Machine Learning: Why Do We Care?

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-635

8/28/2022

Course Page/Syllabus Up

- Syllabus and policy:
- https://www.cs.rit.edu/~ago/courses/635/index.html
- Prerequisites:
- CSCI 630: Foundations of Intelligent Systems
- Or equivalent background

General Tentative Schedule

Week	Topics	Homework	Reading	Special Events and Due Dates	Slides & Lecture Notes
1 (8/28+30,9/1)	Introduction, **Review:** linear algebra		DL (Ch. 1 &) Ch. 2		Slides (1), Slides (2)
2 (9/XX+XX+XX)	**Review:** linear algebra; data & computation graphs		DL Ch. 2 & 3		Slides (1), Slides (2)
3 (9/XX+XX+XX)	**Review:** probability, sampling & stochastic processes		DL Ch. 3		Slides (1), Slides (2)
4 (9/XX+XX+XX)	Probability distributions, information theory, optimization		DL Ch. 5		Slides (1), Slides (2)
5 (9/20+22)	Learning theory, foundational principles			HW #1 [Data/Prob] due 10/5	Slides (1)
6 (9/27+29)	Learning theory, foundational principles				Slides (1), Slides (2)
7 (10/4+10/6)	Linear regression				
8 (10/11+13)	Logistic regression		Paper (1)		Slides (1)
9 (10/18+20)	Discriminative modeling, linear classifiers			HW #2 [Stat/ML] due 11/2	Slides (1), Slides (2)
12 (10/25+27)	Clustering, generative modeling				Slides (1), Slides (2)
11 (11/1+3)	Probabilistic graphical models (naive Bayes), Gaussian mixtures (GMMs)			HW #3 [MaxEnt/NB] due 11/18	Slides (1), Slides (2)
13 (11/8+10)	GMMs, PCA (**Guest: Will Gebhardt**)				Slides (1), Slides (2)
12 (11/15+17)	Decision tree, ensembles: bagging & boosting, artificial neural networks (ANNs)			HW #4 [Nets] due 12/9	Slides (1), Slides (2)
14 (11/22)	*ANNs:* reverse-mode differentiation		DL Ch. 6		Slides (1), Slides (2)
15 (11/29+12/1)	*ANNs*: recurrence, Bayesian neural nets (**Guest: Hong Yang**)			Final Exam/Project	Slides (1), Slides (2)
16 (12/13, 1:30-4pm)	Final Project Presentations			Slides due 11:59pm (Papers due Dec 13, 11:59pm)	

AO Office Hours: Monday, 4:30-5:30pm (general class logistics/issues/concerns), **starting September 4, 2023**

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

Subject to change…

635 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)

635 Course *Theme*:
Machine Learning “Literacy”

Gradient-Based Learning Applied to Document Recognition

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

Invited Paper

... ptical character recognition.
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

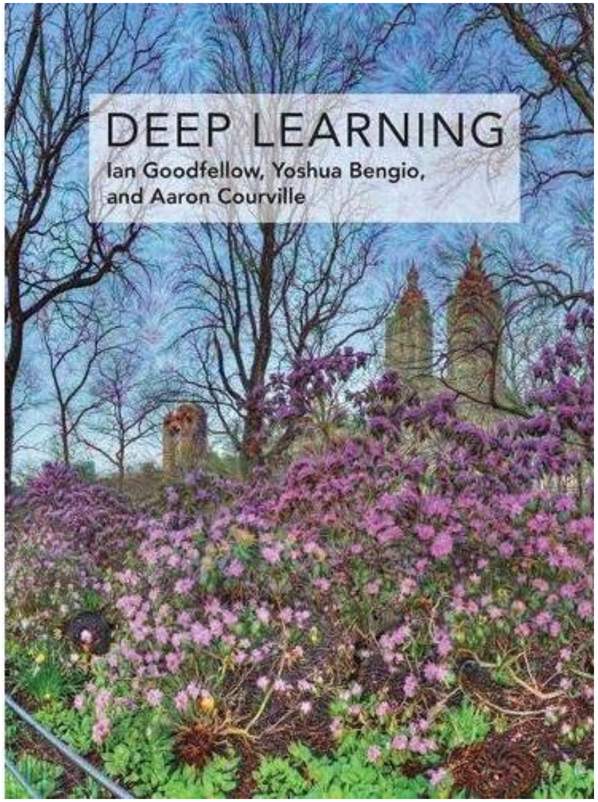
Policies Worth Noting

- **Academic Integrity**

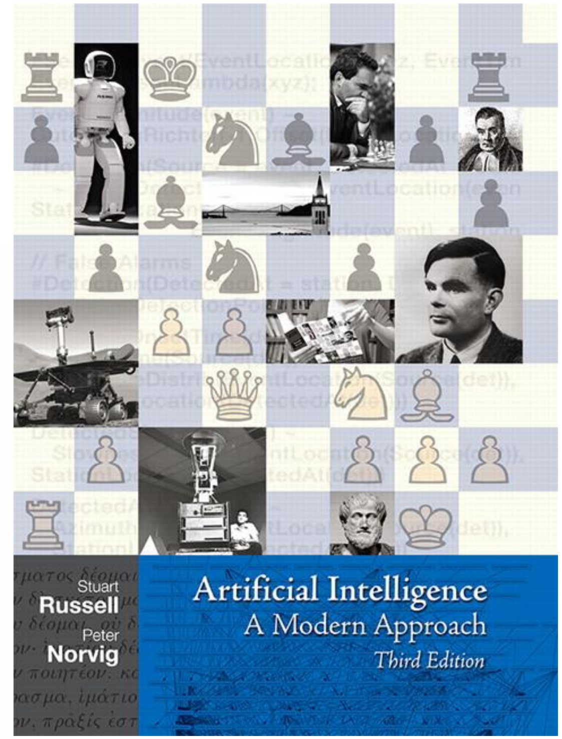
  The DCS Policy on Academic Honesty will be enforced.
  You should only submit work that is completely your own. Failure to do so counts as academic dishonesty and so does being the source of such work. Submitting work that is in large part not completely your own work is a flagrant violation of basic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

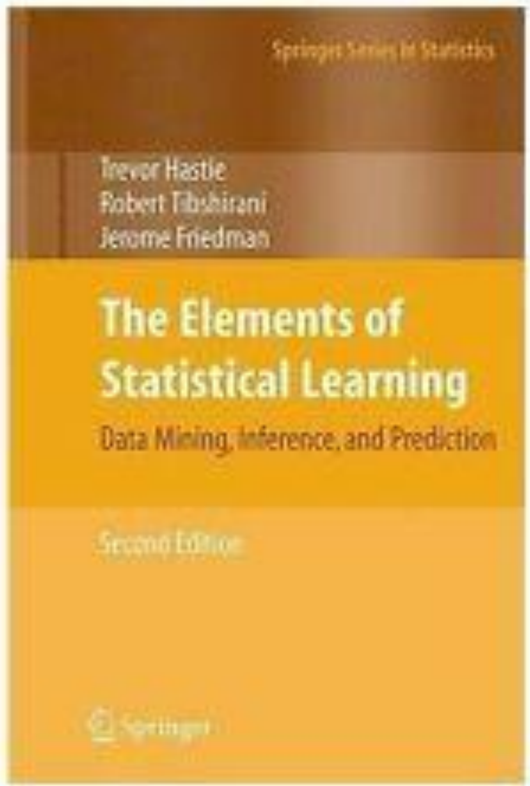
- **Policy on Large Language Models**

  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

From 630!!

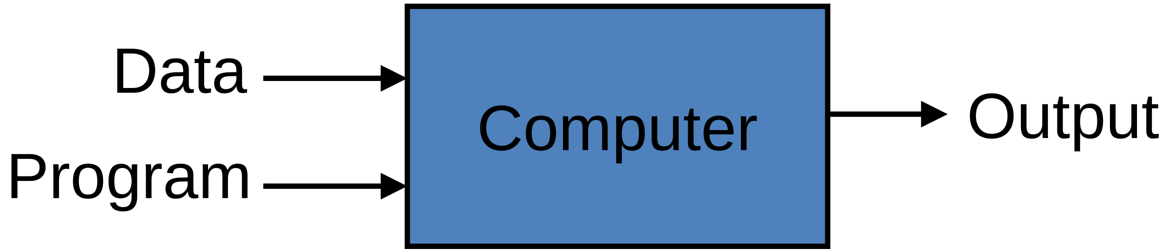

Stuart
Russell
Peter
Norvig
Artificial Intelligence
A Modern Approach
Third Edition


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

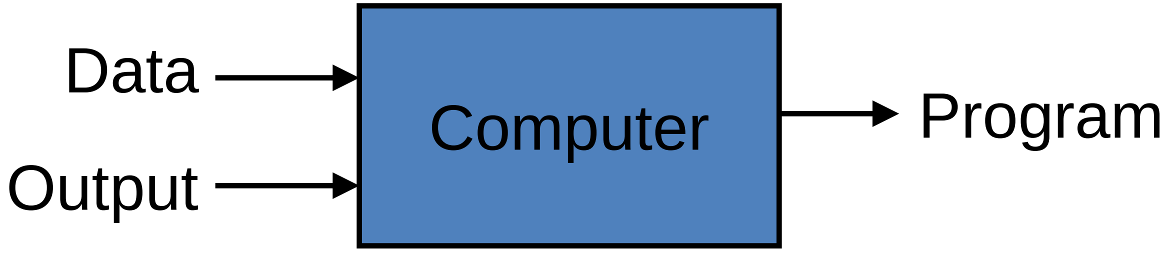
What is Machine Learning (ML)?

- 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

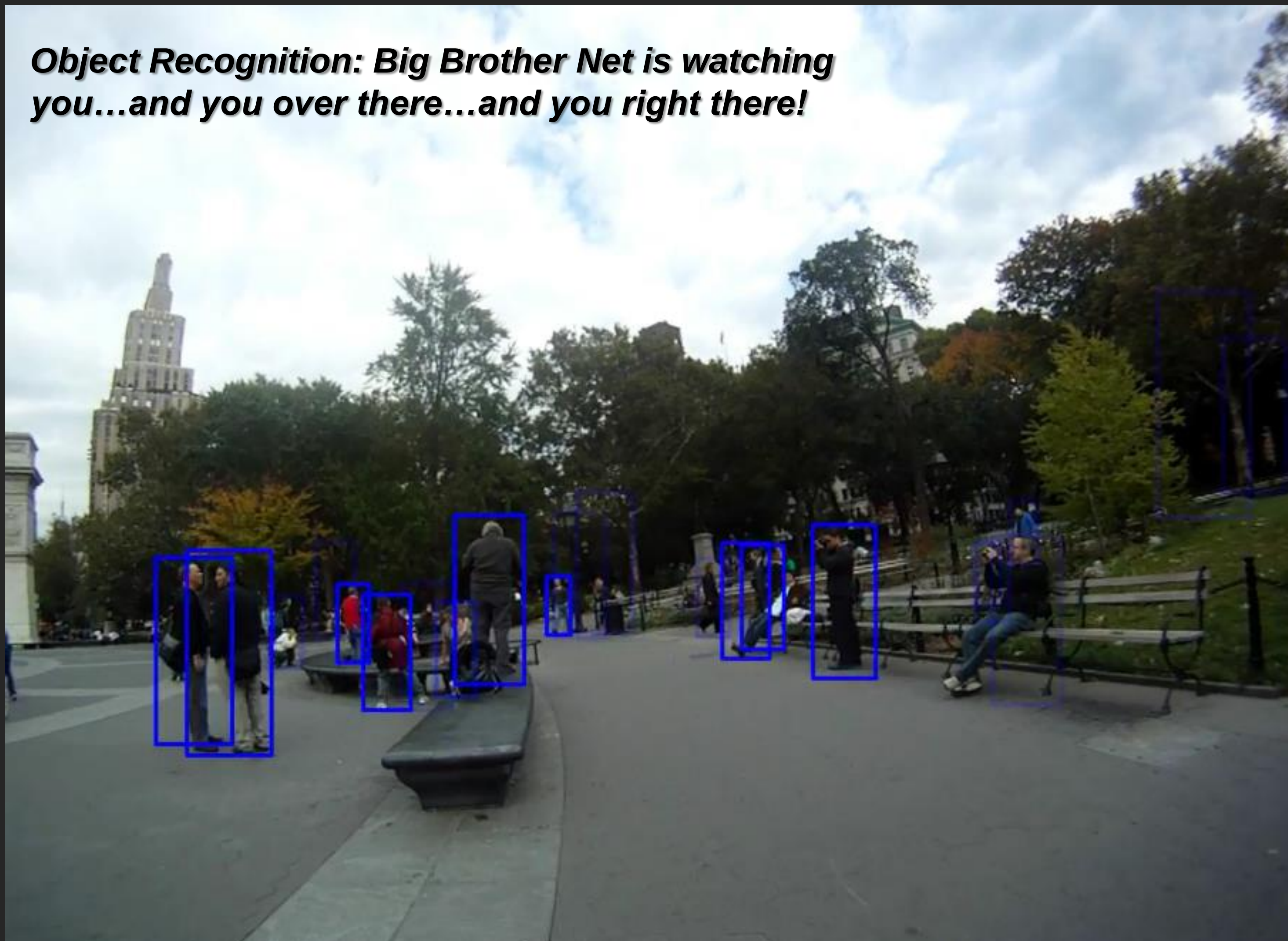
- A galaxy (tens of thousands) of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three key components (*design elements*):
  - **Representation**
  - **Evaluation**
  - **Optimization**

This will be another global concept in this course!

The three "pillars" of machine learning

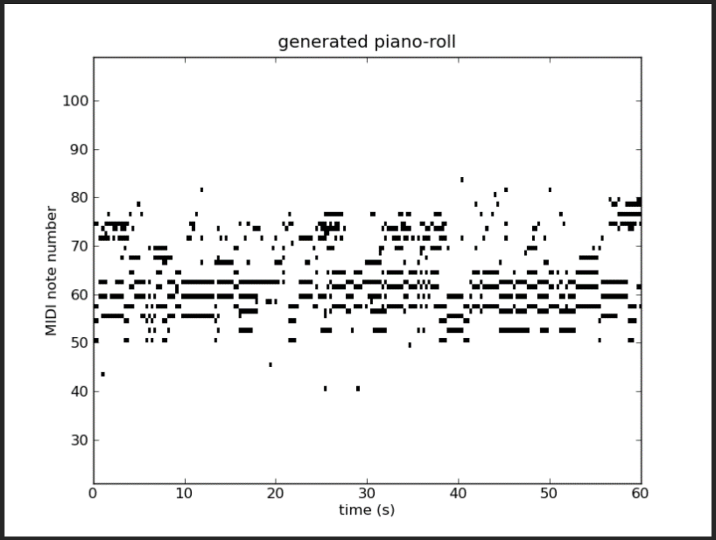
Question:

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


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!

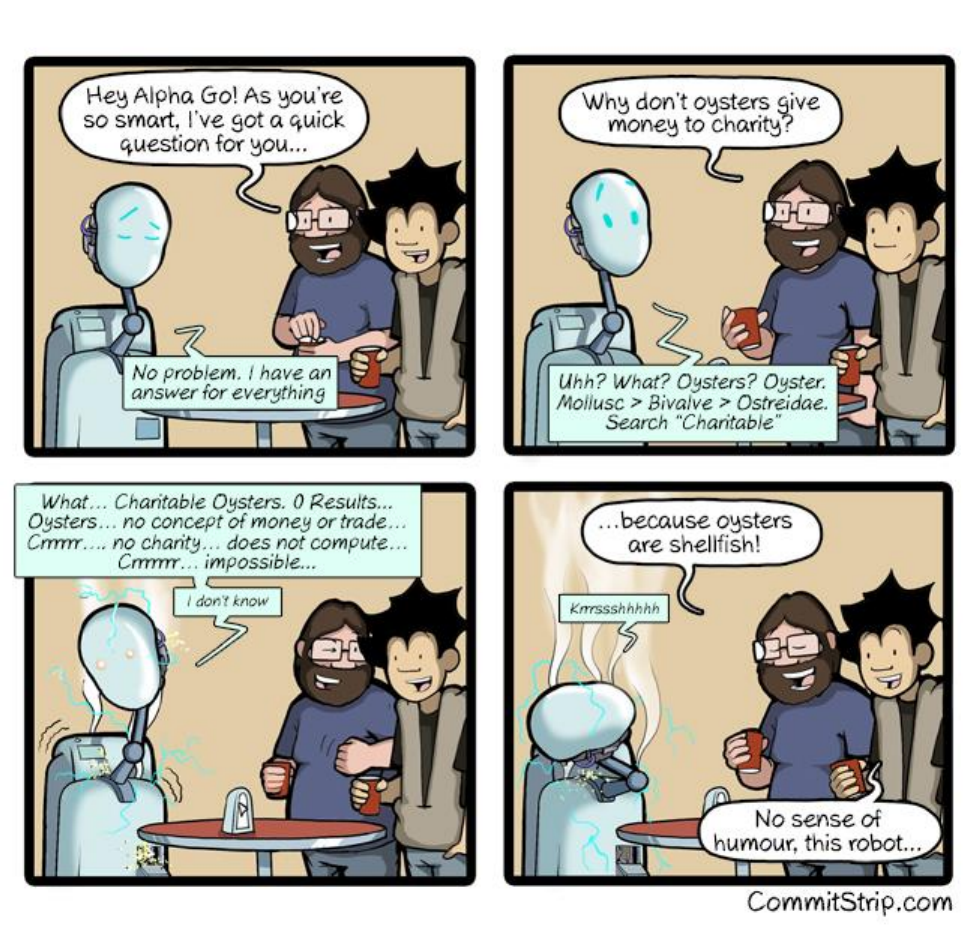
Modeling of human users, e.g., Twitter bots!
The case above is eerily realistic and perfectly accurate!

Deep Drumpf!

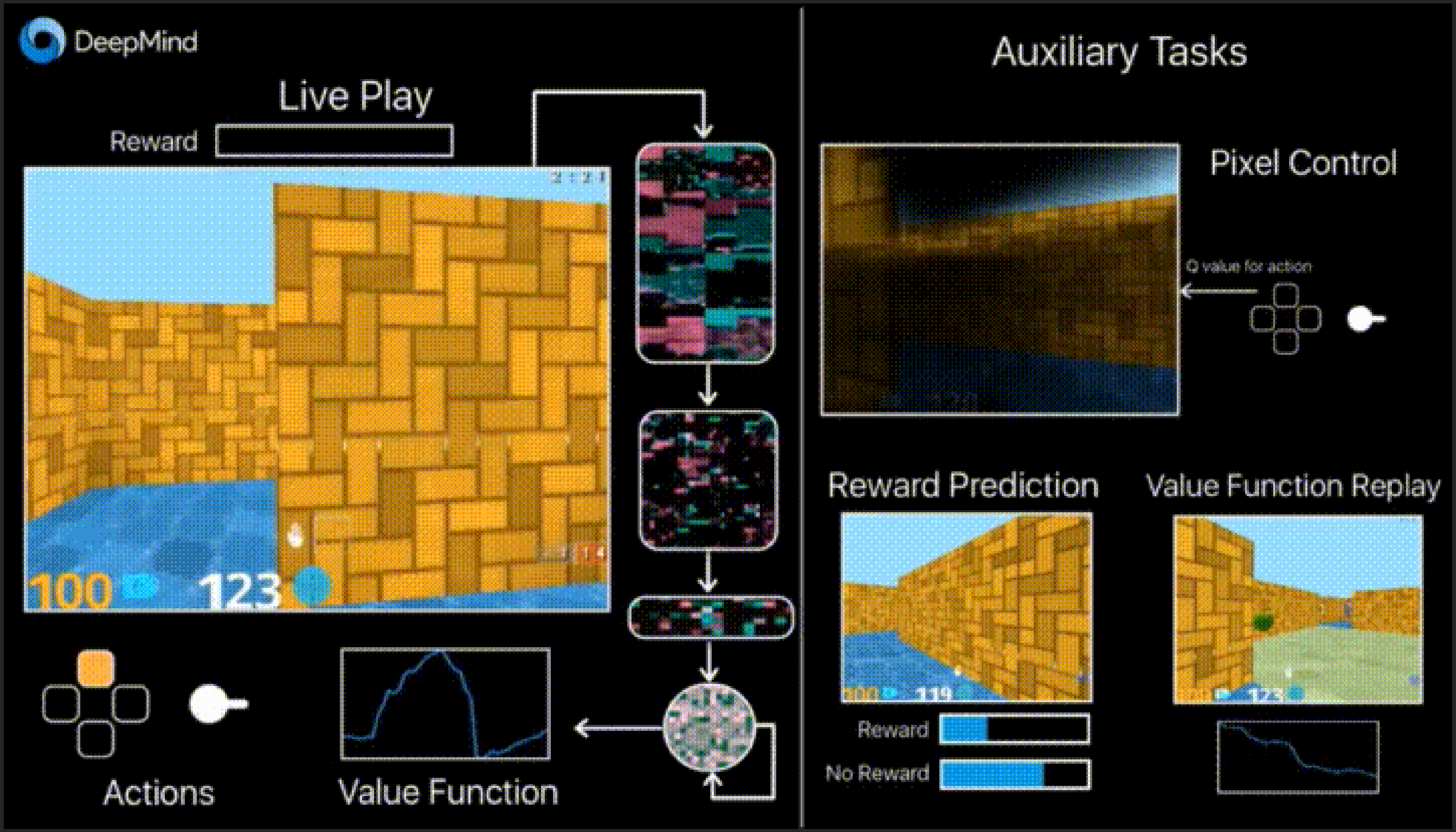
Modeling of human users, e.g., Twitter bots!
The case above is eerily realistic and perfectly accurate!


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


Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
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…

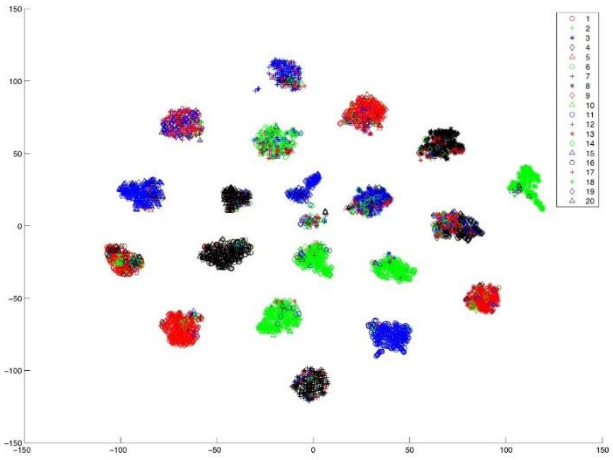
Deep robots!

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

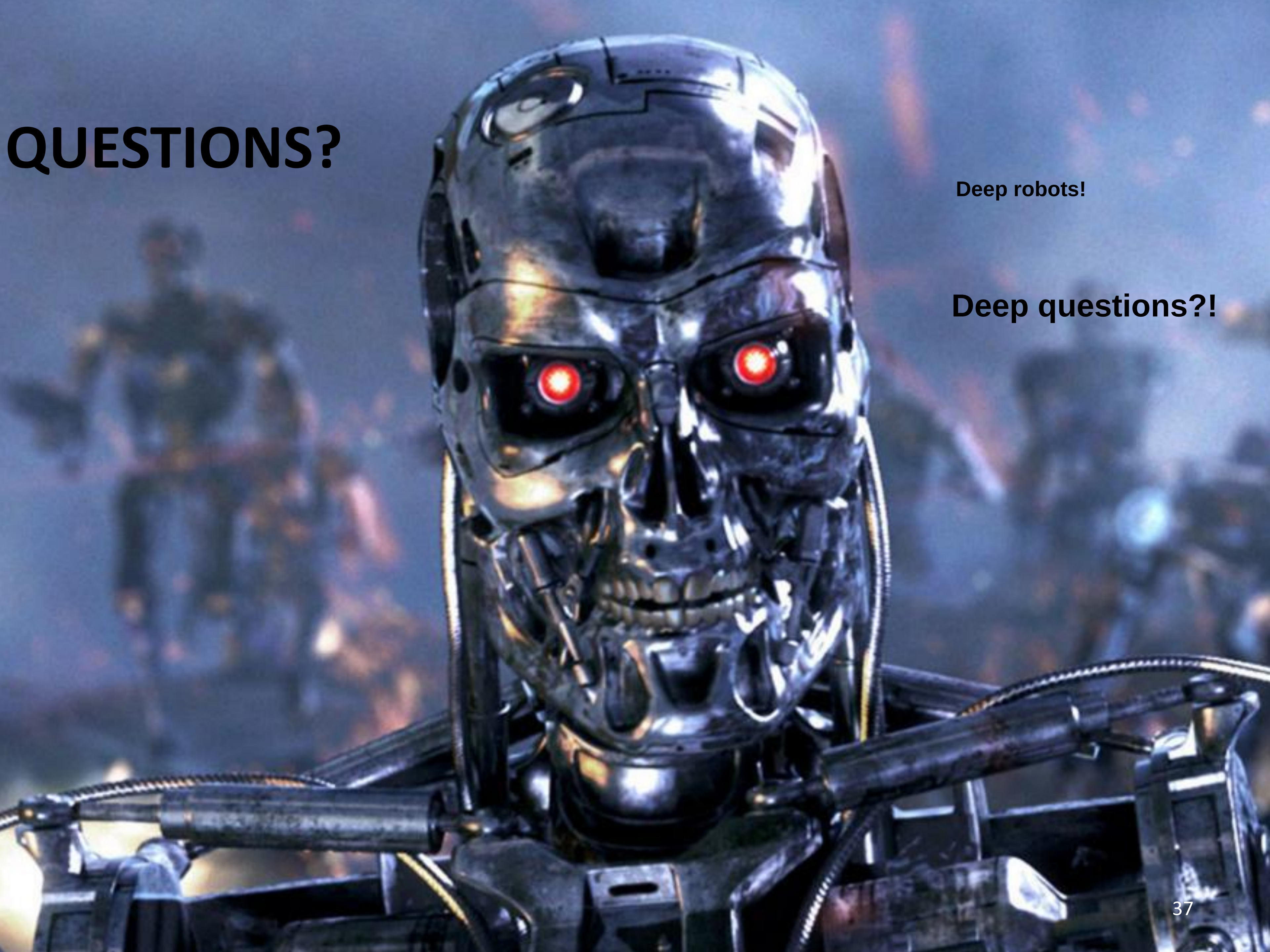


https://lvdmaaten.github.io/tsne/

Conclusions

- Simple overview of general machine learning
- Coverage of some foundational machine learning ideas / concepts
  - These ideas certainly apply to “deep learning”
  - Three key components to an ML system
    - Representation
    - Evaluation
    - Optimization
- Practical considerations!
  - Think about your problem first!

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.