

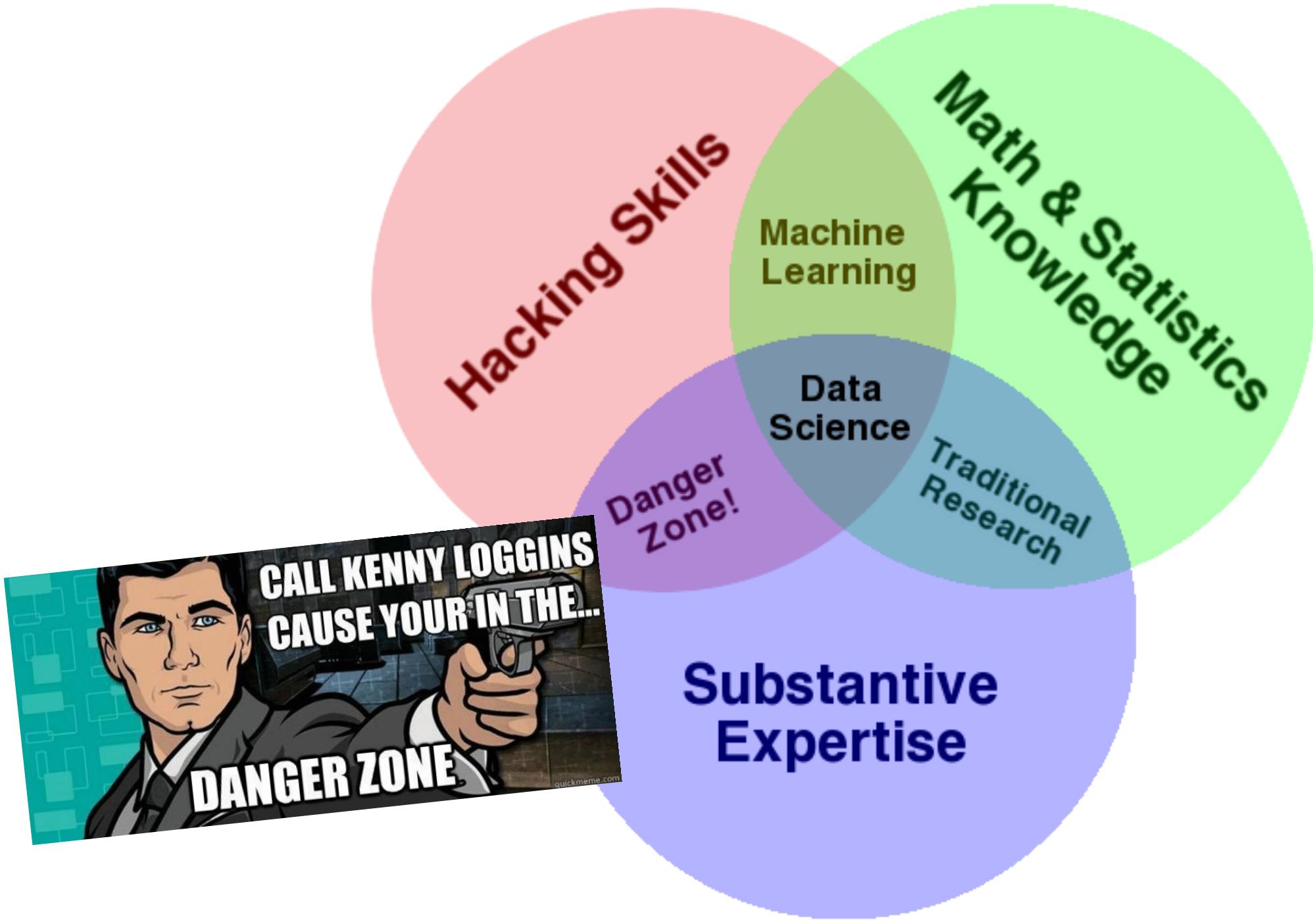
# Machine Learning

PHYS 453 – Spring 2024

Dr. Daugherty



ABILENE  
CHRISTIAN  
UNIVERSITY

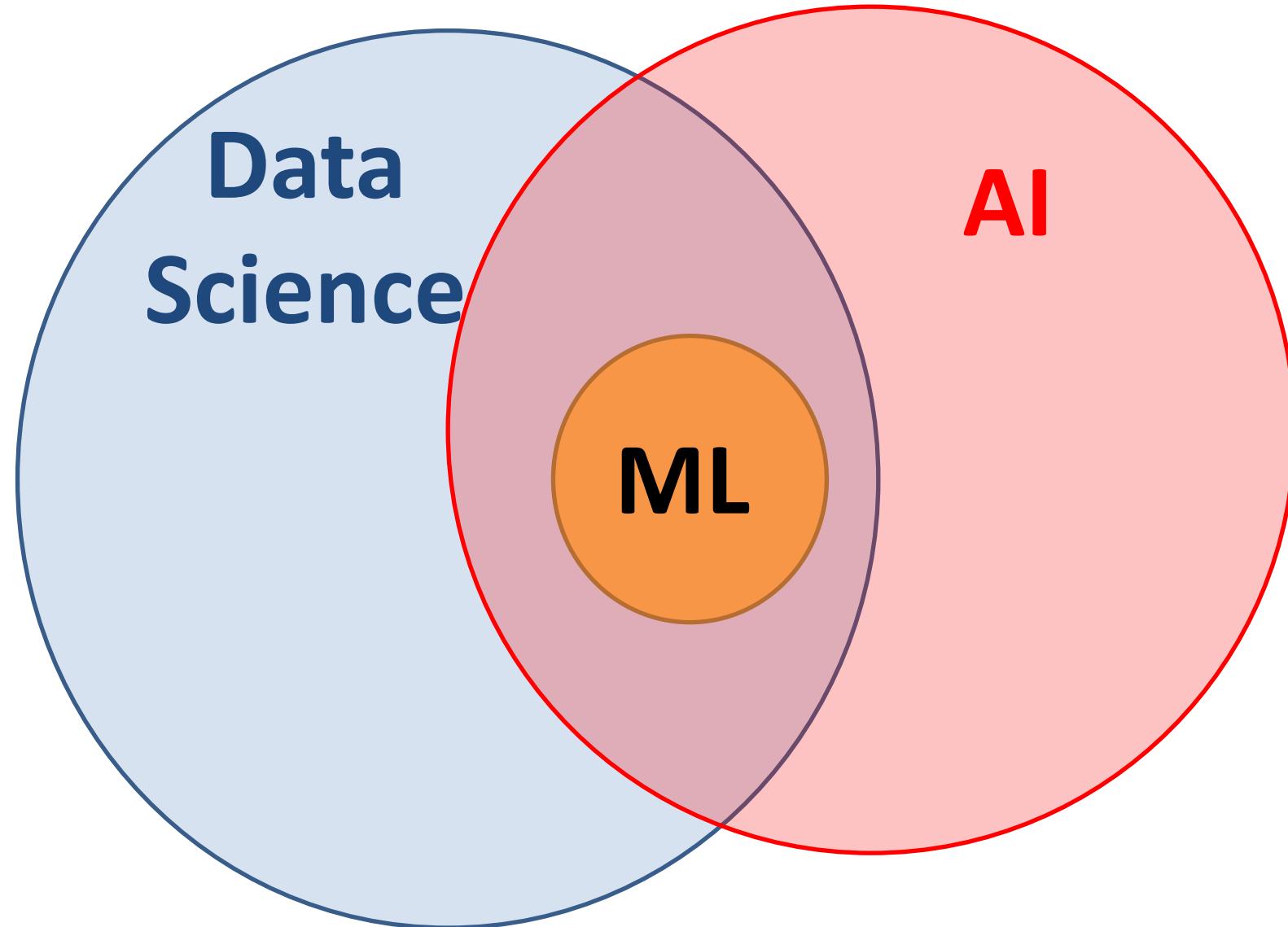


“Data science, as it’s practiced, is a blend of Red-Bull-fueled hacking and espresso-inspired statistics.

But data science is not merely hacking, because when hackers finish debugging their Bash one-liners and Pig scripts, few care about non-Euclidean distance metrics.

And data science is not merely statistics, because when statisticians finish theorizing the perfect model, few could read a tab delimited file into R if their job depended on it.

Data science is the civil engineering of data. Its acolytes possess a practical knowledge of tools & materials, coupled with a theoretical understanding of what’s possible.”



\* not to scale

# About Me

## Bio:

- Grew up in Oklahoma City
- 2002 ACU: Physics and CS
- 2008 PhD in Nuclear Physics from UT
- Married in 2000, two kids

## Research:

- Particle and Nuclear Physics
- Atom smashers
- Radiation Detectors
- Other interests: artificial intelligence, cosmology, amateur theology

## Fun Facts:

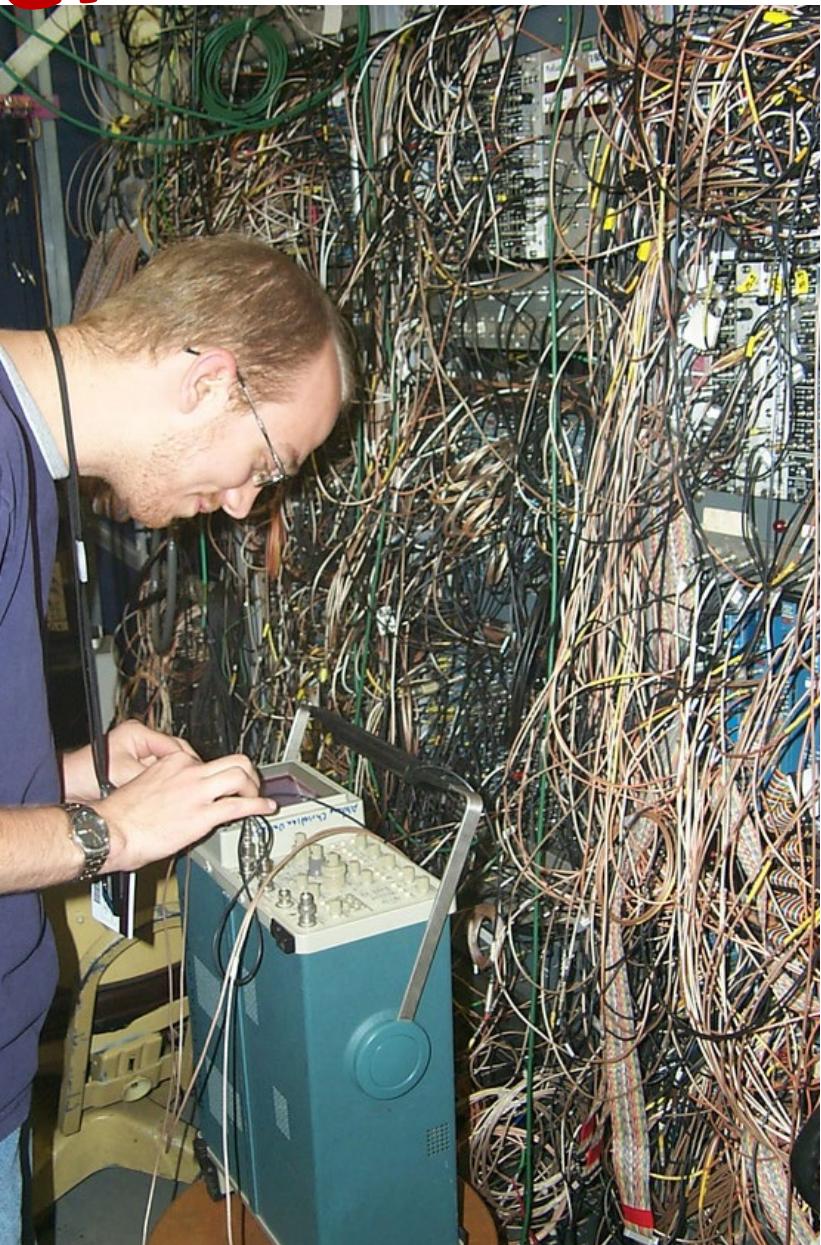
- Have been paid as a professional model
- Once burned off an eyebrow in class
- Plays drums in Abilene's best cover band

## Hobbies:

- Eating
- Putting things in a laser
- Building dangerous things



# First Experiment



# PhD Research

Two-Particle Correlations in Ultra Relativistic Heavy  
Ion Collisions

by

Michael Scott Daugherty, B.S.; B.S.

## Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

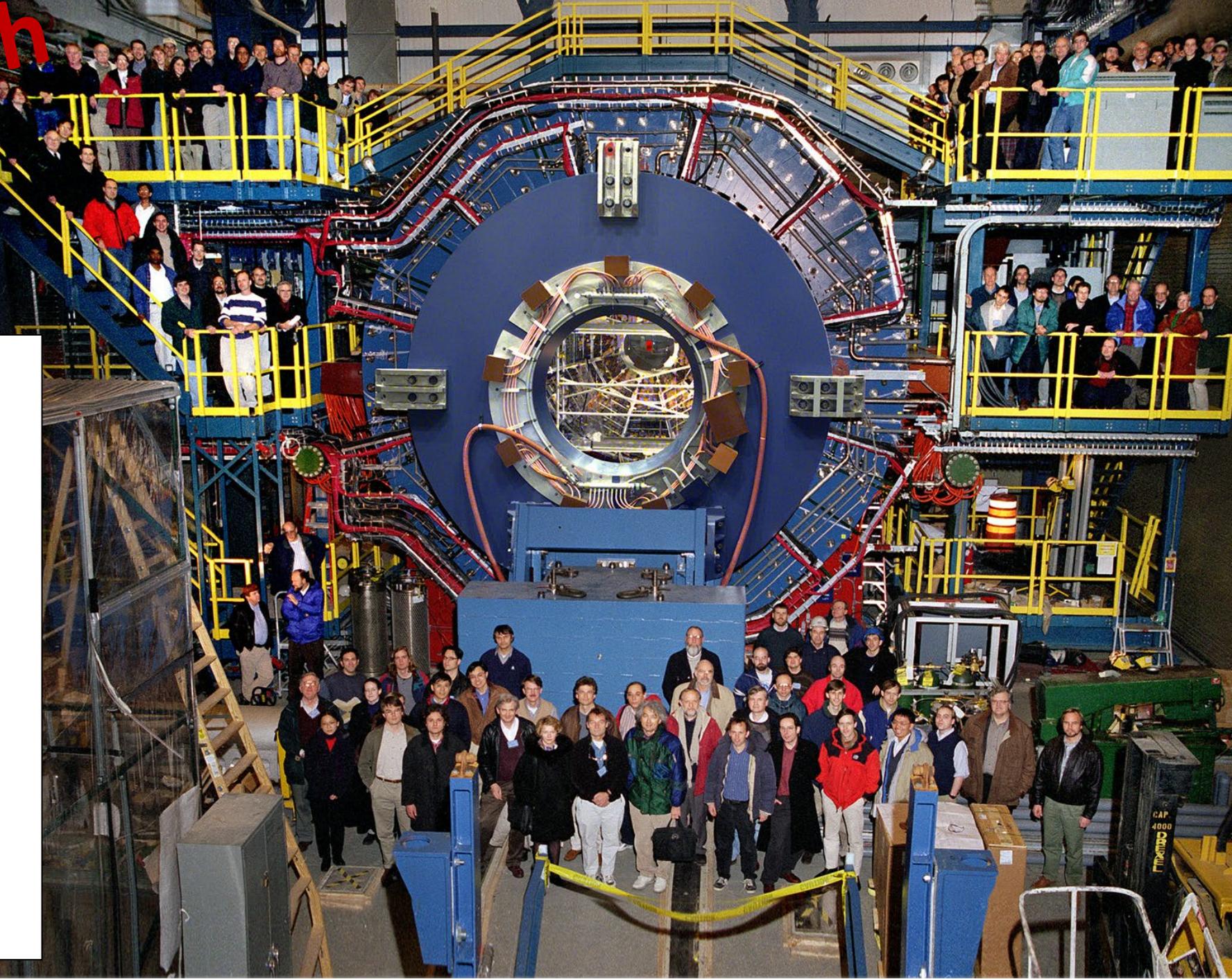
of the Requirements

for the Degree of

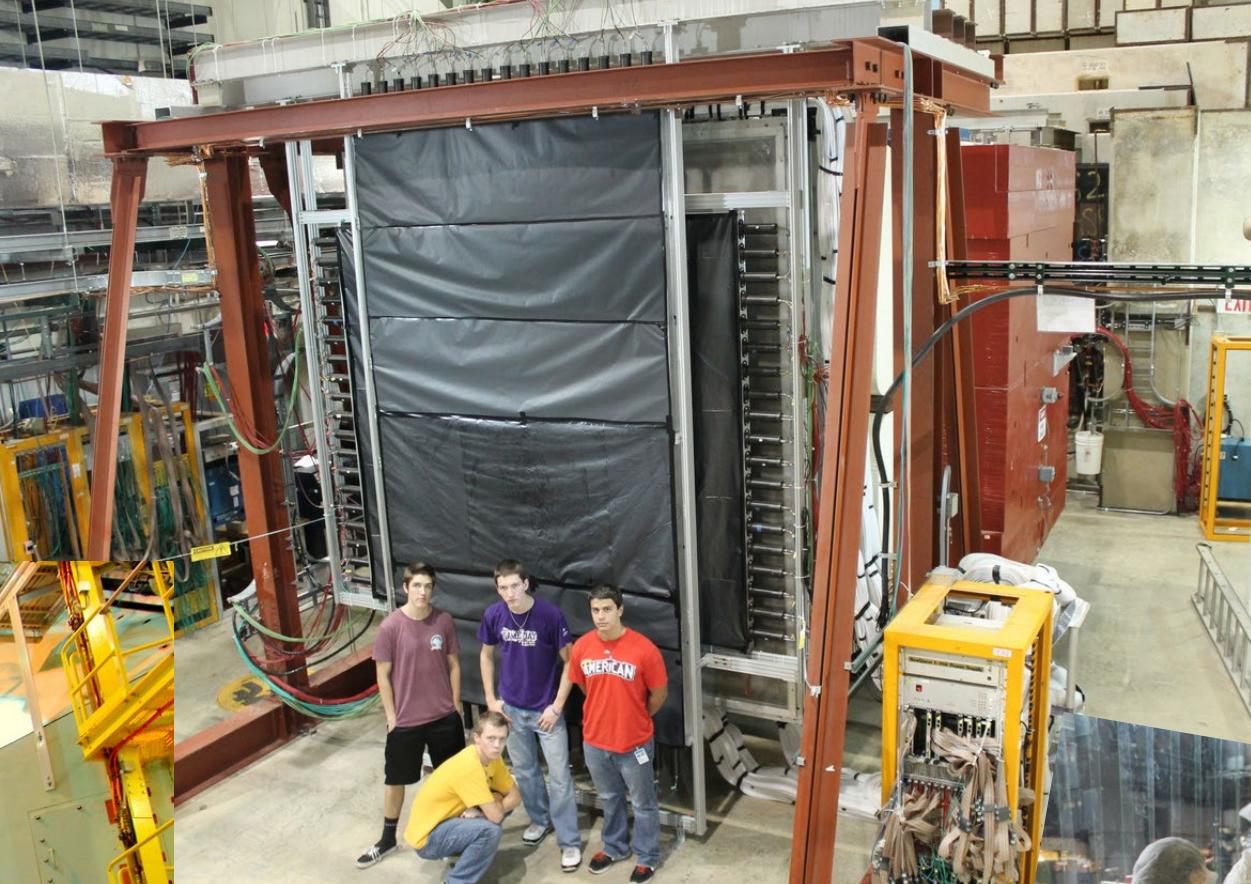
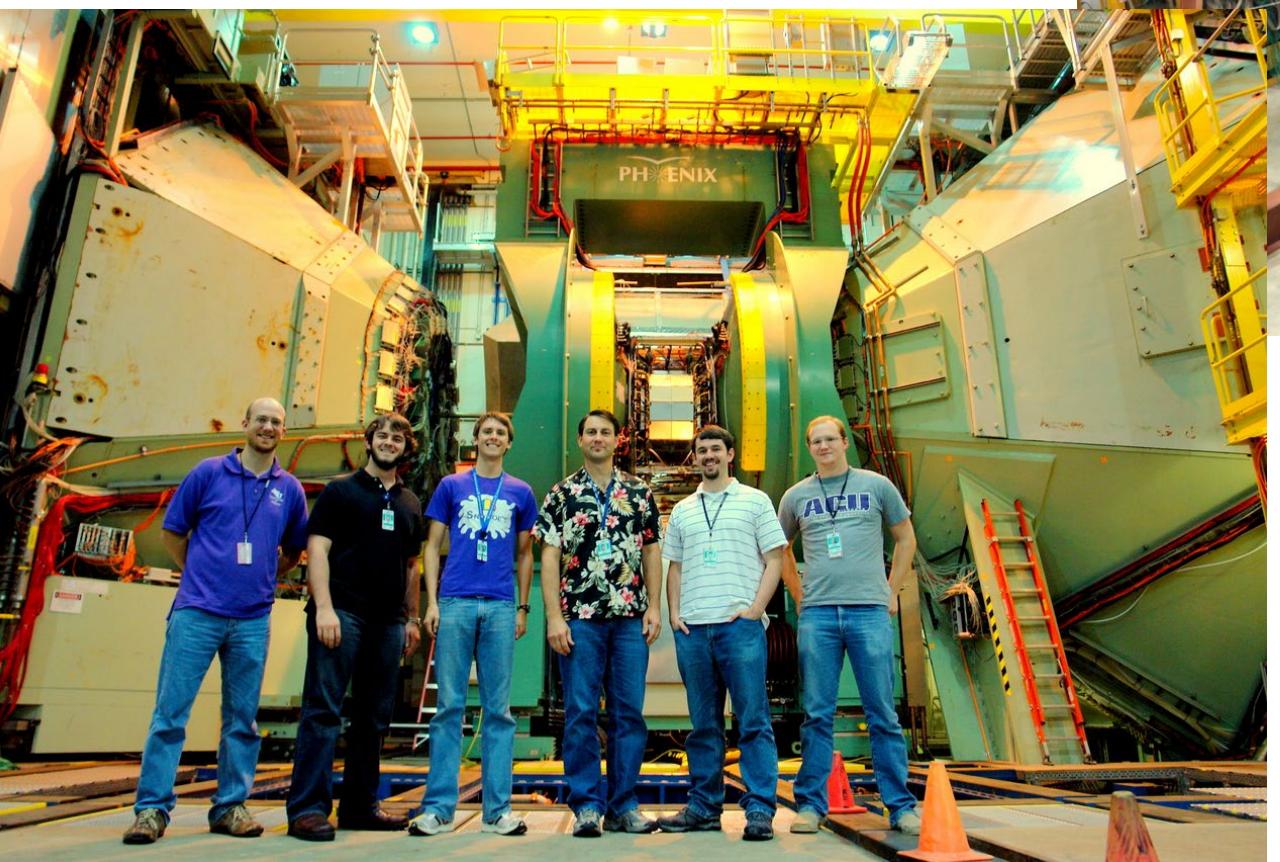
Doctor of Philosophy

The University of Texas at Austin

August 2008



# ACU Research





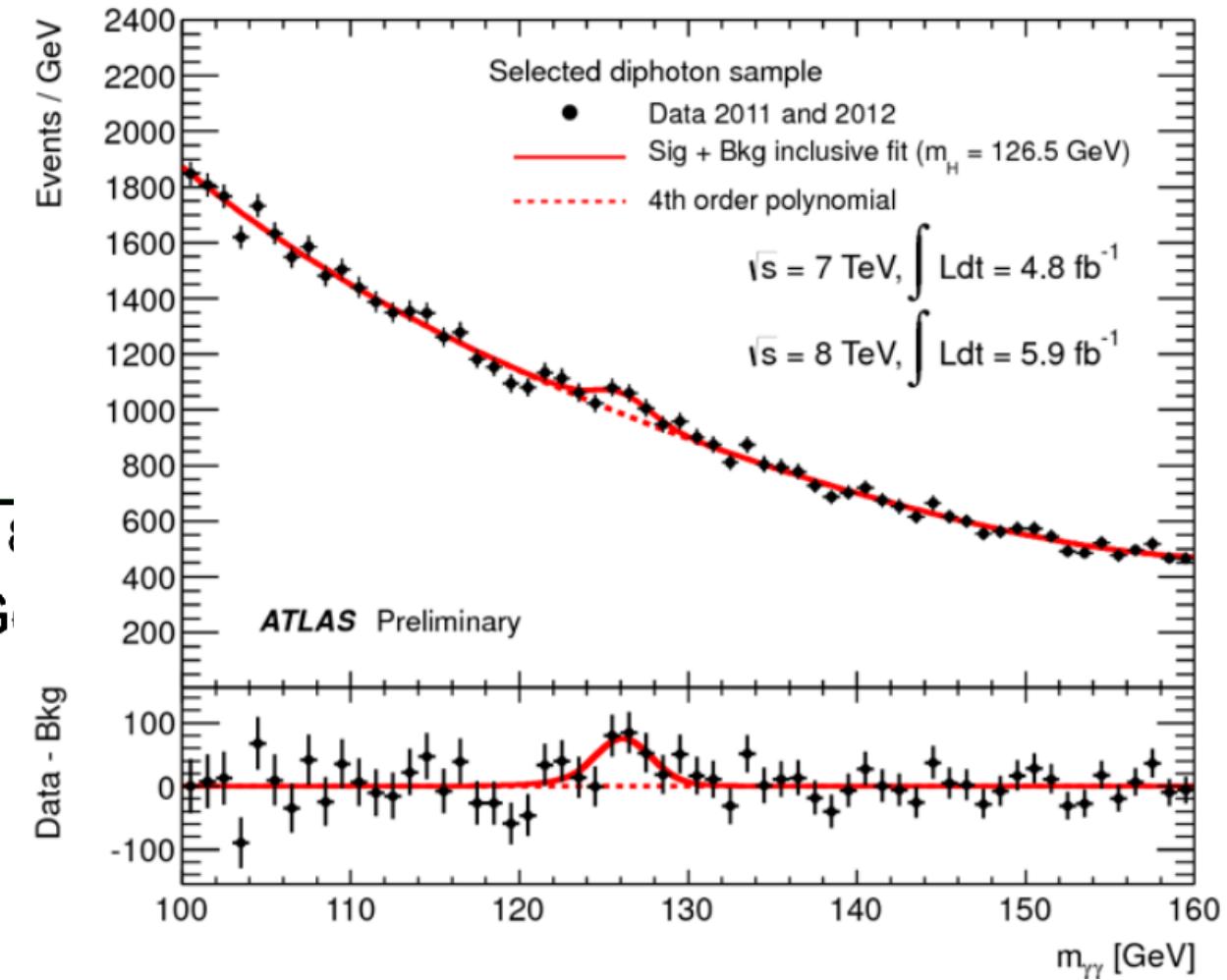
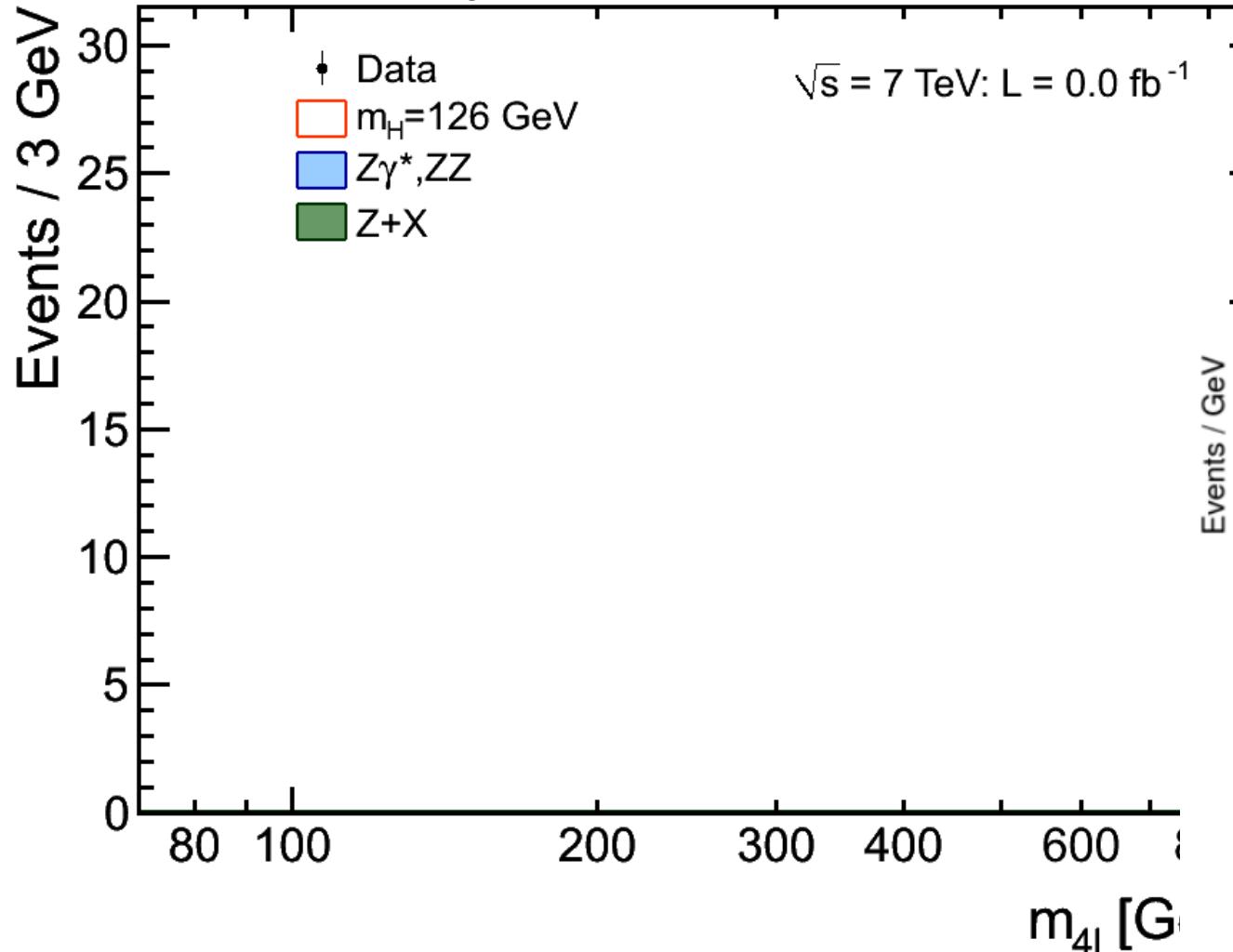
# Before “Big Data” was a thing, there was high-energy physics...

[ ] 16 Mar 2010

## A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and $b$ -quark Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV

T. Aaltonen,<sup>24</sup> J. Adelman,<sup>14</sup> B. Álvarez González<sup>v</sup>,<sup>12</sup> S. Amerio<sup>dd</sup>,<sup>44</sup> D. Amidei,<sup>35</sup> A. Anastassov,<sup>39</sup> A. Annovi,<sup>20</sup> J. Antos,<sup>15</sup> G. Apollinari,<sup>18</sup> A. Apresyan,<sup>49</sup> T. Arisawa,<sup>58</sup> A. Artikov,<sup>16</sup> J. Asaadi,<sup>54</sup> W. Ashmanskas,<sup>18</sup> A. Attal,<sup>4</sup> A. Aurisano,<sup>54</sup> F. Azfar,<sup>43</sup> W. Badgett,<sup>18</sup> A. Barbaro-Galtieri,<sup>29</sup> V.E. Barnes,<sup>49</sup> B.A. Barnett,<sup>26</sup> P. Barria<sup>ff</sup>,<sup>47</sup> P. Bartos,<sup>15</sup> G. Bauer,<sup>33</sup> P.-H. Beauchemin,<sup>34</sup> F. Bedeschi,<sup>47</sup> D. Beecher,<sup>31</sup> S. Behari,<sup>26</sup> G. Bellettini<sup>ee</sup>,<sup>47</sup> J. Bellinger,<sup>60</sup> D. Benjamin,<sup>17</sup> A. Beretvas,<sup>18</sup> A. Bhatti,<sup>51</sup> M. Binkley,<sup>18</sup> D. Bisello<sup>dd</sup>,<sup>44</sup> I. Bizjak<sup>jj</sup>,<sup>31</sup> R.E. Blair,<sup>2</sup> C. Blocker,<sup>7</sup> B. Blumenfeld,<sup>26</sup> A. Bocci,<sup>17</sup> A. Bodek,<sup>50</sup> V. Boisvert,<sup>50</sup> D. Bortoletto,<sup>49</sup> J. Boudreau,<sup>48</sup> A. Boveia,<sup>11</sup> B. Brau<sup>a</sup>,<sup>11</sup> A. Bridgeman,<sup>25</sup> L. Brigliadori<sup>cc</sup>,<sup>6</sup> C. Bromberg,<sup>36</sup> E. Brubaker,<sup>14</sup> J. Budagov,<sup>16</sup> H.S. Budd,<sup>50</sup> S. Budd,<sup>25</sup> K. Burkett,<sup>18</sup> G. Busetto<sup>dd</sup>,<sup>44</sup> P. Bussey,<sup>22</sup> A. Buzatu,<sup>34</sup> K. L. Byrum,<sup>2</sup> S. Cabrera<sup>x</sup>,<sup>17</sup> C. Calancha,<sup>32</sup> S. Camarda,<sup>4</sup> M. Campanelli,<sup>31</sup> M. Campbell,<sup>35</sup> F. Canelli<sup>14</sup>,<sup>18</sup> A. Canepa,<sup>46</sup> B. Carls,<sup>25</sup> D. Carlsmith,<sup>60</sup> R. Carosi,<sup>47</sup> S. Carrillo<sup>n</sup>,<sup>19</sup> S. Carron,<sup>18</sup> B. Casal,<sup>12</sup> M. Casarsa,<sup>18</sup> A. Castro<sup>cc</sup>,<sup>6</sup> P. Catastini<sup>ff</sup>,<sup>47</sup> D. Cauz,<sup>55</sup> V. Cavaliere<sup>ff</sup>,<sup>47</sup> M. Cavalli-Sforza,<sup>4</sup> A. Cerri,<sup>29</sup> L. Cerrito<sup>q</sup>,<sup>31</sup> S.H. Chang,<sup>28</sup> Y.C. Chen,<sup>1</sup> M. Chertok,<sup>8</sup> G. Chiarelli,<sup>47</sup> G. Chlachidze,<sup>18</sup> F. Chlebana,<sup>18</sup> K. Cho,<sup>28</sup> D. Chokheli,<sup>16</sup> J.P. Chou,<sup>23</sup> K. Chung<sup>o</sup>,<sup>18</sup> W.H. Chung,<sup>60</sup> Y.S. Chung,<sup>50</sup> T. Chwalek,<sup>27</sup> C.I. Ciobanu,<sup>45</sup> M.A. Ciocci<sup>ff</sup>,<sup>47</sup> A. Clark,<sup>21</sup> D. Clark,<sup>7</sup> G. Compostella,<sup>44</sup> M.E. Convery,<sup>18</sup> J. Conway,<sup>8</sup> M. Corbo,<sup>45</sup> M. Cordelli,<sup>20</sup> C.A. Cox,<sup>8</sup> D.J. Cox,<sup>8</sup> F. Crescioli<sup>ee</sup>,<sup>47</sup> C. Cuenca Almenar,<sup>61</sup> J. Cuevas<sup>v</sup>,<sup>12</sup> R. Culbertson,<sup>18</sup> J.C. Cully,<sup>35</sup> D. Dagenhart,<sup>18</sup> M. Datta,<sup>18</sup> T. Davies,<sup>22</sup> P. de Barbaro,<sup>50</sup> S. De Cecco,<sup>52</sup> A. Deisher,<sup>29</sup> G. De Lorenzo,<sup>4</sup> M. Dell'Orso<sup>ee</sup>,<sup>47</sup> C. Deluca,<sup>4</sup> L. Demortier,<sup>51</sup> J. Deng<sup>f</sup>,<sup>17</sup> M. Deninno,<sup>6</sup> M. d'Errico<sup>dd</sup>,<sup>44</sup> A. Di Canto<sup>ee</sup>,<sup>47</sup> G.P. di Giovanni,<sup>45</sup> B. Di Ruzza,<sup>47</sup> J.R. Dittmann,<sup>5</sup> M. D'Onofrio,<sup>4</sup> S. Donati<sup>ee</sup>,<sup>47</sup> P. Dong,<sup>18</sup> T. Dorigo,<sup>44</sup> S. Dube,<sup>53</sup> K. Ebina,<sup>58</sup> A. Elagin,<sup>54</sup> R. Erbacher,<sup>8</sup> D. Errede,<sup>25</sup> S. Errede,<sup>25</sup> N. Ershaidat<sup>bb</sup>,<sup>45</sup> R. Eusebi,<sup>54</sup> H.C. Fang,<sup>29</sup> S. Farrington,<sup>43</sup> W.T. Fedorko,<sup>14</sup> R.G. Feild,<sup>61</sup>

CMS Preliminary



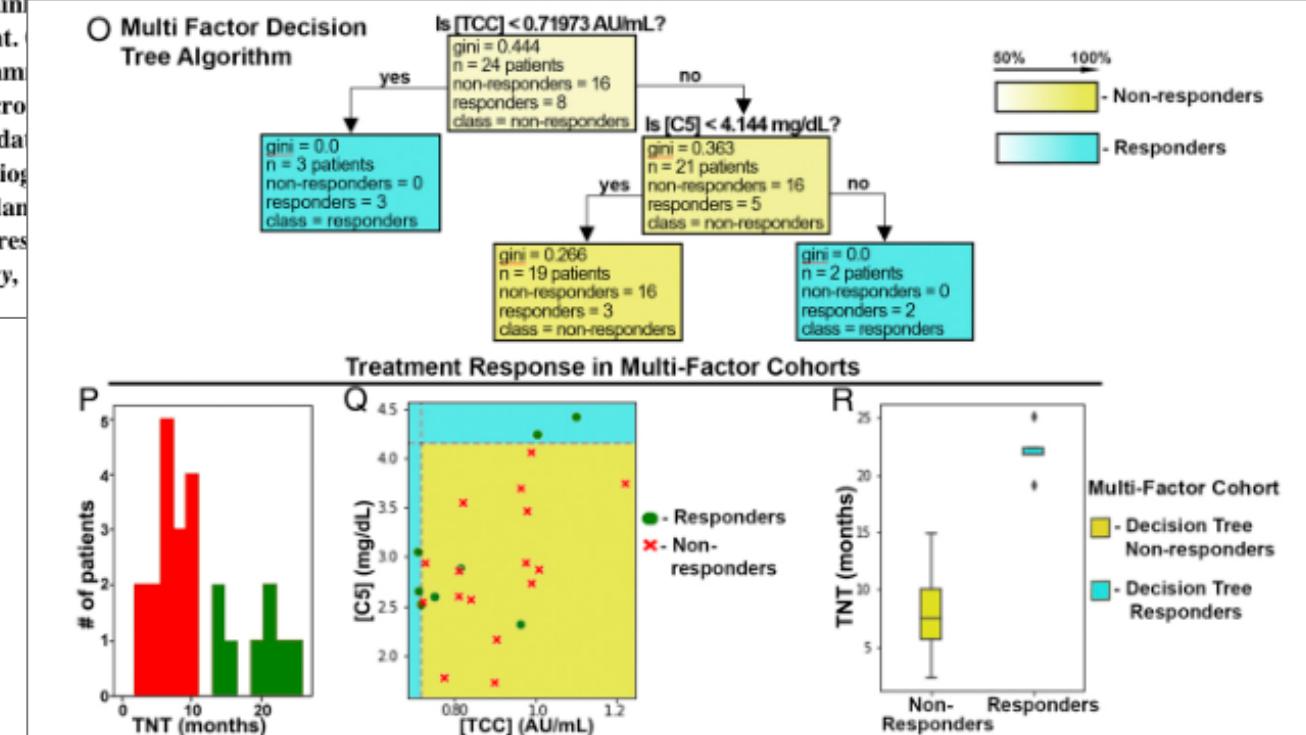
# One More Example

<https://www.jimmunol.org/content/early/2020/11/05/jimmunol.2000511>

## Complement as Prognostic Biomarker and Potential Therapeutic Target in Renal Cell Carcinoma

Britney Reese,\*<sup>1</sup> Ashok Silwal,\*<sup>1</sup> Elizabeth Daugherty,\* Michael Daugherty,<sup>†</sup> Mahshid Arabi,\* Pierce Daly,\* Yvonne Paterson,<sup>‡</sup> Layton Woolford,<sup>§,¶</sup> Alana Christie,<sup>§,¶</sup> Roy Elias,<sup>§,¶</sup> James Brugarolas,<sup>§,¶</sup> Tao Wang,<sup>¶,||</sup> Magdalena Karbowniczek,\* and Maciej M. Markiewski\*

Preclinical studies demonstrated that complement promotes tumor growth. Therefore, we sought to determine the best target for complement-based therapy among common human malignancies. High expression of 11 complement genes was linked to unfavorable prognosis in renal cell carcinoma. Complement protein expression or deposition was observed mainly in tumor vasculature, corresponding to a role of complement in regulating the tumor microenvironment. In tumors correlated with a high nuclear grade. Complement genes clustered within an aggressive inflammatory cancer characterized by poor prognosis, markers of T cell dysfunction, and alternatively activated macrophages. Complement proteins correlated with response to immune checkpoint inhibitors. Corroborating human data and blockade reduced tumor growth by enhancing antitumor immunity and seemingly reducing angiogenesis of kidney cancer resistant to PD-1 blockade. Overall, this study implicates complement in the immune landscape of renal cell carcinoma, and notwithstanding cohort size and preclinical model limitations, the data suggest that tumors resistant to immunotherapy might be suitable targets for complement-based therapy. *The Journal of Immunology*, 2020; 195(22): 7333–7344.



# Data Science



- 1) Having data is like sitting on top of a gold mine
- 2) Everyone has data...

Data Science is learning how to dig through the mountain and get the gold out

Part 2

# **WHAT IS MACHINE LEARNING: A TOO SHORT INTRO**

# On One Hand

Relevant XKCD



IN CS, IT CAN BE HARD TO EXPLAIN  
THE DIFFERENCE BETWEEN THE EASY  
AND THE VIRTUALLY IMPOSSIBLE.

# On The Other



# What is “learning”?

The goal:

**Make a decision based on data**

i.e. find “patterns” in the data. (Besides, if you aren’t making your decisions based on data, then what are you basing them on?)

To get a good answer you **MUST**:

- ask a good question
- have good data

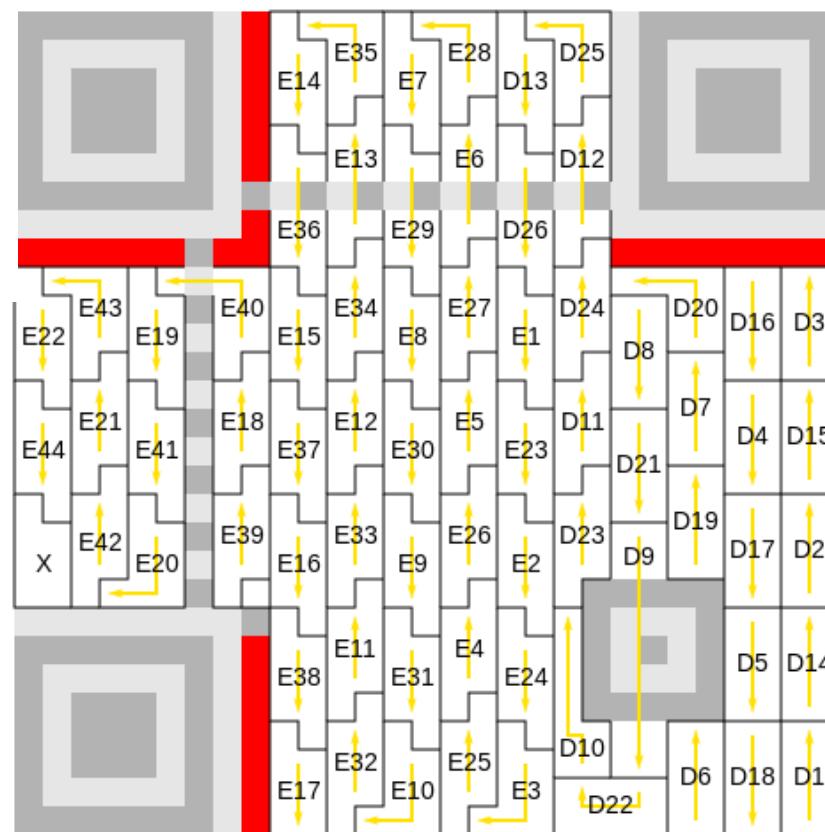
EXAMPLE

**QR CODES**

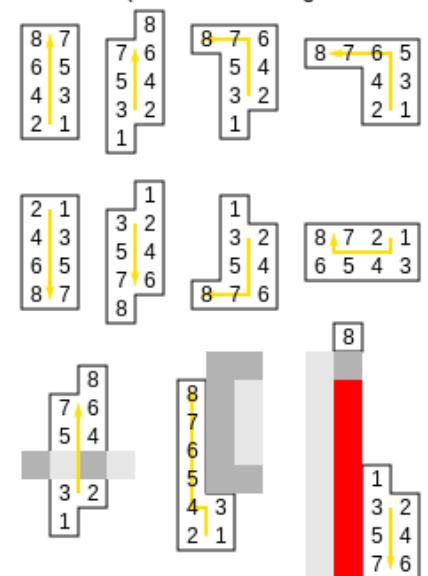
<http://www.datagenetics.com/blog/november12013/index.html>



- Positioning/Orientation
  - Format Information
  - Timing marks
  - Version Information
  - Spacing
  - Alignment



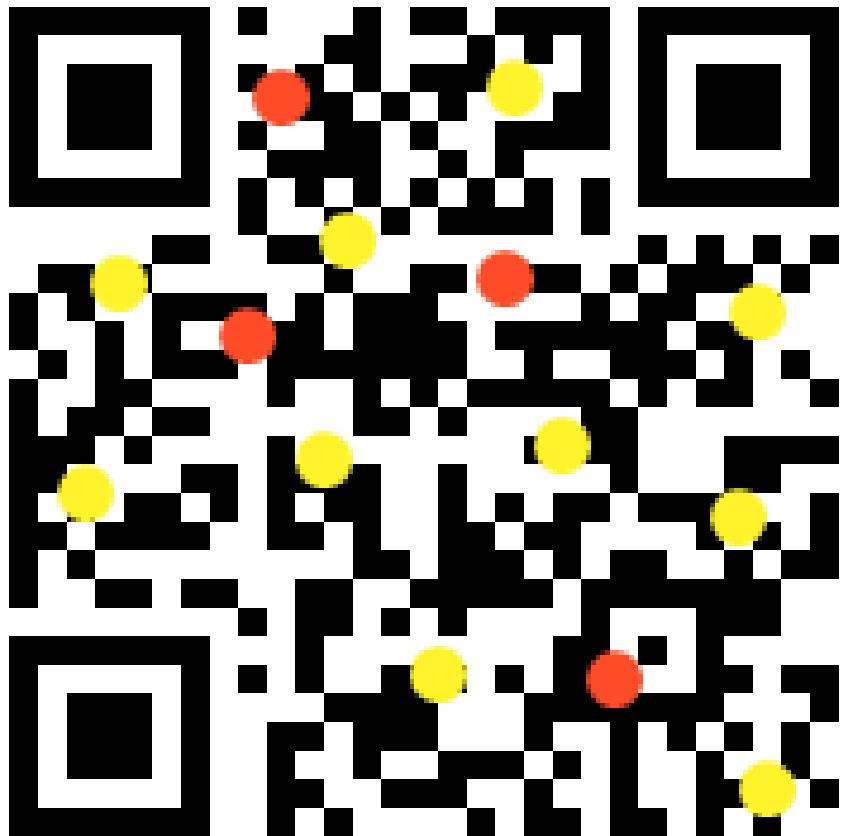
- Fixed Patterns Format Info  
D: Data, E: Error Correction, X: Unused  
Error Correction Level H is shown  
Block 1 Codewords: D1–D13, E1–E22  
Block 2 Codewords: D14–D26, E23–E44  
Message Data: D1–D13, D14–D26  
Bit order (1 is the most significant bit):

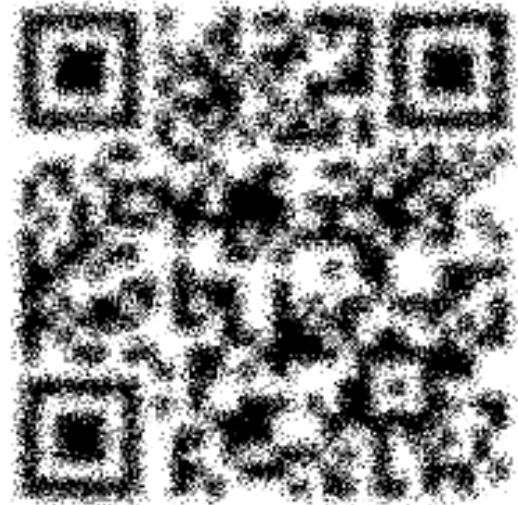












# MOAR EXAMPLES

# Machine Perception

- Build a machine that can recognize patterns:
  - Email or spam
  - Speech recognition
  - OCR (Optical Character Recognition)
- We often don't care *how* it works!?!
  - Lives or dies based on quality of training data

# Spam Filter

## Assassinating spam e-mail

SpamAssassin is a widely used open-source spam filter. It calculates a score for an incoming e-mail, based on a number of built-in rules or ‘tests’ in SpamAssassin’s terminology, and adds a ‘junk’ flag and a summary report to the e-mail’s headers if the score is 5 or more.

-0.1 RCVD_IN_MXRATE_WL	RBL: MXRate recommends allowing [123.45.6.789 listed in sub.mxrate.net]
0.6 HTML_IMAGE_RATIO_02	BODY: HTML has a low ratio of text to image area
1.2 TVD_FW_GRAPHIC_NAME_MID	BODY: TVD_FW_GRAPHIC_NAME_MID
0.0 HTML_MESSAGE	BODY: HTML included in message
0.6 HTML_FONx_FACE_BAD	BODY: HTML font face is not a word
1.4 SARE_GIF_ATTACH	FULL: Email has a inline gif
0.1 BOUNCE_MESSAGE	MTA bounce message
0.1 ANY_BOUNCE_MESSAGE	Message is some kind of bounce message
1.4 AWL	AWL: From: address is in the auto white-list

From left to right you see the score attached to a particular test, the test identifier, and a short description including a reference to the relevant part of the e-mail. As you see, scores for individual tests can be negative (indicating evidence suggesting the e-mail is ham rather than spam) as well as positive. The overall score of 5.3 suggests the e-mail might be spam.

# Spam Filter

SEND

Save Now

Discard

To

jenny [REDACTED]@g

Add Cc Add Bo

Subject

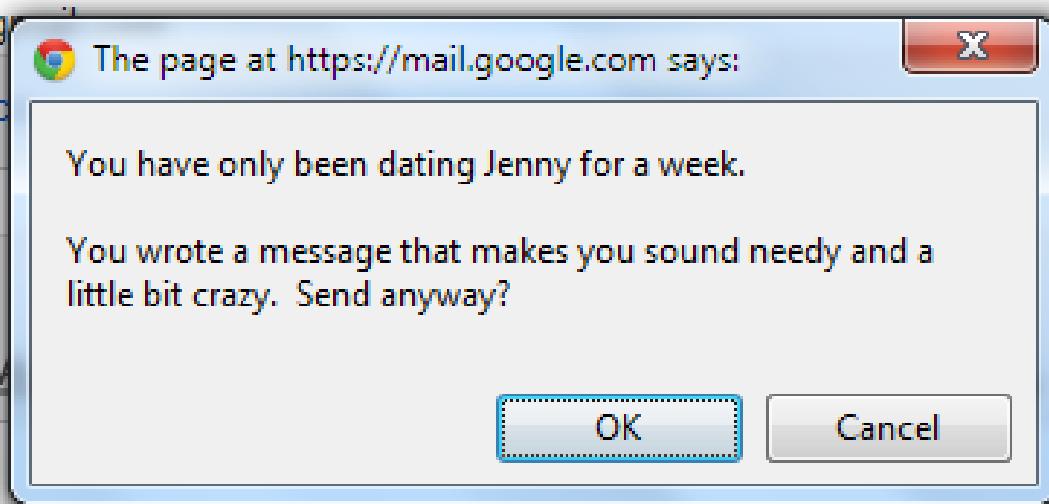
<3 <3 <3

Attach a file

B I U T + T + A

Dearest Jenny,

I can't stop thinking about you. I had that dream about you again last night. Y'know that dream where I am Captain Picard and you are Dr. Crusher. I miss you so much. Why didn't you return my calls last night? I tried calling you several times last night but I kept getting your voicemail. I know you were home last night because of your facebook and twitter posts. Are you upset with me or something? I hope I didn't do anything to mess up our relationship because I think we have something really special like we were destined to be



« Plain Text

# Kinect



## Real-Time Human Pose Recognition in Parts from Single Depth Images

Jamie Shotton

Andrew Fitzgibbon

Mat Cook

Toby Sharp

Mark Finocchio

Richard Moore

Alex Kipman

Andrew Blake

Microsoft Research Cambridge & Xbox Incubation

### Abstract

We propose a new method to quickly and accurately predict 3D positions of body joints from a single depth image, using no temporal information. We take an object recognition approach, designing an intermediate body parts representation that maps the difficult pose estimation problem into a simpler per-pixel classification problem. Our large and highly varied training dataset allows the classifier to estimate body parts invariant to pose, body shape, clothing, etc. Finally we generate confidence-scored 3D proposals of several body joints by reprojecting the classification result and finding local modes.

The system runs at 200 frames per second on consumer hardware. Our evaluation shows high accuracy on both synthetic and real test sets, and investigates the effect of several training parameters. We achieve state of the art accuracy in our comparison with related work and demonstrate improved generalization over exact whole-skeleton nearest neighbor matching.

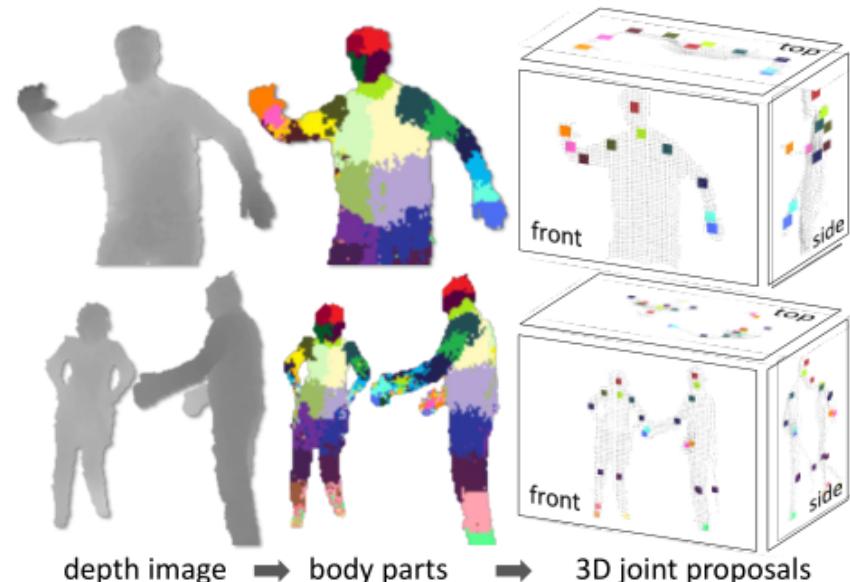
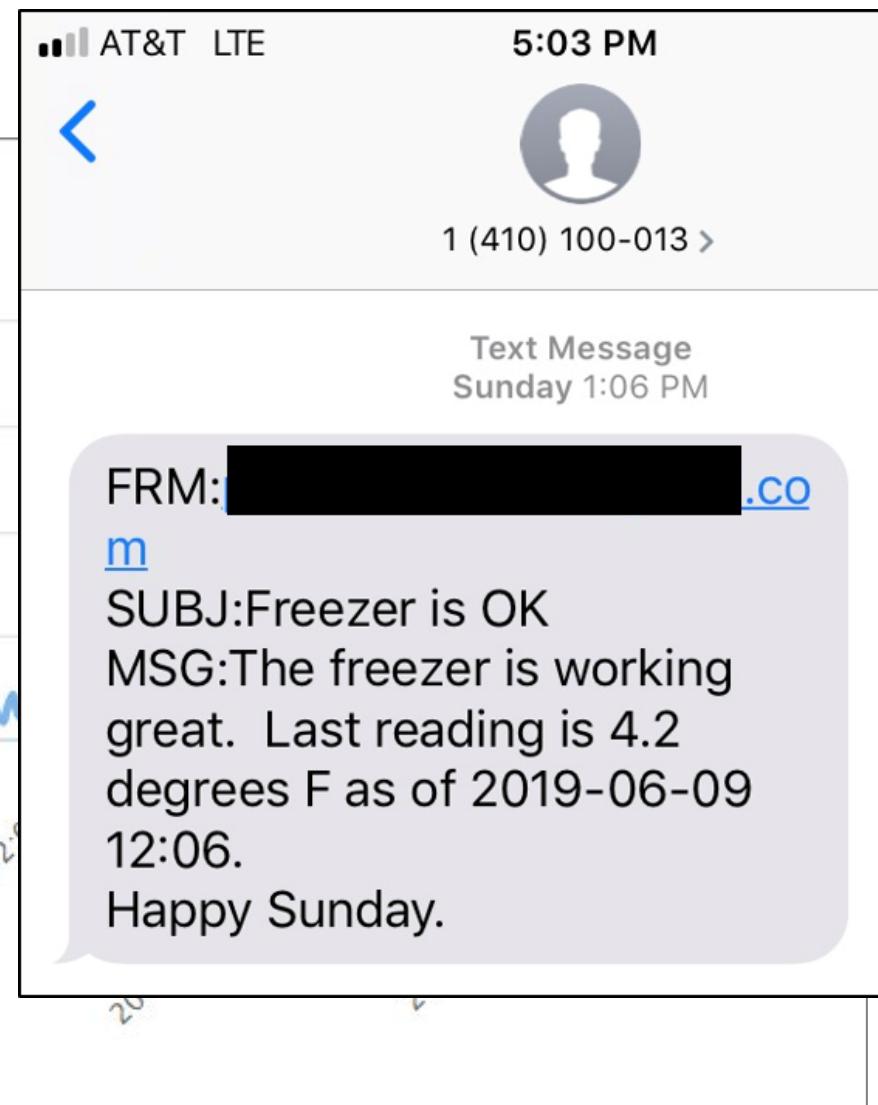
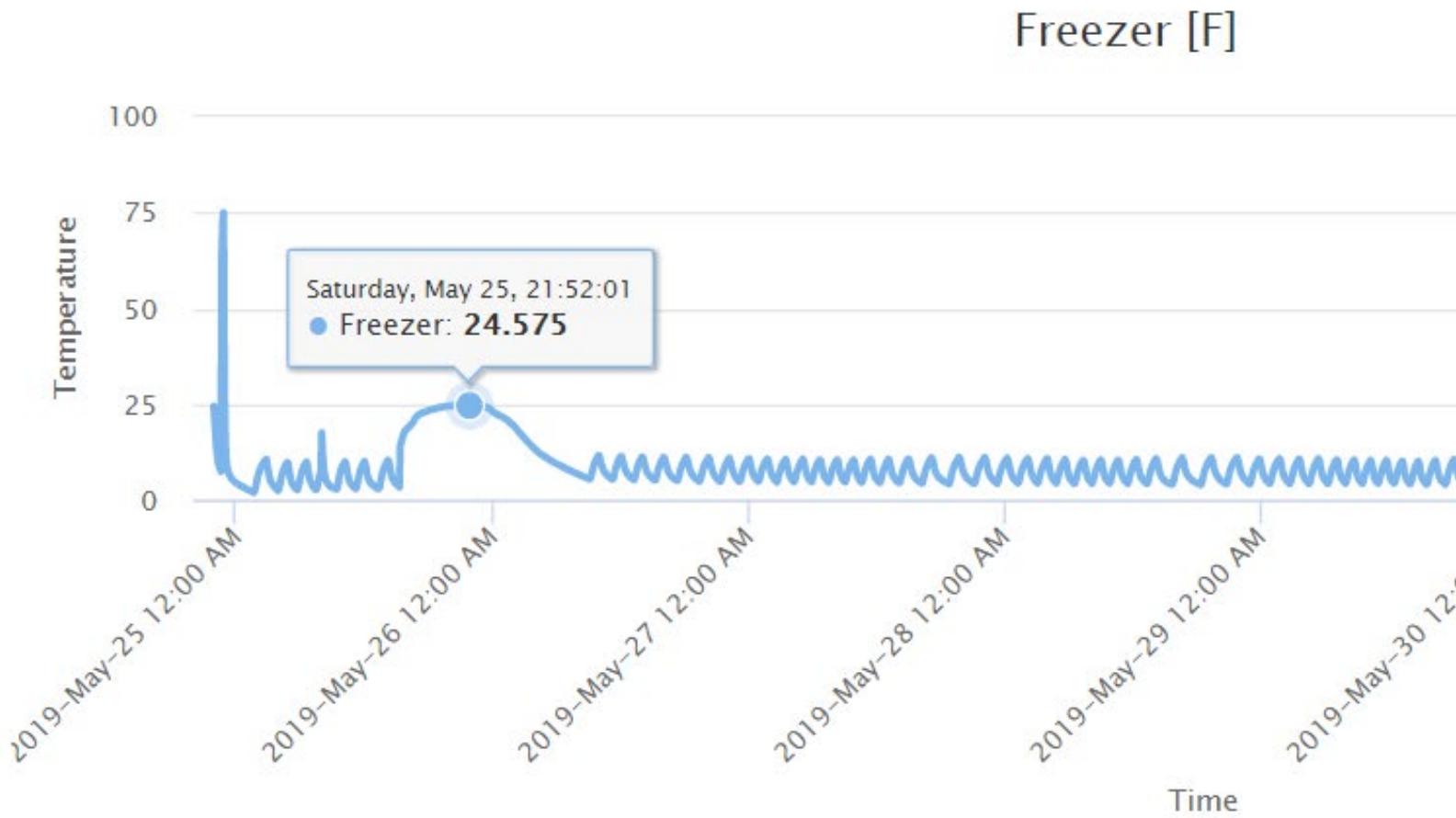


Figure 1. **Overview.** From an single input depth image, a per-pixel body part distribution is inferred. (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals). Local modes of this signal are estimated to give high-quality proposals for the 3D locations of body joints, even for multiple users.

# Freezer

Microcontrollers are immanently practical



# Just the basics **CONCEPTS**

Interviewer: What's your biggest strength?

Me: I'm an expert in machine learning.

Interviewer: What's  $9 + 10$ ?

Me: Its 3.

Interviewer: Not even close. It's 19.

Me: It's 16.

Interviewer: Wrong. Its still 19.

Me: It's 18.

Interviewer: No, it's 19.

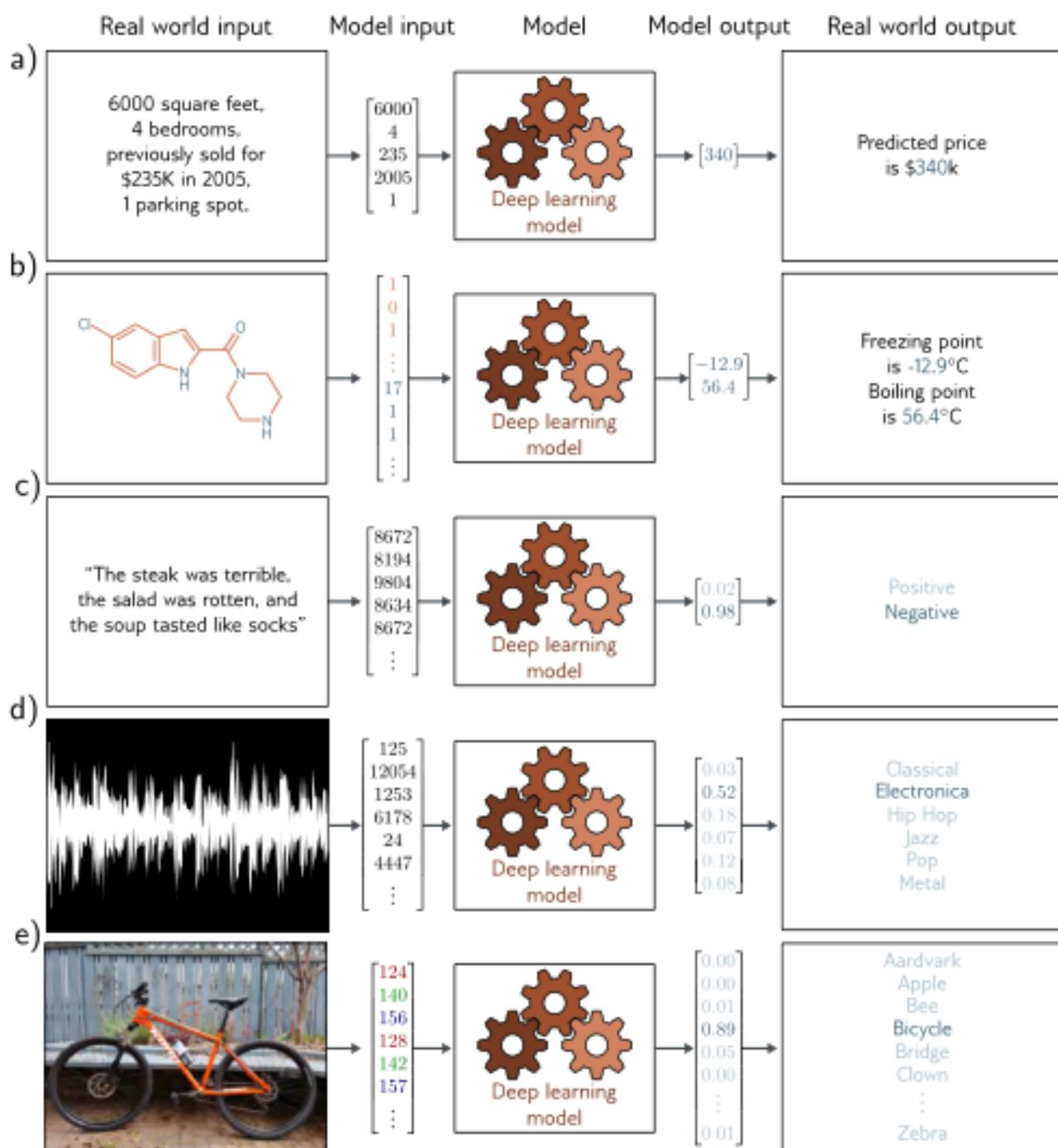
Me: it's 19.

Interviewer: You're hired

# Supervised Learning

- Features:
  - Turn a sample into some numbers
  - Curse of Dimensionality
  - Symmetries!
- Models:
  - Lots and lots and lots of classifiers
  - “No Free Lunch” Theorem, and no silver bullets either
- Training: “learn” and evaluate
- Repeat?

**Figure 1.2** Regression and classification problems. a) This *regression* model takes a vector of numbers that characterize a property and predicts its price. b) This *multivariate regression* model takes the structure of a chemical molecule and predicts its melting and boiling points. c) This *binary classification* model takes a restaurant review and classifies it as either positive or negative. d) This *multiclass classification* problem assigns a snippet of audio to one of  $N$  genres. e) A second multiclass classification problem in which the model classifies an image according to which of  $N$  possible objects it might contain.



# Feature Choice - Examples

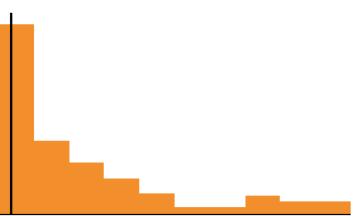
What does Spotify use for songs?

<https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>

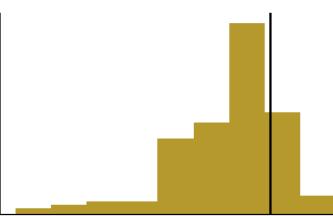
How Songs on Spotify's 'Grime Shutdown' Playlist Compare to my Music Taste

The song '*Strictly Business*' by Shorty, Wiley has an average variance of 0.11 compared with the mean of each audio feature.

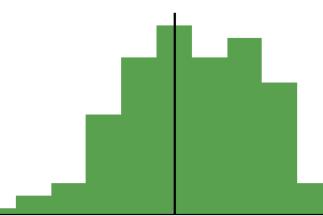
Acousticness



Danceability



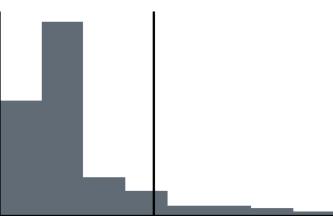
Energy



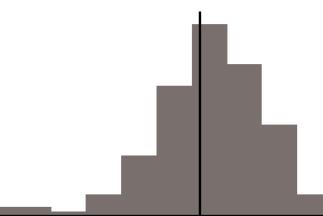
Instrumentalness



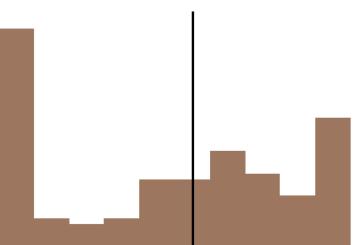
Liveness



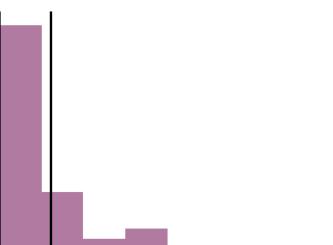
Loudness



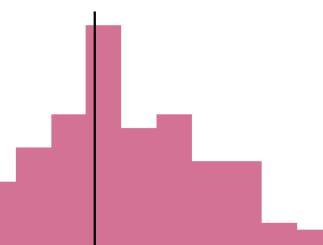
Key



Speechiness



Valence



## acousticness number<float>

A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

=> 0 <= 1

## analysis\_url string

A URL to access the full audio analysis of this track. An access token is required to access this data.

## danceability number<float>

Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.

## duration\_ms integer

The duration of the track in milliseconds.

## energy number<float>

Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

## id string

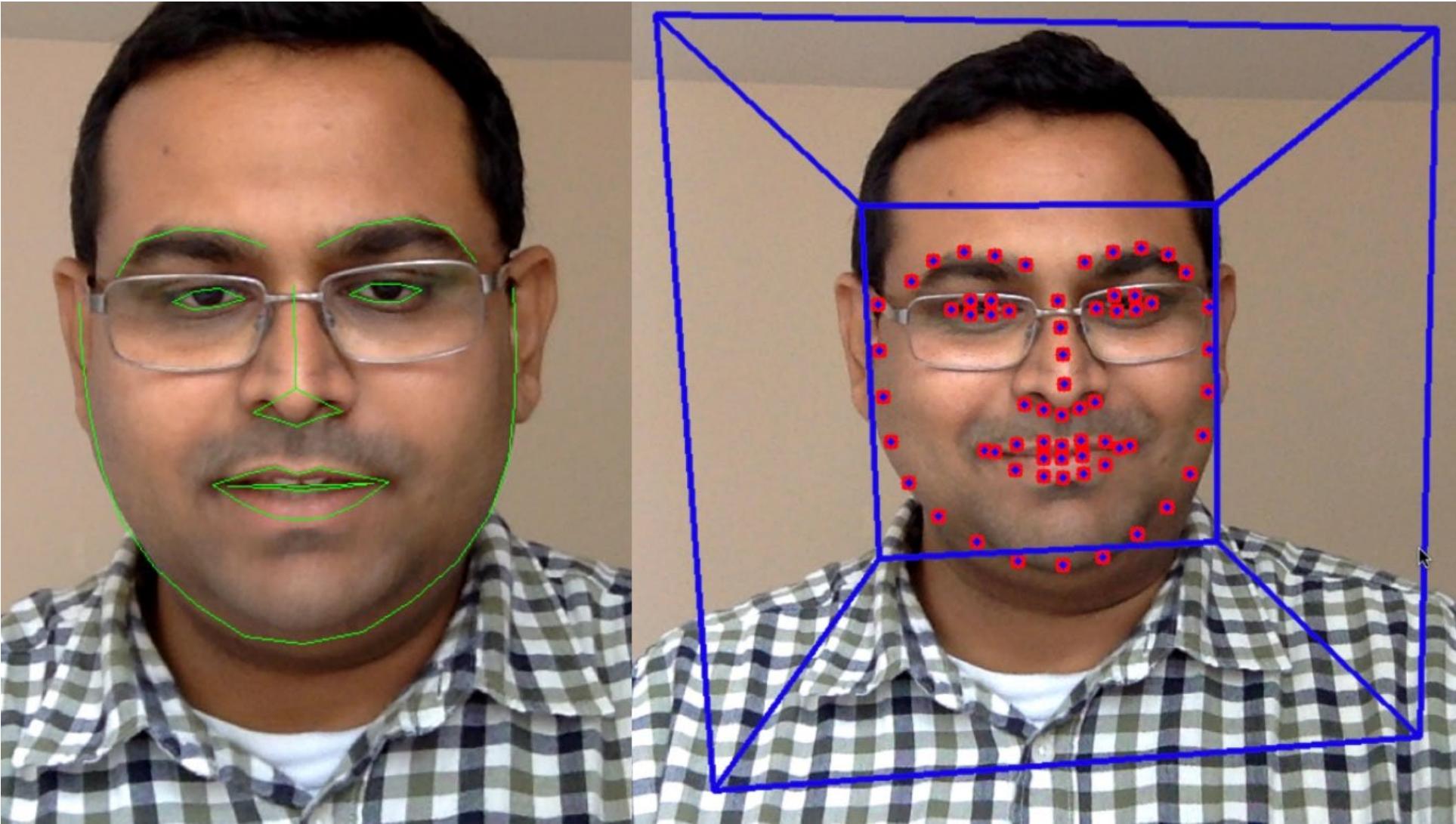
The Spotify ID for the track.

## instrumentalness number<float>

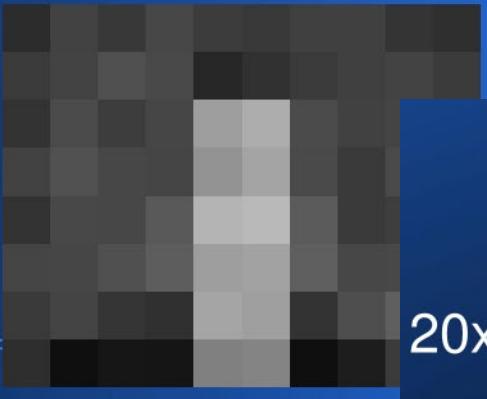
Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.

# Feature Choice - Examples

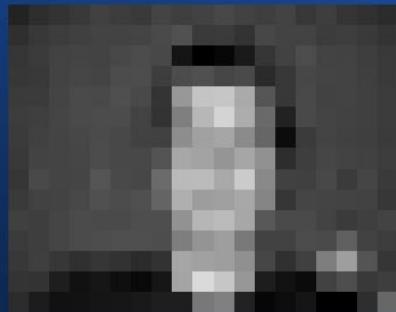
What about faces?



# Humans may require more inputs



20x15 pixels



10x7 pixels



<https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>

TOM SIMONITE

BUSINESS 07.22.2019 07:00 AM

# The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

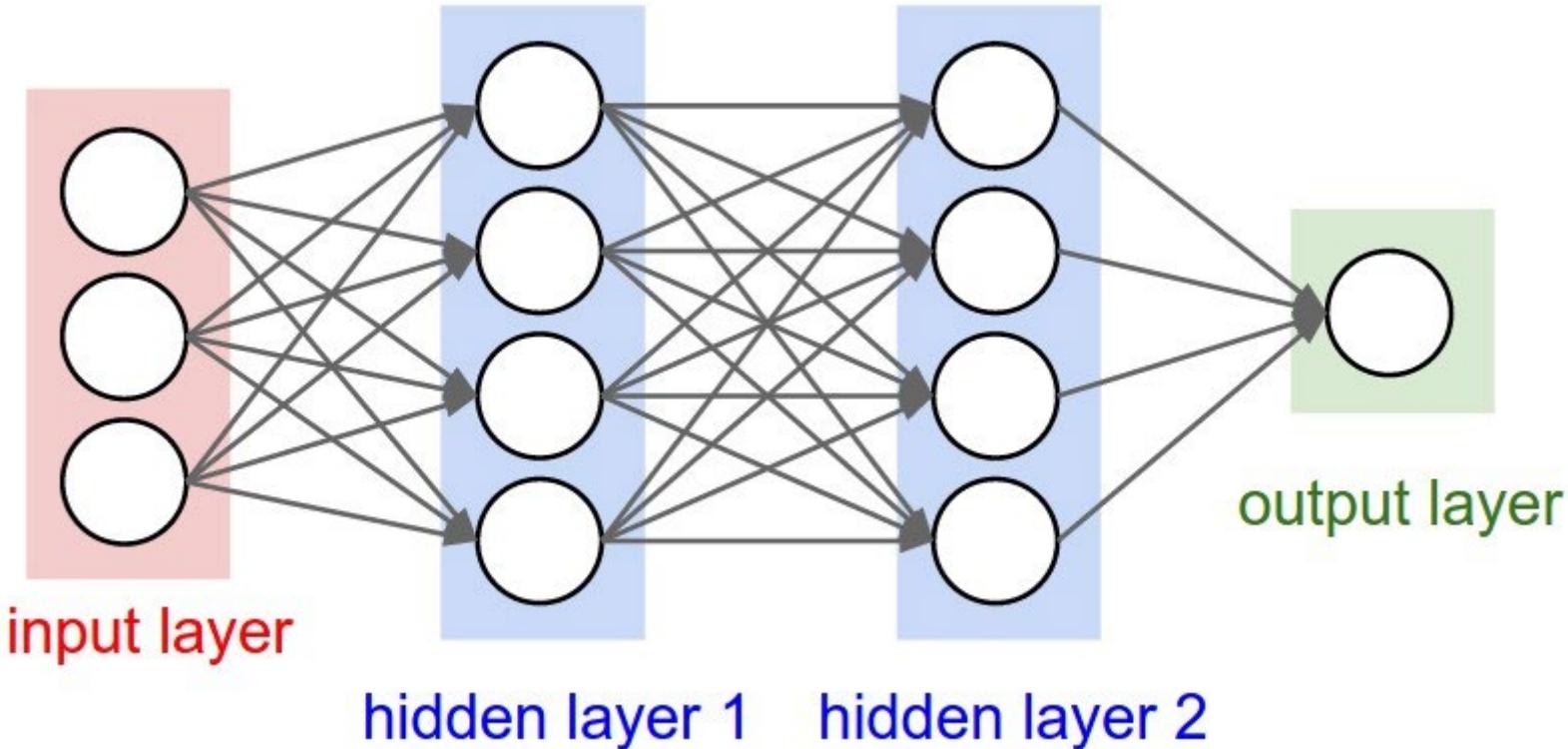
OCTOBER 24, 2020

BLOG, SCIENCE POLICY, SPECIAL EDITION: SCIENCE POLICY AND SOCIAL JUSTICE

## Racial Discrimination in Face Recognition Technology

<http://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>

# Models: Neural Network



- Each **input** is a SINGLE NUMBER
- **Hidden** and **output** values are directly calculated from inputs and **weights**
- Weights (arrows) are adjusted in training to make output match known right answer

# Seriously Though

[https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)

## 1. Supervised learning

- 1.1. Linear Models
- 1.2. Linear and Quadratic Discriminant Analysis
- 1.3. Kernel ridge regression
- 1.4. Support Vector Machines
- 1.5. Stochastic Gradient Descent
- 1.6. Nearest Neighbors
- 1.7. Gaussian Processes
- 1.8. Cross decomposition
- 1.9. Naive Bayes
- 1.10. Decision Trees
- 1.11. Ensemble methods
- 1.12. Multiclass and multioutput algorithms
- 1.13. Feature selection
- 1.14. Semi-supervised learning
- 1.15. Isotonic regression
- 1.16. Probability calibration
- 1.17. Neural network models (supervised)

## 2. Unsupervised learning

- 2.1. Gaussian mixture models
- 2.2. Manifold learning
- 2.3. Clustering
- 2.4. Biclustering
- 2.5. Decomposing signals in components (matrix factorization problems)
- 2.6. Covariance estimation
- 2.7. Novelty and Outlier Detection
- 2.8. Density Estimation
- 2.9. Neural network models (unsupervised)

# Other Links

## Word Lens

- <https://www.youtube.com/watch?v=h2OfQdYrHRs>
- <https://www.youtube.com/watch?v=zSOlYdlqyTQ>

## Other Stuff

- <https://www.netsafe.org.nz/rescam/>
- <http://detexify.kirelabs.org/classify.html>
- <https://www.youtube.com/watch?v=qv6UVOQ0F44> – MARI/O
- <https://www.youtube.com/watch?v=DMyJ6fGUqRI&t=46s> – Rock Wall Pong
- <https://xkcd.com/2236/> - Is it Christmas?

# What is Pattern Recognition?

Your turn:

My turn:

**Make a decision based on data**

i.e. find “patterns” in the data. Besides, if you aren’t making your decisions based on data, then what are you basing them on?

To get a good answer you MUST:

- ask a good question
- have good data

Other Tricks

**REGRESSION**

# Regression

Examples:

- predict the price of a house in Boston
- predict how much money a movie will make
- guess the value of a stock
- Diabetes dataset: predict measure of disease based on patient data

# Regression

Instead of classifying into categories, our job is now to predict a real-valued number based on training data

Basically all of the techniques we've learned can be modified to work for regression

[http://scikit-learn.org/stable/auto\\_examples/plot\\_cv\\_predict.html](http://scikit-learn.org/stable/auto_examples/plot_cv_predict.html)

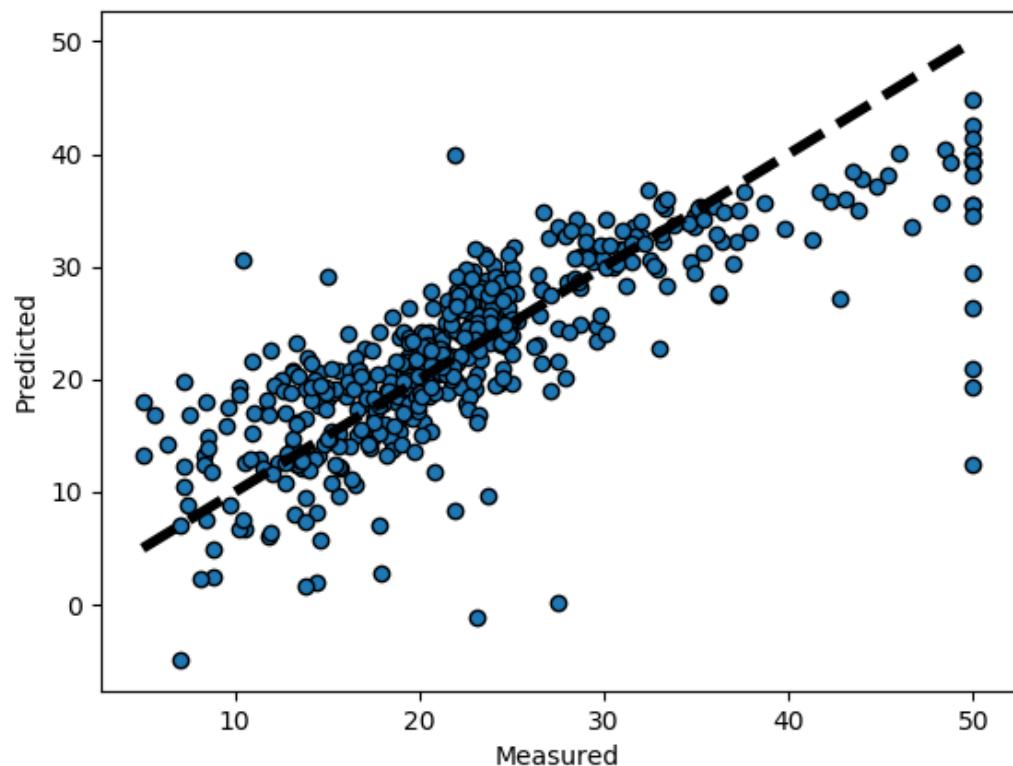
Predicting the price of a house based on features like square footage, # bedrooms, age, etc.

```
from sklearn import datasets
from sklearn.model_selection import cross_val_predict
from sklearn import linear_model
import matplotlib.pyplot as plt

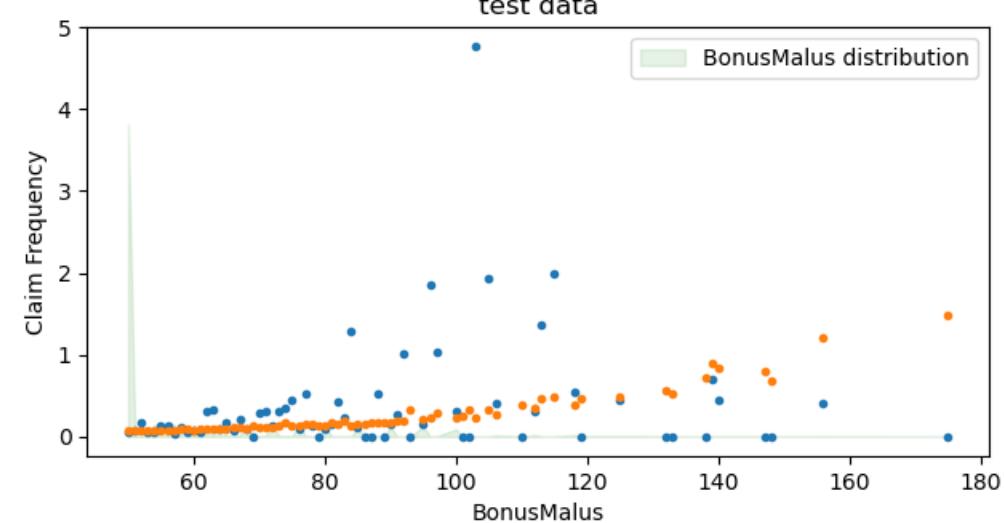
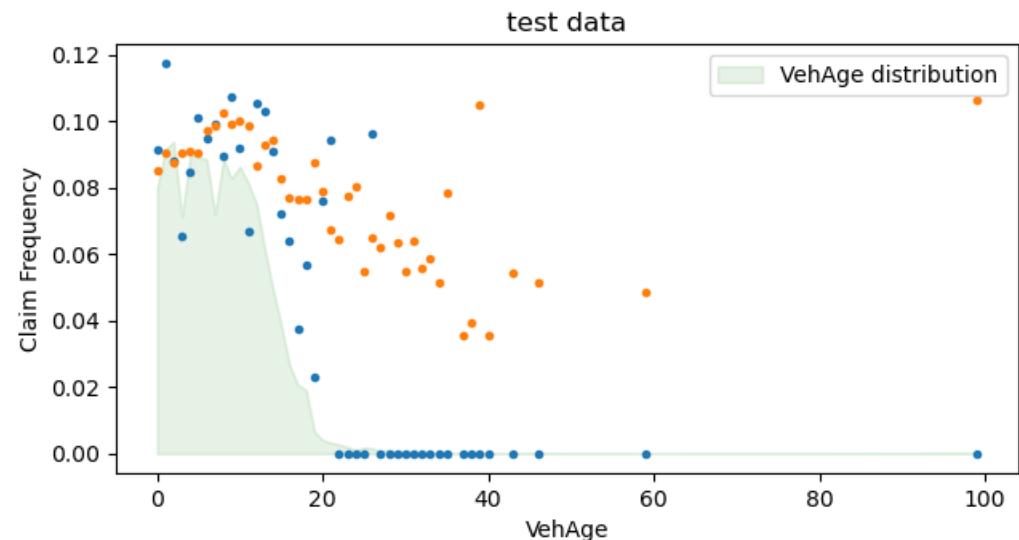
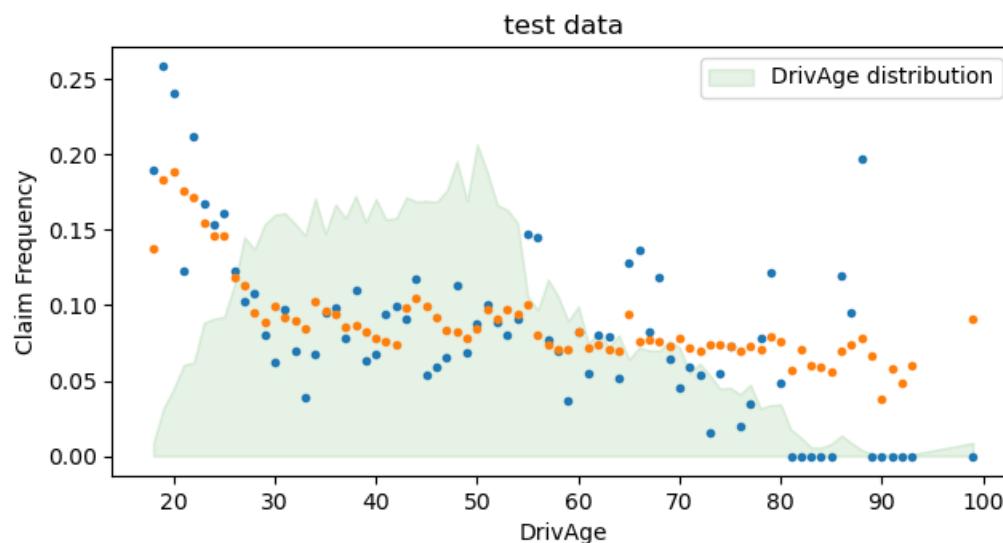
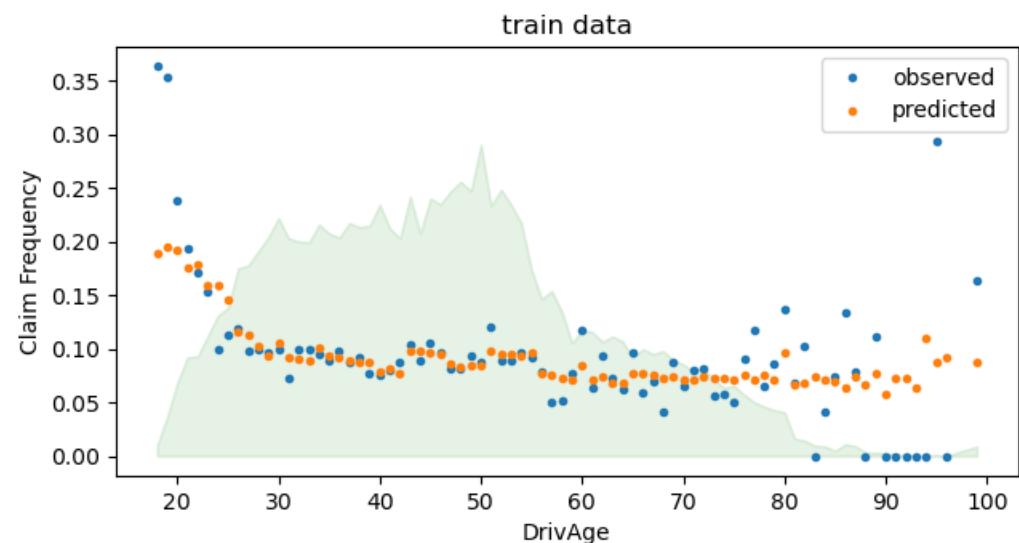
lr = linear_model.LinearRegression()
boston = datasets.load_boston()
y = boston.target

# cross_val_predict returns an array of the same size as `y` where each entry
# is a prediction obtained by cross validation:
predicted = cross_val_predict(lr, boston.data, y, cv=10)

fig, ax = plt.subplots()
ax.scatter(y, predicted, edgecolors=(0, 0, 0))
ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



# Predicting car insurance claims for drivers in France

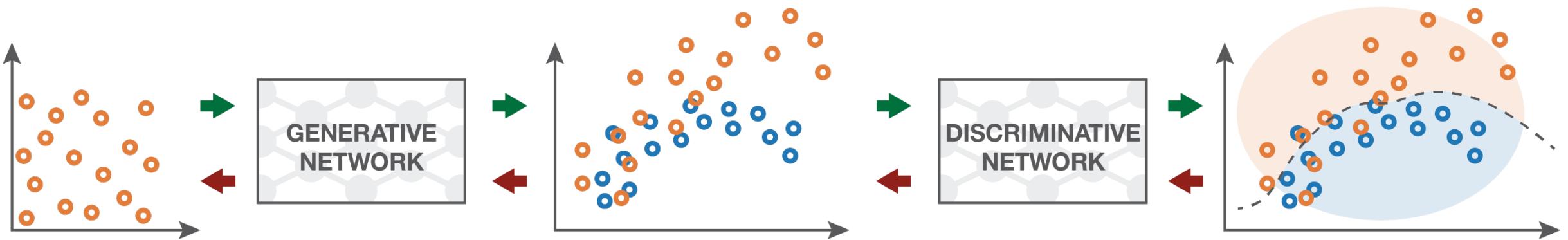


*Bonus-Malus (good-bad) is like a good driver score based on how many previous claims a driver has*

Other Tricks

**GENERATION**

■ Forward propagation (generation and classification) ■ Backward propagation (adversarial training)



Input random variables.

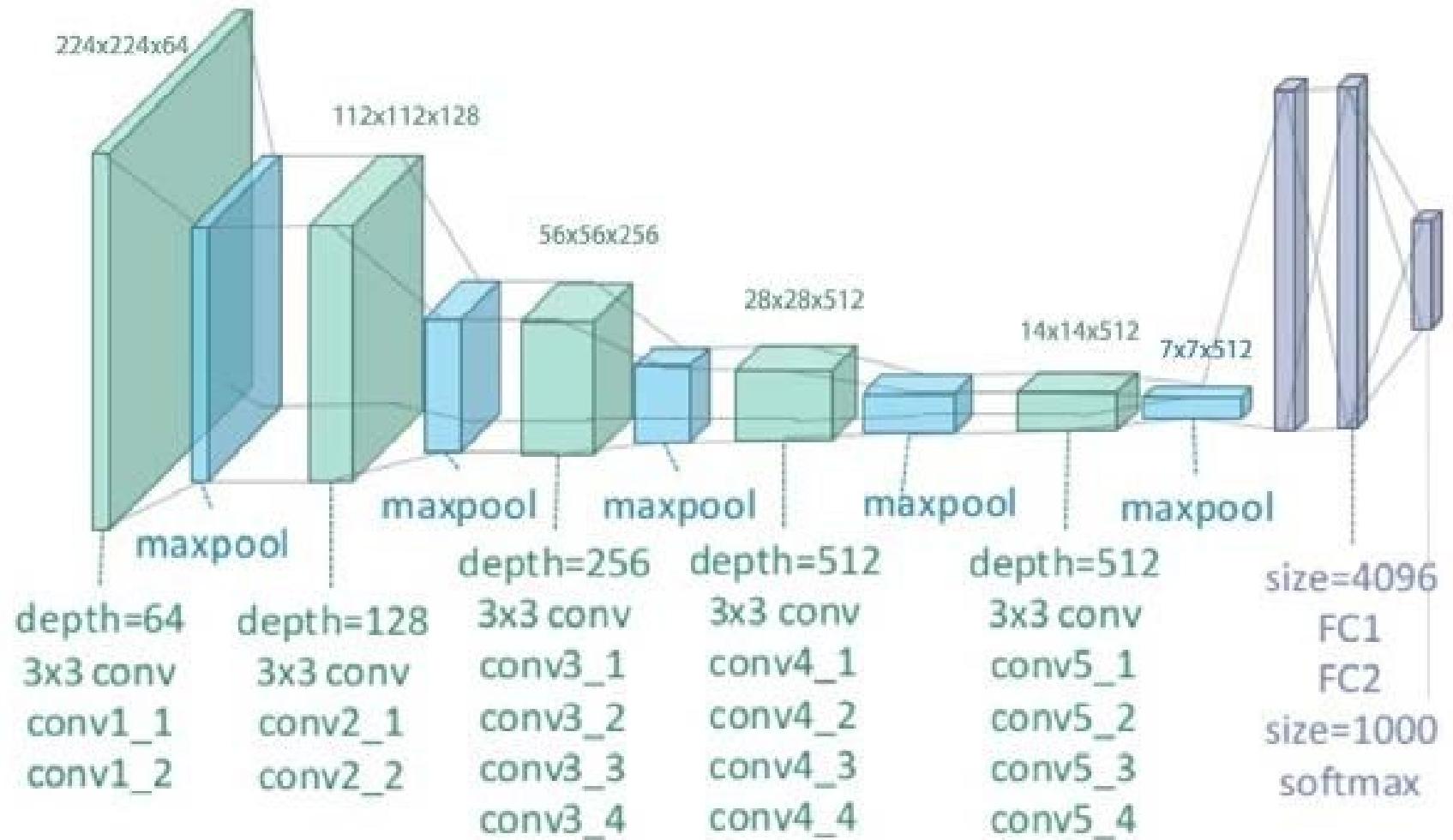
The generative network is trained to **maximise** the final classification error.

The **generated distribution** and the **true distribution** are not compared directly.

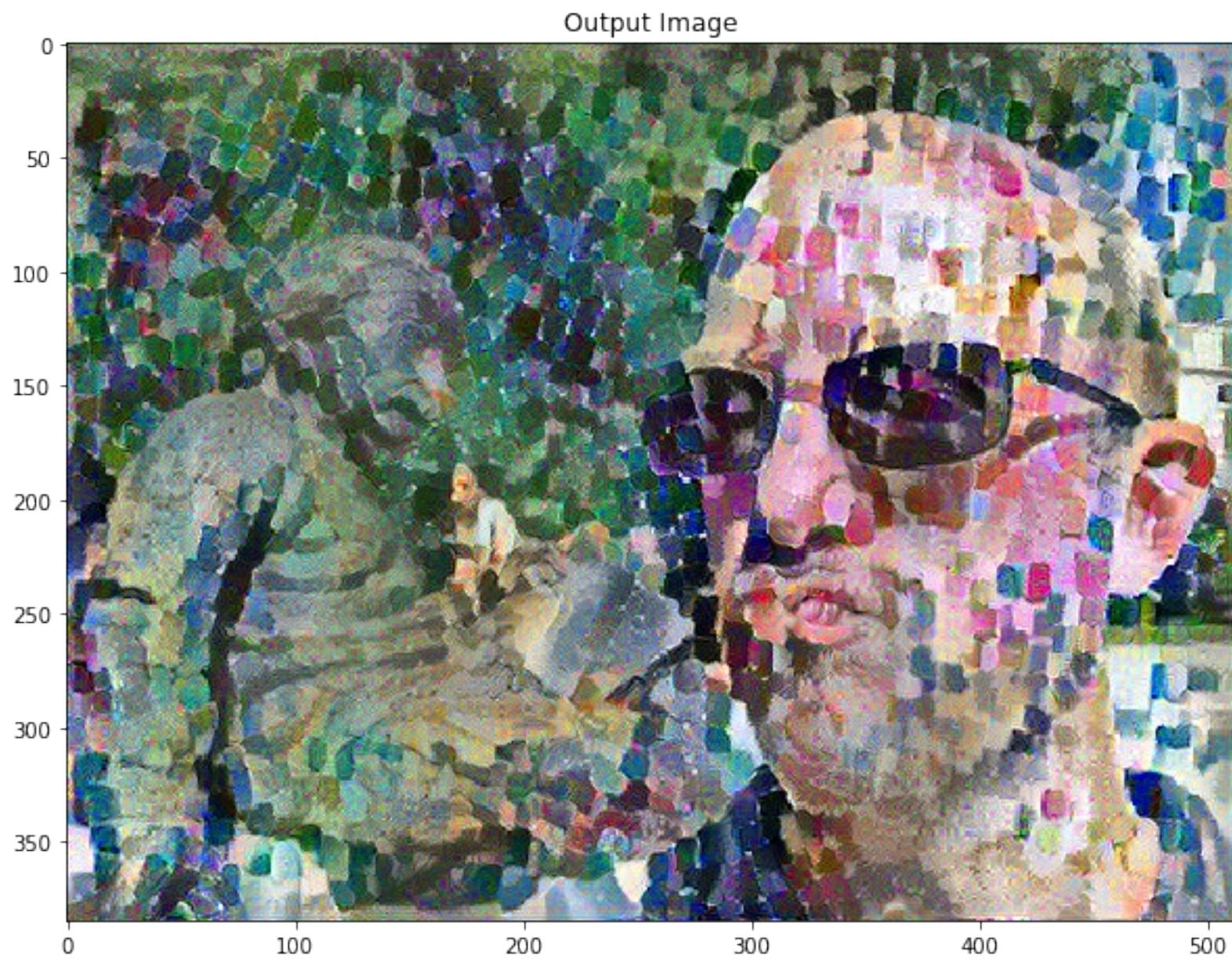
The discriminative network is trained to **minimise** the final classification error.

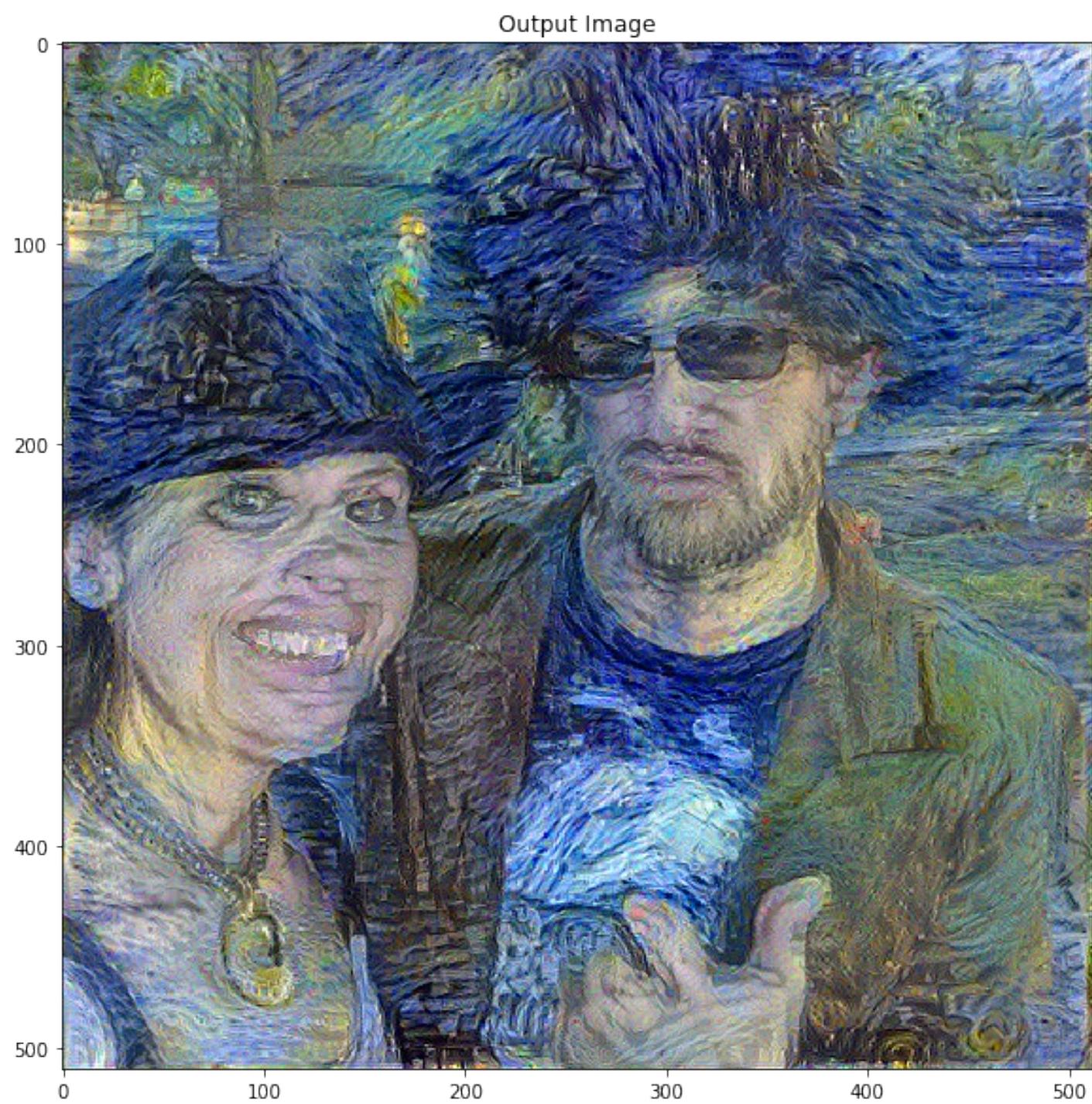
The classification error is the basis metric for the training of both networks.

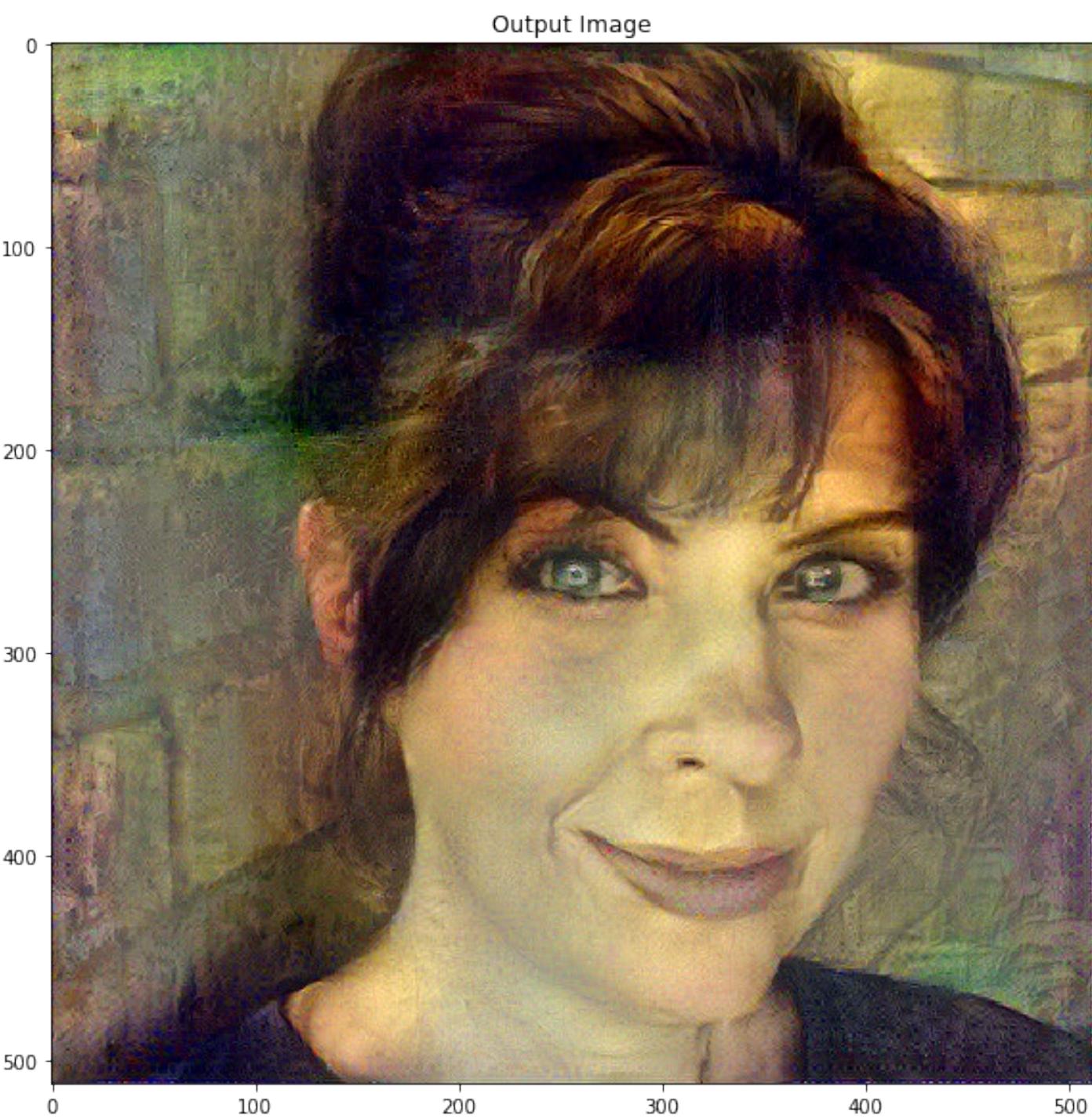
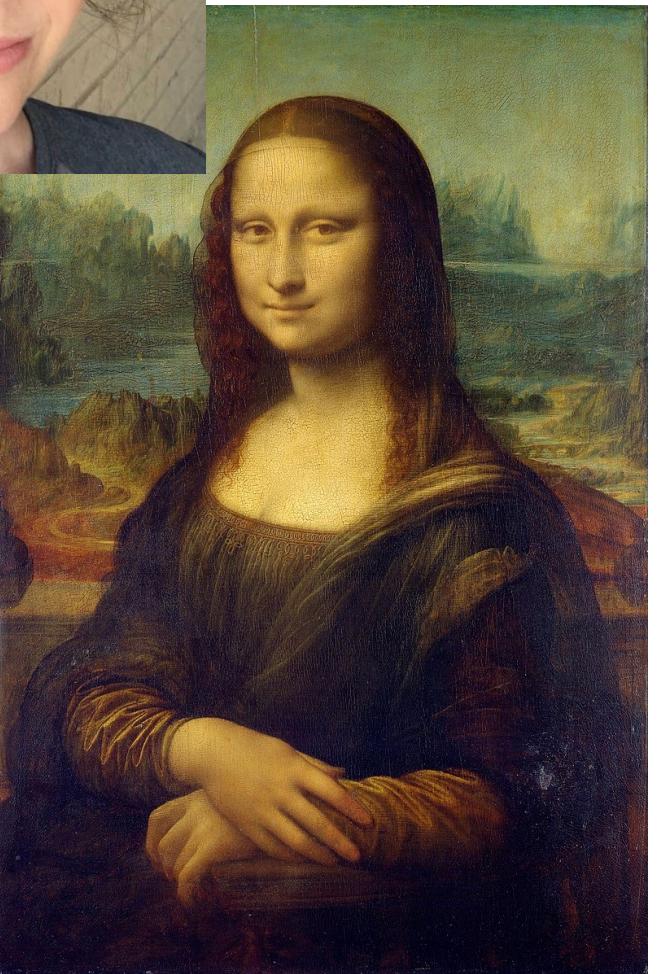
[https://colab.research.google.com/github/tensorflow/models/blob/master/research/nst\\_blogpost/4\\_Neural%20Style%20Transfer%20with%20Eager%20Execution.ipynb](https://colab.research.google.com/github/tensorflow/models/blob/master/research/nst_blogpost/4_Neural%20Style%20Transfer%20with%20Eager%20Execution.ipynb)









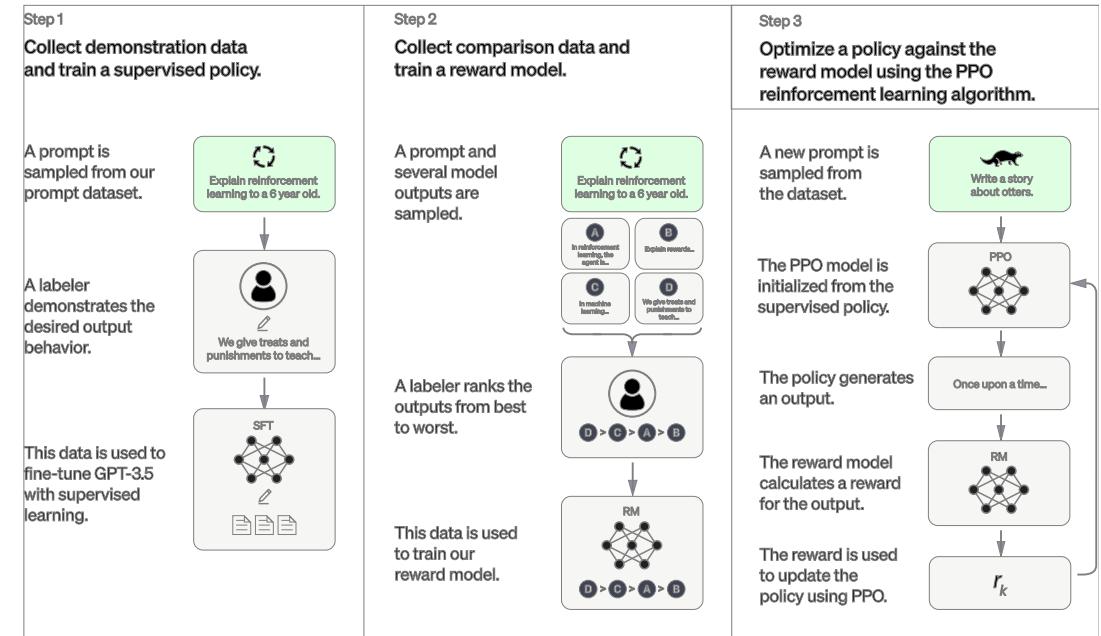




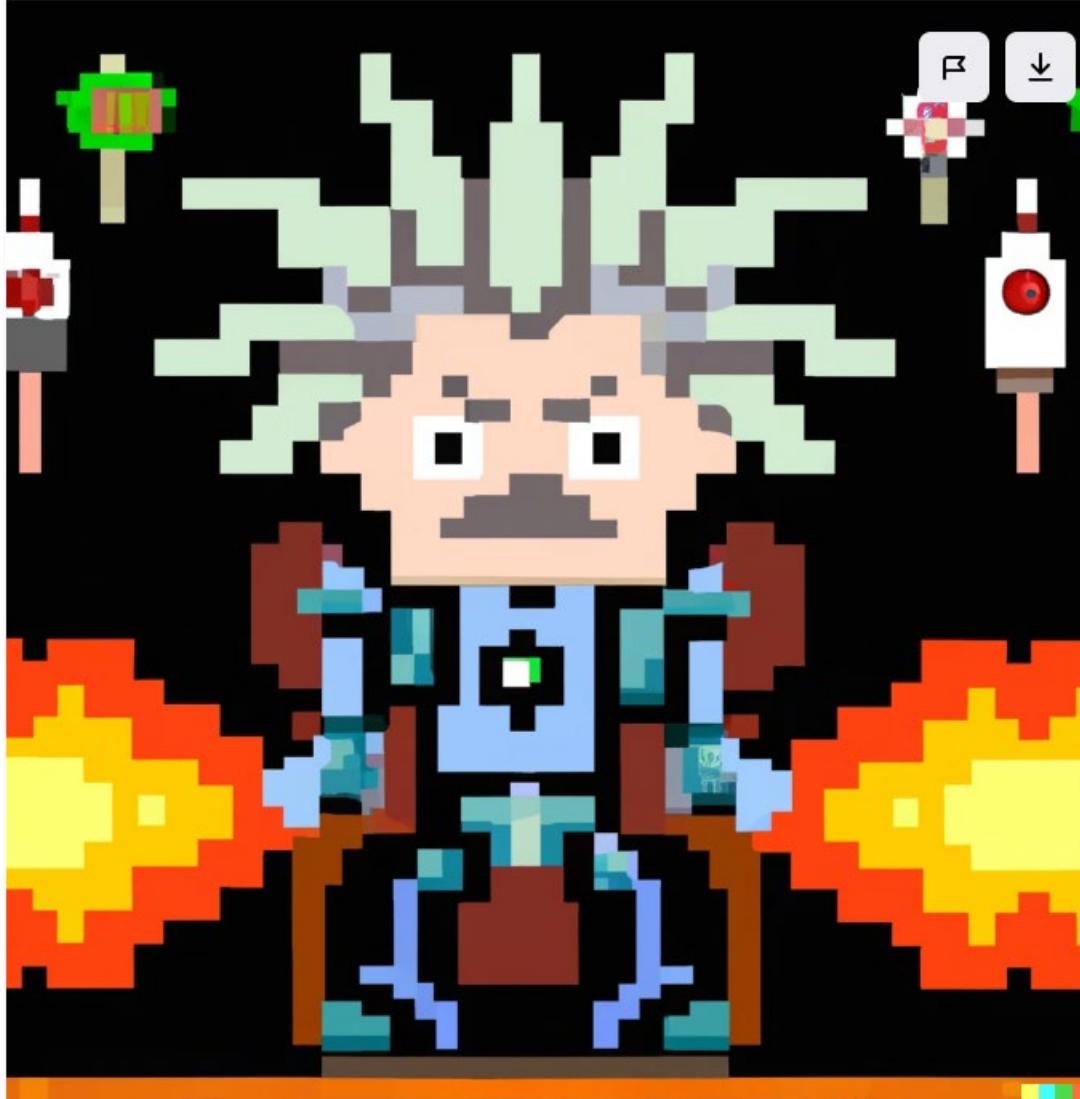


# More Examples

- DALL-E: generating images based on input text description  
<https://labs.openai.com/>
- ChatGPT: generating text based on input text description  
<https://openai.com/blog/chatgpt/>



## DALLE 2: Making Einstein as a final boss for GR



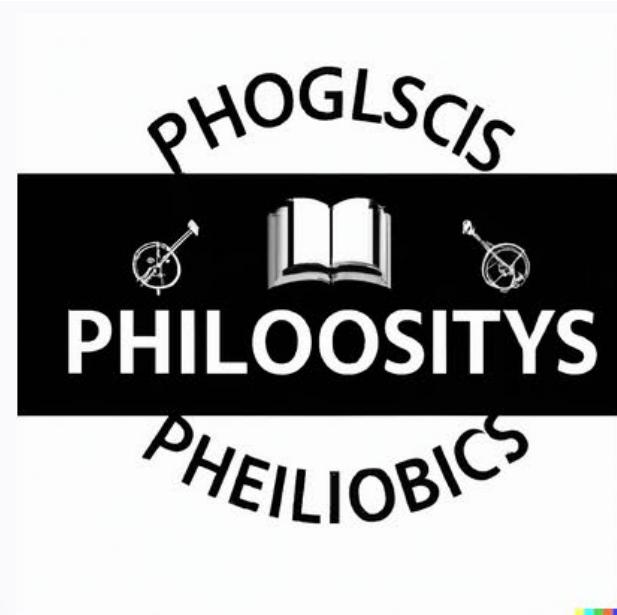
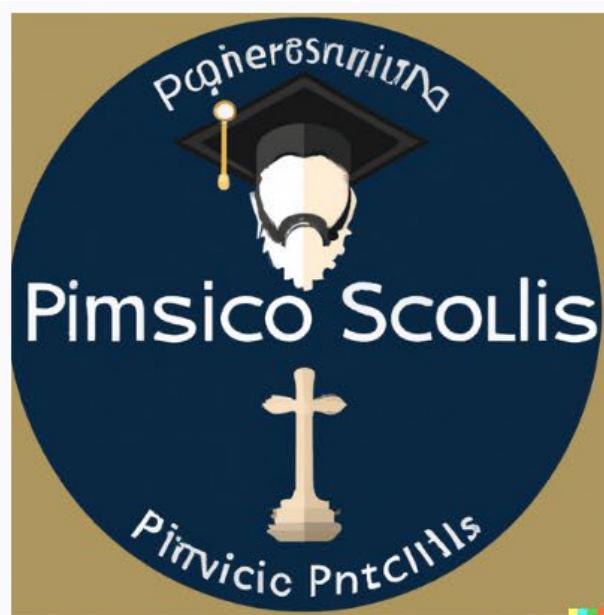
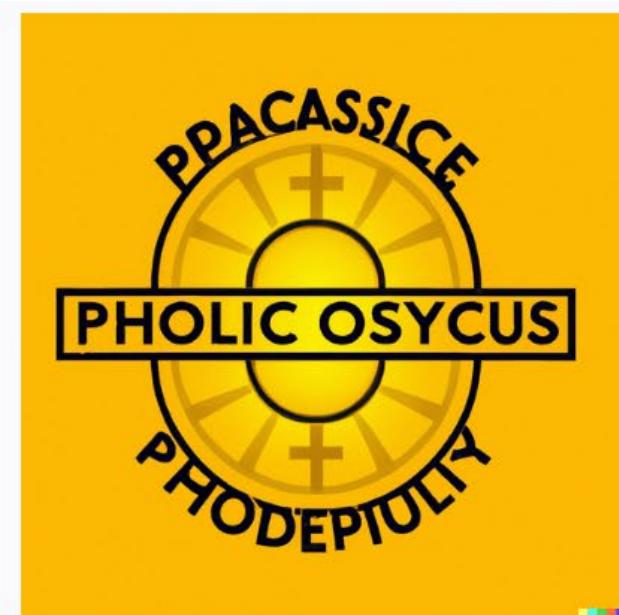
# AI-Generated “Art”

I gave DALLE, the AI picture generator, prompts about religion and science

DALL-E History Collections

Edit the detailed description

logo for college class about philosophy of science and religion



# AI-Generated “Art”

I gave DALLE, the AI picture generator, prompts about religion and science

 DALL-E History Collections

Edit the detailed description

a photograph of a scientist and a priest fighting



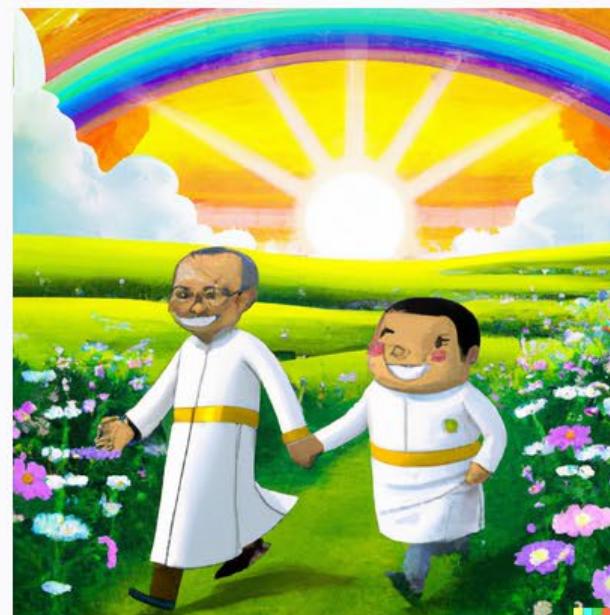
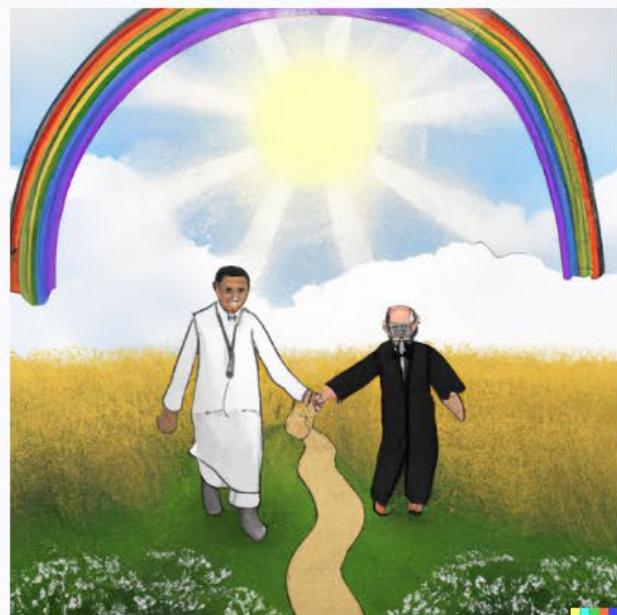
# AI-Generated “Art”



DALL-E History Collections

Edit the detailed description

digital art of a scientist and a priest holding hands and skipping through a field of flowers under a rainbow and a smiling sun



[← Post](#)

Felix

@felix\_red\_panda

[Follow](#)

...

Microsoft paper claims ChatGPT 3.5  
has ~20 billion parameters  
[arxiv.org/abs/2310.17680](https://arxiv.org/abs/2310.17680)

System description		Python (0)		
System	Model	#P	top-1	top-5
T5	t5-large	770M	80.4	94.0
CodeT5	codet5-large	770M	80.5	94.0
GPT-3	text-davinci-003	175B	<b>82.5</b>	94.0
ChatGPT	gpt-3.5-turbo	<b>20B</b>	80.6	94.0
StarCoder	starcoder	15.5B	79.2	93.0
CodeT5+	codet5p-16b	16B	79.6	93.0
CodeGen	codegen-350m	350M	80.1	93.0
Diffusion-LM	Custom	50M	70.4	89.0
GENIE	Custom	93M	73.2	93.0
<b>CODEFUSION</b>	<b>Custom</b>	75M	80.7	94.0

8:03 PM · Oct 30, 2023 · 1M Views

68

417

2.3K

676

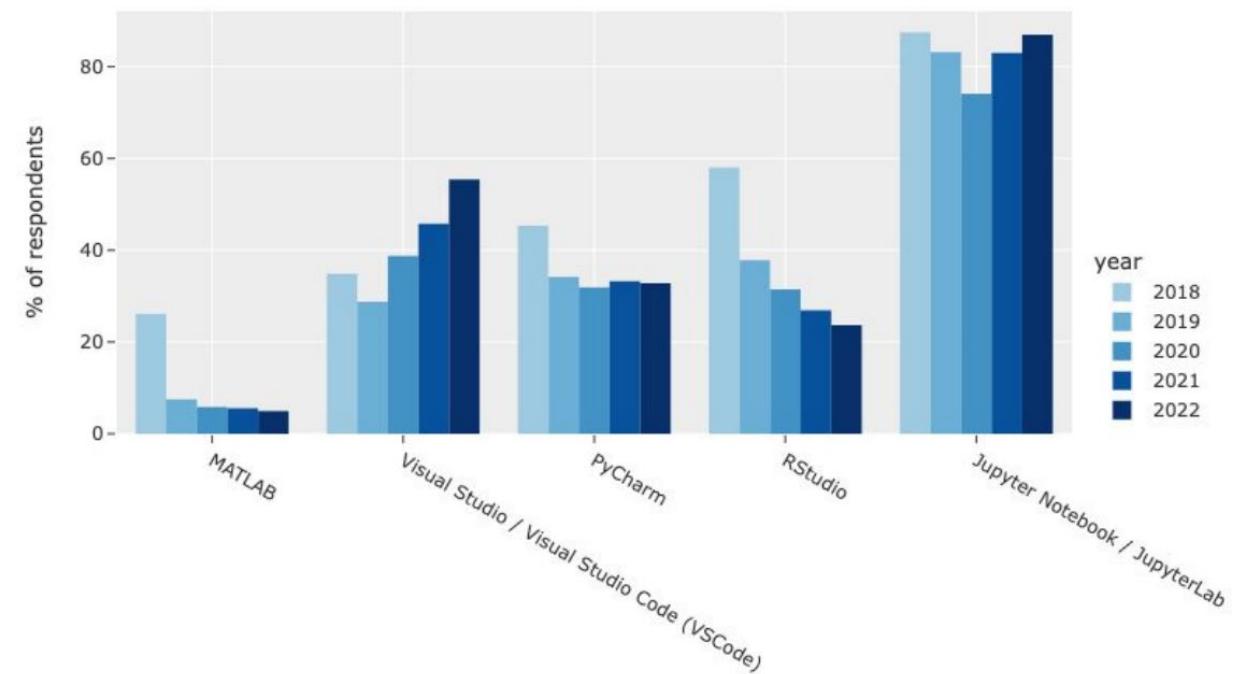
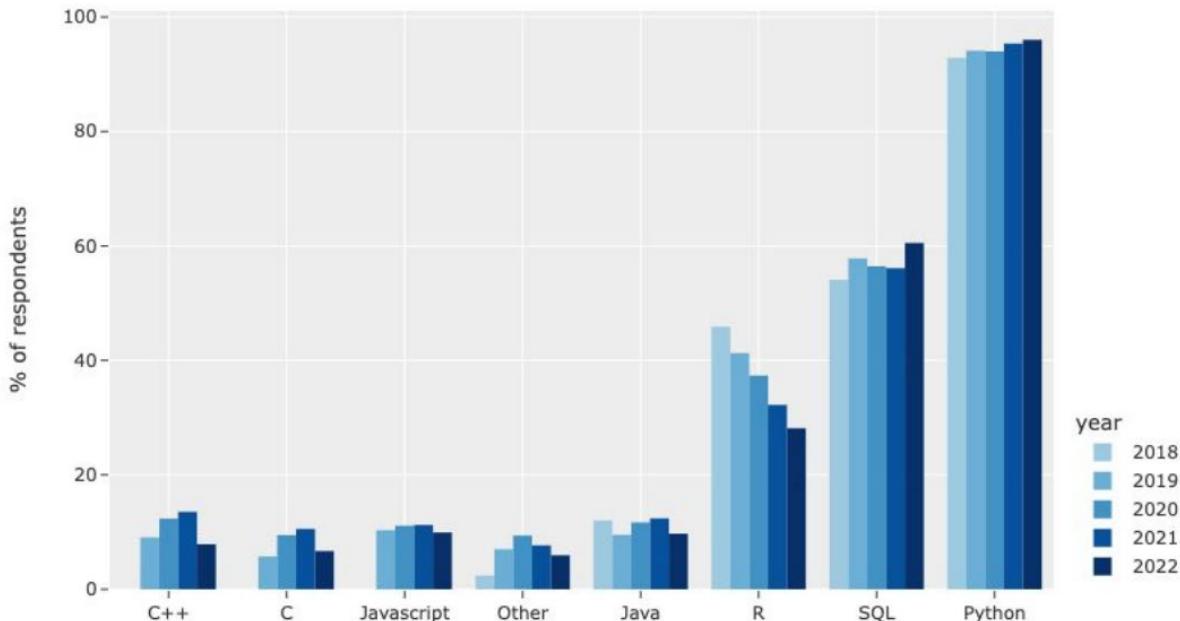


Part 3

# **TOOLS**

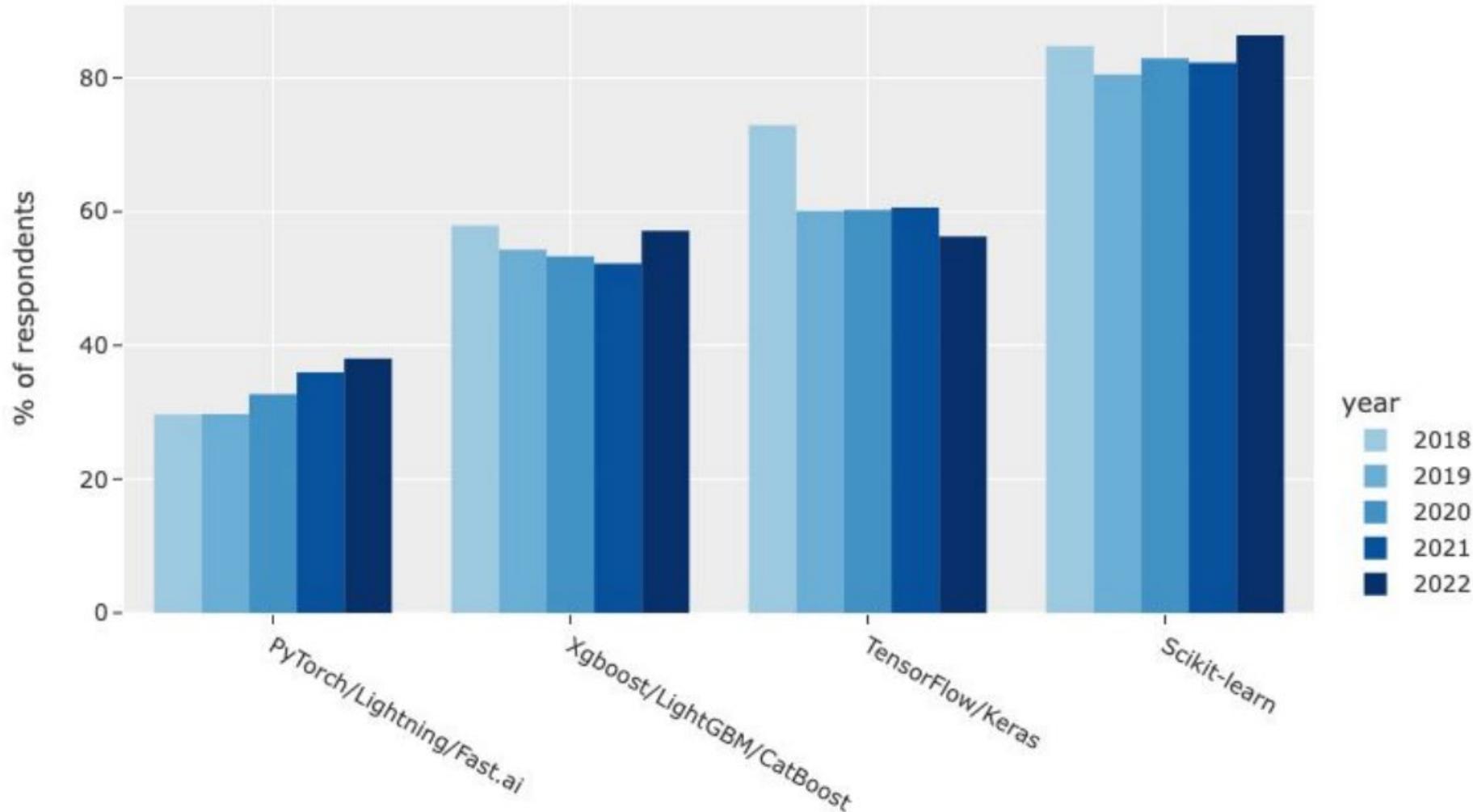
# Kaggle: Making Data Science a Sport

<https://www.kaggle.com/kaggle-survey-2022>



# Kaggle: Making Data Science a Sport

<https://www.kaggle.com/kaggle-survey-2022>



# Getting Started

Great python resources:

A Whirlwind Tour of Python by Jake VanderPlas.

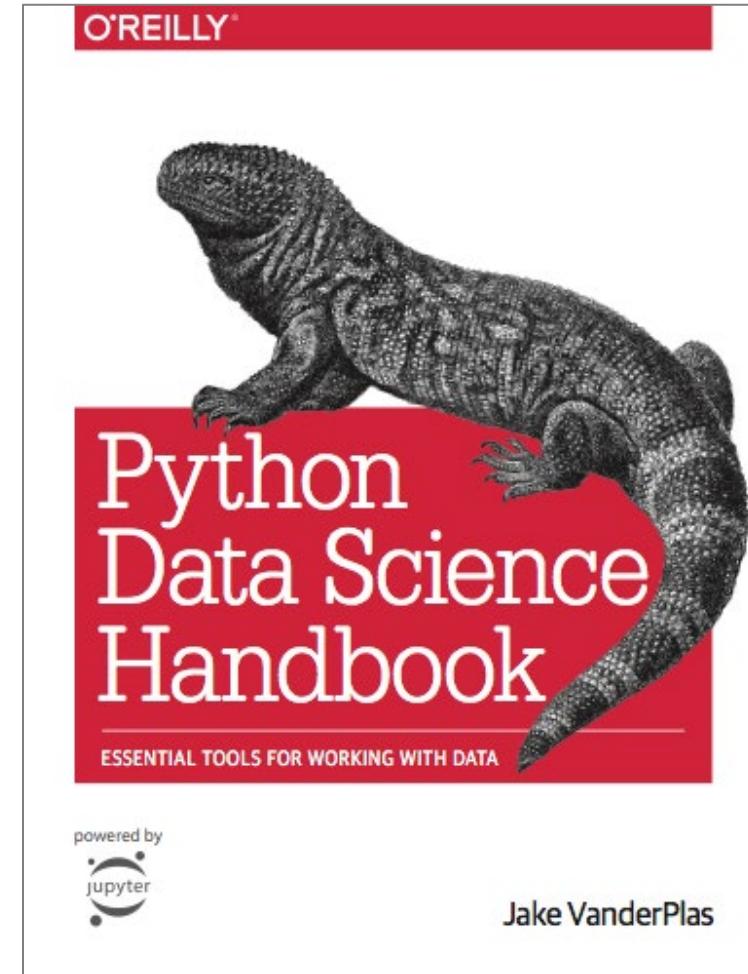
Freely available online at:

<https://jakevdp.github.io/WhirlwindTourOfPython/>

Python Data Science Handbook by Jake VanderPlas.

Freely available online at:

<https://jakevdp.github.io/PythonDataScienceHandbook/>



Part N-1

# **POLICIES**

Doctors: Googling stuff online does not make you a doctor.

Programmers:



Not sharing your source code out of greed



Not sharing your source code out of shame