CS 4644 / 7643-A: Deep Learning

Website: https://www.cc.gatech.edu/classes/AY2023/cs7643_spring/

Zoom: https://gatech.zoom.us/j/91625047484

Piazza:	https://piazza.com/class/lcl94yjxkbb59e/
	(code: DLSPR2023)
Canvas:	https://gatech.instructure.com/courses/294004 (4644)
	https://gatech.instructure.com/courses/293996 (7643)
Gradescope:	https://www.gradescope.com/courses/484319 (4644)
	https://www.gradescope.com/courses/484320 (7643)

Zsolt Kira

School of Interactive Computing

Georgia Tech

Are you in the right place?

- This is CS 4644(DL) / CS 7643-A
  - "On campus" class

- This is NOT CS 7643-O01/OAN/Q/R
  - Online class for OMSCS program

Spring 23 Delivery Format

- In-Person
  - Clough UG Learning Commons 152
- Streaming & Recording
  - We STRONGLY encourage you to attend the lectures in person
  - Lectures will be streamed over zoom (link on canvas/above)
  - Lectures are recorded and available for viewing
- Office hours, HW/project submissions online
- **Remember: Content is free online.**
  - **You are here for the interaction and the insight.**

Outline for Today

- What is Deep Learning, the field, about?

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

What are we here to discuss?

Some of the most exciting developments in

Machine Learning,
Vision, NLP, Speech, Robotics
& AI in general

in the last decade!

Concepts

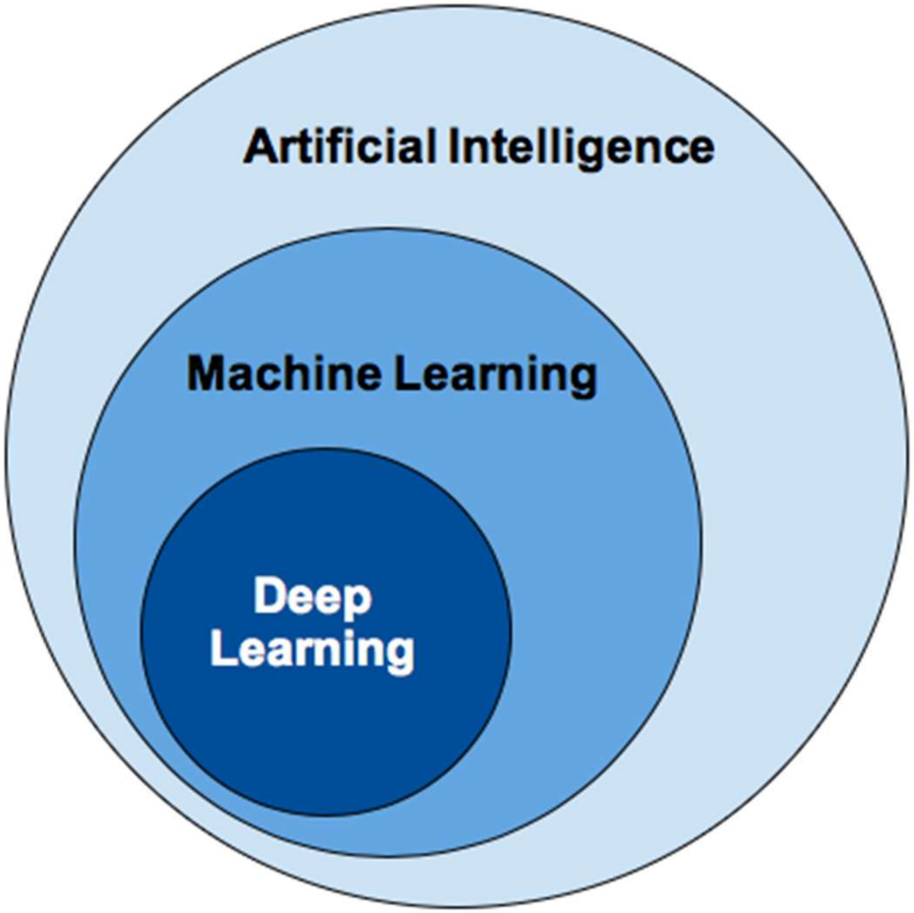


Image Credit: https://www.sumologic.com/blog/machine-learning-deep-learning/

What is (general) intelligence?

- Boring textbook answer

The ability to acquire and apply knowledge and skills

  - Dictionary

- Many others
  - Survival, various types/aspects of intelligence, etc.

What is artificial intelligence?

- Boring textbook answer

Intelligence demonstrated by machines

  – Wikipedia

- What others say:

The science and engineering of making computers behave in ways that, until recently, we thought required human intelligence.

  – Andrew Moore, CMU

What is machine learning?

- A favorite

Study of algorithms that
improve their performance (P)
at some task (T)
with experience (E)

  – Tom Mitchell, CMU

So what *is* Deep (Machine) Learning?

- **Objective:** Representation Learning

- **Model:** Neural Networks

- **Learning Method:** Deep
  Unsupervised/Supervised/Reinforcement/Structured/
  <insert-qualifier-here>
  Learning

- **Simply**: Deep Learning

So what *is* Deep (Machine) Learning?

- A few different ideas:

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations
  - No single neuron "encodes" everything
  - Groups of neurons work together

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

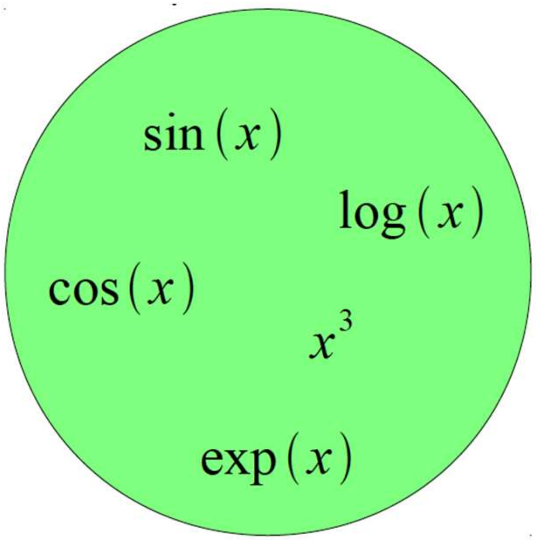
sample → spectral band → formant → motif → phone → word

NLP

character → word → NP/VP/.. → clause → sentence → story

Building A Complicated Function

Given a library of simple functions

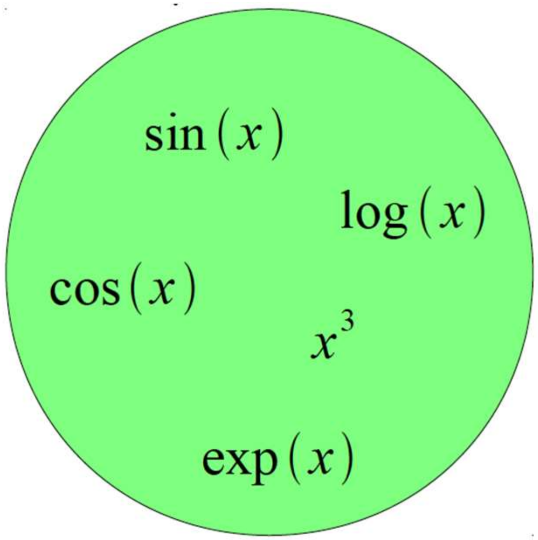


Compose into a

complicate function

Building A Complicated Function

Given a library of simple functions

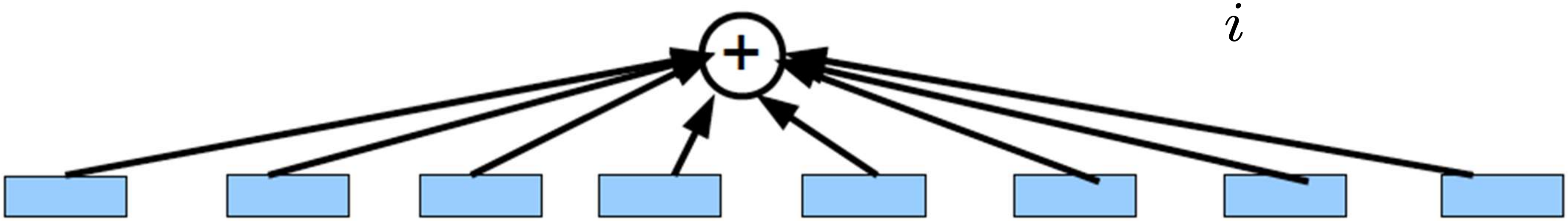


Compose into a complicate function

Idea 1: Linear Combinations

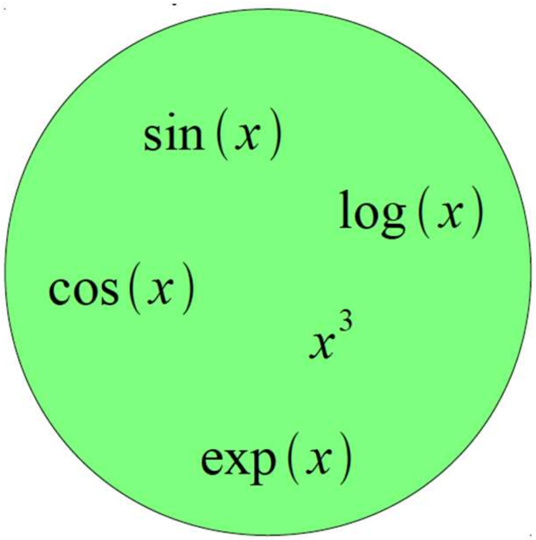
- Boosting
- Kernels
- …

$$f(x) = \sum_i \alpha_i g_i(x)$$



Building A Complicated Function

Given a library of simple functions



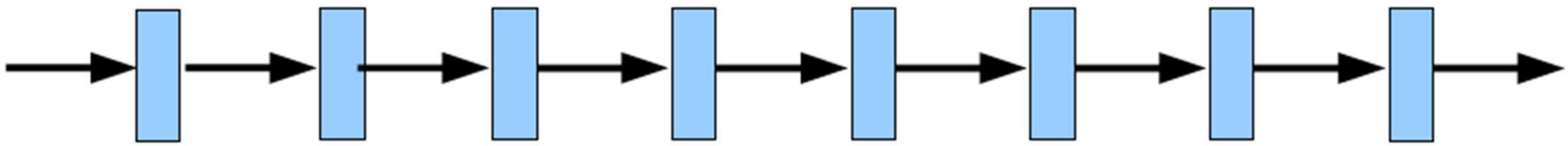
Compose into a

complicate function

Idea 2: Compositions

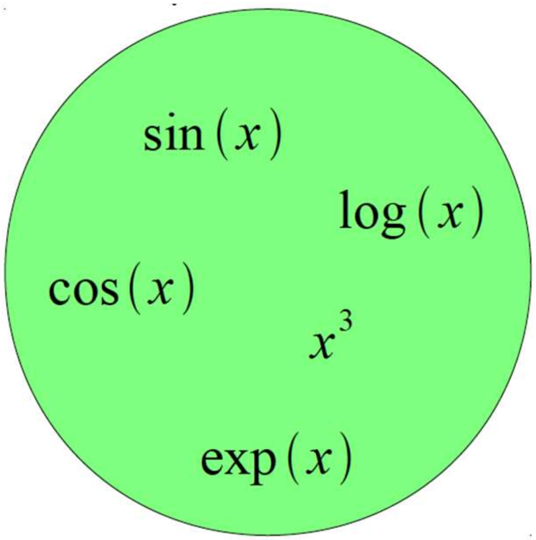
- Deep Learning
- Grammar models
- Scattering transforms…

$$f(x) = g_1(g_2(\ldots(g_n(x)\ldots))$$



Building A Complicated Function

Given a library of simple functions



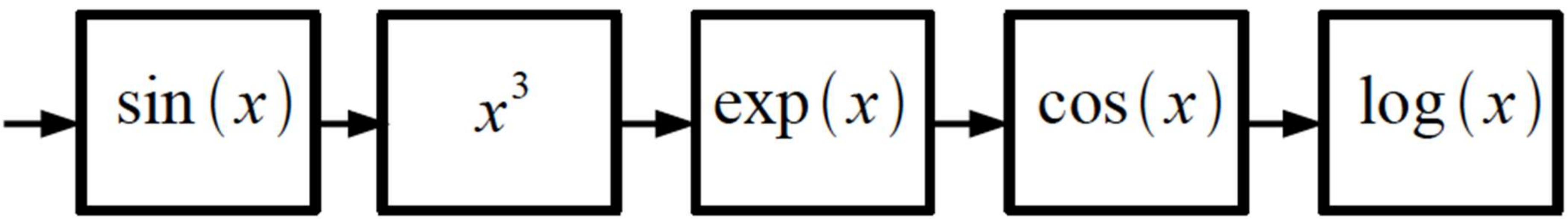
Compose into a

complicate function

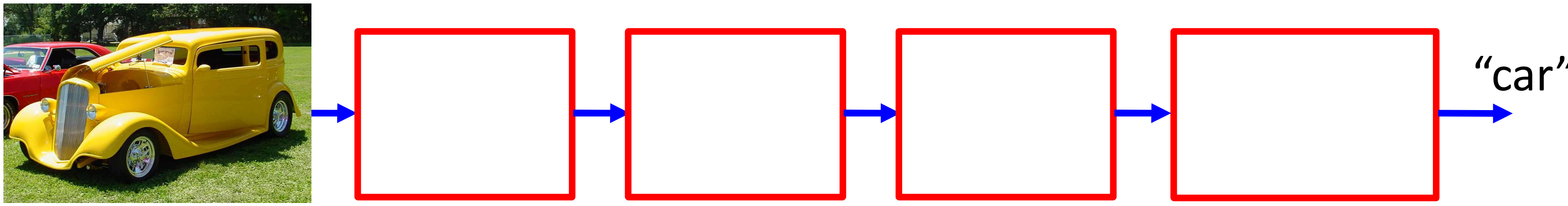
Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

$$f(x) = \log(\cos(\exp(\sin^3(x))))$$

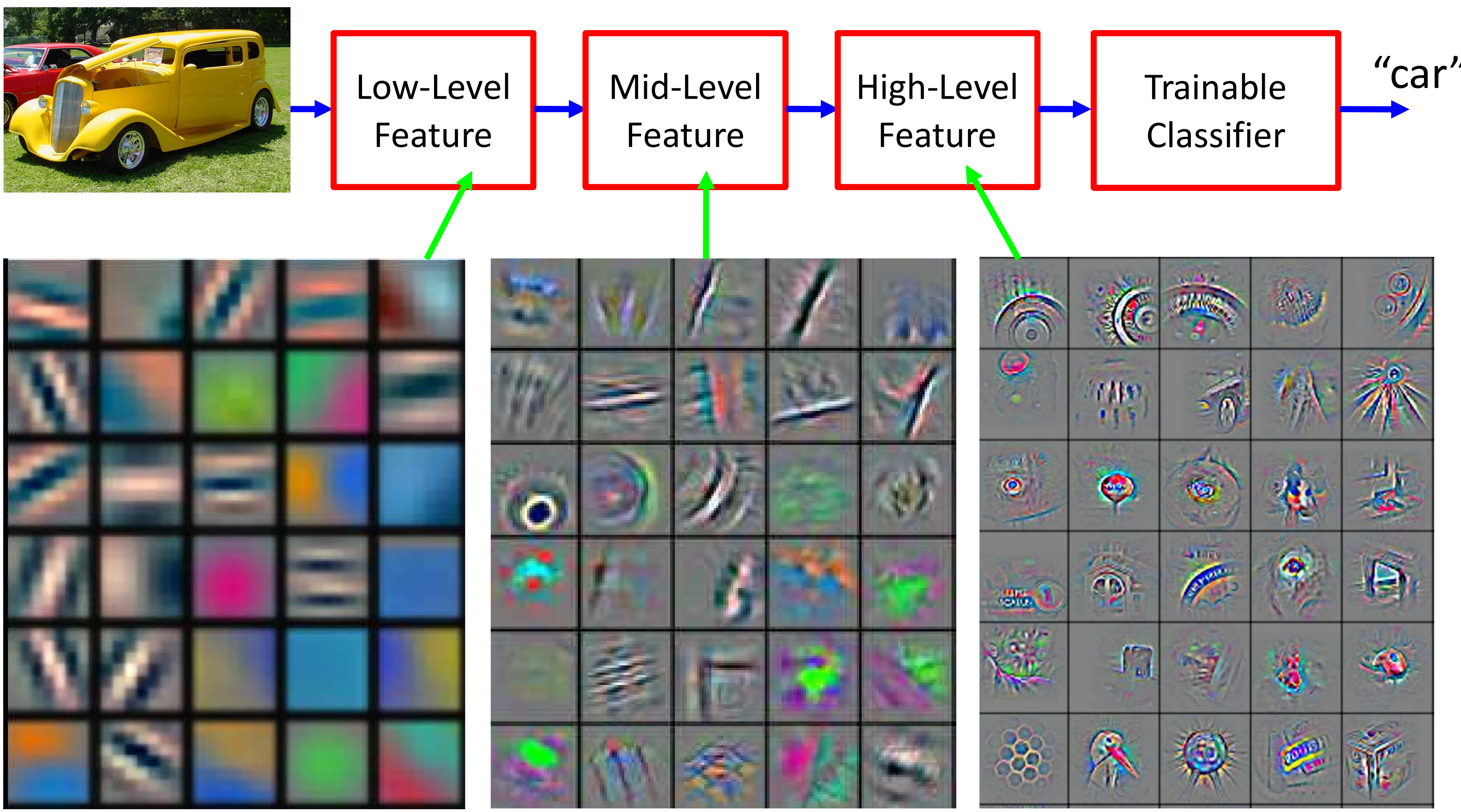


Deep Learning = Hierarchical Compositionality



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

So what *is* Deep (Machine) Learning?

- A few different ideas:

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations
  - No single neuron "encodes" everything
  - Groups of neurons work together

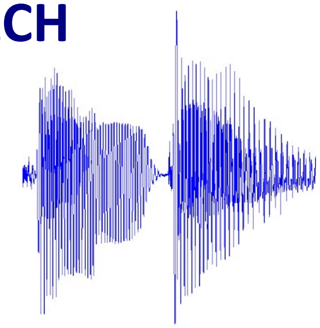
Traditional Machine Learning

VISION

hand-crafted features SIFT/HOG → your favorite classifier → "car"

fixed | learned

SPEECH



hand-crafted features MFCC → your favorite classifier → \'d ē p\

fixed | learned

NLP

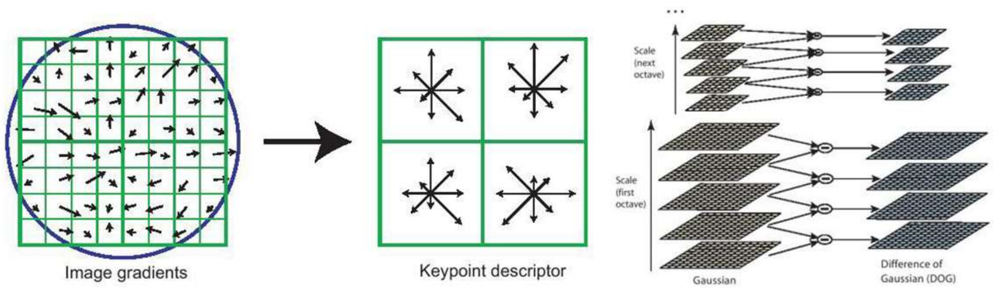
This burrito place is yummy and fun!

hand-crafted features Bag-of-words → your favorite classifier → "+"

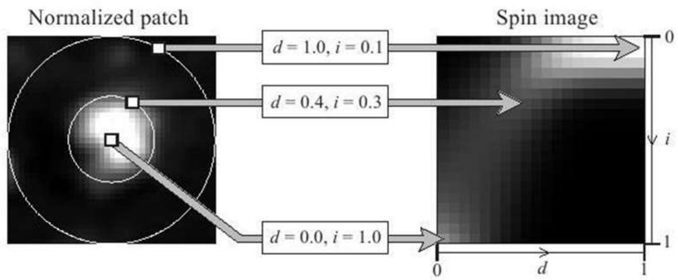
fixed | learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

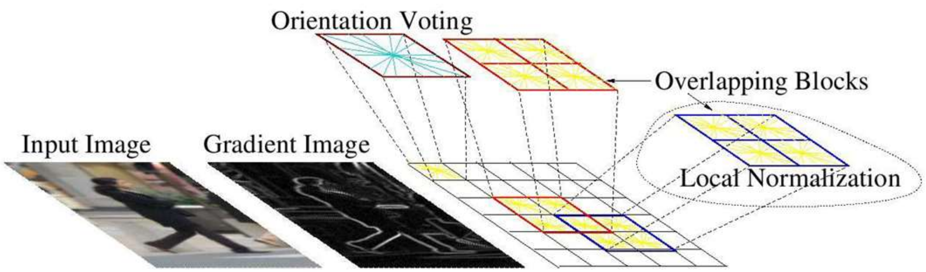
Feature Engineering



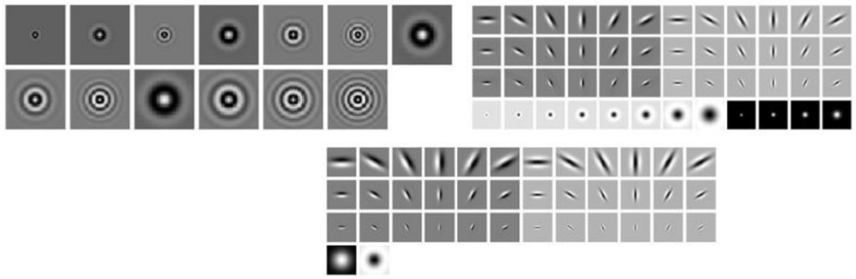
SIFT



Spin Images



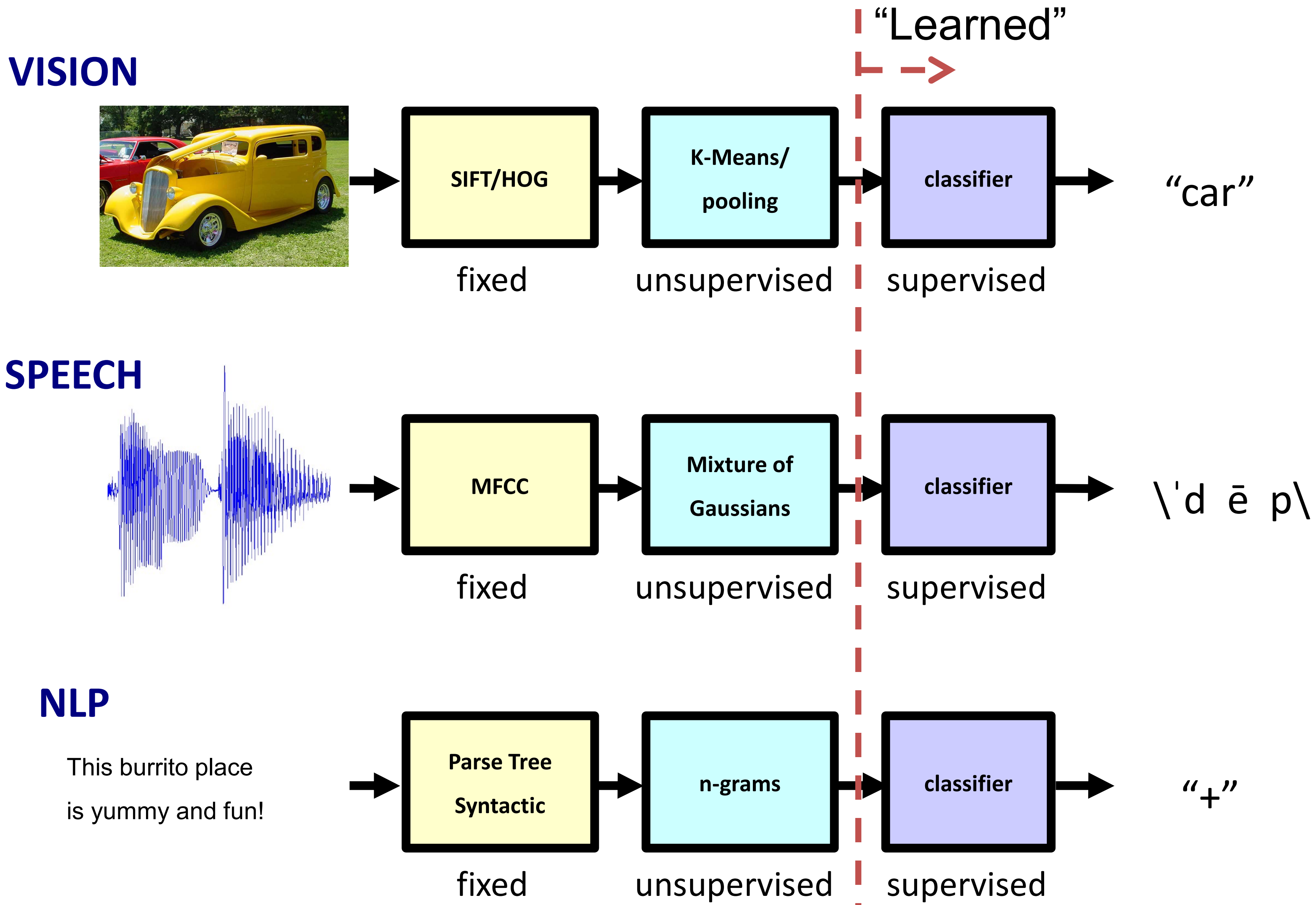
HoG



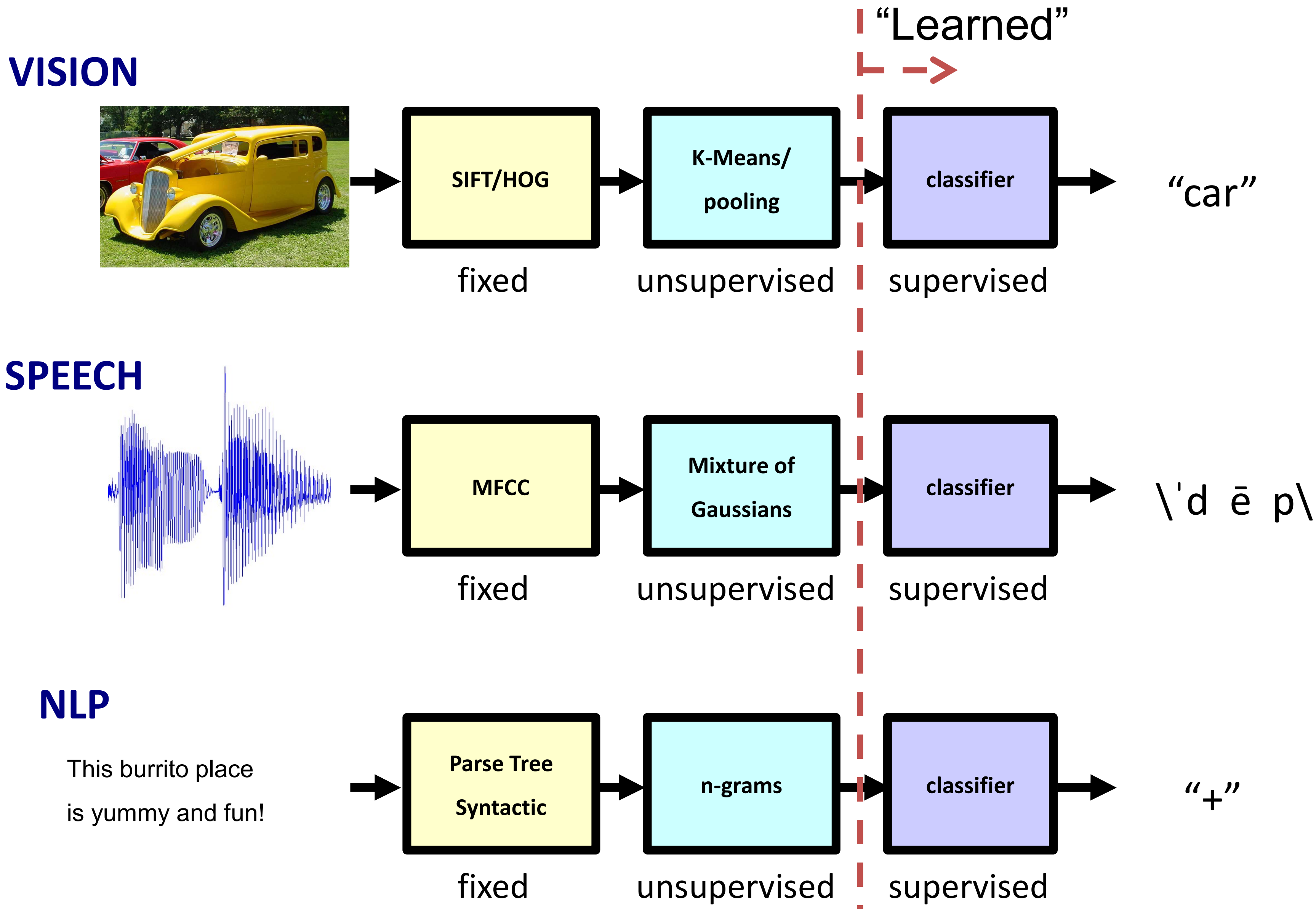
Textons

and many many more….

Traditional Machine Learning (more accurately)



Deep Learning = End-to-End Learning



“Shallow” vs Deep Learning

- “Shallow” models

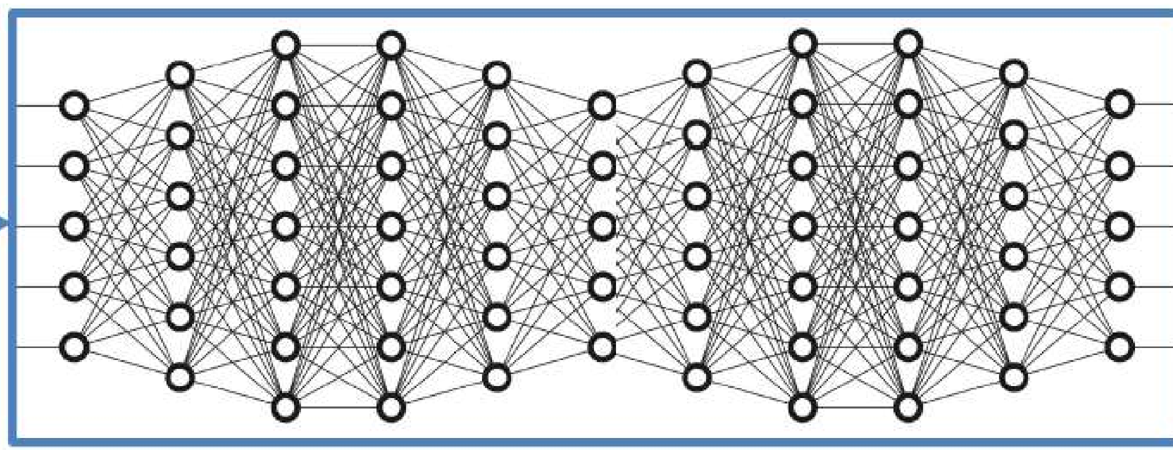
hand-crafted Feature Extractor

fixed

“Simple” Trainable Classifier

learned

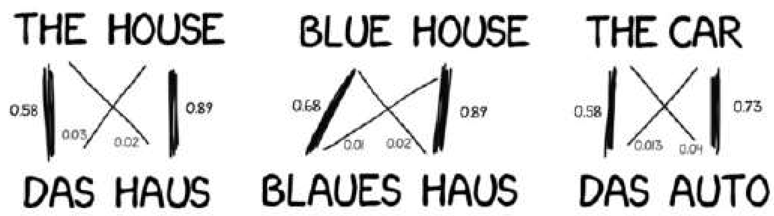
- Deep models

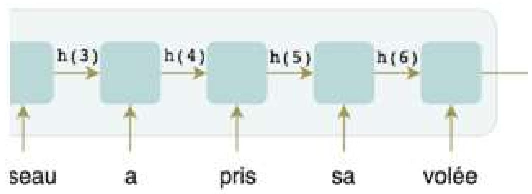


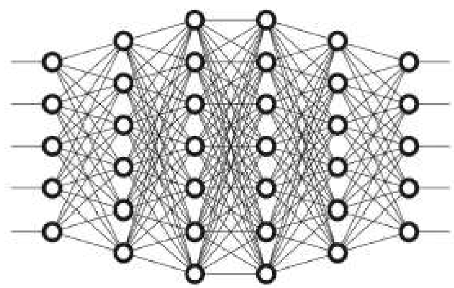
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

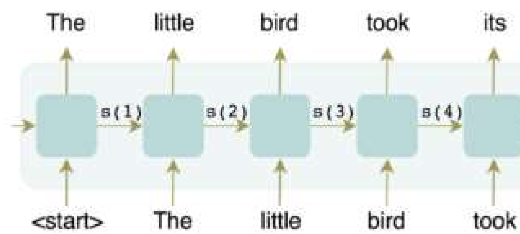
“Shallow” vs Deep Learning

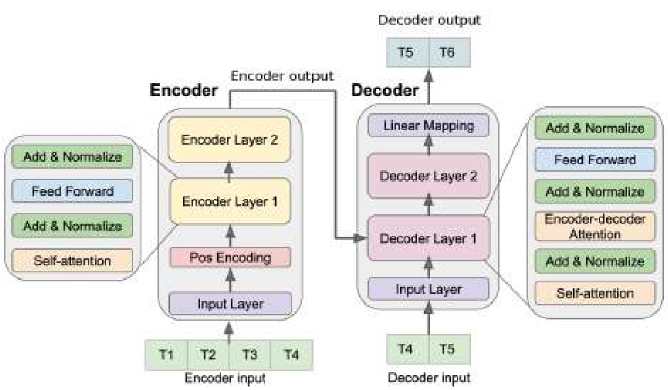
“Shallow” vs. deep language models











Transformer Models
(Vaswani *et al.*, 2017)

GPT3

GPT3 large language model
(Brown *et al.*, 2020)

https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5/

"Expert" intuitions can be misleading

- *"Every time I fire a linguist, the performance of our speech recognition system goes up"*
  – Fred Jelinik, IBM '98

- *"Because gradient descent is better than you"*
  – *Yann LeCun,* CVPR '13

“The Bitter Lesson”

- “The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation.” (Sutton, 2019)

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

So what *is* Deep (Machine) Learning?

- A few different ideas:

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Benefits of Deep/Representation Learning

- (Usually) Better Performance
  - Caveats: given enough data, similar train-test distributions, non-adversarial evaluation, etc., etc.

- New domains without “experts”
  - RGBD/Lidar
  - Multi-spectral data
  - Gene-expression data
  - Unclear how to hand-engineer

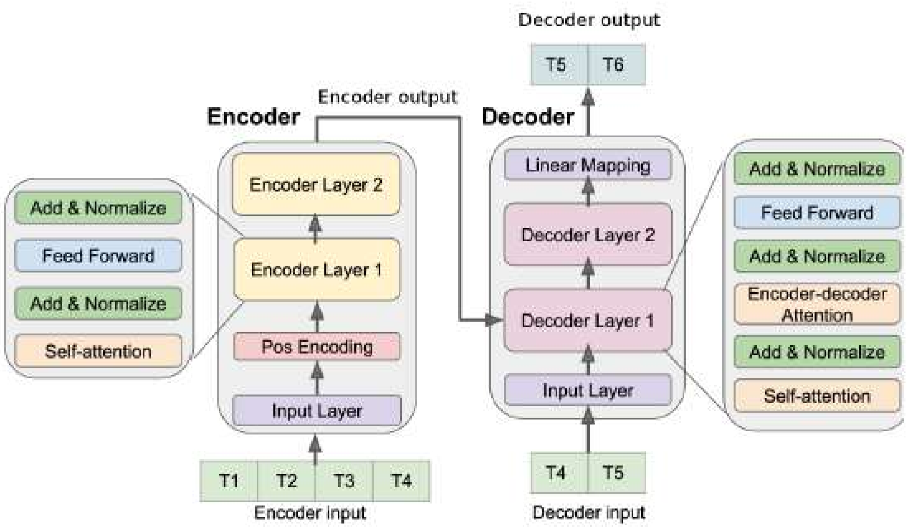
- “Homogenization” of model design
- New abilities emerge with more data and compute

Homogenization of Deep Learning

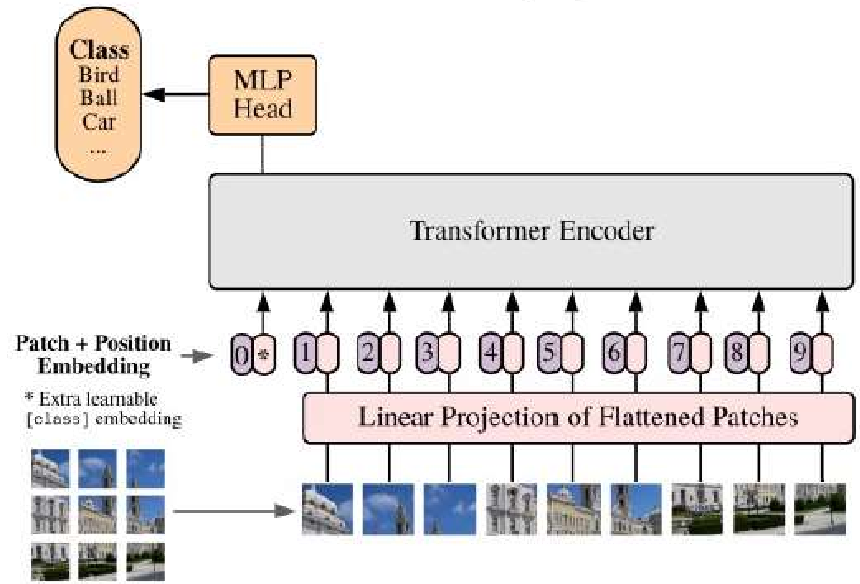
Homogenization is the **consolidation** of methodologies for building machine learning systems across a wide range of applications.

- Enabled by modular, plug-n-play nature of neural networks and training
- Consequence: Multi-modal, unified architectures, unified tasks (next-token prediction)

Example: The Transformer Models (Vaswani et al., 2017)



Transformer Models originally designed for NLP



Almost identical model (Visual Transformers) can be applied to Computer Vision tasks

Emergence of new behaviors

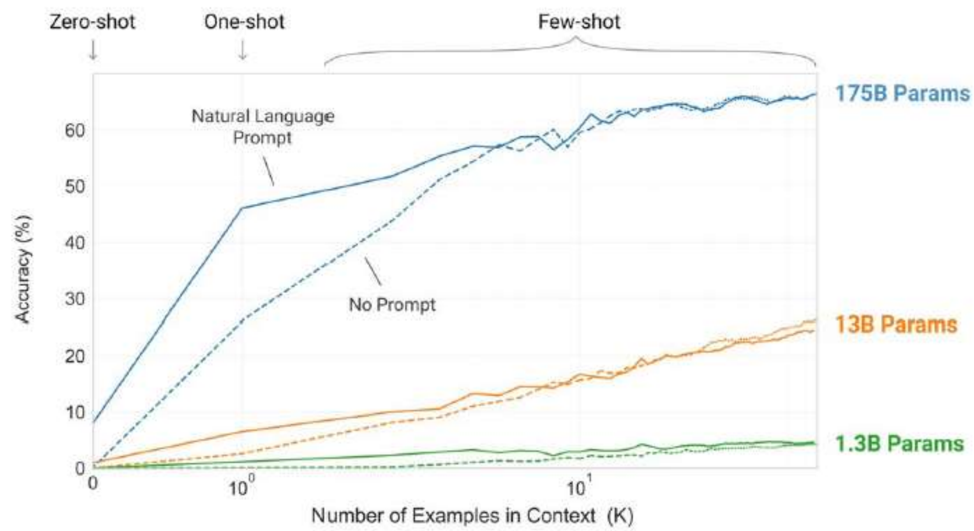
Emergence means that the behavior of a system is implicitly induced rather than explicitly constructed. For Deep Learning, emergence is often induced by larger model & more data.

Example: Compared to GPT-2's 1.5B parameter parameter model, GPT-3's 175-billion model permits "prompting" and "in-context learning", i.e., adapting to a new task simply by describing task.

Example input (prompt):

Ask it to translate French to English

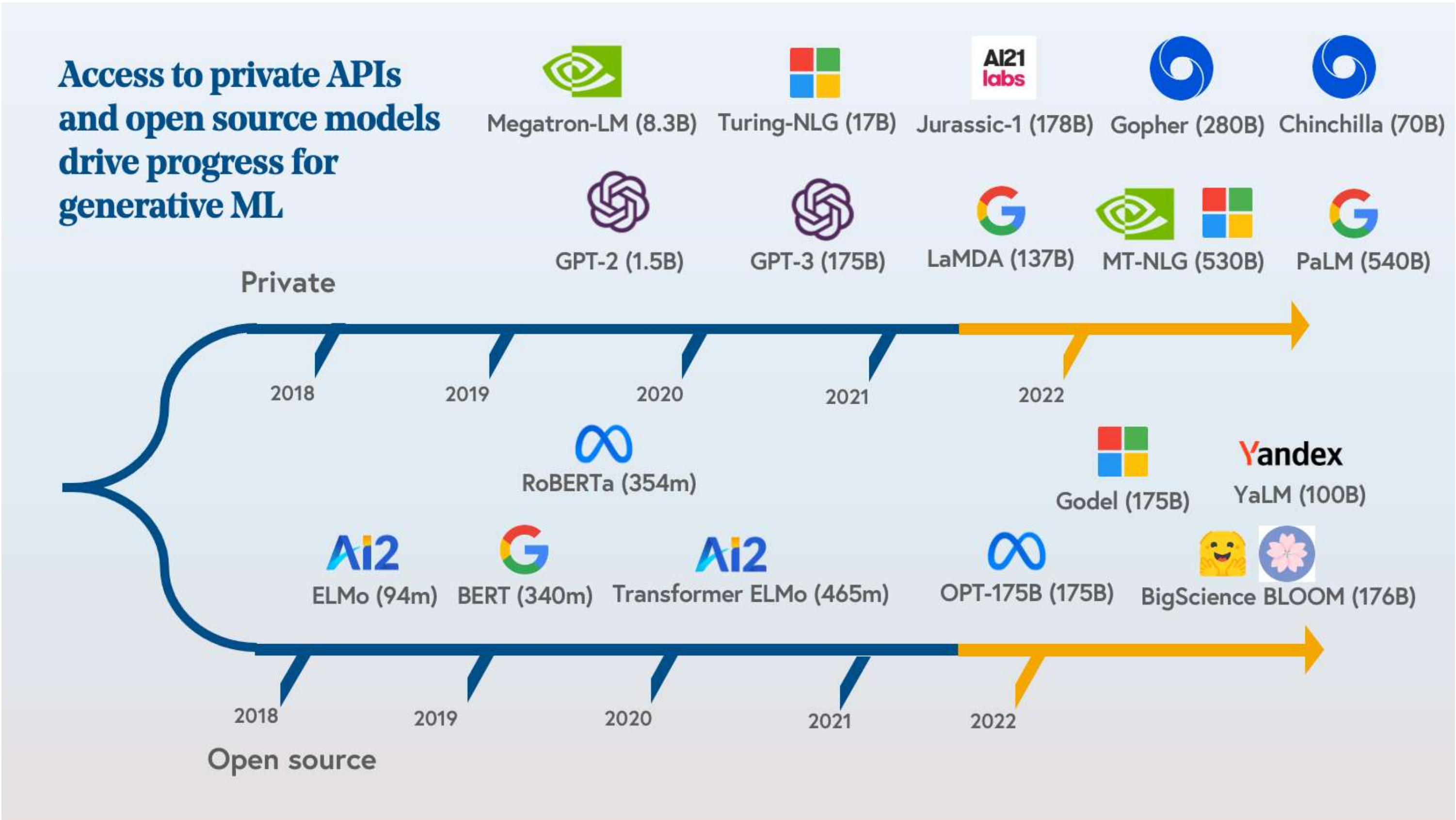
$$\underbrace{\text{maison} \rightarrow \text{house, chat} \rightarrow \text{cat, chien} \rightarrow}_{\text{prompt}} \ \underbrace{\text{dog}}_{\text{completion}} .$$



Bommasani et al., On the Opportunities and Risks of Foundation Models

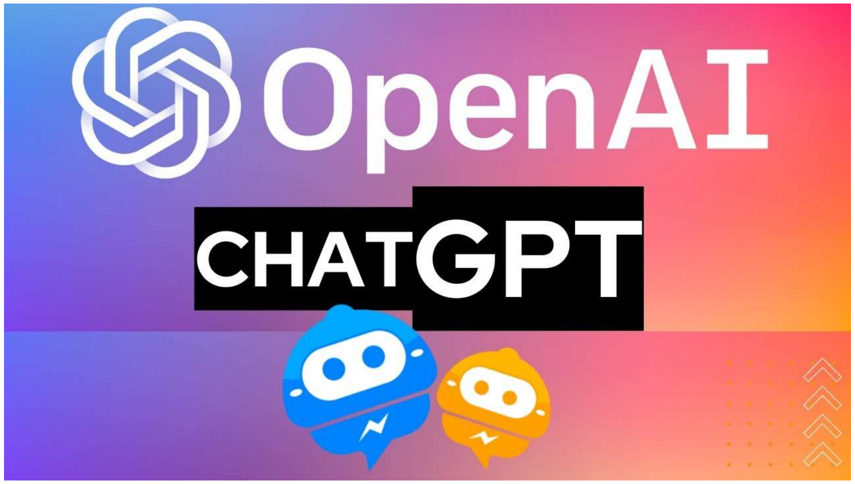
Pace of Progress

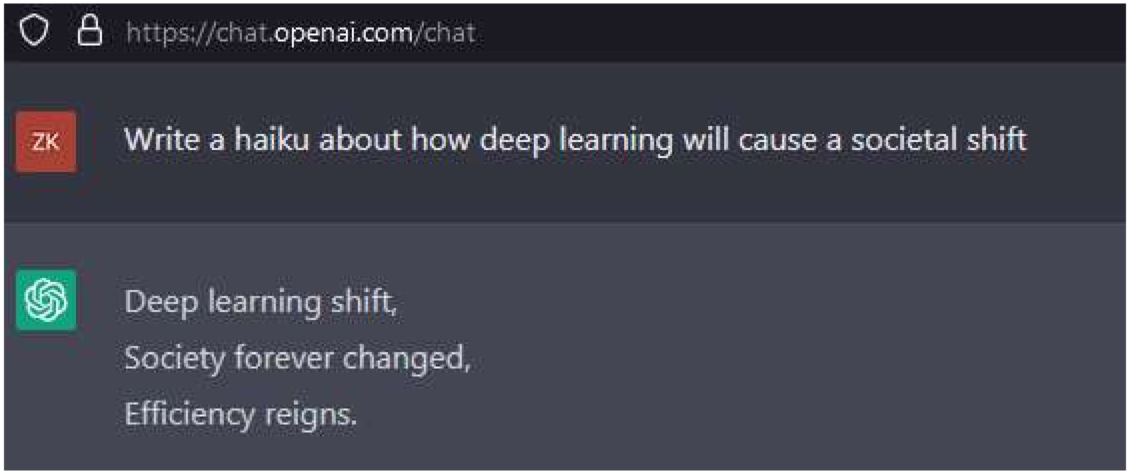
- The pace of progress has been **enormous** driven by:
  - Compute
  - Open knowledge (arXiv)
  - Open source



Societal Change is Coming

- GitHub Copilot, ChatGPT, etc. are now useful enough to **speed up higher-level human work!**





https://gamefromscratch.com/dall-e-vs-stable-diffusion-vs-midjourney/

Problems with Deep Learning

- Problem#1: Lack of a formal understanding
  - Non-Convex! Non-Convex! Non-Convex!
    - Depth>=3: most losses non-convex in parameters
  - Worse still, existing intuitions from classical statistical learning theory don't seem to carry over.
  - Theoretically, we are stumbling in the dark here

- Standard response #1
  - "Yes, but this just means there's new theory to be constructed"
  - "All interesting learning problems are non-convex"
    - For example, human learning
      - Order matters → wave hands → non-convexity

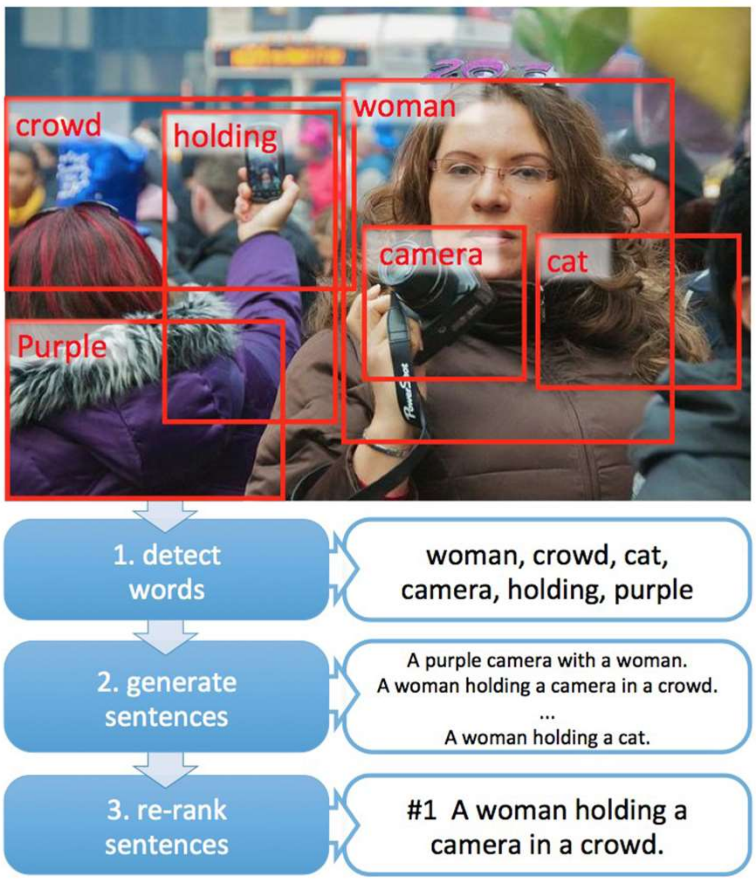
- Standard response #2
  - "Yes, but it often works!"

Problems with Deep Learning

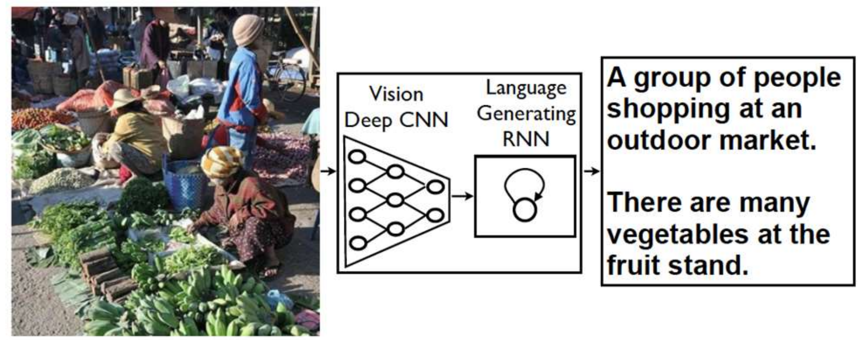
- Problem#2: Lack of interpretability
  - Hard to track down what's failing
  - Pipeline systems have expected performances at each step
  - In end-to-end systems, it's hard to know why things are not working

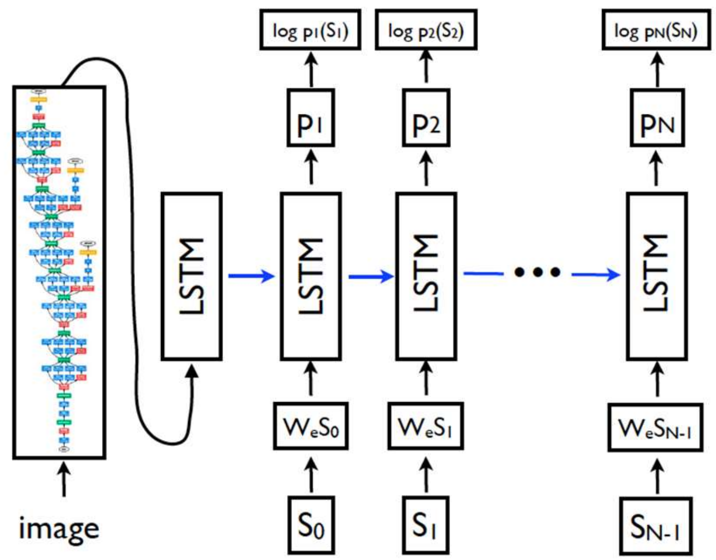
Problems with Deep Learning

- Problem#2: Lack of interpretability



[Fang et al. CVPR15]





[Vinyals et al. CVPR15]

Pipeline

End-to-End

Problems with Deep Learning

- Problem#2: Lack of interpretability
  - Hard to track down what's failing
  - Pipeline systems have "oracle" performances at each step
  - In end-to-end systems, it's hard to know why things are not working
- Standard response #1
  - Tricks of the trade: visualize features, add losses at different layers, pre-train to avoid degenerate initializations…
  - "We're working on it"
- Standard response #2
  - "Yes, but it often works!"

Problems with Deep Learning

- Problem#3: Lack of easy reproducibility
  - Direct consequence of stochasticity & non-convexity
    - different initializations → different local minima
    - Almost everything matters! (hyper-parameters, small design decisions, etc.)
  - More recently: Privatization of unknown models trained on unknown data

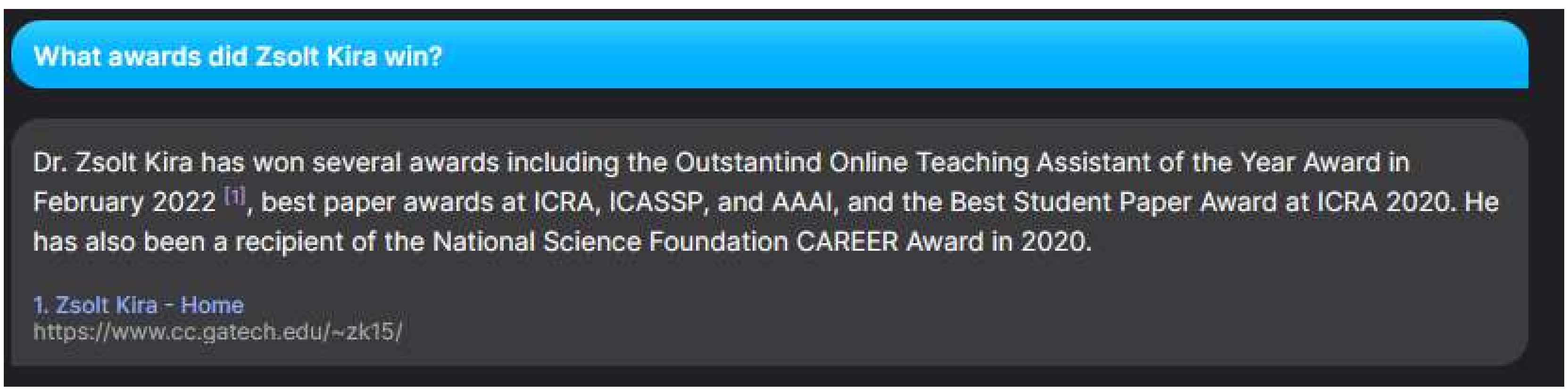
- Standard response #1
  - It's getting much better
  - Standard toolkits/libraries/frameworks now available
  - PyTorch, TensorFlow, MxNet…

- Standard response #2
  - "Yes, but it often works!"

Consequences

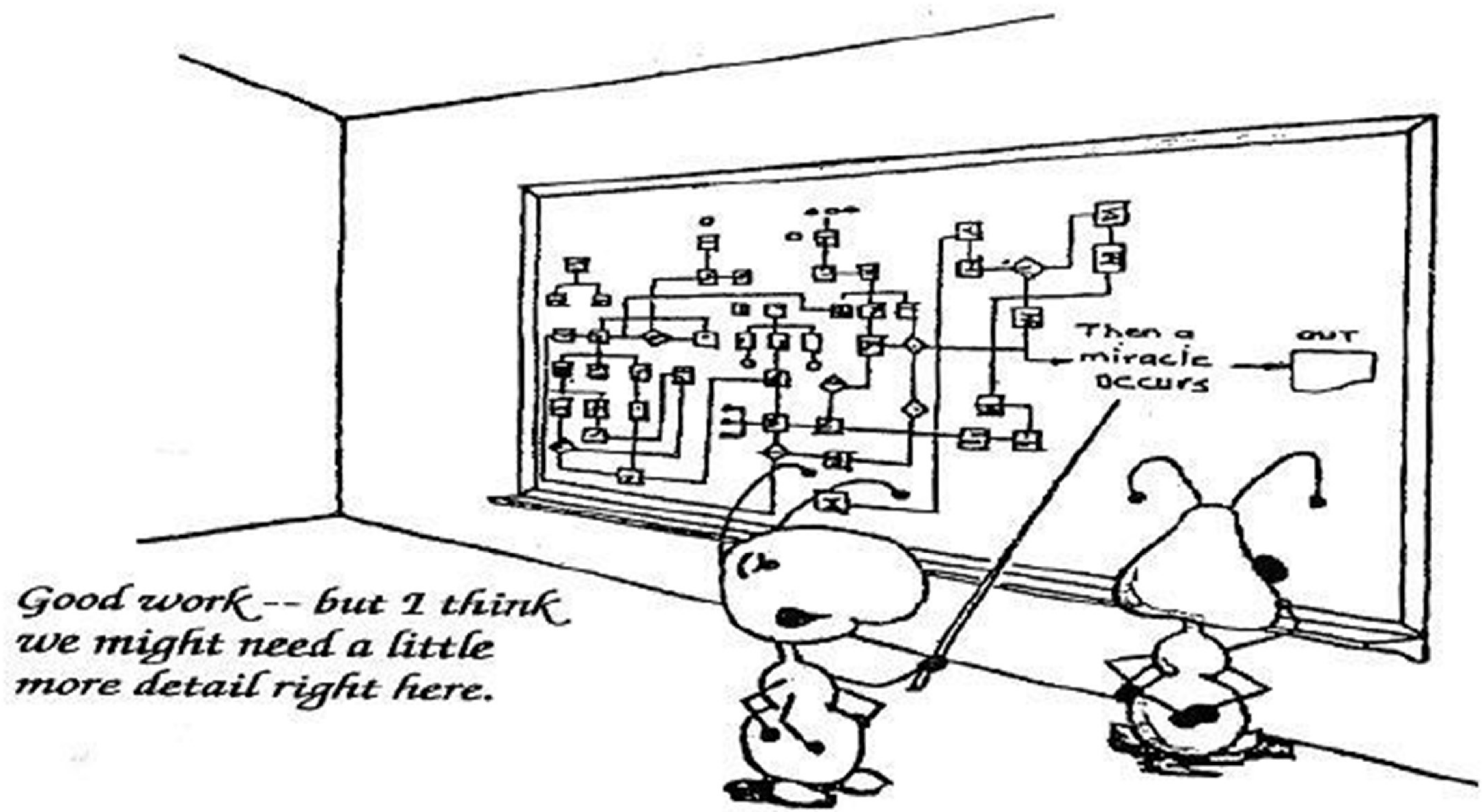
- As a consequence, general issue of **safety and correctness**
  - No explicit reasoning or logical mechanisms

- **Example:**
  - Tesla crashes
  - Language models hallucinating



https://you.com/search?q=who+are+you&tbm=youchat

Yes it works, but how?



Outline

- What is Deep Learning, the field, about?
  - Highlight of some recent projects from my lab
- What is this class about?
  - What to expect?
  - Logistics
- FAQ

Outline

- What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

- What is this class about?
  – What to expect?
  – Logistics

- FAQ

What is this class about?

- Introduction to Deep Learning

- Goal:
  - After finishing this class, you should be ready to get started on your first DL research project.
    - CNNs (Image data)
    - RNNs / Transformers (sequential/graph data)
    - Deep Reinforcement Learning (decision-making)
    - Generative Models (VAEs, Diffusion Models, GANs) (unsupervised learning)

- Target Audience:
  - Senior undergrads, MS-ML, and new PhD students

What this class is NOT

- NOT the target audience:
  - Students without sufficient background knowledge (Python, linear algebra, calculus, basic probability & statistics)
  - Advanced grad-students already working in ML/DL areas
  - People looking for an in-depth understanding of a research
  - area that uses deep learning (3D Vision, Large Language
  - Models, Deep RL, etc.).
- NOT the goal:
  - Intro to Machine Learning
  - Teaching a toolkit. “Intro to TensorFlow/PyTorch”

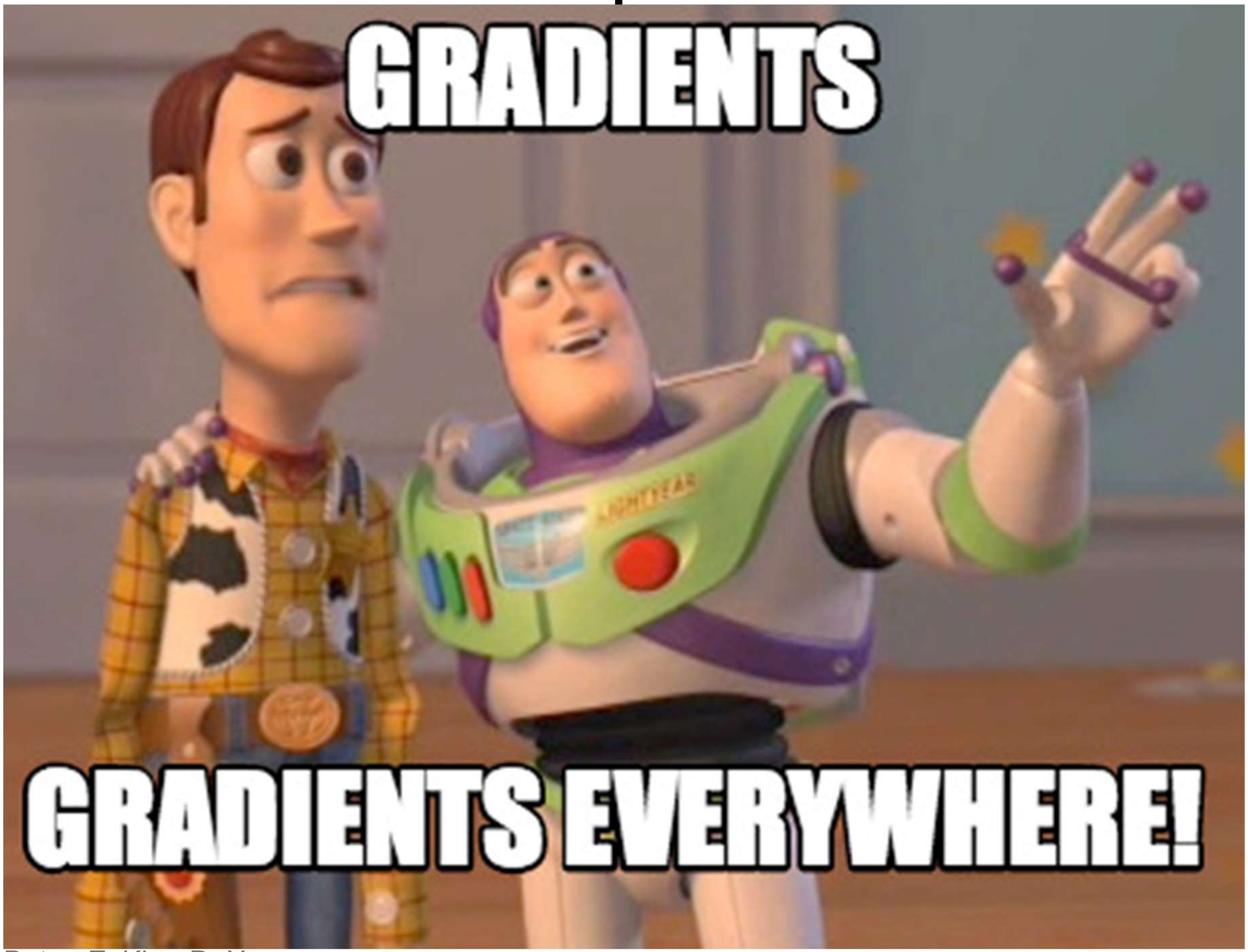
Caveat

- This is an ADVANCED Machine Learning class
  - This should NOT be your first introduction to ML
  - You will need a formal class; not just self-reading/coursera
  - If you took CS 7641/ISYE 6740/CSE 6740 @GT, you're in the right place
  - If you took an equivalent class elsewhere, see list of topics taught in CS 7641 to be sure.

Prerequisites

- Intro Machine Learning
  - Classifiers, regressors, loss functions, MLE, MAP
- Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness…
- Calculus
  - Multi-variate gradients, hessians, jacobians…
- Must read (on W3 reading list): Matrix calculus for deep learning
  - https://explained.ai/matrix-calculus/index.html

Prerequisites



Prerequisites

- Programming!
  - Homeworks will require Python!
  - Libraries/Frameworks: PyTorch
  - HW1 (pure python), HW2 (python + PyTorch), HW3+4 (PyTorch)
  - Your language of choice for project

Course Information

- Instructor: Zsolt Kira
  - [censored]@gatech.edu **(use piazza public/private instead!)**

Zsolt Kira

Assistant Professor

Associate Director, ML@GT

TAs

Ting-Yu Lan

Anshul Ahluwalia

Ahmed Shaikh

Aaditya Sigh

Yash Jain

Yash Jakhotiya

Hoon Lee

Zach Minot

Sumedh Vijay

Ningyuan Yang

Jan Vijay Singh

Aaditya Singh

Organization & Deliverables

- 4 problem-sets+homeworks (72%)
  - Mix of theory (PS) and implementation (HW)
  - First one goes out next week
    - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early

- Final project (28%)
  - Projects done, recommended in groups of 3-4
  - You need a good reason to do a solo project
  - Mid-semester project proposal before project period starts
  - **Find a team ASAP! Talk to people, use Piazza "find a teammate" post**

- (Bonus) Class Participation (1%)
  - Top (endorsed) contributors on Piazza

Plenty of "buffer" built in

- Grace period
  - 2 days grace period
    - Intended for *checking* submission NOT to replace due date
    - No need to ask for grace, no penalty for turning it in within grace period
    - Can NOT use for PS0/HW0
  - After grace period, you get a 0 (no excuses except medical)
    - Send all medical requests to dean of students (https://studentlife.gatech.edu/)
    - Form: https://gatech-advocate.symplicity.com/care_report/index.php/pid224342?
  - **DO NOT SEND US ANY MEDICAL INFORMATION!** We do not need any details, just a confirmation from dean of students

GT Resources for Mental Health

Georgia Tech Police Department
Emergency: Call 911 | 404-894-2500

Dean of Students Office
404-894-2565 | studentlife.gatech.edu
Afterhours Assistance Line & Dean on Call: 404-894-2204

Center for Assessment, Referral and Education (CARE)
404-894-3498 | care.gatech.edu

Collegiate Recovery Program
404-894-2575 | counseling.gatech.edu

Counseling Center
404-894-2575 | counseling.gatech.edu

Health Initiatives
404-894-9980
healthinitiatives.gatech.edu

LGBTQIA Resource Center
404-385-4780 | lgtbqia.gatech.edu

Stamps Psychiatry Center
404-894-1420

VOICE
404-385-4464 |
404-385-4451
24/7 Info Line: 404-894-9000 |
voice.gatech.edu

Women's Resource Center
404-385-0230 |
womenscenter.gatech.edu

Veterans Resource Center
404-894-4953 |
veterans.gatech.edu

Georgia Crisis and Access Line
1-800-715-4225
The crisis line is staffed with professional social workers and counselors 24 hours per day, every day, to assist those with urgent and emergency needs.

Trevor Project
1-866-488-7386
Trained counselors are available to support anyone in need.

National Suicide Prevention Hotline
1-800-273-8255
A national network of local crisis centers that provides free and confidential emotional support to people in suicidal crisis or emotional distress 24/7.

Georgia State Psychology Clinic
404-413-2500
The clinic offers high quality and affordable psychological services to adults, children, adolescents, families and couples from the greater Atlanta area.

PS0/HW0

- Out already; due Sunday Jan 15th
  - Will be available on class webpage + Canvas
  - If not registered yet (on waitlist), see webpage FAQ for form to request gradescope access

- Grading
  - Not counted towards your final grade, but required
  - <=75% means that you might not be prepared for the class
  - We may not be able to grade before registration ends if submit later than Thursday

- Topics
  - PS: probability, calculus, convexity, proving things
  - HW: Python + Numpy

Computing

- Major bottleneck
  - GPUs

- Options
  - Your own / group / advisor's resources

  - Google Colab
    - jupyter-notebook + free GPU instance

  - Google Cloud credits (details TBA)
    - Tutorial on setting up gloud: https://github.com/cs231n/gcloud

4644 vs 7643

- Level differentiation
- HWs
  - Extra credit questions for 4644 students, necessary for 7643
- Project
  - Higher expectations from 7643

Outline

- What is Deep Learning, the field, about?
  - Highlight of some recent projects from my lab
- What is this class about?
  - What to expect?
  - Logistics
- FAQ

Waitlist / Audit / Sit in

- Waitlist
  - Waitlist are mostly full. Class size will likely increase closer to room size
  - Do PS0/HW0 **NOW**. Come to first few classes.
  - Hope people drop.

- "I need this class to graduate"
  - Talk to your degree program advisor. They control the process of making sure you have options to graduate on time.

- Audit or Pass/Fail
  - No. We will give preference to people taking class for credit.

- Sitting in
  - Welcome to if space allows; otherwise free to join remote

What is the re-grading policy?

- Homework assignments
  - **Within 1 week** of receiving grades: see the TAs

- This is an advanced grad class.
  - The goal is understanding the material and making progress towards our research.

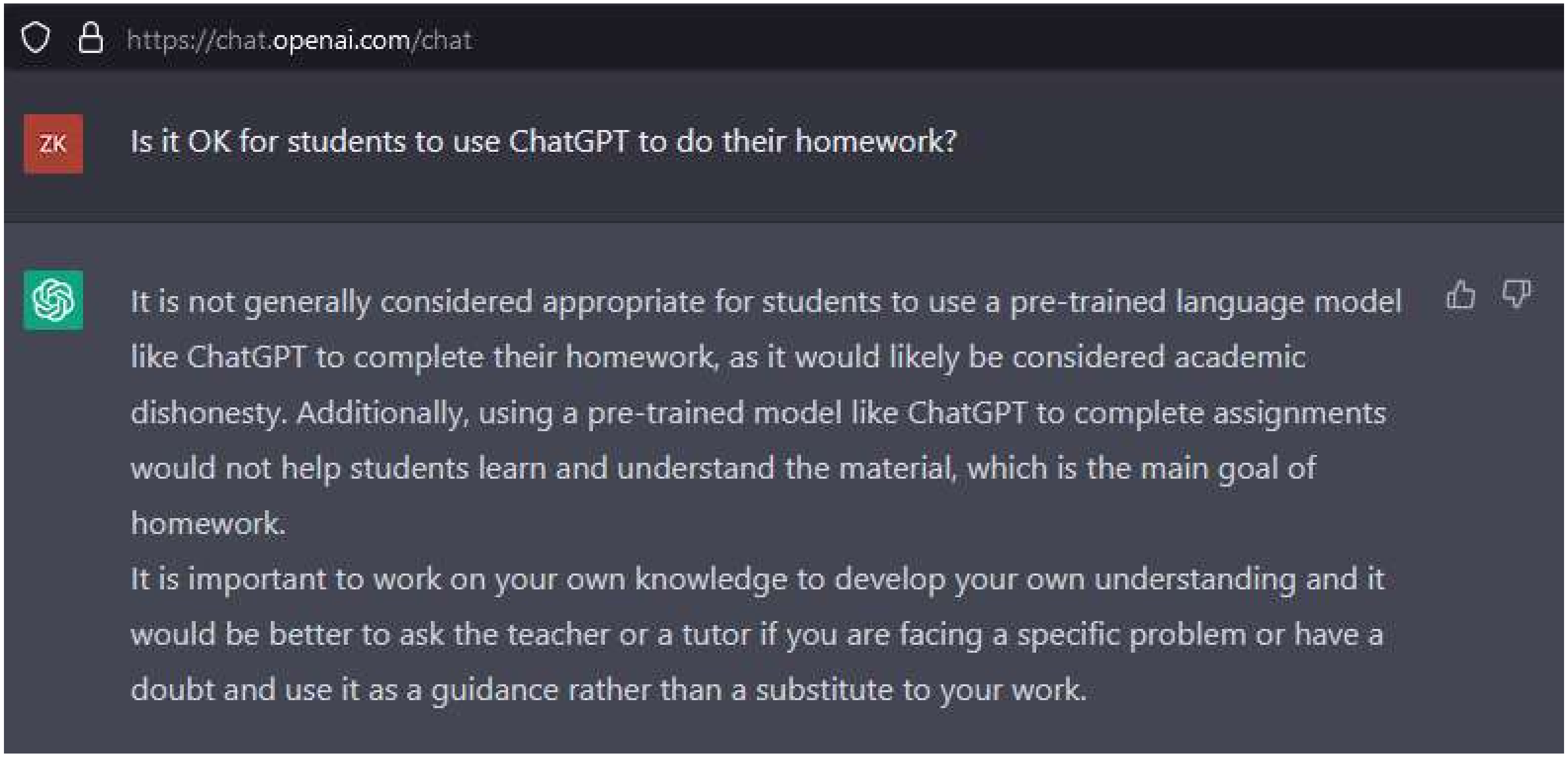
What is the collaboration policy?

- Collaboration
  - Only on HWs and project (not allowed in PS0/HW0).
  - You may discuss the questions
  - **Each student writes their own answers**
  - Write on your homework anyone with whom you collaborate
  - Each student must write their own code for the programming part

- Zero tolerance on plagiarism
  - Neither ethical nor in your best interest
  - Always credit your sources
  - Don’t cheat. We will find out.

Deep Learning is So Good..

- That I had to put this slide in
- Our policy on ChatGPT/Co-Pilot/etc. for now:
  - Don’t use it! We will consider it against the honor code.



https://chat.openai.com/chat

How do I get in touch?

- Primary means of communication -- **Piazza**
  - No direct emails to Instructor unless private information
  - Instructor/TAs can provide answers to everyone on forum
  - Class participation credit for answering questions!
  - No posting answers. We will monitor.
  - Stay respectful and professional

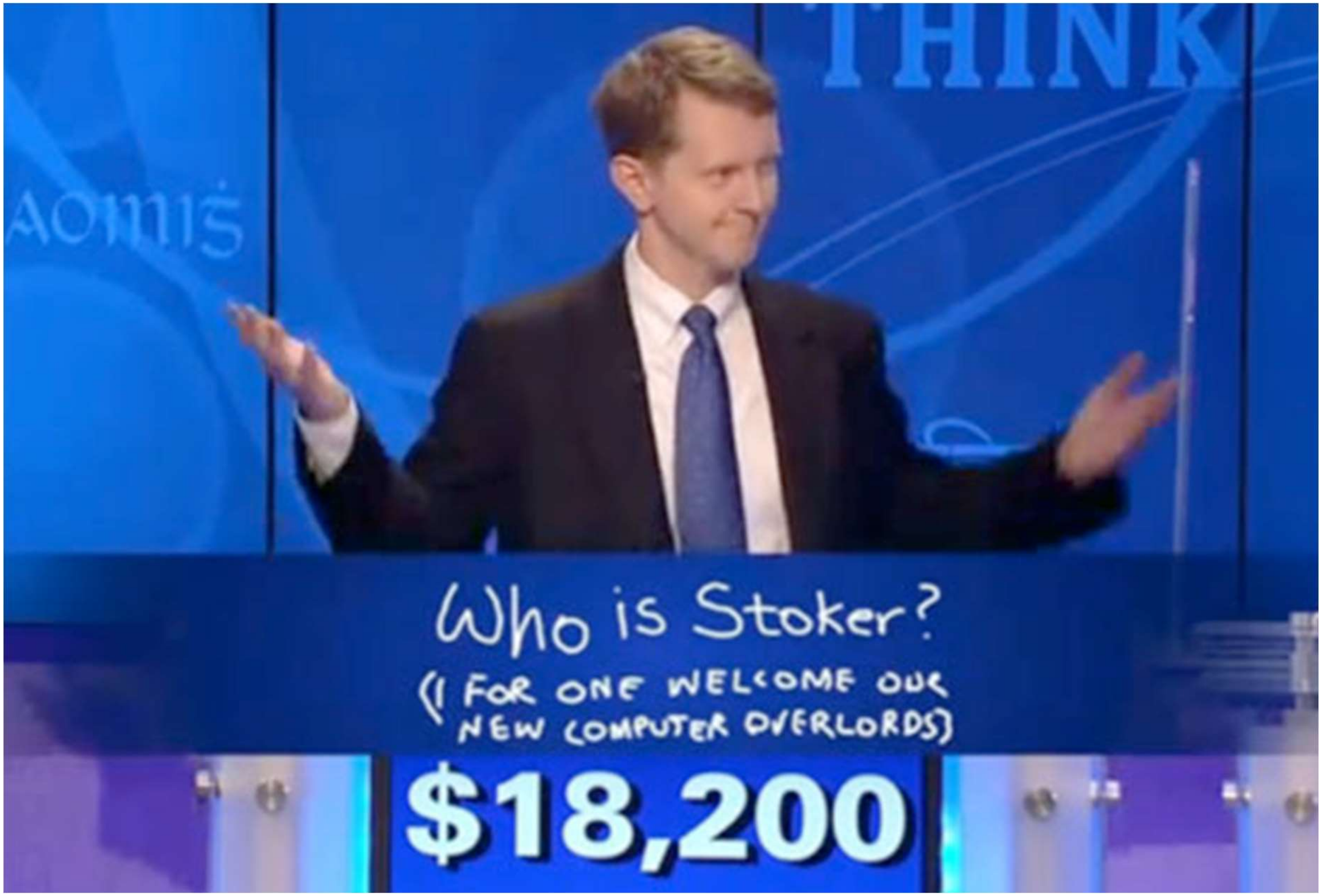
Research

- "Can I work with your group for funding/credits/neither?"
  - Fill out this form, but too late for Spring 2023

Todo

- PS0/HW0
  - Due: Jan 15th 11:59pm

Welcome


THINK
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$18,200