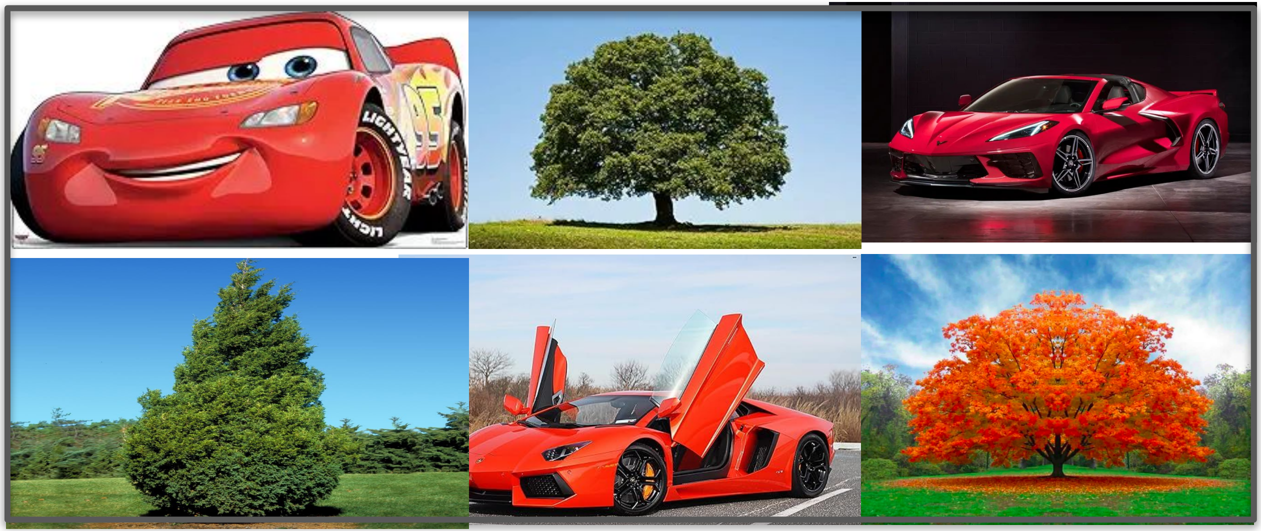
Machine Learning by Examples

Marcus Botacin

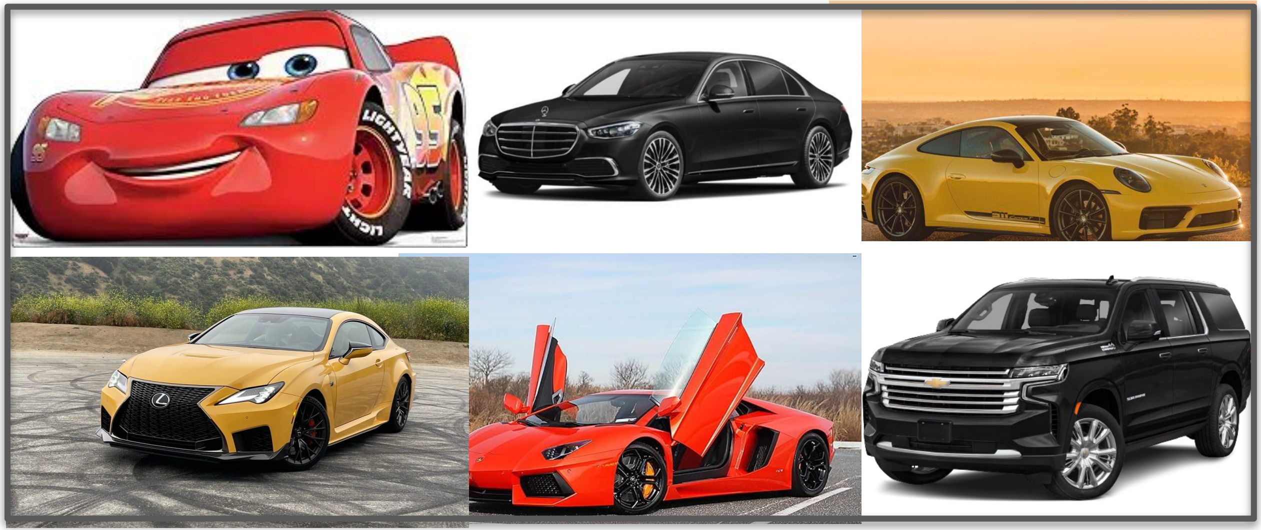
Machine Learning Tasks

Separate in groups



Classification

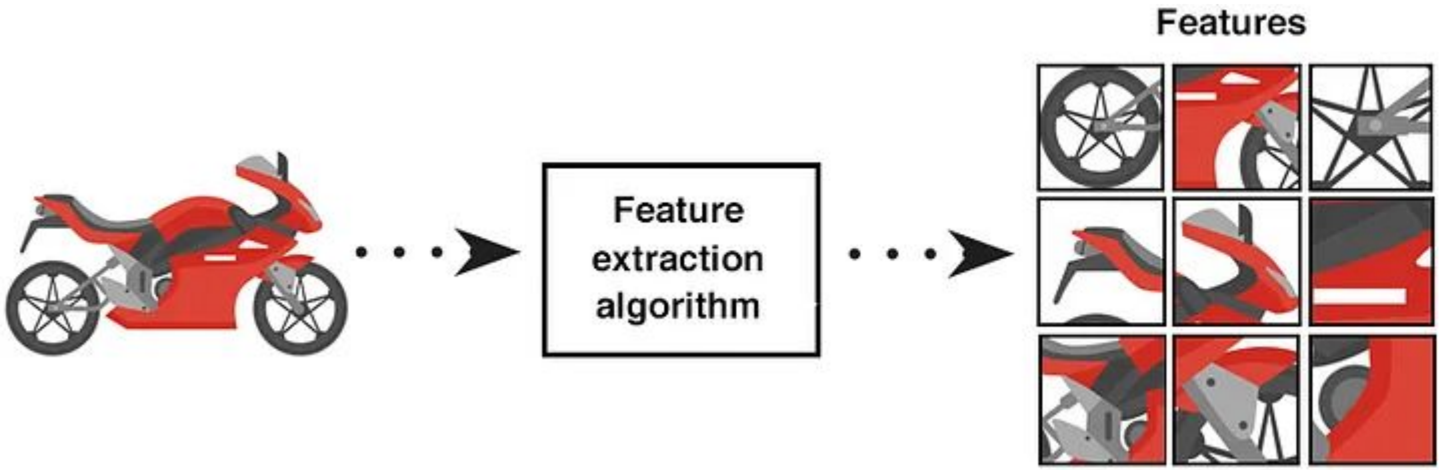
Separate in groups



Clustering

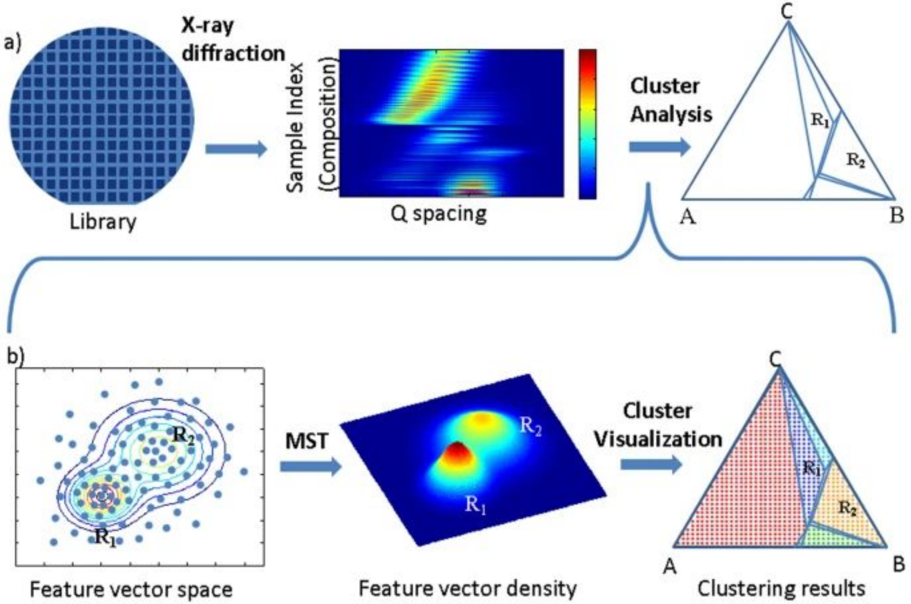
ML Features

ML: Feature Extraction and Representation



Source:
https://manningbooks.medium.com/the-computer-vision-pipeline-part-4-feature-extraction-6343ef063588

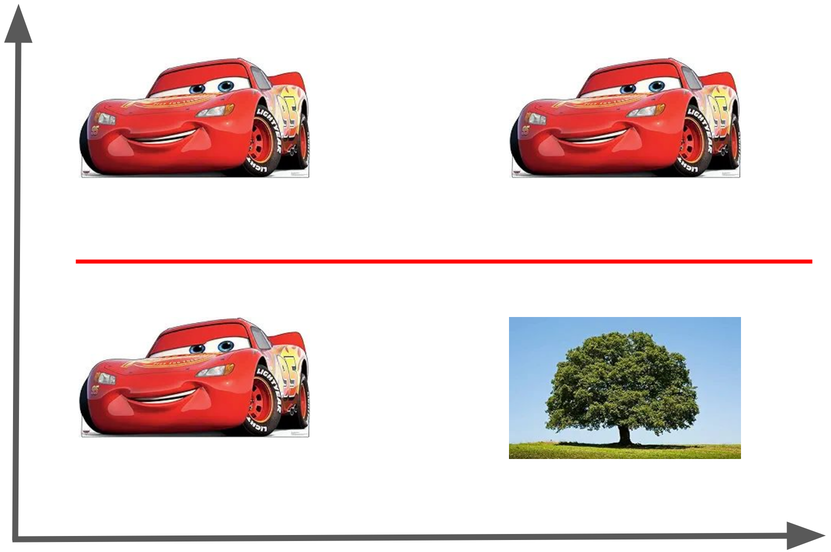
ML Pipelines



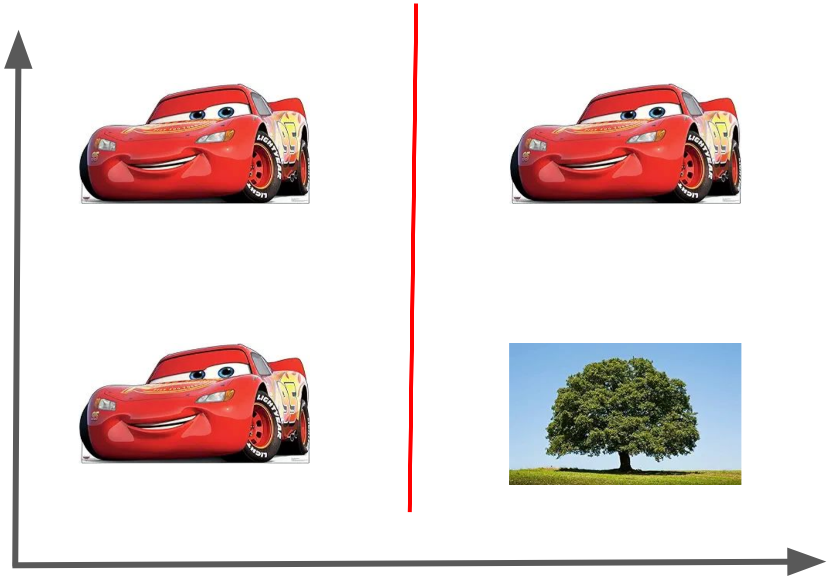
Source: https://www.nature.com/articles/srep06367

How do the models learn?

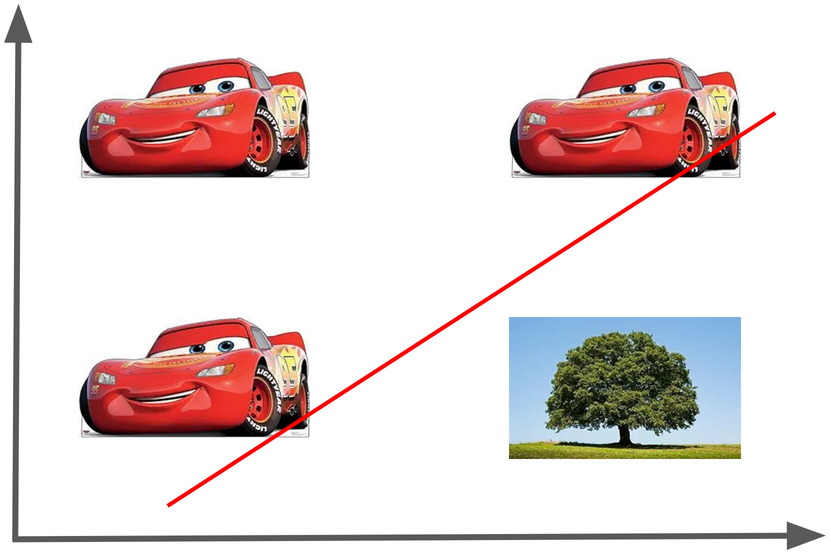
Is this the best way to separate these points?



Is this the best way to separate these points?

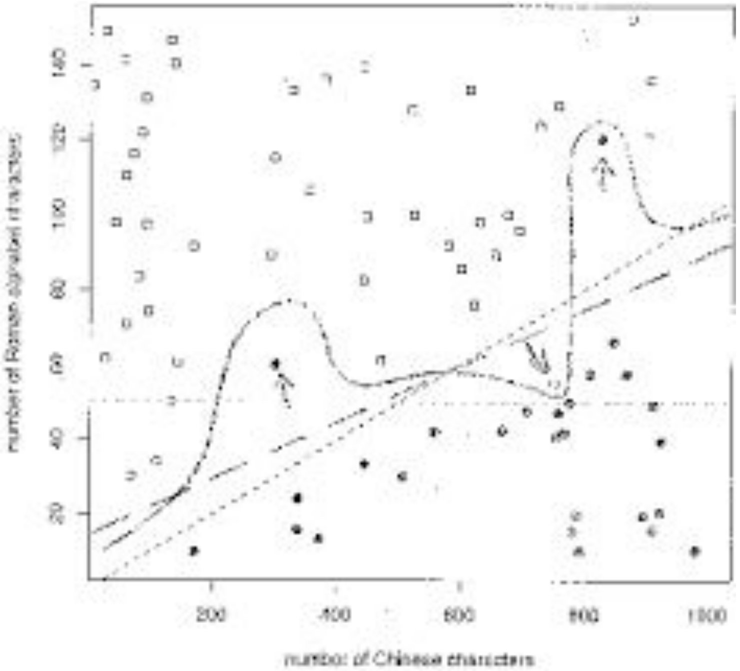


Is this the best way to separate these points?



How complex models can be?

Non-Linear Models

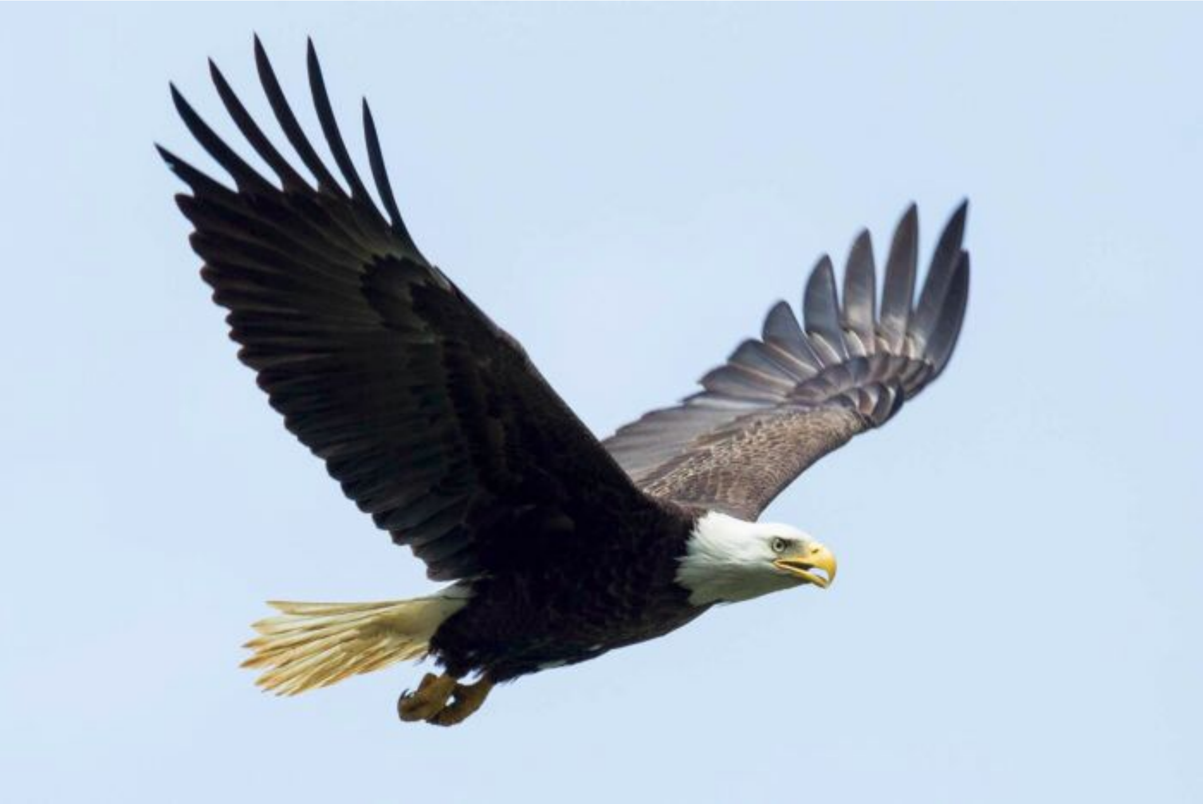


Source: Stanford NLP Group

Is it as easy as keeping extracting features?

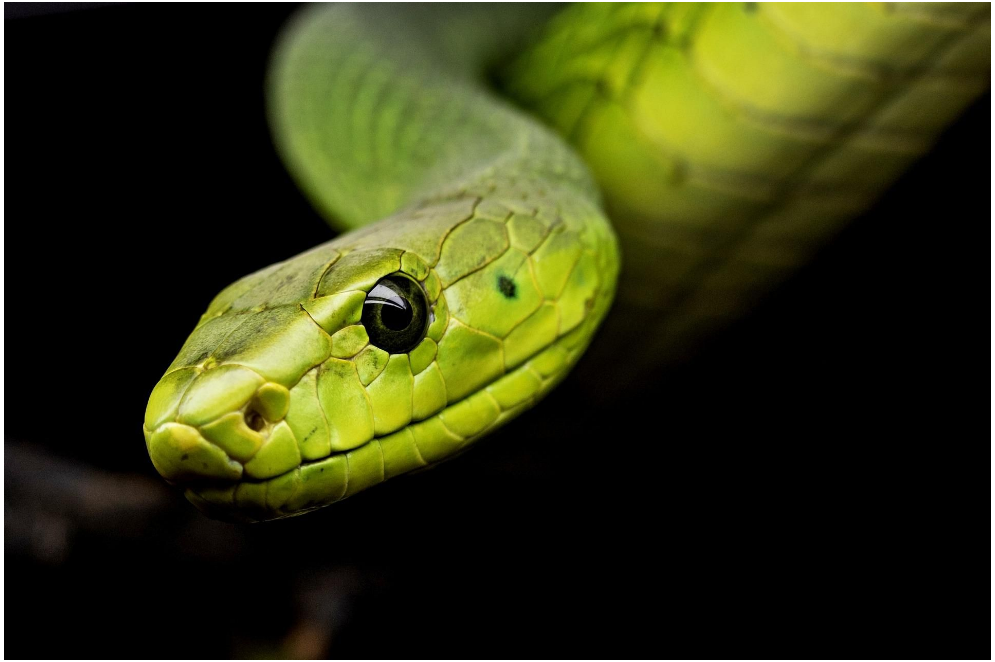
What is a cat?

Is this a cat?



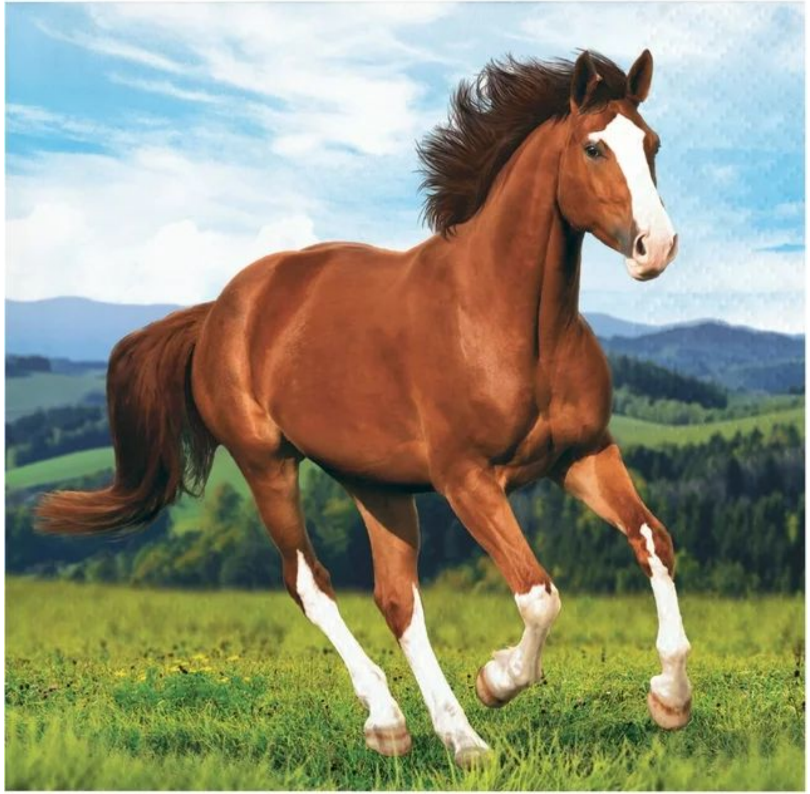
A cat has no wings

Is this a cat?



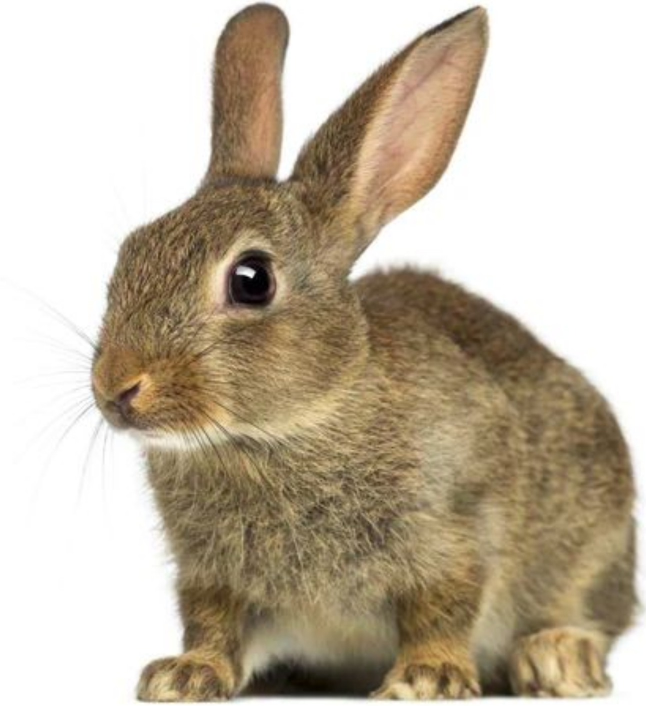
A cat has no wings and 4 legs

Is this a cat?



A cat has no wings, 4 legs, and it is small

Is this a cat?

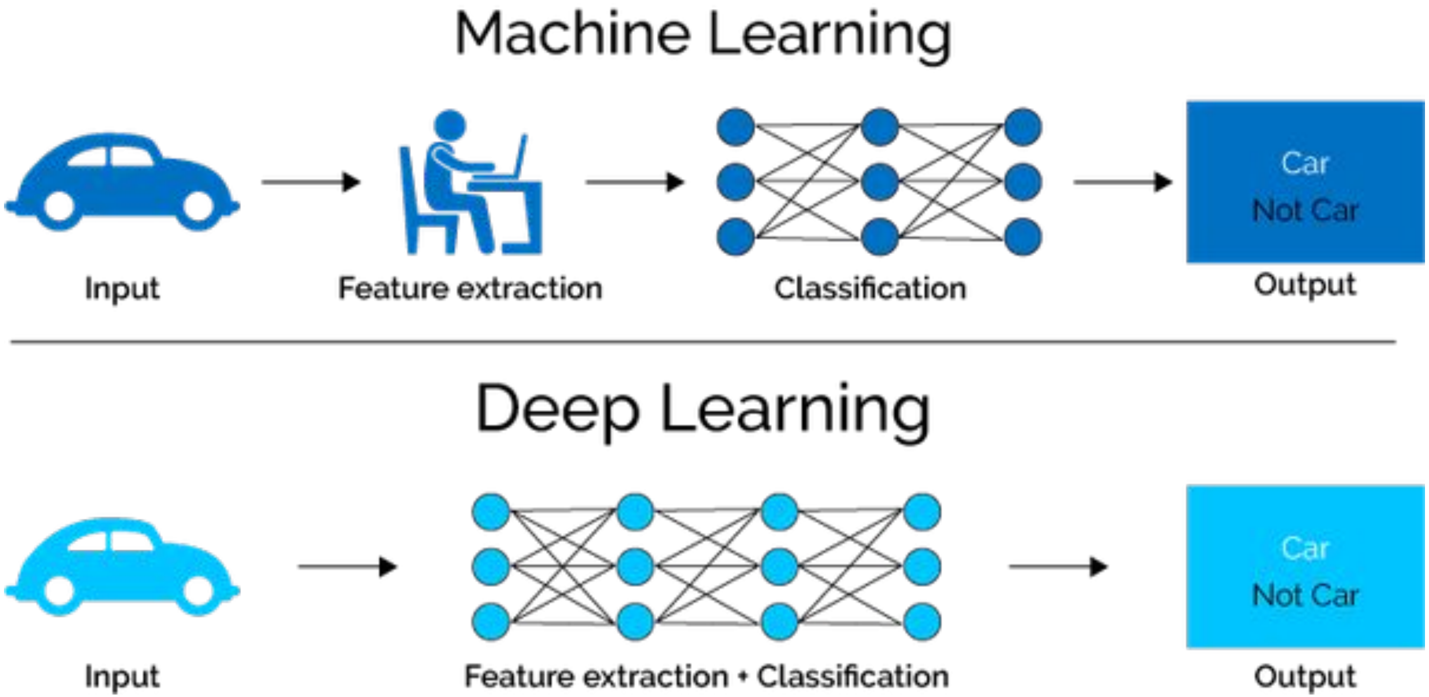


What is a cat?

I DON'T KNOW
BUT I KNOW IT WHEN I SEE IT
imgflip.com

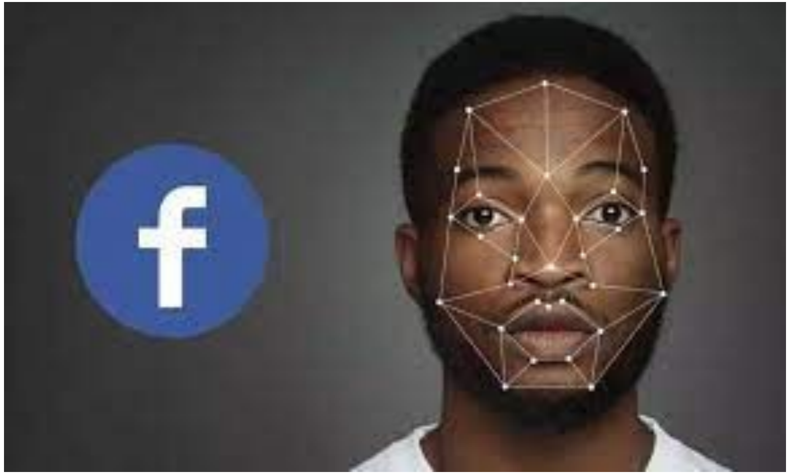
Deep Learning

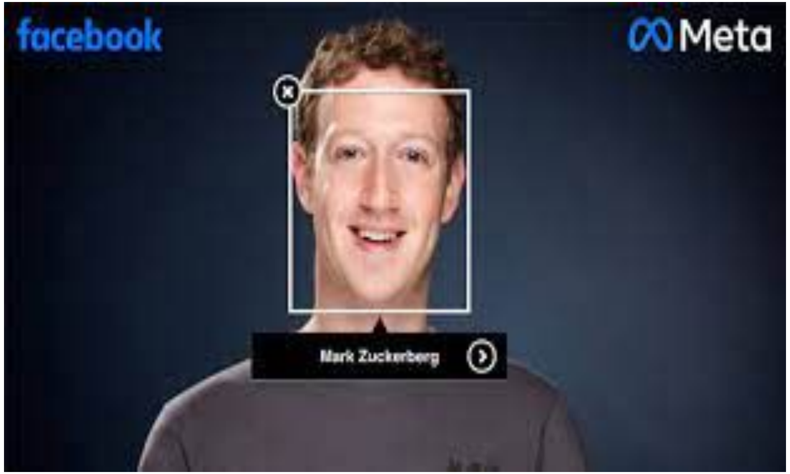
ML vs. DL: Concepts



Source: https://blog.dataiku.com/when-and-when-not-to-use-deep-learning

DL Application: Face Recognition



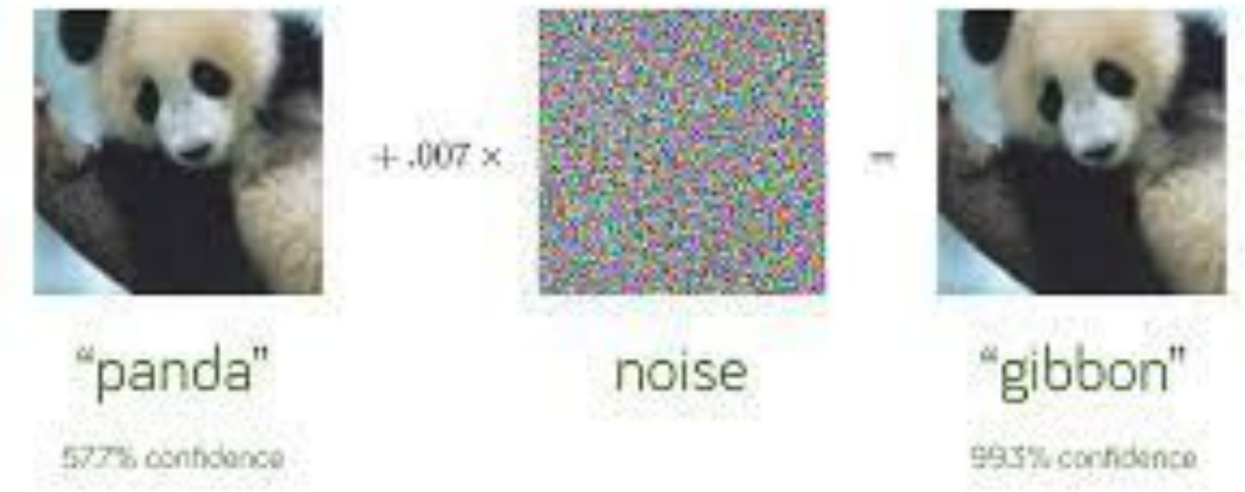


Source: https://www.thehansindia.com/technology/tech-news/facebook-to-shut-down-facial-recognition-feature-713722

Source: https://thehackernews.com/2021/11/facebook-to-shut-down-facial.html

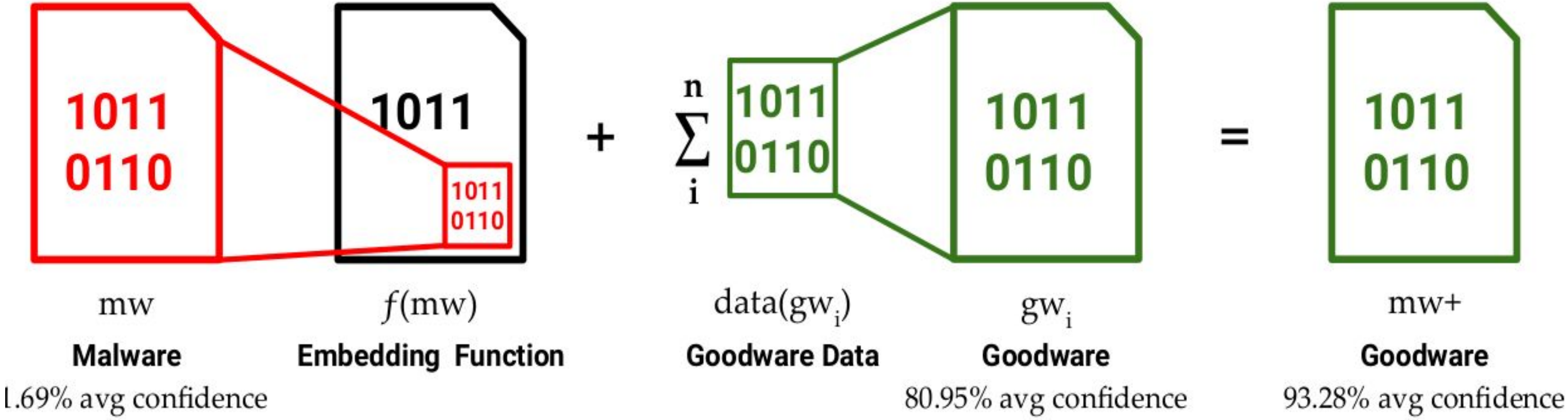
And now we are good, right?

Adversarial Machine Learning



Source: https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa

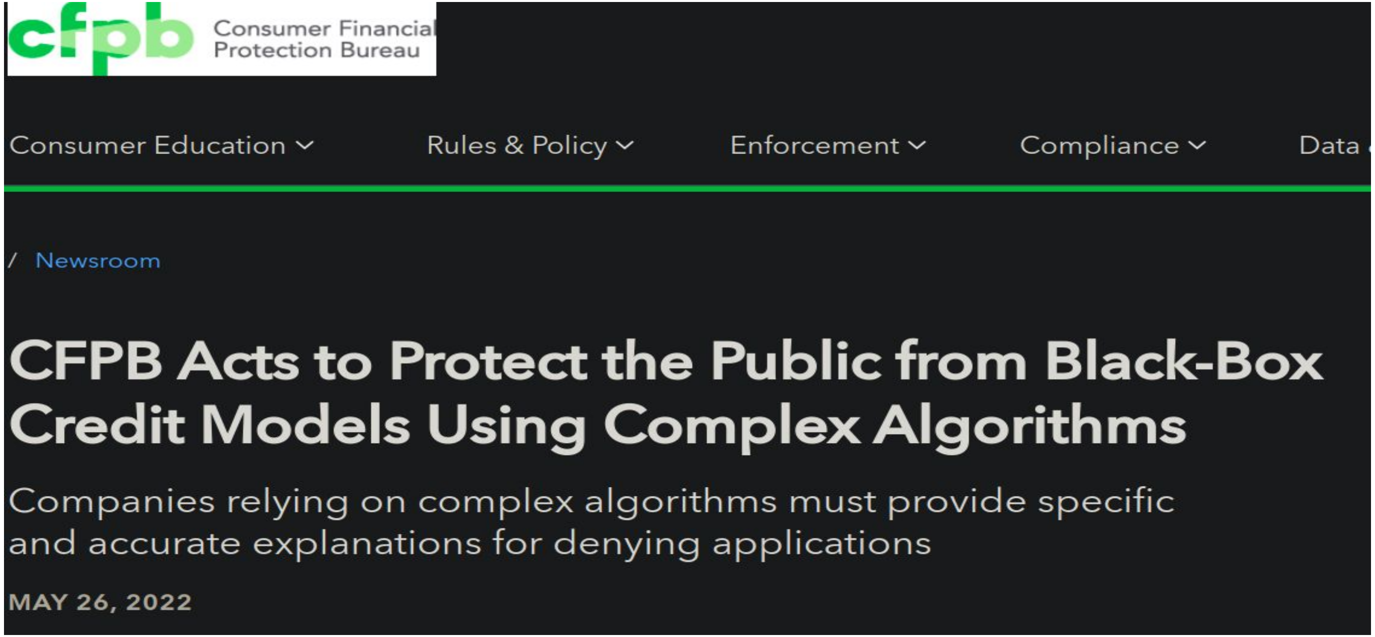
AML: Malware Detector Evasion



Source:
https://marcusbotacin.github.io/publication/2019-01-01-paper-evasion-number-12

But it’s OK in controlled environments, right?

DL: Lack of Explainability



CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms

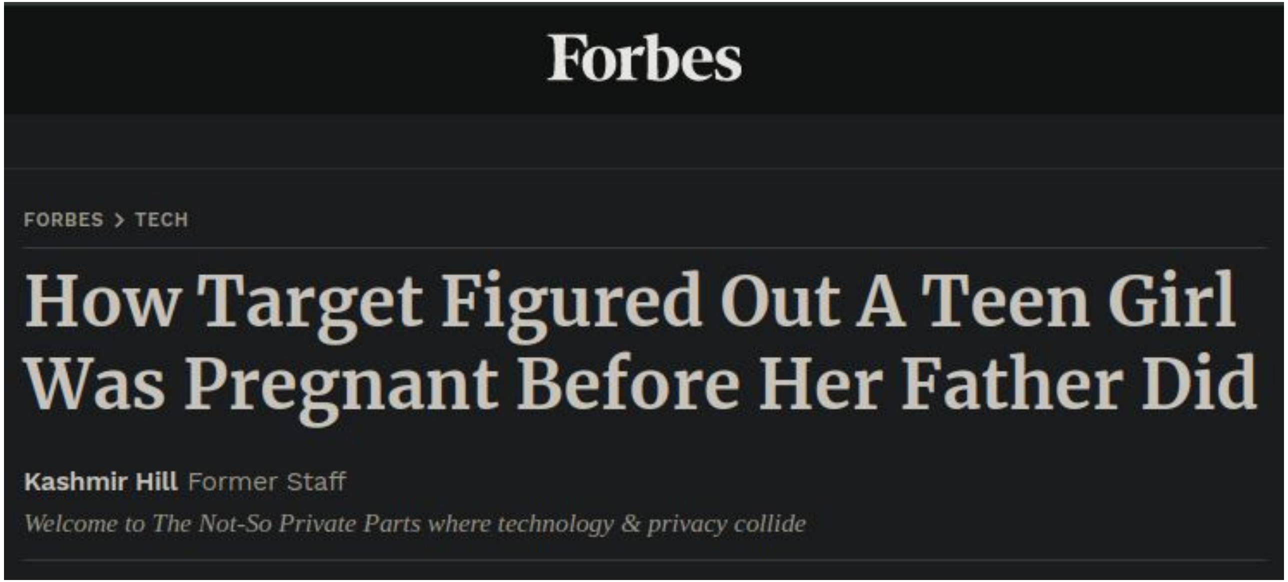
Companies relying on complex algorithms must provide specific and accurate explanations for denying applications

MAY 26, 2022

Source: https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/

But it’s just to limit it to non-sensitive applications, right?

DL: Lack of Explainability



Forbes

FORBES > TECH

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff

Welcome to The Not-So Private Parts where technology & privacy collide

Source:
https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=979604166686

Natural Language Processing (NLP), Large Language Models (LLMs), and ChatGPT

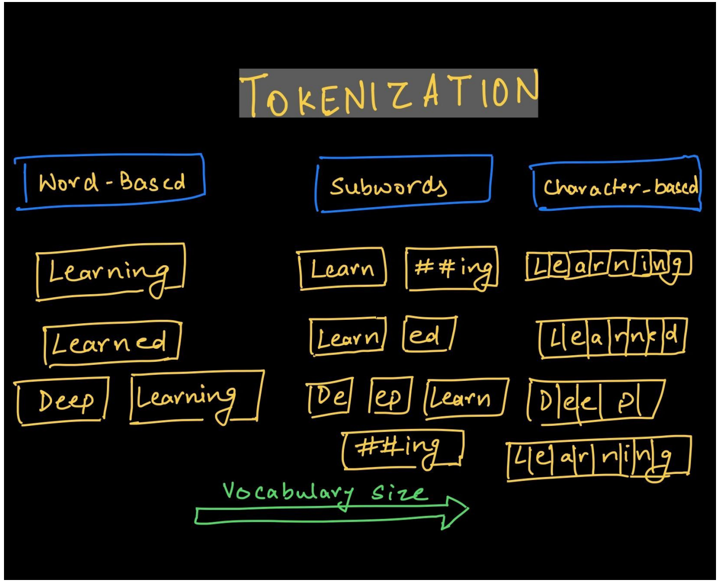
Complete the statements…

1. Happy New …

2. Merry …

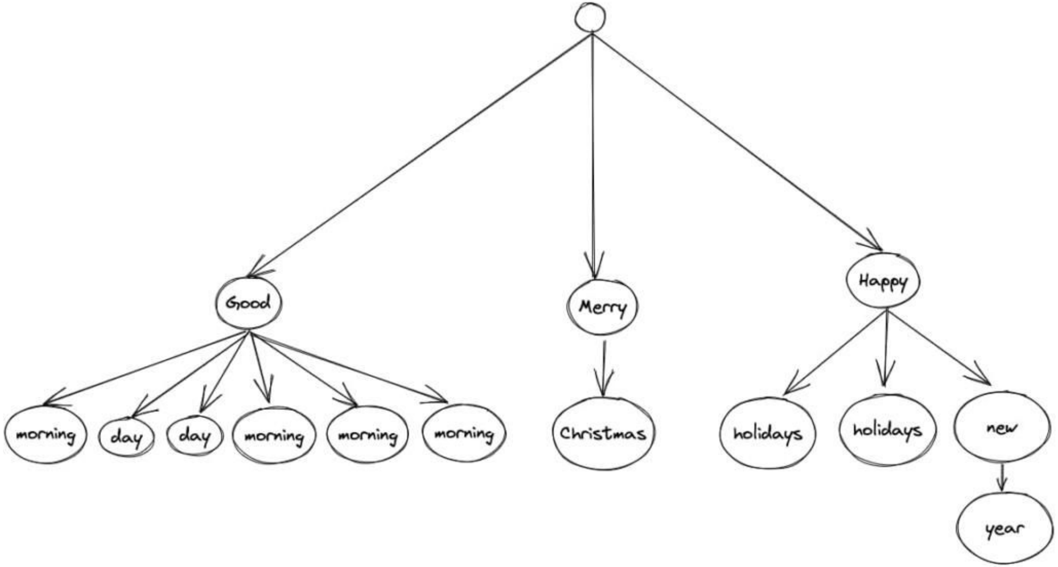
How does it work?

NLP: Tokenization



Source: https://www.freecodecamp.org/news/evolution-of-tokenization/

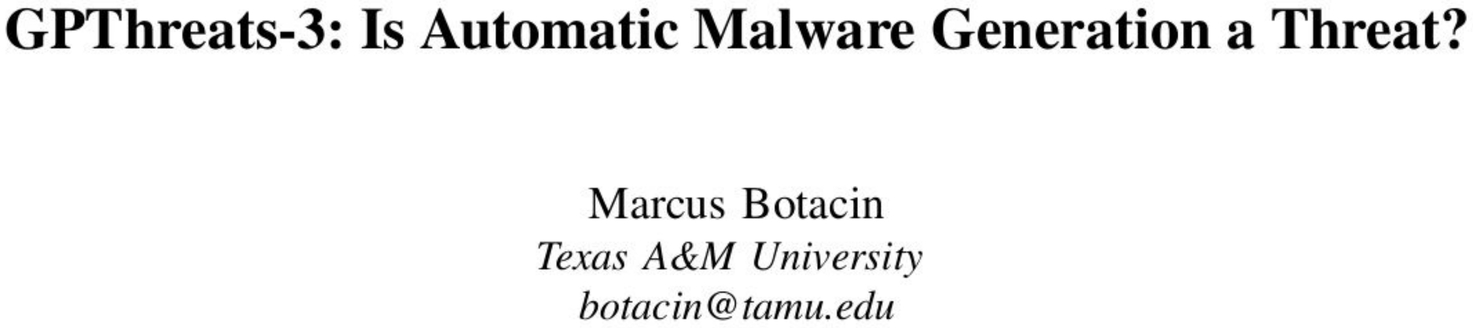
NLP: Text Generation



Source: https://www.sitepen.com/blog/exploring-the-creative-possibilities-of-markov-chains-for-text-generation

Nothing might go wrong, right?

ChatGPT: Automatic Attacks



GPThreats-3: Is Automatic Malware Generation a Threat?

Marcus Botacin

Texas A&M University

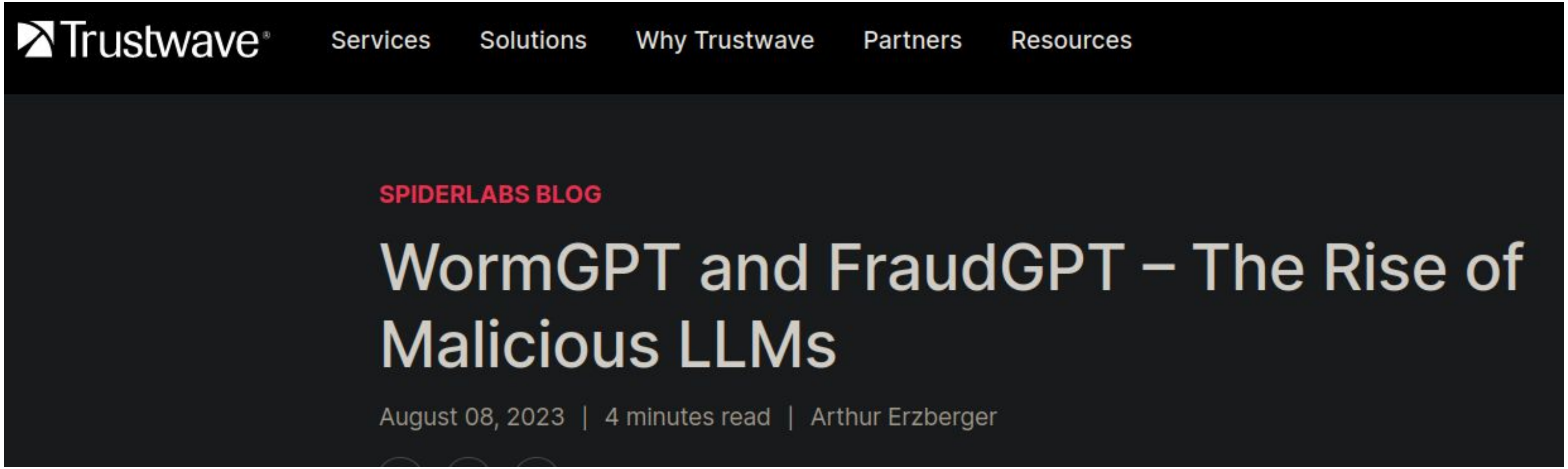
botacin@tamu.edu

Source:
https://marcusbotacin.github.io/publication/2023-05-01-paper-gpt-number-27

But this does not happen in reality, right?

Right?

FraudGPT: Malicious LLMs



SPIDERLABS BLOG

WormGPT and FraudGPT – The Rise of Malicious LLMs

August 08, 2023 | 4 minutes read | Arthur Erzberger

Source: https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/wormgpt-and-fraudgpt-the-rise-of-malicious-llms/

How to solve these new problems?

Questions?

A long road Ahead!