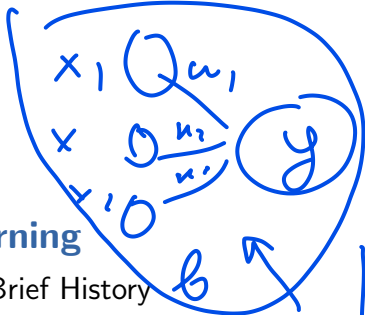


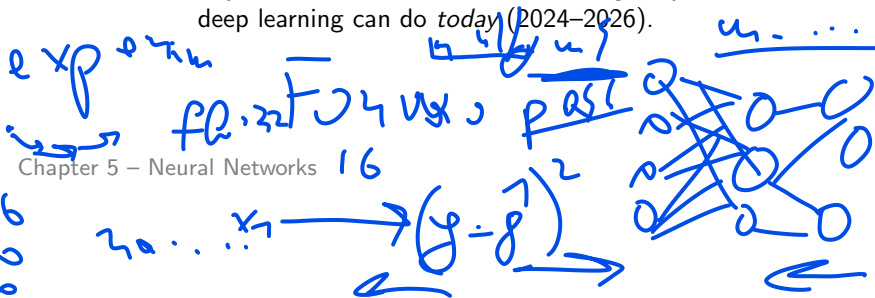
Adam W

weight P_{torch} Deep Learning

Introduction & a Brief History



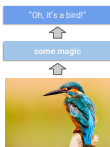
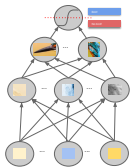
This deck embeds the LMU *Introduction to Deep Learning* intro + history chunks in full, then adds a short gallery of what deep learning can do today (2024–2026).



Chapter 5 – Neural Networks 16

Deep Learning

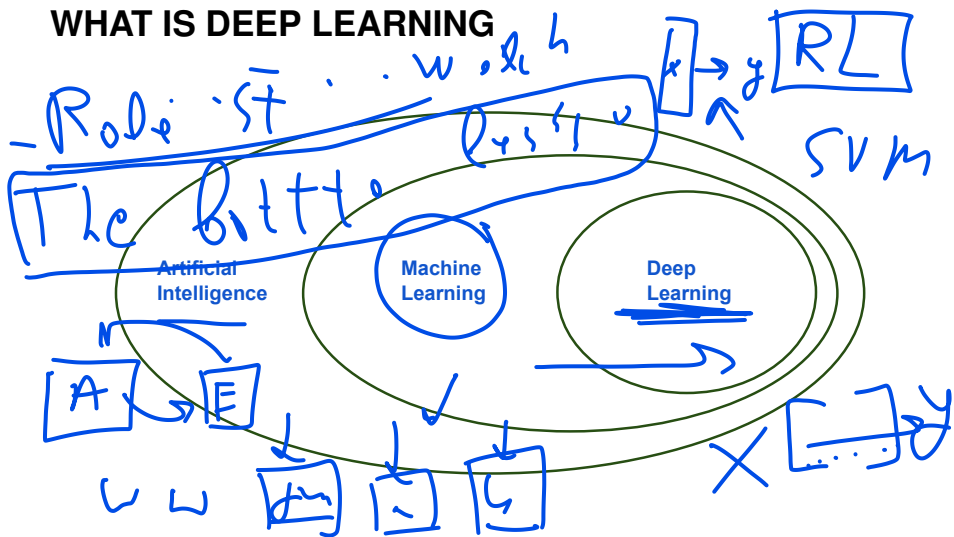
Introduction



Learning goals

- Relationship of DL and ML
- Concept of representation or feature learning
- Use-cases and data types for DL methods

WHAT IS DEEP LEARNING



- Deep learning is a subfield of ML based on artificial neural networks.

$[:]$

DEEP LEARNING AND NEURAL NETWORKS



- Deep learning itself is not *new*:

- Neural networks have been around since the 70s.
- *Deep* neural networks, i.e., networks with multiple hidden layers, are not much younger.

Ho H 3 3 t.u.

- Why everybody is talking about deep learning now:



- 1 Specialized, powerful hardware allows training of huge neural networks to push the state-of-the-art on difficult problems.
- 2 Large amount of data is available.
- 3 Special network architectures for image/text data.
- 4 Better optimization and regularization strategies.

6.3.17
07.0

RaNet GPU

AlexNet
image

IMAGE CLASSIFICATION WITH NEURAL NETWORKS

"Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction."

Y. Bengio

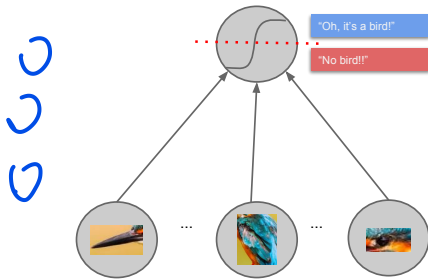


IMAGE CLASSIFICATION WITH NEURAL NETWORKS

"Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction."

Y. Bengio

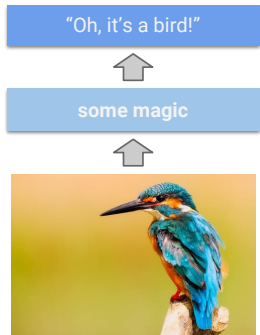
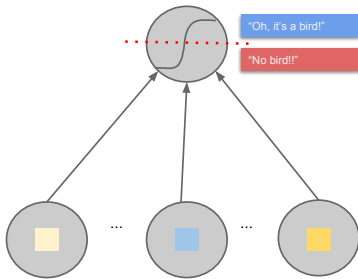


IMAGE CLASSIFICATION WITH NEURAL NETWORKS

"Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction."

Y. Bengio

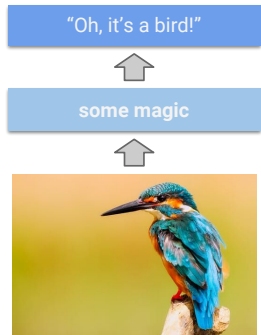
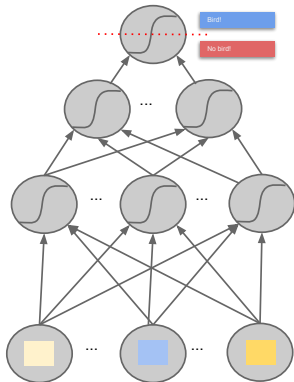
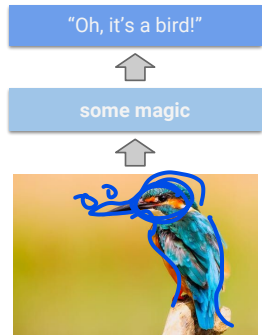
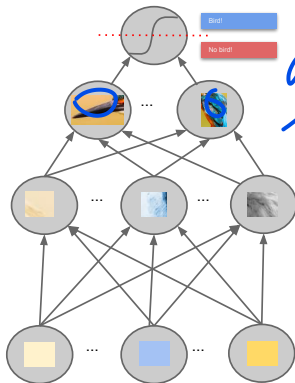


IMAGE CLASSIFICATION WITH NEURAL NETWORKS

"Machine learning algorithms, inspired by the brain, based on learning multiple levels of representation/abstraction."

Y. Bengio



POSSIBLE USE-CASES

Deep learning can be extremely valuable if the data has these properties:

- It is high dimensional. ✓
- Each single feature itself is not very informative but only a combination of them might be.
- There is a large amount of training data. ✓

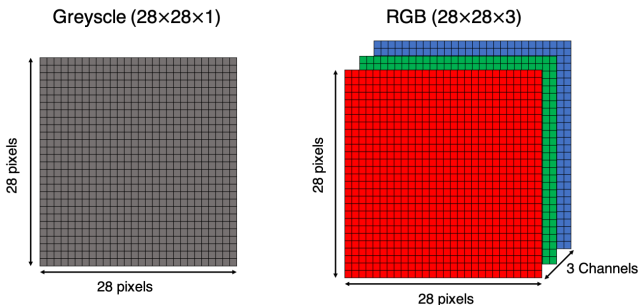
This implies that for tabular data, deep learning is rarely the correct model choice.

- Without extensive tuning, models like random forests or gradient boosting will outperform deep learning most of the time.
- One exception is data with categorical features with many levels.

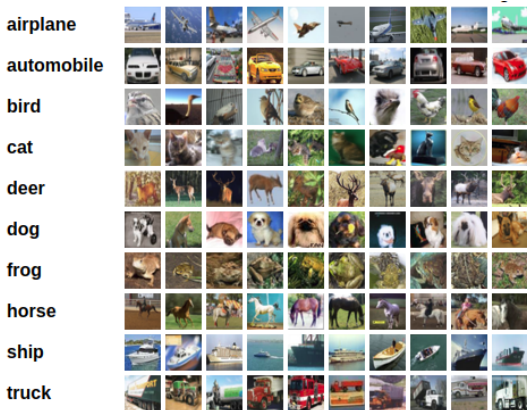
POSSIBLE USE-CASE: IMAGES

- **High Dimensional:** A color image with 255×255 (3 Colors) pixels already has 195075 features.
- **Informative:** A single pixel is not meaningful in itself.
- **Training Data:** Depending on applications huge amounts of data are available.

Architecture: Convolutional Neural Networks (CNN)



POSSIBLE USE-CASE: IMAGES



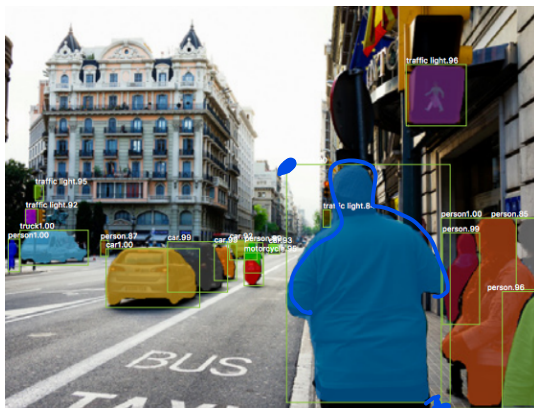
I_k

Figure: Example for image classification (Krizhevsky, 2009)

Image classification tries to predict a single label for each image.

CIFAR-10 is a well-known dataset used for image classification. It consists of 60,000 32x32 color images containing one of 10 object classes, with 6000 images per class.

POSSIBLE USE-CASE: IMAGES



$$\begin{bmatrix} 4 \\ 110 \\ 330 \\ . \\ . \end{bmatrix}$$

Figure: Example for object detection (He et al., 2017)

Object Detection Mask R-CNN is a general framework for instance segmentation, that efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance.

POSSIBLE USE-CASE: IMAGES

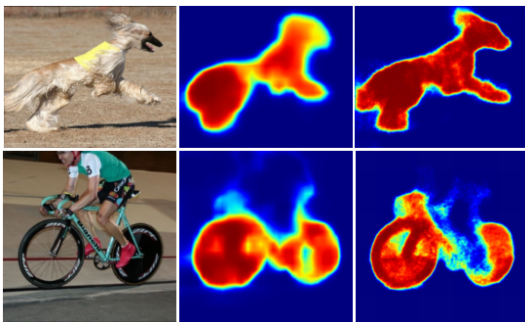


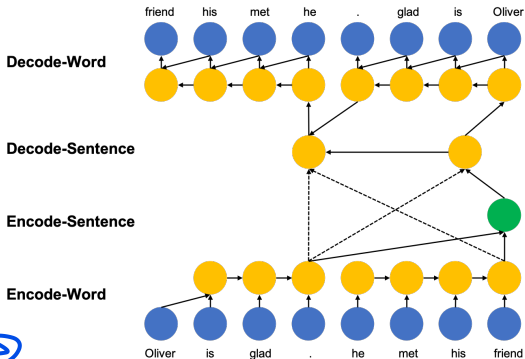
Figure: Example for image segmentation (Noh et al., 2015)

Image segmentation partitions the image into (multiple) segments.

POSSIBLE USE-CASE: TEXT

- **High Dimensional:** Each word can be a single feature (300000 words in the German language).
- **Informative:** A single word does not provide much context.
- **Training Data:** Huge amounts of text data available.

Architecture: Recurrent Neural Networks (RNN)



POSSIBLE USE-CASE: TEXT CLASSIFICATION



Positive

Great job! Your customer support is fantastic.



Neutral

Not bad, but it should be improved in the future.

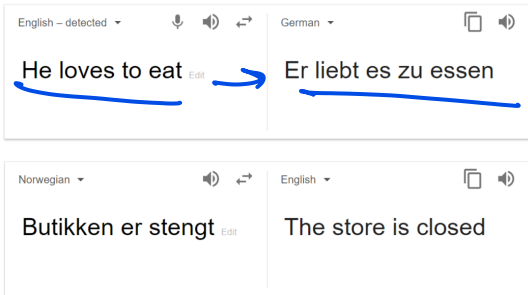


Negative

The worst customer service I have ever seen.

Sentiment Analysis is the application of natural language processing to systematically identify the emotional and subjective information in texts.

POSSIBLE USE-CASE: TEXT



at 1,1t

600

Machine Translation (e.g. google translate) Neural machine translation exploits neural networks to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

APPLICATIONS OF DEEP LEARNING: SPEECH

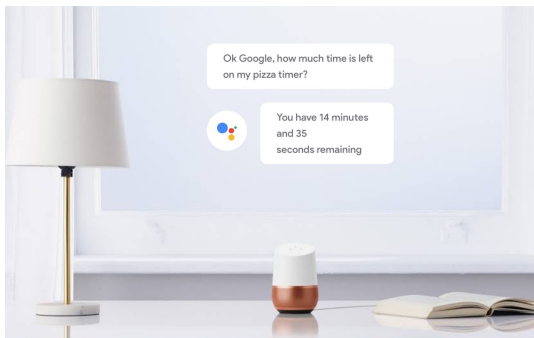


Figure: Example for speech recognition and generation (Google)

Speech Recognition and Generation (e.g. google assistant) Neural network extracts features from audio data for downstream tasks, e.g., to classify emotions in speech.

Deep Learning

Brief History

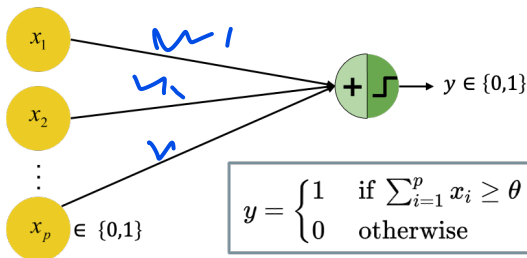


Learning goals

- Predecessors of modern (deep) neural networks
- History of DL as a field

A BRIEF HISTORY OF NEURAL NETWORKS

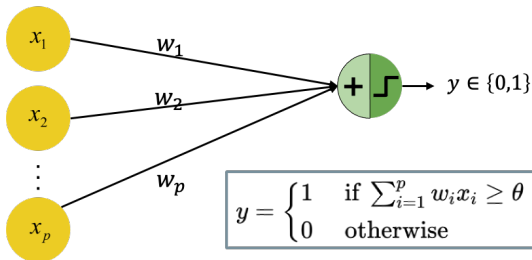
- 1943: The first artificial neuron, the "Threshold Logic Unit (TLU)", was proposed by Warren McCulloch & Walter Pitts.



- The model is limited to binary inputs.
- It fires/outputs +1 if the input exceeds a certain threshold θ .
- The weights are not adjustable, so learning could only be achieved by changing the threshold θ .

A BRIEF HISTORY OF NEURAL NETWORKS / 2

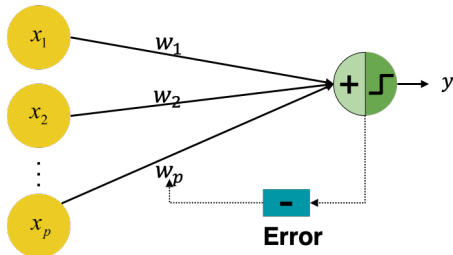
- **1957:** The perceptron was invented by Frank Rosenblatt.



- The inputs are not restricted to be binary.
- The weights are adjustable and can be learned by learning algorithms.
- As for the TLU, the threshold is adjustable and decision boundaries are linear.

A BRIEF HISTORY OF NEURAL NETWORKS / 3

- **1960:** Adaptive Linear Neuron (ADALINE) was invented by Bernard Widrow & Ted Hoff; weights are now adjustable according to the weighted sum of the inputs.

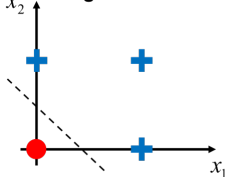


- **1965:** Group method of data handling (also known as polynomial neural networks) by Alexey Ivakhnenko. The first learning algorithm for supervised deep feedforward multilayer perceptrons.

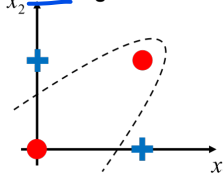
A BRIEF HISTORY OF NEURAL NETWORKS / 4

- 1969: The first “AI Winter” kicked in.
 - Marvin Minsky & Seymour Papert proved that a perceptron cannot solve the XOR-Problem (linear separability).
 - Less funding $\Rightarrow$ Standstill in AI/DL research.

OR Logical Function



XOR Logical Function



- 1985: Multilayer perceptron with backpropagation by David Rumelhart, Geoffrey Hinton, and Ronald Williams.
 - Efficiently compute derivatives of composite functions.
 - Backpropagation was developed already in 1970 by Linnainmaa.

A BRIEF HISTORY OF NEURAL NETWORKS / 5

- **1985:** The second “AI Winter” kicked in.
 - Overly optimistic expectations concerning potential of AI/DL.
 - The phrase “AI” even reached a pseudoscience status.
 - Kernel machines and graphical models both achieved good results on many important tasks.
 - Some fundamental mathematical difficulties in modeling long sequences were identified.

SVM

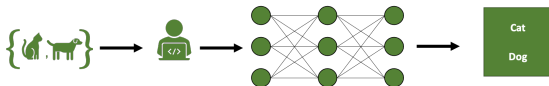


Figure: Symbolic image of an “AI Winter” (Fagella, 2019)

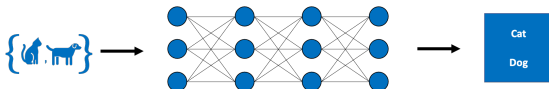
A BRIEF HISTORY OF NEURAL NETWORKS / 6

- **2006:** Age of deep neural networks began.
 - Geoffrey Hinton showed that a deep belief network could be efficiently trained using *greedy layer-wise pretraining*.
 - This wave of research popularized the use of the term deep learning to emphasize that researchers were now able to train deeper neural networks than had been possible before.
 - At this time, deep neural networks outperformed competing AI systems based on other ML technologies as well as hand-designed functionality.

Machine Learning



Deep Learning



A BRIEF HISTORY OF NEURAL NETWORKS / 7

RVW

Deep Learning Timeline

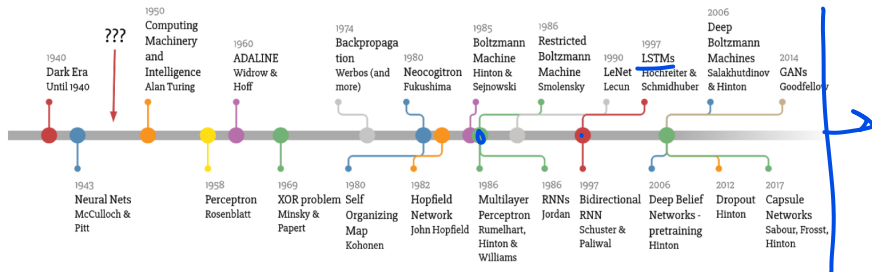
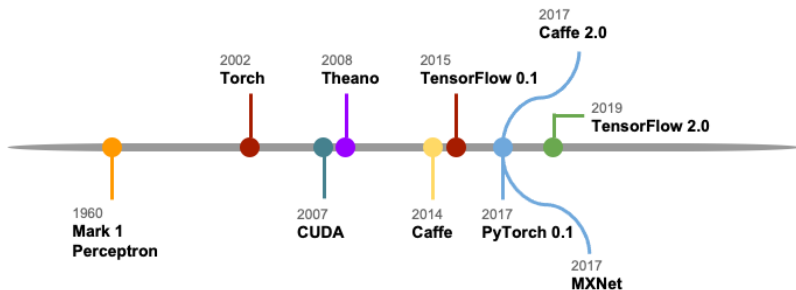


Figure: (Vazquez, 2018)

History of DL Tools



A BRIEF HISTORY OF NEURAL NETWORKS / 9



Figure: IBM Supercomputer Watson (IBM, 2011)

- Watson is a question-answering system capable of answering questions posed in natural language, developed in IBM's DeepQA project.
- In 2011, Watson competed on Jeopardy! against champions Brad Rutter and Ken Jennings, winning the first place prize of \$1 million.

A BRIEF HISTORY OF NEURAL NETWORKS / 10



Figure: Google's self driving car (McCabe, 2016)

- Google's development of self-driving technology began on January 17, 2009, at the company's secretive X lab.
- By January 2020, 20 million miles of self-driving on public roads had been completed by Waymo.

A BRIEF HISTORY OF NEURAL NETWORKS / 11



Figure: Protein structure by AlphaFold (DeepMind, 2022)

- **AlphaFold** is a deep learning system, developed by Google DeepMind, for determining a protein's 3D shape from its amino-acid sequence.
- In 2018 and 2020, AlphaFold placed first in the overall rankings of the Critical Assessment of Techniques for Protein Structure Prediction (CASP).

A BRIEF HISTORY OF NEURAL NETWORKS / 12

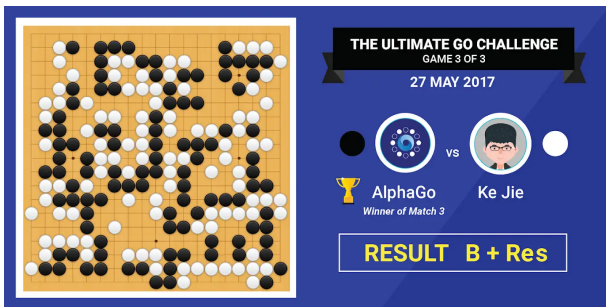


Figure: AlphaGo (DeepMind, 2022)



DOTA-2

- **AlphaGo**, originally developed by DeepMind, is a deep learning system that plays the board game Go. In 2017, the Master version of AlphaGo beat Ke Jie, the number one ranked player in the world at the time.
- While there are several extensions to AlphaGo (e.g., Master AlphaGo, AlphaGo Zero, AlphaZero, and MuZero), the main idea is the same: search for optimal moves based on knowledge acquired by machine learning.

A BRIEF HISTORY OF NEURAL NETWORKS / 13



Figure: Logo OpenAI (OpenAI, 2023)

- **Generative Pre-trained Transformer 3 (GPT-3)** is the third generation of the GPT model, introduced by OpenAI in May 2020, to produce human-like text.
- There are 175 billion parameters to be learned by the algorithm, but the quality of the generated text is so high that it is hardly possible to distinguish it from a human-written text.

WHAT THE FIELD LOOKS LIKE TODAY

- A major shift in the 2020s was the move from highly specialized models to **foundation models**: large pretrained models that can be adapted to many downstream tasks.
- The frontier is increasingly **multimodal**: models can jointly process text, images, audio, video, and actions.
- Modern systems are also becoming more **interactive**: they can write code, search, retrieve documents, and call external tools.
- Examples of this broader landscape include **GPT-5** and **Claude Opus** for general-purpose assistants, **Stable Diffusion** for image generation, **DINOv3** for self-supervised visual representations, and **TabPFN** for tabular prediction.

WHAT THE FIELD LOOKS LIKE TODAY / 2

- Progress is now often driven by a combination of
 - better architectures and training objectives,
 - much larger datasets,
 - more compute and specialized hardware,
 - stronger software tooling and engineering.

The modern DL pipeline is frequently: **pretrain at scale, then adapt with prompting, finetuning, retrieval, or tool use.**

Deep learning today

A few results from 2024–2026 that would have looked like science fiction back when these methods were first invented. The rest of this chapter builds the machinery behind them, starting from a single neuron.

The state of the field (2026)

Op. . .

- ▶ **Foundation models rule.** A handful of huge pretrained models, *adapted* (prompting, fine-tuning, tools) to thousands of tasks – not one model per task.
- ▶ **Multimodal by default.** One model handles text, images, audio, and video together.
- ▶ **Reasoning models.** Models now “think” longer at inference (*test-time compute*) – big gains on math and code.
- ▶ **Open weights, near-frontier. DeepSeek-R1** (Jan 2025) matched top closed models on math/code – built on a base model reportedly trained for ~\$6M – with weights released.
- ▶ **Agents.** Models don’t just answer – they browse, run code, and call tools.

cfw! . . . -

Two 2024 Nobel Prizes went to neural-network work – **Physics** (Hopfield & Hinton) and **Chemistry** (AlphaFold). The field the intro called “not new” is now mainstream science.

The state of AI today: the industry

- ▶ **NVIDIA sells the shovels.** Almost every model trains on its GPUs; it holds ~80% of the AI-chip market and in **July 2025 became the first company ever worth \$4 trillion** (past Apple & Microsoft). Its Blackwell chips are sold out.
- ▶ **Open weights caught up.** Open models (**DeepSeek, Qwen, Kimi, Llama, Mistral**) now trail the best closed models by only *months*, at a fraction of the price – with much of the open-weight frontier now coming from Chinese labs.
- ▶ **The next frontier is “physical AI”.** NVIDIA’s Jensen Huang: “every industrial company will become a robotics company” – pitching humanoid robots as a ~\$40 trillion market and selling the stack (world models) to train them.
- ▶ **Frontier models keep leaping.** Anthropic’s **Claude Fable 5** (2026) can work *for days* in an agent harness – planning, delegating to sub-agents, and checking its own work.

The gold-rush pattern: so far the surest profits go to whoever sells the “picks and shovels” (the compute), more than to the model builders themselves.

Case study: an AI rewrote Bun in 11 days

Bun is a fast JavaScript runtime, written in **Zig**. In May 2026 its creator used a *pre-release* **Claude Fable 5** to rewrite the whole thing in **Rust**:

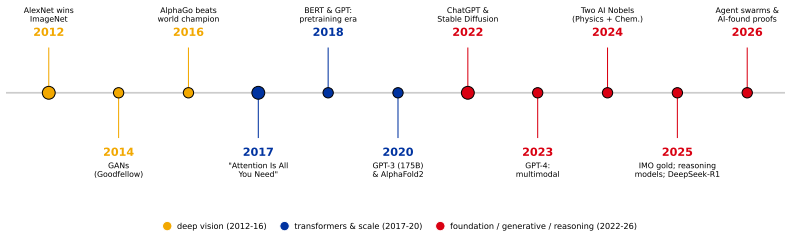
- ▶ **~64 Claude Code agents** at a time, across **~50 dynamic workflows** – peaking at **~1,300 lines of code per minute**, over **1 million lines** of Rust in **11 days**.
- ▶ Cost **~\$165,000** in API usage.
- ▶ The result passed **100% of Bun's 1M+ test assertions** on every supported platform.
- ▶ *Why Rust*: mixing Zig's manual memory with JavaScript's garbage collector kept producing use-after-free and memory-leak bugs.
- ▶ **It shipped. Bun 1.4** (Aug 20, 2026) is the first Rust release; Claude Code itself had already run on it for months. Idle CPU use dropped 5×.

Reality check. Not everyone is convinced – Zig's creator **Andrew Kelley** called the output “unreviewed slop” (July 2026). But a task that would take a team *months* ran, and passed its tests, in under two weeks.

A timeline of the recent explosion

The history chunk stopped around 2012. Here is what happened *since* – barely 14 years:

Deep learning, 2012-2026: the recent explosion



Notice the acceleration. From “beats a human at a board game” (2016) to “gold medal at the Mathematical Olympiad” (2025) took under a decade, and to a claimed solution of a \$1M Millennium Prize Problem (2026) about ten years.

Everyday AI: language assistants

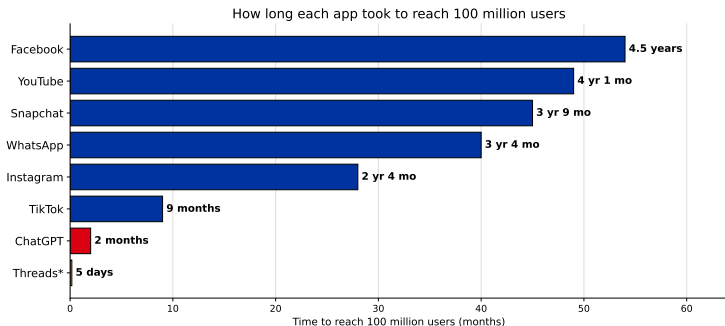
- ▶ Chat assistants – **ChatGPT, Claude, Gemini** – are the fastest-adopted software in history: ChatGPT reached **100 million** users in about **two months** (early 2023).
- ▶ **One** general model writes essays, debugs code, tutors mathematics, translates, and drafts email – no task-specific training.
- ▶ The jump in scale: **GPT-3** (2020, 175B params) → today's **GPT-5 / Claude Opus** generation.



ChatGPT / OpenAI

Under the hood this is still a neural network predicting the next token – the same “stack of affine maps + nonlinearities” idea, scaled up enormously and trained on much of the public internet.

The fastest adoption in history



* Threads launched on Instagram's existing user base. Source: Visual Capitalist / UBS.

ChatGPT reached 100 million users in about two months – the fastest ever for a standalone app. (Threads hit it in 5 days, but launched straight onto Instagram's existing billions of users.)

Generative media: text → image, audio, video

- ▶ Type a sentence, get a photorealistic **image** (Midjourney, Stable Diffusion) – or now a **video**.
- ▶ 2025–26 video models (**Veo 3**, **Sora 2**) produce 8–20s clips with **synchronized audio**, consistent characters, and plausible physics – “the *GPT-3.5 moment* for video”.
- ▶ The same family of models clones voices and generates music.



“Théâtre D’opéra Spatial”:
AI-generated (Midjourney), 2022

The flip side – deepfakes. When anyone can synthesize convincing video, “seeing is believing” breaks. Provenance and detection are now serious open problems.

From chatbots to agents

- ▶ The newest systems don't just *answer* – they **act**: browse the web, run code, call external tools, and complete multi-step tasks on their own.
- ▶ **Coding agents** write and edit whole codebases; “**computer-use**” models click through real apps; “**deep-research**” agents read many sources and write a report.

The model shifts from an *oracle* (you ask, it answers) to a *worker* (it does the task). This is exactly why “tool use” and “agents” are now core research topics – and, in fact, how the slides in front of you were assembled.

Embodied AI: robots that learn

- ▶ The hard part of robotics was always **control** – hand-writing rules for balance and manipulation.
- ▶ **Figure's Helix** (2026): a single **10-million-parameter** neural network replaced roughly **109,000 lines** of hand-tuned C++ balance code.
- ▶ Humanoids (**Tesla Optimus**, **Figure**) now do real factory tasks; NVIDIA calls this wave **“physical AI”**.

Same recipe as vision and language: **replace hand-engineered rules with a network trained on data**. Representation learning, now for motor control.

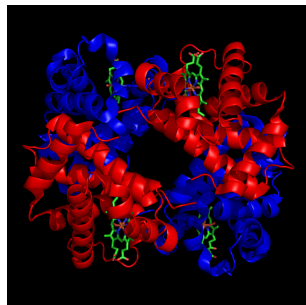


Tesla Optimus (2025)

AI for science I – biology's 50-year problem

- ▶ Predicting a protein's **3D shape** from its amino-acid sequence stumped biology for about **50 years**.
- ▶ **AlphaFold2** (DeepMind, 2020) cracked it, and has since predicted **200M+ structures** – essentially every protein known to science – used by **2M+ researchers** in 190 countries.
- ▶ **2024 Nobel Prize in Chemistry:** Hassabis & Jumper (shared with David Baker for protein *design*).

The **first Nobel** for an *applied* deep-learning system – it directly speeds up drug and vaccine design.



A protein's 3D structure
(ribbon model, hemoglobin)

AI for science II – mathematics

- ▶ **IMO 2025:** Google DeepMind and OpenAI both reached **gold-medal level** (35/42), with full proofs **in natural language**.
- ▶ **Jan 2026: GPT-5.2 Pro** started settling open **Erdős problems**, checked in **Lean**. Terence Tao: real progress, but mostly on the easier ones.
- ▶ **May 2026:** an OpenAI model *disproved* Erdős's 1946 **unit distance conjecture**: among n points in the plane, how many pairs can be exactly 1 apart? It found point sets with far more than Erdős believed possible.



Terence Tao
(Fields Medal 2006)

Nine mathematicians checked the unit distance proof. Fields medalist **Tim Gowers**: “if a human had written the paper and submitted it to the *Annals of Mathematics* [...] I would have recommended acceptance without any hesitation.”

Case study: can a fluid blow up? (Navier-Stokes, 1/2)

- ▶ The **Navier-Stokes** equations describe flowing fluids: air over a wing, blood in a vessel, the weather.
- ▶ One of the seven **Millennium Prize Problems** (**Clay Institute**, 2000, \$1M each): can a smooth 3D flow **blow up**, its speed becoming infinite at a point in finite time?
- ▶ Either answer wins: prove flows always stay smooth, *or* build one that blows up. The official statement lets that blow-up use a smooth external **force**.

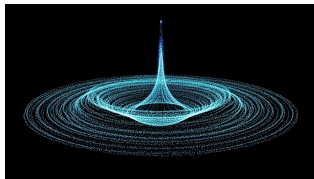


Illustration: DeepMind (2025)

7

Neural networks enter (Sept 2025). DeepMind, with Brown, NYU and Stanford, trained **physics-informed neural networks (PINNs)**: no data, the loss is how badly the network violates the equation. Trained to near machine precision (with a second-order Gauss-Newton optimizer), they found new families of *unstable* singularities in related fluid equations. A network, a loss, an optimizer: this chapter's machinery.

Case study: an AI swarm and Navier-Stokes (2/2)

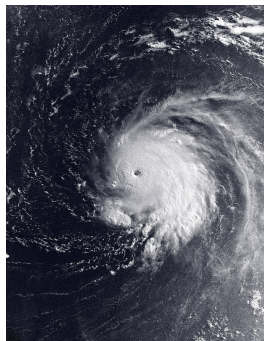
September 2026, four days:

- ▶ **Sept 7-8: Tristan Buckmaster** (NYU) and **Levent Alpöge** (Anthropic) announce AI-assisted, Lean-checked blow-ups with a smooth force for three simpler equations, including **3D Euler** (Navier-Stokes without viscosity). Tao: the method may well extend to Navier-Stokes.
- ▶ **Sept 8, ~12 hours later: OpenAI** claims forced 3D **Navier-Stokes** itself: a 166-page paper plus a **Lean** formalization.
- ▶ **The run: ~10,000 agents** of an unreleased model for **88 hours**: 2.7 million messages, ~130 billion output tokens. 105
- ▶ **Sept 11:** the Clay Institute says the problem “has apparently been settled”; its review will be deliberately unhurried. OpenAI says it will not claim the \$1M.

Reality check. Still a *claimed* proof, not yet peer reviewed. It uses a chosen external force: whether a fluid can blow up *on its own* is still open, and many mathematicians see that as the deeper question. Buckmaster says news of his work reached OpenAI first; OpenAI denies it. And Tao warns that “strip-mining” open problems could destroy the ecosystem that grows new techniques and new mathematicians.

AI for science III – the physical world

- ▶ **Weather: GraphCast** makes 10-day forecasts faster *and* more accurate than the gold-standard physics simulator – it called Hurricane Lee's landfall **3 days** ahead of traditional models.
- ▶ **Materials: GNoME** predicted **2.2 million** new stable crystals – roughly **800 years** of discovery at once.
- ▶ **Energy:** DeepMind used reinforcement learning to **control the plasma** inside a fusion reactor (now partnered with Commonwealth Fusion).
- ▶ **Games (the spark): AlphaGo** beat the world's #1 Go player in 2016–17 – long thought a decade away.



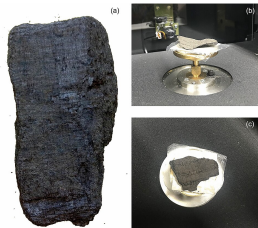
Hurricane Lee (2023) –
GraphCast called its track
early

AI for science IV – it sharpens its own tools

- ▶ AlphaTensor (DeepMind, 2022) found **faster ways to multiply matrices**, e.g. 4×4 matrices with 47 multiplications in arithmetic mod 2, beating **Strassen's 49 from 1969**.
- ▶ AlphaDev (2023) found **faster sorting routines**, now shipping in the **C++ standard library** – run billions of times a day on real devices.
- ▶ AlphaEvolve (DeepMind, May 2025), a Gemini-powered coding agent, did it for ordinary (complex-valued) 4×4 matrices: **48** multiplications. It also recovers **0.7%** of Google's worldwide compute, and a kernel it sped up cut **Gemini's own training time by 1%**.

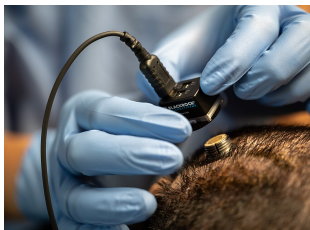
The operation they sped up – **matrix multiplication** – is exactly what every neural network runs on (each layer is one big matrix multiply). AI is now optimizing **the very math it runs on**, and in AlphaEvolve's case, the training of the model behind it.

...and a few that sound like science fiction



Reading burnt Roman scrolls

ML read a charred scroll buried by Vesuvius in 79 AD
– *without* unrolling it
(Vesuvius Challenge, 2024–26).



Giving a voice back

Brain implants + AI decode a paralyzed person's intended speech in real time, in their own voice (~97%).



WAYMO

Driverless taxis at scale

Waymo: ~500k paid robotaxi rides a week (2026), scaling toward 1M.

And still more: AI-designed drugs in human trials, models decoding **animal communication**, and autonomous drones that beat human **racing champions**.

The people behind it

Deep learning is the work of a stubborn few who kept going through two “AI winters”. A handful of the names worth knowing.

The three “godfathers” (2018 Turing Award)



Geoffrey Hinton

Backprop, Boltzmann machines,

AlexNet. **Nobel Physics**
2024.



Yann LeCun

Convolutional networks
(LeNet) –

the CNN that reads digits.



Yoshua Bengio

Deep learning for language,
attention, generative
models.

The three who kept neural nets alive when the field had written them off. The **2018 Turing Award** (the “Nobel of computing”) named them the fathers of the deep-learning revolution.

More key figures



Frank Rosenblatt
the perceptron (1958)



Vladimir Vapnik
SVMs, VC dimension



John Hopfield
Hopfield nets; Nobel Phys.
2024



Jürgen Schmidhuber
LSTM (long-memory RNNs)



Fei-Fei Li
ImageNet – sparked the
2012 leap



Demis Hassabis
DeepMind; AlphaGo/Fold;
Nobel 2024

And the teachers who spread it

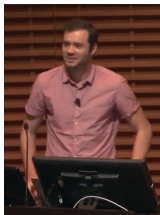
Research is only half the story – these three put deep learning within reach of *millions* through free courses:



st

Andrew Ng ✓

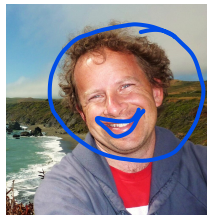
Coursera co-founder; the ML course taken by **millions**; deeplearning.ai.



Andrej Karpathy

Stanford CS231n; “Zero to Hero”; OpenAI/Tesla; Eureka Labs.

Re-..



Jeremy Howard

fast.ai co-founder; “Practical Deep Learning for Coders”.

If you learn deep learning online, you are probably learning from one of these three. **Karpathy's** from-scratch videos and **Ng's** courses are a great next step after this chapter.

01

4

GPT-2

Keep learning: two YouTube channels and a newsletter



Welch Labs

Stephen Welch's visual explainers. *Neural Networks Demystified* (2014) builds a network in Python; stills later in this chapter are his.



Yannic Kilcher

ETH Zurich PhD. Long walkthroughs of new ML papers, figure by figure: the way to learn to read research.



TLDR AI (tldr.tech/ai)

Free weekday email: new models, papers and tools in a few sentences each, about a 5-minute read. 1.1M readers.

Welch Labs for intuition, **Yannic Kilcher** for papers, **TLDR** to keep up. The field moves weekly: the Navier-Stokes story earlier took four days.

Where this chapter goes next

All of the above are **applications**. The rest of the chapter builds the **machinery** that makes them possible:

- ▶ **The single neuron** – a weighted sum plus a nonlinearity (you already know it as logistic regression).
- ▶ **Hidden layers** – stacking neurons so the model learns its own features (representation learning).
- ▶ **Training** – backpropagation and gradient descent, the algorithm that fits the millions of weights.

Everything in the gallery is the same core idea – **a stack of affine maps and nonlinearities, trained on data** – scaled up and specialized. We start from one neuron.

Sources

Modern-applications & people slides (web-verified July 2026, updated Sept 2026):

- ▶ **Nobels/awards:** 2024 Nobel Prize in Physics (Hopfield & Hinton) & Chemistry (AlphaFold: Hassabis & Jumper); 2018 Turing Award – NobelPrize.org, ACM.
- ▶ **Math:** IMO 2025 gold (DeepMind; OpenAI); Erdős problems (Quanta, Aug 2026); unit distance disproof (OpenAI, May 2026; Alon, Gowers et al., arXiv 2605.20695).
- ▶ **Navier-Stokes:** DeepMind et al., arXiv 2509.14185 (Sept 2025); Tao's blog (7 Sept 2026); OpenAI (8 Sept); Clay Institute (11 Sept); Fortune; Sci. American.
- ▶ **Robotics/media:** Figure Helix; video gen (Veo 3, Sora 2); reasoning & open weights (DeepSeek-R1) – vendor/lab announcements, 2025–26.
- ▶ **AI for science:** GraphCast, GNoME, fusion plasma, AlphaGo, AlphaTensor/AlphaDev (Nature 2022–23); AlphaEvolve (May 2025); Vesuvius scrolls; speech neuroprostheses; Waymo – DeepMind/UC Davis, 2016–26.
- ▶ **Industry:** NVIDIA \$4T (Jul 2025); open weights; physical AI (Huang); Claude Fable 5 and the Bun rewrite (Sumner; The Register; Bun 1.4 notes, Aug 2026). Adoption speeds – Visual Capitalist / UBS.
- ▶ **Images:** Wikimedia Commons; YouTube channel and TLDR logos; DeepMind blog. TLDR reader count per tldr.tech.

Embedded slides: LMU *Introduction to Deep Learning* (01_introduction, 02_brief-history), reproduced in full.