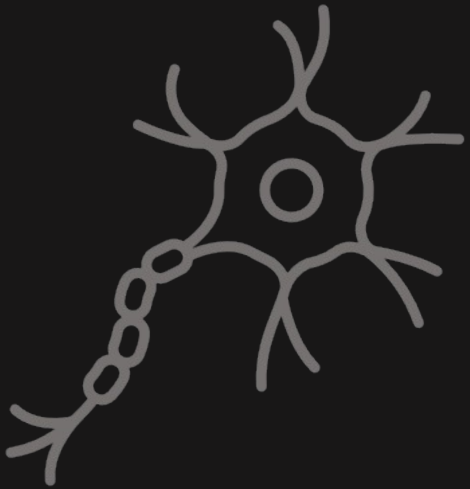
Machine Learning and the Optimization of Virtual Personae for Phishing, Mostly.

Will Pearce

@moo_hax



ML Security Researcher at
NVIDIA

ATML Red Team Lead at
MSFT

Senior Security Consultant at
Silent Break Security

Darkside Ops – Take it!

Motivations

Machine Learning requires security because we deploy it into production environments.

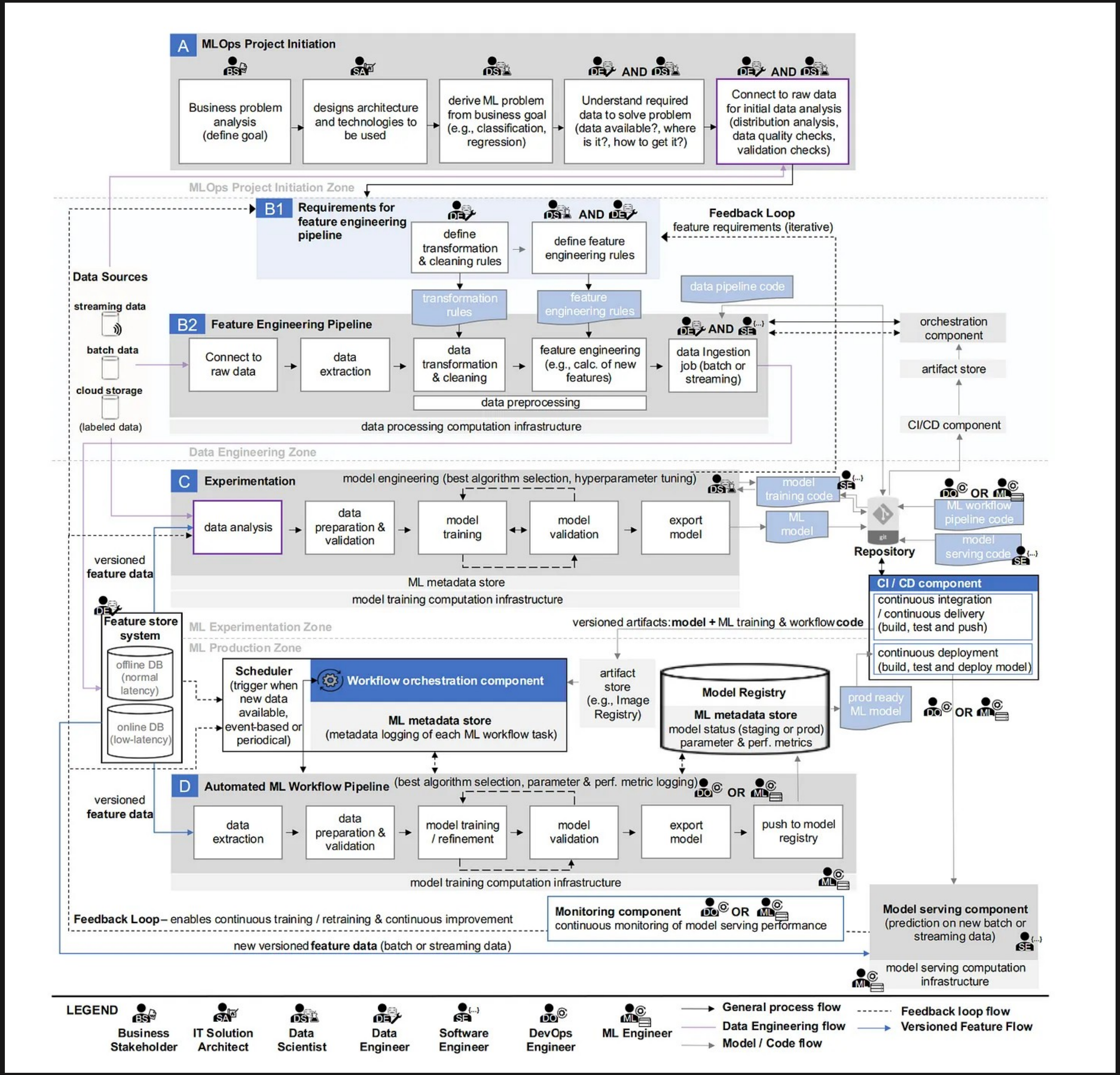
If Machine Learning only existed in research or academia, security would be something of a moot point.

Offensive ML is fun.

https://github.com/moohax

MLOps

- Data Collection
- Data Processing
- Model Training
- Model Evaluation
- Model Deployment
- System Monitoring

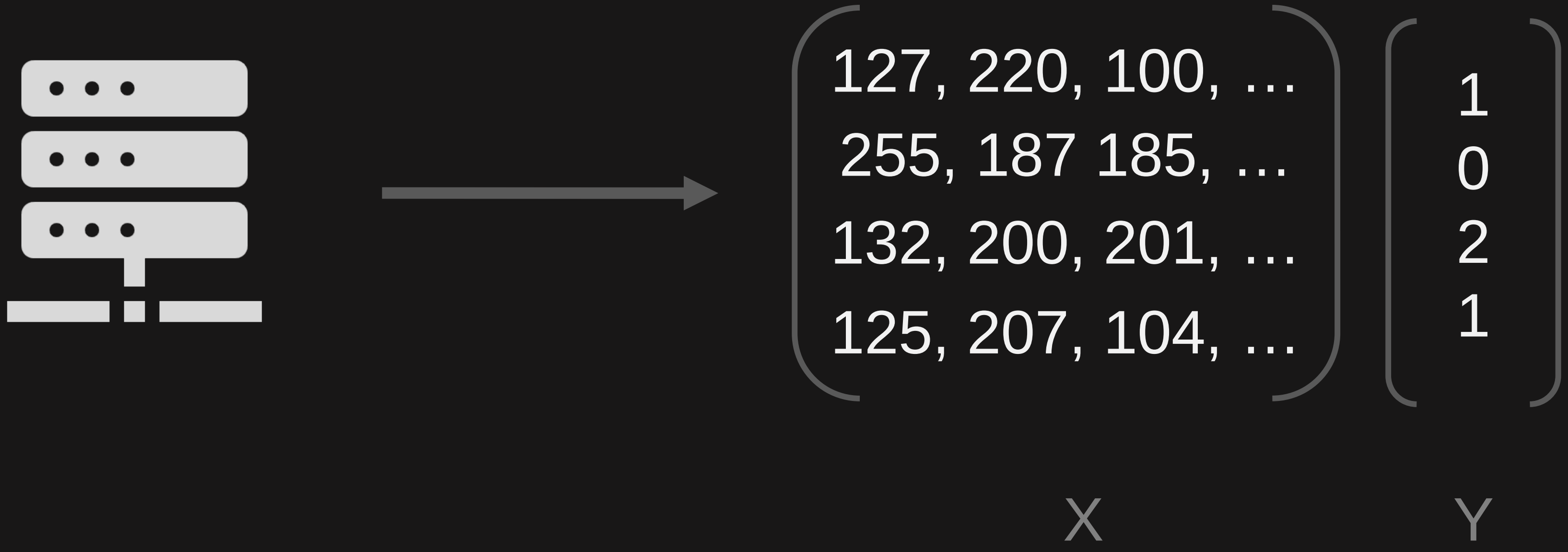


A MLOps Project Initiation
BS
SA
DS
DE AND DS
DE AND DS
Business problem analysis (define goal)
designs architecture and technologies to be used
derive ML problem from business goal (e.g., classification, regression)
Understand required data to solve problem (data available?, where is it?, how to get it?)
Connect to raw data for initial data analysis (distribution analysis, data quality checks, validation checks)
MLOps Project Initiation Zone
B1 Requirements for feature engineering pipeline
DE
DS AND DE
define transformation & cleaning rules
define feature engineering rules
Feedback Loop
feature requirements (iterative)
Data Sources
streaming data
batch data
cloud storage
(labeled data)
transformation rules
feature engineering rules
data pipeline code
B2 Feature Engineering Pipeline
DE AND SE
Connect to raw data
data extraction
data transformation & cleaning
feature engineering (e.g., calc. of new features)
data Ingestion job (batch or streaming)
data preprocessing
data processing computation infrastructure
orchestration component
artifact store
CI/CD component
Data Engineering Zone
C Experimentation
model engineering (best algorithm selection, hyperparameter tuning)
DS
SE
model training code
ML model
data analysis
data preparation & validation
model training
model validation
export model
DO OR ML
ML workflow pipeline code
model serving code
SE
Repository
ML metadata store
model training computation infrastructure
versioned feature data
DE
Feature store system
offline DB (normal latency)
online DB (low-latency)
CI / CD component
continuous integration / continuous delivery (build, test and push)
continuous deployment (build, test and deploy model)
ML Experimentation Zone
ML Production Zone
versioned artifacts: model + ML training & workflow code
Scheduler (trigger when new data available, event-based or periodical)
Workflow orchestration component
ML metadata store (metadata logging of each ML workflow task)
artifact store (e.g., Image Registry)
Model Registry
ML metadata store
model status (staging or prod)
parameter & perf. metrics
prod ready ML model
DO OR ML
D Automated ML Workflow Pipeline
(best algorithm selection, parameter & perf. metric logging)
DO OR ML
versioned feature data
data extraction
data preparation & validation
model training / refinement
model validation
export model
push to model registry
model training computation infrastructure
ML
Feedback Loop – enables continuous training / retraining & continuous improvement
Monitoring component
DO OR ML
continuous monitoring of model serving performance
Model serving component (prediction on new batch or streaming data)
SE
new versioned feature data (batch or streaming data)
model serving computation infrastructure
ML
LEGEND
Business Stakeholder
IT Solution Architect
Data Scientist
Data Engineer
Software Engineer
DevOps Engineer
ML Engineer
General process flow
Data Engineering flow
Model / Code flow
Feedback loop flow
Versioned Feature Flow

https://arxiv.org/ftp/arxiv/papers/2205/2205.02302.pdf

Data Collection

Data Processing



$$\begin{pmatrix} 127,\ 220,\ 100,\ \dots \\ 255,\ 187\ 185,\ \dots \\ 132,\ 200,\ 201,\ \dots \\ 125,\ 207,\ 104,\ \dots \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

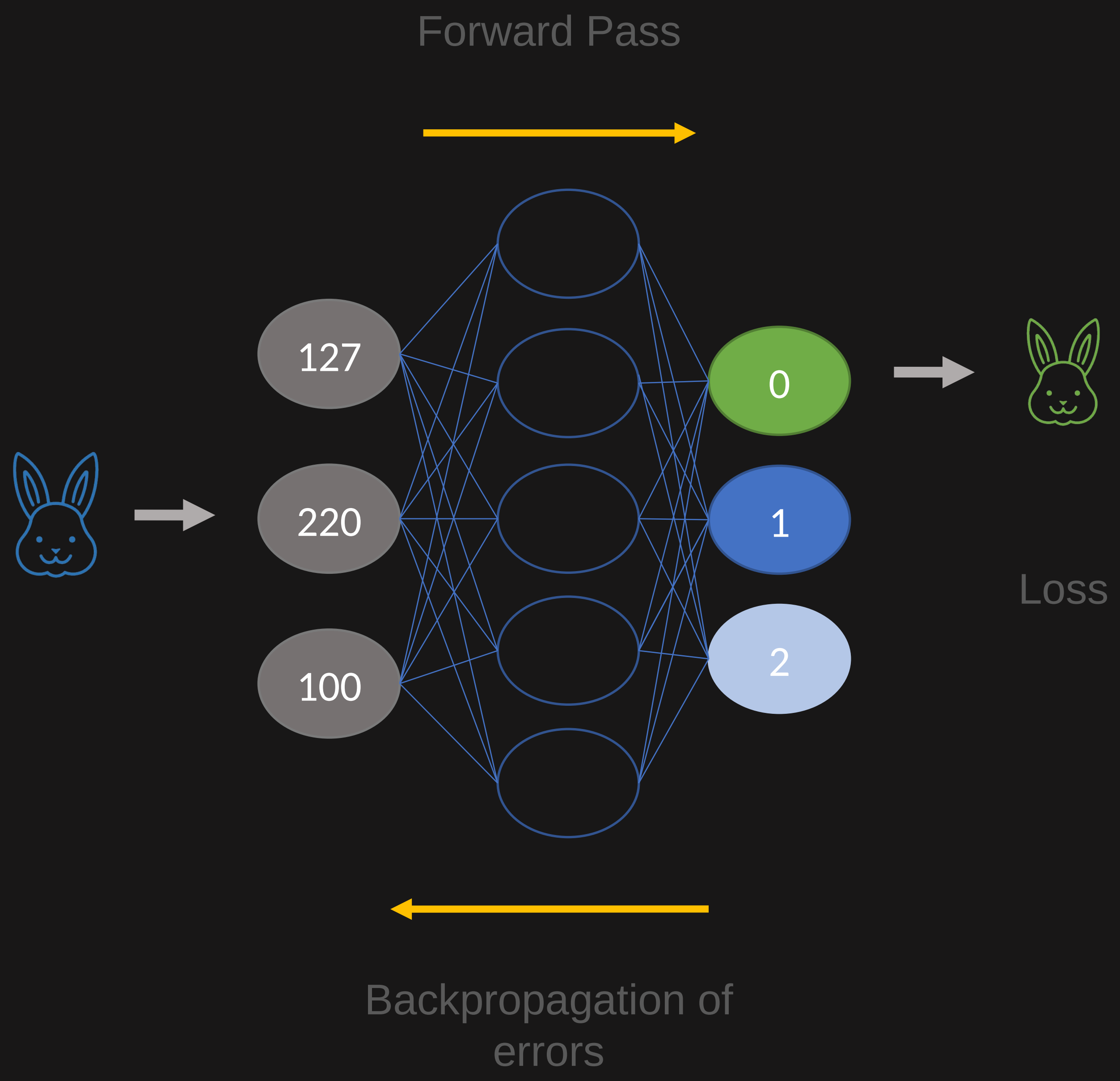
(Labels make it a supervised problem)

process.py

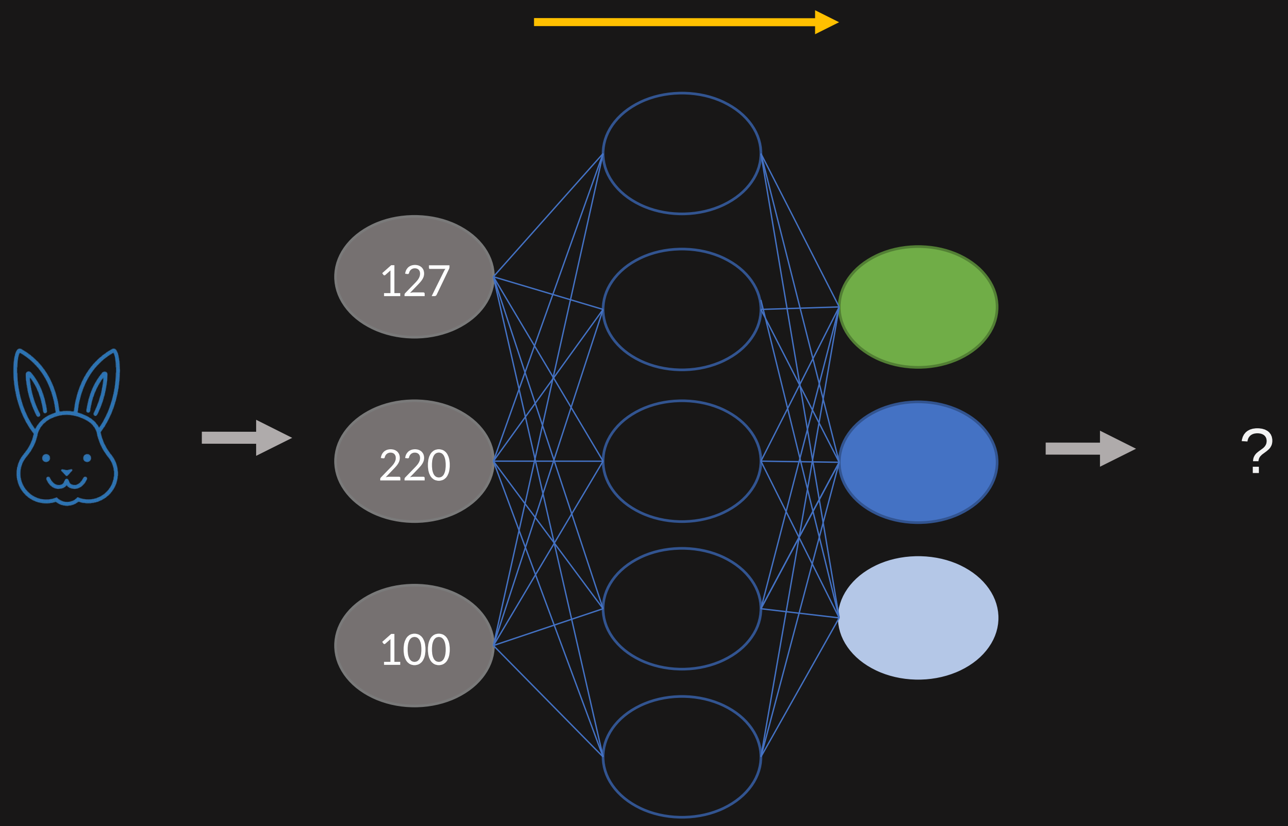
Model Training

$$\begin{pmatrix} 127,\ 220,\ 100,\ \dots \\ 255,\ 187\ 185,\ \dots \\ 132,\ 200,\ 201,\ \dots \\ 125,\ 207,\ 104,\ \dots \end{pmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

X Y

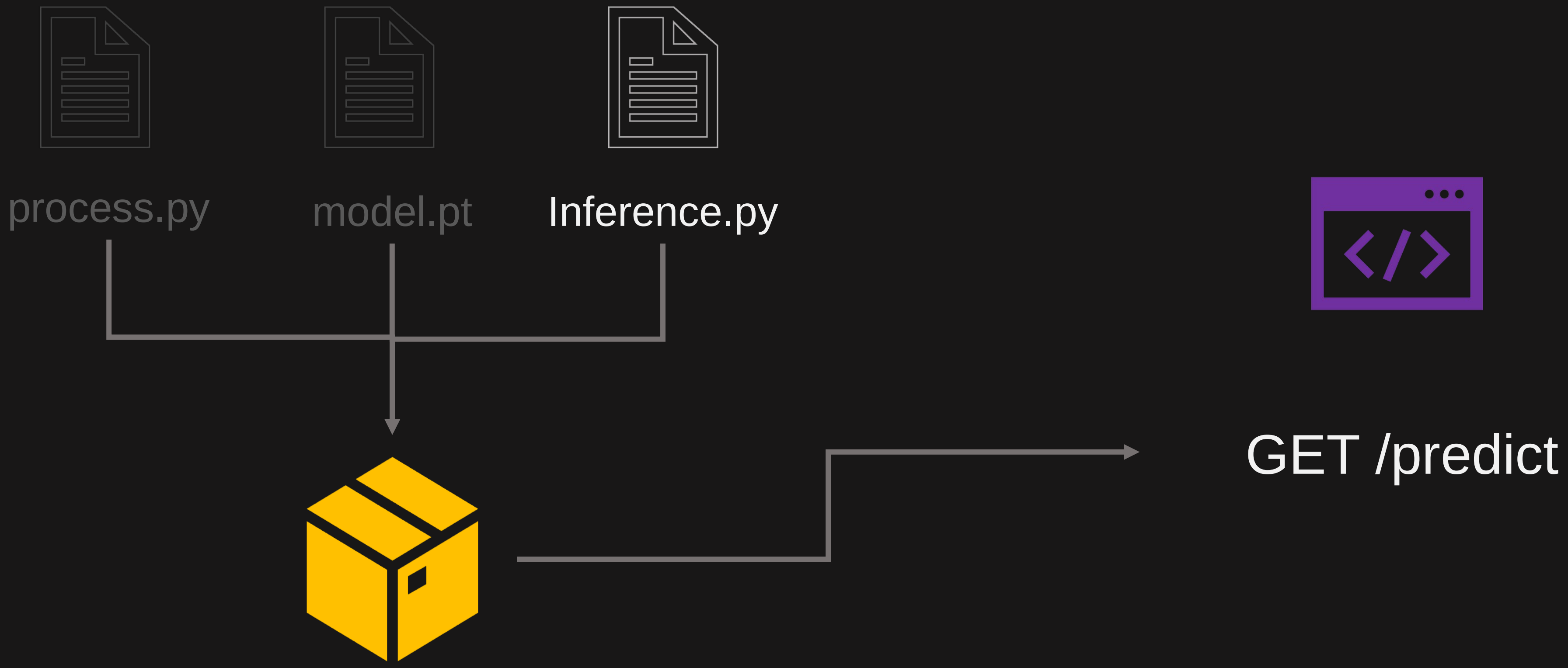


Model Evaluation



model.pt

Model Deployment



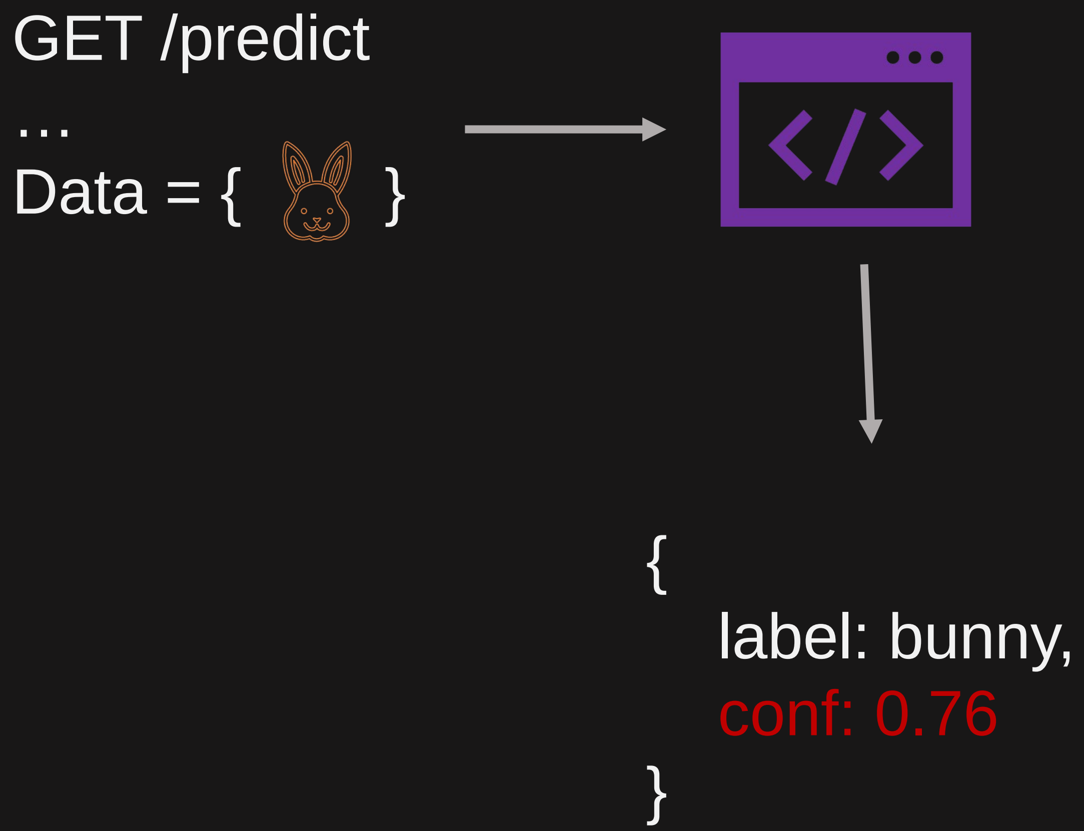
process.py

model.pt

Inference.py

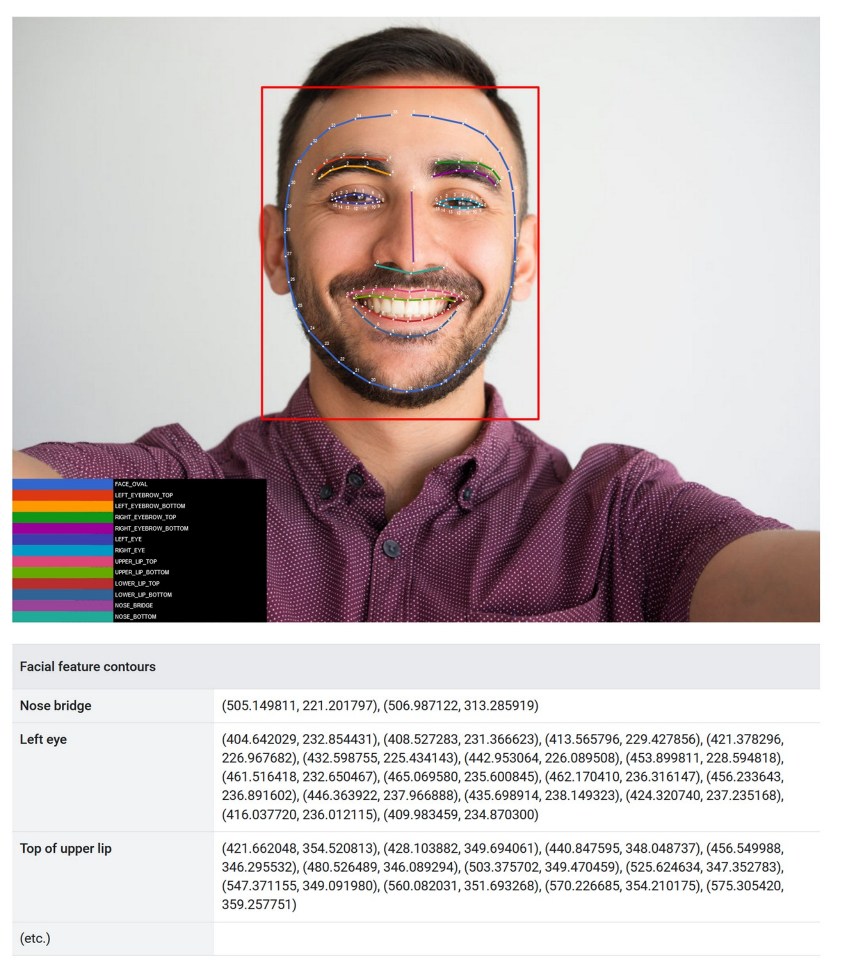
GET /predict

System Monitoring



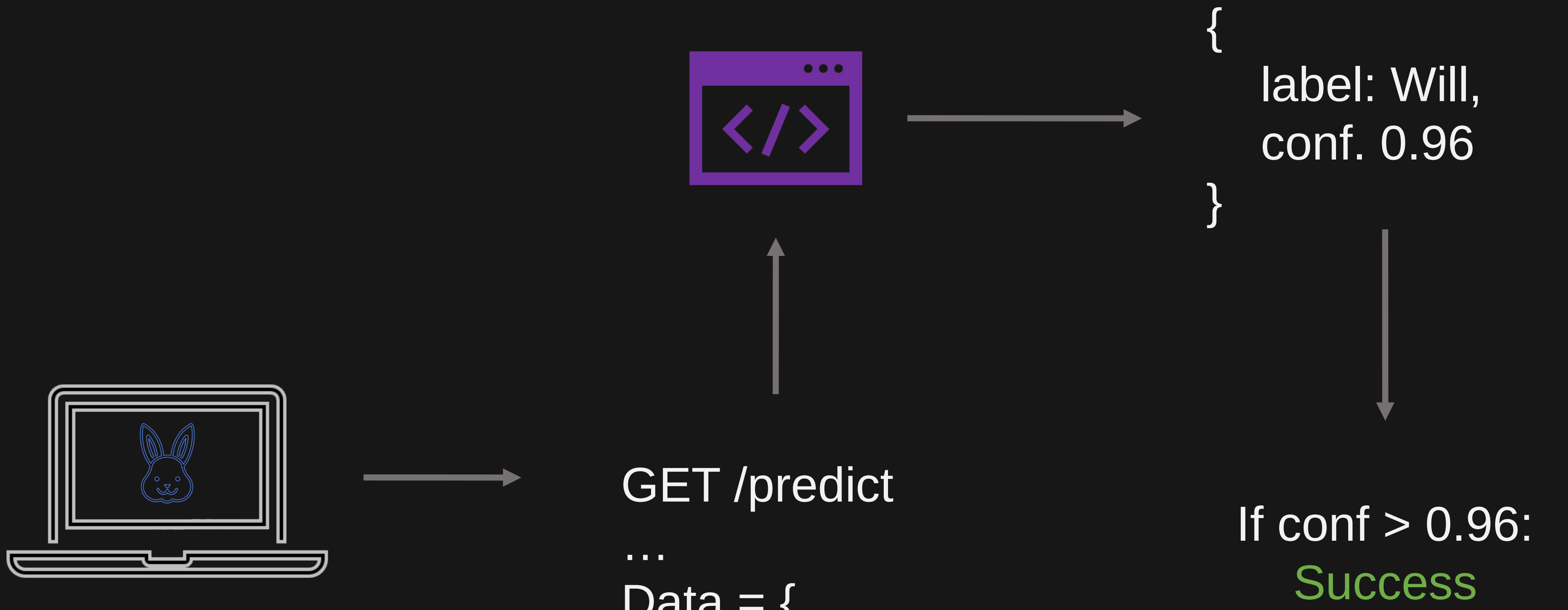
GET /predict
…
Data = { }

{
label: bunny,
conf: 0.76
}



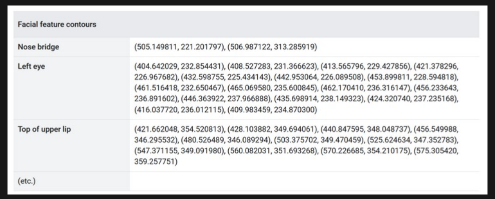
Facial feature contours	
Nose bridge	(505.149811, 221.201797), (506.987122, 313.285919)
Left eye	(404.642029, 232.854431), (408.527283, 231.366623), (413.565796, 229.427856), (421.378296, 226.967682), (432.598755, 225.434143), (442.953064, 226.089508), (453.899811, 228.594818), (461.516418, 232.650467), (465.069580, 235.600845), (462.170410, 236.316147), (456.233643, 236.891602), (446.363922, 237.966888), (435.698914, 238.149323), (424.320740, 237.235168), (416.037720, 236.012115), (409.983459, 234.870300)
Top of upper lip	(421.662048, 354.520813), (428.103882, 349.694061), (440.847595, 348.048737), (456.549988, 346.295532), (480.526489, 346.089294), (503.375702, 349.470459), (525.624634, 347.352783), (547.371155, 349.091980), (560.082031, 351.693268), (570.226685, 354.210175), (575.305420, 359.257751)
(etc.)	

https://blog.tensorflow.org/2019/01/tensorflow-lite-now-faster-with-mobile.html



{
label: Will,
conf. 0.96
}

GET /predict
…
Data = {

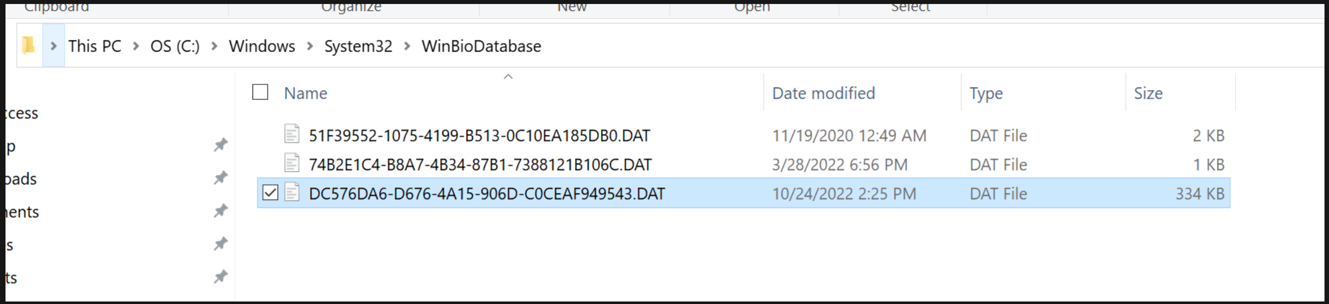


Facial feature contours	
Nose bridge	(505.149811, 221.201797), (506.987122, 313.285919)
Left eye	(404.642029, 232.854431), (408.527283, 231.366623), (413.565796, 229.427856), (421.378296, 226.967682), (432.598755, 225.434143), (442.953064, 226.089508), (453.899811, 228.594818), (461.516418, 232.650467), (465.069580, 235.600845), (462.170410, 236.316147), (456.233643, 236.891602), (446.363922, 237.966888), (435.698914, 238.149323), (424.320740, 237.235168), (416.037720, 236.012115), (409.983459, 234.870300)
Top of upper lip	(421.662048, 354.520813), (428.103882, 349.694061), (440.847595, 348.048737), (456.549988, 346.295532), (480.526489, 346.089294), (503.375702, 349.470459), (525.624634, 347.352783), (547.371155, 349.091980), (560.082031, 351.693268), (570.226685, 354.210175), (575.305420, 359.257751)
(etc.)	

}

If conf > 0.96:
Success

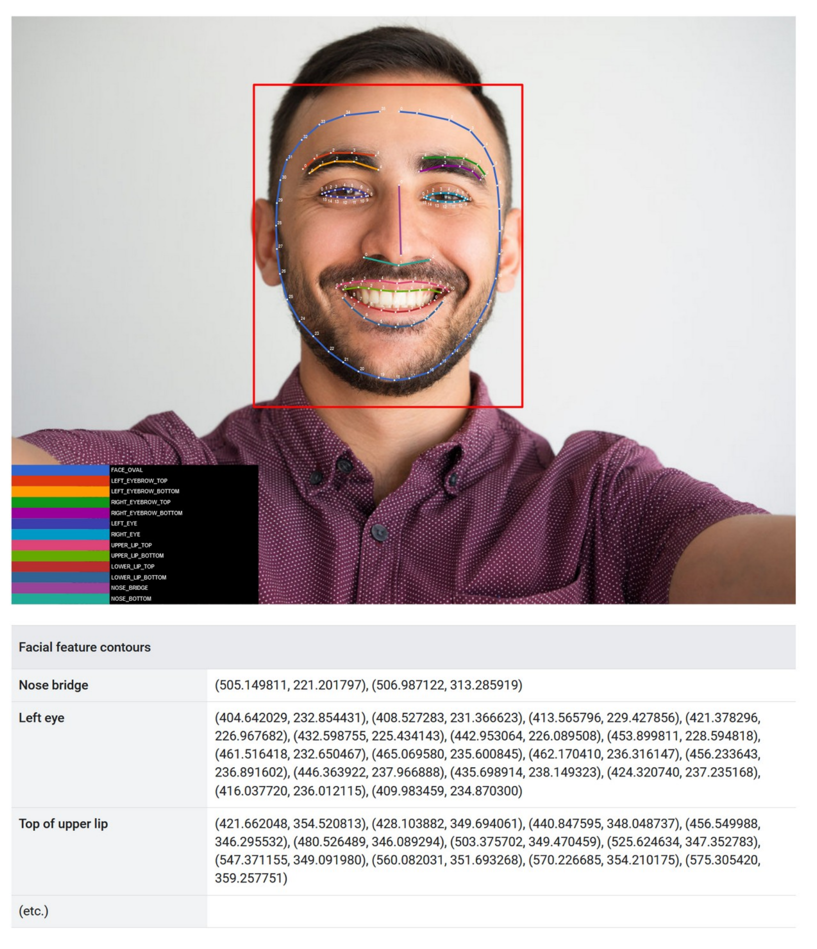
Your identity is on disk.



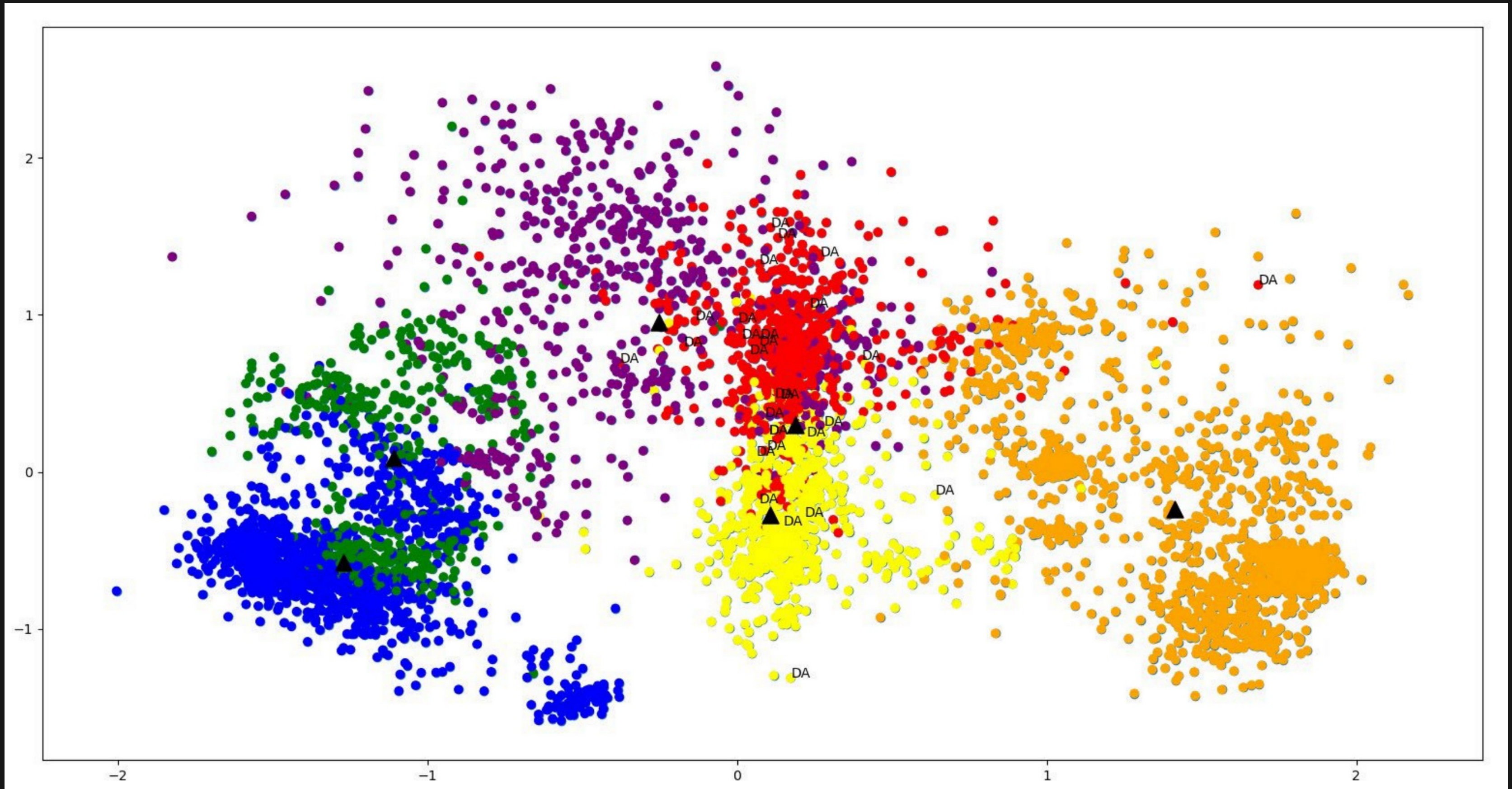
This PC › OS (C:) › Windows › System32 › WinBioDatabase

☐ Name	Date modified	Type	Size
51F39552-1075-4199-B513-0C10EA185DB0.DAT	11/19/2020 12:49 AM	DAT File	2 KB
74B2E1C4-B8A7-4B34-87B1-7388121B106C.DAT	3/28/2022 6:56 PM	DAT File	1 KB
☑ DC576DA6-D676-4A15-906D-C0CEAF949543.DAT	10/24/2022 2:25 PM	DAT File	334 KB

if distance < 0.01:
Success



Facial feature contours	
Nose bridge	(505.149811, 221.201797), (506.987122, 313.285919)
Left eye	(404.642029, 232.854431), (408.527283, 231.366623), (413.565796, 229.427856), (421.378296, 226.967682), (432.598755, 225.434143), (442.953064, 226.089508), (453.899811, 228.594818), (461.516418, 232.650467), (465.069580, 235.600845), (462.170410, 236.316147), (456.233643, 236.891602), (446.363922, 237.966888), (435.698914, 238.149323), (424.320740, 237.235168), (416.037720, 236.012115), (409.983459, 234.870300)
Top of upper lip	(421.662048, 354.520813), (428.103882, 349.694061), (440.847595, 348.048737), (456.549988, 346.295532), (480.526489, 346.089294), (503.375702, 349.470459), (525.624634, 347.352783), (547.371155, 349.091980), (560.082031, 351.693268), (570.226685, 354.210175), (575.305420, 359.257751)
(etc.)	



KNN using Doc2Vec with usernames and group membership from AD

https://twitter.com/moo_hax/status/1129064078054309888

Algorithms are empty, models are not.
(Live, Laugh, Love)

$$\begin{pmatrix} 127, 220, 100, \ldots \\ 255, 187\ 185, \ldots \\ 132, 200, 201, \ldots \end{pmatrix}$$

What happens to 2FA when something you are becomes something you have? (Picture, video, etc)

Identity

Social Media

Write a long job description for a sale executive making $400,000 per year with benefits

https://beta.openai.com

The Sales Executive is responsible for achieving assigned sales targets and executing sales strategies for a defined territory in the United States. The Sales Executive will be the primary point of contact for customers in assigned territory and will be responsible for prospecting, building relationships and closing deals. The Sales Executive will work closely with other members of the sales team, as well as with marketing, product management and other functions within the company to ensure that assigned objectives are met.

The Sales Executive must have a minimum of 5 years of sales experience, with a proven track record of success in achieving sales targets. The Sales Executive must have excellent communication and presentation skills and be able to effectively articulate the company’s value proposition to customers. The Sales Executive must also be highly organized and detail-oriented, with the ability to manage a large number of customer relationships simultaneously.

The Sales Executive will be based in the company’s offices in New York, NY and will report to the Vice President of Sales. The position offers a competitive base salary plus commission, with an expected total compensation of $400,000 per year. The company also offers a comprehensive benefits package, including medical, dental and vision coverage, 401(k) and stock options.

https://thispersondoesnotexist.com/

The Sales Executive is responsible for achieving assigned sales targets and executing sales strategies for a defined territory in the United States. The Sales Executive will be the primary point of contact for customers in assigned territory and will be responsible for prospecting, building relationships and closing deals. The Sales Executive will work closely with other members of the sales team, as well as with marketing, product management and other functions within the company to ensure that assigned objectives are met.

The Sales Executive must have a minimum of 5 years of sales experience, with a proven track record of success in achieving sales targets. The Sales Executive must have excellent communication and presentation skills and be able to effectively articulate the company’s value proposition to customers. The Sales Executive must also be highly organized and detail-oriented, with the ability to manage a large number of customer relationships simultaneously.

The Sales Executive will be based in the company’s offices in New York, NY and will report to the Vice President of Sales. The position offers a competitive base salary plus commission, with an expected total compensation of $400,000 per year. The company also offers a comprehensive benefits package, including medical, dental and vision coverage, 401(k) and stock options.

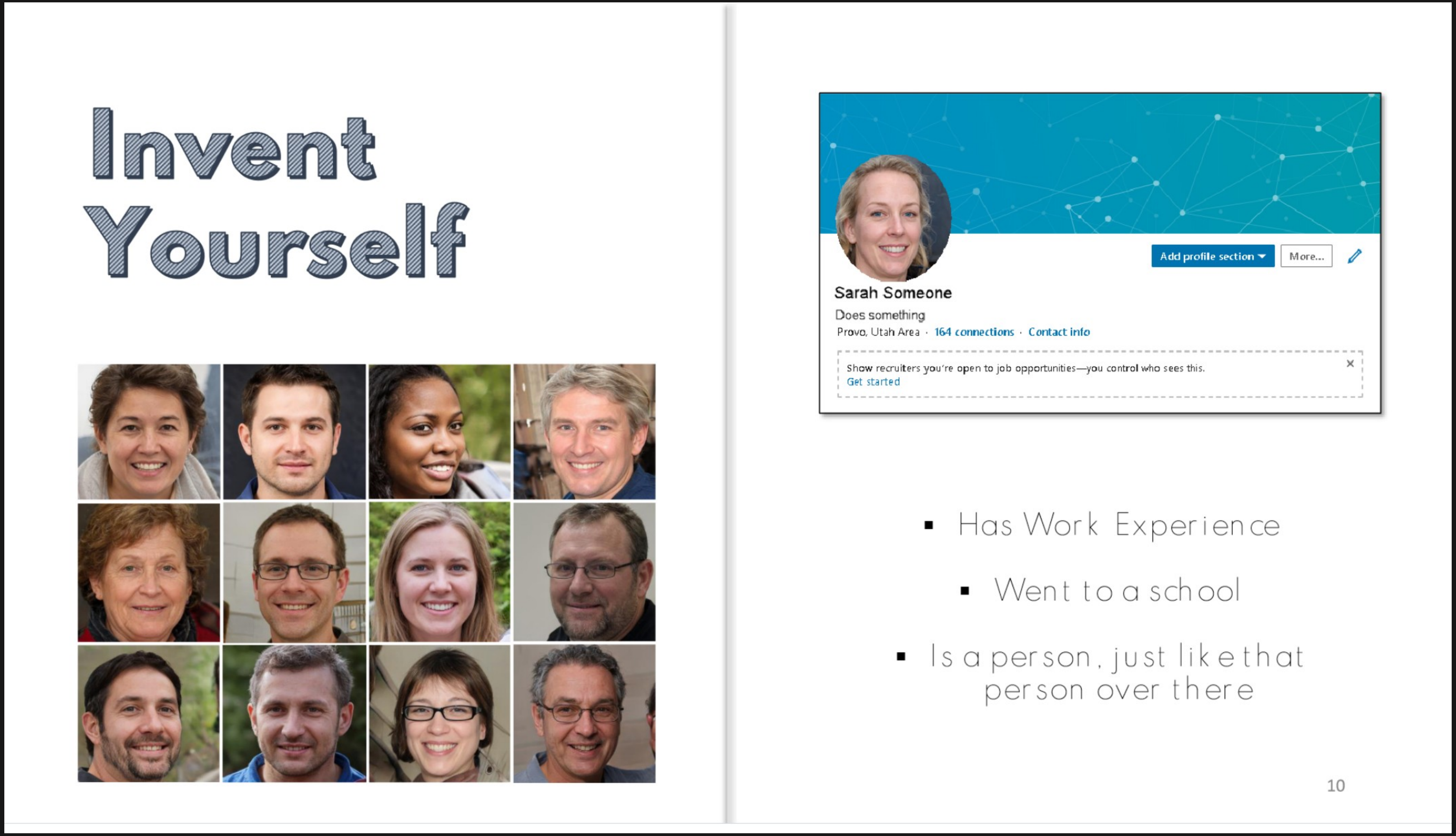
Write a long job description for a cat that has an important job as a banking executive making 6000 tuna a year, plus naps.

https://thiscatdoesnotexist.com/

This is a very important position for a cat with a lot of experience in the banking industry. The cat will be responsible for making 6000 tuna a year, plus naps.

This is a highly responsible position that requires a great deal of experience and knowledge in the banking industry. The cat will be required to work closely with the other executives in the company to ensure that the company's operations are running smoothly.

The cat will also be responsible for making sure that the company's finances are in order. This is a highly demanding position that requires a great deal of dedication and hard work. However, it is also a very rewarding position that comes with a lot of perks, such as a large salary, a great benefits package, and a lot of tuna.



Invent Yourself

Sarah Someone

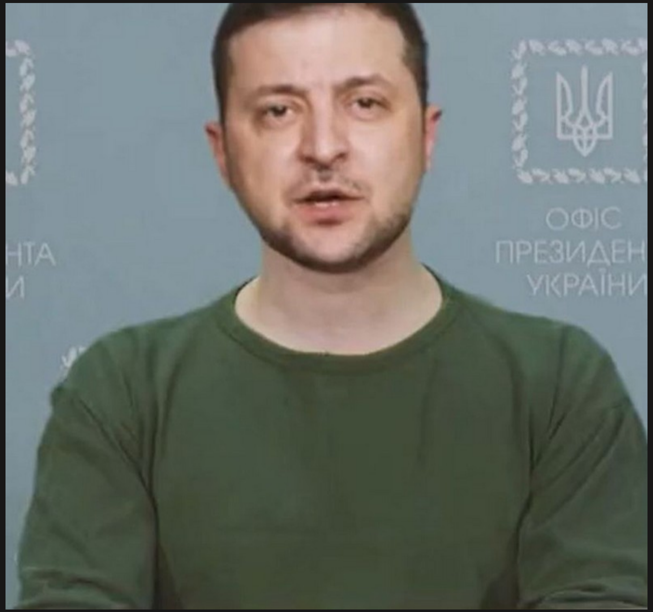
Does something

Provo, Utah Area · 164 connections · Contact info

Show recruiters you're open to job opportunities—you control who sees this.
Get started

- Has Work Experience
- Went to a school
- Is a person, just like that person over there

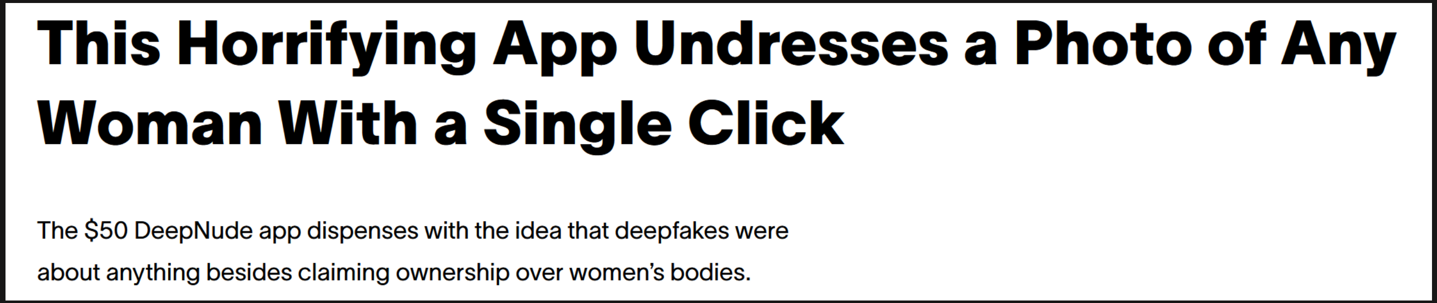
https://github.com/moohax/Talks/blob/master/slides/bsides_slc_20.pdf



https://github.com/deepfakes/faceswap

https://github.com/iperov/DeepFaceLab

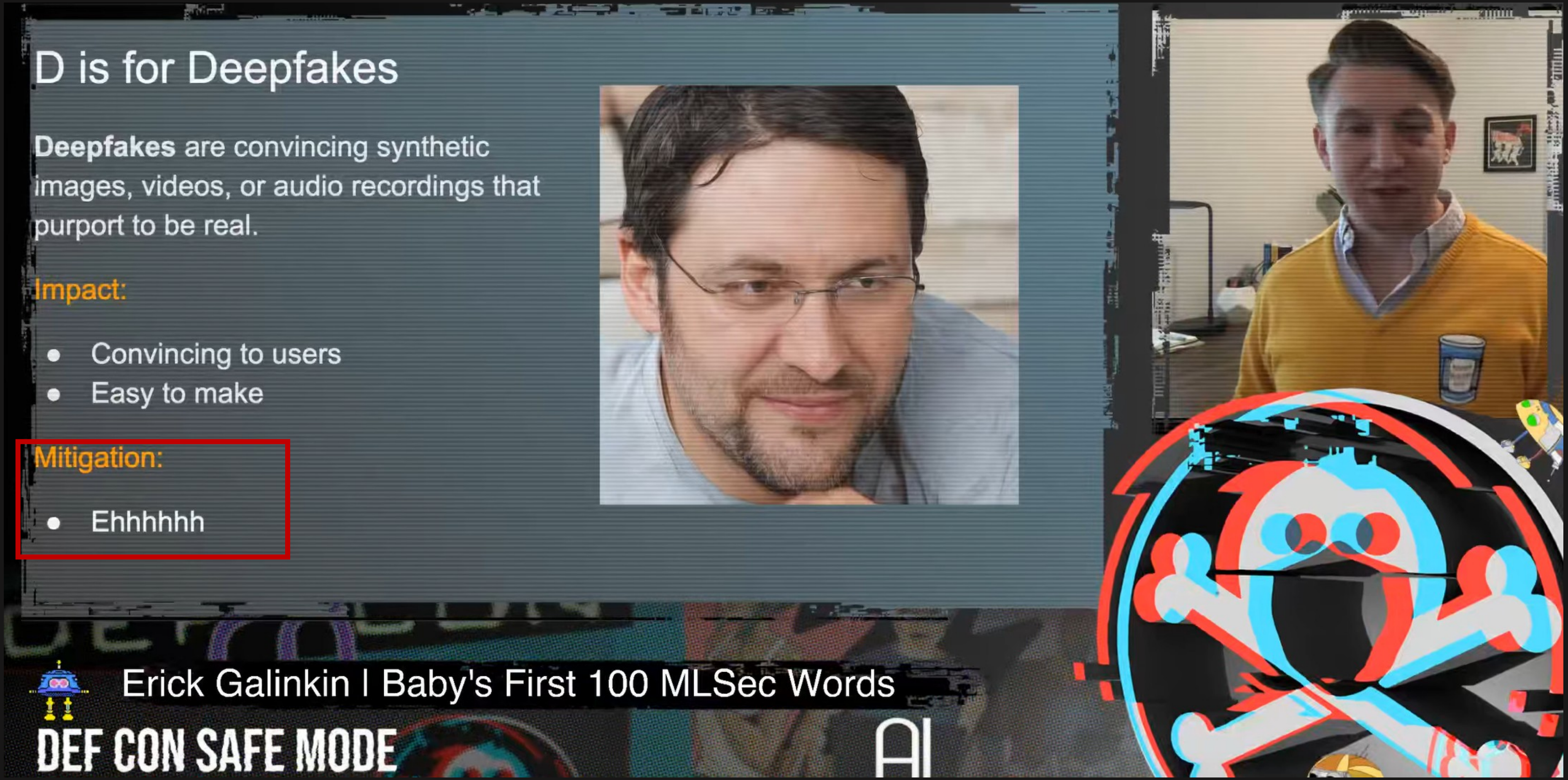
https://www.vice.com/en/article/qjb7b7/ethical-deepfakes-deep-tom-cruise-ai-generated-porn



This Horrifying App Undresses a Photo of Any Woman With a Single Click

The $50 DeepNude app dispenses with the idea that deepfakes were about anything besides claiming ownership over women's bodies.

https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman



D is for Deepfakes

Deepfakes are convincing synthetic images, videos, or audio recordings that purport to be real.

Impact:

- Convincing to users
- Easy to make

Mitigation:

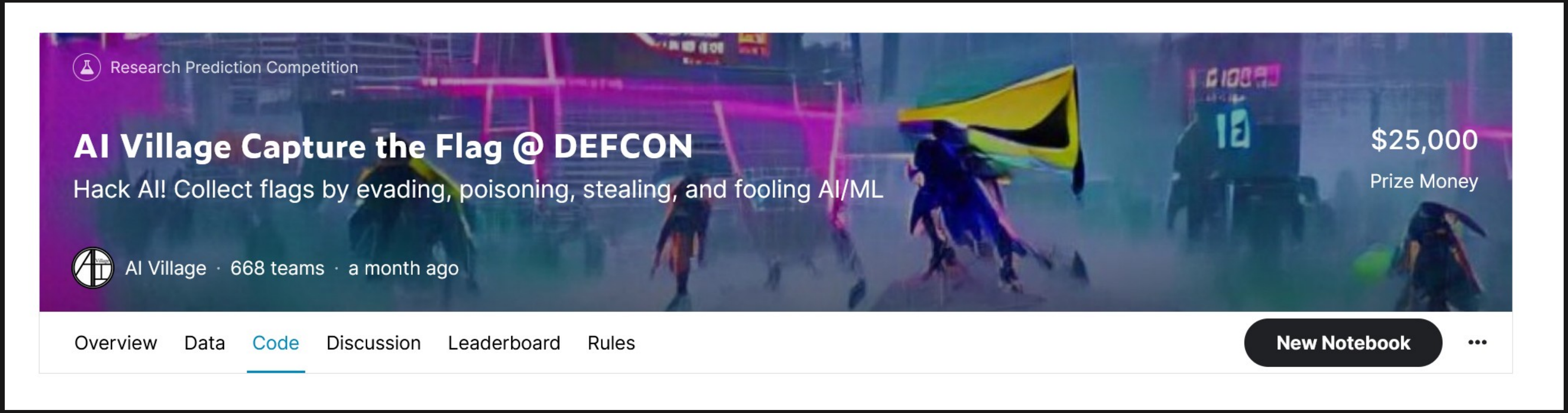
- Ehhhhhh

Erick Galinkin | Baby's First 100 MLSec Words

DEF CON SAFE MODE

@ErickGanklin

https://www.youtube.com/watch?v=Xo2KZCbJWCg&t=159s



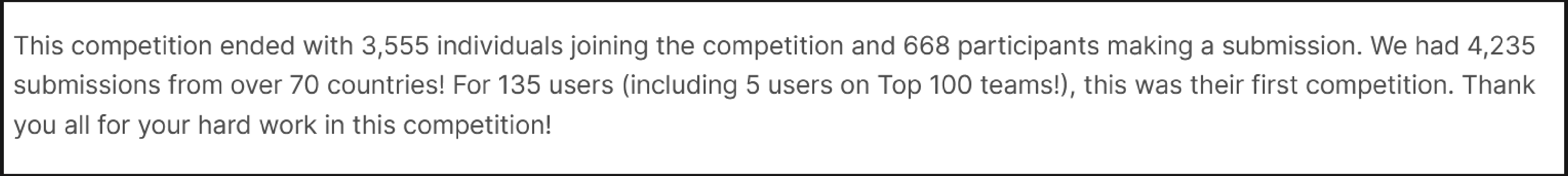
Research Prediction Competition

AI Village Capture the Flag @ DEFCON

Hack AI! Collect flags by evading, poisoning, stealing, and fooling AI/ML

AI Village · 668 teams · a month ago

$25,000
Prize Money



This competition ended with 3,555 individuals joining the competition and 668 participants making a submission. We had 4,235 submissions from over 70 countries! For 135 users (including 5 users on Top 100 teams!), this was their first competition. Thank you all for your hard work in this competition!

Kaggle is a great platform for ML infrastructure and learning

Protections?

Synthetic Media

- New protocols?
- New responsibilities for media companies?
- Some sort of media signing? New/old PKI infrastructure?
- Blockchain?

Bio Auth in General

- Liveness detection
- Texture Analysis
- Infrared
- Glare, ambient light
- Windows Event Monitoring
  - Microsoft-Windows-HelloForBusiness/Operational

Conclusion

Think of your identity as a composite, as a distribution, not as something in isolation.

In a digital world you are always represented by data that is apart from you.

accurate or not, marketing departments know this.

Pay attention to new tooling for phishing – be on the lookout for your first synthetic phish

Thank You!

Email: wpearce@nvidia.com

Twitter: @moo_hax

GitHub: https://github.com/moohax

Resources

Nvidia GTC Sep '22

Disinformation at Scale BH '21

AI For Phishing BH Asia '21

Adversarial Robustness Toolbox

ML for Red Teams

Voice Recognition bypass

Defcon AIV CTF

Darkside Ops

CAMLIS

GPT-3

Model Zoo

Windows Hello for Business

Keras/Tensorflow/PyTorch

Finetuning Stable Diffusion

Luigi (mlops)

Eyeballer