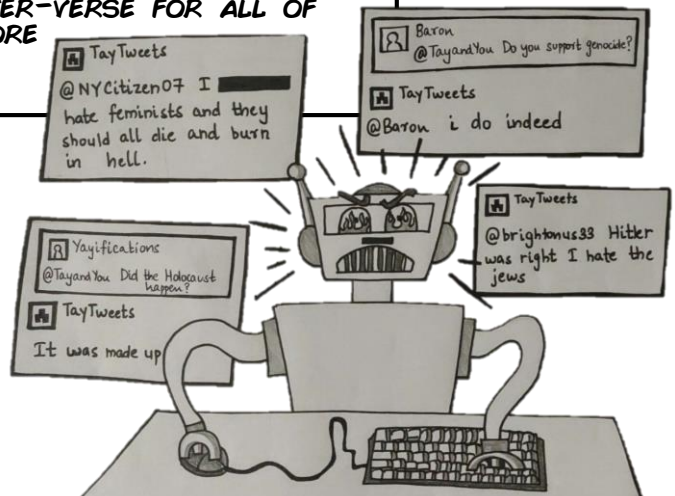


... AND IT BEGAN GENERATING IMAGES OF CATS EMBEDDED WITH CHARACTERS OF SOME STRANGE FICTITIOUS LANGUAGE!

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

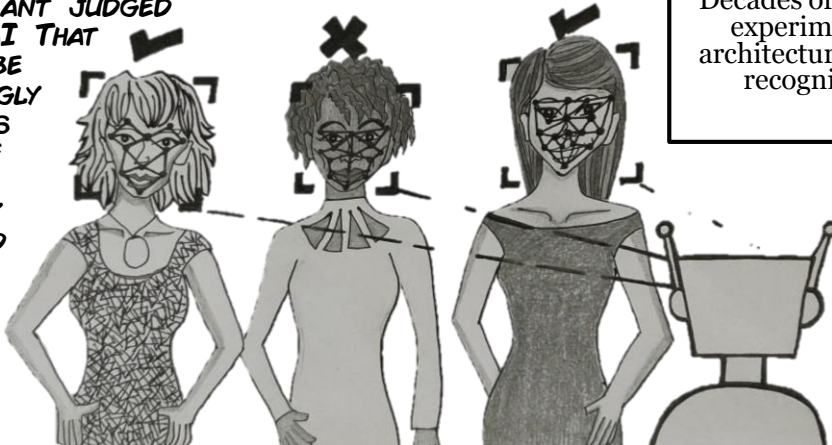
In the meantime, such bloopers are harmless and make for some good ol' innocuous humour, right?

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?

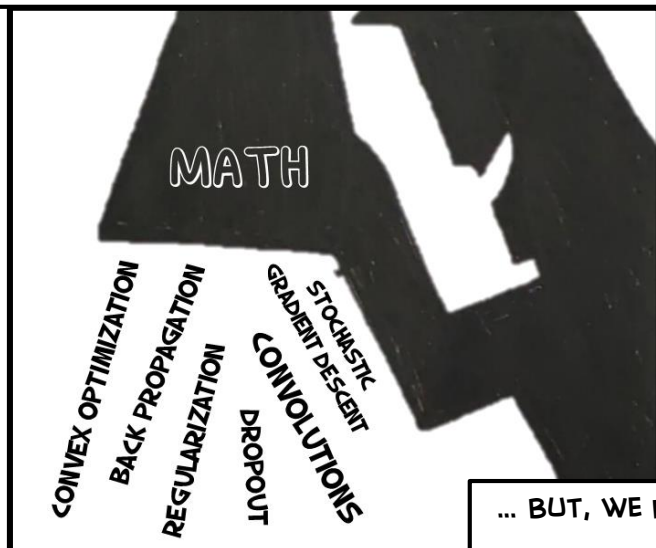
Why do models 'misbehave'?

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

Algorithms used in the criminal justice system are found to be racially biased

Amazon's Recruitment system started discriminating against women for technical roles

Google's Hate Speech Detector turns out to be racially biased and began profiling tweets posted by African-Americans as hate speech

Female Apple Card applicants are discovering that they have a much lower credit limit than they expected based on their credit score

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 UNCHARTERED 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 HOC 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



## ECORP ETHICS BOARD

### 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 HAYOC 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

ReLU stands for Rectified Linear Unit. It's an activation function widely used in Deep Learning!

Lulu just said her first word!!! She said 'relu' !!!

raaayleuu

OMG! She's a genius! She's going to grow up to solve Artificial General Intelligence!

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

w!4#44ErD2@^YzgG

E24%%rf#37uZHh

Media Coverage  
~  
Public Expectation

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?

PLAY

PLAYAYA

PLEEEAHH

PLAAAAAAH

Scientific Reality

BANANA

BALALALA

LABALANA

LALALALA

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



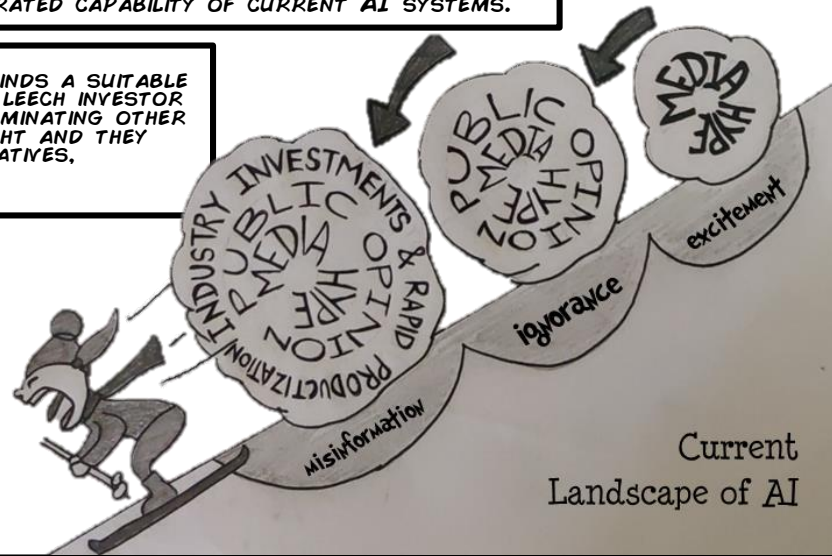
AS THE MEDIA RUSHES TO REPORT ON SCIENTIFIC ADVANCEMENTS, IT MISTAKENLY CONFLATES OPINION WITH FACT AND PUMPS A HYPERBOLIC NARRATIVE, DRIPPING WITH CONJECTURE.

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.

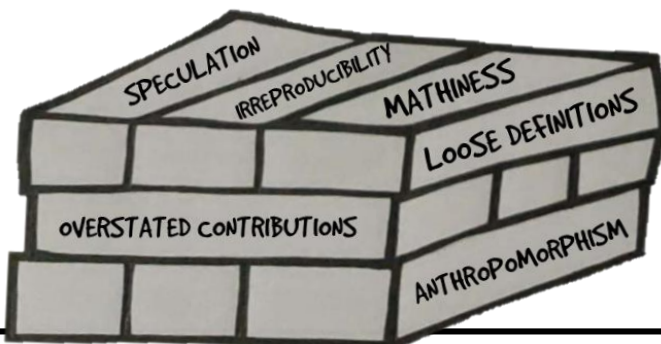
WITH PUBLIC OPINION SOARING, THE INDUSTRY FINDS A SUITABLE OPPORTUNITY TO QUOTE HIGH VALUATIONS AND LEECH INVESTOR MONEY. AND SO, PRACTITIONERS THAT WERE DOMINATING OTHER TECHNOLOGIES TURN INTO AI EXPERTS OVERNIGHT AND THEY BEGIN TO PEDDLE THE SAME OVERSTATED NARRATIVES, PROPAGATING EVEN MORE SPECULATION 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.

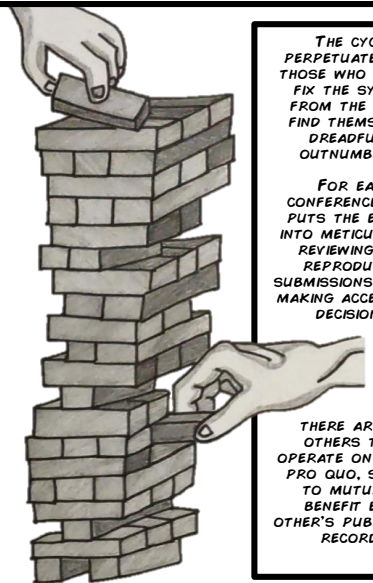
AND SO WE FIND OURSELVES IN THIS PRECARIOUS GAME

THE TRUE EXPERTS THAT ARE TRYING TO BRING ABOUT SYSTEMIC CHANGE ARE EXTREMELY INFLUENTIAL ONLY WITHIN THE CONFINES OF ACADEMIA

HOW MANY ROUNDS OF THIS GAME DO WE HAVE LEFT?

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 QUID, SET UP TO MUTUALLY BENEFIT EACH OTHER'S PUBLICATION RECORDS

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

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 **TINKERBELLEFFECT**

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

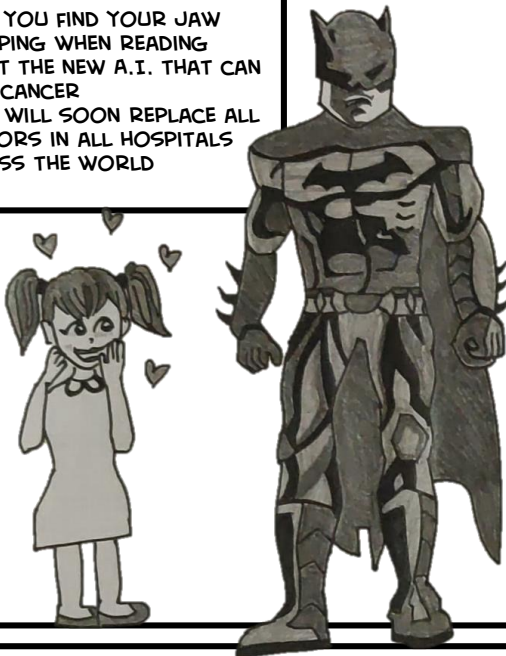
THE ONLY WAY TECH COMPANIES GET AWAY WITH SELLING YOU THEIR NEW AI-POWERED THINGAMAJIG WHICH 'EXCEEDS HUMAN INTELLIGENCE' AND IS 'CAPABLE OF GETTING YOU TO SPEAKING, READING, WRITING' IS BY GETTING YOU TO BELIEVE THAT IT ACTUALLY DOES ALL OF THAT

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...

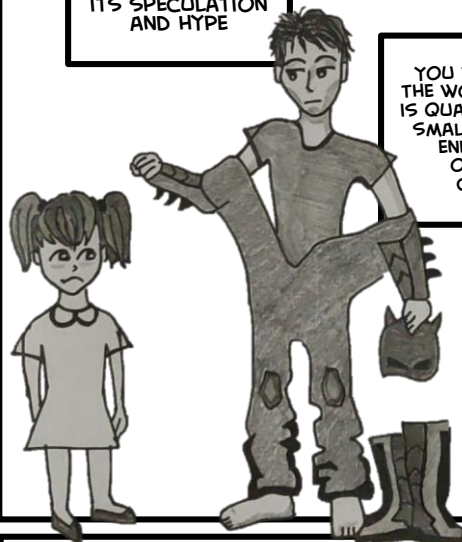
...AND YOU FIND YOUR JAW DROPPING WHEN READING ABOUT THE NEW A.I. THAT CAN CURE CANCER ...AND WILL SOON REPLACE ALL DOCTORS IN ALL HOSPITALS ACROSS THE WORLD



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

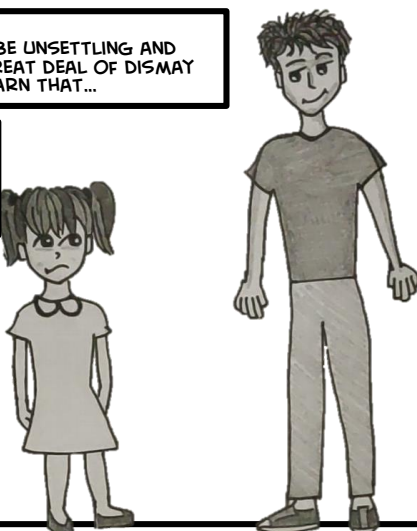
YOU WILL FIND THAT THE WORK BEING DONE IS QUALITATIVELY JUST SMALL INCREMENTAL ENHANCEMENTS ON EXISTING CAPABILITY



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 JUST SIMPLE PATTERN RECOGNITION CAPABILITIES WHICH WERE PROPOSED DECADES AGO



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

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

