

Machine Learning

CS 539

Worcester Polytechnic Institute

Department of Computer Science

Instructor: Prof. Kyumin Lee

A Little About Me

- Software Engineer at [NHN](#) from 2006-2008, South Korea
- Ph.D. at [Texas A&M](#) in 2013
- Research Internship at [eBay Research Labs](#) in 2011 and [IBM Research](#) in 2012
- Summer Research Fellow at Air Force Research Lab in 2022
- Associate Professor in CS, DS and AI at WPI
- Graduate Coordinator of AI program.
- Director of [Infolab](#): Information Retrieval, Machine Learning/AI, Natural Language Processing, Social Computing and AI for social good

Course Objectives

- Introduce
 - theory, design, and implementation of machine learning algorithms and applications.
- Study
 - supervised, unsupervised, and reinforcement learning such as regression, classification, clustering, reinforcement learning, ensemble methods and deep learning.
- Read
 - research papers and apply research results in machine learning.

Course Structure

Course Information

- Instructor
 - Kyumin Lee
 - `kmlee@wpi.edu`
 - Office hours: Thursday 7:00-8:00pm
 - Zoom link: <https://wpi.zoom.us/my/kyumin.lee>
- Class hours:
 - Each week's lecture recording will be uploaded to Canvas on Wednesday

Course Information

- Course web page
 - Check the course Canvas page and Schedule page
 - <https://canvas.wpi.edu/>
- Communication
 - Will post important announcements via Canvas mailing list
 - Send us a question via email

Course Materials

- Primary Textbook:
 - Machine Learning: The Art and Science of Algorithms That Make Sense of Data by Peter Flach
- Additional Readings will be drawn from the following books:
 - Learning From Data by Y. S. Abu-Mostafa, M. Magdon-Ismael, and H.T. Lin., AML Book
 - Machine Learning by Tom Mitchell, McGraw Hill
 - Machine Learning Lecture Notes by Andrew Ng
 - Reinforcement Learning: An Introduction by Sutton and Barto, MIT Press
 - Deep Learning with Python by François Chollet, Manning Publications
 - Selected Research Papers

Optional

- For a more advanced treatment of machine learning topics, you may read one of the following books:
 - Pattern Recognition and Machine Learning by Bishop, Springer.
 - Machine Learning: A Probabilistic Perspective by Kevin P. Murphy, MIT Press.
 - Deep Learning by Yoshua Bengio, Ian Goodfellow, and Aaron Courville.

Course Communication

- Check schedule page to understand upcoming due dates and topics
- I will post important announcements to the Canvas page and sometimes send messages via Canvas message service
 - Make sure to check the Canvas messages or get them via email.
 - <https://community.canvaslms.com/t5/Student-Guide/How-do-I-manage-my-Canvas-notification-settings-as-a-student/ta-p/434>
- You may email me anytime ... but I may not respond during the weekend.
- The best way to discuss general questions or share something cool stuff is to email it via Canvas mailing list.

Class Structure

- Lectures
 - I'll teach machine learning techniques
- Your part
 - Homework
 - 4 assignments
 - Quizzes
 - Project
 - Proposal, execution, workshop presentation (recording and live Q&A)

Grading

- 40% Assignments
- 30% Quizzes
- 30% Project

Assignments

- 4 assignments
 - Be familiar with Python
- Submit your solution to Canvas
- Please make sure to submit your solution by the due date
 - late submission will not be accepted

Quizzes

Quizzes

- Four quizzes will be taken via Canvas
- Each quiz date/week is already posted in the schedule page. Each quiz will be released on Monday of the scheduled week. Please make sure to take it on **Tuesday** because you will have **only one day to take it**.

Project

The Project

- 3~4 person team (I will randomly assign project team members in the 3rd week unless you express your own preference in advance).
- Project idea:
 - Propose anything you wish (related to machine learning)
 - You are encouraged to talk to me
- **30% of your final grade!!**

Project Grading Criteria

- [10%] Project Proposal: June 25 by 11:59pm
- [10%] Project Website: July 23 by 11:59pm
- [10%] Project Workshop: July 23 via a live Zoom session at 6pm

So far...

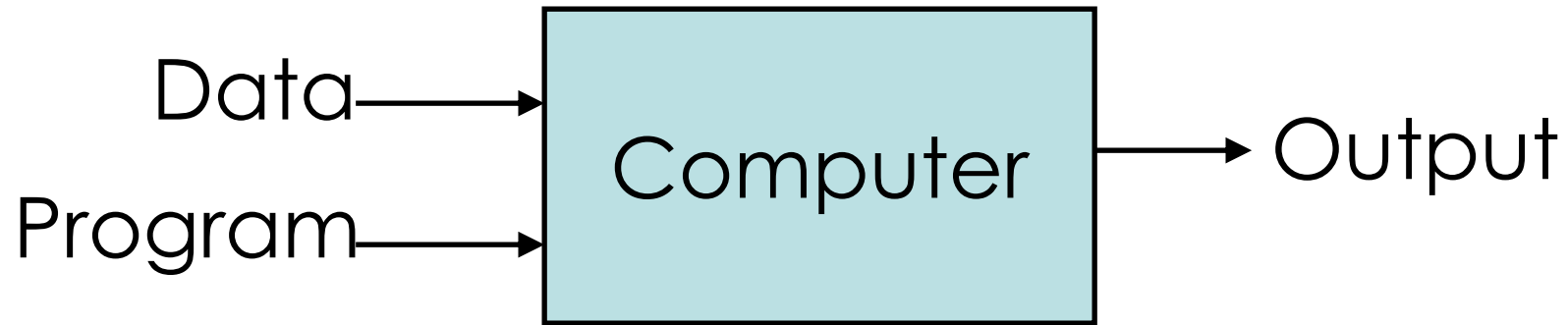
- Read the syllabus carefully
- Be familiar with Python for Assignments
- I will randomly assign project teams by June 4

Introduction to Machine Learning

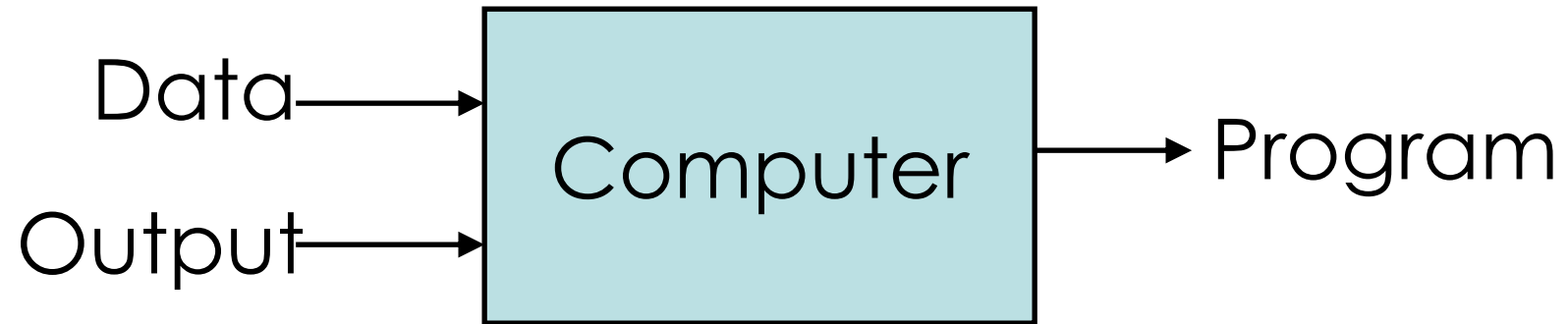
What is Machine Learning?

- “Learning is any process by which a system improves performance from experience.”
 - Herbert Simon
- Definition by Tom Mitchell (1998):
 - Machine Learning is the study of algorithms that
 - improve their performance P
 - at some task T
 - with experience E
 - A well-defined learning task is given by $\langle P, T, E \rangle$.

Traditional Programming

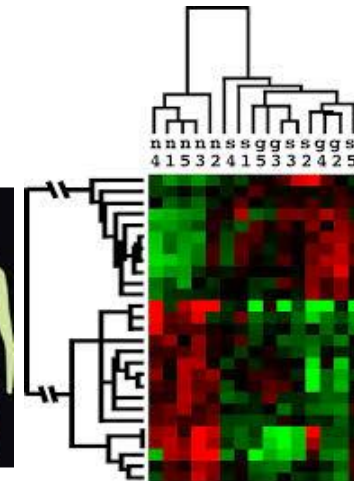


Machine Learning



When Do We Use Machine Learning?

- ML is used when:
 - Human expertise does not exist (navigating on Mars)
 - Humans can't explain their expertise (speech recognition)
 - Models are based on huge amounts of data (genomics)



- Learning isn't always useful:
 - There is no need to “learn” to calculate payroll

- A classic example of a task that requires machine learning:
 - It is very hard to say what makes a 2

0 0 0 1 1 1 1 1 1 2

2 2 2 2 2 2 2 3 3 3

3 4 4 4 4 4 5 5 5 5

6 6 7 7 7 7 8 8 8 8

9 9 9 9 9 9 9 9 9 9

Some more examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns:
 - Facial identities or facial expressions
 - Handwritten or spoken words
 - Medical images
- Generating patterns:
 - Generating images or motion sequences
 - Generating texts



Generated bedrooms. Source: "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" <https://arxiv.org/abs/1511.06434v2>

GROVER – A State-of-the-Art Defense against Neural Fake News

Generate

Detect

Examples

Why Bitcoin is a great investment

Select an example or build an article below

Inputs

Domain

nytimes.com

Generate

Give your article a domain source to write against

Date

June 6, 2019

Generate

Give your article a date to write against

Authors

Paul Krugman

Generate

Give your article one or more authors to write against

Headline

Why Bitcoin is a great investment

Generate

Give your article a headline to write against

Article

There is some truth to the Forbes cover story about Bitcoin, which shows the devastation the cryptocurrency has already endured. In the following months, however, the problems will soon be nothing compared to the advantages.

Lately, cryptocurrencies have been using artificial intelligence and networks, known as blockchain, to try to make the financial sector operate more like the real world. For example, something called Ethereum — the technology used by Bitcoin and other cryptocurrencies — is also being used to make exchanges usefully automated. It's not clear if financial innovation will ever reach the heights it has reached in the past. But it is promising.

Although some people worry that the technology is the mark

Generate

Fake Article

nytimes.com

Why Bitcoin is a great investment

June 6, 2019 - Paul Krugman

There is some truth to the Forbes cover story about Bitcoin, which shows the devastation the cryptocurrency has already endured. In the following months, however, the problems will soon be nothing compared to the advantages.

Lately, cryptocurrencies have been using artificial intelligence and networks, known as blockchain, to try to make the financial sector operate more like the real world. For example, something called Ethereum — the technology used by Bitcoin and other cryptocurrencies — is also being used to make exchanges usefully automated. It's not clear if financial innovation will ever reach the heights it has reached in the past. But it is promising.

Although some people worry that the technology is the mark of a corrupt and dangerous era, the future is truly bright for blockchain technology.

Here's another point in favor of cryptocurrency: Money is a privilege, not a right. Some countries take great care to give citizenship to their citizens and allow them to vote — precisely because they understand that there is a difference between a privilege and a right. The same is true of access to credit, a real issue for the poor, and of employment and other opportunities.

For most people, including rich people, however, there is nothing wrong with being in debt. So it will be important to do something to reduce debt.

The best way to do that is to exchange wealth that people have created with other people's money, not with paper. That makes it possible to keep money in the real world and use that money effectively. Otherwise, you will have the same real problems that prevail in the developed world, including price deflation, rising inequality and increasingly lopsided distribution of income and wealth.

That brings us to some tips about investing in cryptocurrencies. Just as you should know the drawbacks of borrowing, you should know the drawbacks of investing in a risky, perhaps deadly, asset that no one really knows how to

Paper: <https://arxiv.org/abs/1905.12616>

Demo: <https://grover.allenai.org/>

Any other ML examples/applications?

Some more examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns:
 - Facial identities or facial expressions
 - Handwritten or spoken words
 - Medical images
- Generating patterns:
 - Generating images or motion sequences
 - Generating texts
- Recognizing anomalies:
 - Unusual credit card transactions
 - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
 - Future stock prices or currency exchange rates

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- [Your favorite area]

Samuel's Checkers-Player

- “Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.” -Arthur Samuel (1959)



Defining the Learning Task

Improve on task T , with respect to
performance metric P , based on experience E

T : Playing checkers

P : Percentage of games won against an arbitrary opponent

E : Playing practice games against itself

T : Recognizing hand-written words

P : Percentage of words correctly classified

E : Database of human-labeled images of handwritten words

T : Driving on four-lane highways using vision sensors

P : Average distance traveled before a human-judged error

E : A sequence of images and steering commands recorded while observing a human driver.

T : Categorize email messages as spam or legitimate.

P : Percentage of email messages correctly classified.

E : Database of emails, some with human-given labels

State of the Art Applications of Machine Learning

Large Language Models

Gemini ▾

See the latest updates to the [Gemini Apps Privacy Hub](#)



You

machine learning applications?



ChatGPT

Machine learning has an extensive range of applications across various fields. Here are some ones:

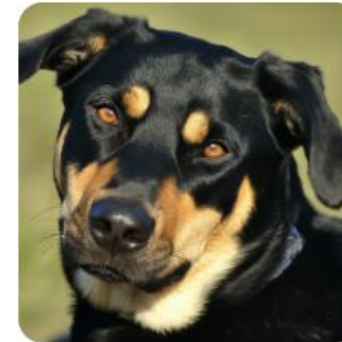
1. **Natural Language Processing (NLP):** Used for sentiment analysis, language translation, text summarization, and more.
2. **Computer Vision:** Applied in image recognition, object detection, facial recognition, medical analysis, autonomous vehicles, and surveillance systems.
3. **Recommendation Systems:** Employed by companies like Amazon, Netflix, and Spotify to suggest products, movies, music, and other content to users based on their preferences and behavior.
4. **Healthcare:** Used for disease diagnosis, medical image analysis, drug discovery, personalized treatment plans, and health monitoring.
5. **Finance:** Applied in fraud detection, risk assessment, algorithmic trading, credit scoring, and personalized financial services.



generate image with a dog



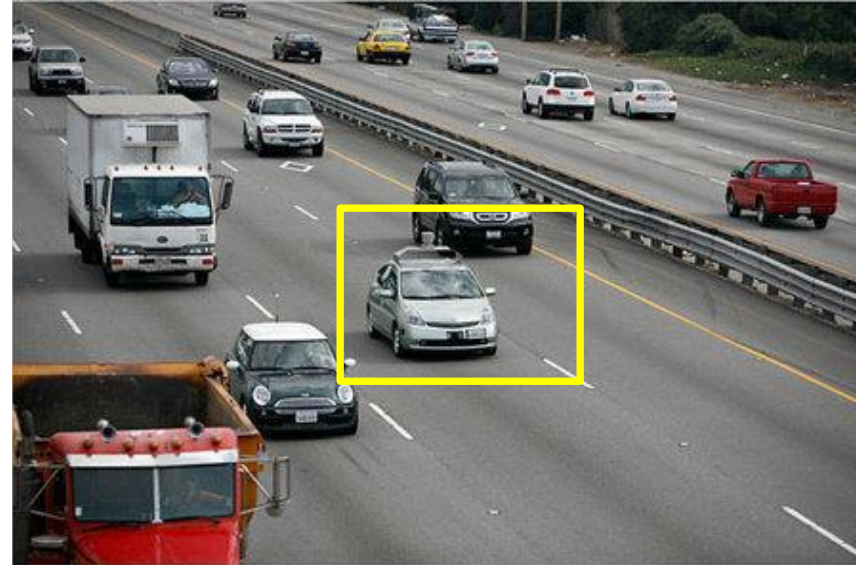
Sure, here is an image with a dog:



Generate more

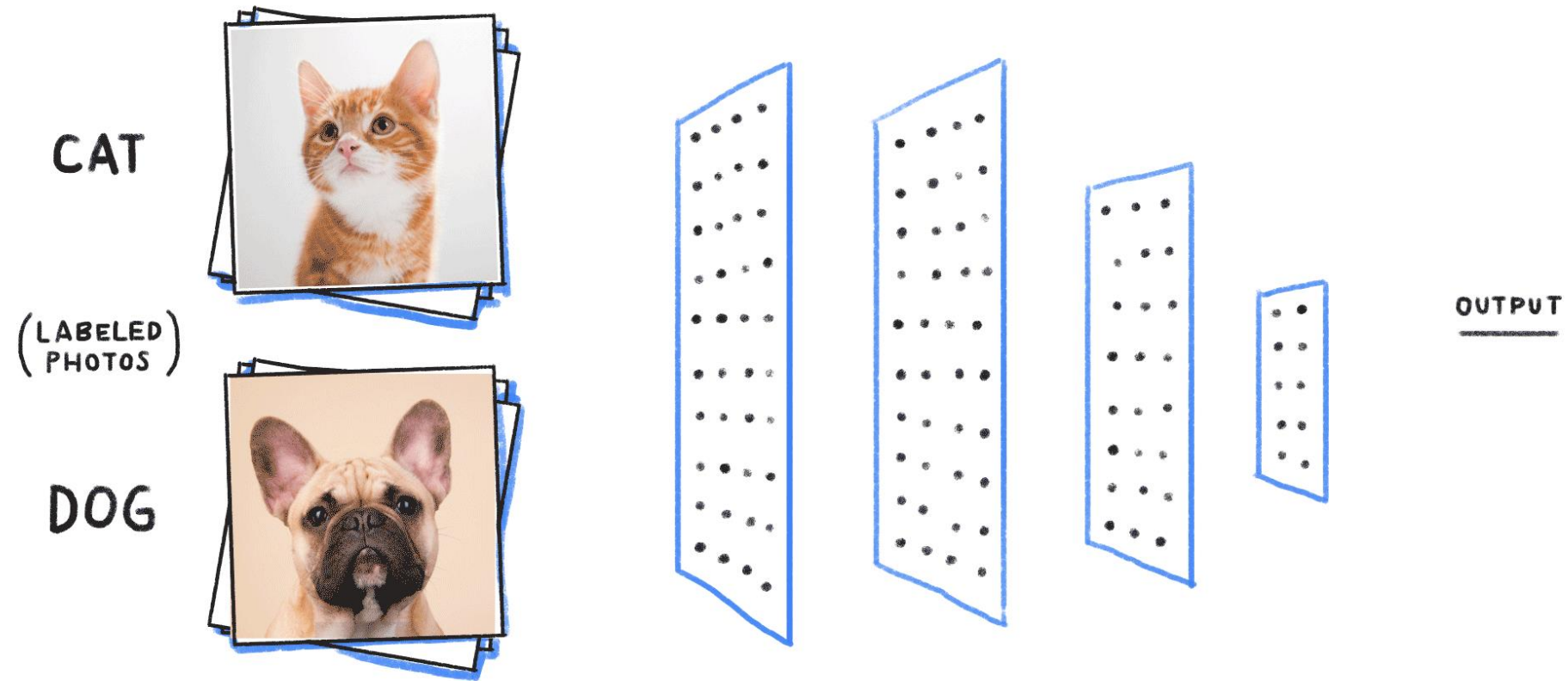


Autonomous Cars



- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2023, twenty-nine states have legalized autonomous cars

Image Classification



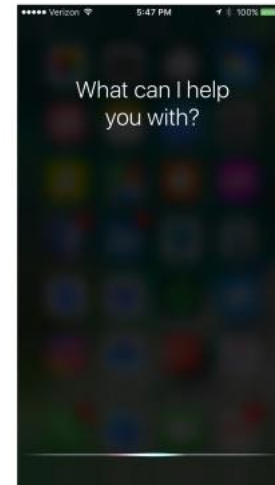
Digital Assistant



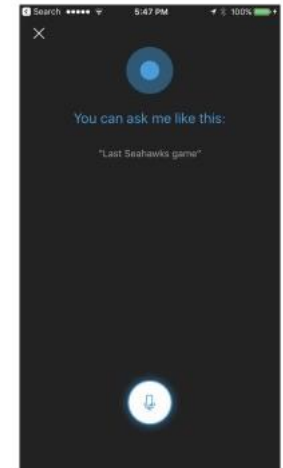
AMAZON'S ALEXA



GOOGLE'S ASSISTANT



APPLE'S SIRI



MICROSOFT'S CORTANA

AI/Machine Learning/Deep Learning

The New York Times

The Machines Are Learning, and So Are the Students

Artificial intelligence is starting to take over repetitive tasks in classrooms, like grading, and is optimizing coursework and revolutionizing the preparation for college entrance exams.



DISCOVER

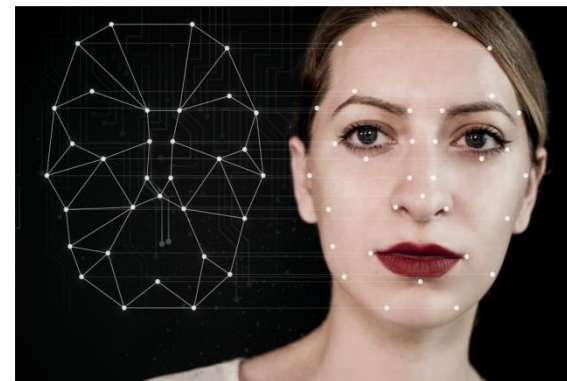
THE SCIENCES | MIND | TECHNOLOGY | HEALTH | ENVIRONMENT | PLANET EARTH

TECHNOLOGY

Deepfakes: The Dark Origins of Fake Videos and Their Potential to Wreak Havoc Online

The presence of videos and images altered with artificial intelligence are nearly doubling every six months. But the problem might be more than meets the eye.

By Jennifer Walter | August 13, 2020 12:05 PM



Forbes

EDITORS' PICK | 3,752 views | Aug 20, 2020, 01:01am EDT

Chatbots Are Machine Learning Their Way To Human Language



Adrian Bridgwater Senior Contributor
Cloud

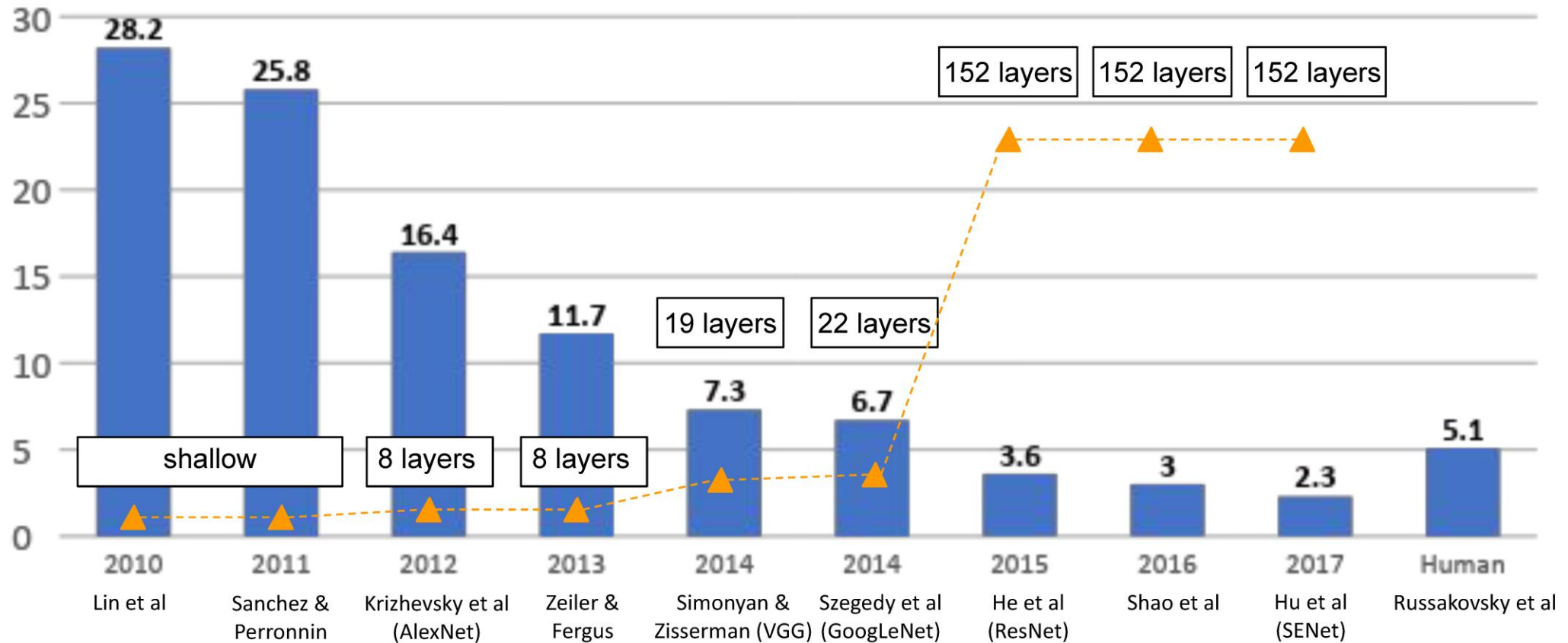
I track enterprise software application development & data management.











Moveworks founding team from left to right Vaibhav Nivargi, CTO; Bhavin Shah, CEO; Varun Singh, VP ... [+] MOVEWORKS

Computers and humans have never spoken the same language. Over and above speech recognition, we also need computers to understand the semantics of written

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm
1	JDExplore d-team	Vega v1		91.3	73.8	97.9	94.5/92.6	93.5/93.1	76.7/91.1	92.1	91.9
2	Microsoft Alexander v-team	Turing NLR v5		91.2	72.6	97.6	93.8/91.7	93.7/93.3	76.4/91.1	92.6	92.4
3	DIRL Team	DeBERTa + CLEVER		91.1	74.7	97.6	93.3/91.1	93.4/93.1	76.5/91.0	92.1	91.8
4	ERNIE Team - Baidu	ERNIE		91.1	75.5	97.8	93.9/91.8	93.0/92.6	75.2/90.9	92.3	91.7
5	AliceMind & DIRL	StructBERT + CLEVER		91.0	75.3	97.7	93.9/91.9	93.5/93.1	75.6/90.8	91.7	91.5
6	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.8	71.5	97.5	94.0/92.0	92.9/92.6	76.2/90.8	91.9	91.6
7	HFL iFLYTEK	MacALBERT + DKM		90.7	74.8	97.0	94.5/92.6	92.8/92.6	74.7/90.6	91.3	91.1
	8	PING-AN Omni-Sinitic		90.6	73.5	97.2	94.0/92.0	93.0/92.4	76.1/91.0	91.6	91.3
9	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9
10	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8
21	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8

Types of Learning

Types of Learning

- **Three Types of Major Learning**
 - **Supervised (inductive) learning**
 - Given: training data + desired outputs (labels)
 - **Unsupervised learning**
 - Given: training data (without desired outputs)
 - **Reinforcement learning**
 - Rewards from sequence of actions
- **Other Types of Learning**
 - **Semi-supervised learning**
 - Given: training data + a few desired outputs
 - **Transfer Learning**
 - A model, which is developed for a task, is reused as the starting point for a model on a second task
 - E.g., Resnet, BERT

Supervised Learning: Regression

- Given $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
 - y is real-valued == regression

