Gen AI & Classic ML

when to use what

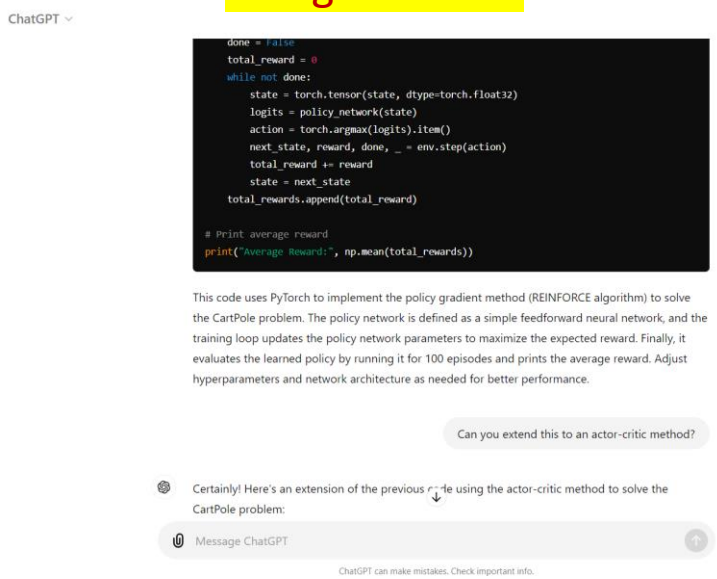
October 2024

Felix Wick

Classic ML vs Deep Learning

Generative vs Predictive Models

text generation



[ChatGPT]

image synthesis

Prompt: Epic anime artwork of a wizard atop a mountain at night casting a cosmic spell into the dark sky that says "Stable Diffusion 3" made out of colorful energy

[Stable Diffusion 3 — Stability AI]

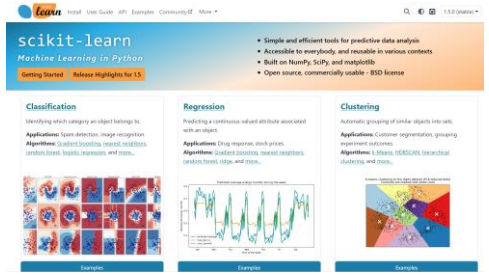
text-to-video

Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She... +

[Sora | OpenAI]

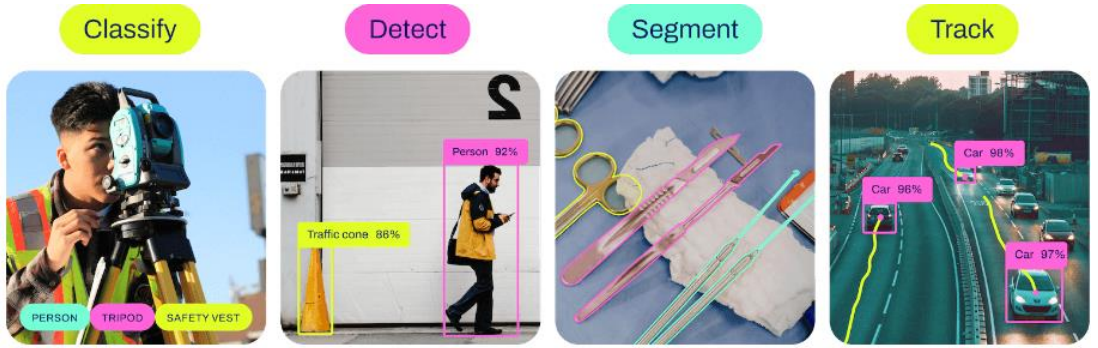
BERT family

tabular data



dmlc XGBoost

computer vision



[YOLO]

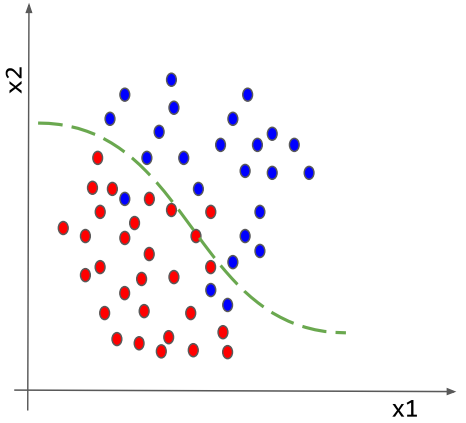
Deep Learning & Foundation Models

Classic Supervised Learning

input-output mapping

specific models for each task and data set

feature engineering



Example: Spam Filtering

Classify emails as spam or no spam

*use accordingly **labeled emails as training set***

use information like **occurrence of specific words** or **email length** as **features**

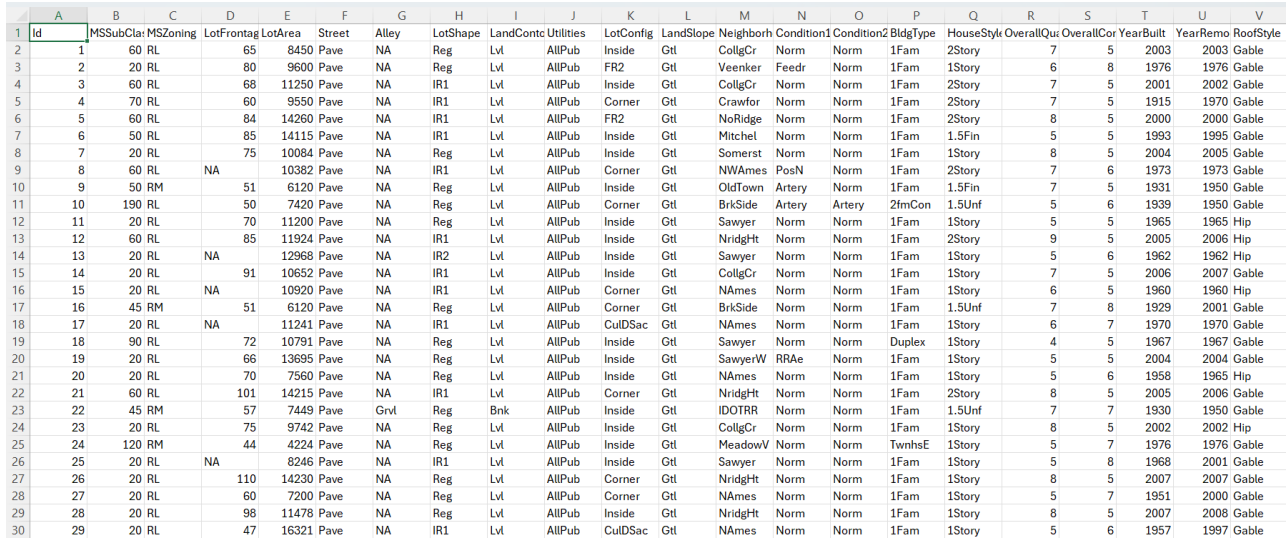
features x1 and x2

spam, no spam

Classic ML Use Cases

plenty of applications:

- prediction of house prices
- energy consumption prediction
- demand forecasting
- churn prediction
- forecast of traffic conditions
- weather forecasts
- predictive maintenance
- …



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Id	MSSubCla	MSZoning	LotFrontag	LotArea	Street	Alley	LotShape	LandConto	Utilities	LotConfig	LandSlope	Neighborh	Condition1	Condition2	BldgType	HouseStyle	OverallQua	OverallCor	YearBuilt	YearRemo	RoofStyle
2	1	60	RL	65	8450	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	2Story	7	5	2003	2003	Gable
3	2	20	RL	80	9600	Pave	NA	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm	1Fam	1Story	6	8	1976	1976	Gable
4	3	60	RL	68	11250	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	2Story	7	5	2001	2002	Gable
5	4	70	RL	60	9550	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	Norm	1Fam	2Story	7	5	1915	1970	Gable
6	5	60	RL	84	14260	Pave	NA	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	Norm	1Fam	2Story	8	5	2000	2000	Gable
7	6	50	RL	85	14115	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	Mitchel	Norm	Norm	1Fam	1.5Fin	5	5	1993	1995	Gable
8	7	20	RL	75	10084	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	Somerst	Norm	Norm	1Fam	1Story	8	5	2004	2005	Gable
9	8	60	RL	NA	10382	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl	NWAmes	PosN	Norm	1Fam	2Story	7	6	1973	1973	Gable
10	9	50	RM	51	6120	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	OldTown	Artery	Norm	1Fam	1.5Fin	7	5	1931	1950	Gable
11	10	190	RL	50	7420	Pave	NA	Reg	Lvl	AllPub	Corner	Gtl	BrkSide	Artery	Artery	2fmCon	1.5Unf	5	6	1939	1950	Gable
12	11	20	RL	70	11200	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	Sawyer	Norm	Norm	1Fam	1Story	5	5	1965	1965	Hip
13	12	60	RL	85	11924	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	NridgHt	Norm	Norm	1Fam	2Story	9	5	2005	2006	Hip
14	13	20	RL	NA	12968	Pave	NA	IR2	Lvl	AllPub	Inside	Gtl	Sawyer	Norm	Norm	1Fam	1Story	5	6	1962	1962	Hip
15	14	20	RL	91	10652	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	1Story	7	5	2006	2007	Gable
16	15	20	RL	NA	10920	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl	NAmes	Norm	Norm	1Fam	1Story	6	5	1960	1960	Hip
17	16	45	RM	51	6120	Pave	NA	Reg	Lvl	AllPub	Corner	Gtl	BrkSide	Norm	Norm	1Fam	1.5Unf	7	8	1929	2001	Gable
18	17	20	RL	NA	11241	Pave	NA	IR1	Lvl	AllPub	CulDSac	Gtl	NAmes	Norm	Norm	1Fam	1Story	6	7	1970	1970	Gable
19	18	90	RL	72	10791	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	Sawyer	Norm	Norm	Duplex	1Story	4	5	1967	1967	Gable
20	19	20	RL	66	13695	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	SawyerW	RRAe	Norm	1Fam	1Story	5	5	2004	2004	Gable
21	20	20	RL	70	7560	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	NAmes	Norm	Norm	1Fam	1Story	5	6	1958	1965	Hip
22	21	60	RL	101	14215	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl	NridgHt	Norm	Norm	1Fam	2Story	8	5	2005	2006	Gable
23	22	45	RM	57	7449	Pave	Grvl	Reg	Bnk	AllPub	Inside	Gtl	IDOTRR	Norm	Norm	1Fam	1.5Unf	7	7	1930	1950	Gable
24	23	20	RL	75	9742	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	1Story	8	5	2002	2002	Hip
25	24	120	RM	44	4224	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	MeadowV	Norm	Norm	TwnhsE	1Story	5	7	1976	1976	Gable
26	25	20	RL	NA	8246	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl	Sawyer	Norm	Norm	1Fam	1Story	5	8	1968	2001	Gable
27	26	20	RL	110	14230	Pave	NA	Reg	Lvl	AllPub	Corner	Gtl	NridgHt	Norm	Norm	1Fam	1Story	8	5	2007	2007	Gable
28	27	20	RL	60	7200	Pave	NA	Reg	Lvl	AllPub	Corner	Gtl	NAmes	Norm	Norm	1Fam	1Story	5	7	1951	2000	Gable
29	28	20	RL	98	11478	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl	NridgHt	Norm	Norm	1Fam	1Story	8	5	2007	2008	Gable
30	29	20	RL	47	16321	Pave	NA	IR1	Lvl	AllPub	CulDSac	Gtl	NAmes	Norm	Norm	1Fam	1Story	5	6	1957	1997	Gable

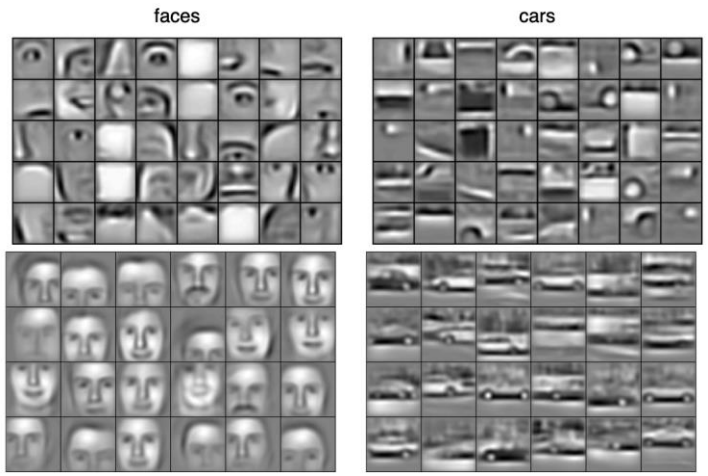
You got it, these are predictive models.

Ladder of Generalization

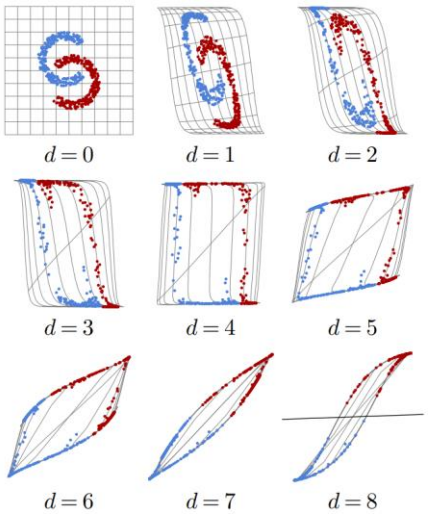
classic ML: feature engineering

deep learning: feature learning

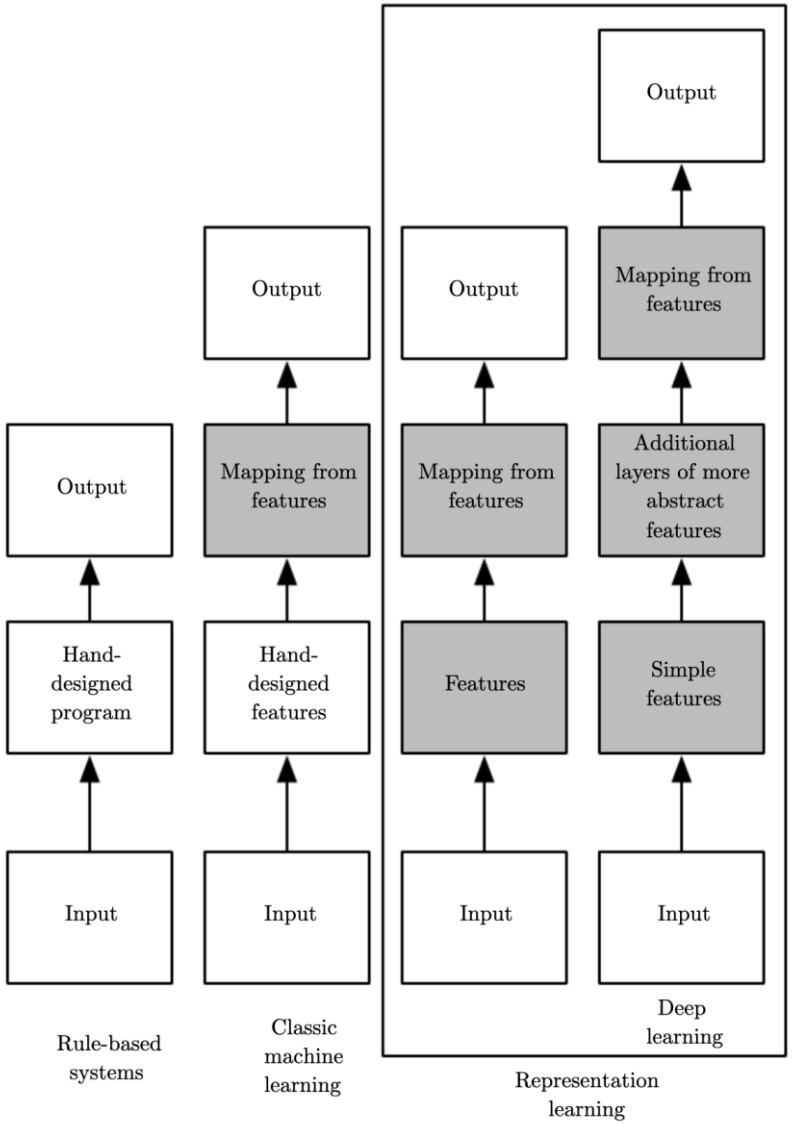
(hierarchy of concepts learned from raw data in deep graph with many layers)



source



source



source

Structured/Tabular vs Unstructured Data

unstructured data: homogenous

→ deep learning rules

→ allows transfer learning (foundation models in CV and NLP)

ImageNet

The Lord of the Rings

Article Talk

From Wikipedia, the free encyclopedia

(Redirected from Lord of the rings)

This article is about the book. For other uses, see The Lord of the Rings (disambiguation). "War of the Ring" redirects here. For other uses, see War of the Ring (disambiguation).

The Lord of the Rings is an epic[1] high fantasy novel[2] by the English author and scholar J. R. R. Tolkien. Set in Middle-earth, the story began as a sequel to Tolkien's 1937 children's book *The Hobbit*, but eventually developed into a much larger work. Written in stages between 1937 and 1949, ***The Lord of the Rings*** is one of the best-selling books ever written, with over 150 million copies sold.[3]

The title refers to the story's main antagonist,[4] Sauron, the Dark Lord who in an earlier age created the One Ring to rule the other Rings of Power given to Men, Dwarves, and Elves, in his campaign to conquer all of Middle-earth. From homely beginnings in the Shire, a hobbit land reminiscent of the English countryside, the story ranges across Middle-earth, following the quest to destroy the One Ring, seen mainly through the eyes of the hobbits Frodo, Sam, Merry, and Pippin. Aiding Frodo are the Wizard Gandalf, the Men Aragorn and Boromir, the Elf Legolas, and the Dwarf Gimli, who unite in order to rally the Free Peoples of Middle-earth against Sauron's armies and give Frodo a chance to destroy the One Ring in the fire of Mount Doom.

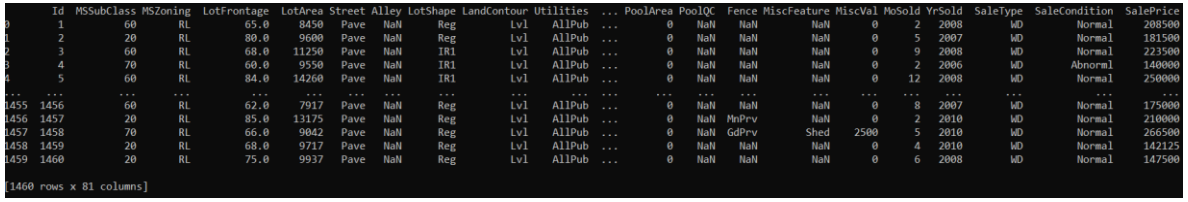
Although often mistakenly called a trilogy, the work was intended by Tolkien to be one volume in a two-volume set along with *The Silmarillion*.[nb 1] For economic reasons, ***The Lord of the Rings*** was first published over the course of a year from 29 July 1954 to 20 October 1955 in three volumes rather than one[nb 2] under the titles *The Fellowship of the Ring*, *The Two Towers*, and *The Return of the King*; *The Silmarillion* appeared only after the author's death. The work is divided internally into six books, two per volume, with several appendices of background material.[5] These three volumes were later published as a boxed set, and even finally as a single volume, following the author's original intent.

structured data: heterogenous

→ feature engineering needed

→ deep learning loses its advantage over shallow methods

→ e.g., gradient boosting still used a lot



	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	9	2008	WD	Normal	223500
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	140000
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	8	2007	WD	Normal	175000
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	MnPrv	NaN	0	2	2010	WD	Normal	210000
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	GdPrv	Shed	2500	5	2010	WD	Normal	266500
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	4	2010	WD	Normal	142125
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	6	2008	WD	Normal	147500

[1460 rows x 81 columns]

Transfer Learning

idea:

- generic pre-training of foundation models on huge data sets
- subsequent finetuning for specific tasks on small(er) data sets

(usually done with deep learning methods, using its compositional nature)

very successful for:

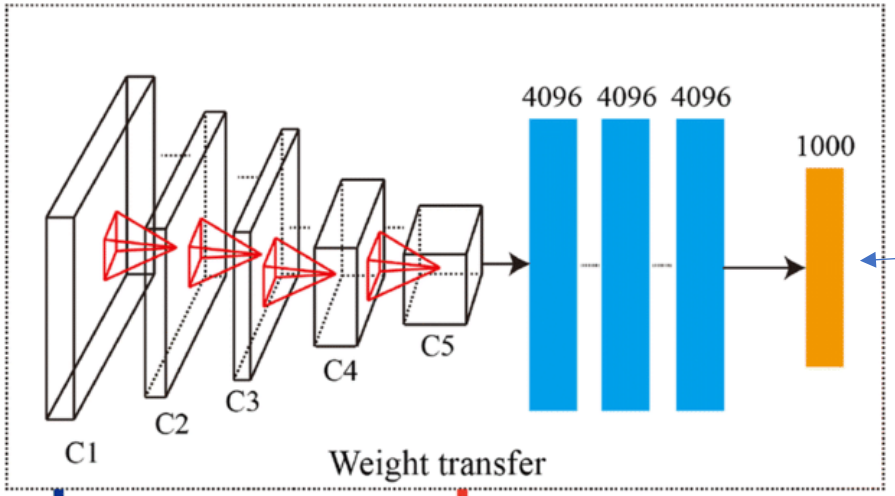
- computer vision (e.g., object classification)
- language models (e.g., BERT, GPT)

not (yet) for tabular data (due to its notorious heterogeneity)

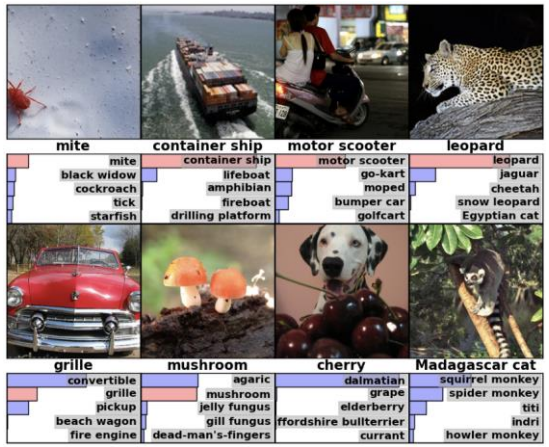
(but some promising research, e.g., Chronos for univariate time series)

CNN Finetuning

Pretrained on ImageNet



object classification

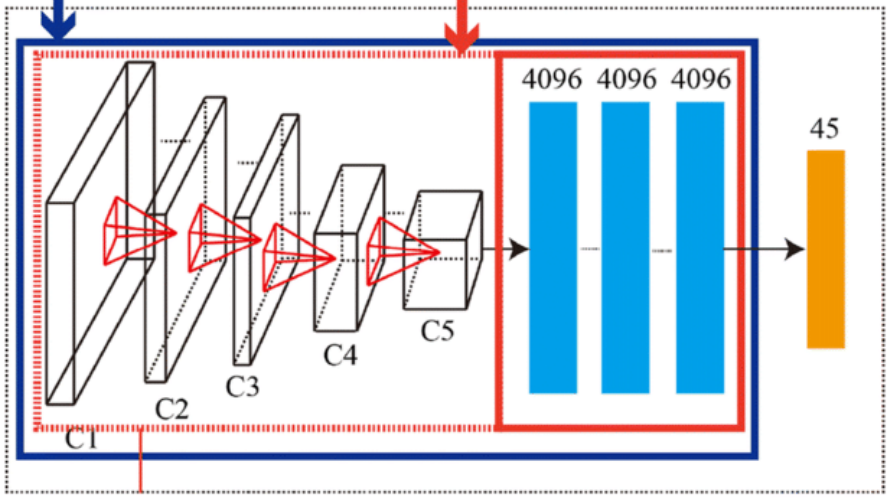


source

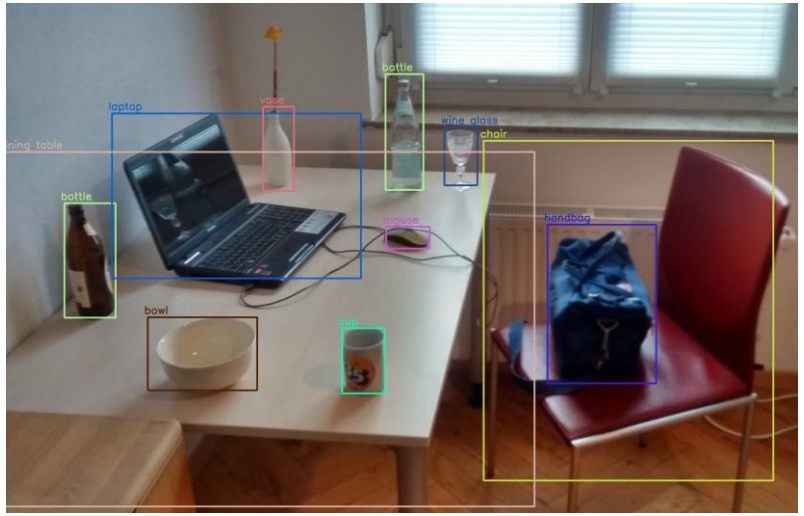
Finetune all layers

Finetune the last few layers

Finetuned on NWPU



source

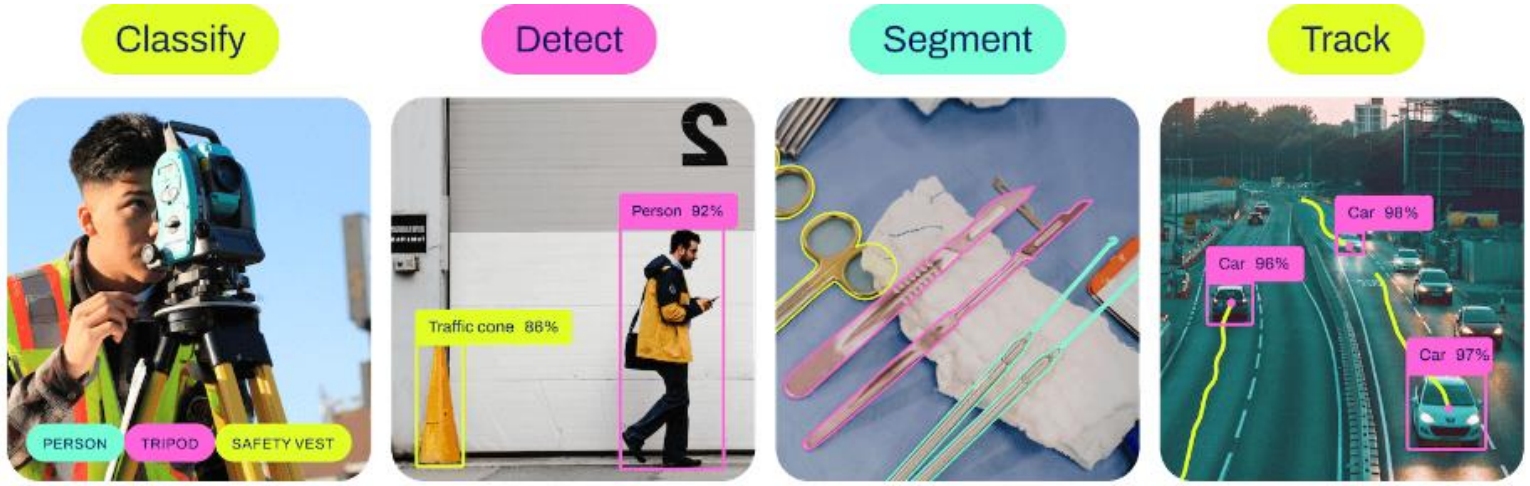


from wikipedia

Computer Vision Use Cases

many use cases:

- visual defect detection
- face recognition
- perception for autonomous driving, drone control, …
- cellular segmentation
- support chip design
- …



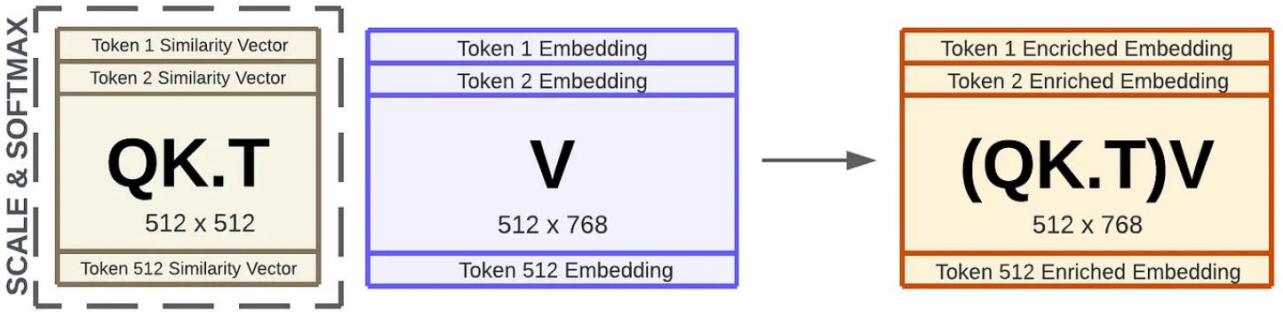
YOLO

Again, these are no generative, but predictive models.

But aspects of it can be applied in generative ones (e.g., U-Net in diffusion models).

Language Models: Contextual Semantics

- self-supervised learning: e.g., next/masked-word prediction
- tokenization: split text into chunks (e.g., words)
- semantics by means of vector embeddings: e.g., via bag-of-words (or end-to-end in transformer)
- positional encoding & embeddings: order of sequence
- contextual embeddings: (self-)attention (weighted averages: influence from other tokens)

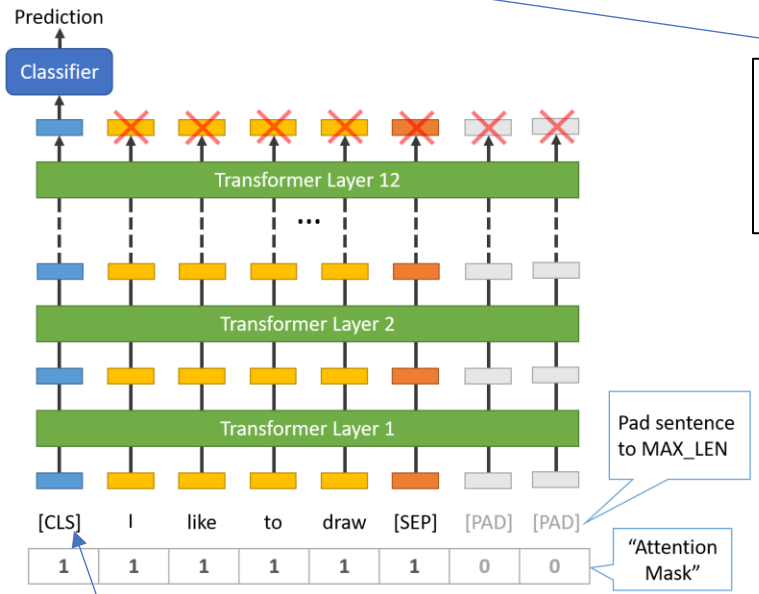


source

Language Model Finetuning

- self-supervised pre-training on massive data sets (→ foundation models like GPT or BERT)
- subsequent supervised finetuning on specific data sets (adapting parameters or/and adding layers)

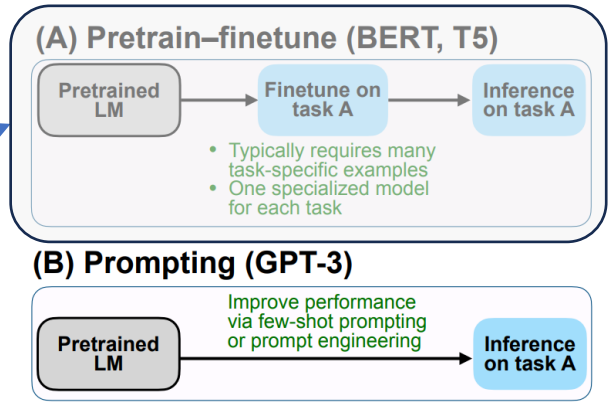
example: sequence classification



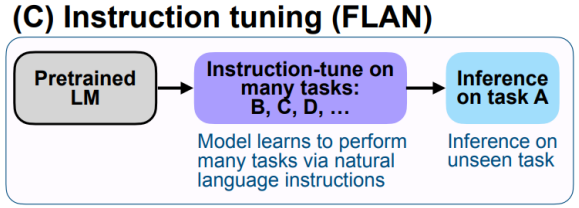
source

sentence classification via [CLS] token

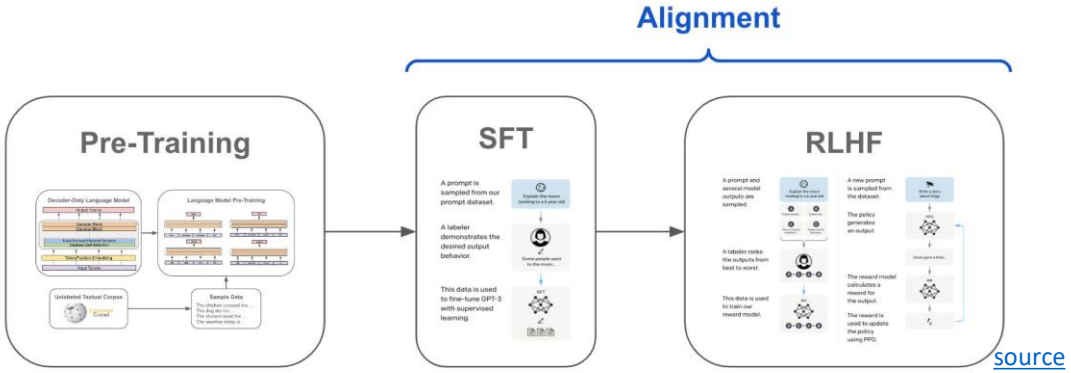
purely predictive



source



tuning predictive, output generative



source

Generative AI

Generative vs Predictive/Discriminative Models

discriminative models:

predict conditional probability $P(Y|\boldsymbol{X})$

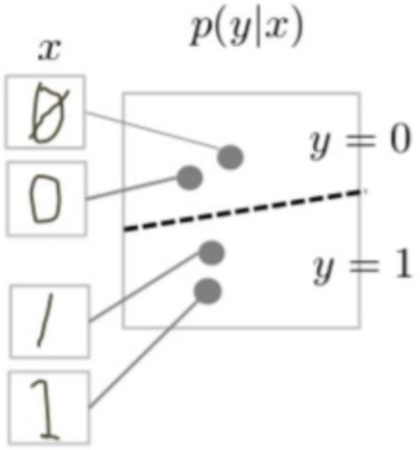
generative models:

predict joint probability $P(Y, \boldsymbol{X})$

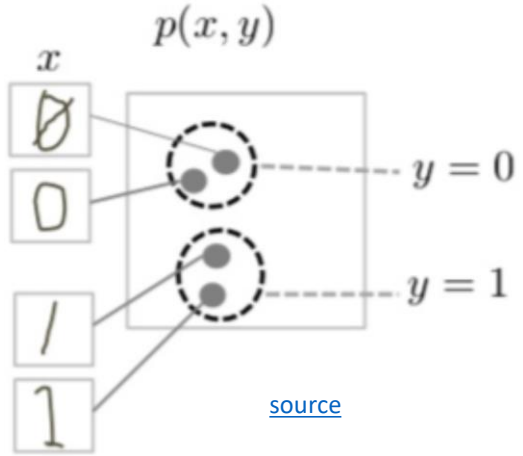
(or just $P(\boldsymbol{X})$ → unsupervised learning)

→ allow to generate new data samples

discriminative model



generative model



source

task of generative models more difficult: need to model full data distribution rather than merely find patterns in inputs to distinguish outputs

Generative models can be used for predictive tasks (Bayes theorem).
But predictive models are usually better at it.

Deep Learning for Generative AI

Depending on the application, there are currently two dominant approaches for generative AI:

- text generation: decoder LLMs

- image synthesis: diffusion models

note the difference between image synthesis and multimodal understanding in LLMs (images as additional input sequences to transformer, tokenized by splitting into patches)

most new LLMs decoder-only

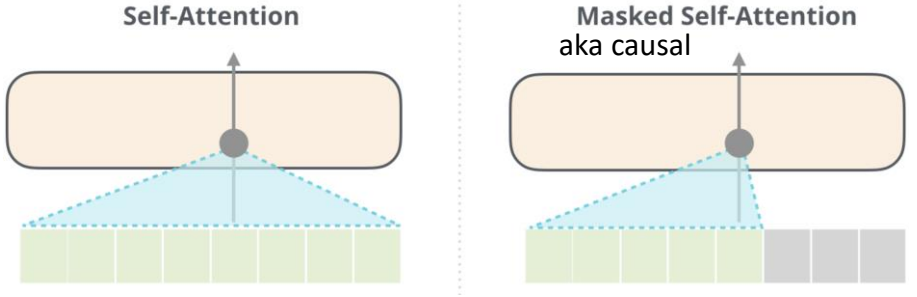
Encoder vs Decoder LLMs

encoder-only LLMs

- prime example: BERT
- self-supervised pre-training: masked-word prediction
- finetuning on downstream tasks (e.g., sequence classification)
- can’t generate text
- can’t be prompted

decoder-only LLMs

- prime example: GPT
- self-supervised pre-training: next-word prediction
- instruction tuning (e.g., RL from human feedback)
- generate text: chat bots
- prompt engineering (zero-/few-shot)



source

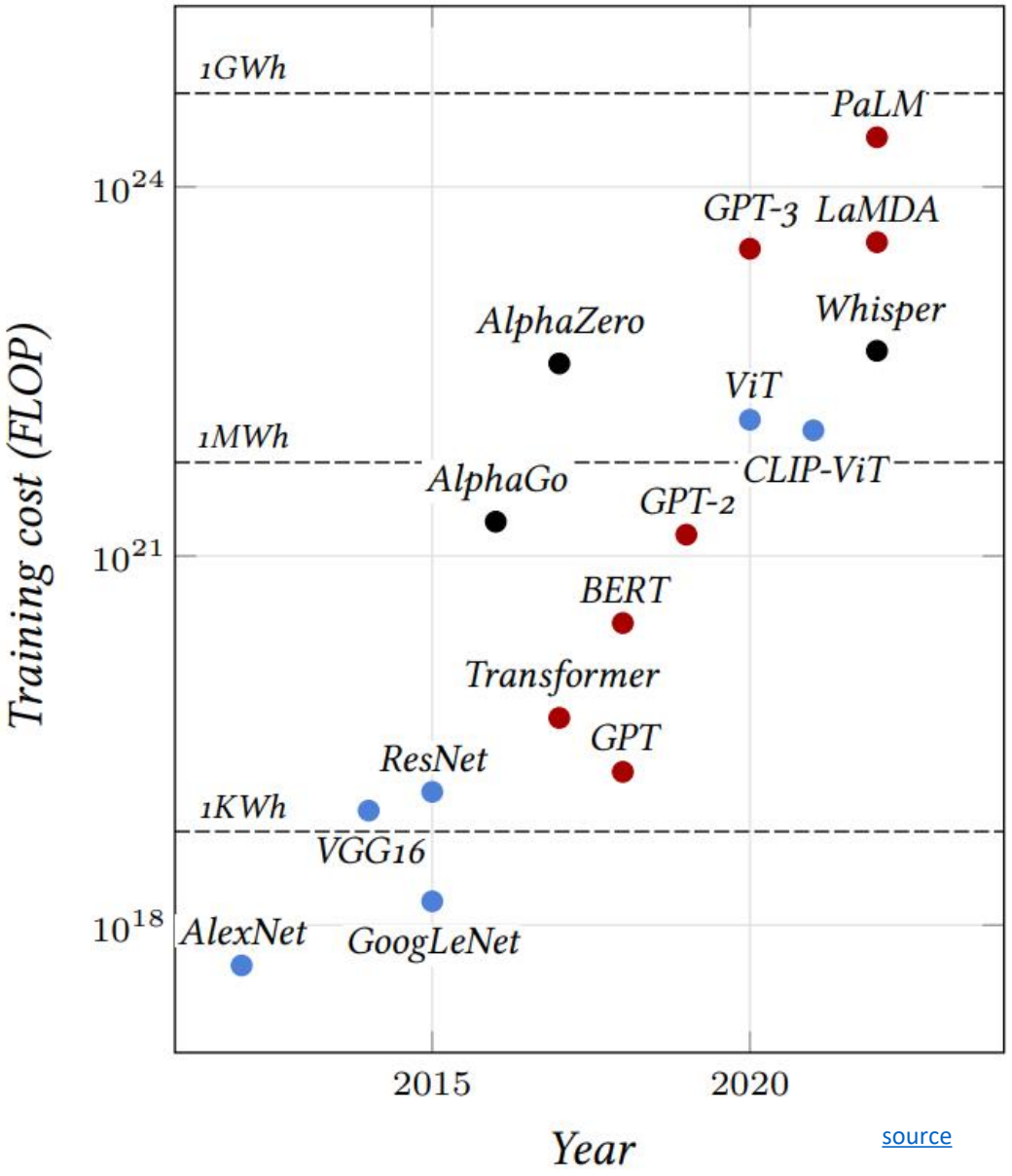
Text Generation

in-context learning as alternative to finetuning (new paradigm):
feed information into LLM via input prompt (decoder LLMs)
→ attention to context

typical prompt:
instructions, context (potentially retrieved), query, output indicator

enables multi-task capabilities (all of ML before was only narrow tasks)
→ assistants

some deep learning history:



source

(open-source) LLM SOTA (July 2024): Llama 3

some LLM numbers (example Llama 3 405B):

- vocabulary size (tokens): 128K
- embedding/model dimensions: 16,384
- parameters: 405B
- training tokens: 15.6T
- context length/window (tokens): 128K
- training hardware: 16K GPUs (H100)

factor less than 40
→ a lot of memorizing

another trend: small language models on edge devices

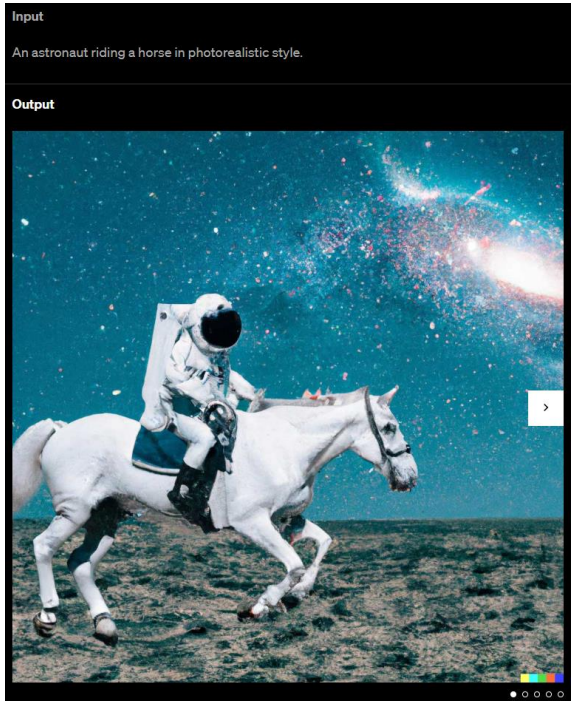
Image Synthesis

idea: generate new images as variations of training data

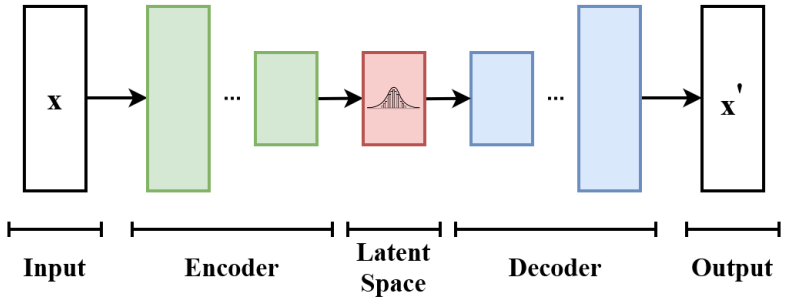
usually conditioned on text (prompt) by transformers

compared to text generation, additional mechanism needed (e.g., diffusion) to create more complex structures

example: DALL-E 2



Variational AutoEncoder:



from wikipedia

Diffusion models:
Gradually add Gaussian noise and then reverse

chain of denoising autoencoders

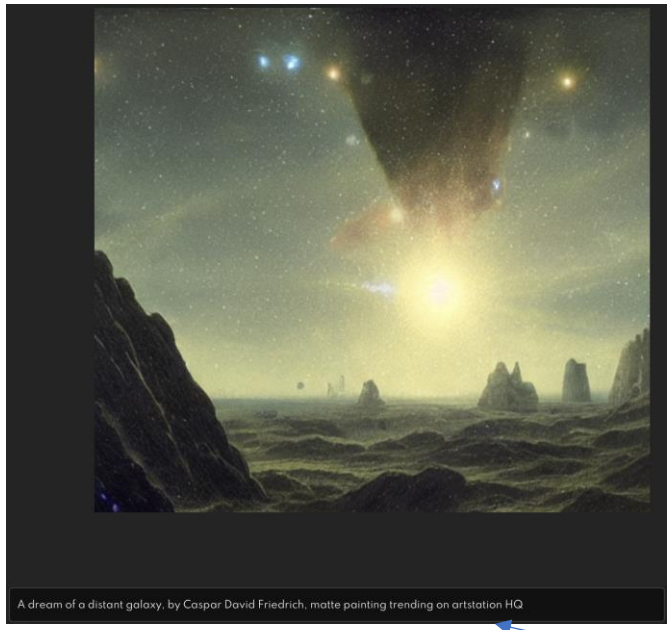
source

plenty of applications: DALL-E, Stable Diffusion, ImageGen, Midjourney, ...

rather "translations"

also text-to-speech (VALL-E, Speech T5, …) (and speech recognition, e.g., Whisper), text-to-video (Make-A-Video, Lumiere, Sora, …) → dynamics/physics understanding/simulation

web app for Stable Diffusion: DreamStudio



inpainting example (GLIDE):

"zebras roaming in the field"

source

prompt

at some point maybe also proteins, materials, structured data, …

Decision Making

predictive

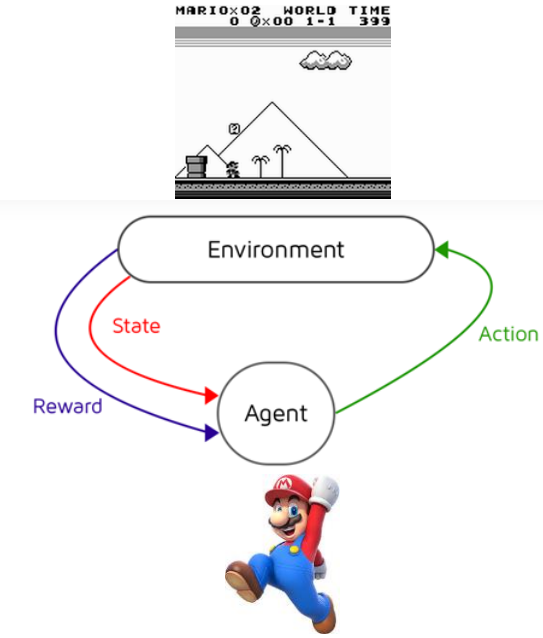
Sequential Decision Making

no concrete input-output examples like in supervised learning

typically, domain of

- reinforcement learning (e.g., Q-learning or policy-gradient methods → model-free)
- optimal control (e.g., model predictive control → model-based)

plenty of use cases: e.g., games or complex control mechanisms



generative

Alternative: Sequence Modeling

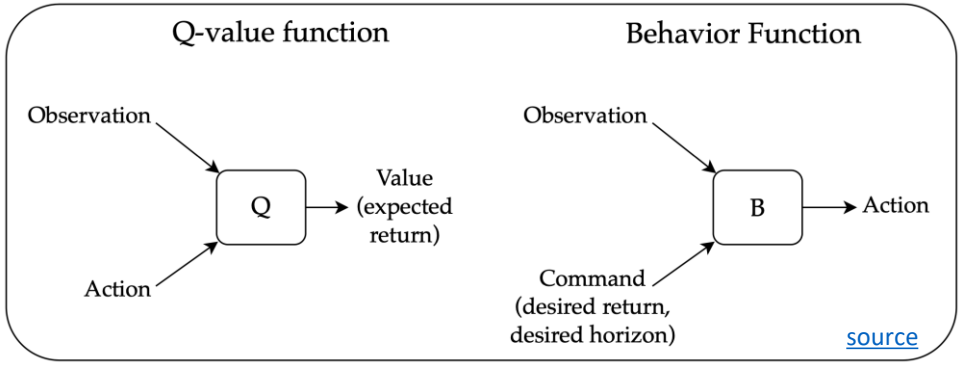
offline method

not restricted by Markov property

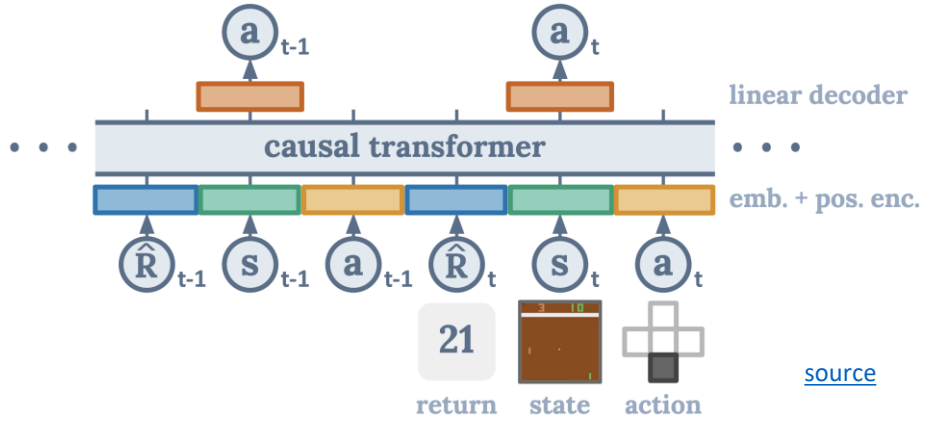
Decision Transformer:

- generative: transformer to autoregressively model trajectories
- credit assignment directly via self-attention: state-return associations
- desired return tokens as prompt for action generation (but extrapolation is difficult)

overcoming the deadly triad of deep RL:



source



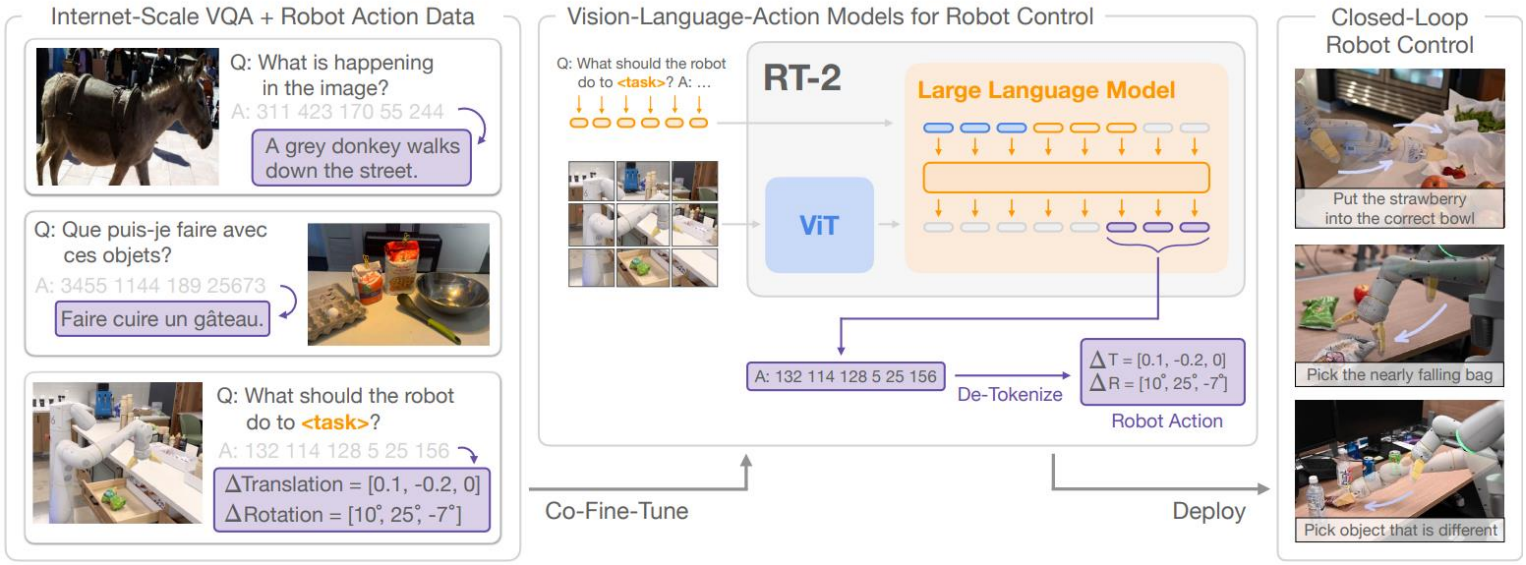
source

or apply language models directly ...

Example Use Case: Robotic Control

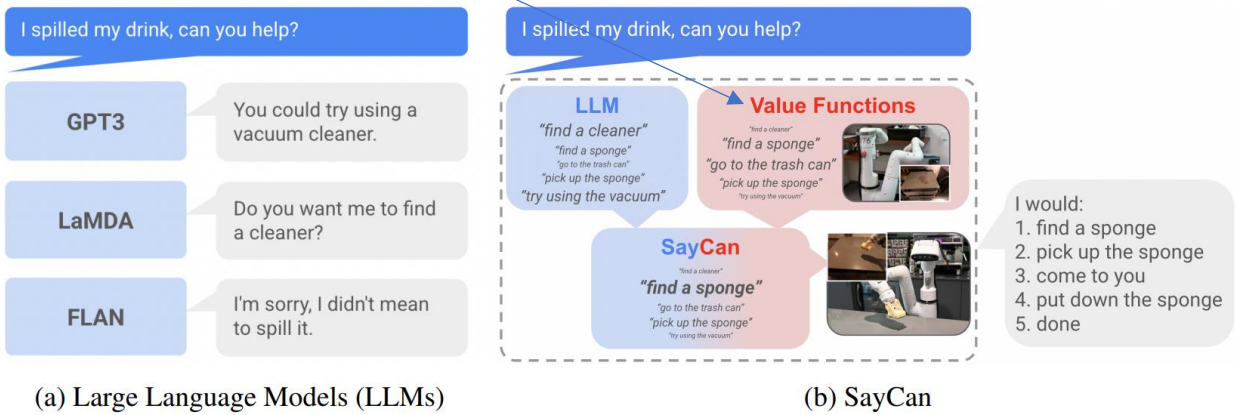
generated by LLMs/VLMs

RT-2:



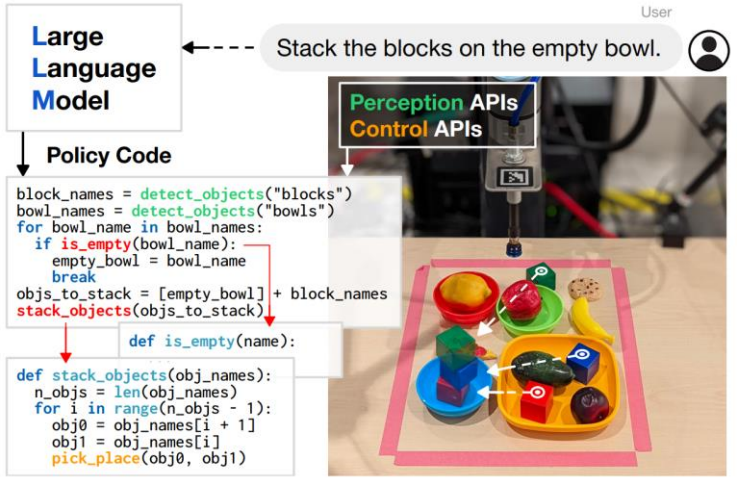
diffusion objective for multimodal outputs

SayCan (grounding with pre-trained skills):



(a) Large Language Models (LLMs) (b) SayCan

Code as Policies:



Agents

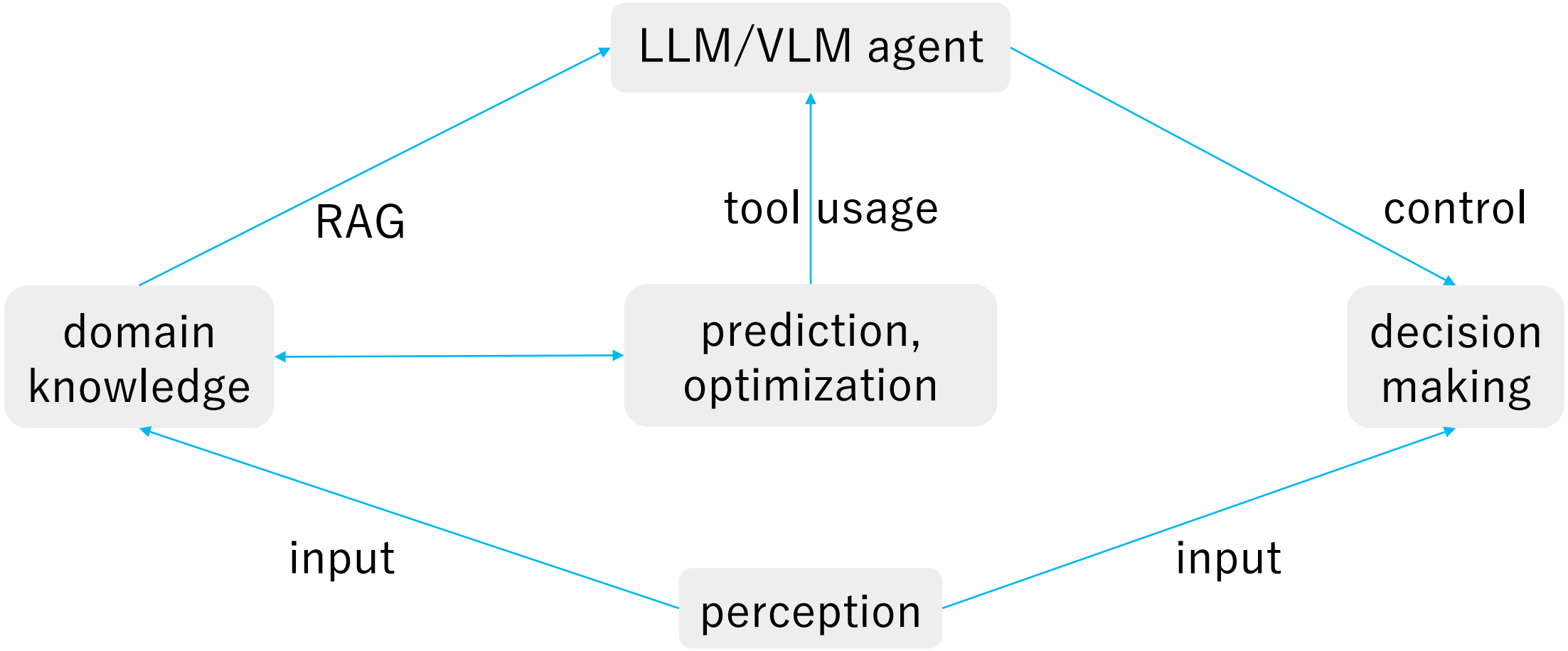
LLM Agents

desired agent capabilities:

- planning (LLMs: decomposition of complex issue in multiple simple steps)
- tool use (LLMs: use predictive models for numerical/optimization tasks)
- reflection (LLMs: clever prompting schemes)
- collaboration with other agents (LLMs: mutual prompting)

(but capabilities still limited, e.g., reasoning)

Goal: Autonomous End-to-End Workflow



So, when to use what, now?

Deep learning (especially as foundation models) rules for unstructured, classic ML (specific models) is (still) a good choice for tabular data.

Generative models allow to create new data (text, code, images, video) and can be used in a very generic way (prompting).

But predictive models are usually better for predictive tasks, and business problems are often predictive (numerical, optimization, …).

They can be combined: generative model (e.g., an LLM) as orchestrator, using predictive models as tools.

And some Philosophical Thoughts

Current AI (SOTA: LLM) is good at learning statistical patterns and making predictions.

But there is no sign of real “understanding”. → stochastic parrot?

Anyway, generative AI is an interesting, slightly twisted mirror of our creations.