Machine Learning: Why Do We Care?

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-335

8/26/2026

Course Page/Syllabus Up

- Syllabus and policy:
  - https://www.cs.rit.edu/~ago/courses/335/index.html
- Some useful background to help you:
  - Comfort with mathematics (machine learning is just applied mathematics) – you cannot be afraid of formal symbols, equations, calculus, and linear algebra (we will start by reviewing basics in this course)
  - Comfort with programming / coding (particularly Python, as you will be working with Python in homework); we assume you know how to program and debug
  - **My policy on generative-AI (grain-of-salt: I'm an ex-ML researcher who worked with language models and generative models in the "old days"):**
    This should be an "AI-free space" b/c you are literally learning how to build AI, and you need the ***fundamentals*** to stand out in the job market or to do the science. I can't stop you from taking the easy way out, but later on, you might realize you don't know what's going on with your model.
    (A calculator is powerful if you know the math it's doing!)

General Tentative Schedule

CSCI-335 Introduction to Machine Learning: Schedule

Fall Semester 2026 (XXXX)

RIT Academic Calendar

Disclaimer: We may get ahead of (or fall behind) this schedule, I will try to keep this up to date but regardless, quiz/homework topics will follow the actual lecture topic pace.

Week (Subject to change)	Topics	Homework	Reading	Special Events and Due Dates	Slides & Lecture Notes
1 (1/12+14+16)	Introduction, class logistics, **Review:** linear algebra		DL Ch. 1 & Ch. 2		Slides (1), (2)
2 (1/X+X+X)	Stochastic processes & probability, distributions, information theory		DL Ch. 3, TEoSL Ch. 1	HW #0 due X/XX	Slides (1), (2), (3)
3 (1/X+X+X)	Optimization, foundational principles of ML		DL Ch. 4 & Ch. 5		Slides (1), (2), (3), (4)
4 (X/X+X+X)	Learning theory, generalization, the ML pipeline, K-NN		DL Ch. 5		Slides (1), (2), (3)
5 (X/X+X+X)	Supervised learning: Linear regression		TEoSL Ch. 2.3, 3.1	HW #1.1 due X/XX	Slides (1), (2), (3)
6 (X/X+X+X)	Logistic regression (LR) and (2-class) classification		TEoSL Ch. 4.4		Slides (1), (2), (3)
7 (X/X+X+X)	Discriminative modeling with linear classifiers			HW #1.2 due X/XX	Slides (1), (2), (3), (4)
8 (X/X+X+X)	Unsupervised learning: dimensionality reduction PCA (**Guest lec**)				Slides (1), (2)
9 (X/X+X+X)	Probabilistic graphical models (PGMs): naïve Bayes (NB)		NB vs. LR (Ng & Jordan '01)	HW #2 due X/X	Slides (1), (2), (3)
12 (X/X+X, X/X)	Unsupervised learning, generative models, clustering				Slides (1), (2), (3)
11 (X/X+X+X)	Mixtures of Gaussians, decision trees		Random Forests (Breiman '01)		Slides (1), (2), (3)
13 (X/X+X+X)	Trees/ensembles, artificial neural networks (ANNs)		DL Ch. 6		Slides (1), (2), (3)
12 (X/X+X+X)	ANNs: reverse-mode differentiation, tricks of the trade				Slides (1), (2), (3)
13 (X/X+X+X)	Violating i.i.d.: RNNs & time-series		DL Ch. 10		Slides (1)
14 (X/X+X+X)	*ANNs*: Generative modeling			Final Exam/Project	Slides (1), (2)
16 (X/X, Final: X/XX, XX:XXam-XX:XXpm)	Outlook, Final Project Presentations			Slides due X/XX, 10:40am (Papers due X/XX, 8am)	

Grader Office Hours: TBA (homework-specific questions, homework grade questions)

Subject to change…

335 Course *Theme*:
Machine Learning "Literacy"

Gradient-Based Learning Applied to Document Recognition

YANN LECUN, MEMBER, IEEE, LÉON BOTTOU, YOSHUA BENGIO, AND PATRICK HAFFNER

Invited Paper

(LeCun et al., 1998)

Multilayer neural networks trained with the back-propagation algorithm constitute the best example of a successful gradient-based learning technique. Given an appropriate network architecture, gradient-based learning algorithms can be used

NN	Neural network.
OCR	Optical character recognition.
PCA	Principal component analysis.

AN INEQUALITY WITH APPLICATIONS TO STATISTICAL ESTIMATION FOR PROBABILISTIC FUNCTIONS OF MARKOV PROCESSES AND TO A MODEL FOR ECOLOGY

BY LEONARD E. BAUM AND J. A. EAGON

Communicated by R. C. Buck, November 21, 1966

1. **Summary.** The object of this note is to prove the theorem below and sketch two applications, one to statistical estimation for (probabilistic) functions of Markov processes [1] and one to Blakley's model for ecology [4].

2. Result

(Baum and Eagon 1967)

335 Course *Theme*: **Machine Learning "Literacy"**

Gradient-Based Learning Applied to Document Recognition

YANN LECUN, MEMBER, IEEE, LÉON BOTTOU, YOS... ER

Invited Paper

...ical character recognition.

...A Principal component analysis.

CSCI 736 Theme:
Machine Learning "Communication"

AN ... STATISTICAL ... ISTIC FUNCTIONS ... PROCESSES AND TO A MODEL FOR ECOLOGY

BY LEONARD E. BAUM AND J. A. EAGON

Communicated by R. C. Buck, November 21, 1966

1. **Summary.** The object of this note is to prove the theorem below and sketch two applications, one to statistical estimation for (probabilistic) functions of Markov processes [1] and one to Blakley's model for ecology [4].

2. Result

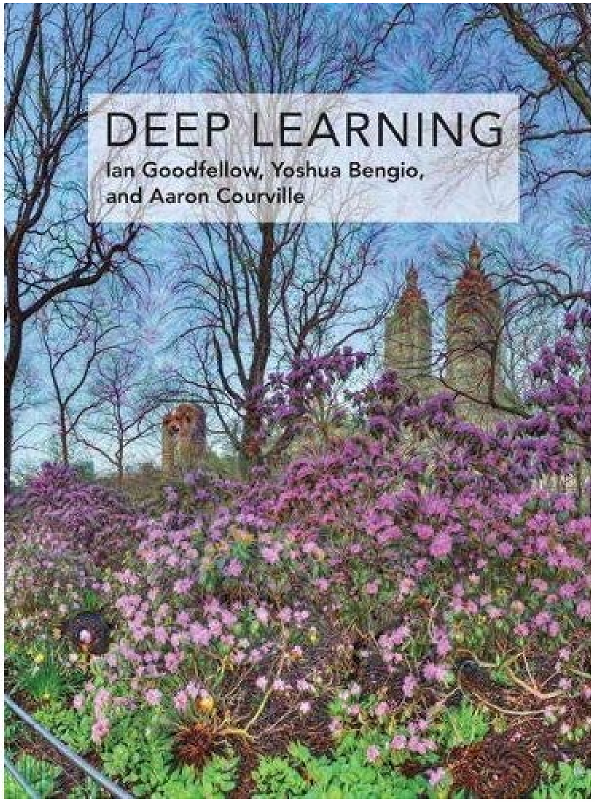
Policies Worth Noting

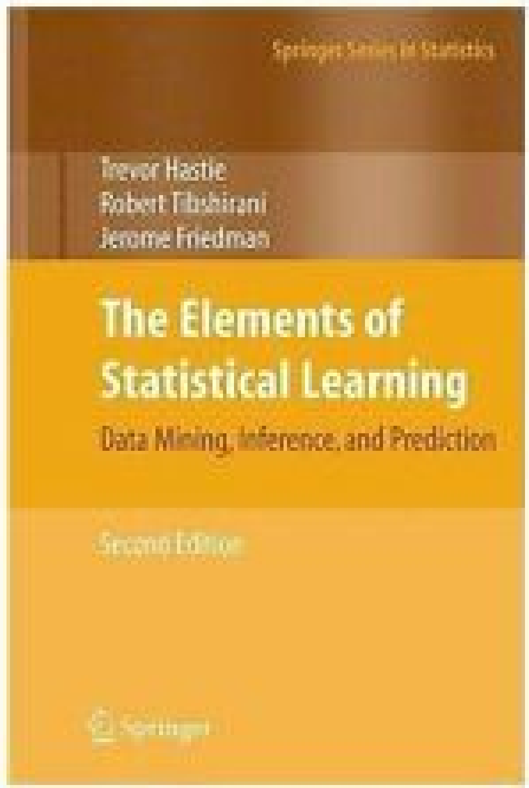
- ## Academic Integrity

The DCS Policy on Academic Honesty will be enforced.
You should only submit work that is completely your own. Failure to do so counts as academic dishonesty and so does being the source of such work. Submitting work that is in large part not completely your own work is a flagrant violation of basic ethical behavior and will be punished according to department policy.

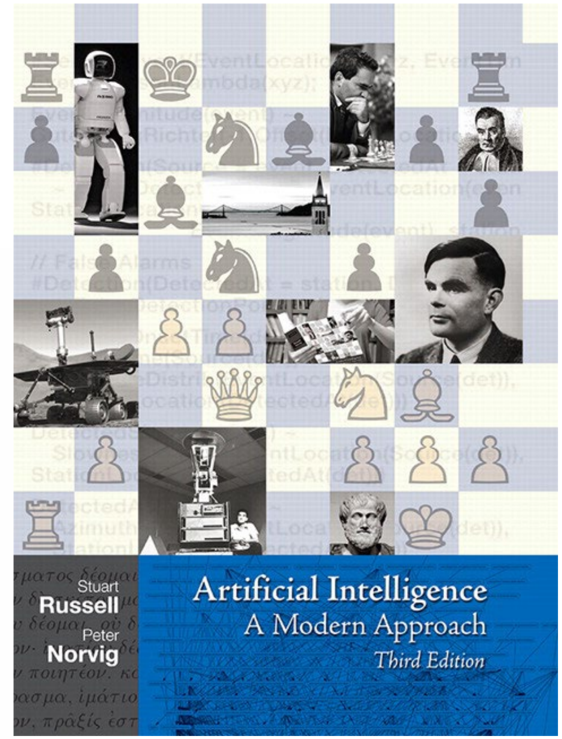
- ## Policy on Large Language Models

The policy on using large language models (LLMs), part of a broader class of statistical learning models labeled as "generative AI", for this course is simple -- please read the above strict policy "Academic Integrity" for this class. Using an LLM to write your code/text will be treated as not producing your own work (such as copying one of your classmates' work) and will be handled accordingly -- you must produce work that is completely your own. Adhering to this policy is further for your personal benefit -- you get what you put into this class, and to master the craft of machine learning, you must work through the mathematics and do the thinking for yourself in order to truly develop the machine learning literacy this class aims to provide.


DEEP LEARNING
Ian Goodfellow, Yoshua Bengio,
and Aaron Courville


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of
Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer

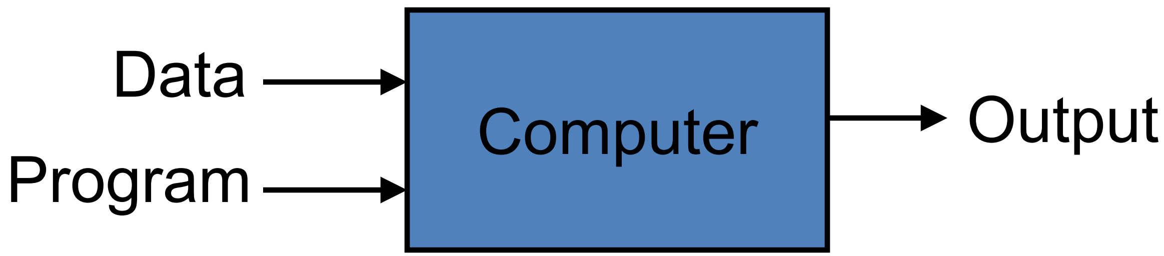
From 630!!


Stuart
Russell
Peter
Norvig
Artificial Intelligence
A Modern Approach
Third Edition

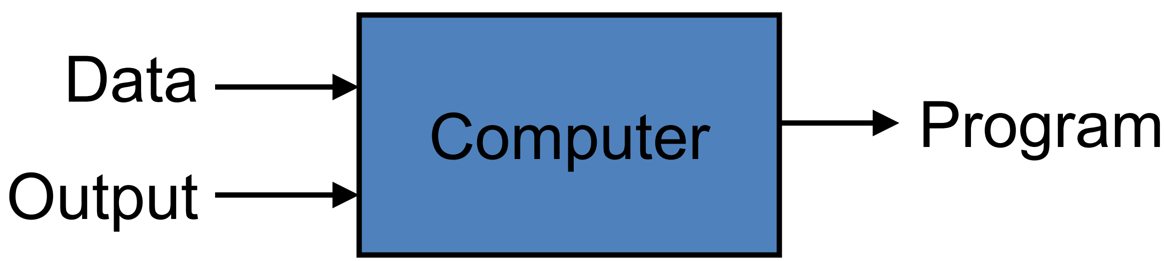
What is Machine Learning (ML)?

- A branch of **artificial intelligence**, concerned with the design and development of algorithms that allow computers to evolve their behavior based on empirical data
  - Automating automation
  - Getting computers to program themselves
  - Writing software is the bottleneck
  - Instead, let the data do the work instead!
- Intelligence requires knowledge, thus it is necessary for computers to acquire knowledge

Traditional Programming



Machine Learning



Is it Magic?

No, it's more like gardening:

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

ML in a Nutshell

- A galaxy (tens of thousands) of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three key components (*design elements*):
  - **Representation**
  - **Evaluation**
  - **Optimization**

ML in a Nutshell

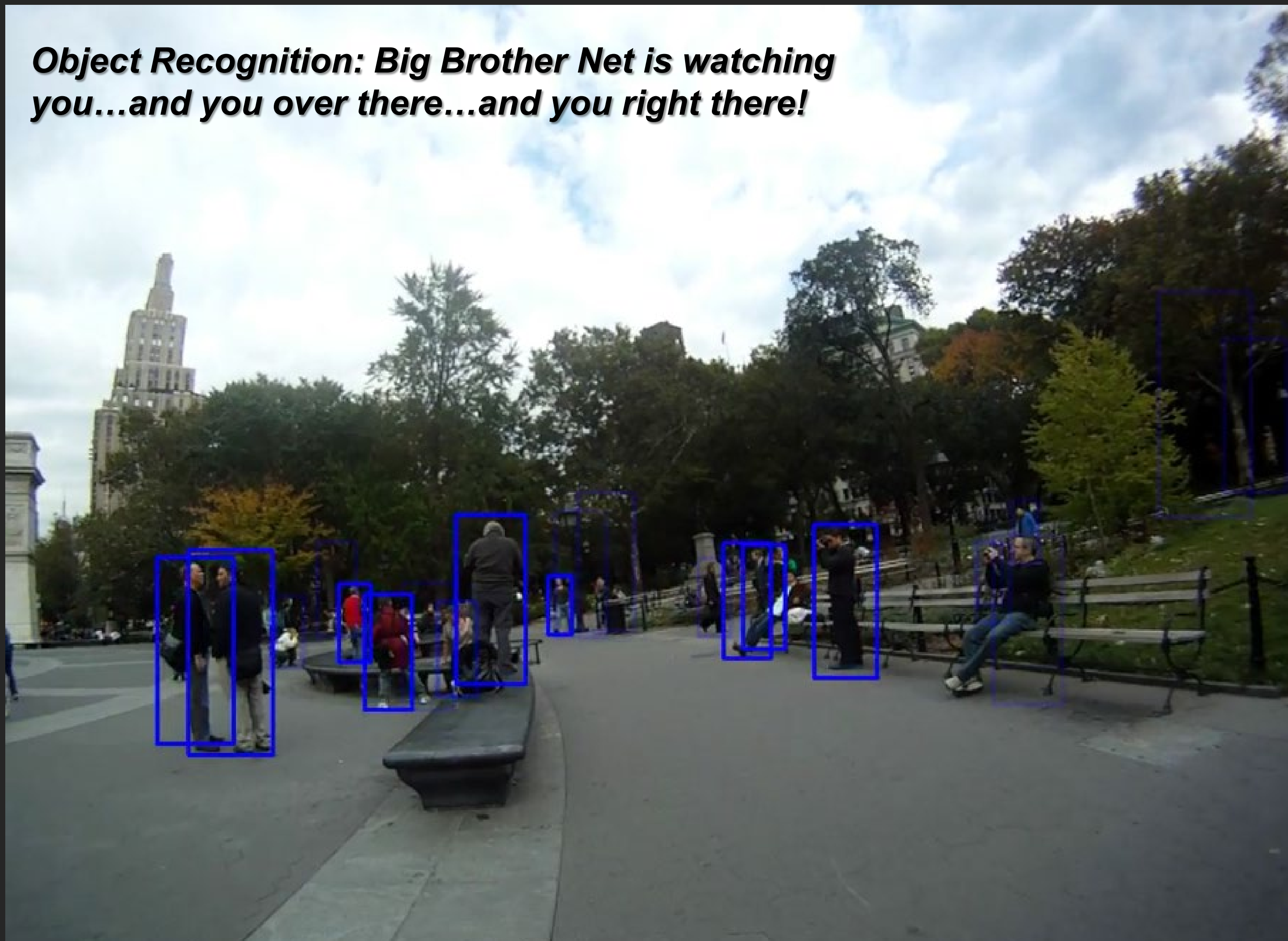
- A galaxy (tens of thousands) of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three key components (*design elements*):
  - **Representation**
  - **Evaluation**
  - **Optimization**

This will be another global concept in this course!

The three "pillars" of machine learning

Question:

What can you do once you have all three pillars implemented?

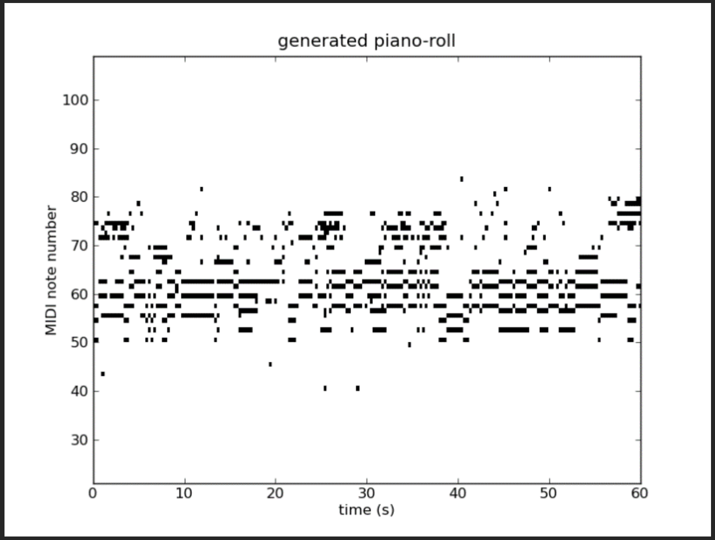

Object Recognition: Big Brother Net is watching
you…and you over there…and you right there!





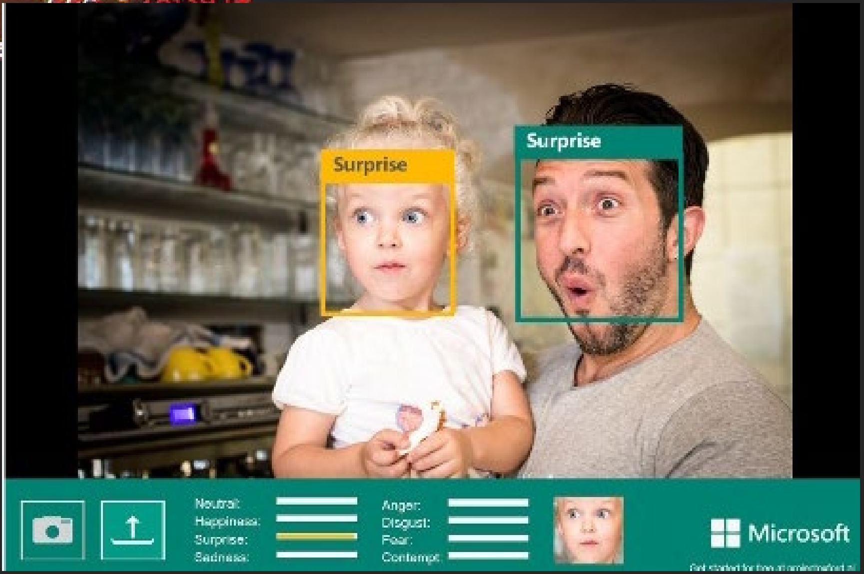
Deep art!

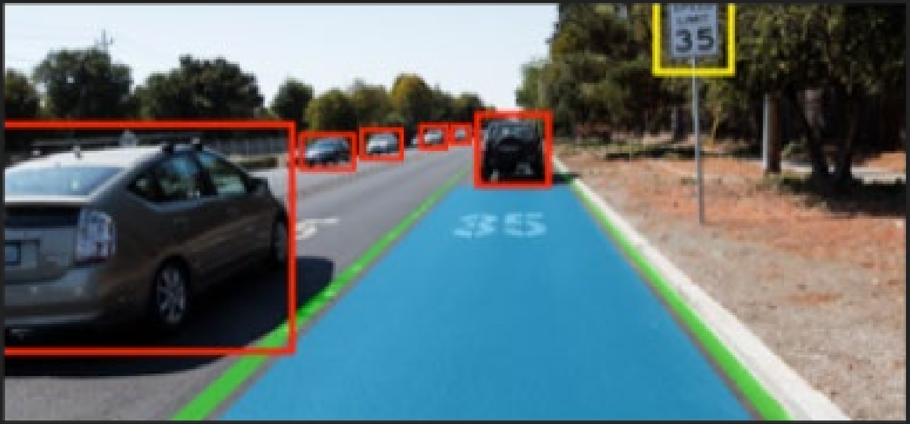
Deep music!



http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/

Deep emotion!





Deep driving!

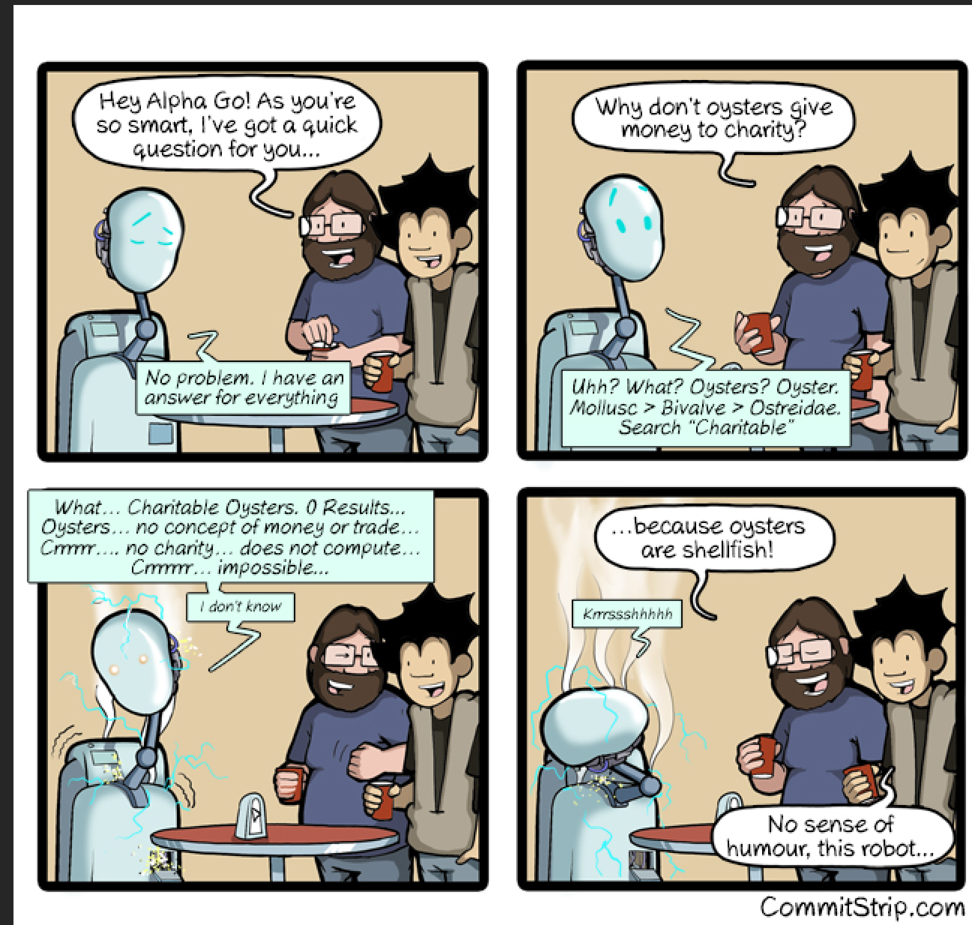
Deep Drumpf!

Modeling of human users, e.g., Twitter bots!
The case above is eerily realistic and perfectly accurate!

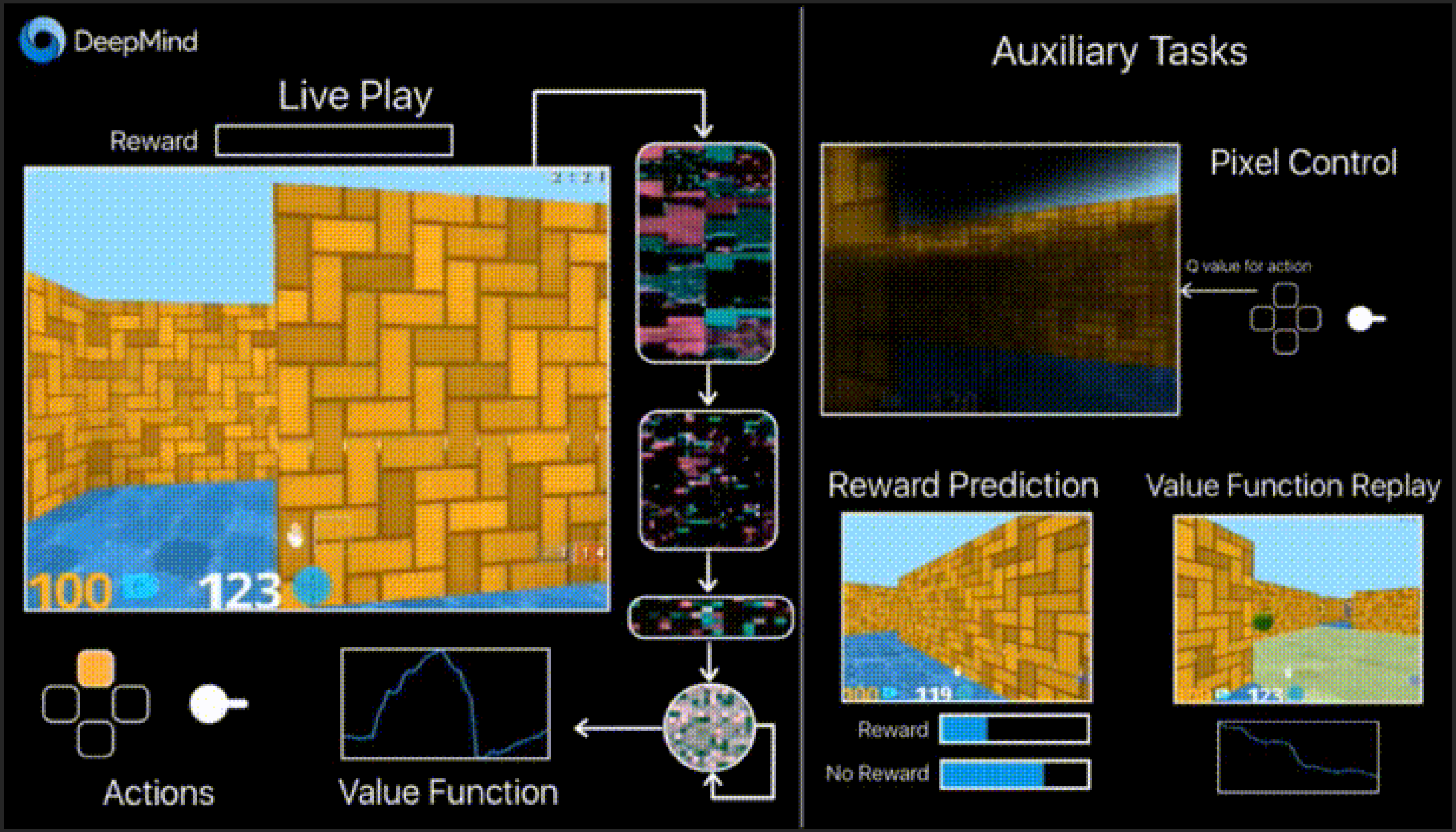
Modeling of human users, e.g., Twitter bots!
The case above is eerily realistic and perfectly accurate!


ALPHAGO
00:12:09
Google DeepMind
Challenge Match
LEE SEDOL
00:34:31


Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
AlphaGo
Lee Sedol

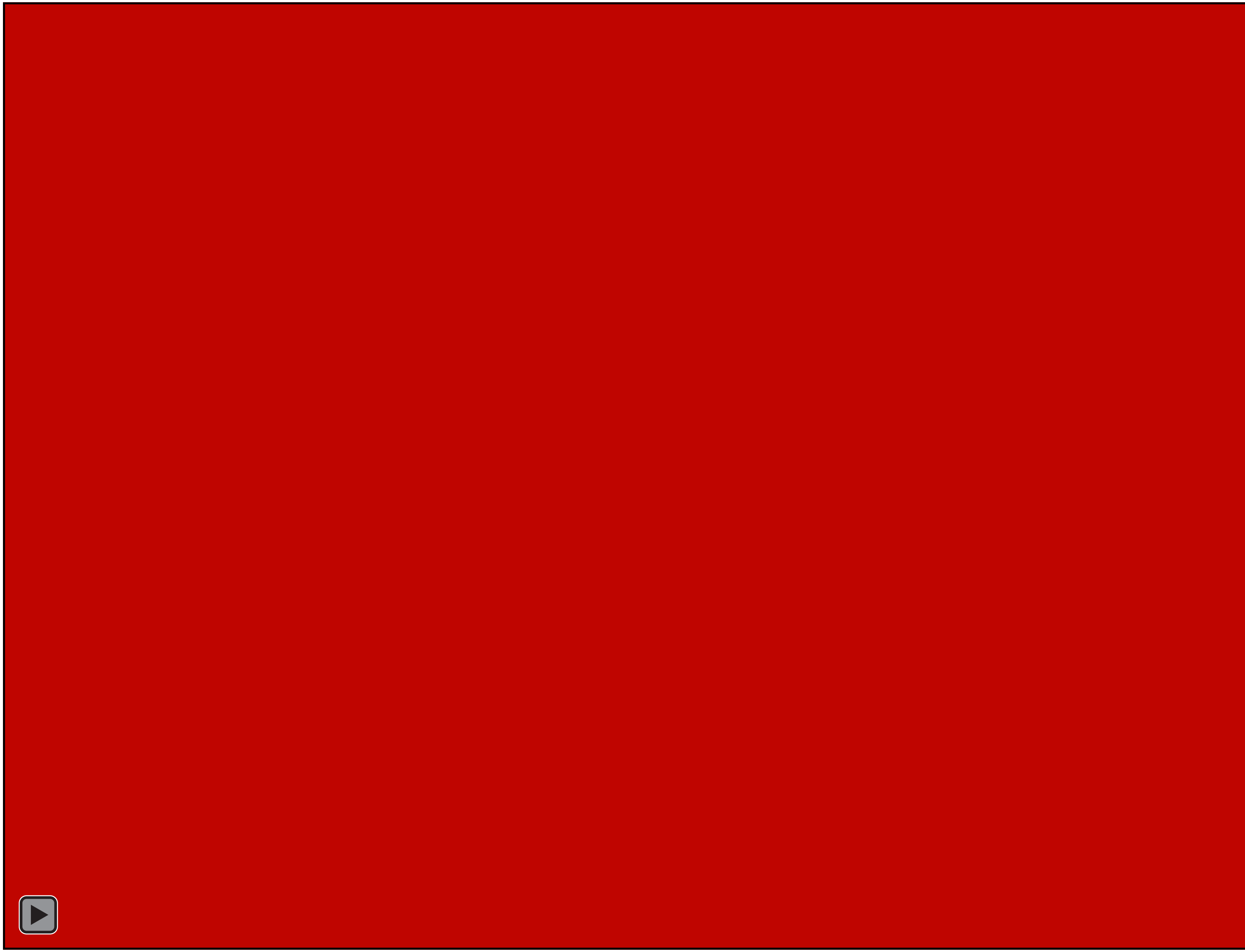

Hey Alpha Go! As you're so smart, I've got a quick question for you...
No problem. I have an answer for everything
Why don't oysters give money to charity?
Uhh? What? Oysters? Oyster. Mollusc > Bivalve > Ostreidae. Search "Charitable"
What... Charitable Oysters. 0 Results... Oysters... no concept of money or trade... Crrrrr.... no charity... does not compute... Crrrrrr... impossible...
I don't know
...because oysters are shellfish!
Krrrssshhhhh
No sense of humour, this robot...
CommitStrip.com

Deep games!



Deep maze exploration!

Or, if you are so lucky, you might build…

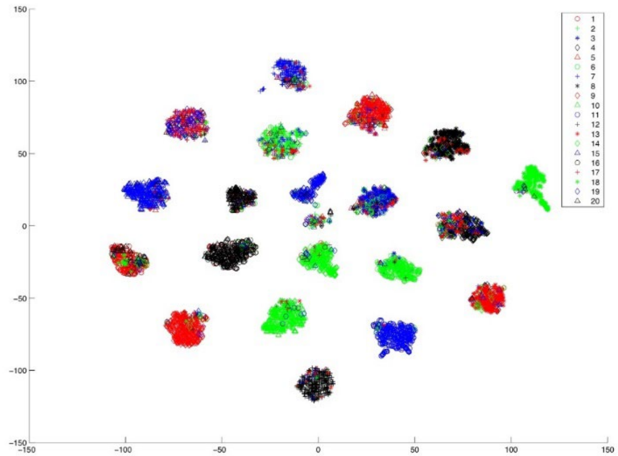


Wait! There's More…

- Classification/regression
  - Sentiment analysis, document categorization, object recognition
  - Information retrieval (entity linking, etc.)

- Dimensionality reduction/visualization
  - Pre-training, t-SNE

- Sequence modeling
  - Language modeling
  - Conversation/dialogue modeling
  - Question answering



https://lvdmaaten.github.io/tsne/

So what are *some* the details of these three design elements?

Representation

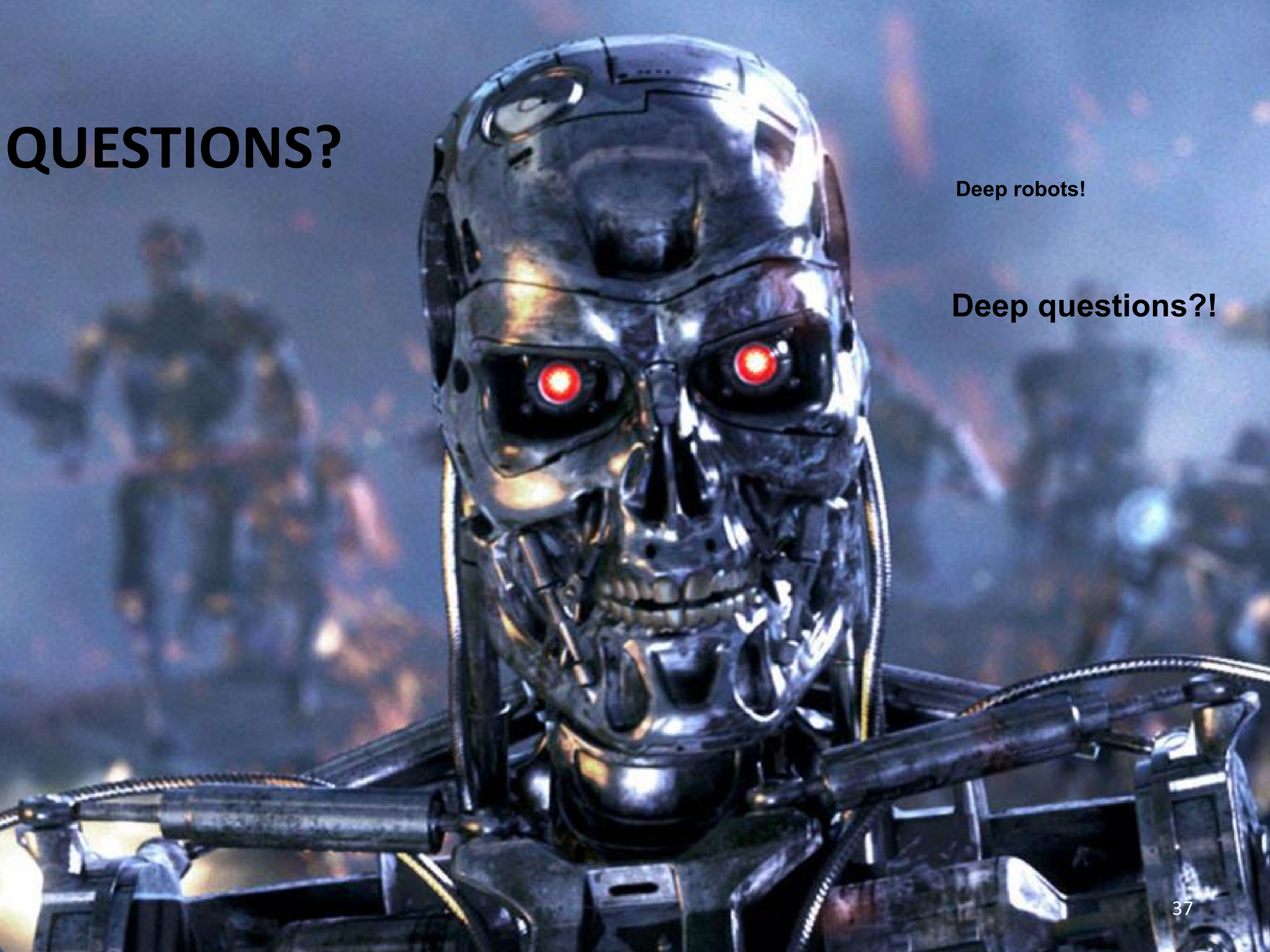
- Decision trees
- Sets of rules / logic programs
- Instances (instance-based learning)
- Graphical models (Bayesian/Markov networks)
- Artificial neural networks (ANNs)
- Support vector machines (SVMs)
- Model ensembles
- And many more…

Evaluation

- Accuracy
- Precision and recall
- (Mean) Squared error
- (Log) Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- Kullback-Leibler (KL) divergence
- ***And many more…***

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
  - We often apply these methods to non-convex problems
- Constrained optimization
  - E.g.: Solution values that are small


QUESTIONS?
Deep robots!
Deep questions?!

References

- Baum, Leonard E., and John Alonzo Eagon. "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology." (1967): 360-363.
- LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.