Creating an Artificial Intelligence to Play Trivia Games

Pedro Rodriguez
CU Boulder Computer Science PhD Student
Artificial Intelligence Group advised by Jordan Boyd-Graber

About Me

- Pedro Rodriguez

  - 1st year C.S. PhD Student at CU
  - Data scientist at Trulia, AMPLab Undergrad Research
  - UC Berkeley 2014 Graduate in Computer Science
  - Research: large scale systems for machine learning
  - Ski, climb, hike, games, open source in free time

University of Colorado Boulder amplab Berkeley UNIVERSITY OF CALIFORNIA trulia

QANTA Project

- Jordan Boyd-Graber, CU Professor of Computer Science
- Mohit Iyyer, PhD Student at University of Maryland
- Pedro Rodriguez
- Hal Daumé III, Anupam Guha, He He, Brianna Satinoff, Manjhunath Ravi, Danny Bouman

University of Colorado Boulder

UNIVERSITY OF MARYLAND

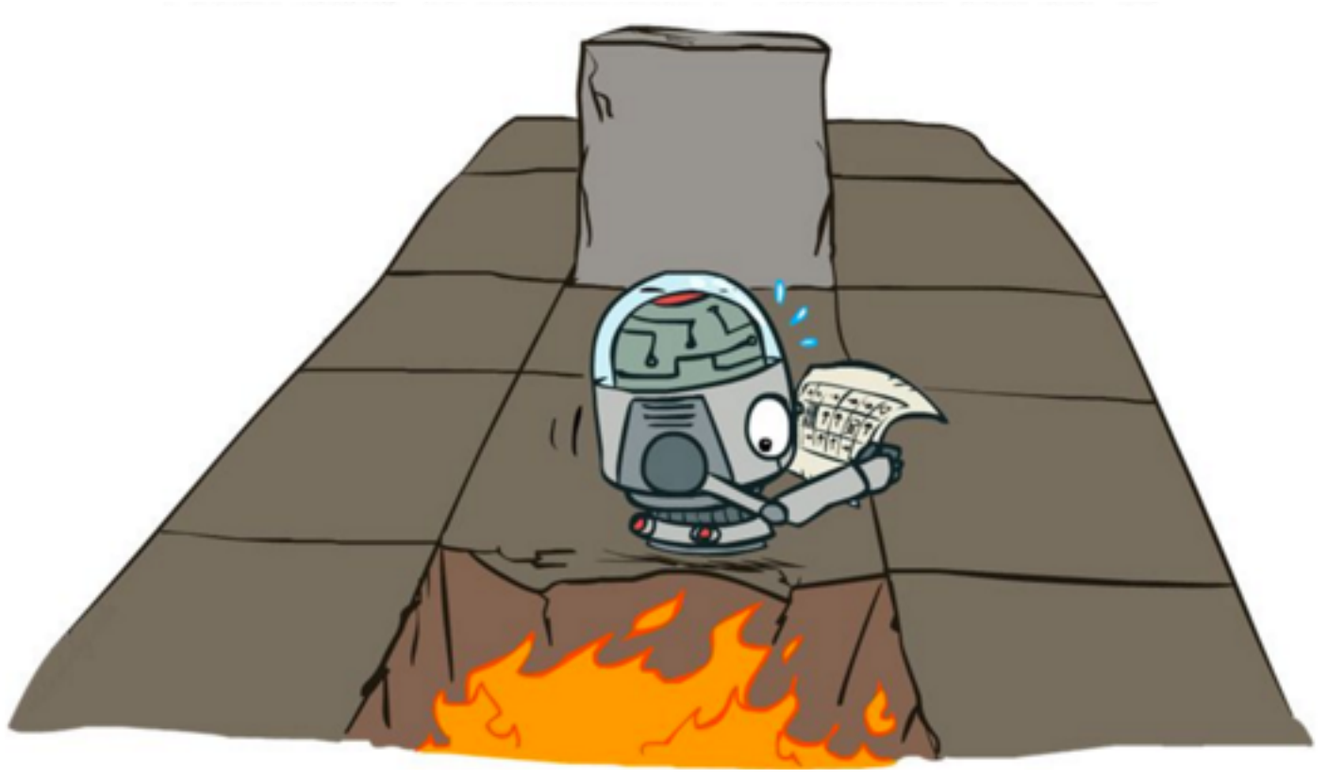
Outline

- Artificial Intelligence and Machine Learning Introduction
- Quiz Bowl Introduction
- Compare IBM Watson and QANTA
- Linear and Logistic Regression
- Incremental Learning
- Words as Numbers and Deep Learning
- Feature Extraction

What is Artificial Intelligence?

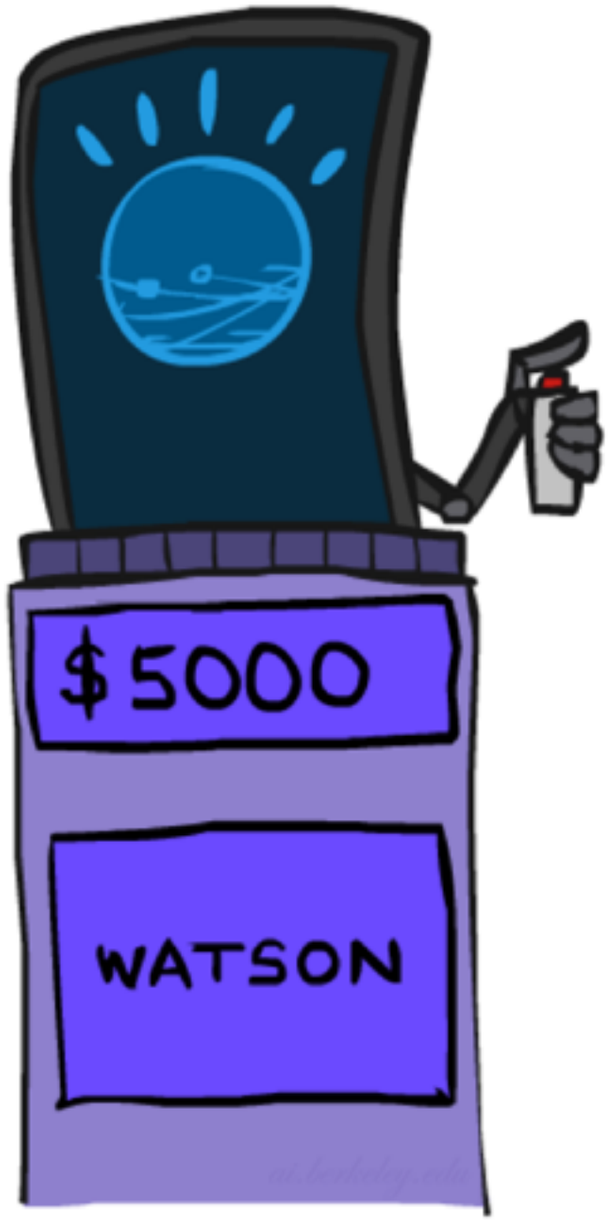
What is Artificial Intelligence?

- Wiki: AI is the **intelligence** exhibited by machines or software
- John McCarthy: science and engineering of making **intelligent** machines
- What is intelligence?
  - Ability to learn or understand the world to make decisions in new or difficult situations

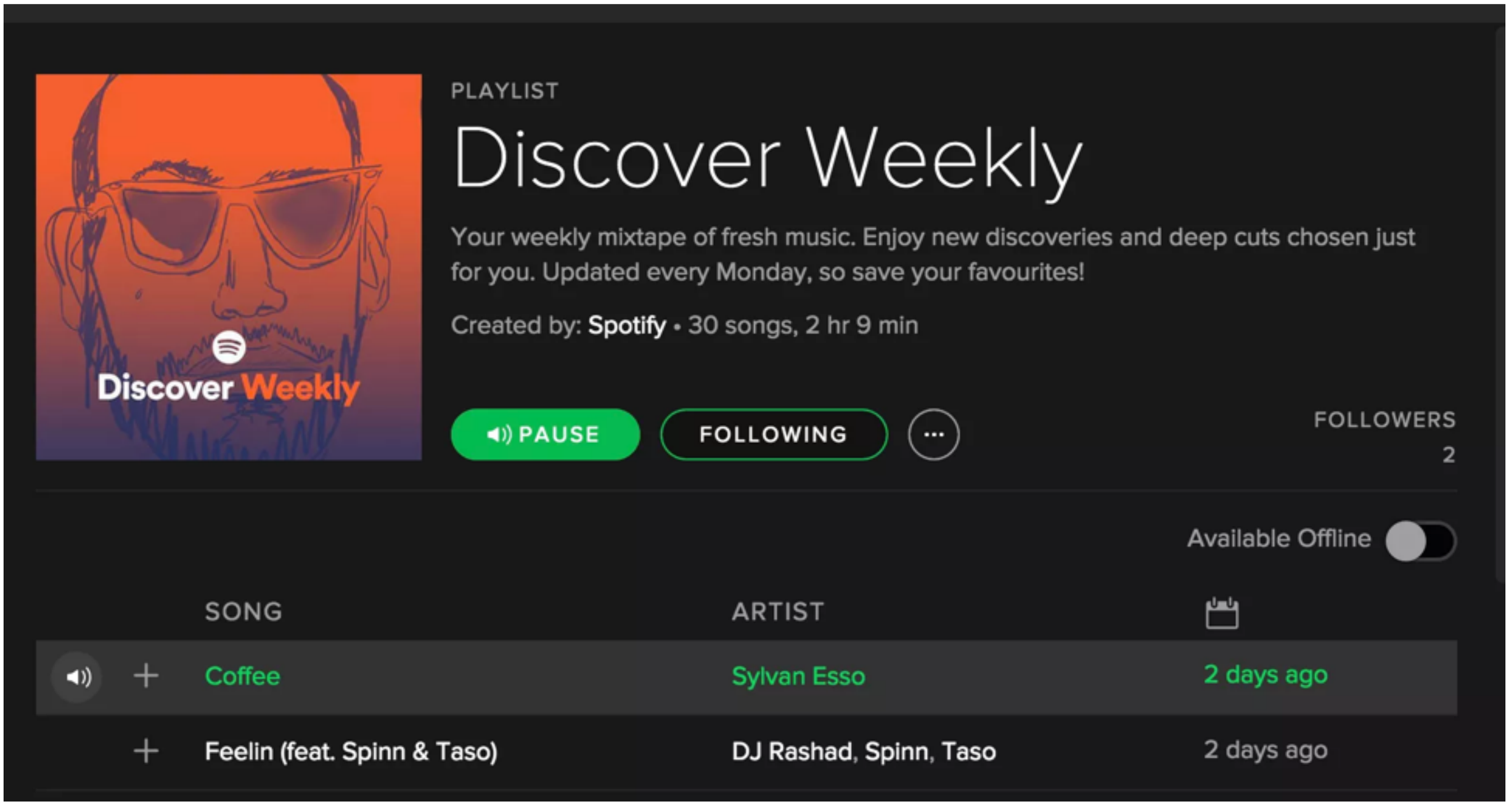


Games: Intelligent Agents

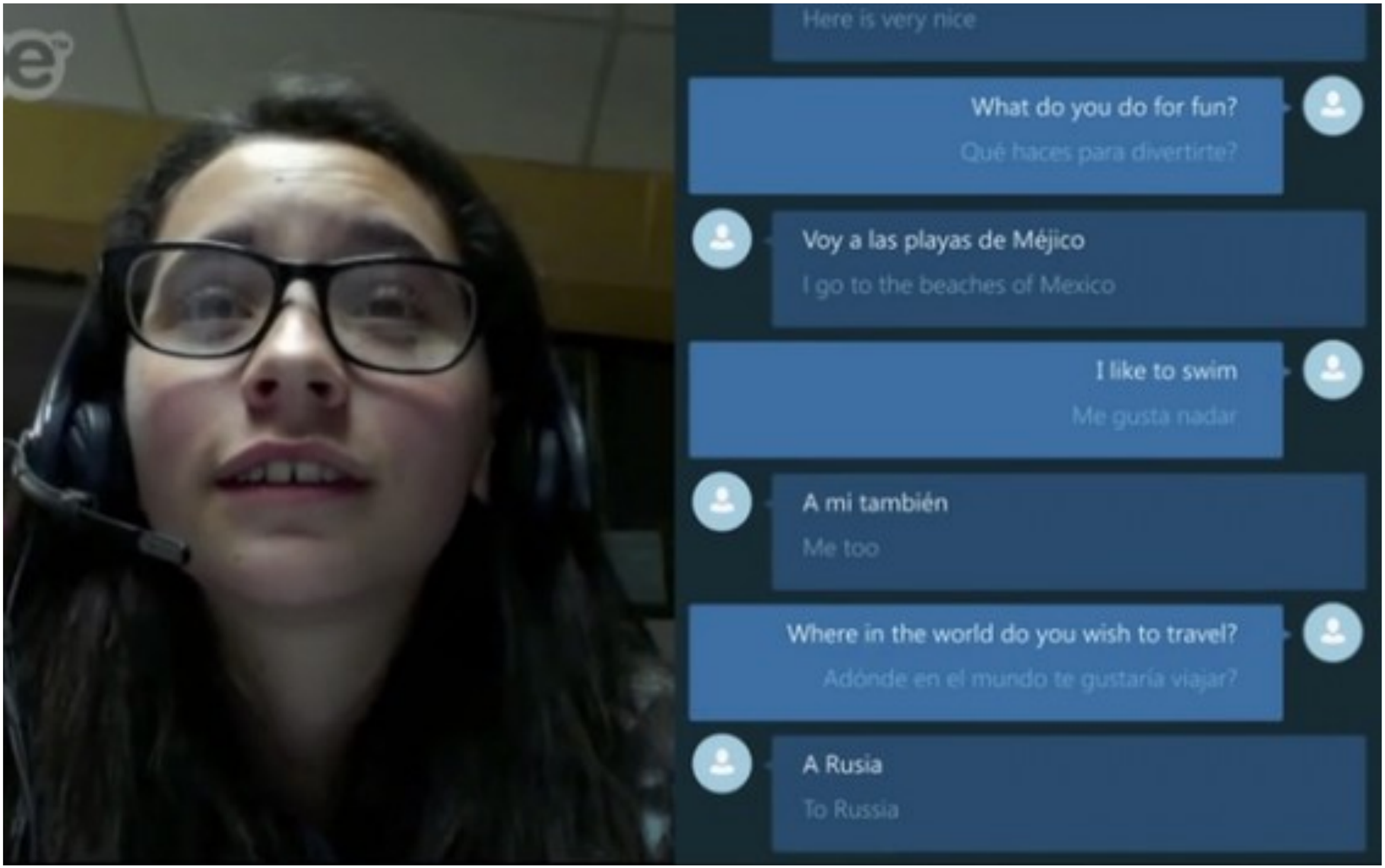
Given knowledge of gameplay and current game
How do you play?



Recommendations: Suggest Music



Machine Translation: Skype

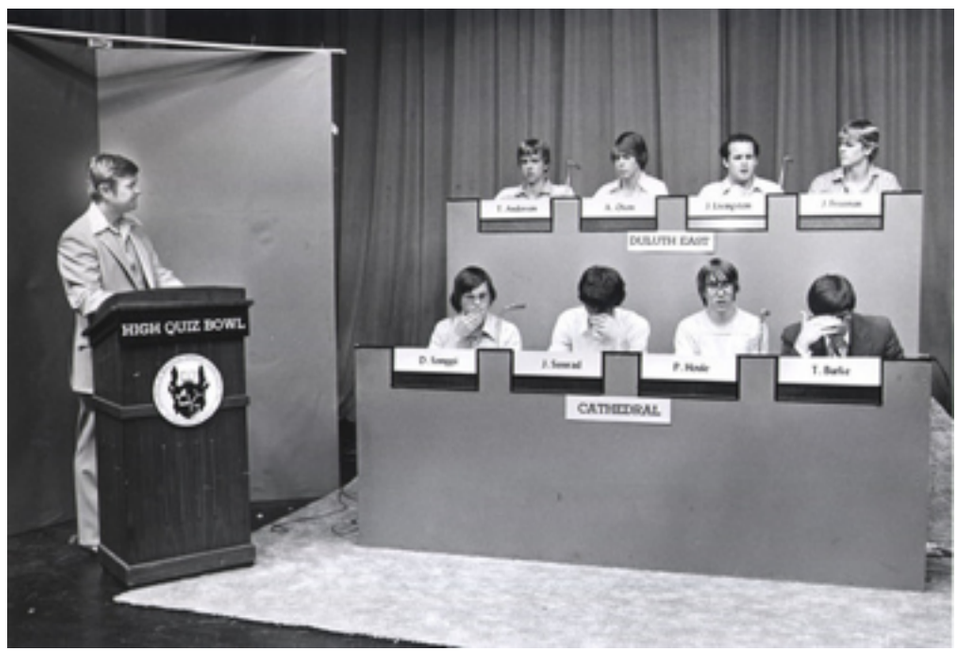


Classification: Spam Email

Machine Learning: design and development of algorithms to evolve behavior based on data

Quiz Bowl

- Two teams play against each other
  - Moderator reads question
  - When team knows the answer “buzz” in
    - Correct guesses award points, wrong guesses let other team see entire question
- Thousands of teams in US



Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

Who is this question talking about?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

- Albert Einstein

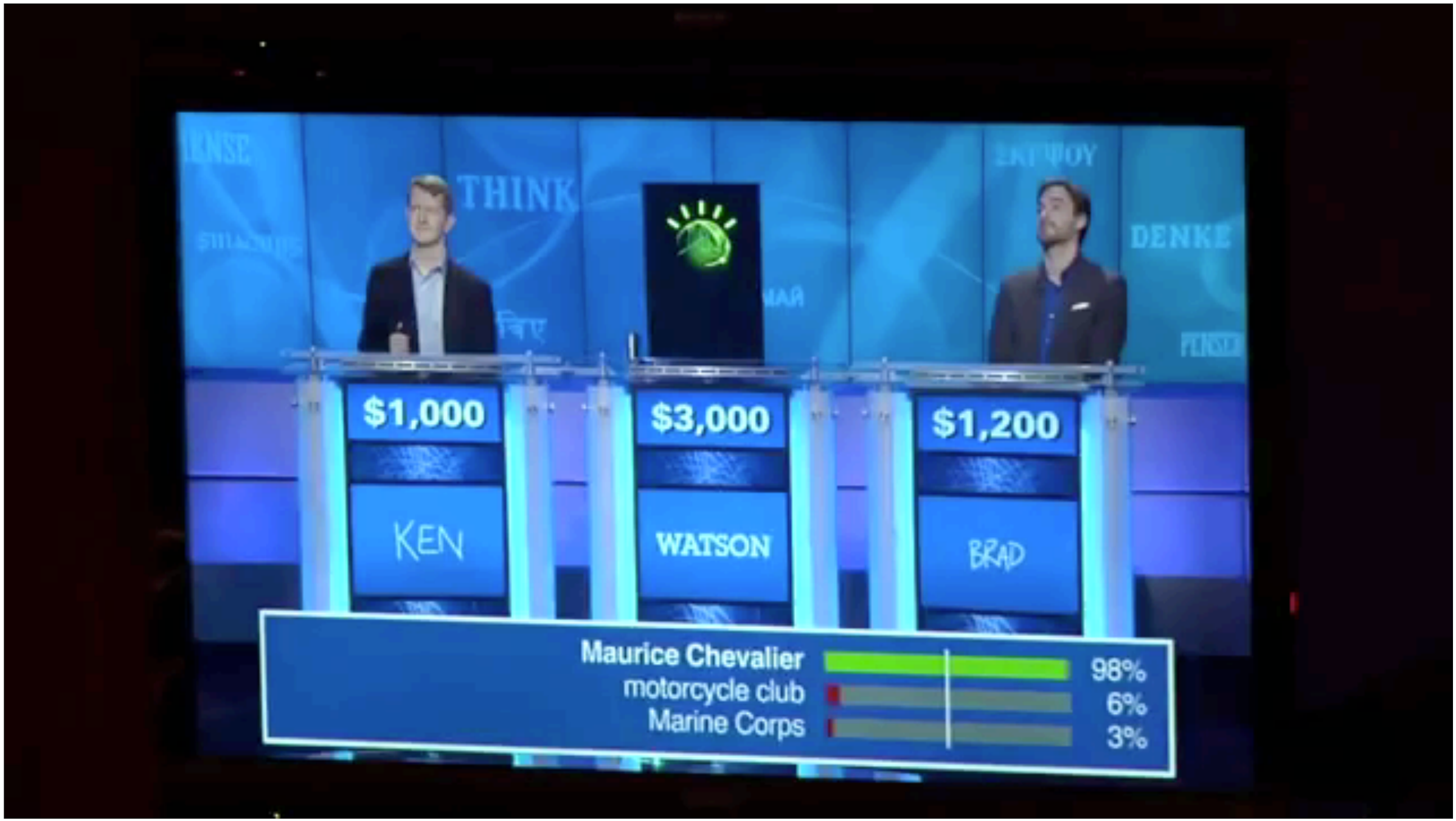
Pyramidal Clues

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

- Albert Einstein

Quiz Bowl vs Jeopardy? IBM Watson vs QANTA?

IBM Watson


THINK
DENKE
$1,000
$3,000
$1,200
KEN
WATSON
BRAD
Maurice Chevalier
motorcycle club
Marine Corps
98%
6%
3%

QANTA: **Q**uestion **A**nswering is **N**ot a **T**rivial **A**ctivity

Similarities and Differences

- Differences
  - Jeopardy: answer questions only at the end
  - Quiz Bowl: decide after each word
- Quiz Bowl is pyramidal
- Humans think more like QANTA than Watson
  - **Why?**

Quiz Bowl Task

- As a sentence is read word per word
  - Track the best guess
  - When the best guess is “good enough”, buzz in

QANTA Overview

Deep Learning

Vector

Guesser

Guesses	Score
Einstein	0.4
Heisenburg	0.3
Alan Turing	0.002
CU Boulder	0.0001

Top Guess

Question

Features

Language

Yes/No

Mentions

Vector

Vector

Logistic Regression

Action

Buzz

Wait

QANTA Overview

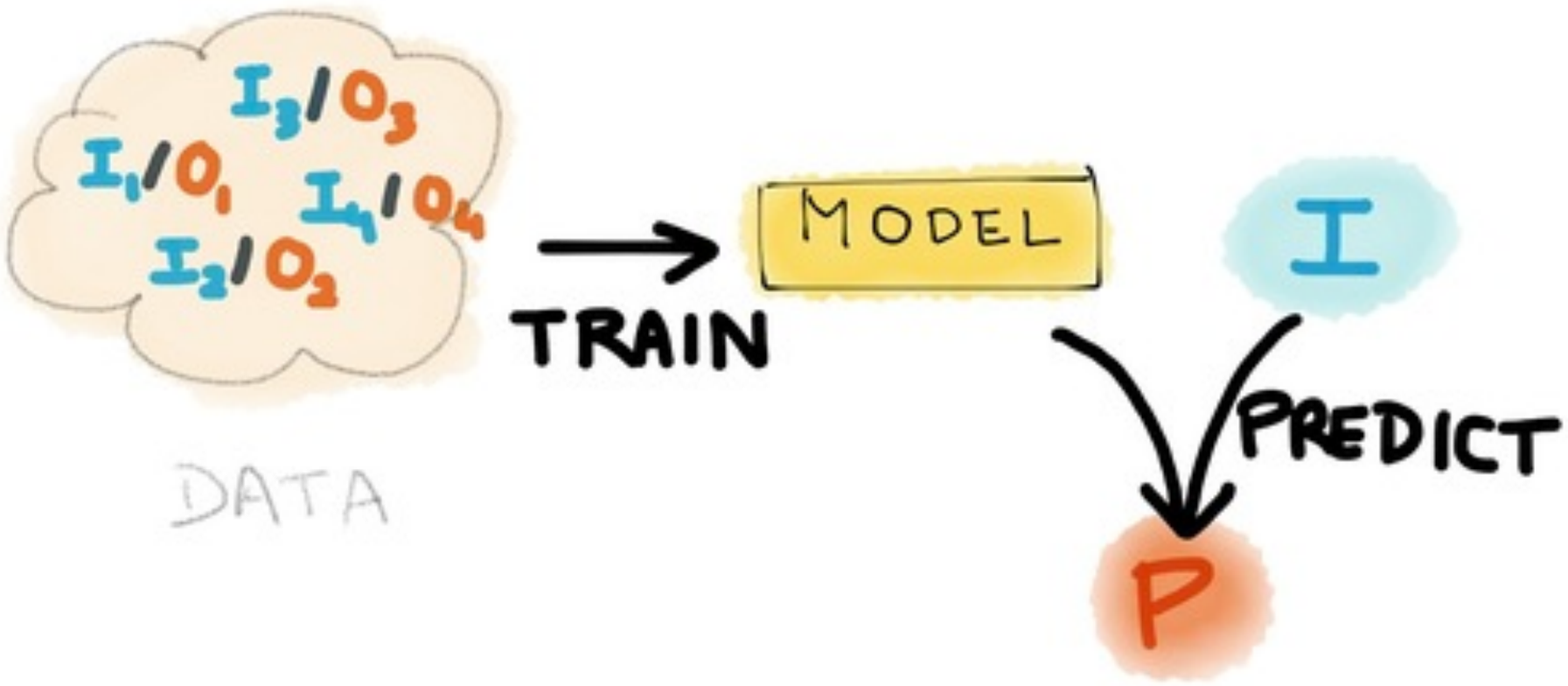
Datasets

- Quiz Bowl Questions
  - ~200,000 questions
  - Many questions per answer
  - Used to generate guesses
- Wikipedia
  - Help to decide when to answer
  - 50GB Decompressed (text)

WIKIPEDIA
The Free Encyclopedia

Machine Learning: Data based approach

- Given data X, predict Y
  - Given a question, predict the answer
  - Lets see an example of one ML algorithm that we will use later too



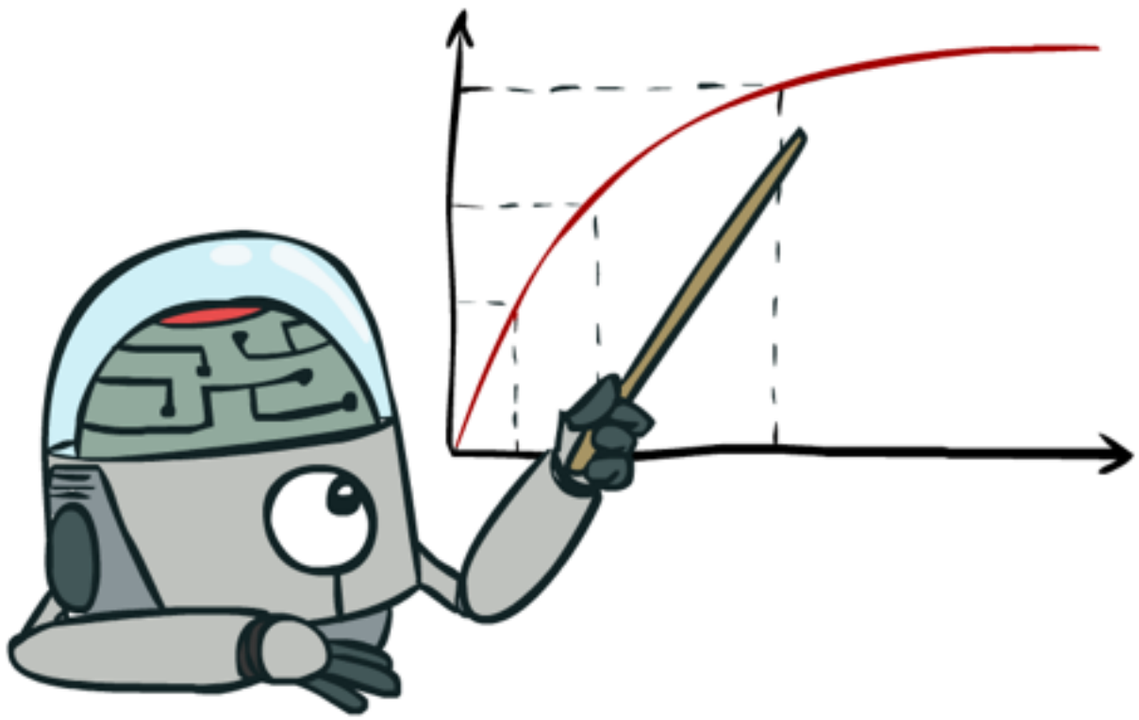
Linear Regression: aka Ax+b=y

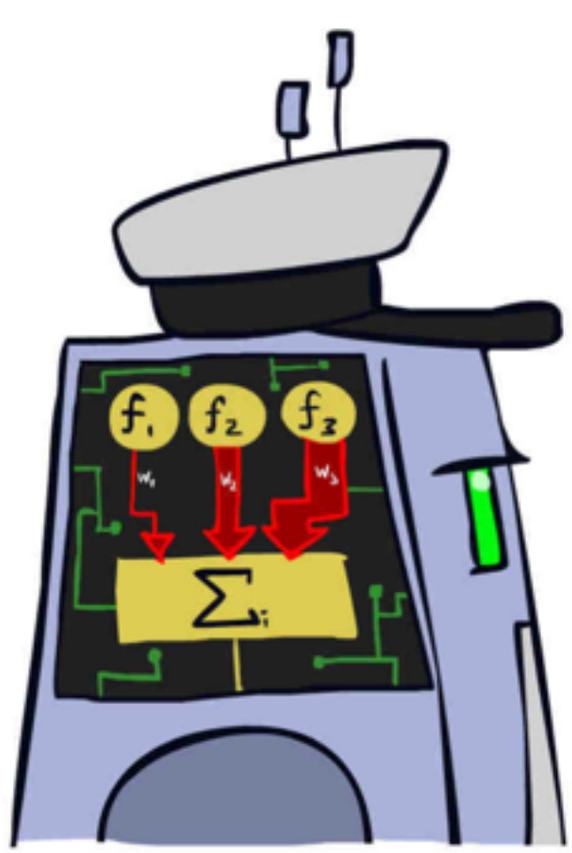
Prediction

Data

Weights

$$y = h_\Theta(x) = \sum_i \Theta_i x_i = \Theta^T x$$





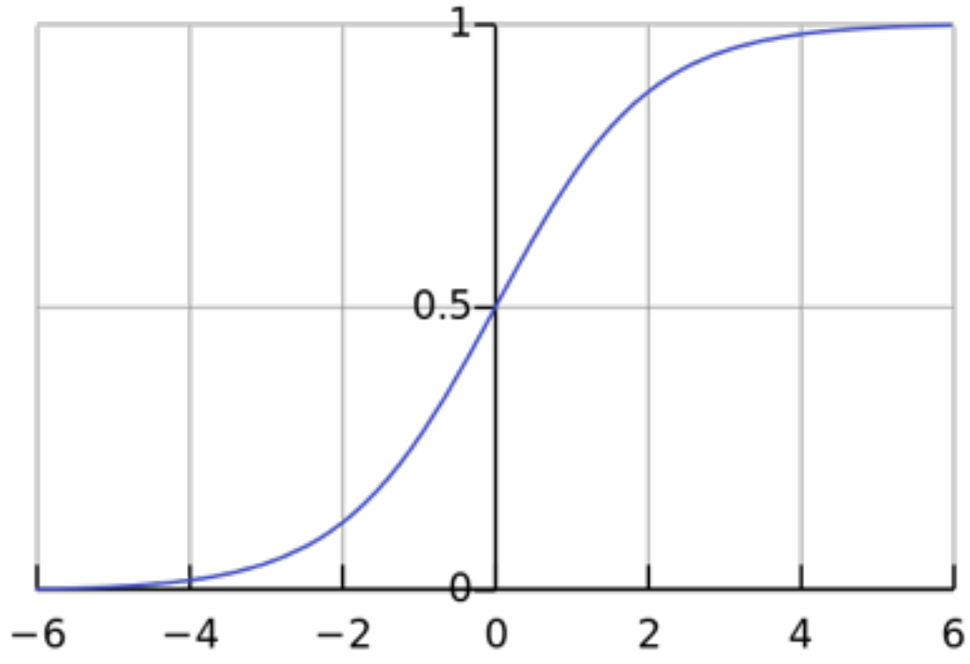
Logistic Regression: binary decision

- EG: Given medical tests X, predict Y=patient has cancer

$$h_\Theta(x) = \frac{1}{1 + e^{-\Theta^T x}} \equiv \sigma(\Theta^T x)$$

Sigmoid/Logistic Function

$$\sigma(z) \equiv \frac{1}{1 + e^{-z}}$$



Logistic Regression

$$P(y = 1|x) = h_\Theta = \sigma(\Theta^T x)$$

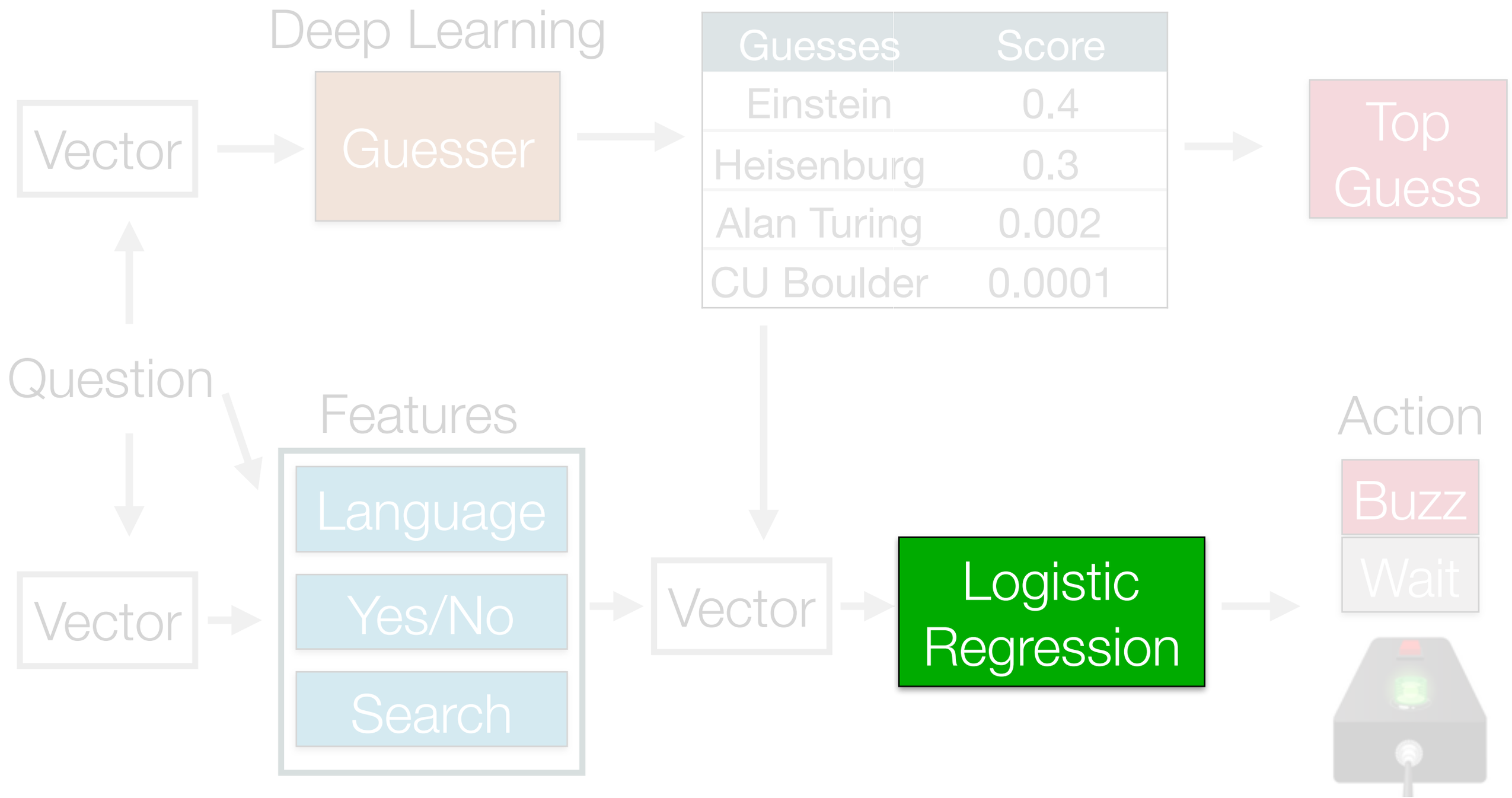
$$P(y = 0|x) = 1 - P(y = 1|x)$$

Given the data X, which is more likely?

$$P(y = 1|x) > .5$$ Output 1

$$P(y = 1|x) < .5$$ Output 0

QANTA Overview



Incremental Learning

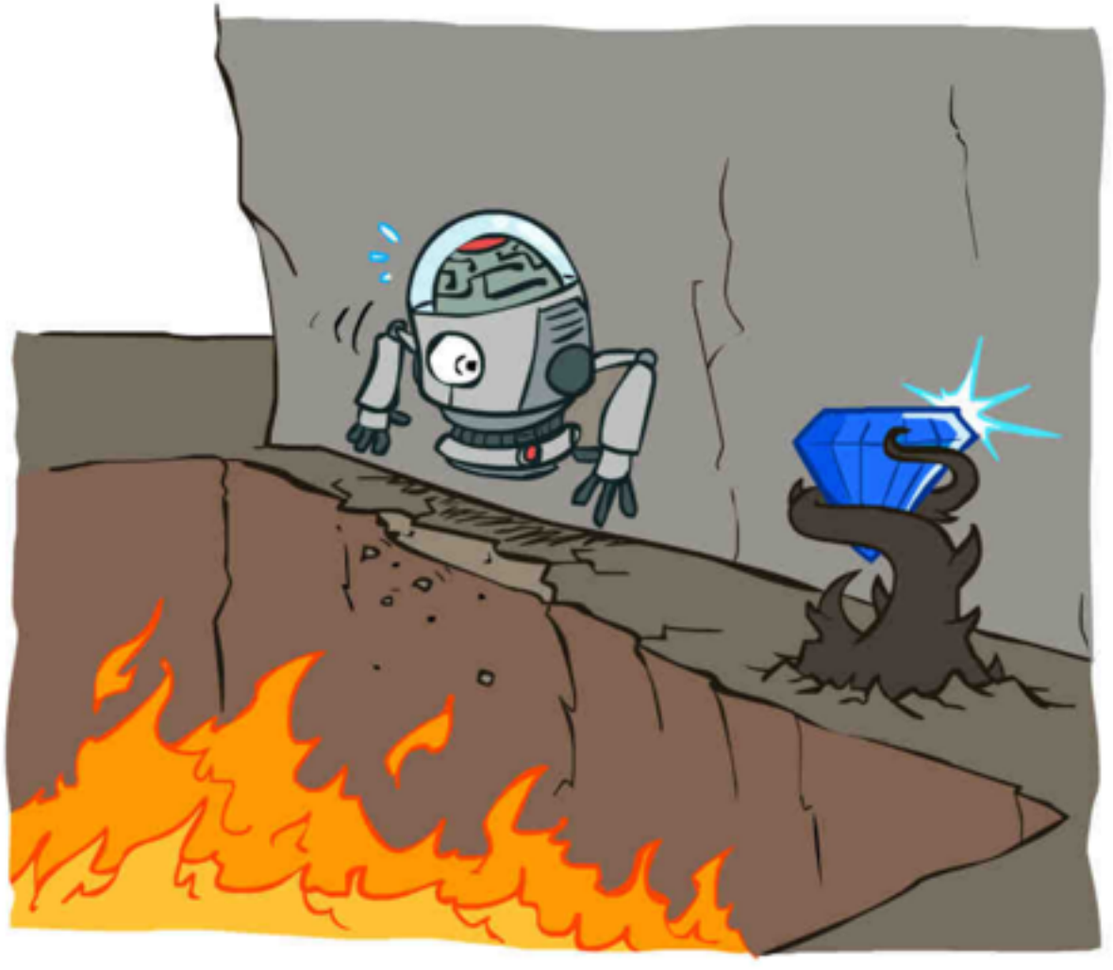
- These algorithms assume **all** of **X** is **known** at **once**
- How does Quiz Bowl differ?
  - Receive input X “incrementally”, one word at a time
  - Quiz Bowl Task doesn’t cleanly fit into single prediction
  - How do we deal with that?

Quiz Bowl Task

- Overall: return the correct answer as soon as possible
- Break the problem down
  - What should we answer with?
  - When should we answer the question?
- Treat as Markov Decision Process

Markov Decision Process

- Framework for modeling decision-making in situations where outcomes are partly random and partly under control of decision-maker



Quiz Bowl Markov Decision Process

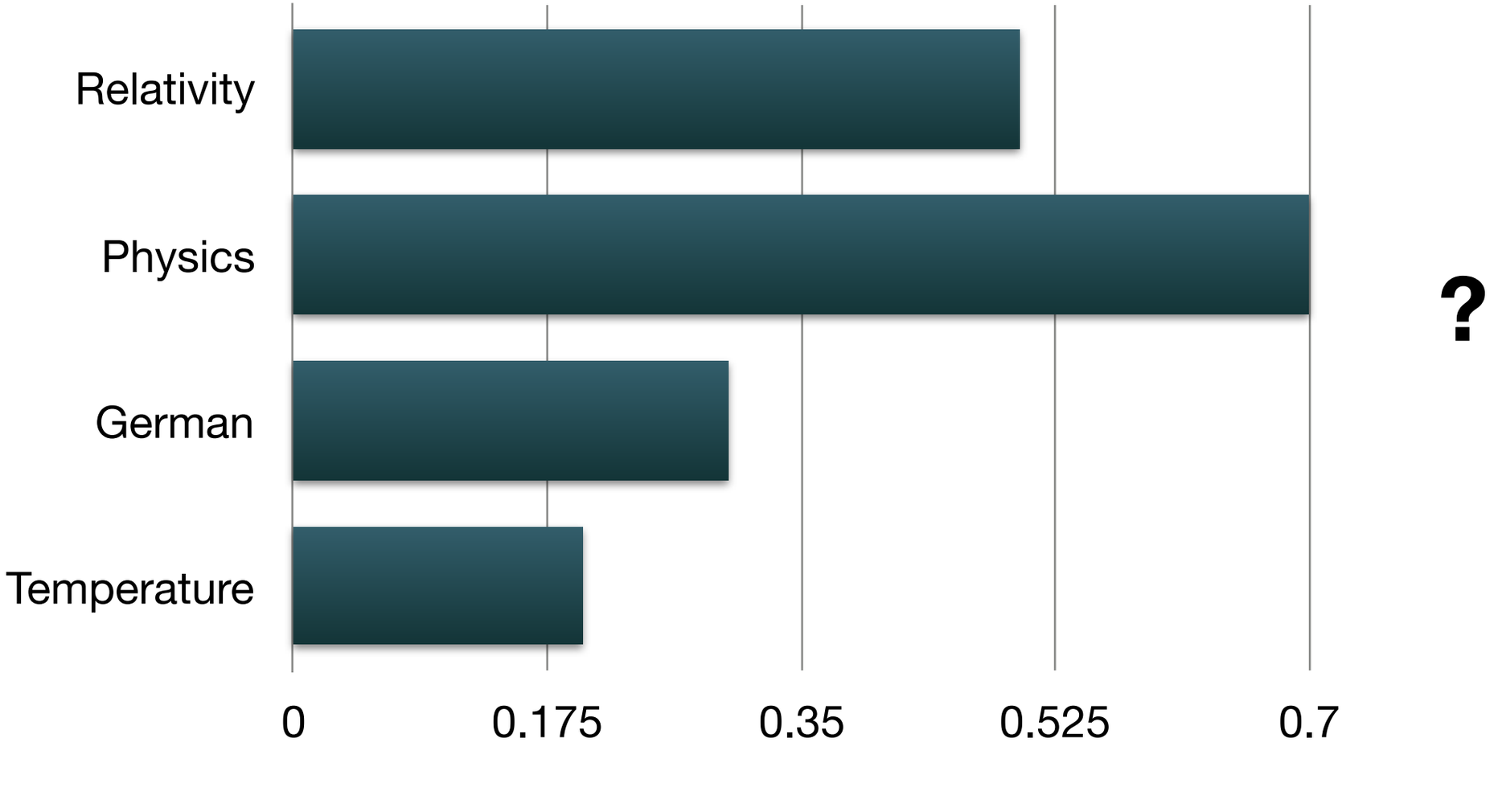
- We have control of when we can buzz
- We have control of what we answer
- Don’t have control over when opponent answers
- Don’t have control over quality of next “clue”
- Do I risk letting opponent answer for more information?

Two Steps to Answering Questions

- Given a question
  - Generate a set of **guesses** (deep learning)
  - **Buzz** if confident (features + classifier)
- Deep learning?
- Features? Classifiers (Logistic Regression)?

Can we represent text numerically?

- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?



Basic Vector Model

Text Vector

$$[0, \ldots, 1, \ldots, 0]$$

- Represent text as a |V| size vector (vocabulary size)
- 1s mark word presence, 0s absence
- Problems?
  - Sparse, wasteful representation
  - No notion of similarity of words with each other

Context Matters!

- From Wikipedia (with some editing)
  - "Albert **Einstein** developed the general **theory** of **relativity**"
  - "This led to the development of Einstein's special **theory** of **relativity**"
- For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?
- The wiki page on Einstein is pretty similar to the question so the "distance" between the vectors should be small

Context Dependent Vectors

- "You shall know a word by the company it keeps" -J. R. Firth
- Track co-occurrence of words
- Problem: expensive to compute
- Recent solution: word2vec

government debt problems turning into **banking** crises as has happened in

saying that Europe needs unified **banking** regulation to replace the hodgepodge

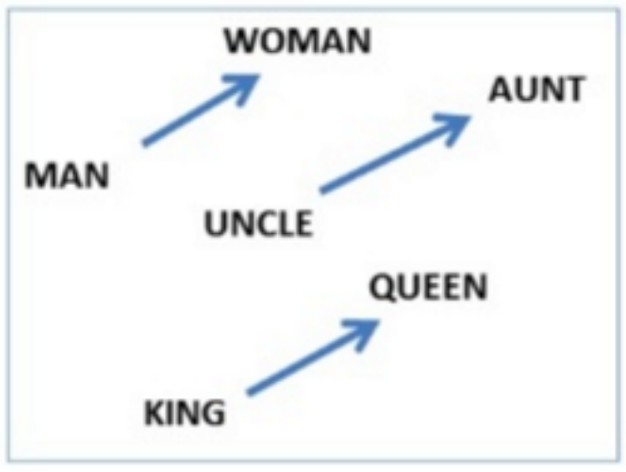
↖ These words will represent *banking* ↗

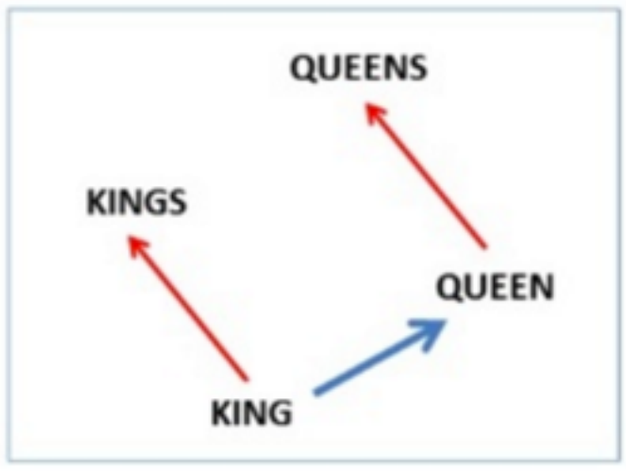
Word2Vec (Mikolov 2013)

- Idea: sliding “context” window around current center word
- Maximize probability of any context given center word

$$J(\Theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{-c,j\neq 0} \log p(w_{t+j}|w_t)$$

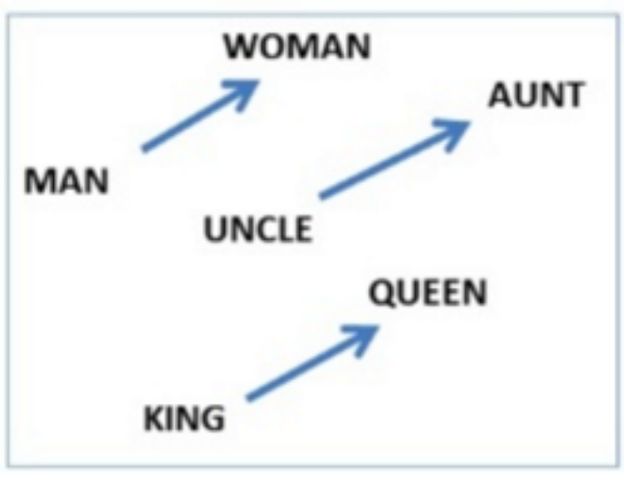
$$vec(\text{"}king\text{"}) - vec(\text{"}man\text{"}) + vec(\text{"}woman\text{"}) \approx vec(\text{"}queen$$

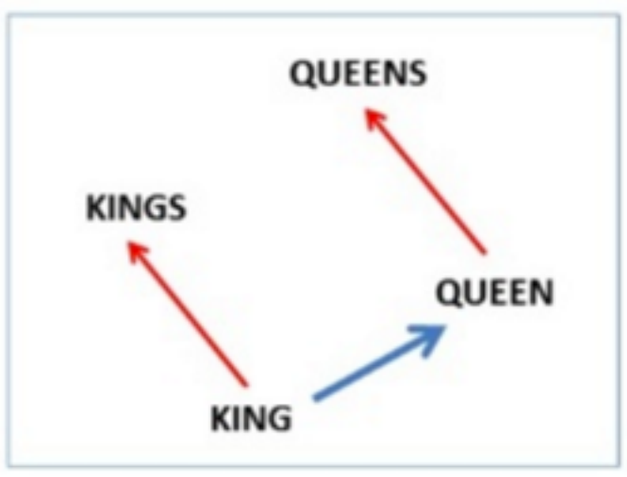




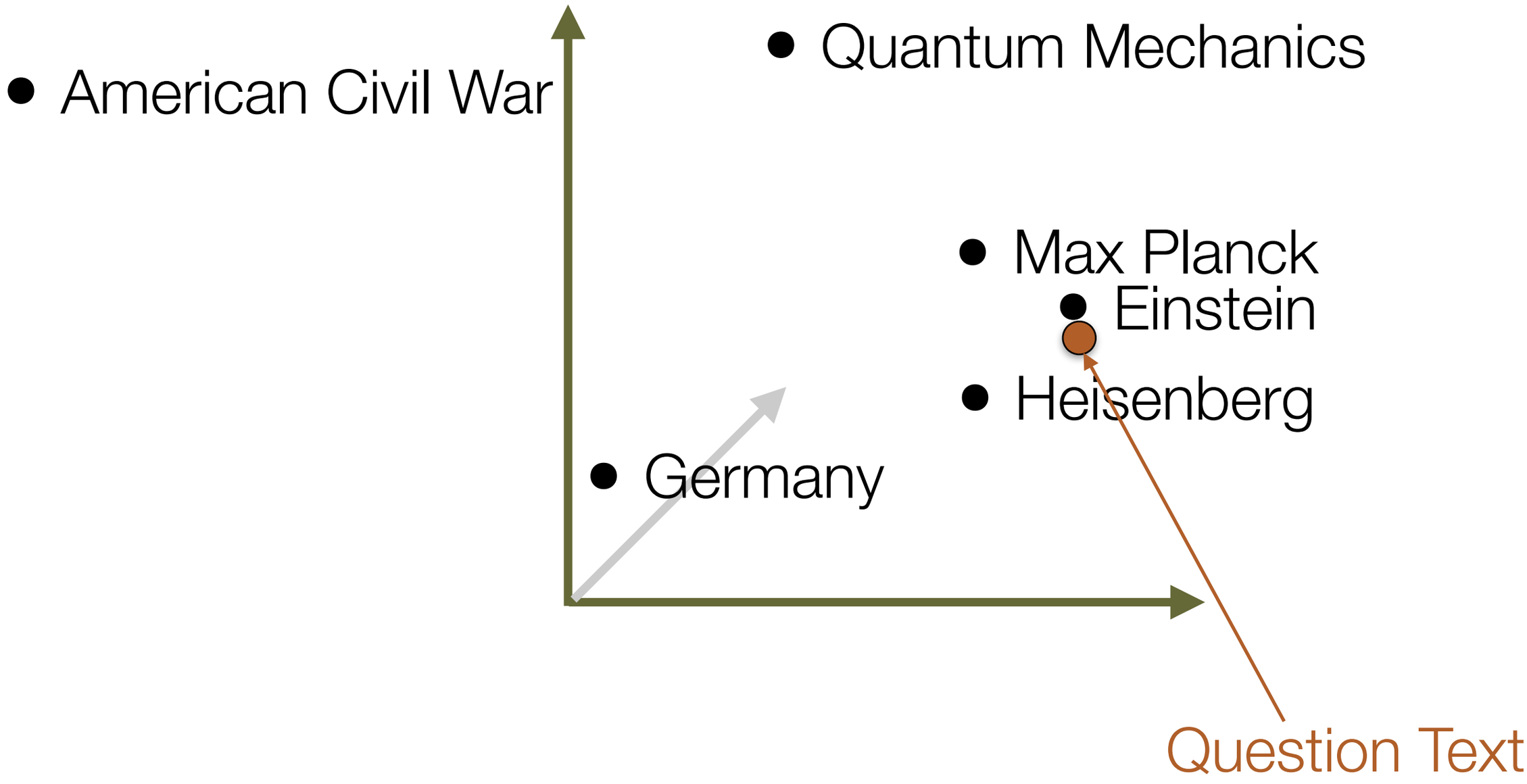
Word2Vec (Mikolov 2013)

$$vec(\text{"}king\text{"}) - vec(\text{"}man\text{"}) + vec(\text{"}woman\text{"}) \approx vec(\text{"}queen\text{"})$$



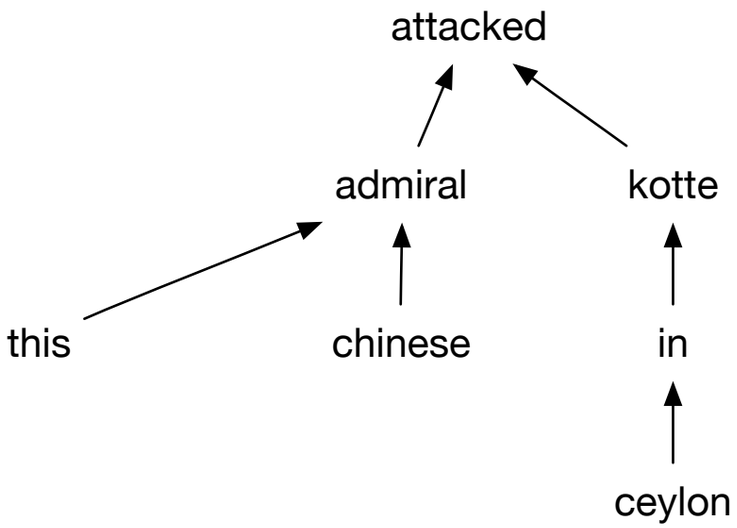


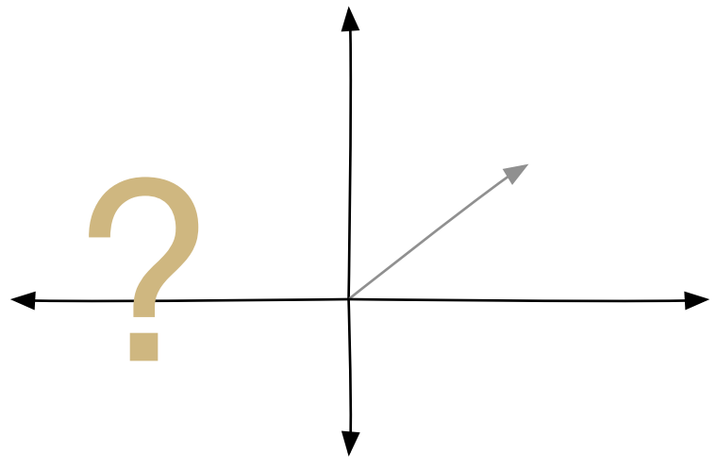
QANTA Embedding



Quiz Bowl Embedding

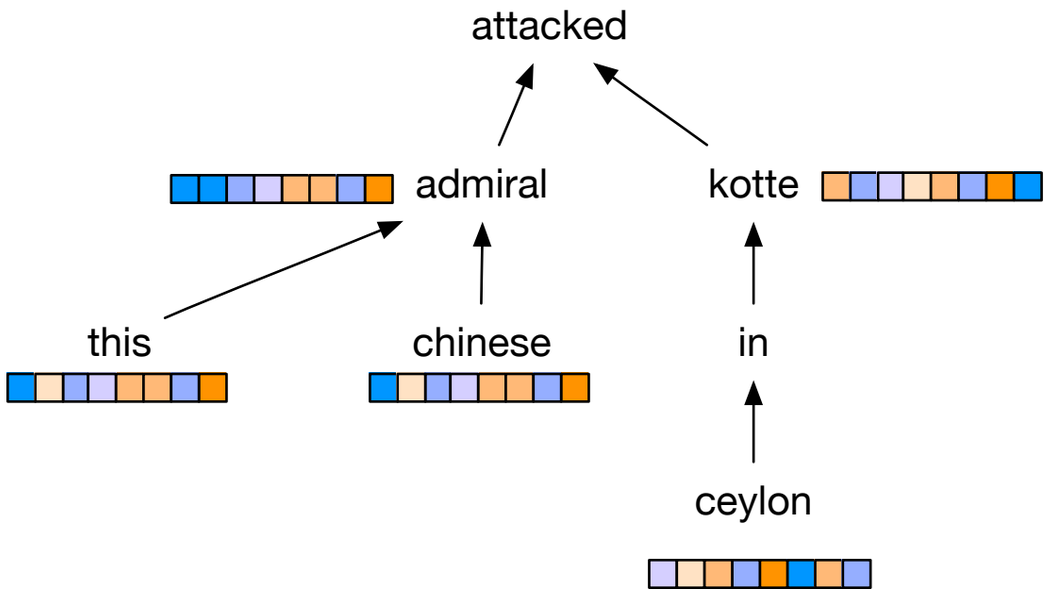
This Chinese admiral attacked motte in ceylon

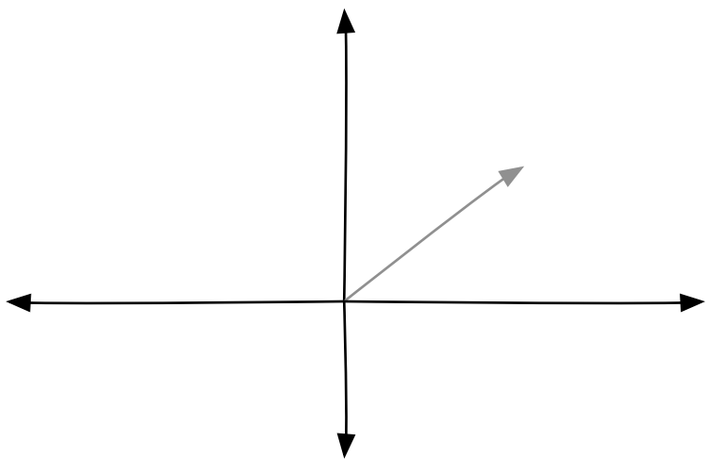




Quiz Bowl Embedding

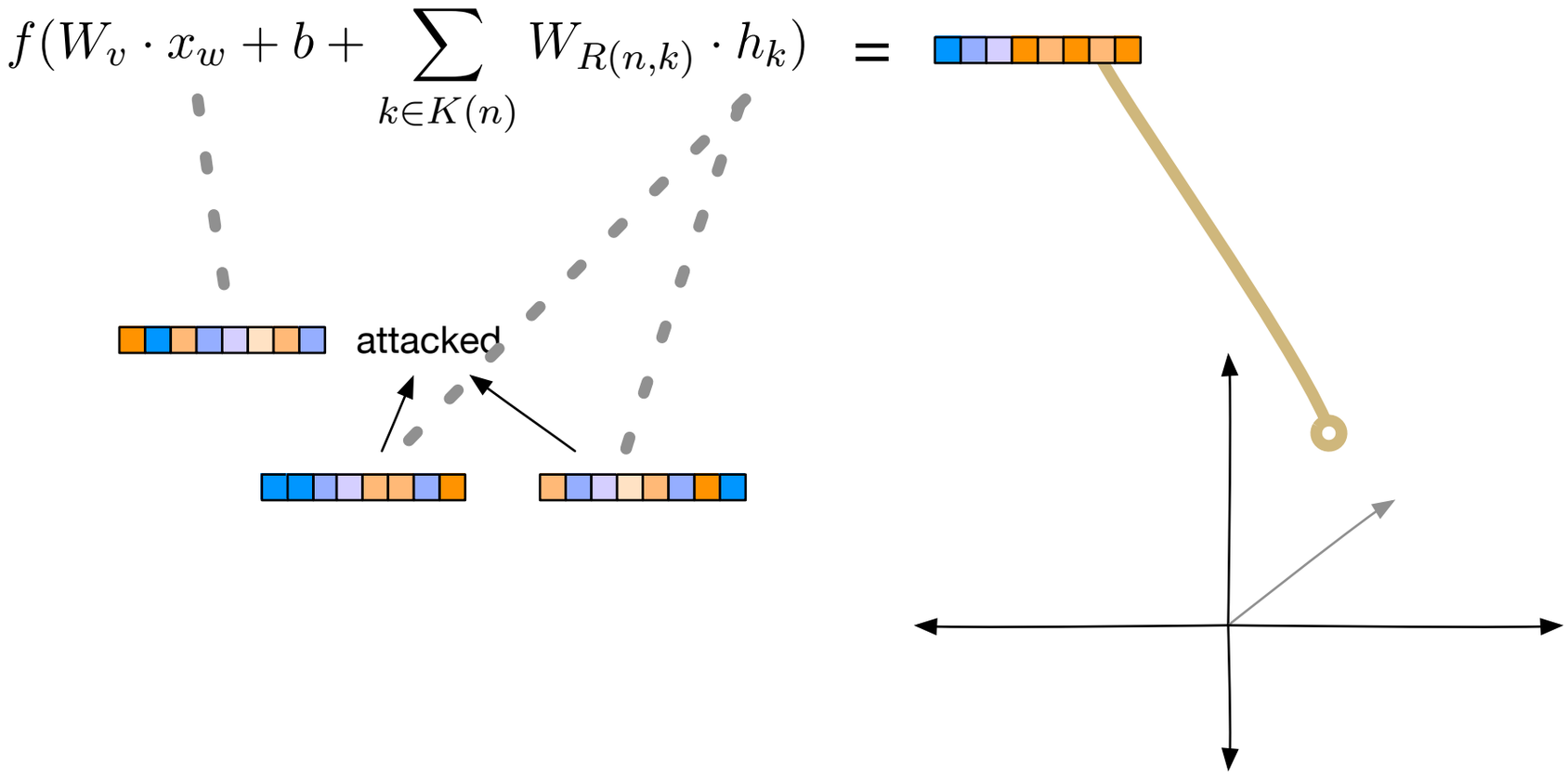
This Chinese admiral attacked motte in ceylon





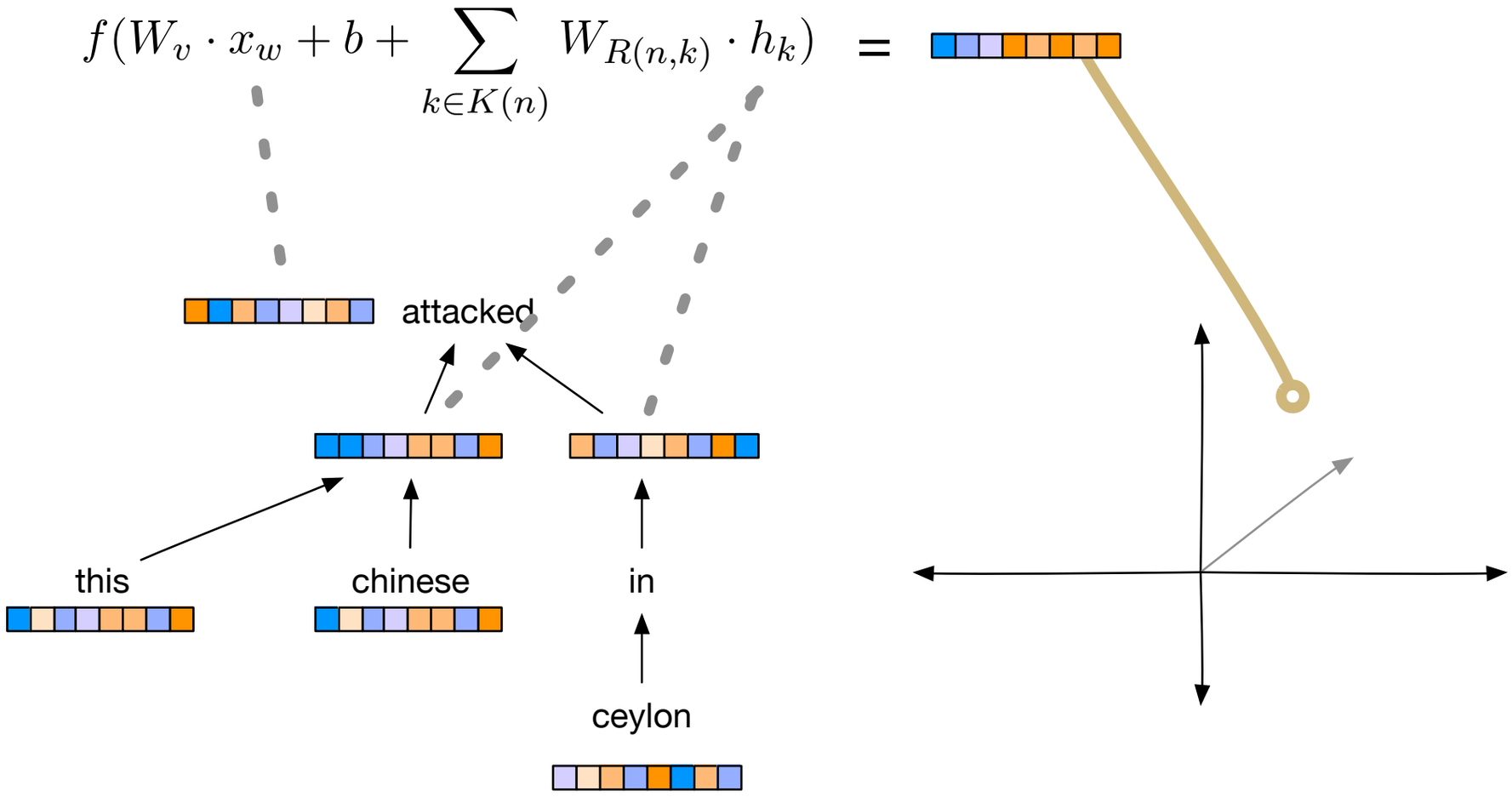
Quiz Bowl Embedding

This Chinese admiral attacked motte in ceylon

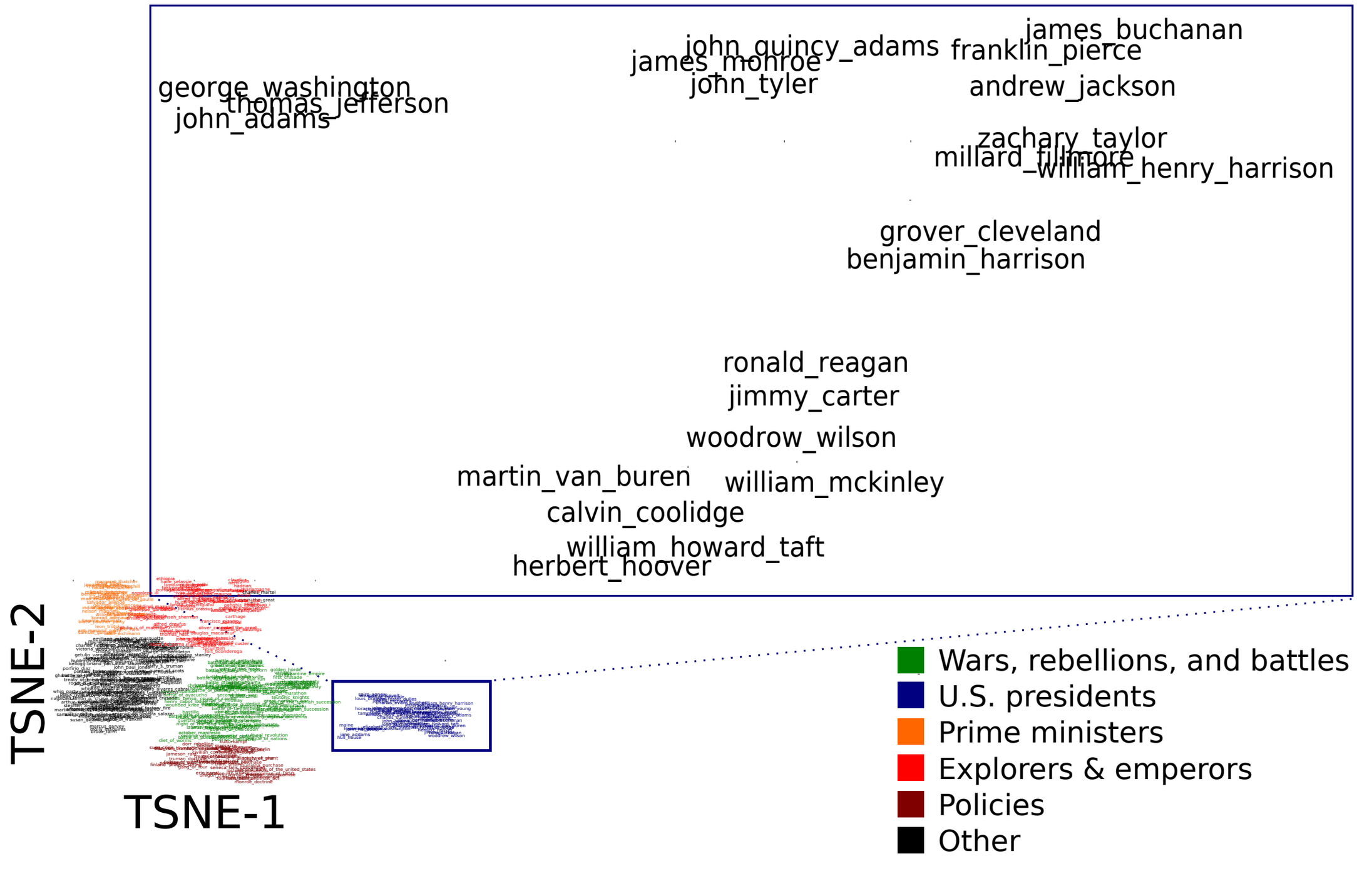


Quiz Bowl Embedding

This Chinese admiral attacked motte in ceylon



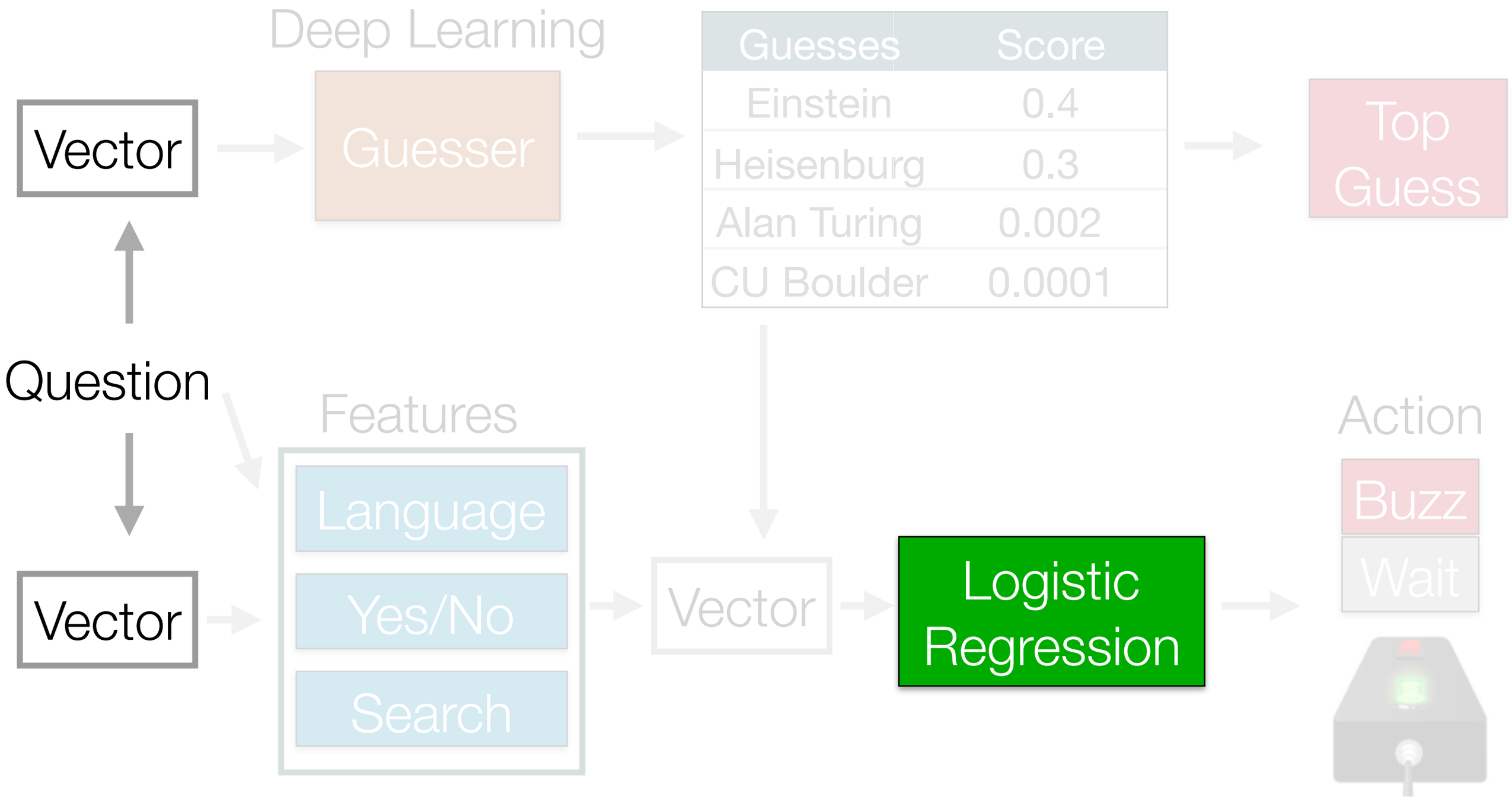
QANTA Vector Space



DAN: Deep Averaging Network

- This works for one sentence, but a paragraph?
- Idea: average the vectors together, then use deep learning to rank guesses
- Benefits
  - Preserves clues from prior sentences
  - Performs well
- Deep Learning next

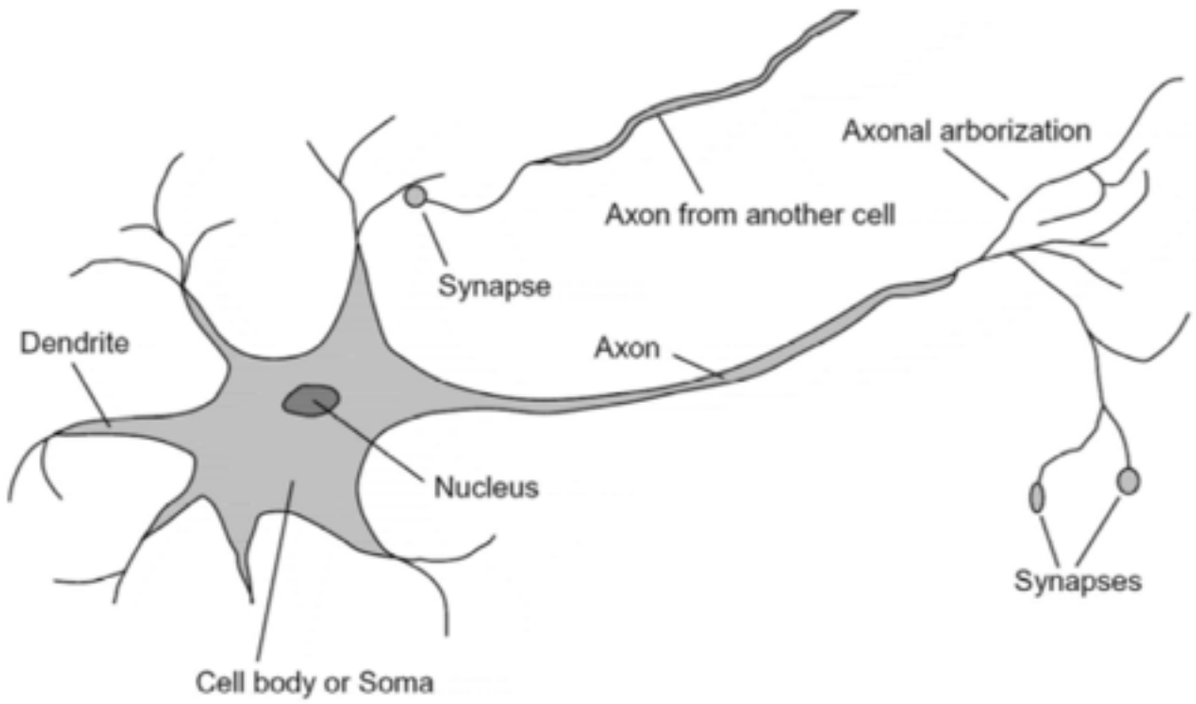
QANTA Overview

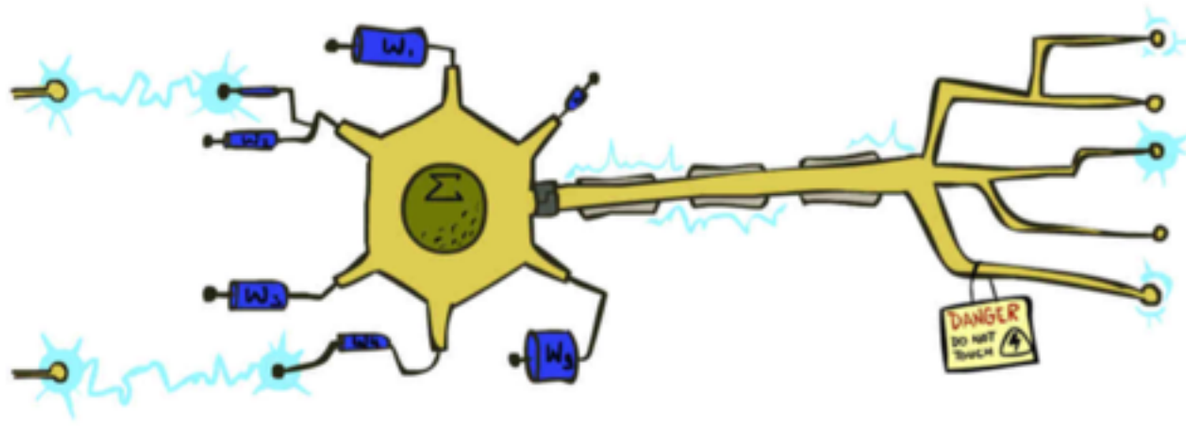


Perceptron (like Logistic Regression)

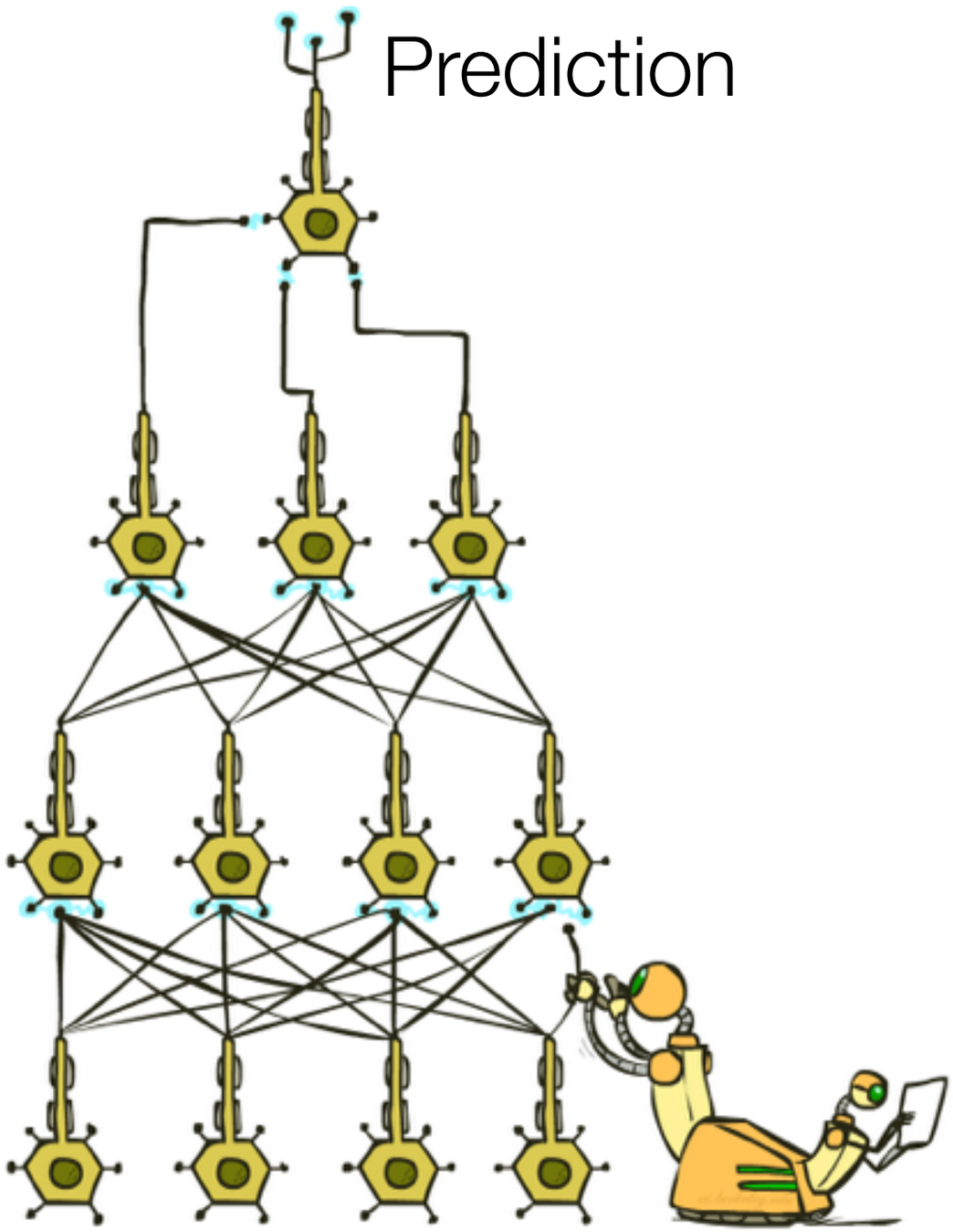
$$P(y=1|x) = h_\Theta = \sigma(\Theta^T x)$$

$$P(y=0|x) = 1 - P(y=1|x)$$

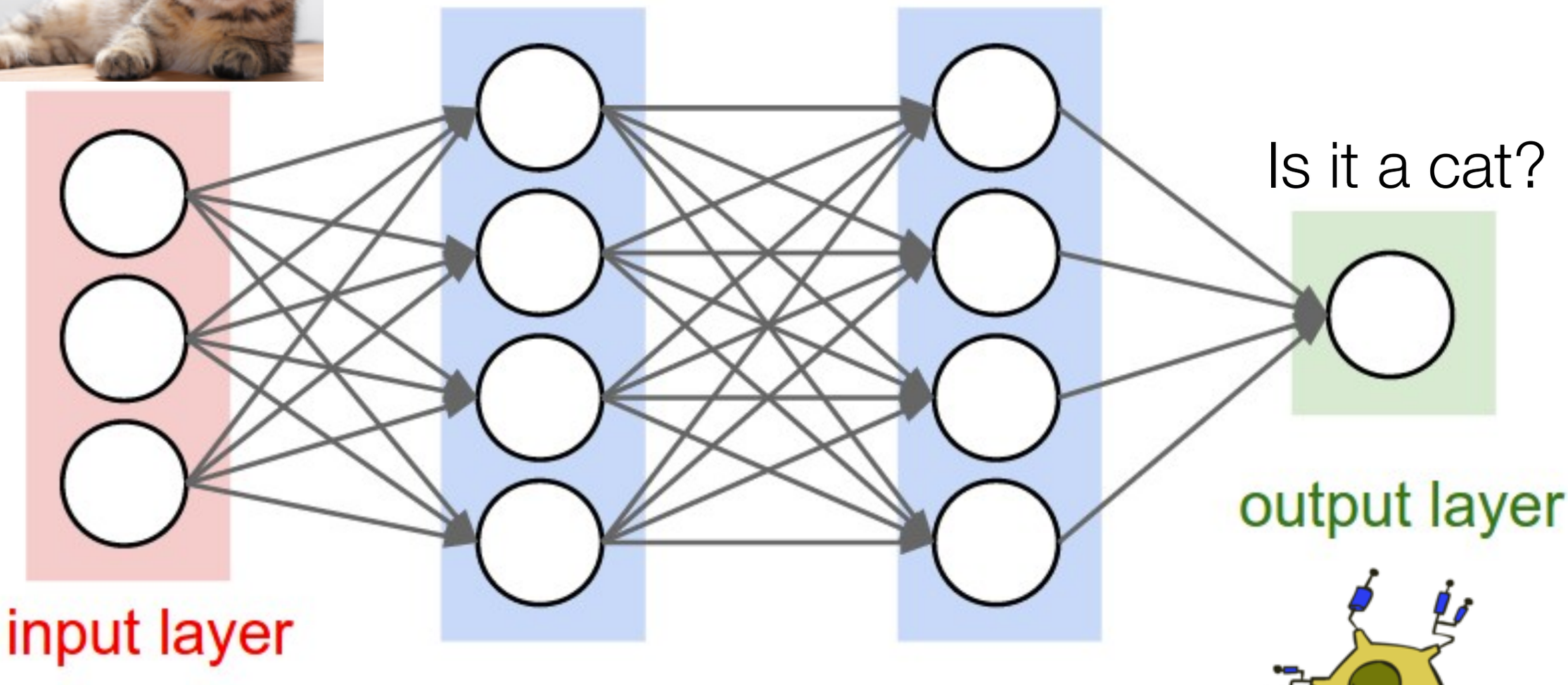




Neural Network

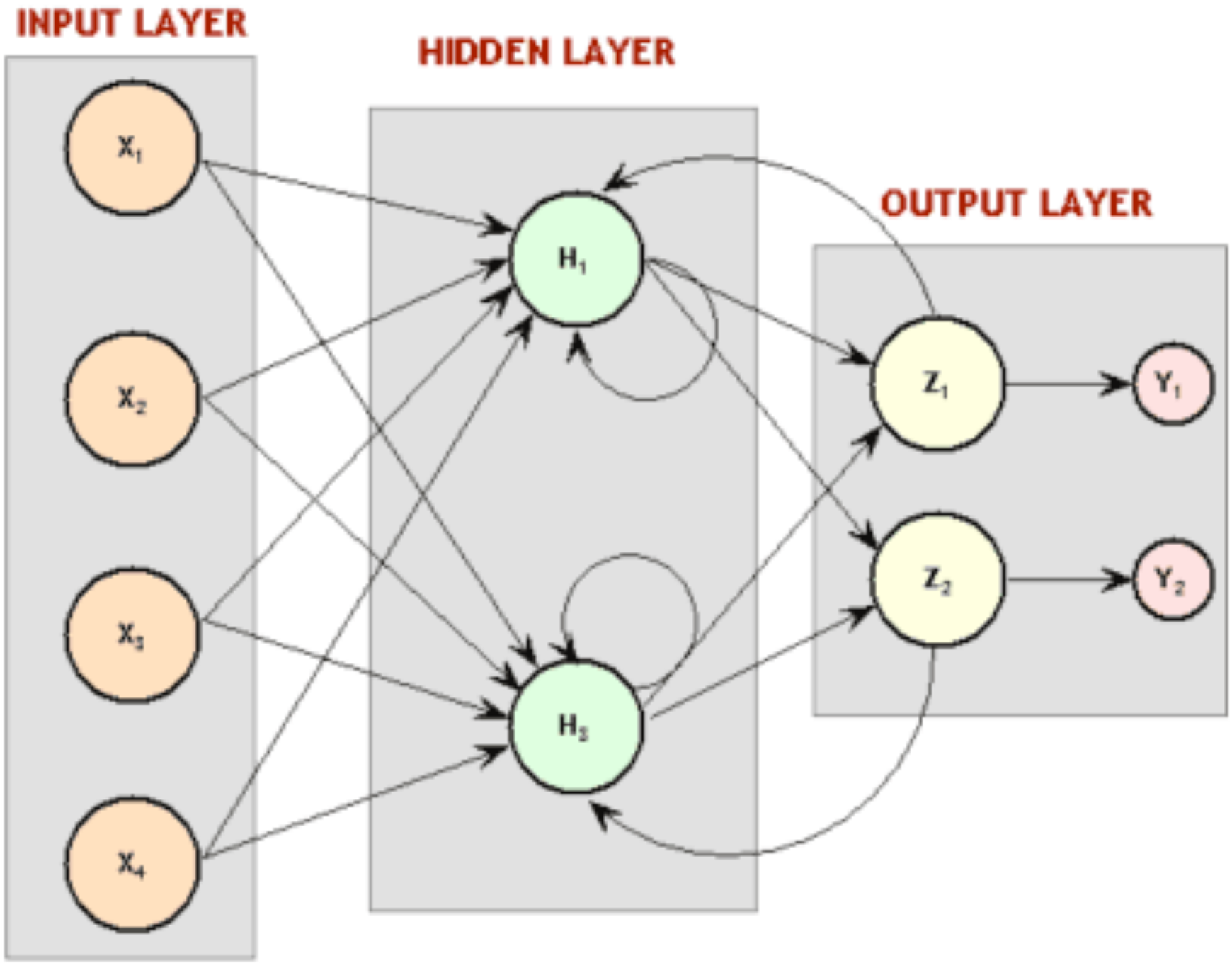


Neural Networks: Compose many perceptrons



Is it a cat?

Recurrent Neural Network

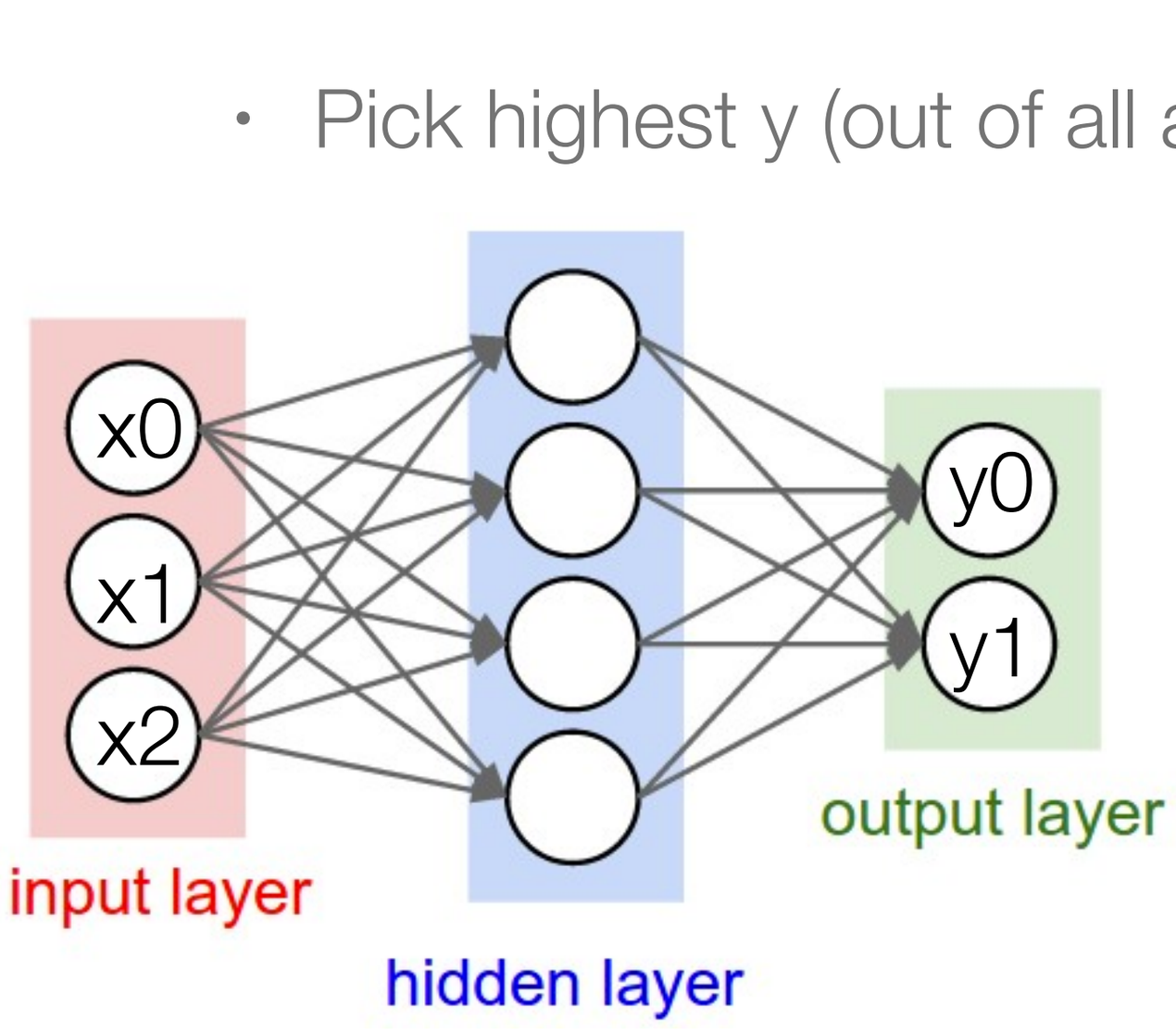


Generating Guesses

Einstein Question

- Vectorize question
- Feed into recurrent neural network
- Pick highest y (out of all answers)

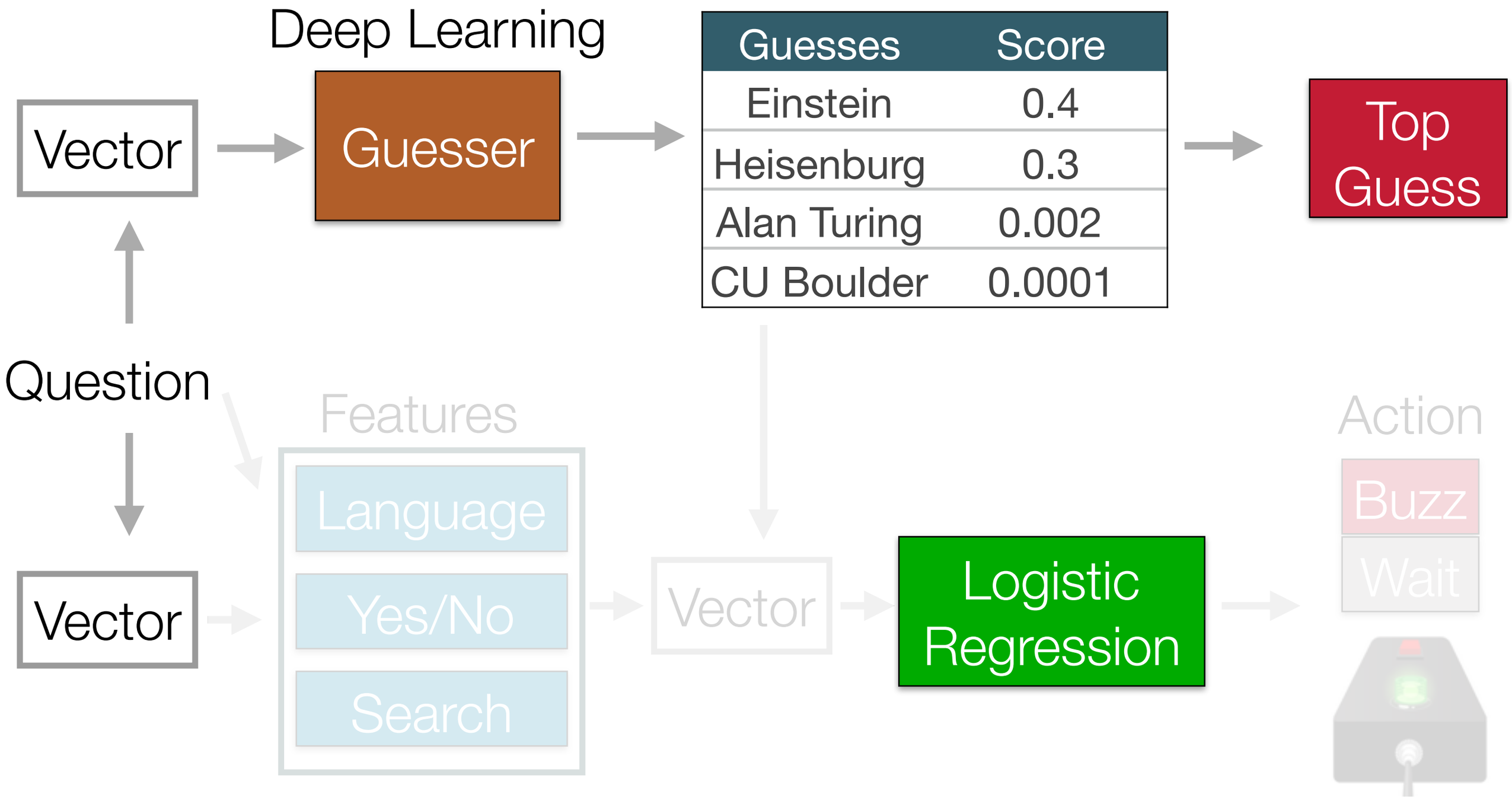
[x0,x1,…xn]



Einstein

Heisenberg

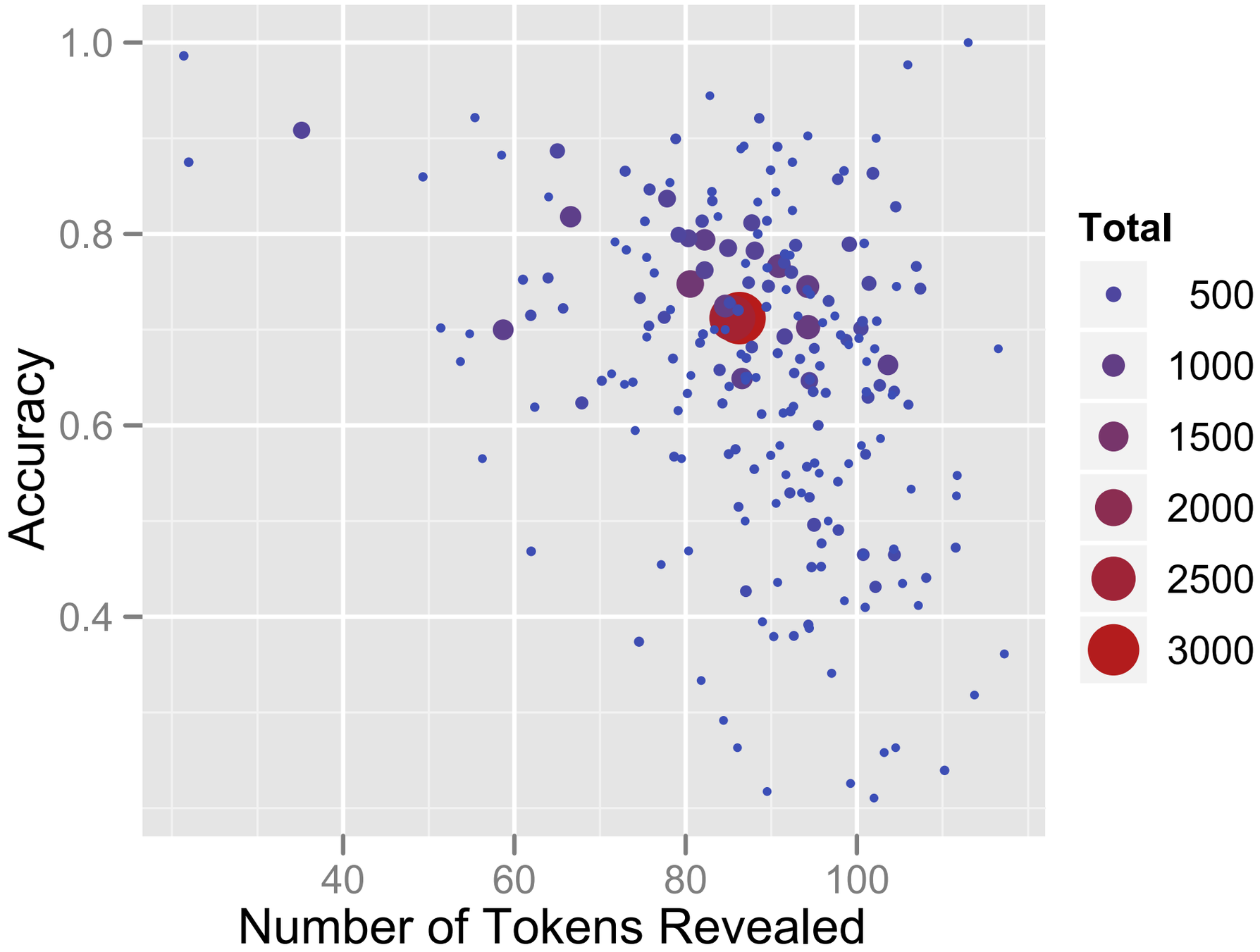
QANTA Overview



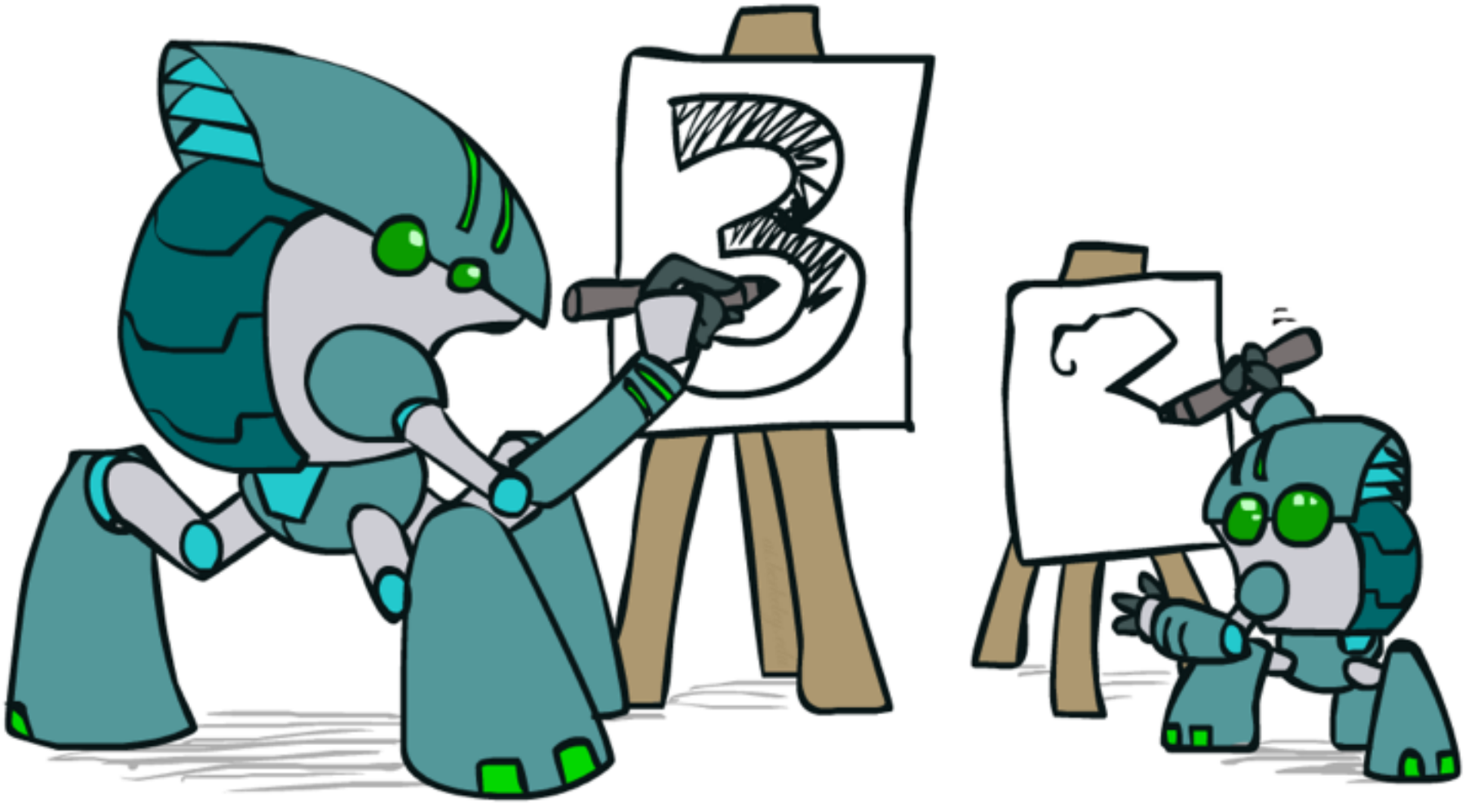
When to Guess?

- Given X predict Y
  - X: all data about the question
  - Y: guess or wait
- Example: when should I sell my stock?
- Quiz Bowl:when should we buzz?

Human Guesses



Learn from best humans



QANTA Features

- Features
  - Something extraction from question that helps discriminate correct/wrong time to guess
  - Language Model
  - Binary Features: gender, answer present,…
  - Wikipedia “lookup”

Language Model

- Probabilistic Model
- How likely is the sequence of words?
- Assuming a bag of words model (order doesn't matter)
- Use Markov Property, why?

Unigram

$$P(w_0, ..., w_n) = P(w_0) \cdot ... \cdot P(w_n)$$

Language Model

- How does this help quiz bowl?
- Condition on what guess is being considered

$$P(w_0, ..., w_n, guess) = P(w_0, ..., w_n | guess) P(guess)$$

$$\max_i P(w_0, \ldots, w_n | guess_i)$$

Score

Classification Features

- Is the answer present in the question?
- Is the question about a male or female?
- What is the question category?

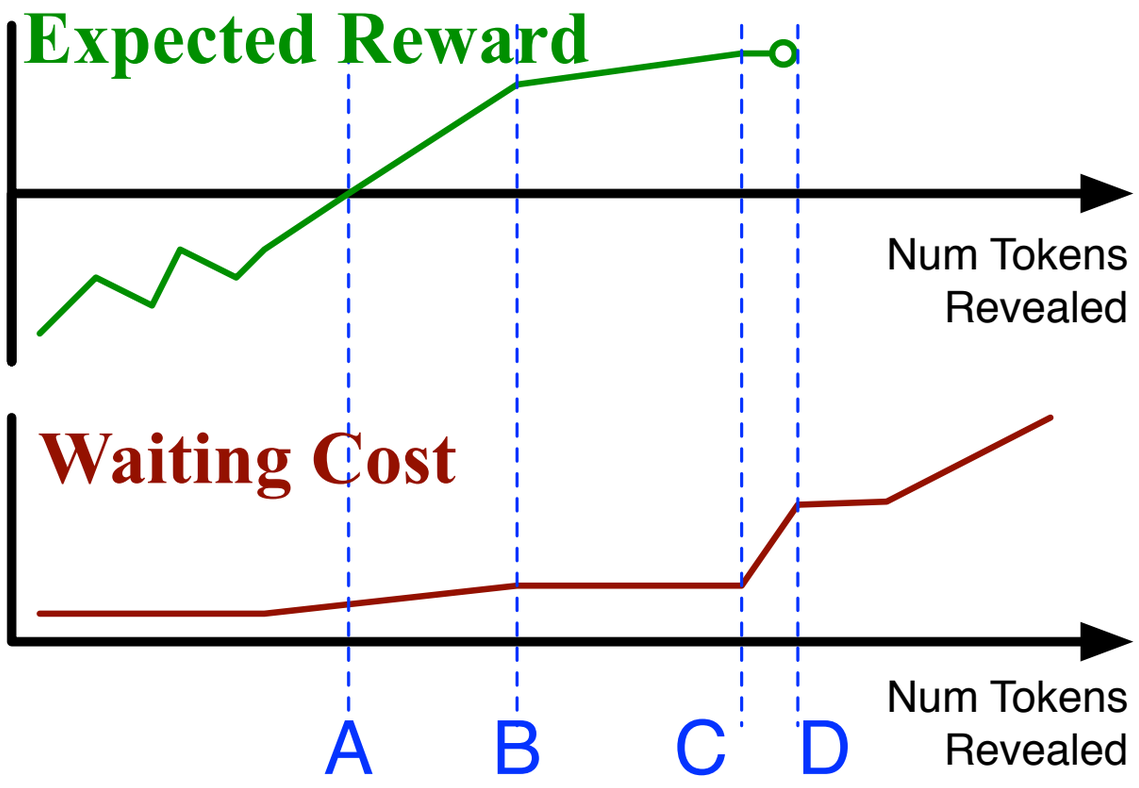
Wikipedia

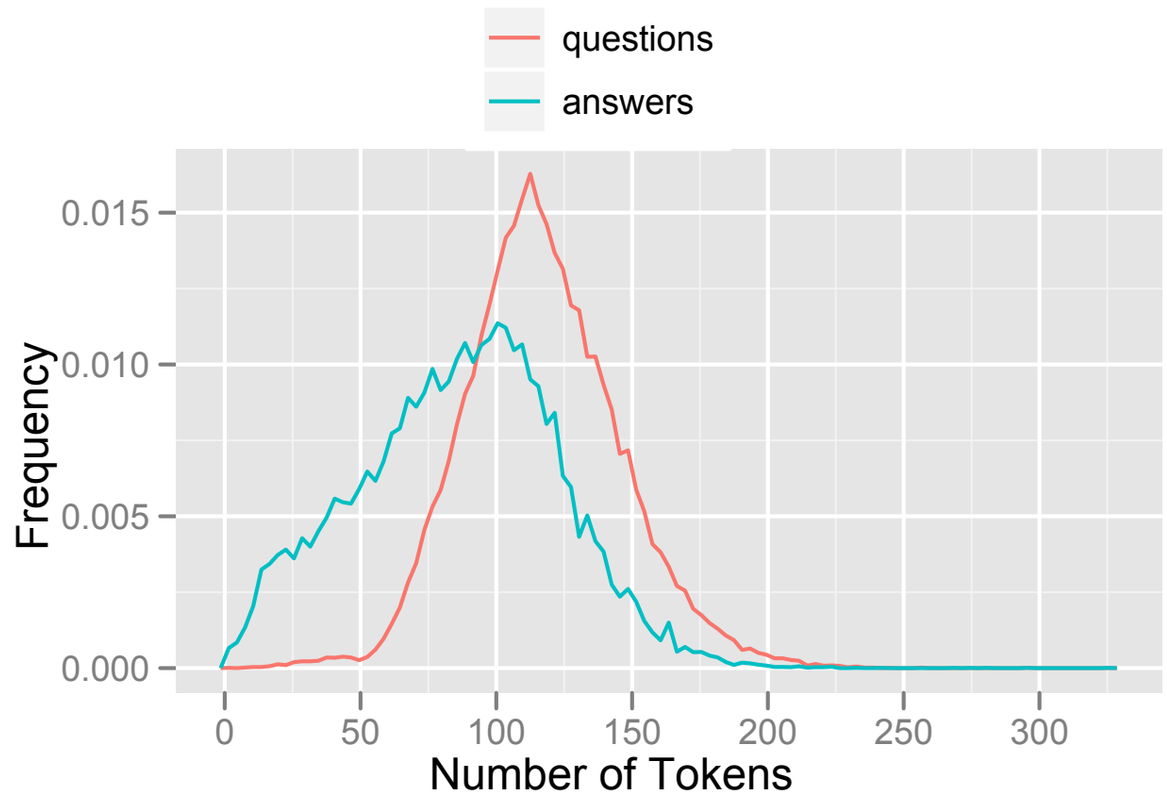
- Idea: “Wikify” the text and look for potential answer
- With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

- Albert Einstein

Features -> Logistic Regression

- Merge all features to large vector
- Train logistic regression model
  - Use data of when guessing is correct





QANTA Overview

Deep Learning

Vector

Guesser

Guesses	Score
Einstein	0.4
Heisenburg	0.3
Alan Turing	0.002
CU Boulder	0.0001

Top Guess

Question

Features

Language

Yes/No

Mentions

Vector

Vector

Logistic Regression

Action

Buzz

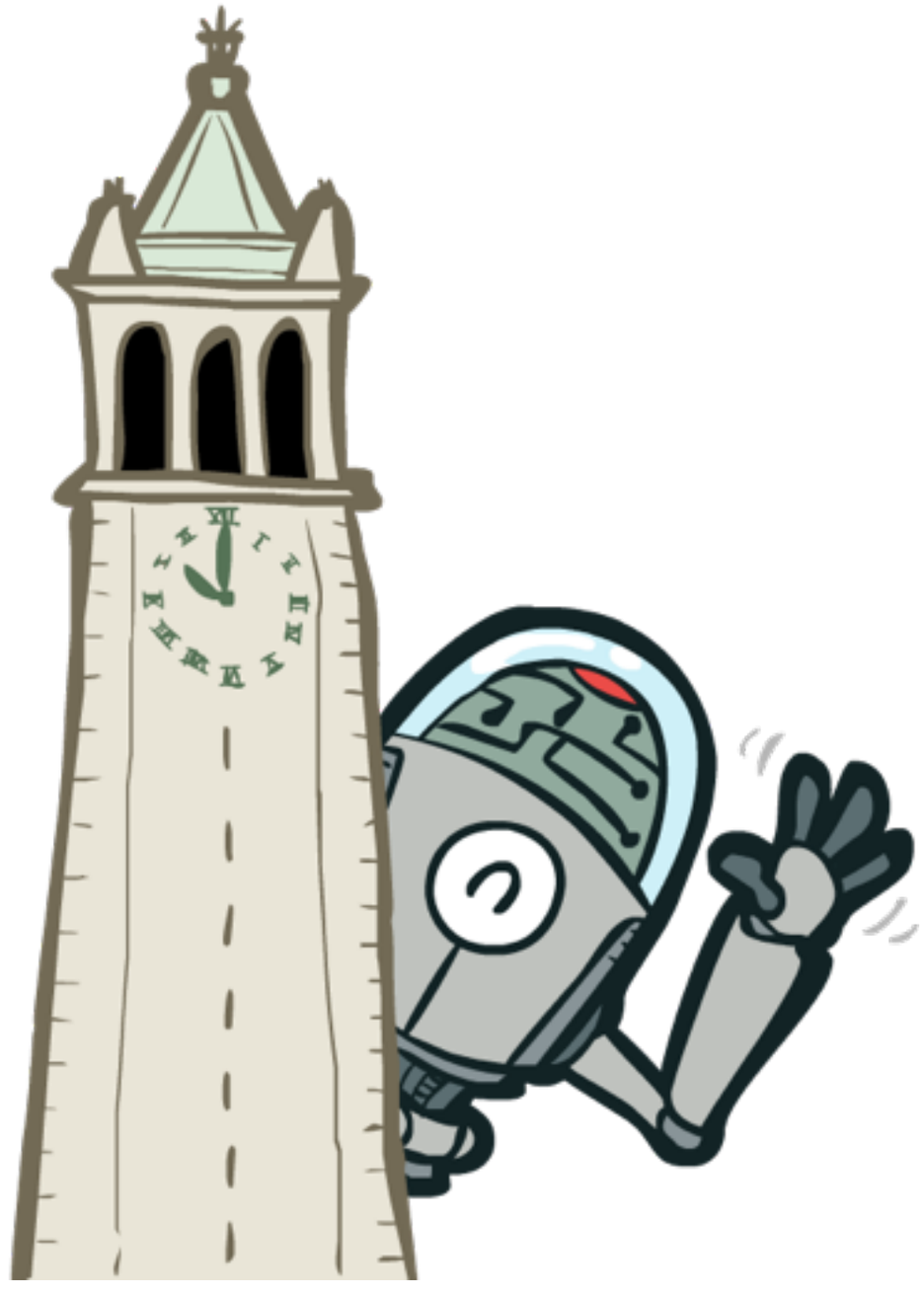
Wait

Future Work

- QANTA
  - Wikipedia topic model similarity
  - Larger knowledge bases
  - Zero-shot learning
- Other Research Interests
  - Detecting player misbehavior in online games
  - Efficient and accurate stream machine learning on big data

Thanks!

- NSF: Bayesian Thinking on Your Feet
- github.com/EntilZha
- QANTA: github.com/Pinafore/qb
- UC Berkeley CS188 Course Materials
- About Me: pedrorodriguez.io
- Contact: p.rodriguez@colorado.edu



References

- Besting the Quiz Master: Crowdsourcing Incremental Classification Games, EMNLP 2012

- A Neural Network for Factoid Question Answering over Paragraphs, EMNLP 2014

- Deep Unordered Composition Rivals Syntactic Methods for Text Classification, ACL 2015

- Removing the training wheels: A coreference dataset that entertains humans and challenges computers, ACL 2015