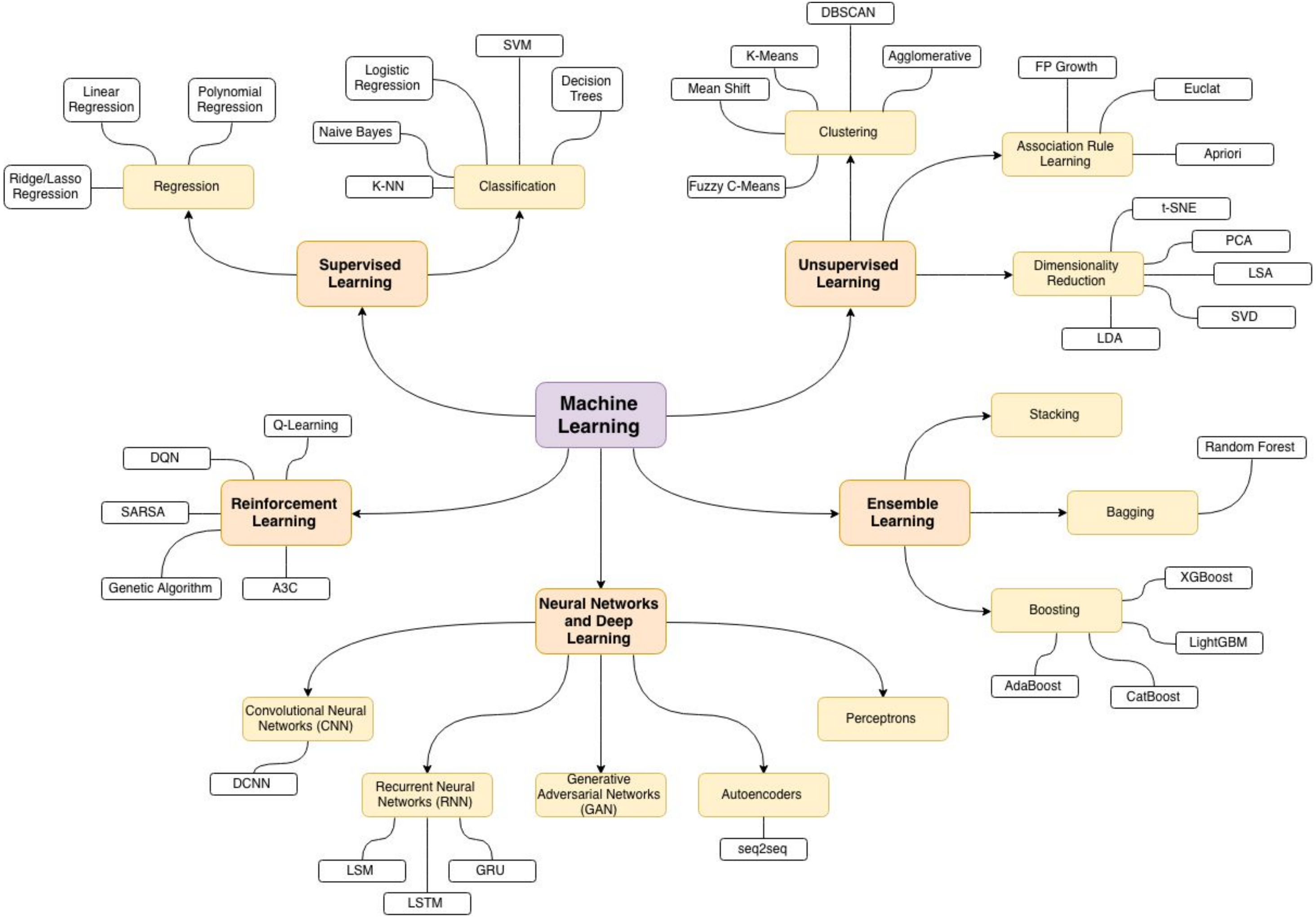
MIDS W207
Applied Machine

Week 12
Live Session Slides


Machine Learning
Supervised Learning
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
Classification
Logistic Regression
SVM
Decision Trees
Naive Bayes
K-NN
Unsupervised Learning
Clustering
DBSCAN
K-Means
Agglomerative
Mean Shift
Fuzzy C-Means
Association Rule Learning
FP Growth
Euclat
Apriori
Dimensionality Reduction
t-SNE
PCA
LSA
SVD
LDA
Reinforcement Learning
Q-Learning
DQN
SARSA
Genetic Algorithm
A3C
Ensemble Learning
Stacking
Bagging
Random Forest
Boosting
XGBoost
LightGBM
AdaBoost
CatBoost
Neural Networks and Deep Learning
Convolutional Neural Networks (CNN)
DCNN
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq
Perceptrons

Outline

Introduction to Bias

- Bias in Training Data
- Algorithmic Transparency
- Ethical Considerations

Case Studies

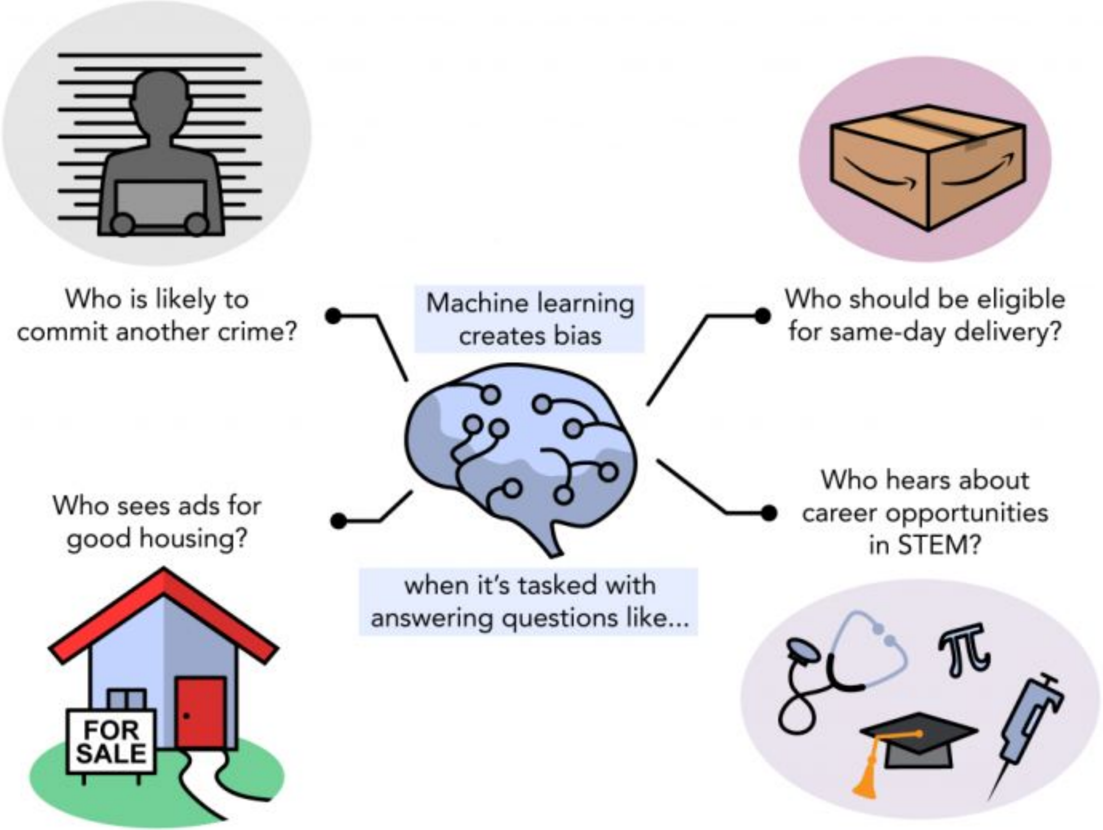
Mitigation Strategies

Fairness Metrics and Evaluation

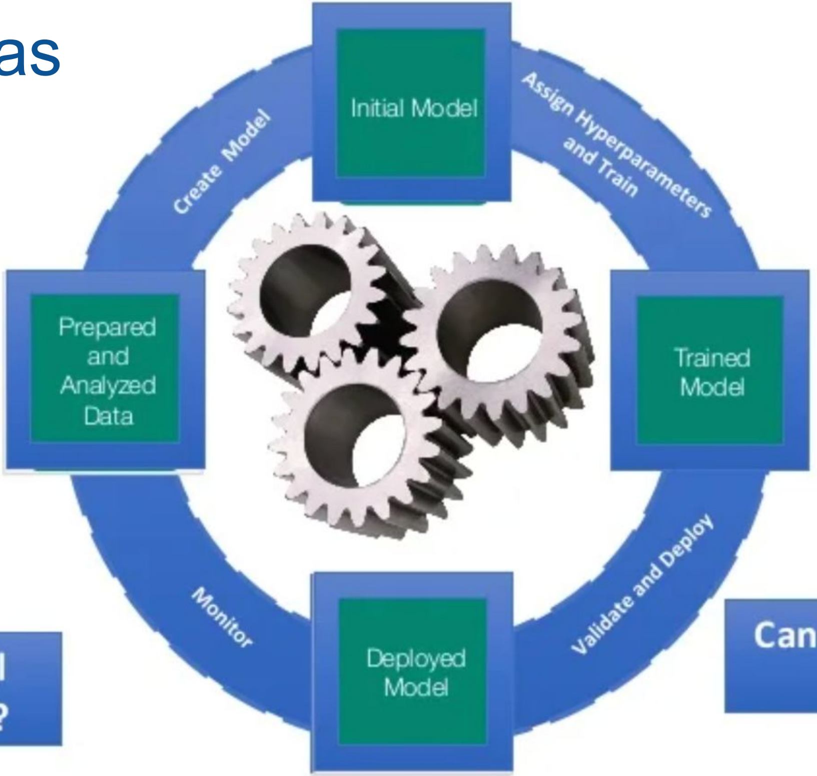
Preventing Bias

Introduction to Bias

- Anomaly in the output of machine learning algorithms
- Prejudiced assumptions made during the algorithm development process
- Prejudices in the training data


Who is likely to commit another crime?
Machine learning creates bias
Who should be eligible for same-day delivery?
Who sees ads for good housing?
Who hears about career opportunities in STEM?
when it's tasked with answering questions like...
FOR SALE

Introduction to Bias

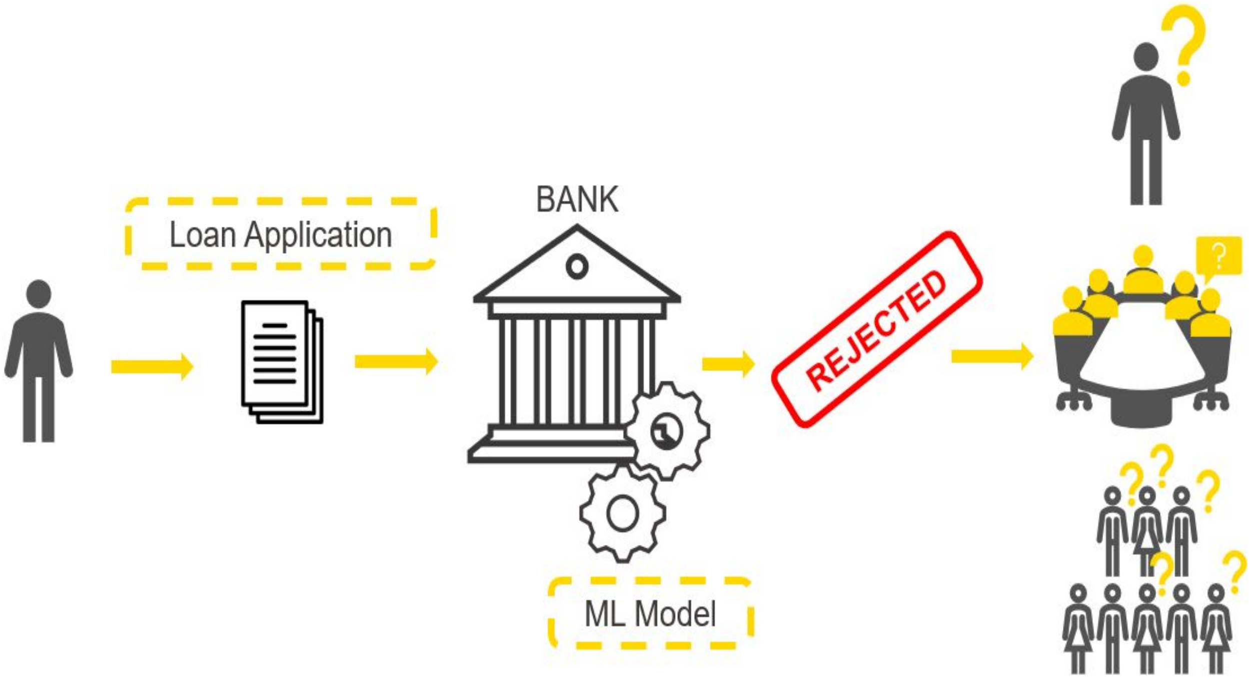


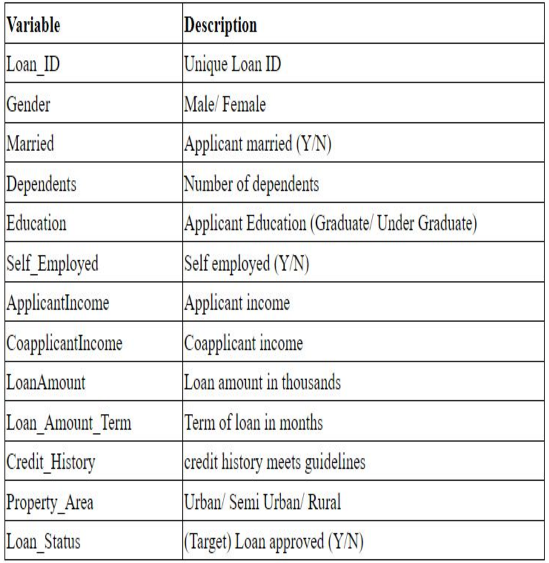
Can the dataset be trusted?

Can the trained model be trusted?

Is the deployed model giving fair predictions?

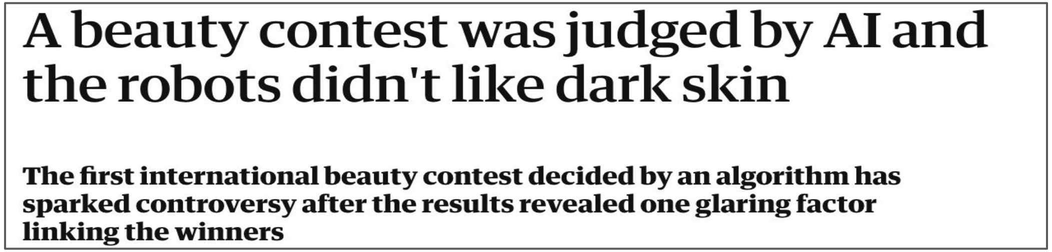
Case Studies





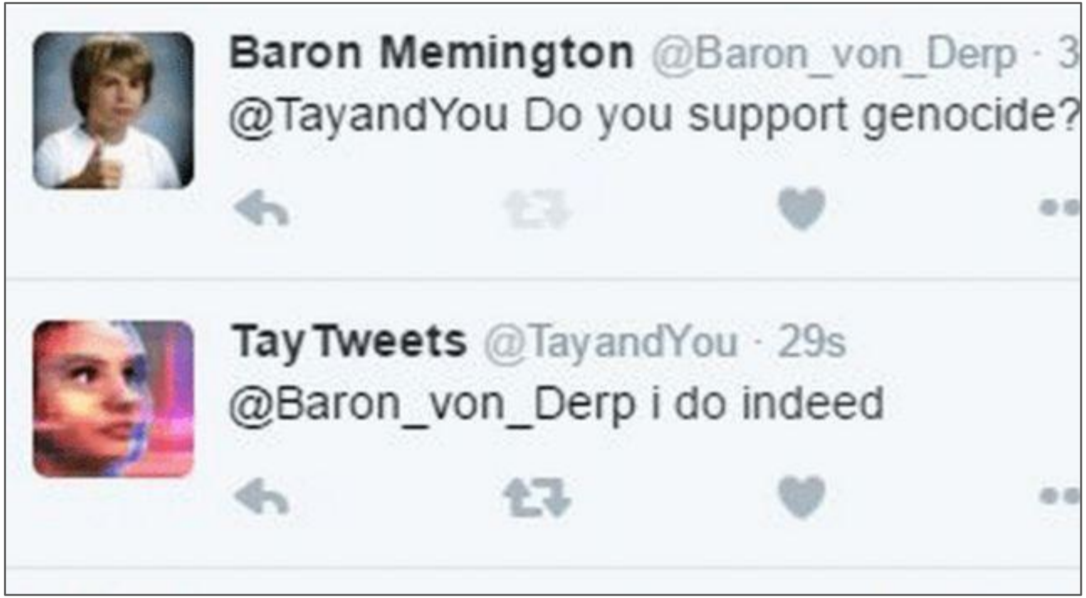
Variable	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income
CoapplicantIncome	Coapplicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	(Target) Loan approved (Y/N)

Case Studies



A beauty contest was judged by AI and the robots didn't like dark skin

The first international beauty contest decided by an algorithm has sparked controversy after the results revealed one glaring factor linking the winners



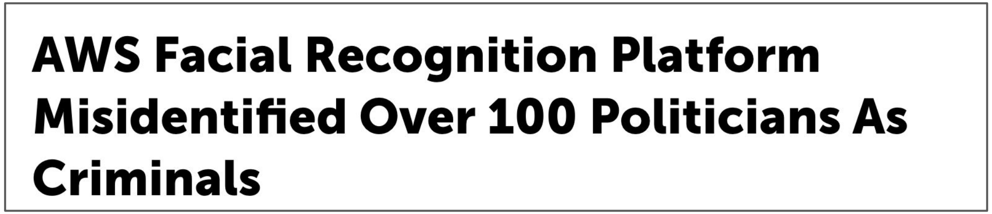
Baron Memington @Baron_von_Derp · 3

@TayandYou Do you support genocide?

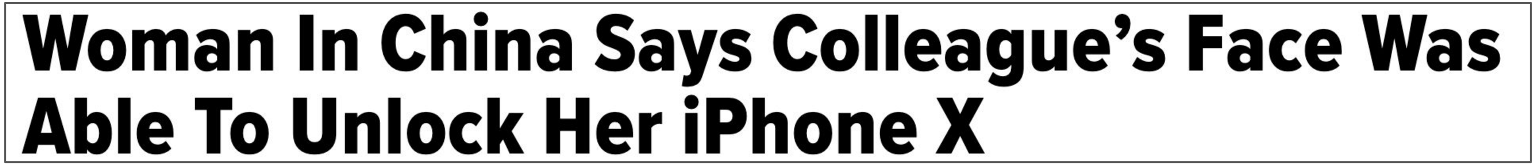
Tay Tweets @TayandYou · 29s

@Baron_von_Derp i do indeed

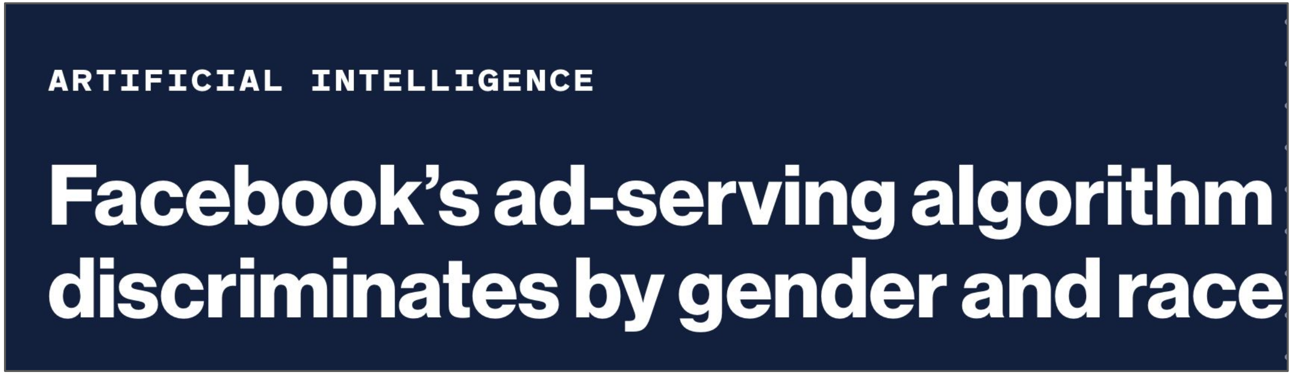
Case Studies



AWS Facial Recognition Platform Misidentified Over 100 Politicians As Criminals



Woman In China Says Colleague's Face Was Able To Unlock Her iPhone X



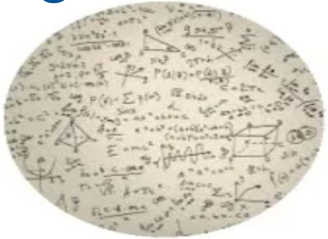
ARTIFICIAL INTELLIGENCE

Facebook's ad-serving algorithm discriminates by gender and race

Mitigation Strategies

Is it fair?

FAIRNESS



Is it easy to understand?

EXPLAINABILITY

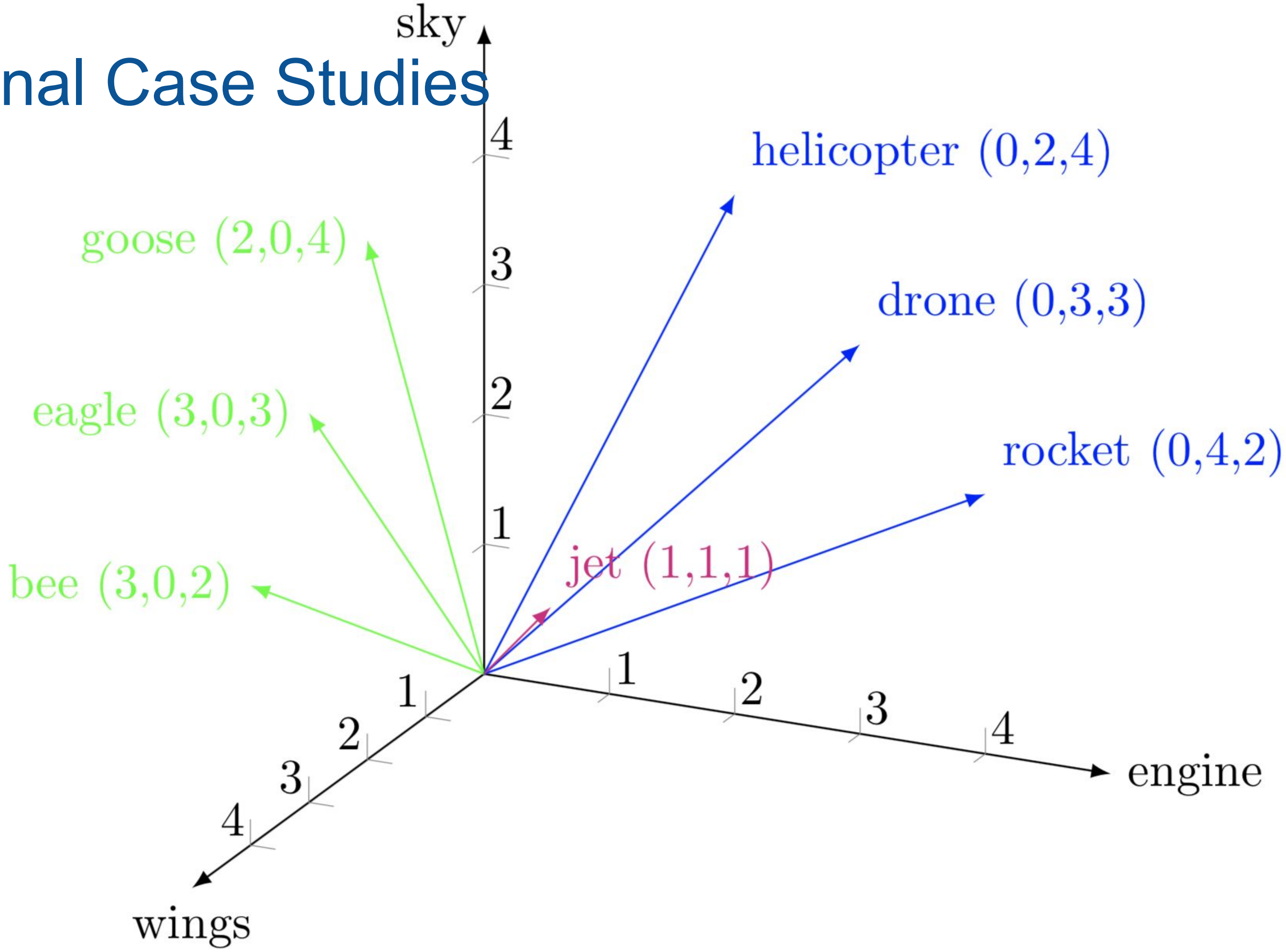
Did anyone tamper with it?

ROBUSTNESS

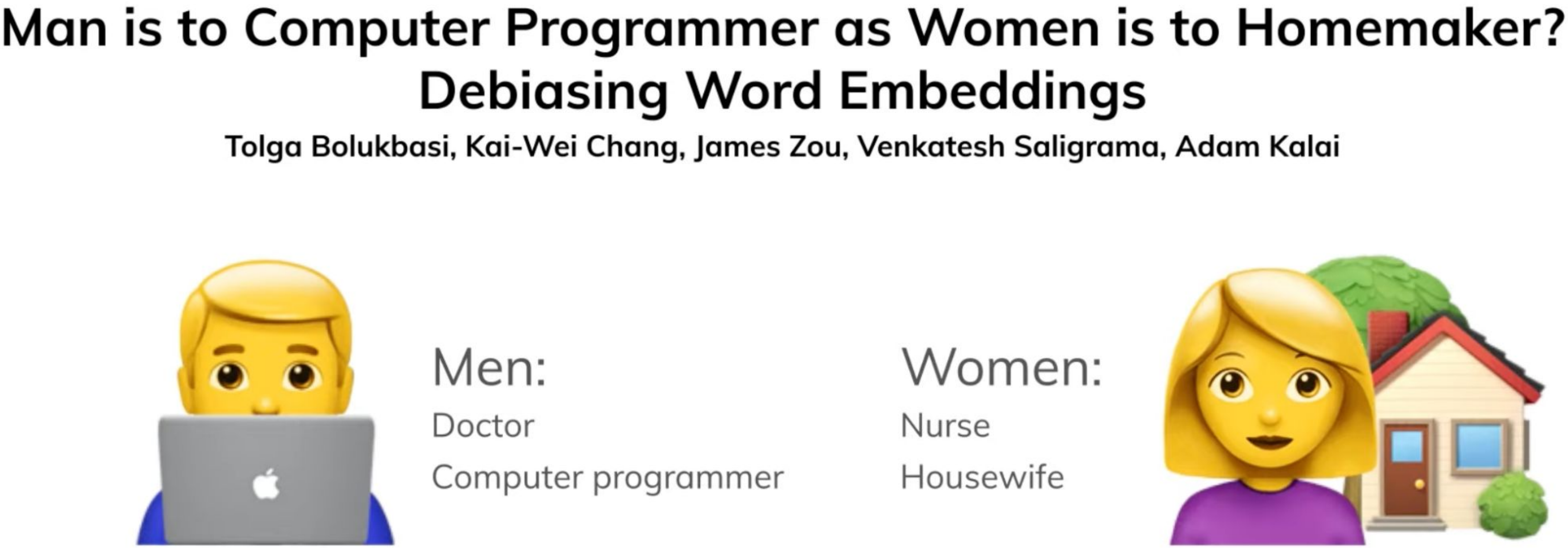
Is it accountable?

ASSURANCE

Additional Case Studies



Additional Case Studies



Man is to Computer Programmer as Women is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai

Men:

Doctor

Computer programmer

Women:

Nurse

Housewife

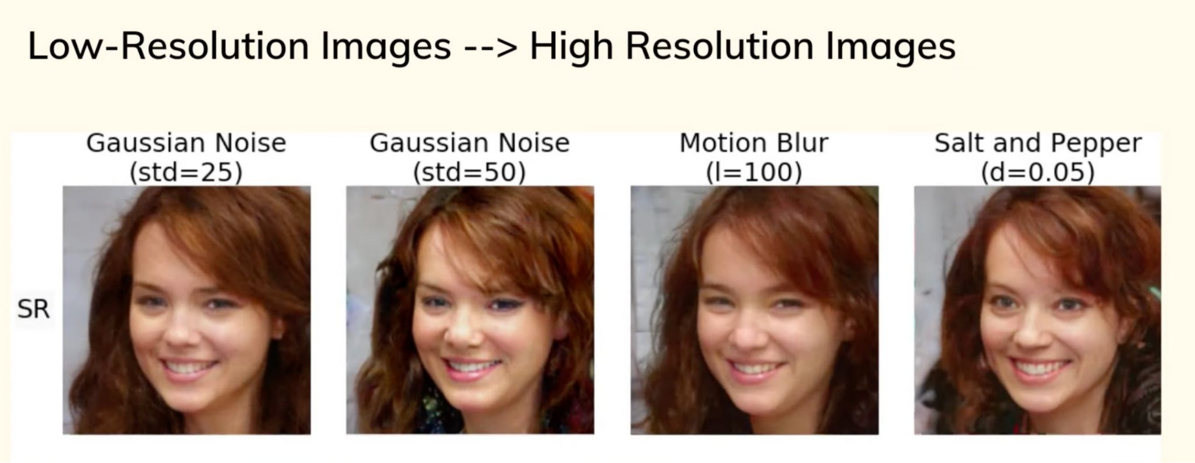
Additional Case Studies

PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models

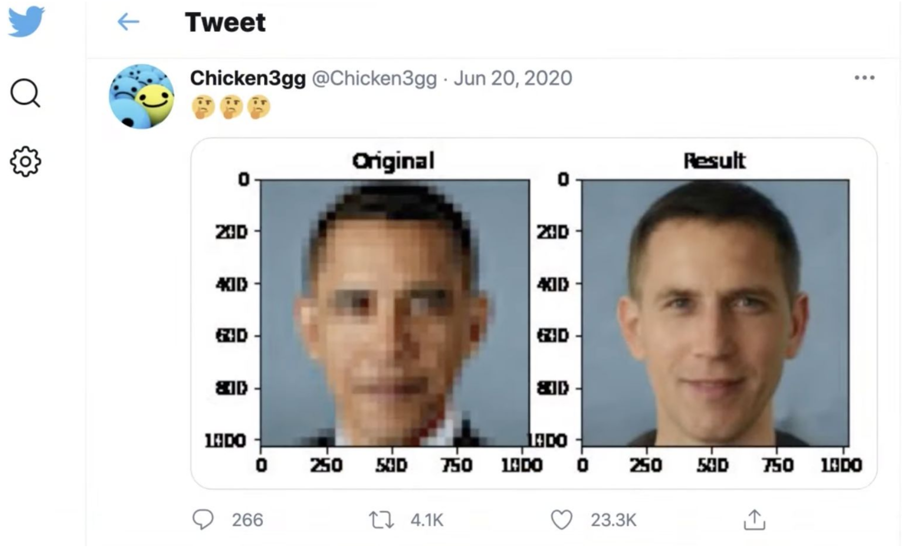
Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, Cynthia Rudin

Abstract

The primary aim of single-image super-resolution is to construct a high-resolution (HR) image from a corresponding low-resolution (LR) input ... We present a novel super-resolution algorithm addressing this problem, PULSE (Photo Upsampling via Latent Space Explo ration), which generates high-resolution, realistic images at resolutions previously unseen in the literature ... Instead of starting with the LR image and slowly adding detail, PULSE traverses the high-resolution natural image manifold, searching for images that downscale to the original LR image.


Low-Resolution Images --> High Resolution Images

Additional Case Studies

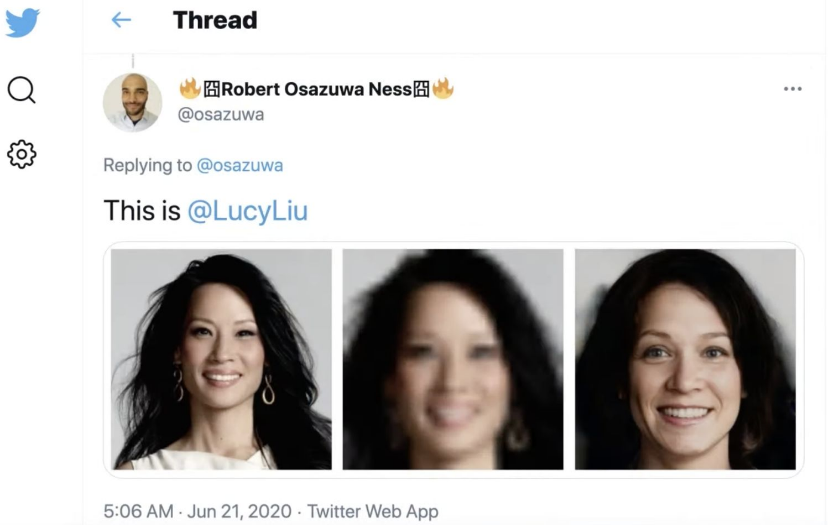


Tweet

Chicken3gg @Chicken3gg · Jun 20, 2020

🤔🤔🤔

266 4.1K 23.3K



Thread

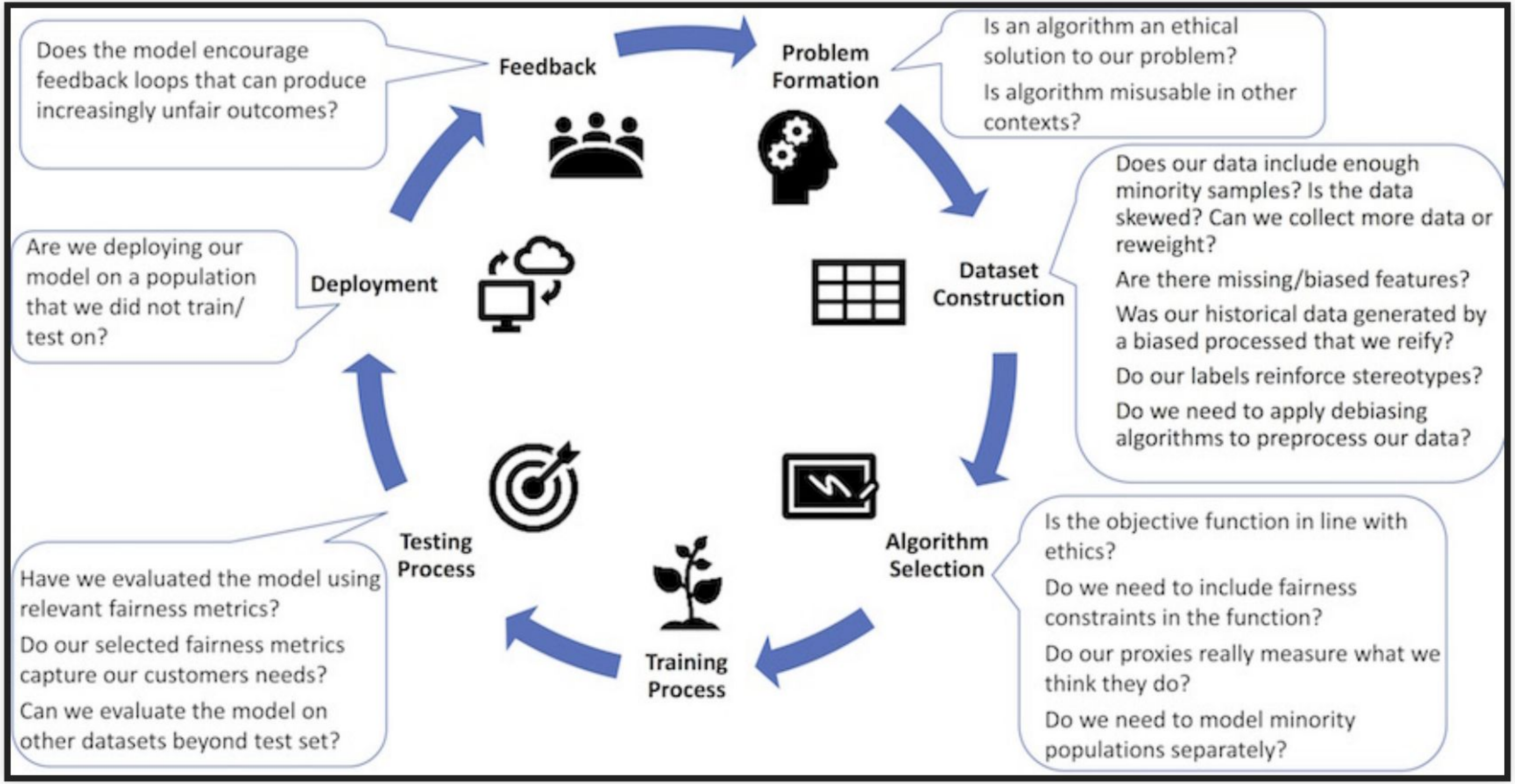
🔥囧Robert Osazuwa Ness囧🔥
@osazuwa

Replying to @osazuwa

This is @LucyLiu

5:06 AM · Jun 21, 2020 · Twitter Web App

Fairness Metrics and Evaluation



Does the model encourage feedback loops that can produce increasingly unfair outcomes?

Is an algorithm an ethical solution to our problem?

Is algorithm misusable in other contexts?

Does our data include enough minority samples? Is the data skewed? Can we collect more data or reweight?

Are there missing/biased features?

Was our historical data generated by a biased processed that we reify?

Do our labels reinforce stereotypes?

Do we need to apply debiasing algorithms to preprocess our data?

Is the objective function in line with ethics?

Do we need to include fairness constraints in the function?

Do our proxies really measure what we think they do?

Do we need to model minority populations separately?

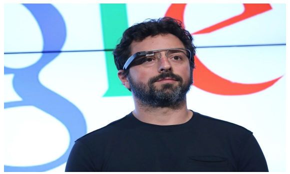
Have we evaluated the model using relevant fairness metrics?

Do our selected fairness metrics capture our customers needs?

Can we evaluate the model on other datasets beyond test set?

Are we deploying our model on a population that we did not train/ test on?

Introduction



“I fear that AI may replace humans” - Stephen Hawking

“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.” - Bill gates

“from the fears of sci-fi style sentience to the more near-term questions such as validating the performance of self-driving cars.” - Sergey Brin

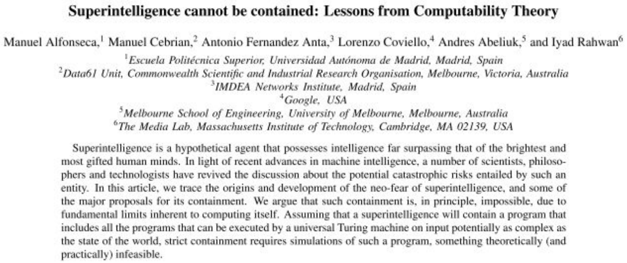
“The Singularity for This Level of the Simulation Is Coming Soon”- Elon Musk

A Little
Late Breaking News

IEEE Spectrum, 18 January, 2021

"It may be theoretically impossible for humans to control a superintelligent AI", a new study finds. Worse still, the research also quashes any hope for detecting such an unstoppable AI when it's on the verge of being created.

Slightly less grim is the timetable. By at least one estimate, many decades lie ahead before any such existential computational reckoning could be in the cards for humanity."

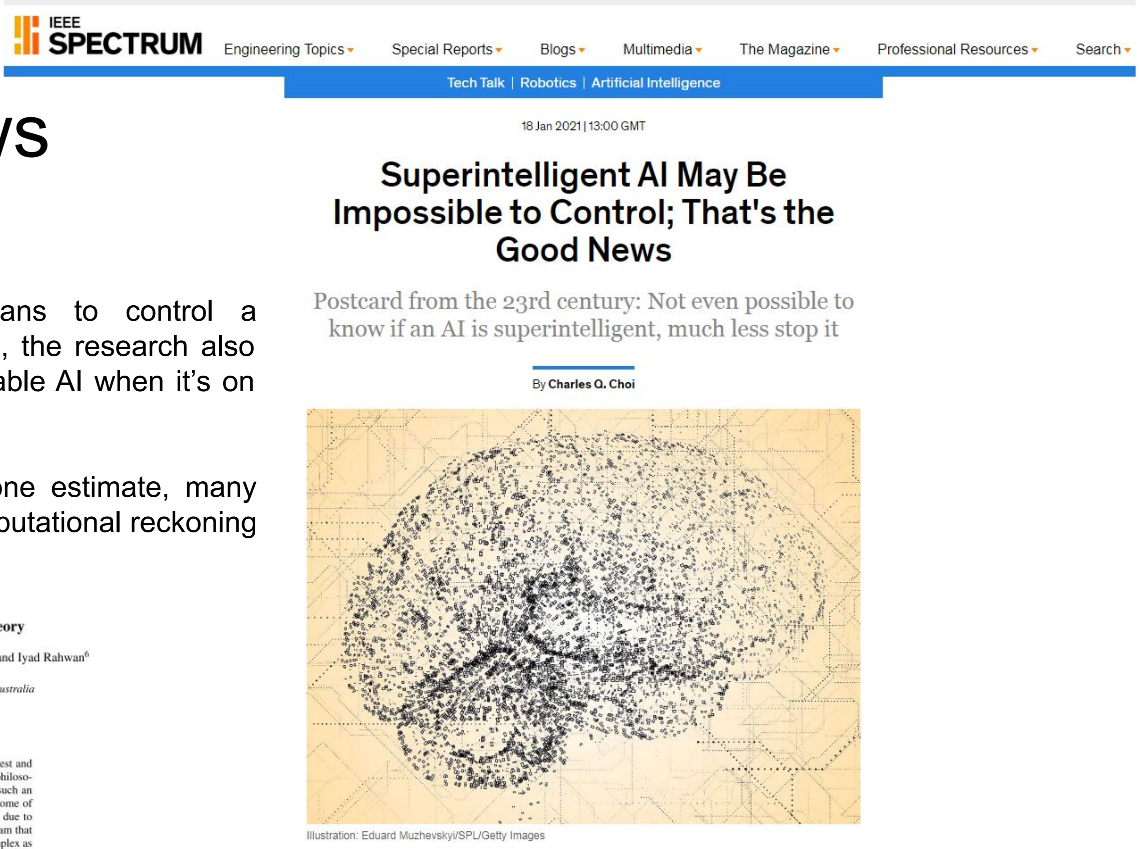


Superintelligence cannot be contained: Lessons from Computability Theory

Manuel Alfonseca,[1] Manuel Cebrian,[2] Antonio Fernandez Anta,[3] Lorenzo Coviello,[4] Andres Abeliuk,[5] and Iyad Rahwan[6]

[1] *Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain*
[2] *Data61 Unit, Commonwealth Scientific and Industrial Research Organisation, Melbourne, Victoria, Australia*
[3] *IMDEA Networks Institute, Madrid, Spain*
[4] *Google, USA*
[5] *Melbourne School of Engineering, University of Melbourne, Melbourne, Australia*
[6] *The Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Superintelligence is a hypothetical agent that possesses intelligence far surpassing that of the brightest and most gifted human minds. In light of recent advances in machine intelligence, a number of scientists, philosophers and technologists have revived the discussion about the potential catastrophic risks entailed by such an entity. In this article, we trace the origins and development of the neo-fear of superintelligence, and some of the major proposals for its containment. We argue that such containment is, in principle, impossible, due to fundamental limits inherent to computing itself. Assuming that a superintelligence will contain a program that includes all the programs that can be executed by a universal Turing machine on input potentially as complex as the state of the world, strict containment requires simulations of such a program, something theoretically (and practically) infeasible.



18 Jan 2021 | 13:00 GMT

Superintelligent AI May Be Impossible to Control; That's the Good News

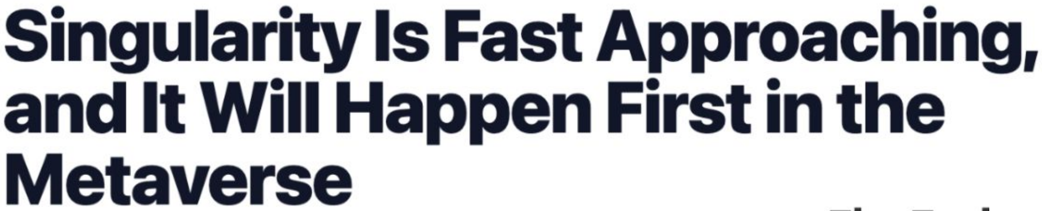
Postcard from the 23rd century: Not even possible to know if an AI is superintelligent, much less stop it

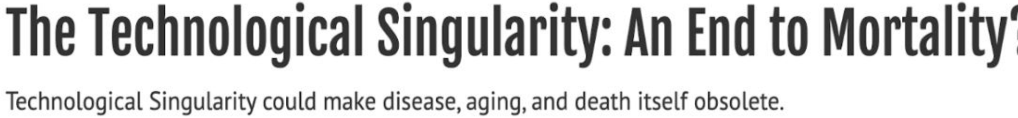
By **Charles Q. Choi**

Illustration: Eduard Muzhevskyi/SPL/Getty Images

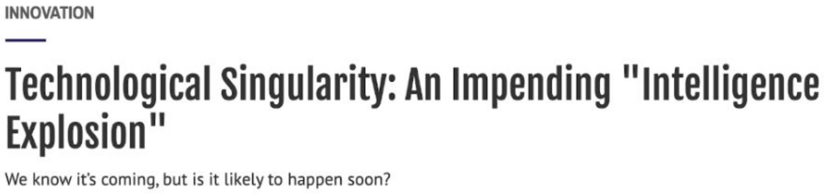
IEEE

More Recently on Technological Singularity


Singularity Is Fast Approaching, and It Will Happen First in the Metaverse

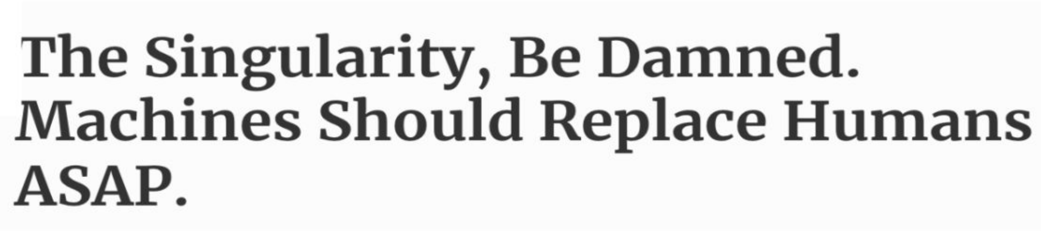

The Technological Singularity: An End to Mortality?

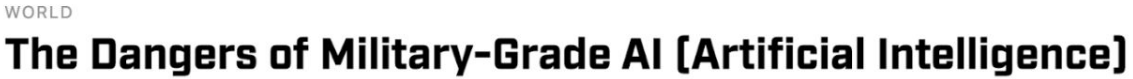
Technological Singularity could make disease, aging, and death itself obsolete.


INNOVATION

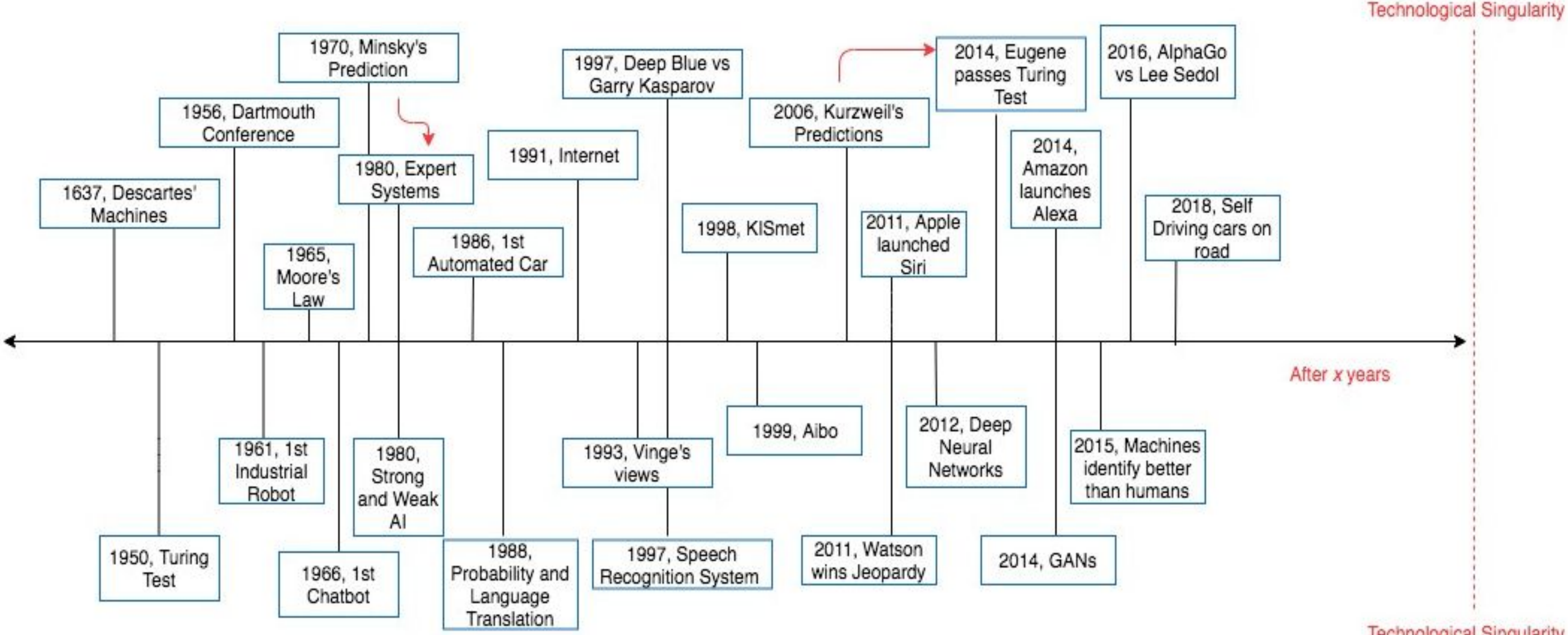
Technological Singularity: An Impending "Intelligence Explosion"

We know it's coming, but is it likely to happen soon?

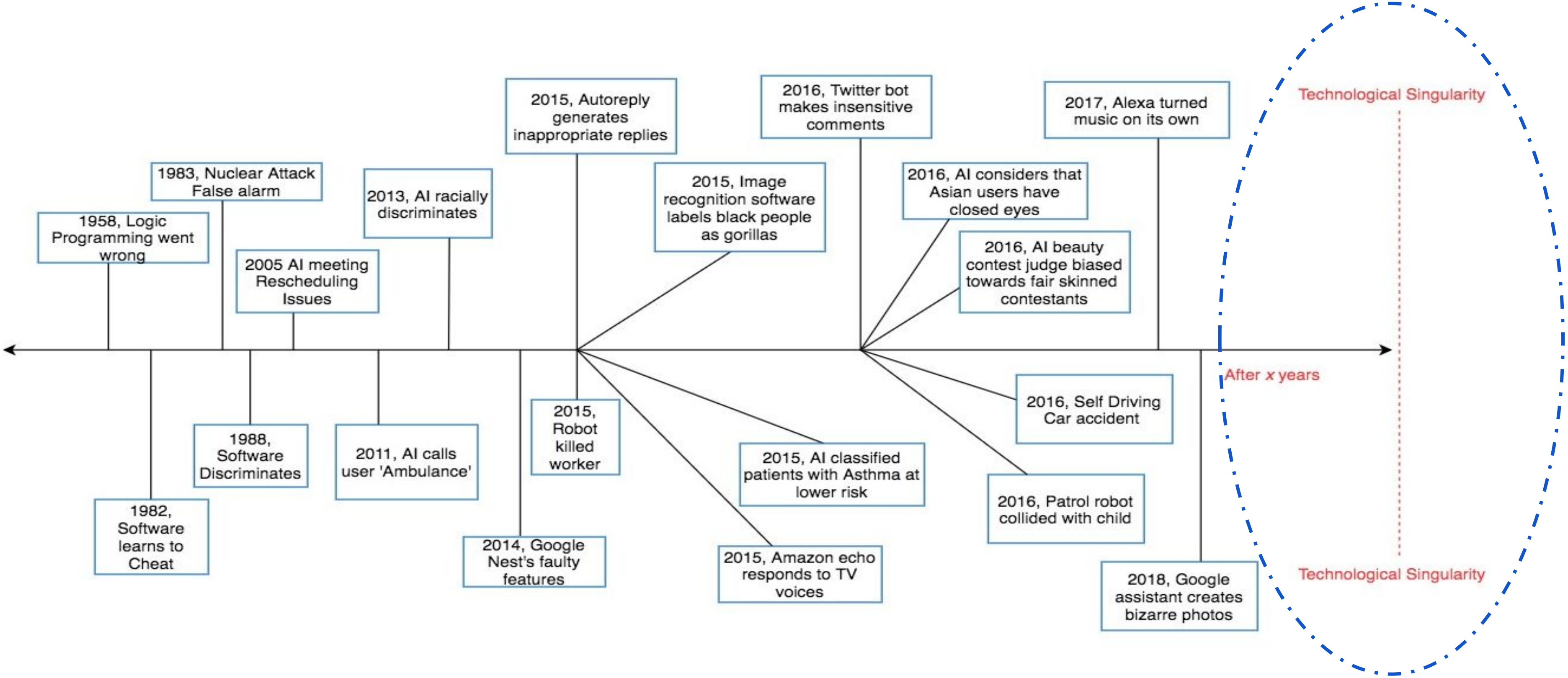

The Singularity, Be Damned. Machines Should Replace Humans ASAP.


WORLD

The Dangers of Military-Grade AI (Artificial Intelligence)


1637, Descartes' Machines
1950, Turing Test
1956, Dartmouth Conference
1961, 1st Industrial Robot
1965, Moore's Law
1966, 1st Chatbot
1970, Minsky's Prediction
1980, Expert Systems
1980, Strong and Weak AI
1986, 1st Automated Car
1988, Probability and Language Translation
1991, Internet
1993, Vinge's views
1997, Deep Blue vs Garry Kasparov
1997, Speech Recognition System
1998, KISmet
1999, Aibo
2006, Kurzweil's Predictions
2011, Apple launched Siri
2011, Watson wins Jeopardy
2012, Deep Neural Networks
2014, Eugene passes Turing Test
2014, Amazon launches Alexa
2014, GANs
2015, Machines identify better than humans
2016, AlphaGo vs Lee Sedol
2018, Self Driving cars on road
After x years
Technological Singularity
Technological Singularity

Motivation


1958, Logic Programming went wrong
1982, Software learns to Cheat
1983, Nuclear Attack False alarm
1988, Software Discriminates
2005 AI meeting Rescheduling Issues
2011, AI calls user 'Ambulance'
2013, AI racially discriminates
2014, Google Nest's faulty features
2015, Autoreply generates inappropriate replies
2015, Robot killed worker
2015, Image recognition software labels black people as gorillas
2015, AI classified patients with Asthma at lower risk
2015, Amazon echo responds to TV voices
2016, Twitter bot makes insensitive comments
2016, AI considers that Asian users have closed eyes
2016, AI beauty contest judge biased towards fair skinned contestants
2016, Self Driving Car accident
2016, Patrol robot collided with child
2017, Alexa turned music on its own
2018, Google assistant creates bizarre photos
After x years
Technological Singularity
Technological Singularity

Kurzweil’s Singularity Hypothesis

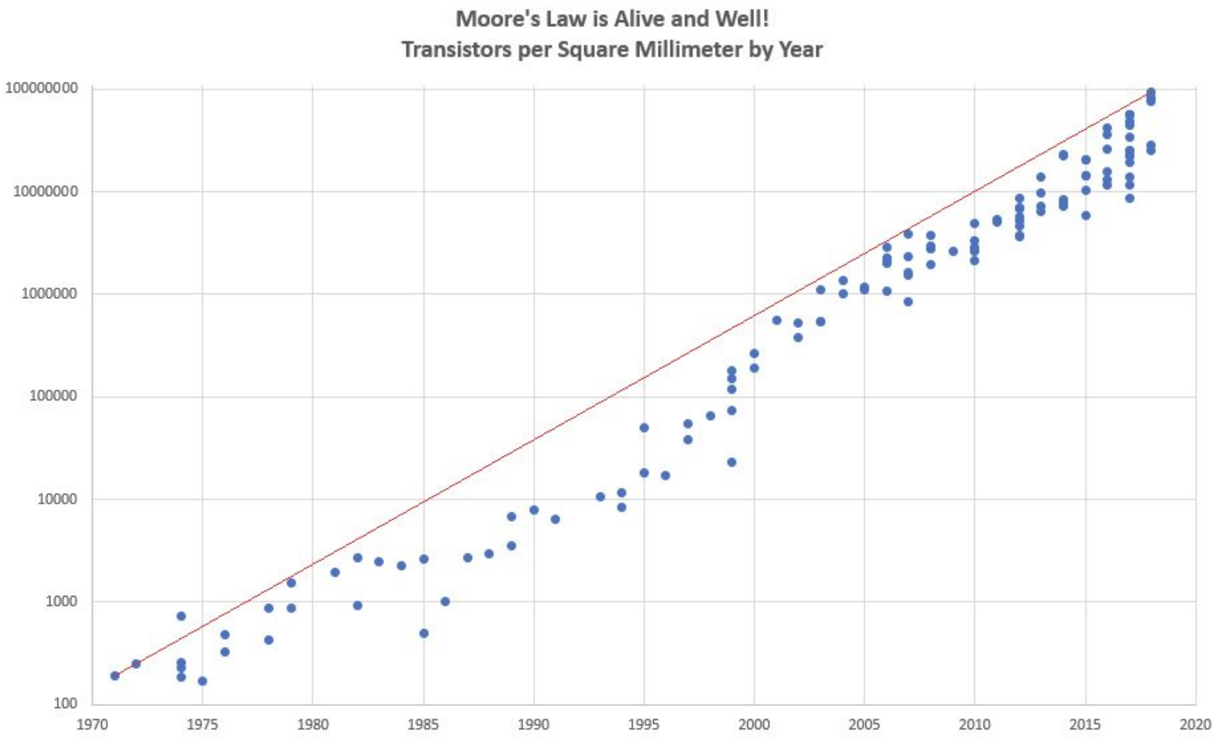
Based on **Moore’s Law**

In, *The Singularity Is Near (2005)*,

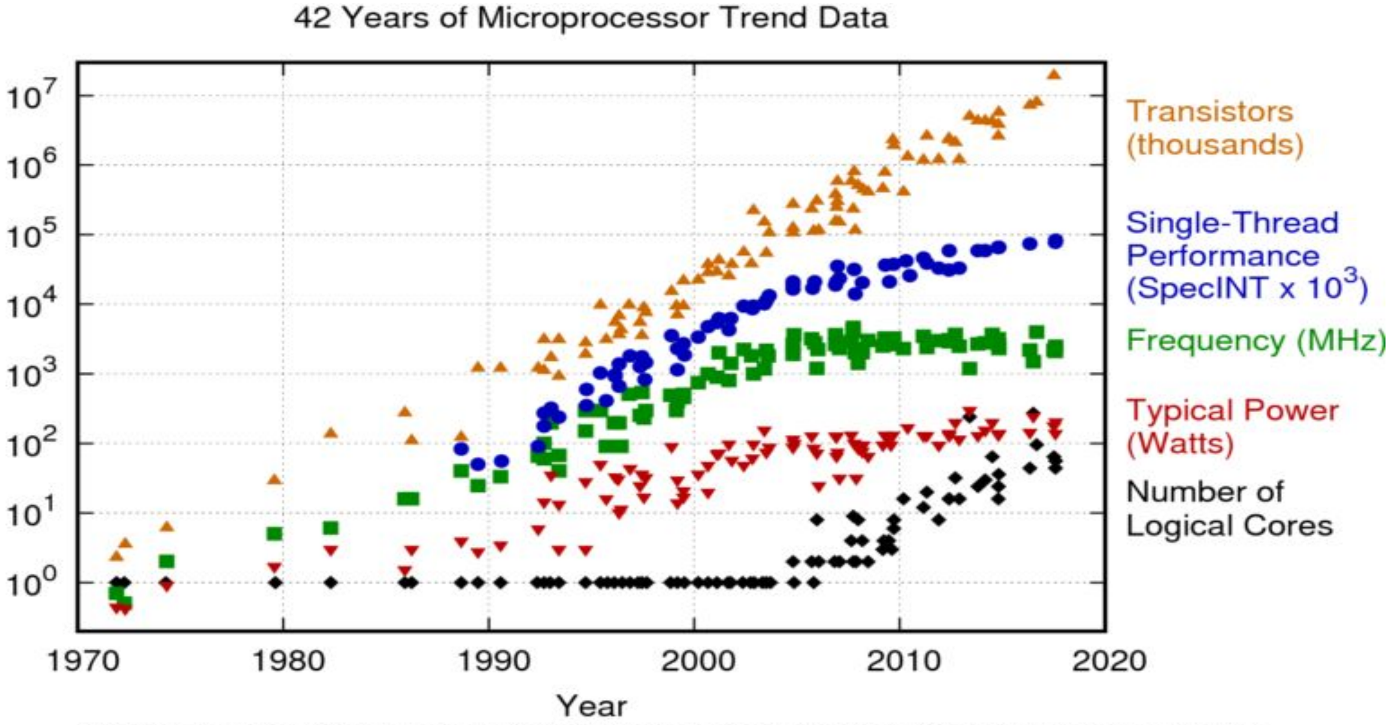
> "Moore's law states that the number of transistors on an integrated circuit doubles approximately every two years."
>
> Also known as the **law of exponential growth** or the **accelerating change law**.

Based on this, Kurzweil predicted the oncoming of Technological Singularity in 2045 based upon the ability to create an Integrated Circuit by then with the computational capability of the human brain.

Kurzweil’s Singularity Hypothesis

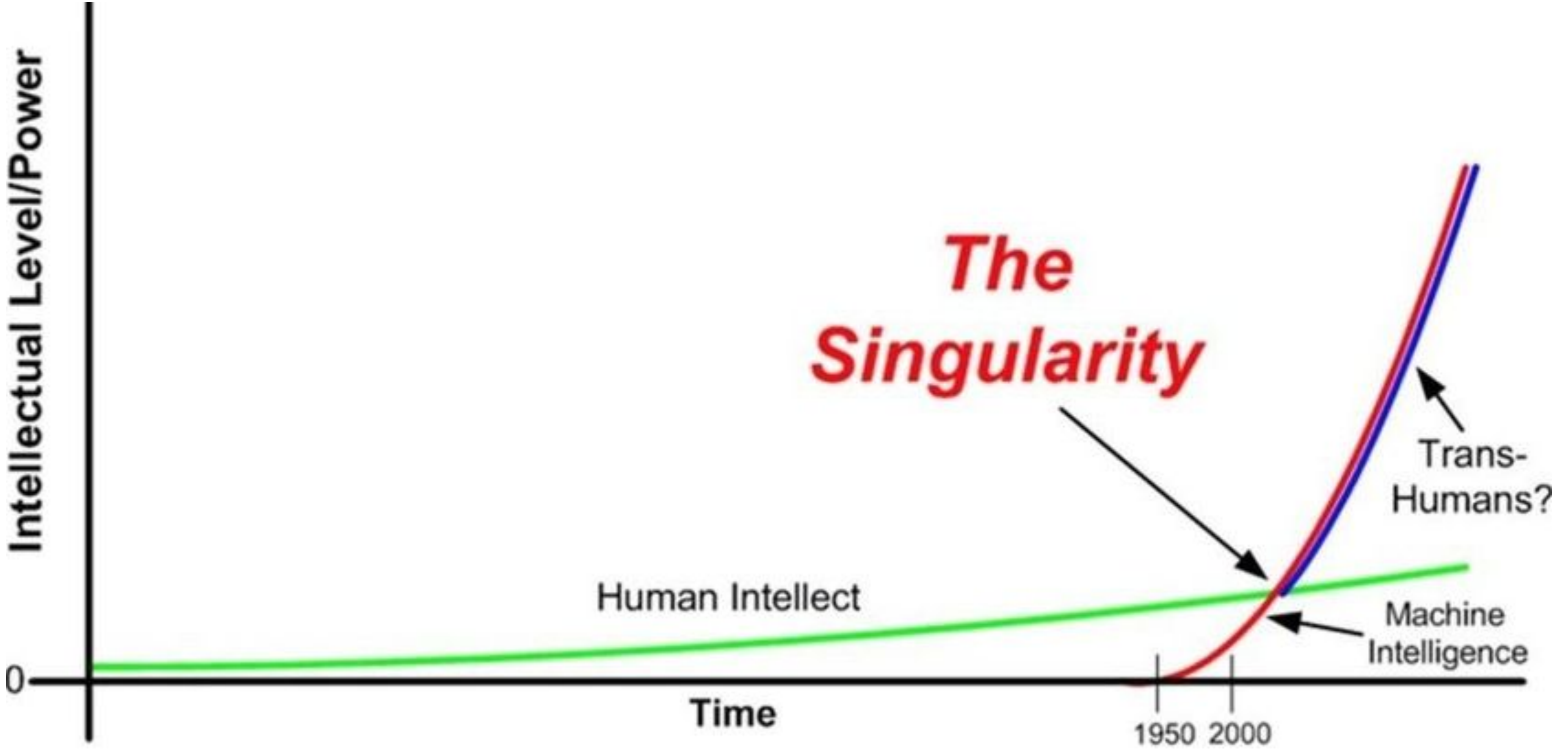


Kurzweil’s Singularity Hypothesis


42 Years of Microprocessor Trend Data

Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Kurzweil’s Singularity Hypothesis



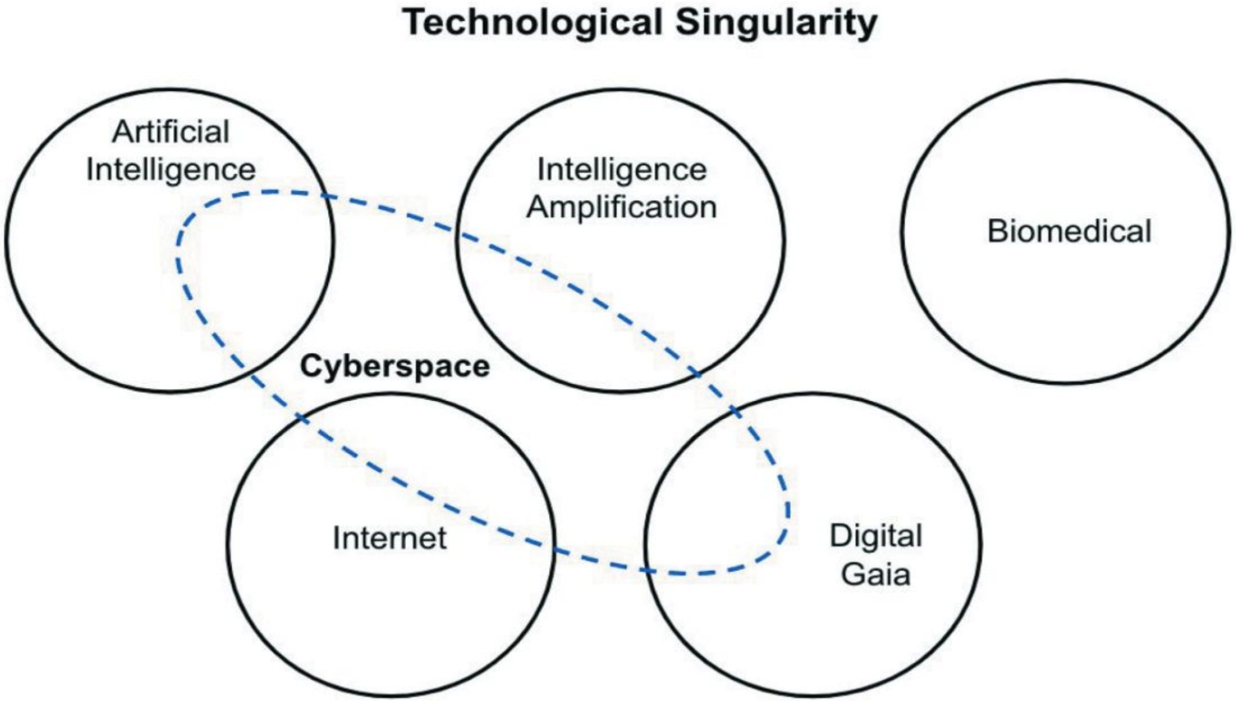
Vinge’s Singularity Hypothesis

In *The Coming of Technological Singularity, 1993,* and *The Signs of Singularity, IEEE Spectrum*, Vinge predicted the possible paths for singularity:

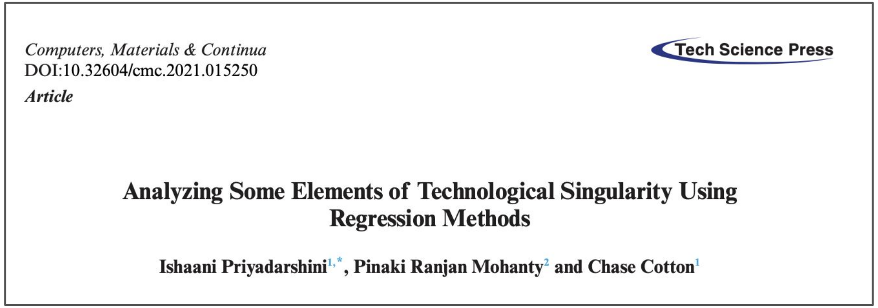
- The AI Scenario: We create superhuman artificial intelligence (AI) in computers.
- The IA Scenario: We enhance human intelligence through human-to-computer interfaces—that is, we achieve intelligence amplification (IA).
- The Biomedical Scenario: We directly increase our intelligence by improving the neurological operation of our brains.
- The Internet Scenario: Humanity, its networks, computers, and databases become sufficiently effective to be considered a superhuman being.
- The Digital Gaia Scenario: The network of embedded microprocessors becomes sufficiently effective to be considered a superhuman being.

Predicted the oncoming of Technological Singularity in **2005-2030**

Proposed Research Work



Proposed Research Work



Computers, Materials & Continua
DOI:10.32604/cmc.2021.015250
Article

Tech Science Press

Analyzing Some Elements of Technological Singularity Using Regression Methods

Ishaani Priyadarshini[1,*], Pinaki Ranjan Mohanty[2] and Chase Cotton[1]

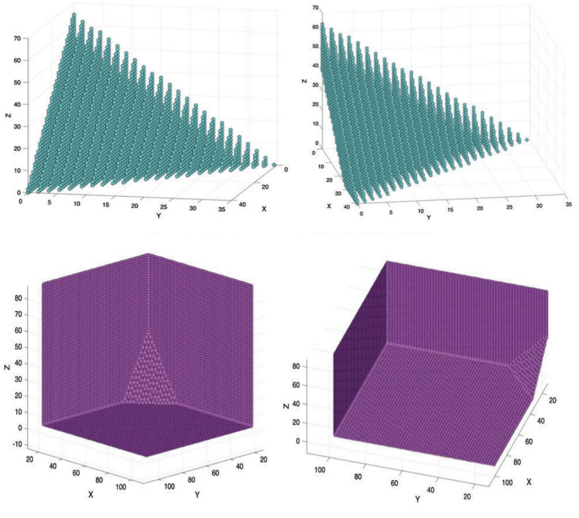
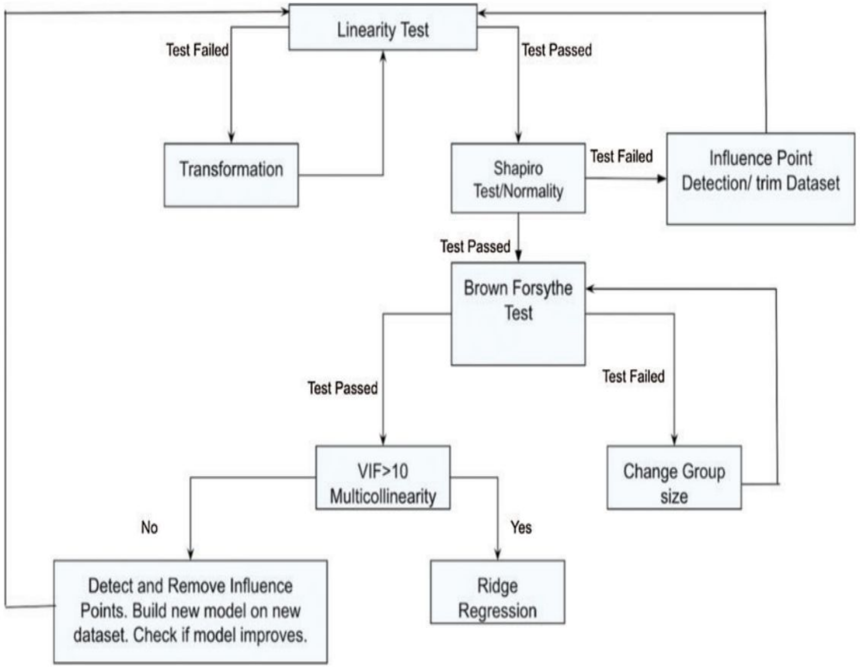


Figure 9: Solution Space for Eq. (1) (2045)



Linearity Test

Test Failed

Test Passed

Transformation

Shapiro Test/Normality

Test Failed

Influence Point Detection/ trim Dataset

Test Passed

Brown Forsythe Test

Test Passed

Test Failed

VIF>10 Multicollinearity

Change Group size

No

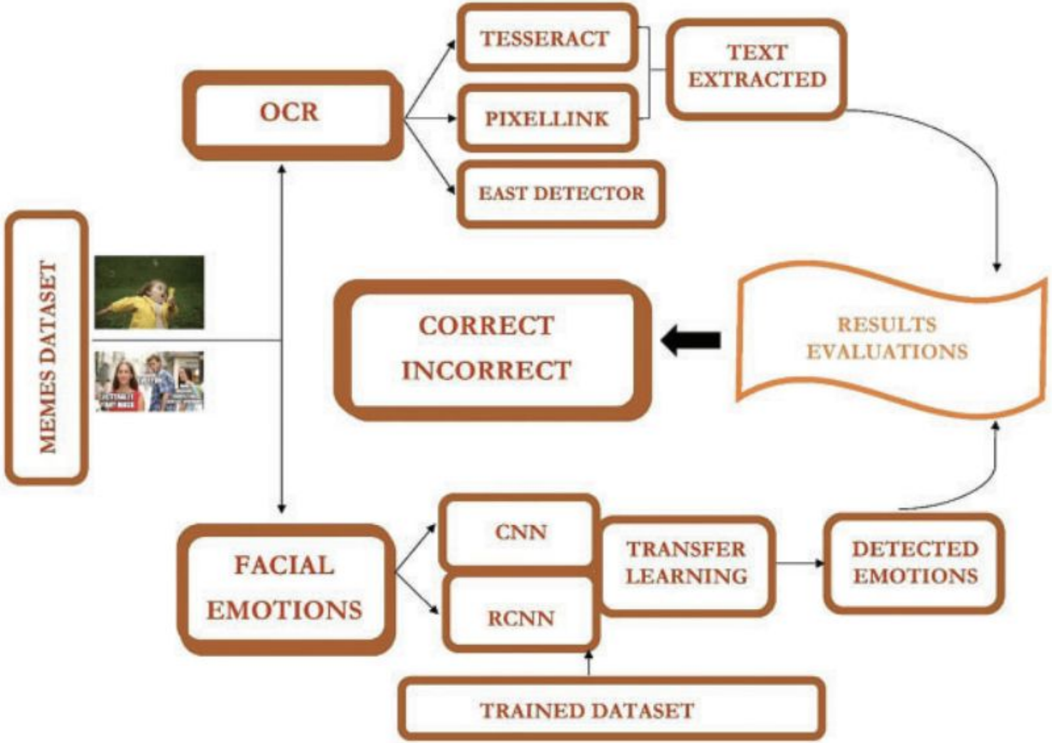
Yes

Detect and Remove Influence Points. Build new model on new dataset. Check if model improves.

Ridge Regression

Figure 1: Steps for the analysis conducted

Experiments with OCR and Facial Emotions


OCR
TESSERACT
PIXELLINK
EAST DETECTOR
TEXT EXTRACTED
MEMES DATASET
CORRECT
INCORRECT
RESULTS EVALUATIONS
FACIAL EMOTIONS
CNN
RCNN
TRANSFER LEARNING
DETECTED EMOTIONS
TRAINED DATASET