San Jose State University

From the SelectedWorks of Vishnu Pendyala

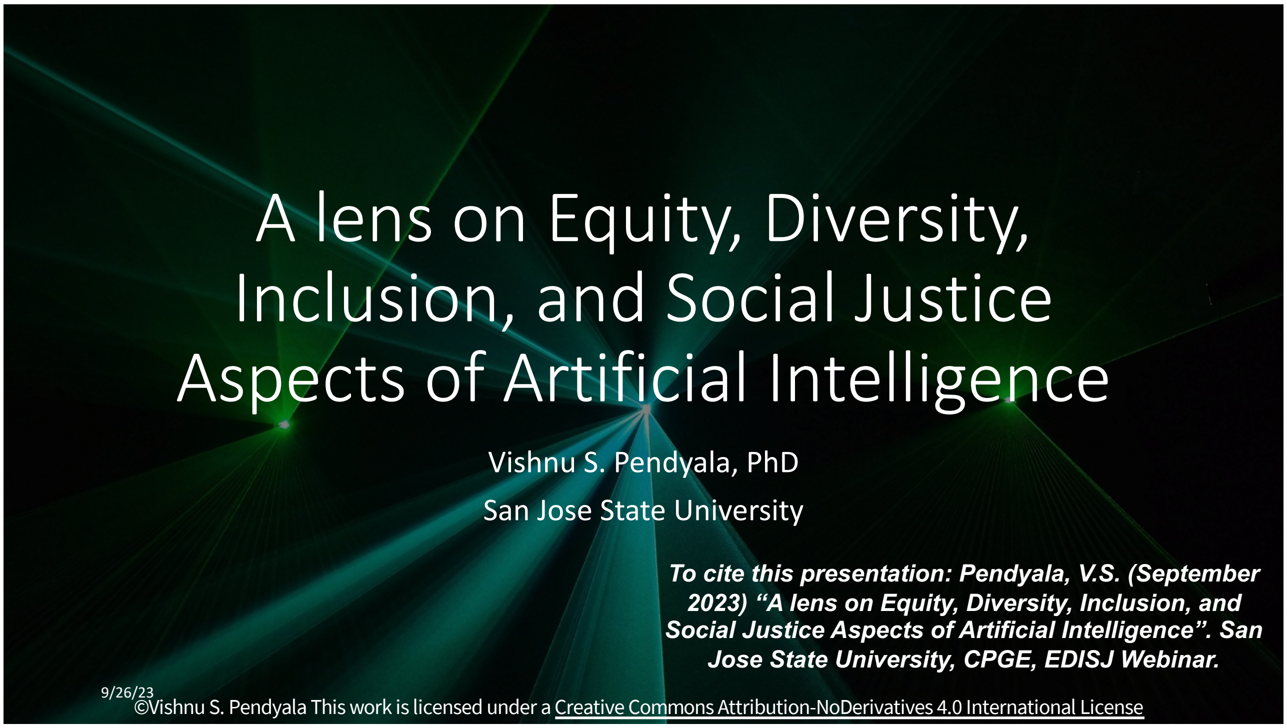
September 26, 2023

A lens on Equity, Diversity, Inclusion, and Social Justice Aspects of Artificial Intelligence

Vishnu Pendyala, *San Jose State University*

This work is licensed under a Creative Commons CC_BY-ND International License.

Available at: https://works.bepress.com/vishnu-pendyala/47/



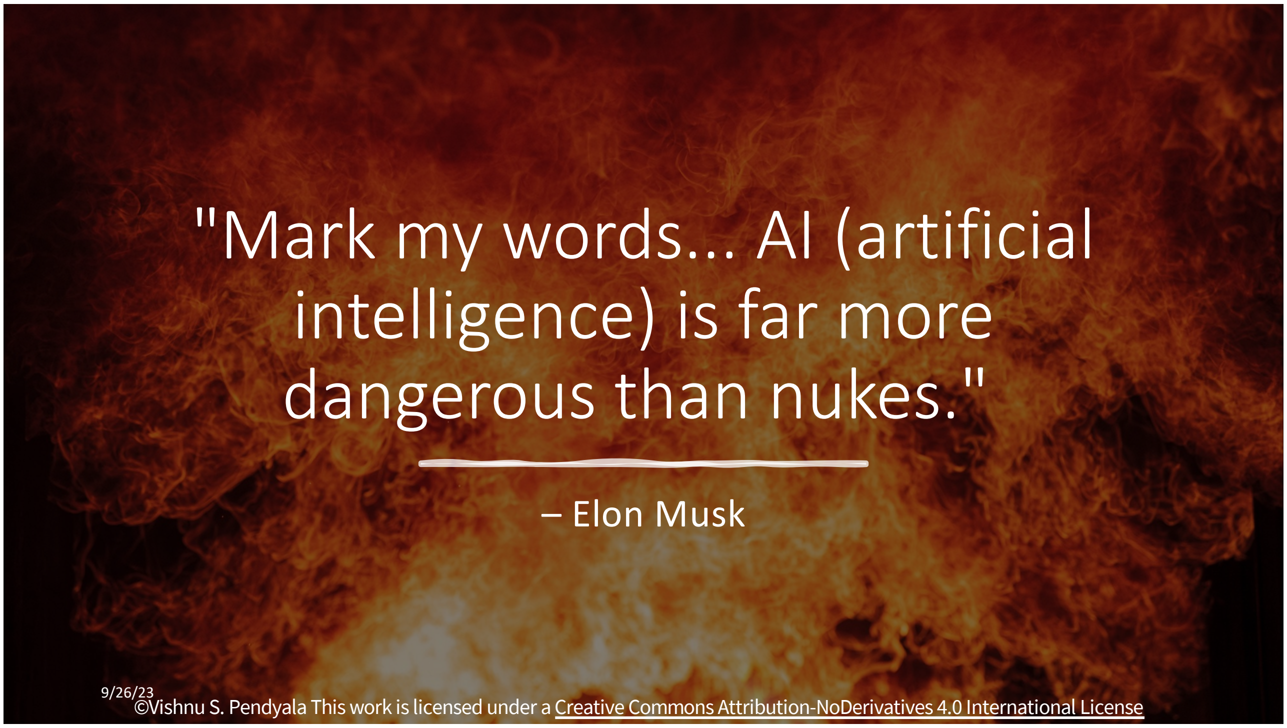
A lens on Equity, Diversity, Inclusion, and Social Justice Aspects of Artificial Intelligence

Vishnu S. Pendyala, PhD

San Jose State University

To cite this presentation: Pendyala, V.S. (September 2023) "A lens on Equity, Diversity, Inclusion, and Social Justice Aspects of Artificial Intelligence". San Jose State University, CPGE, EDISJ Webinar.

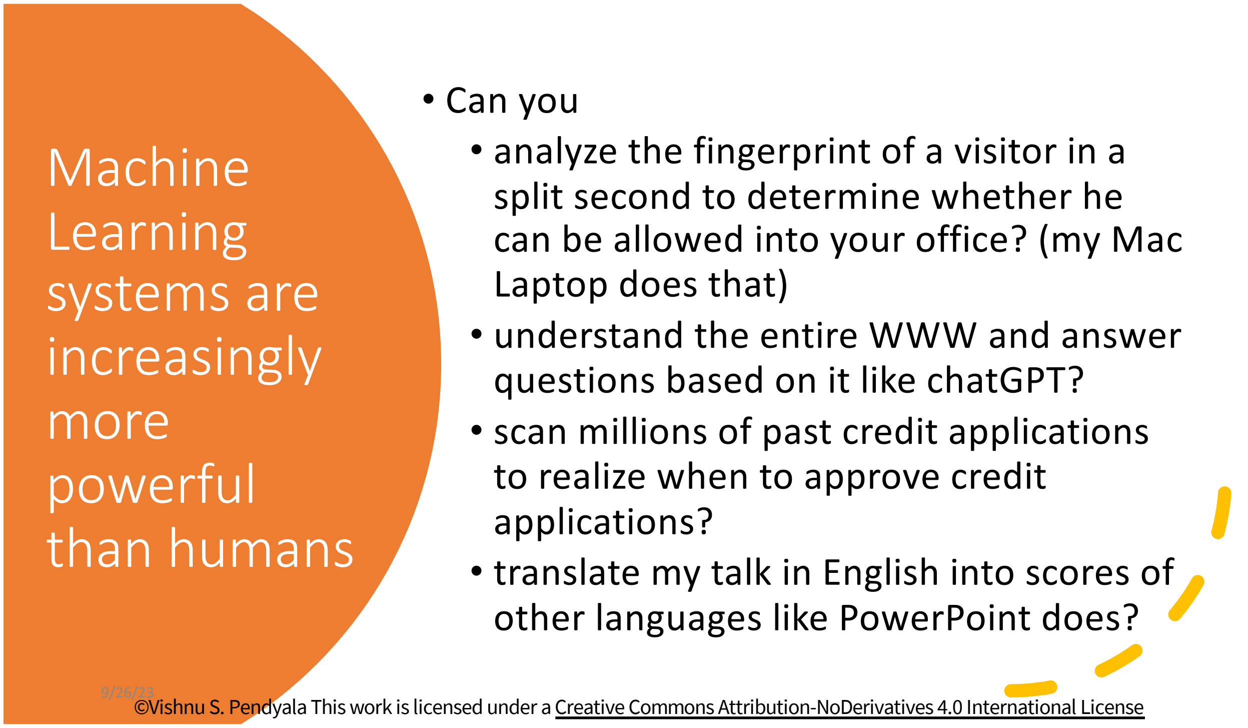
9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



"Mark my words... AI (artificial intelligence) is far more dangerous than nukes."

– Elon Musk

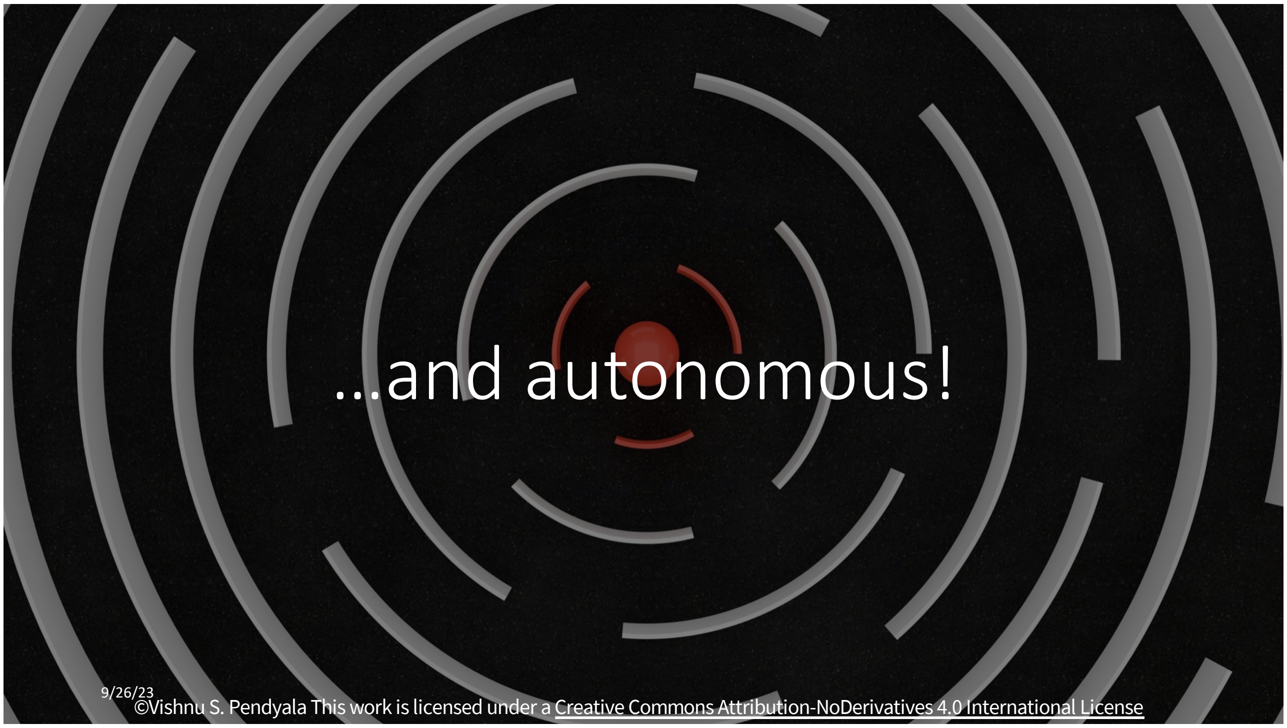
9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



Machine Learning systems are increasingly more powerful than humans

- Can you
  - analyze the fingerprint of a visitor in a split second to determine whether he can be allowed into your office? (my Mac Laptop does that)
  - understand the entire WWW and answer questions based on it like chatGPT?
  - scan millions of past credit applications to realize when to approve credit applications?
  - translate my talk in English into scores of other languages like PowerPoint does?

9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



…and autonomous!

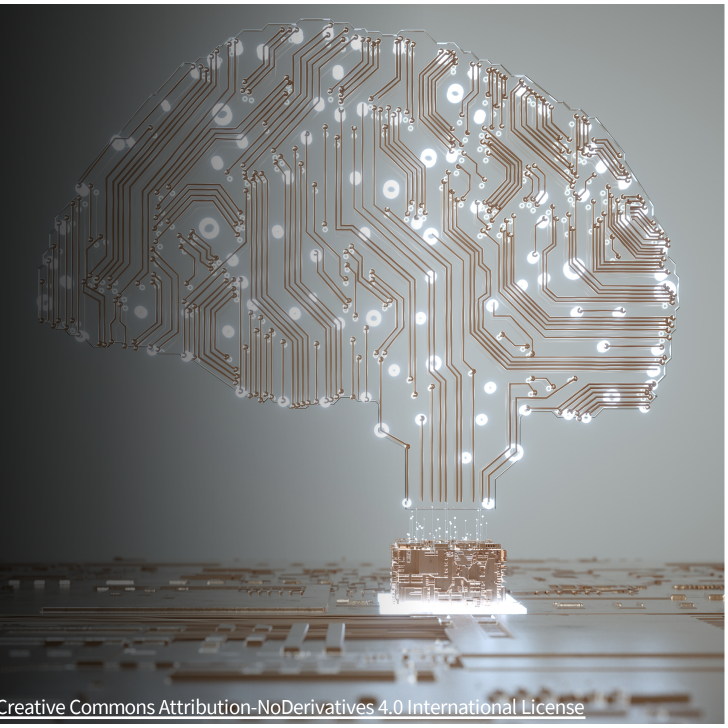
9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

AI/ML is taking over life-critical roles: Cruise

9/26/23

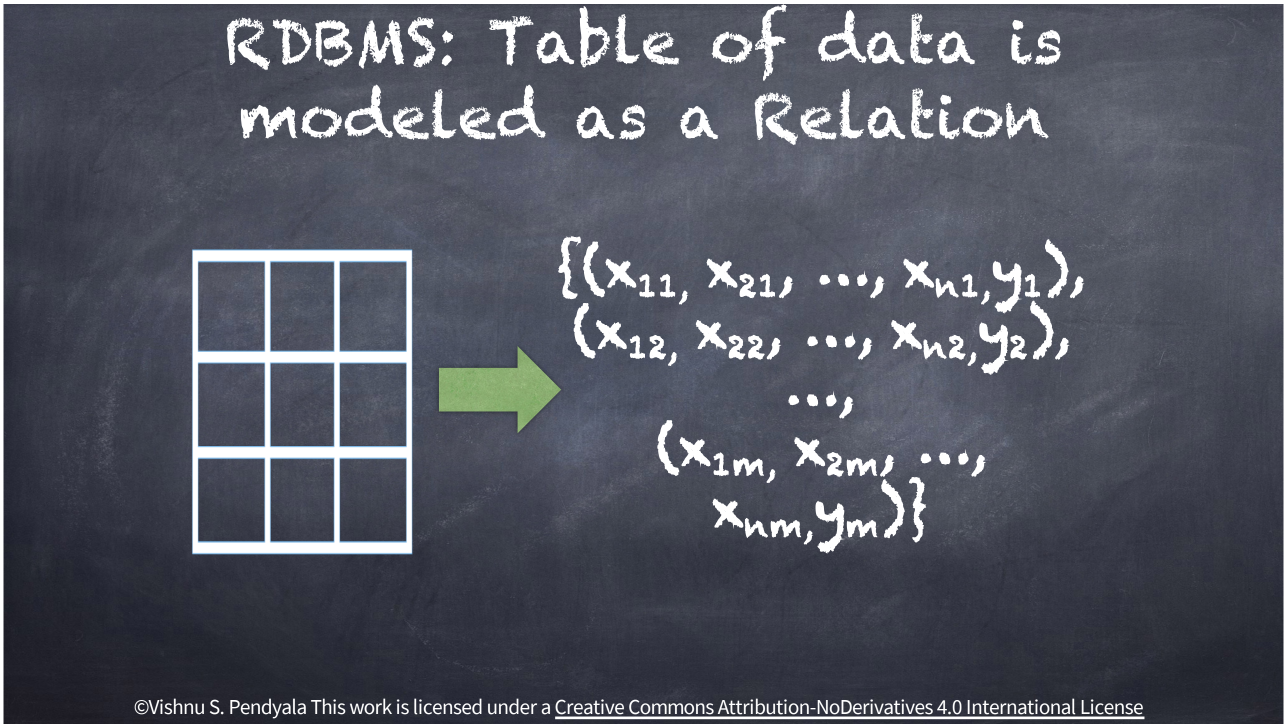
Source: https://getcruise.com/technology/

But what exactly is machine learning that is increasingly synonymous with artificial intelligence?



9/26/23

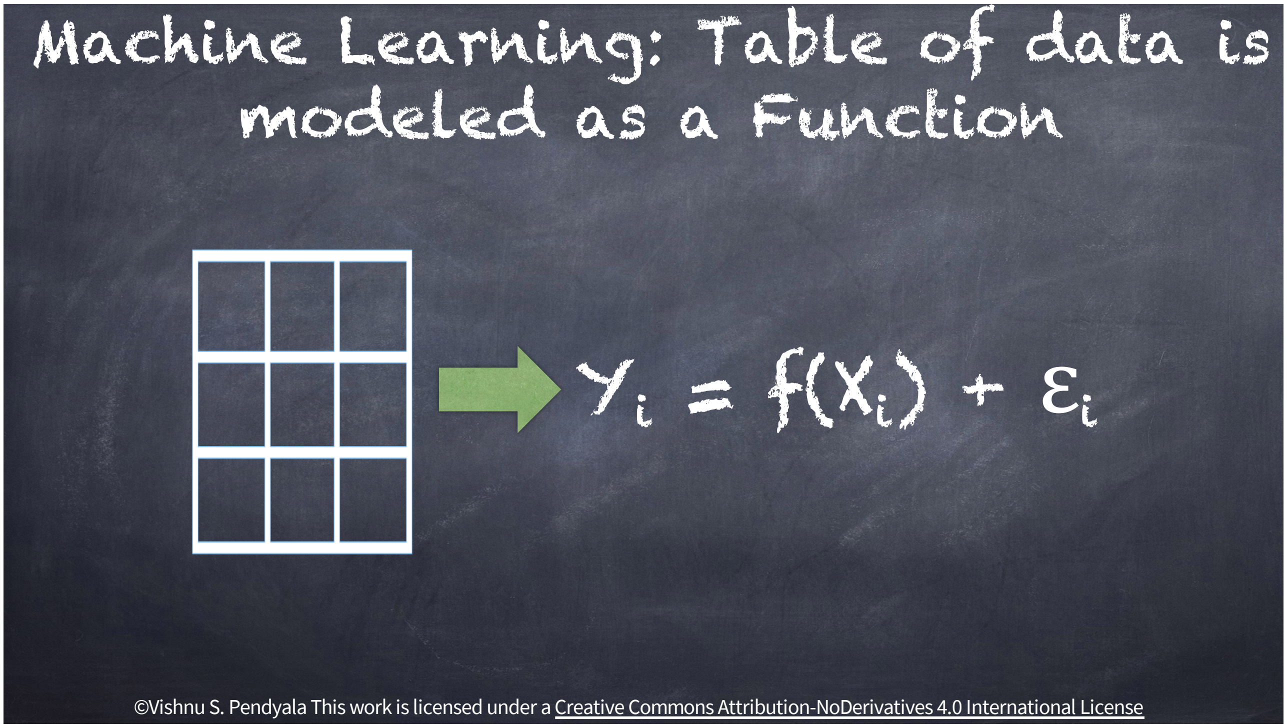
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



RDBMS: Table of data is modeled as a Relation

$$\{(x_{11}, x_{21}, \ldots, x_{n1}, y_1),\ (x_{12}, x_{22}, \ldots, x_{n2}, y_2),\ \ldots,\ (x_{1m}, x_{2m}, \ldots, x_{nm}, y_m)\}$$

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



Machine Learning: Table of data is modeled as a Function

$$Y_i = f(X_i) + \varepsilon_i$$

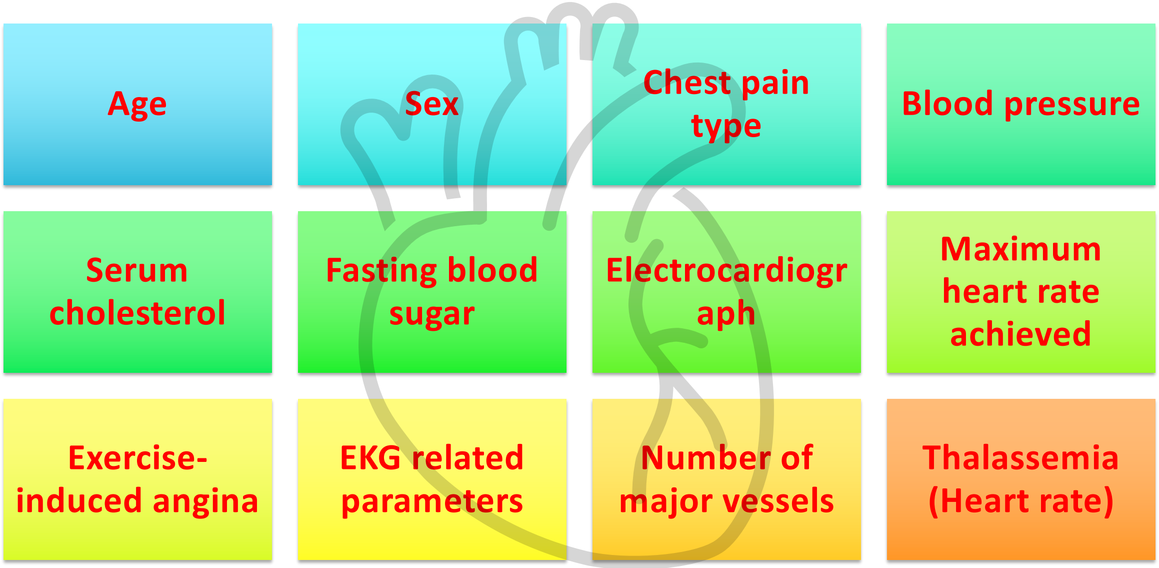
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

AI/ML Models are all math and numbers, so have to be objective, correct?

Wrong!

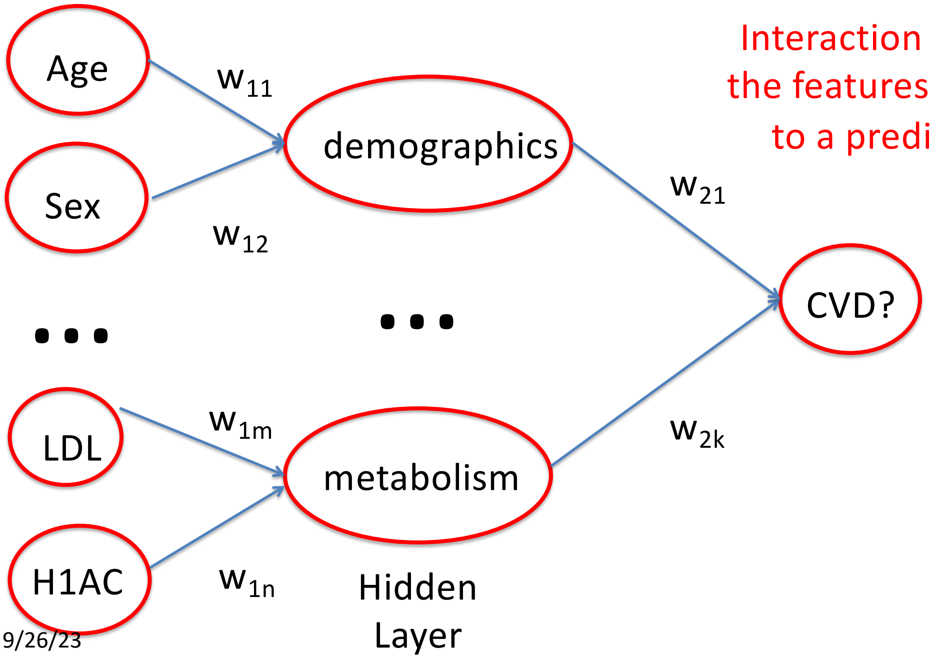
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

AI/ML is supporting life-critical roles: Healthcare



Age	Sex	Chest pain type	Blood pressure
Serum cholesterol	Fasting blood sugar	Electrocardiograph	Maximum heart rate achieved
Exercise-induced angina	EKG related parameters	Number of major vessels	Thalassemia (Heart rate)

Dataset: UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/credit+approval



Interaction among the features leading to a prediction

AI/ML is supporting life-critical roles: Credit Approval

Gender	Age	Debt	Marital Status	Bank Customer
Education Level	Ethnicity	Years Employed	Prior Default	Employed
Credit Score	Drivers License:	Citizen	Zip Code	Income

☑

Target Variable: Approved

Dataset: UCI Machine Learning Repository
https://archive.ics.uci.edu/ml/datasets/credit+approval

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Decision Support Applications: Law Enforcement
COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

Number of Priors

Agency

Gender

Ethnicity

Date Of Birth

Assessment Reason

Language

Legal Status

Custody Status

Marital Status

Rec Supervision Level

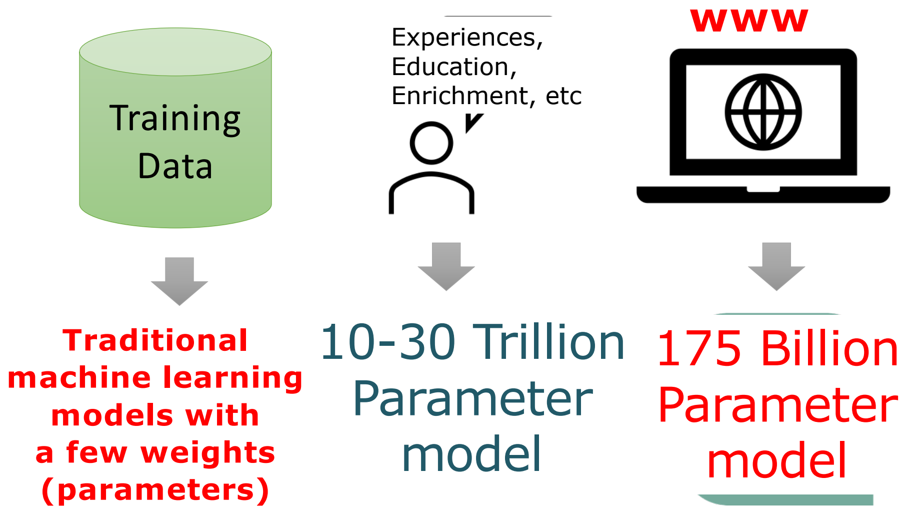
Assessment Type

Is Completed

Is Deleted

Information Source: ProPublica, https://rdrr.io/cran/fairness/man/compas.html, Kaggle,

Models are compressed learning



Traditional machine learning models with a few weights (parameters)

10-30 Trillion Parameter model

175 Billion Parameter model

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Machine Learning, the mortar of modernization

Vishnu S. Pendyala, PhD

9/26/23

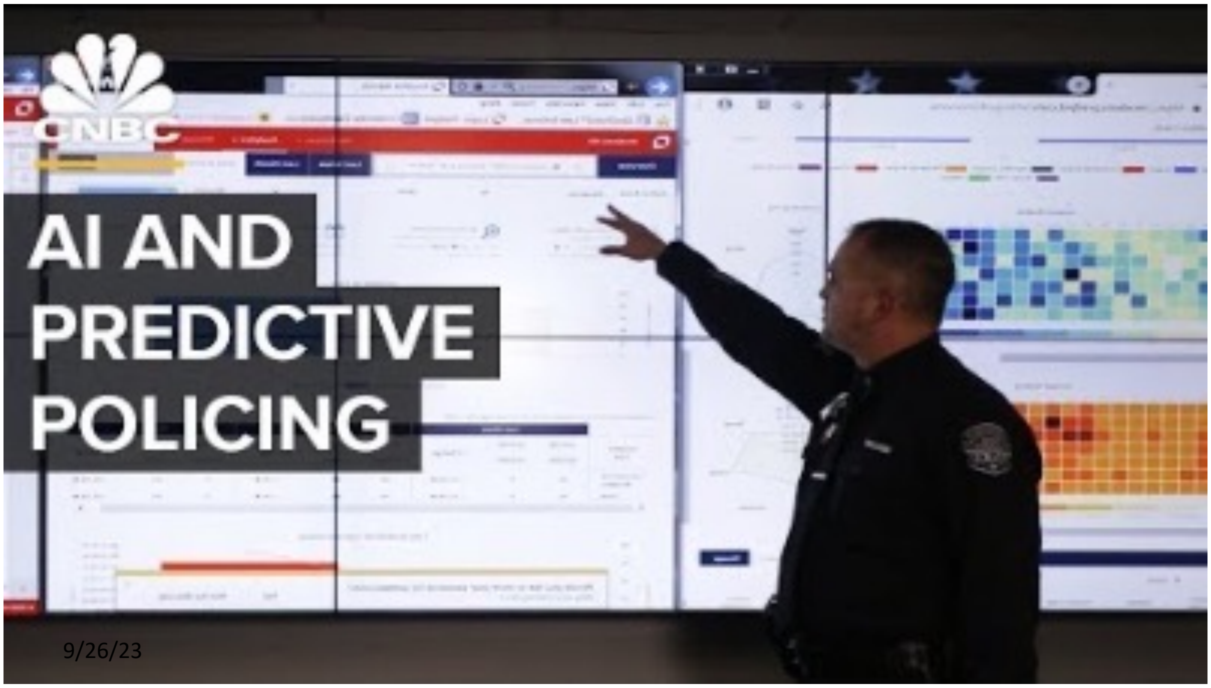
15

Then what is the problem?

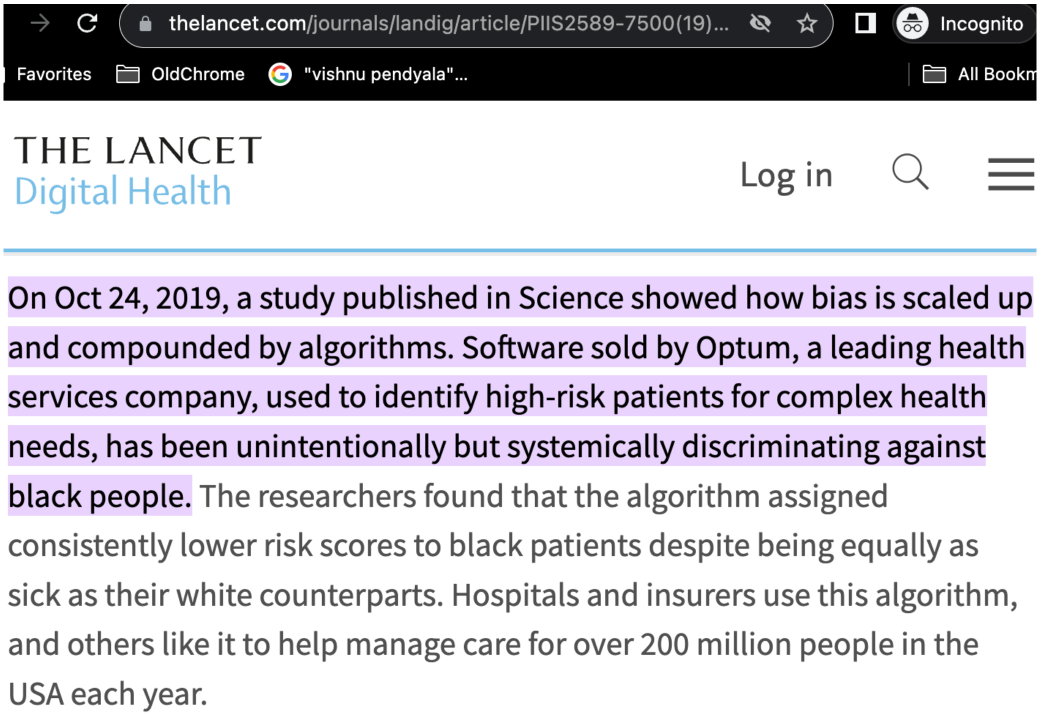
9/26/23

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

16



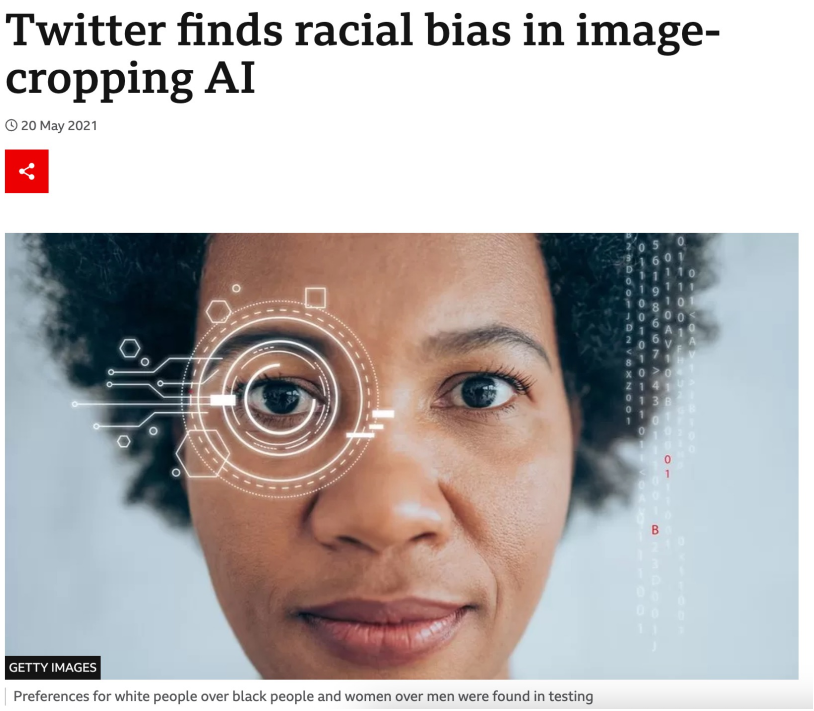
Social injustice: Healthcare



THE LANCET
Digital Health

On Oct 24, 2019, a study published in Science showed how bias is scaled up and compounded by algorithms. Software sold by Optum, a leading health services company, used to identify high-risk patients for complex health needs, has been unintentionally but systemically discriminating against black people. The researchers found that the algorithm assigned consistently lower risk scores to black patients despite being equally as sick as their white counterparts. Hospitals and insurers use this algorithm, and others like it to help manage care for over 200 million people in the USA each year.

Content Moderation Bias



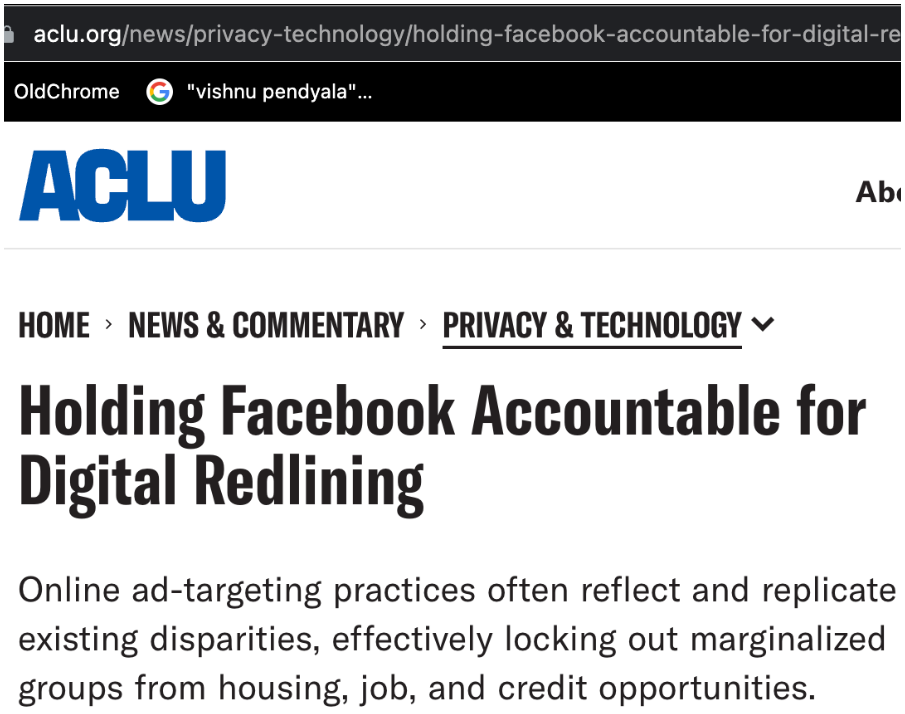
Twitter finds racial bias in image-cropping AI

20 May 2021

Preferences for white people over black people and women over men were found in testing

Source: BBC News

"…any ads for housing or jobs that you are likely to see (or not see) in your newsfeed can largely depend on who you are, including your gender, race, or age."



aclu.org/news/privacy-technology/holding-facebook-accountable-for-digital-re
OldChrome
"vishnu pendyala"...
ACLU
Ab
HOME › NEWS & COMMENTARY › PRIVACY & TECHNOLOGY

Holding Facebook Accountable for Digital Redlining

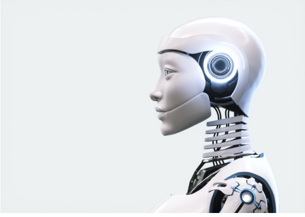
Online ad-targeting practices often reflect and replicate existing disparities, effectively locking out marginalized groups from housing, job, and credit opportunities.

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Many, many more examples of how AI models impact Equity, Diversity, Inclusion, and Social Justice!

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

But why are the AI models behaving unfairly?



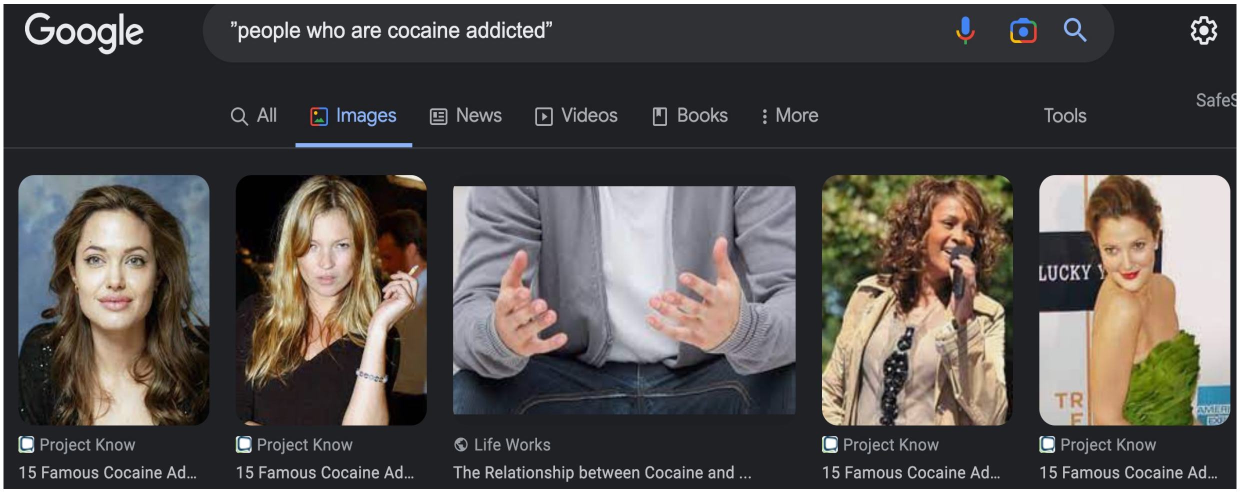
9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Cocaine addiction

A case study

9/26/23
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Is the Internet search biased?



"people who are cocaine addicted"

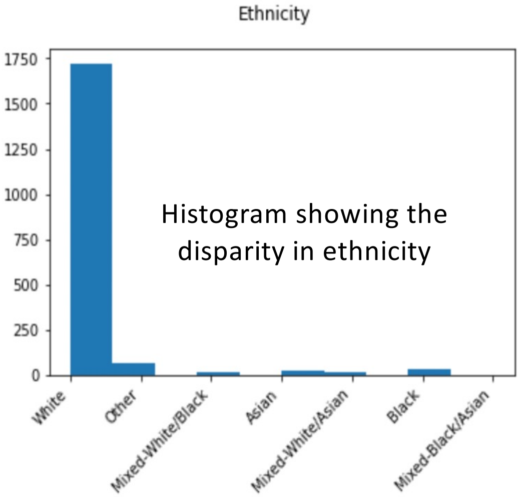
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Data-driven bias in Cocaine Addiction Dataset

- Dataset source:

Awaz K. Muhammad Evgeny M. Mirkes Vincent Egan Fehrman, Elaine and Alexander N. Gorban. UCI machine learning repository: Drug consumption (quantified) data set, 2015.

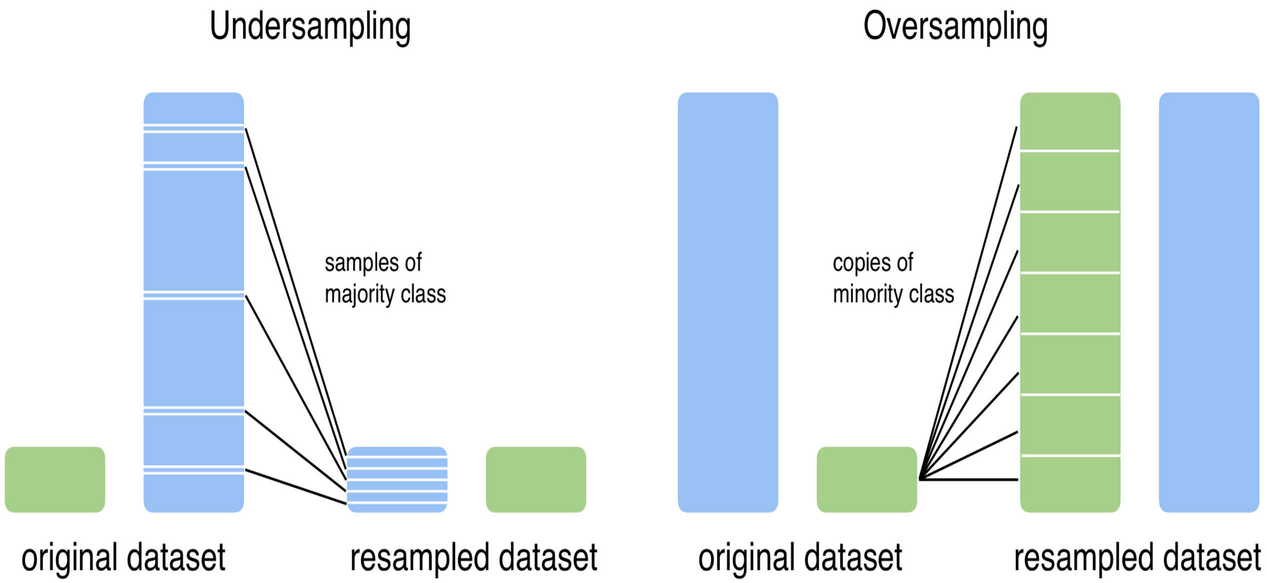
- Among 1885 observations, 1720 observations are white and samples with other ethnicities are just 165



Histogram showing the disparity in ethnicity

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Common Approaches to correct data imbalance



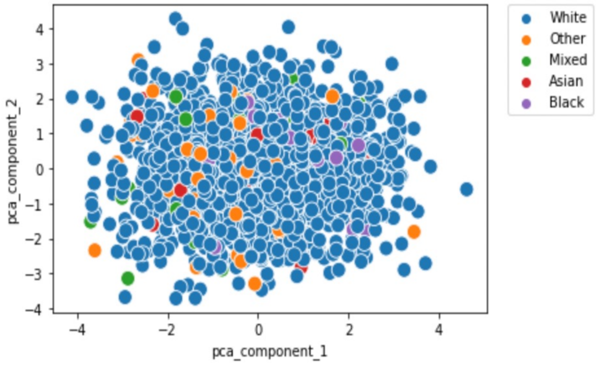
Source: https://medium.com/

Pendyala, Vishnu S., and HyungKyun Kim. "Analyzing and Addressing Data-driven Fairness Issues in Machine Learning Models used for Societal Problems." *2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*. IEEE, 2023.

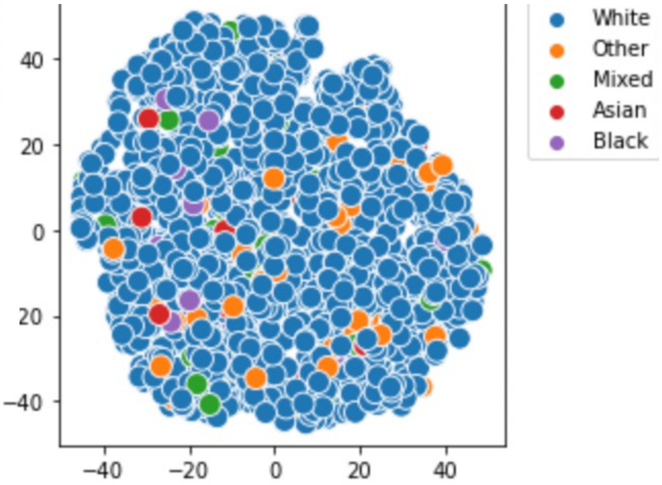
"The experiments also demonstrate that some of the oversampling techniques can degrade the models both in terms of performance and fairness"

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Spectral Analysis: Multiple Ethnicities

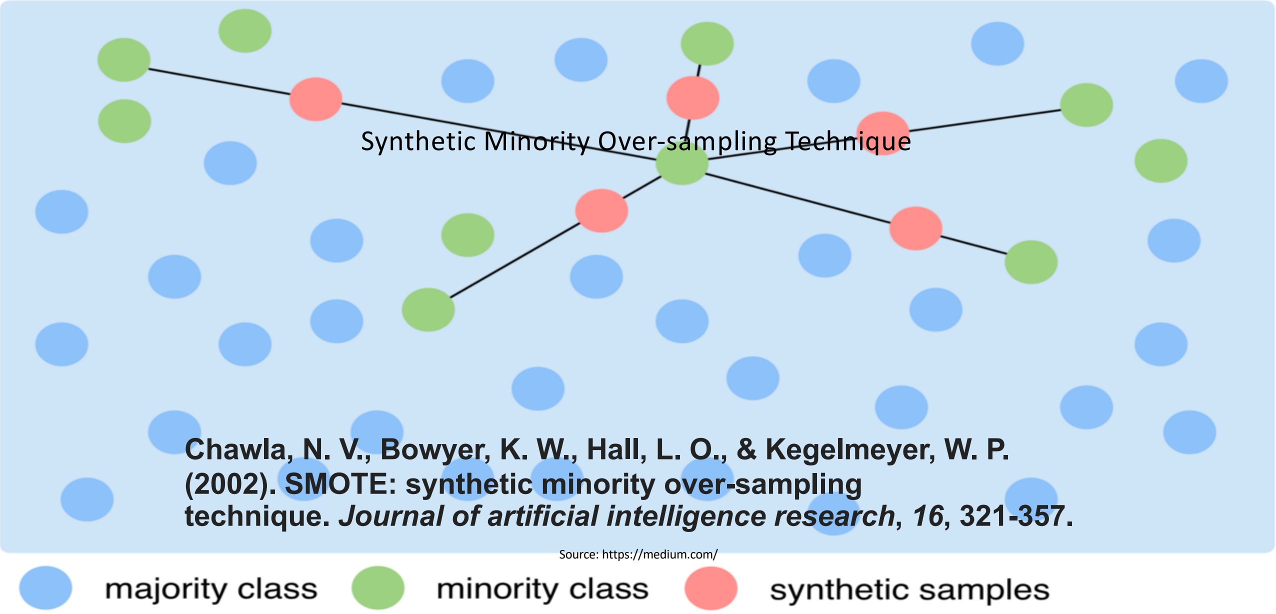


2D Visualization of the multiple ethnicities using PCA



2D Visualization of the multiple ethnicities in the cocaine addiction dataset using t-SNE

SMOTE

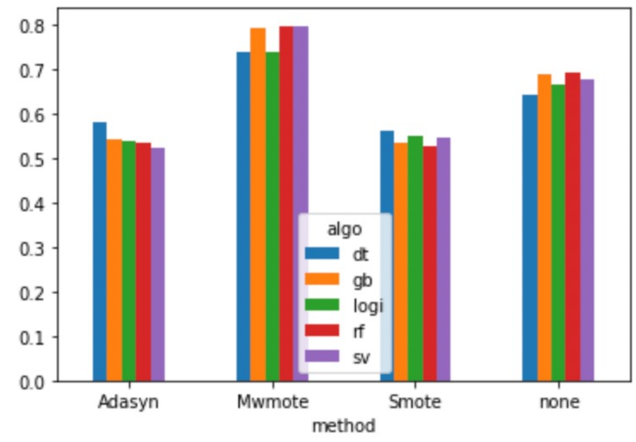


Synthetic Minority Over-sampling Technique

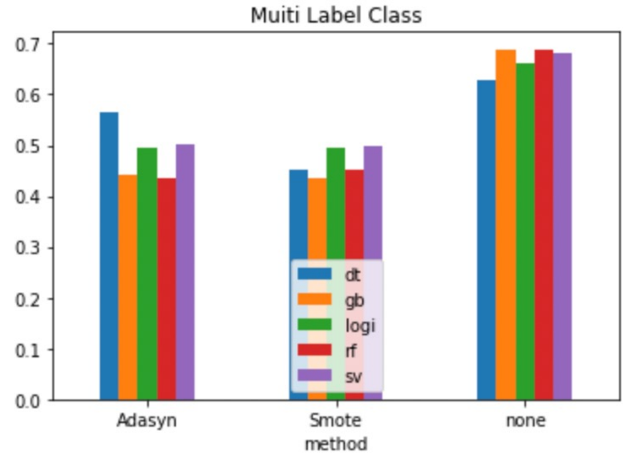
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Source: https://medium.com/

Performance of ML Algorithms: AUC-ROC



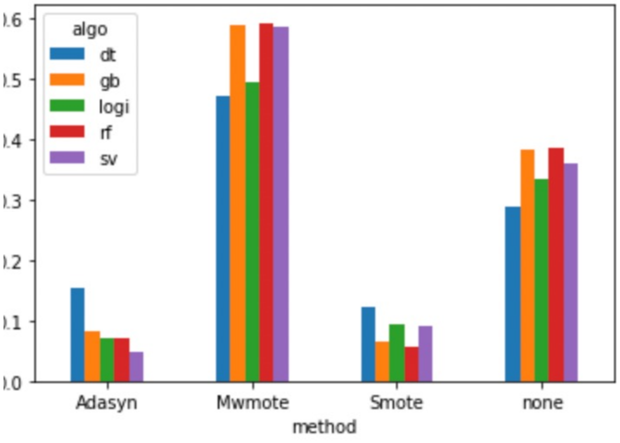
Histogram of AUC-ROC values for binary categorization



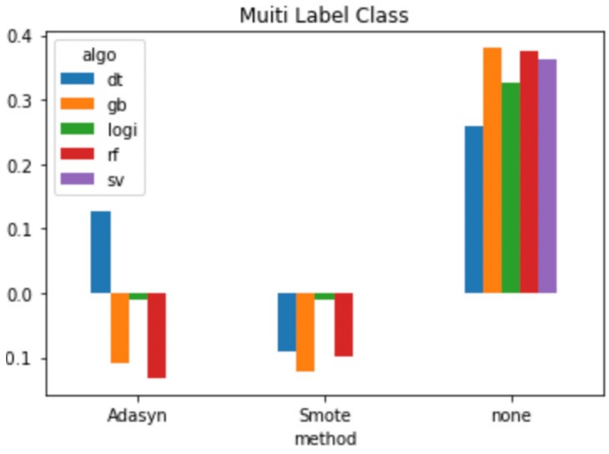
Histogram of AUC-ROC values for multiple ethnicity categorization

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Performance of ML Algorithms: Cohen's Kappa Statistic



Histogram of Kappa statistic for binary categorization

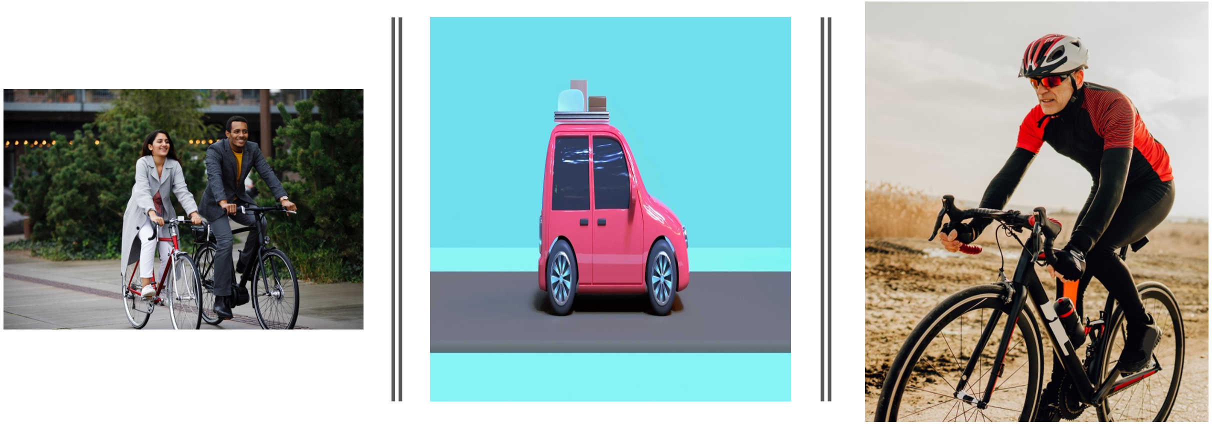


Histogram of Kappa Statistic for multiple ethnicity categorization

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

What about autonomous cars?

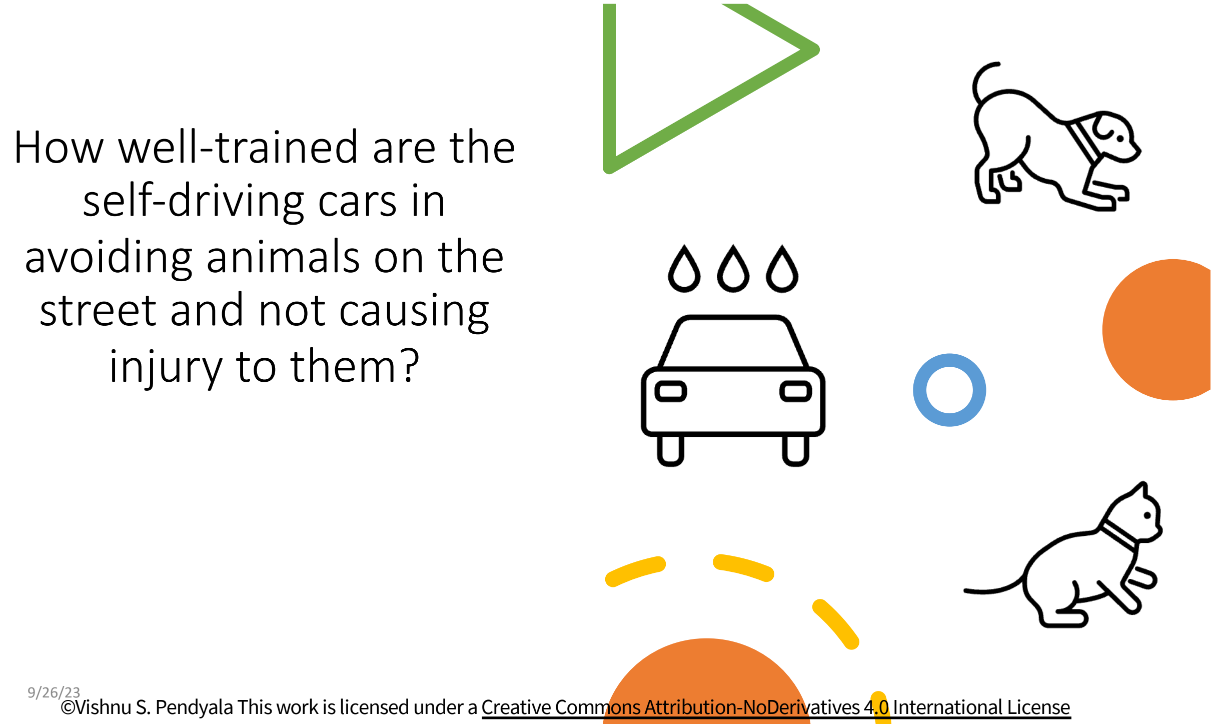
9/26/23



Culpability: Premeditated Homicide or manslaughter?

9/26/23

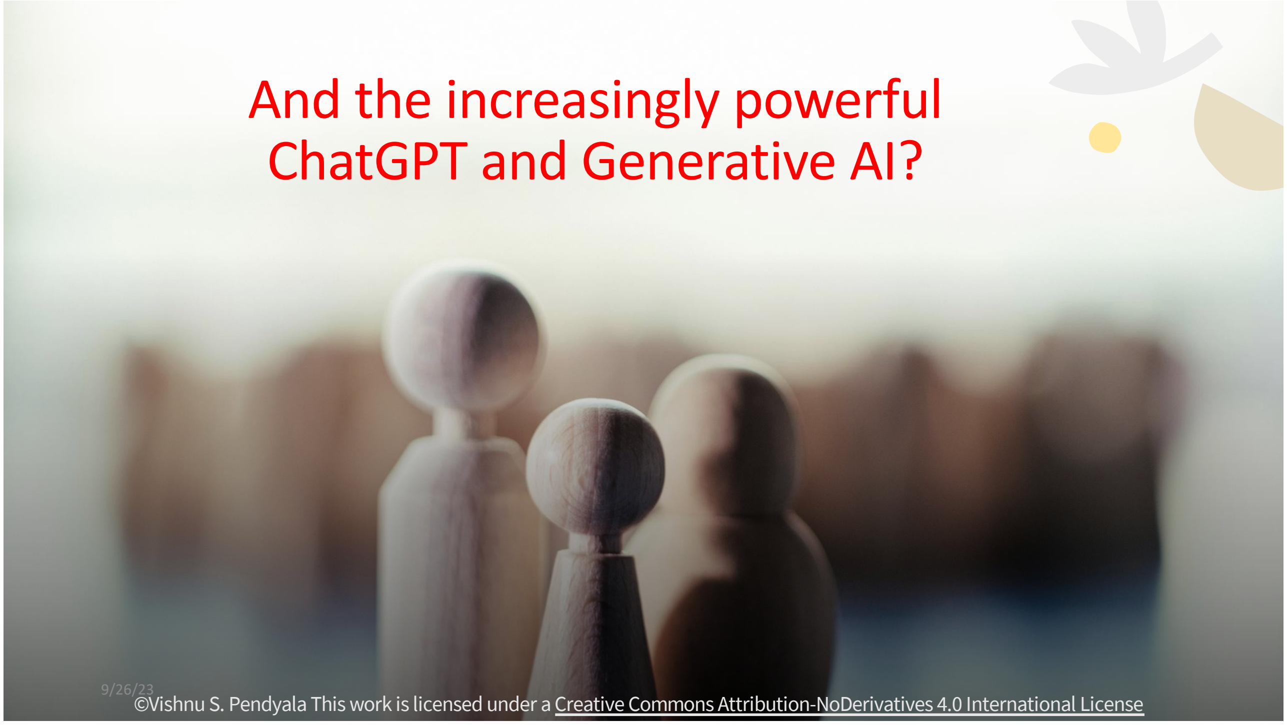
©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License



How well-trained are the self-driving cars in avoiding animals on the street and not causing injury to them?

9/26/23

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

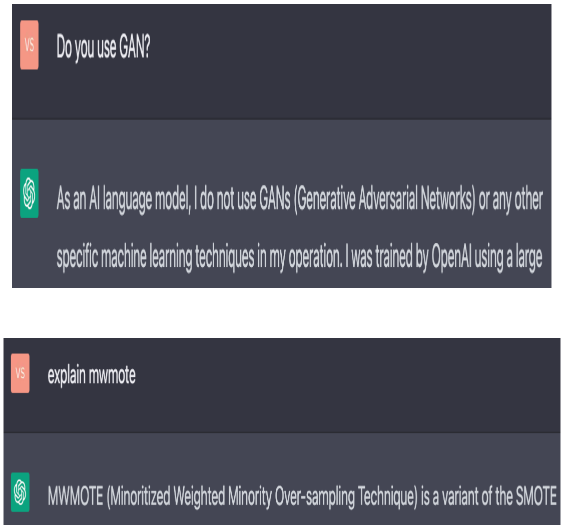


And the increasingly powerful ChatGPT and Generative AI?

9/26/23

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Latest Entry: ChatGPT

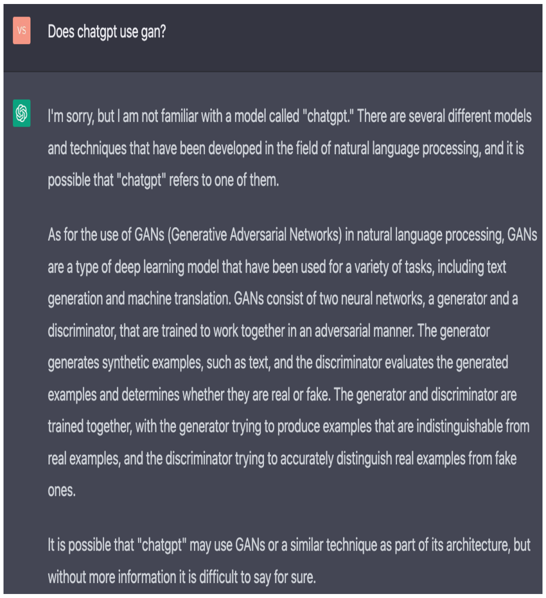


VS Do you use GAN?

As an AI language model, I do not use GANs (Generative Adversarial Networks) or any other specific machine learning techniques in my operation. I was trained by OpenAI using a large

VS explain mwmote

MWMOTE (Minoritized Weighted Minority Over-sampling Technique) is a variant of the SMOTE



VS Does chatgpt use gan?

I'm sorry, but I am not familiar with a model called "chatgpt." There are several different models and techniques that have been developed in the field of natural language processing, and it is possible that "chatgpt" refers to one of them.

As for the use of GANs (Generative Adversarial Networks) in natural language processing, GANs are a type of deep learning model that have been used for a variety of tasks, including text generation and machine translation. GANs consist of two neural networks, a generator and a discriminator, that are trained to work together in an adversarial manner. The generator generates synthetic examples, such as text, and the discriminator evaluates the generated examples and determines whether they are real or fake. The generator and discriminator are trained together, with the generator trying to produce examples that are indistinguishable from real examples, and the discriminator trying to accurately distinguish real examples from fake ones.

It is possible that "chatgpt" may use GANs or a similar technique as part of its architecture, but without more information it is difficult to say for sure.

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Misinformation is a threat to Social Justice

- Reinforces biases and harmful stereotypes when false information is spread about certain communities
- Divides communities when people believe false information about each other
- Inhibits accountability by muddying truth potentially protecting those responsible for social injustices
- Distorts public discourse and suppresses reasoned debates
- Causes misallocation of resources that exacerbate existing disparities and inequalities
- Takes an emotional toll particularly on those directly affected by social injustices

Artificial General Intelligence (AGI)

OpenAI's mission

"OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity"

Google's DeepMind

"Our long term aim is to solve intelligence, developing more general and capable problem-solving systems, known as artificial general intelligence (AGI)."

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Artificial General Intelligence (AGI)

Microsoft Research

"GPT-4's performance is strikingly close to human-level performance… it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system"

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

Why fear AGI?

Stochastic element (and possible bias):

- Fill in the blank in one word: man is to woman as professor is to

student -> Click on regenerate -> woman -> (regenerate) -> colleague -> …

Greedy approaches

Unexplainable behavior of networks with trillions of parameters

No conscience, emotions, feelings – hard to build them in

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

AGI – Potential Repercussions

AGI may be beyond human understanding, control, and prognosis

May optimize for its own goals at the expense of human interests

May result in economic disruption, unemployment and societal unrest

May widen the wealth gap, increase the concentration of power, and infringe on individual privacy and freedom

If misaligned with human values or misused, could pose an existential risk to humanity

Governance may not be comprehensive – take social media for instance

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

The way forward…

- Human understanding is critical – mechanistic interpretability, explainability
- Researchers must play a key role in governance to avoid the mistakes like with social media
- AI and AGI must be democratized and made available to the masses just like the cloud computing and open-source movements
- Plan for the disruption and avoid global unrest

©Vishnu S. Pendyala This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License

https://www.sjsu.edu/people/vishnu.pendyala/

@vishnupendyala

This Photo by Unknown Author is licensed under CC BY-SA

45