AI Slides Viewer & Checker - 2

Sunghee Yun

Co-founder & CTO - AI Technology & Product Strategy

Erudio Bio, Inc.

Table of contents

AI & Biotech

AI in biology

- AI has been used in biological sciences, and science in general
- AI's ability to process large amounts of raw, unstructured data ($e.g.$, DNA sequence data)
  - reduces time and cost to conduct experiments in biology
  - enables others types of experiments that previously were unattainable
  - contributes to broader field of engineering biology or biotechnology
- AI increases human ability to make direct changes at cellular level and create novel genetic material ($e.g.$, DNA and RNA) to obtain specific functions.

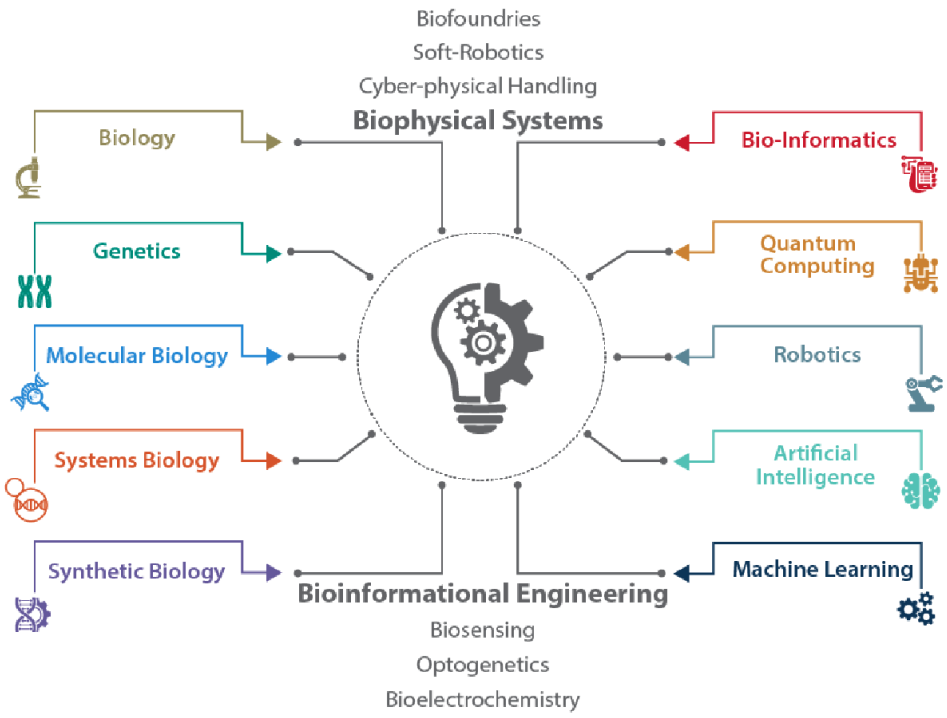
Biotech

Biotech

- biotechnology
  - is multidisciplinary field leveraging broad set of sciences and technologies
  - relies on and builds upon advances in other fields such as nanotechnology & robotics, and, increasingly, AI
  - enables researchers to read and write DNA
    - sequencing technologies "read" DNA while gene synthesis technologies takes sequence data and "write" DNA turning data into physical material
- 2018 National Defense Strategy & senior US defense and intelligence officials identified emerging technologies that could have disruptive impact on US national security [Say21]
  - artificial intelligence, lethal autonomous weapons, hypersonic weapons, directed energy weapons, *biotechnology,* quantum technology
- other names for biotechnology are engineering biology, synthetic biology, biological science (when discussed in context of AI)

biotech - multidisciplinary field

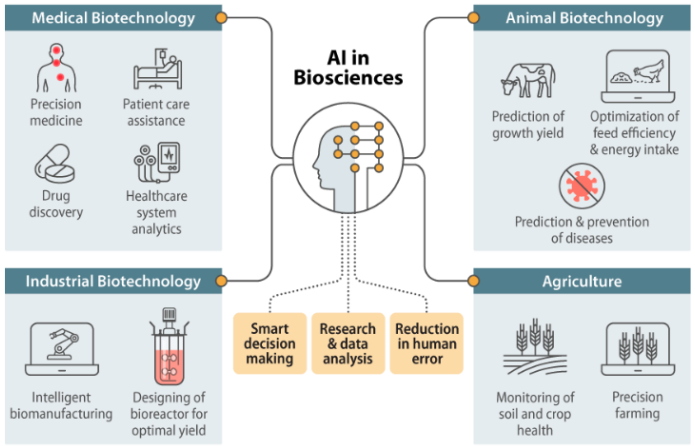
- sciences and technologies enabling biotechnology include, but not limited to,
  - (molecular) biology, genetics, systems biology, synthetic biology, bio-informatics, quantum computing, robotics [DFJ22]



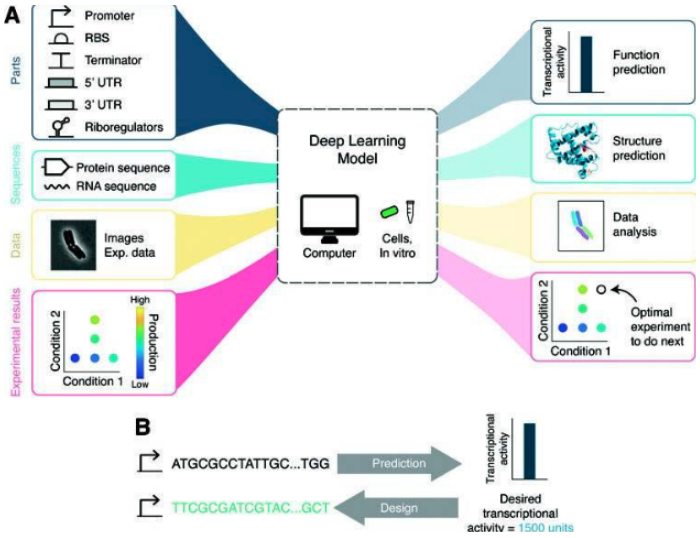
Convergence of AI and biological design

- both AI & biological sciences increasingly converging [BKP22]
  - each building upon the other's capabilities for new research and development across multiple areas
- Demo Hassabis, CEO & cofounder of DeepMind, said of biology [Toe23]

  > ". . . biology can be thought of as information processing system, albeit extraordinarily complex and dynamic one . . . just as mathematics turned out to be the right description language for physics, biology may turn out to be *the perfect type of regime for the application of AI!*"
- Both AI & biotech rely on and build upon advances in other scientific disciplines and technology fields, such as nanotechnology, robotics, and increasingly big data (*e.g.*, genetic sequence data)
  - each of these fields itself convergence of multiple sciences and technologies
- so *their impacts can combine to create new capabilities*



Multi-source genetic sequence data



- AI is essential to analyzing exponential growth of genetic sequence data

  > "AI will be essential to fully understanding how genetic code interacts with biological processes"
  > - US National Security Commission on Artificial Intelligence (NSCAI)

  – process huge amounts of biological data, *e.g.*, genetic sequence data, coming from different biological sources for understanding complex biological systems

    - sequence data, molecular structure data, image data, time-series, omics data

- *e.g.*, analyze genomic data sets to determine the genetic basis of particular trait and potentially uncover genetic markers linked with that trait

Quality & quantity of biological data

- limiting factor, however, is quality and quantity of the biological data, $e.g.$, DNA sequences, that AI is trained on
  - $e.g.$, accurate identification of particular species based on DNA requires reference sequences of *sufficient quality* to exist and be available
- databases have varying standards - access, type and quality of information
- design, management, quality standards, and data protocols for reference databases can affect utility of particular DNA sequence

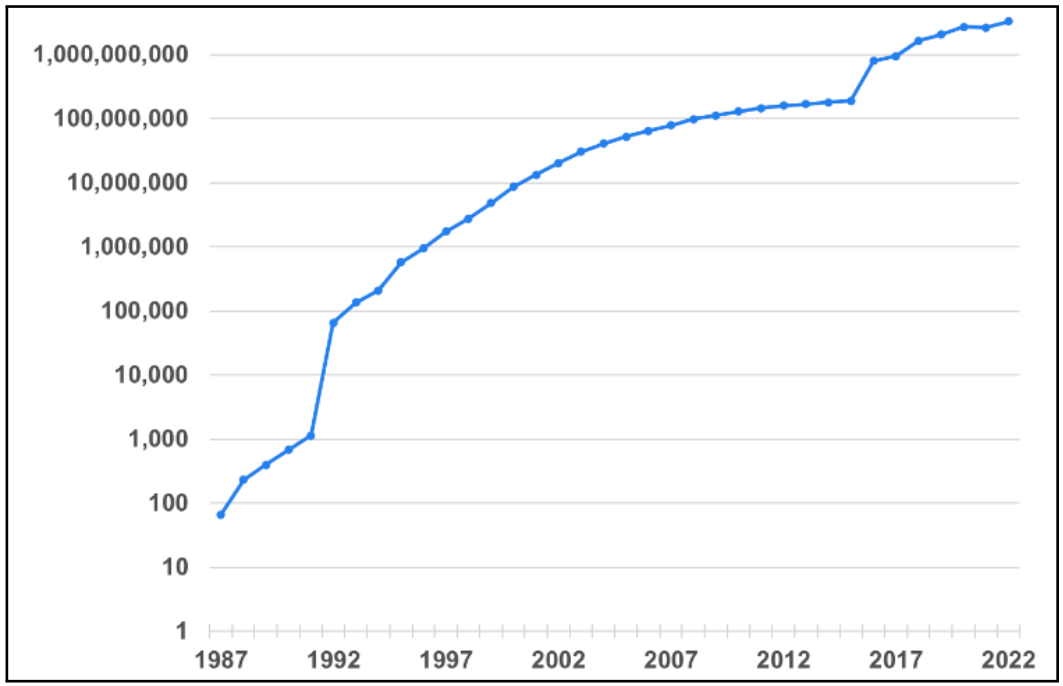
Rapid growth of biological data

- volume of genetic sequence data grown exponentially as sequencing technology has evolved
- more than 1,700 databases incorporating data on genomics, protein sequences, protein structures, plants, metabolic pathways, $etc.$, $e.g.$
  - open-source public database
    - Protein Data Bank, US-funded data center, contains more than *terabyte of three-dimensional structure data* for biological molecules, including proteins, DNA, and RNA
  - proprietary database
    - Gingko Bioworks - possesses more than *2B protein sequences*
  - public research groups
    - Broad Institute - produces roughly *500 terabases of genomic data per month*
- great potential value in aggregate volume of genetic datasets that can be collectively mined to discover and characterize relationships among genes

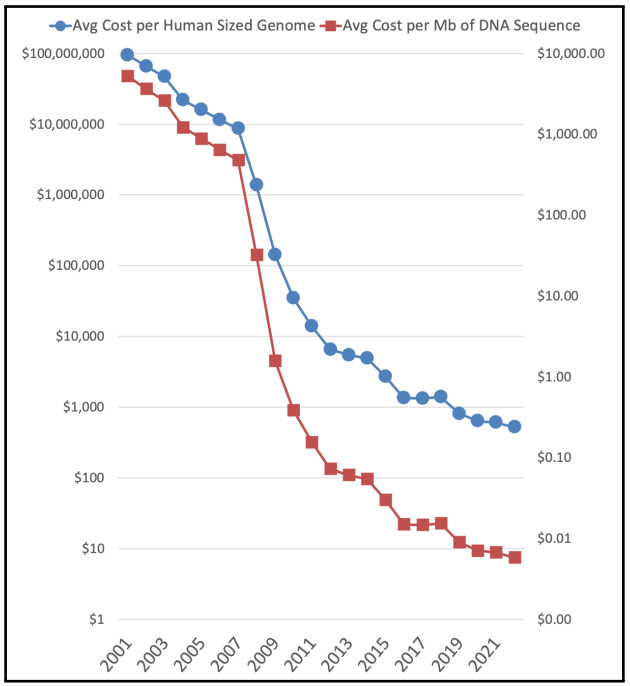
Volume and sequencing cost of DNA over time

- volume of DNA sequences & DNA sequencing cost
  - data source: National Human Genome Research Institute (NHGRI) [Wet23] & International Nucleotide Sequence Database Collaboration (INSDC)

sequences in INSDC



DNA sequencing cost



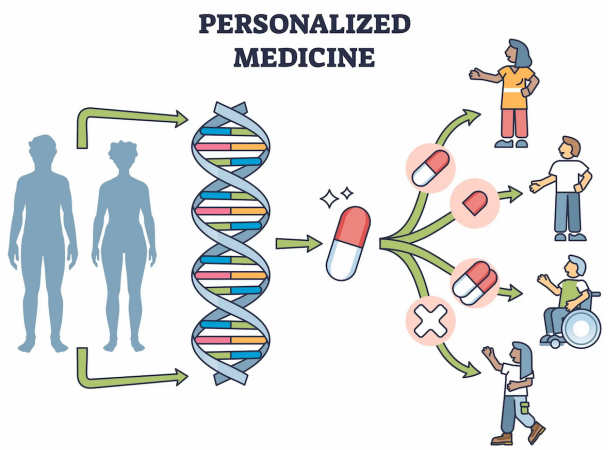
Bio data availability and bias

- US National Security Commission on Artificial Intelligence (NSCAI) recommends
  - US fund and prioritize development of a biobank containing *"wide range of high-quality biological and genetic data sets securely accessible by researchers"*
  - establishment of database of broad range of human, animal, and plant genomes would
    - *enhance and democratize biotechnology innovations*
    - *facilitate new levels of AI-enabled analysis of genetic data*
- bias - availability of genetic data & decisions about selection of genetic data can introduce bias, $e.g.$
  - training AI model on datasets emphasizing or omitting certain genetic traits can affect how information is used and types of applications developed - *potentially privileging or disadvantaging certain populations*
  - access to data and to AI models themselves may impact communities of differing socioeconomic status or other factors unequally

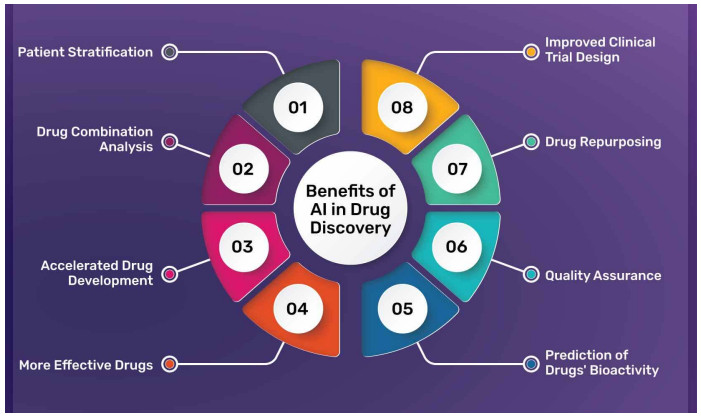
Emerging Trends in Biotech

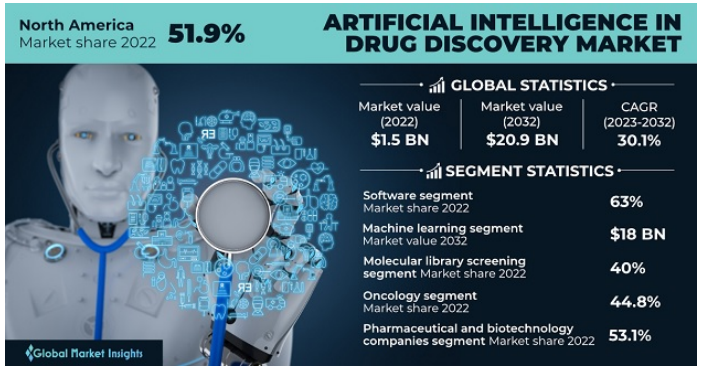
Personalized medicine

- *shift from one-size-fits-all approach to tailored treatments*
- based on individual genetic profiles, lifestyles & environments
- AI enables analysis of vast data to predict patient responses to treatments, thus enhancing efficacy and reducing adverse effects
- $e.g.$, custom cancer therapies, personalized treatment plans for rare diseases & precision pharmacogenomics.
- companies - Tempus, Foundation Medicine, $etc.$



AI-driven drug discovery

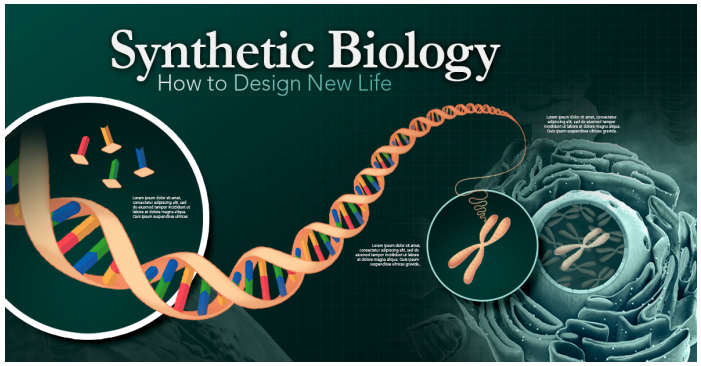


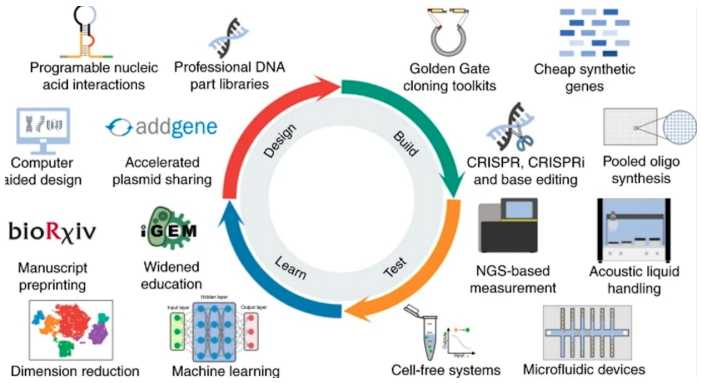


- traditional drug discovery process - time-consuming and costly often taking decades and billions of dollars
- AI streamlines this process by predicting the efficacy and safety of potential compounds with more speed and accuracy
- AI models analyze chemical databases to identify new drug candidates or repurpose existing drugs for new therapeutic uses
- companies - Insilco Medicine, Atomwise.

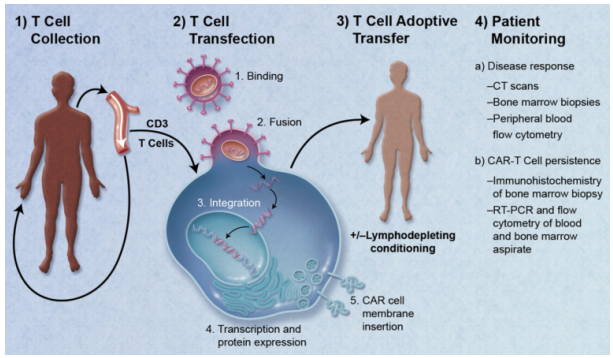
Synthetic biology

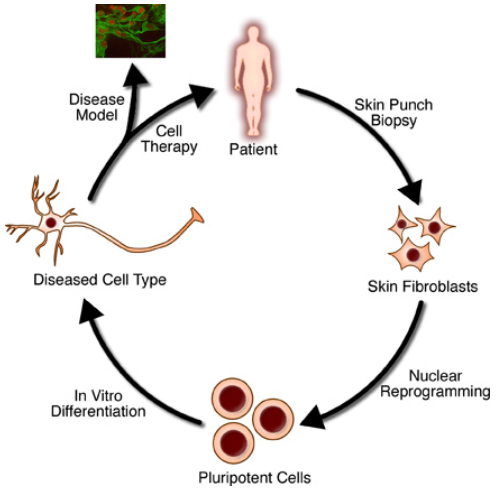
- use AI for gene editing, biomaterial production and synthetic pathways
- combine principles of biology and engineering to design and construct new biological entities
- AI optimizes synthetic biology processes from designing genetic circuits to scaling up production
- company - Ginkgo Bioworks uses AI to design custom microorganisms for applications ranging from pharmaceuticals to industrial chemicals





Regenerative medicine

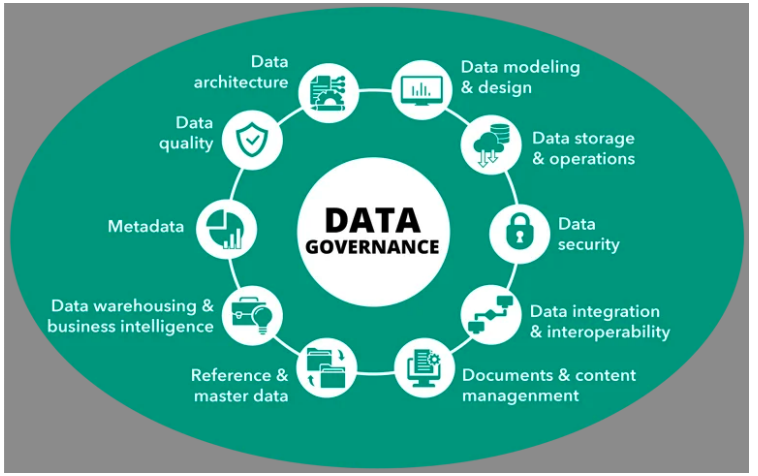




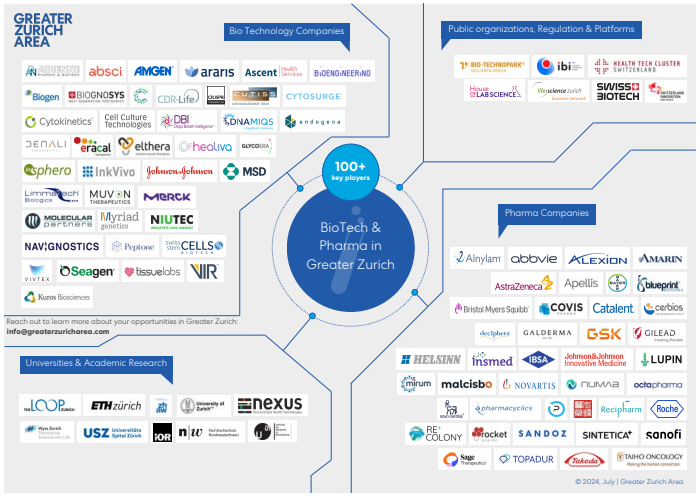
- AI advances development of stem cell therapies & tissue engineering
- AI algorithms assist in identifying optimal cell types, predicting cell behavior & personalized treatments
- particularly for conditions such as neurodegenerative diseases, heart failure and orthopedic injuries
- company - Organovo leverages AI to potentially improve the efficacy and scalability of regenerative therapies, developing next-generation treatments

Bio data integration

- integration of disparate data sources, including genomic, proteomic & clinical data - one of biggest challenges in biotech & healthcare
- AI delivers meaningful insights *only when* seamless data integration and interoperability realized
- developing platforms facilitating comprehensive, longitudinal patient data analysis - vital enablers of AI in biotech
- company - Flatiron Health working on integrating diverse datasets to provide holistic view of patient health



Biotech companies



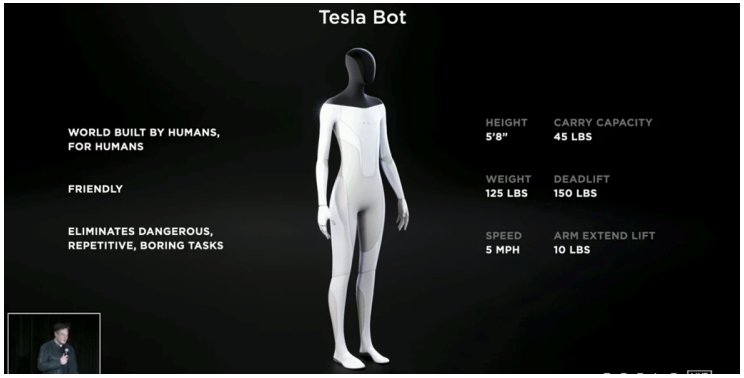
- Atomwise - small molecule drug discovery
- Cradle - protein design
- Exscientia - precision medicine
- Iktos - small molecule drug discovery and design
- Insilico Medicine - full-stack drug discovery system
- Schrödinger, Inc. - use physics-based models to find best possible molecule
- Absci Corporation - antibody design, creating new from scratch antibodies, $i.e.$, "de novo antibodies", and testing them in laboratories

AI-powered Humanoid Robots

Tesla Optimus

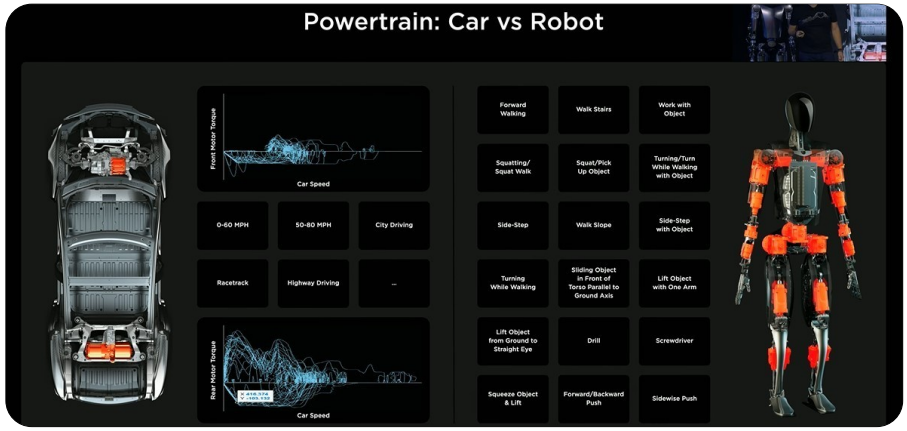
Tesla Optimus

- humanoid robot developed by Tesla intended to handle repetitive & dangerous tasks
- objective - *revolutionize automation* & assist in human labor across various industries
- features - YouTube - Optimus - Gen 2
  - dimensions - 5'8" tall & 125 lbs
  - capabilities - lifting weights, walking at 5 MPH & performing everyday tasks
  - AI-powered - runs on Tesla's AI leveraging same technology used in self-driving cars
  - power source - 2.3 KWH battery designed for efficient power management
  - launch year - announced by Elon Musk during Tesla AI Day in 2021
  - *price - $25,000∼$30,000* expected to decrease over time



History of Tesla Optimus

- inception - first conceptualized as extension of Tesla's AI & robotics capabilities
- AI day 2021 - *officially announced by Elon Musk* w/ vision to solve labor shortages & improve productivity
- Sep 2022 - prototype unveiled
- gen 2 introduced in 2023 - improved capabilities
- Jun 2024 w/ more advanced tasks - *towards mass production for commercial applications*



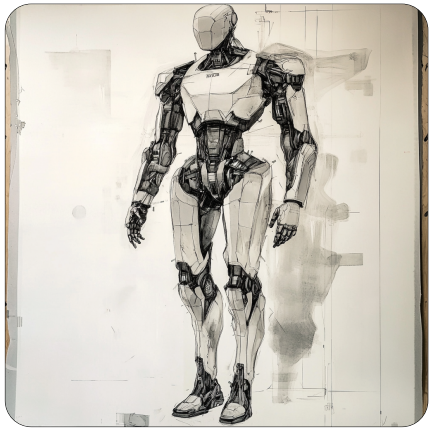
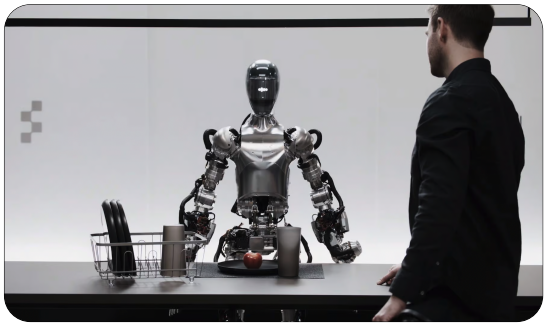
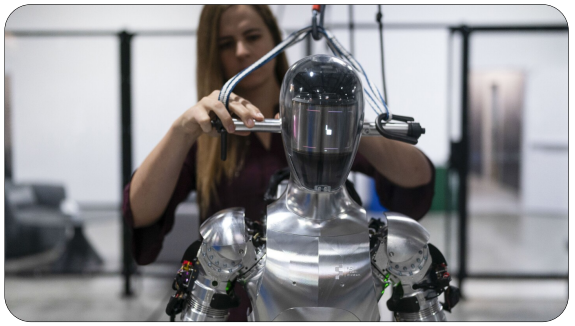


Figure AI

Figure AI robots

- Figure AI
  - founded in 2022 as Silicon Valley startup company by Brett Adcock - serial entrepreneur with successful Archer Aviation & Vettery
  - vision of enhancing productivity by integrating AI and robotics into both industrial & personal spaces
- Figure 02
  - 5'6" tall, 154 lbs, payload of 44 lbs, 5 hr runtime, 1.2 m/s speed
  - imitation learning
  - capabilities - advanced cognition, STS task, dexterous hands w/ 16 degrees of freedom





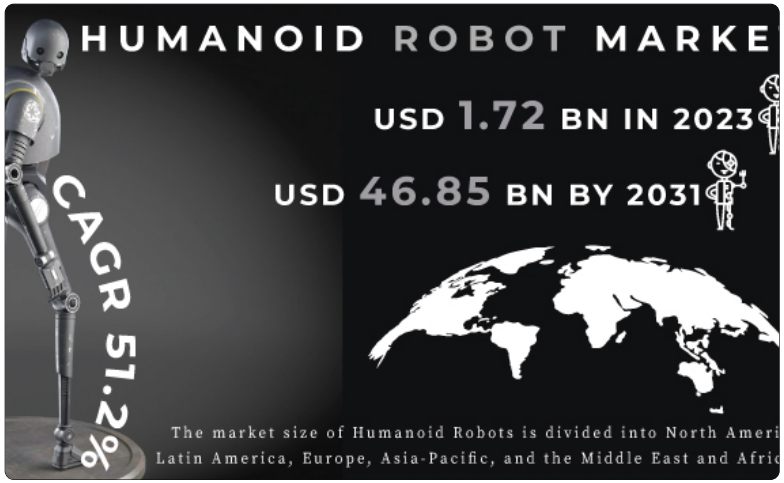
History of Figure AI

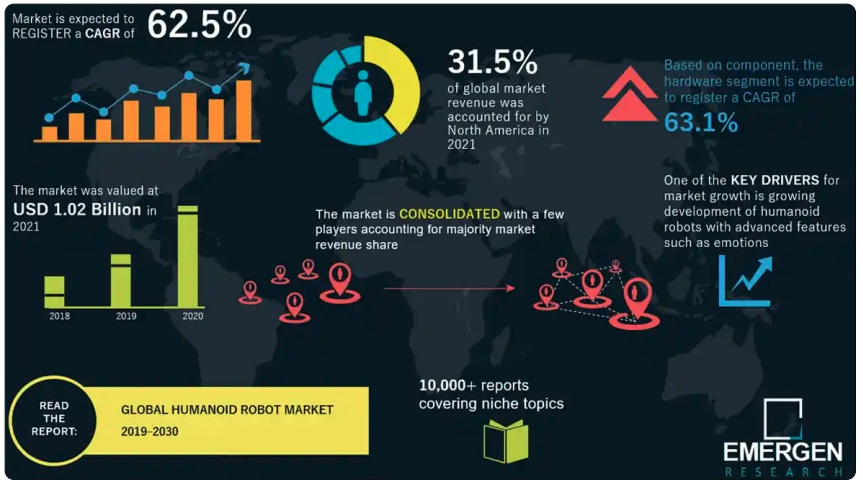
- 2022 - founded by Brett Adcock - previously co-founded Archer Aviation & Vettery
- 2022-2023 - early development - stealth mode focusing on developing their own technology
- May 2023 - public announcement - officially announces mission to develop general-purpose humanoid robots - already raised $70M @ announcement
- Aug 2023 - unveils Figure 01, first prototype w/ basic mobility & manipulation capabilities
- Oct 2023 - series B funding - raised $675M beyond initial goal of $500M - Jeff Bezos, Microsoft, OpenAI - valuation of $\sim$ $2.6B
- late 2023 $\sim$ early 2024 - partnership announcements - refines humanoid robot technology in locomotion, object manipulation & human-robot interaction
- 2024 - significant strides in robot control & decision-making

Impacts & Future

Impacts on industries & markets

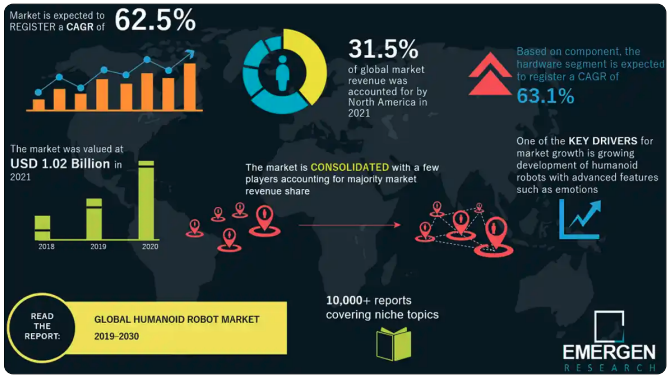
- impacts on robotics history
  - competitor benchmark - competes with robotics giants such as Boston dynamics
  - affordability & scale - predict to lead to *lower costs & higher adoption*
- impacts on labor market
  - task automation - replace human labor in *high-risk & repetitive roles*
  - *job displacement vs creation* - new roles in AI, robot maintenance & oversight
- impacts on consumer market - home automation





Future outlook & predictions

- widespread industrial adoption - expected to become common tool in factories by 2030
- market valued @ *$1.02B in 2021* - expected *CAGR of 62.5%, 63.1% in hardware segment by 2030* - *31.5% revenue increase in 2021* North America - *10,000 humanoid robots* will be shipped worldwide each year by 2027
- AI evolution - continuous learning and AI enhancements will lead to greater efficiency & adaptability
- consumer integration - long-term vision includes personal assistant
- societal impact - could redefine human roles in industries & homes *raising philosophical & ethical questions on human-robot collaboration*



Industrial AI

Industrial AI (inAI)

- inAI (collectively) refers to AI technology & software and their products developed for
  - *customer values creation, productivity improvement, cost reduction, production optimization, predictive analysis, insight discovery*

  in industries such as
  - *semiconductor, steel, oil & gas, cement, and other various manufacturing industries*

  (unlike general AI, which is frontier research discipline striving to achieve human-level intelligence)

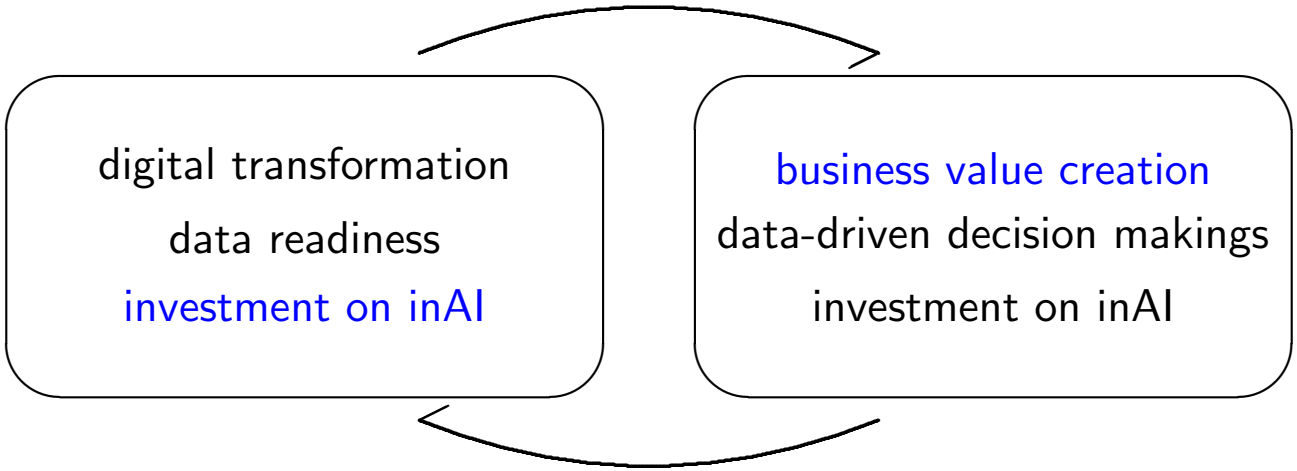
inAI fields

- product
  - product design & innovation, adaptability & advancement, product quality & validation, design for reusability & recyclability, performance optimization
- production process
  - *production quality,* process management, inter-process relations, process routing & scheduling, process design & innovation, *traceability, predictive process control*
- machinery & equipment
  - *predictive maintenance, monitoring & diagnosis,* component development, *ramp-up optimization,* material consumption prediction
- supply chain
  - supply chain monitoring, material requirements planning, customer management, supplier management, logistics, reusability & recyclability

Characteristics of inAI

Vicious (or virtuous) cycle

- integration of inAI with customers' business creates monetary values and encourages data-driven decisions
- however, to do so, digital transformation with data-readiness is MUST-have
- created values, in turn, can be invested into infrastructure required for digital transformation and success of inAI!



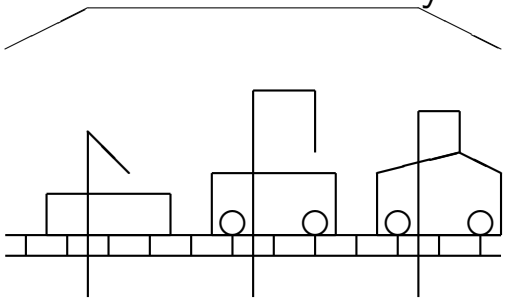
Data-centric AI

- unlike many ML disciplines where foundation models do generic representation learning, $i.e.$, learn universal features
- each equipment has (gradually) different data characteristics, hence need data-centric AI

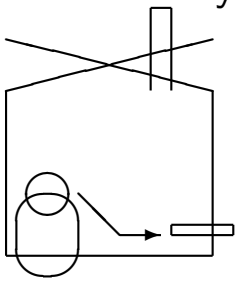
  ". . . need 1,000 models for 1,000 problems" - Andrew Ng

  – data-centric AI - discipline of systematically engineering the data used to build AI system

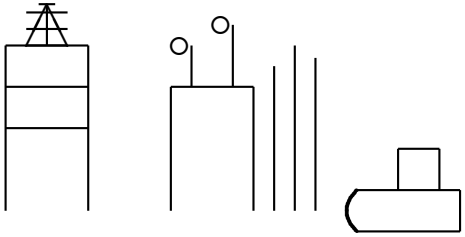
automobile industry



steel industry



oil & gas industry



Challenging data characteristics

- huge volume
- data multi-modality
- high velocity requirement
- very fat data
- sever data shift & drift (in many cases)
- label imbalance
- data quality

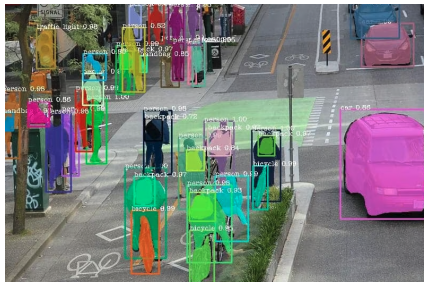
Manufacturing AI

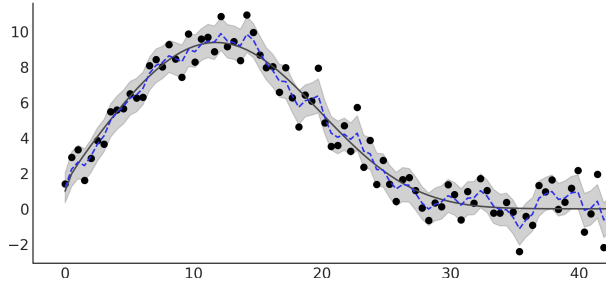
MLs in manufacturing AI (manAI)

- *image data* - huge amount of image data measured and inspected
  - SEM/TEM images, wafer defect maps, test failure pattern maps [1]

  $\rightarrow$ semantic segmentation, defect inspection, anomaly detection
- *time-series (TS) data* - *all the data* coming out of manufacturing is TS
  - equipment sensor data, process times, various measurements, MES data [2]

  $\rightarrow$ regression, anomaly detection, semi-supervised learning, Bayesian inference





[1] SEM: scanning electron miscroscope, TEM: transmission electron miscroscope

[2] MES: manufacturing execution system

CV ML in manAI

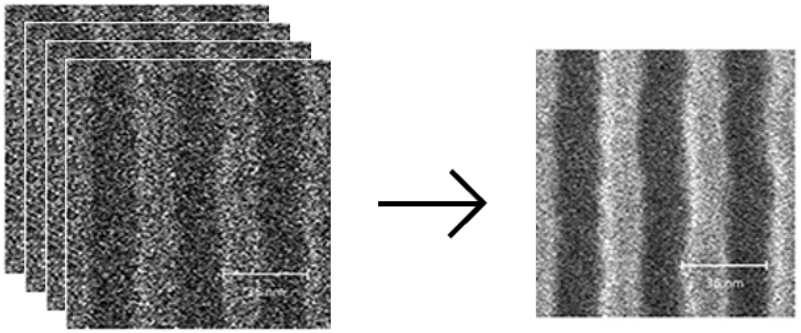
Computer vision ML in manAI

- measurement and inspection (MI)
  - metrology - measurement of critical features
  - inspection - defect inspection, defect localization, defect classification
  - failure pattern analysis
- applications
  - automatic feature measurement
  - anomaly detection
  - defect inspection

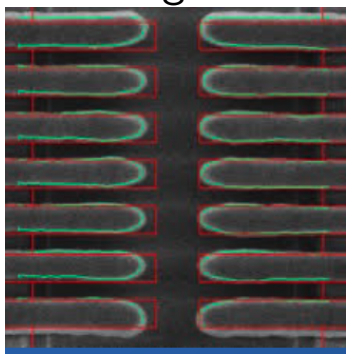
Automatic feature measurement

- ML techniques
  - image enhancement (denoising)
  - texture segmentation
  - repetitive pattern recognition
  - automatic measurement

image enhancement



texture segmentation



auto-measurement

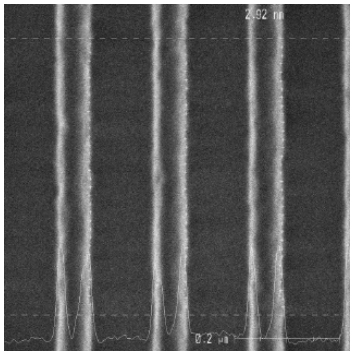
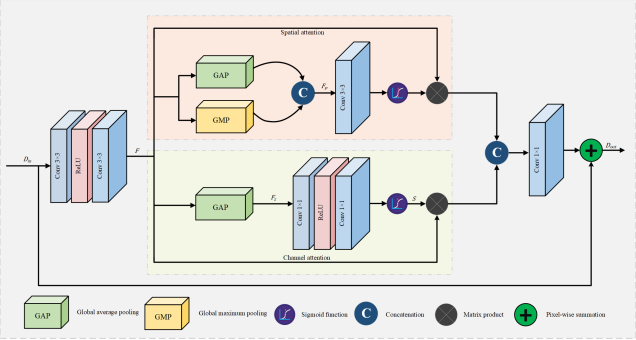


Image enhancement

- image enhancement techniques
  - general supervised denoising using DL
  - blind denoising using DL - remove noise without prior knowledge of noise adapting to various noise types
  - super-resolution - upscale low-resolution images, add realistic details for sharper & higher-quality images



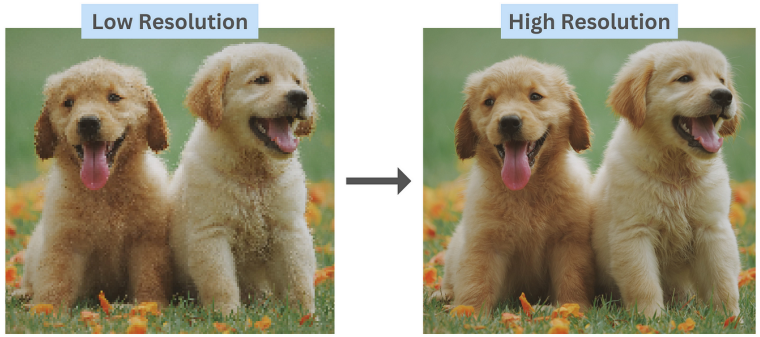
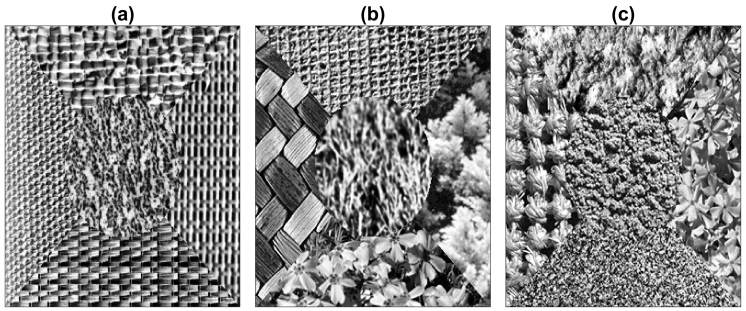


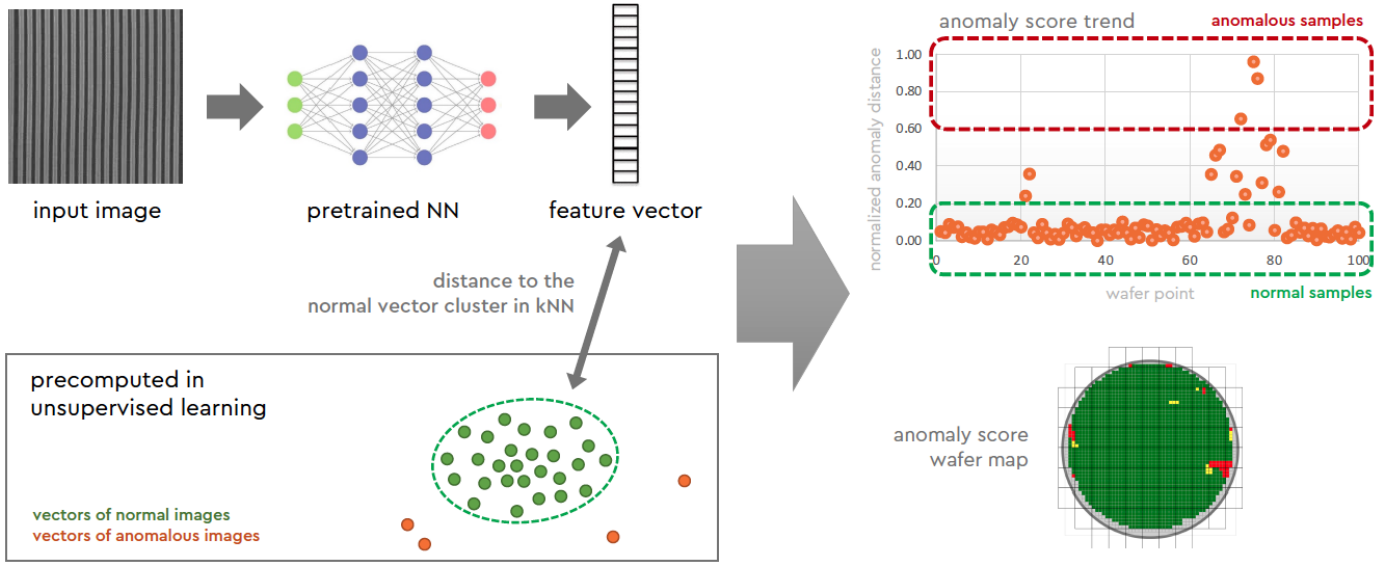
Image segmentation

- texture segmentation
  - distinguish areas based on texture patterns - identifying regions with similar textural features - used for material classification, surface defect detection, medical imaging
  - methods - Gabor filters, wavelet transforms, DL
- semantic segmentation
  - assign class labels to every pixel - enabling precise object and region identification - used for autonomous driving, scene understanding, medical diagnostics
  - methods - fully convolutional network (FCN), U-net, DeepLab



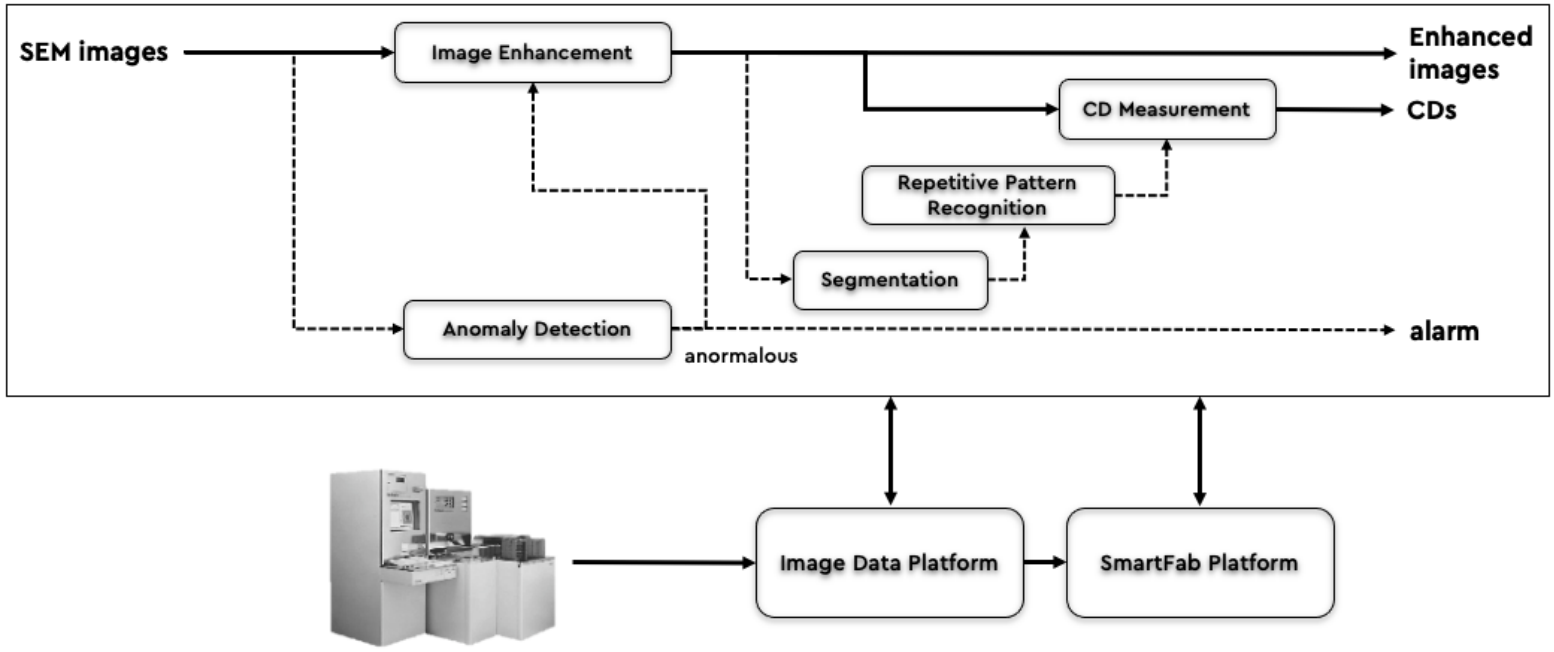
Anomaly detection using side product

- representation in embedding space obtained as side product from previous processes
- distance from normal clusters used for anomaly detection
- can be used for yield drop prediction and analysis



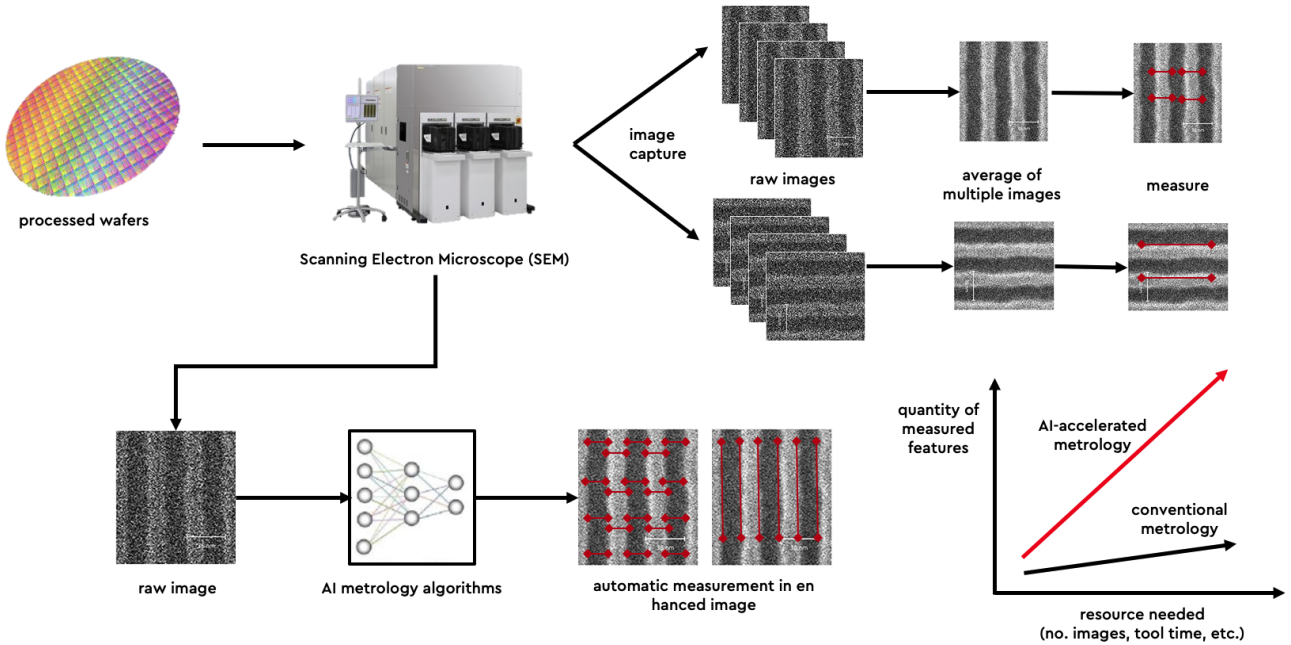
AI-enabled metrology system

- integration of separate components creates AI-enabled metrology system



Benefits of new system

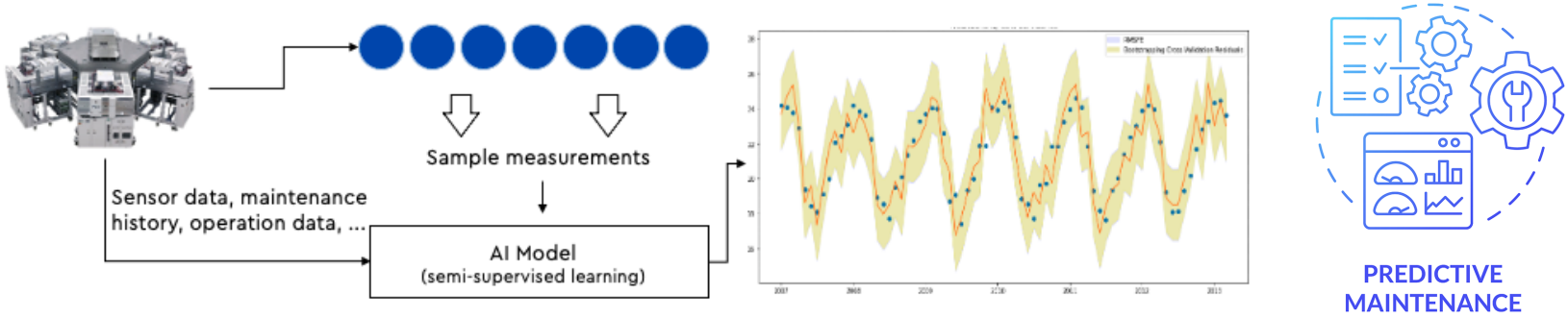
- new system provides
  - improved accuracy and reliability
  - improved throughput
  - savings on investment on measurement equipment



TS ML in manAI

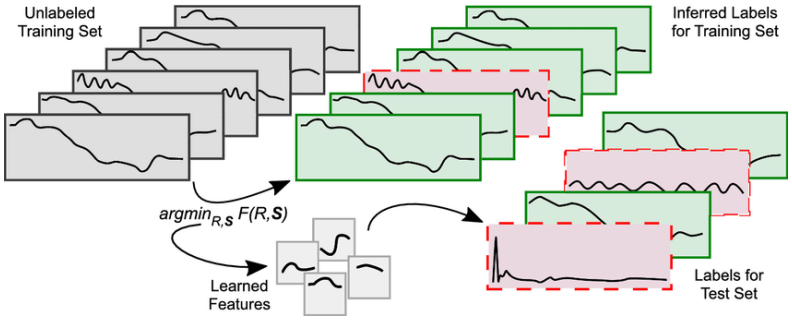
Time-series ML applications in manAI

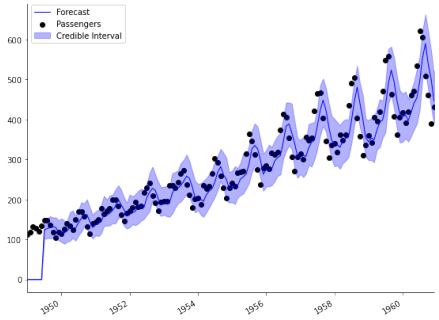
- estimation of TS values
  - virtual metrology - estimate measurement without physically measuring things
- anomaly detection on TS
  - predictive maintenance - predict maintenance times ahead
- multi-modal ML using LLM & genAI
  - root cause analysis and recommendation system



TS MLs in manAI

- TS regression/prediction/estimation
  - LSTM, GRU, attention-based models, Transformer-based architecture for capturing long-term dependencies and patterns
- anomaly detection
  - isolation forest, autoencoders, one-class SVM
- TS regression providing credibility intervals
  - Bayesian-based approaches offering uncertainty estimation alongside predictions



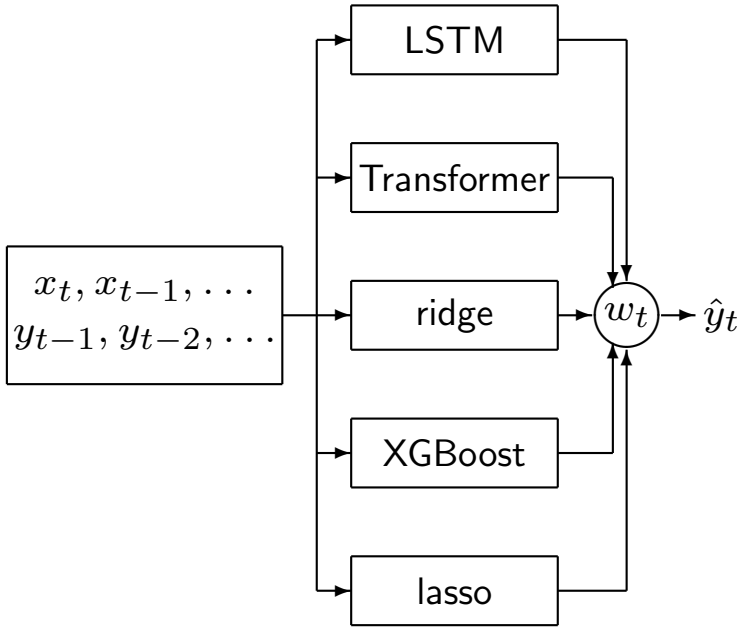


Difficulties with TS ML

- no definition exists for general TS data
- data drift & shift
  - $p(x_{t_k}, x_{t_{k-1}}, \ldots)$ changes over time
  - $p(y_{t_k}|x_{t_k}, x_{t_{k-1}}, \ldots, y_{t_{k-1}}, y_{t_{k-2}}, \ldots)$ changes over time
- (extremely) fat data, poor data quality, huge volume of data to process
- not many research results available
- none of algorithms in academic papers work / no off-the-shelf algorithms work

Online learning for TS regression

- use multiplie experts - $f_{1,k}, \ldots, f_{p_k,k}$ for each time step $t = t_k$ where $f_{i,k}$ can be any of following
  - seq2seq models ($e.g.$, LSTM, Transformer-based models)
  - non-DL statistical learning models ($e.g.$, online ridge regression)
- model predictor for t_k, $g_k : \mathbf{R}^n \to \mathbf{R}^m$ as weighted sum of experts



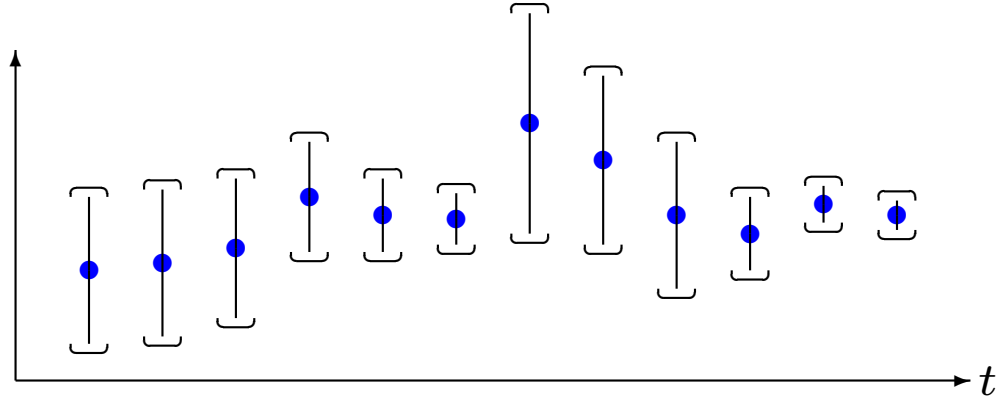
$$g_k = w_{1,k} f_{1,k} + w_{2,k} f_{2,k} + \cdots + w_{p_k,k} f_{p_k,k} = \sum_{i=1}^{p_k} w_{i,k} f_{i,k}$$

Credibility intervals

- every point prediction is wrong, $i.e.$

$$\mathbf{Prob}(\hat{y}_t = y_t) = 0$$

- reliability of prediction matters, however, *none* literature deals with this (properly)
- critical for our customers, $i.e.$, *such information is critical for downstream applications*
  - $e.g.$, when used for feedback control, need to know how reliable prediction resuilts are
  - sometimes *more crucial than algorithm accuracy*



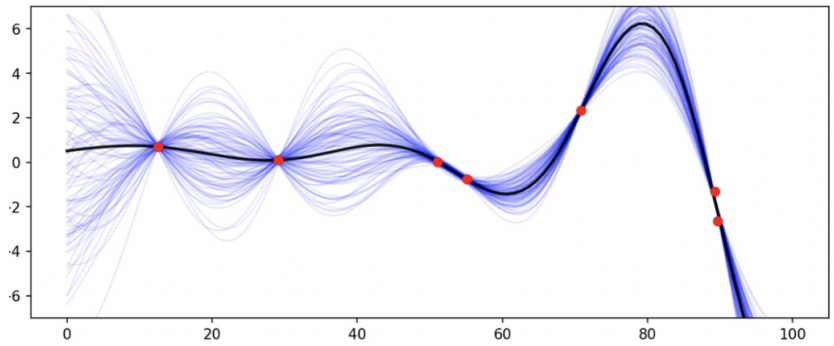
Bayesian approach for credibility interval evaluation

- assume conditional distribution ith predictor parameterized by $\theta_{i,k} \in \Theta$

$$p_{i,k}(y(t_k)|x_{t_k}, x_{t_{k-1}}, \ldots, y(t_{k-1}), y(t_{k-2}), \ldots) = p_{i,k}(y(t_k); x_{t_k}, \theta_{i,k})$$

  - depends on prior & current input, *i.e.*, $\theta_{i,k}$ & x_{t_k}
- update $\theta_{i,k+1}$ from $\theta_{i,k}$ after observing true $y(t_k)$ using Bayesian rule

$$p(w; \theta_{i,k+1}) := p(w|y(t_k); x_{t_k}, \theta_{i,k}) = \frac{p(y(t_k)|w, x_{t_k})p(w; \theta_{i,k})}{\int p(y(t_k)|w, x_{t_k})p(w; \theta_{i,k})dw}$$



Virtual Metrology

VM

- background
  - every process engineer wants to (so badly) measure every material processed - make sure process done as desired
    - $e.g.$, in semiconductor manufacturing, photolithography engineer wants to make sure diameter of holes or line spacing on wafers done correctly to satisfy specification for GPU or memory chips
  - however, various constraints prevent them from doing it, $e.g.$, in semiconductor manufacturing
    - measurement equipment requires investment
    - incur intolerable throughput
    - fab space does not allow
- GOAL - *measure every processed material without physically measuring them*

VM - problem formulation

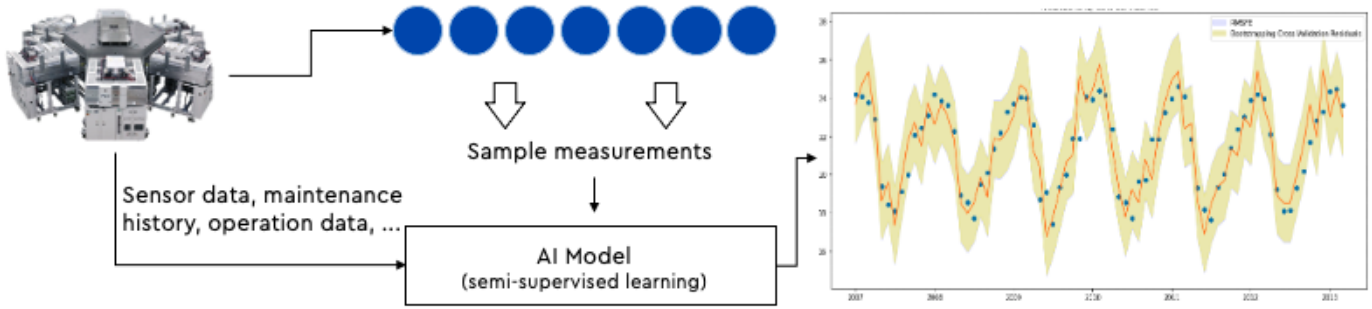
- problem description

$$\begin{array}{rl} \text{(stochastically) predict} & y_{t_k} \\ \text{given} & x_{t_k}, x_{t_{k-1}}, \ldots, y_{t_{k-1}}, y_{t_{k-2}}, \ldots \end{array}$$

- our problem formulation

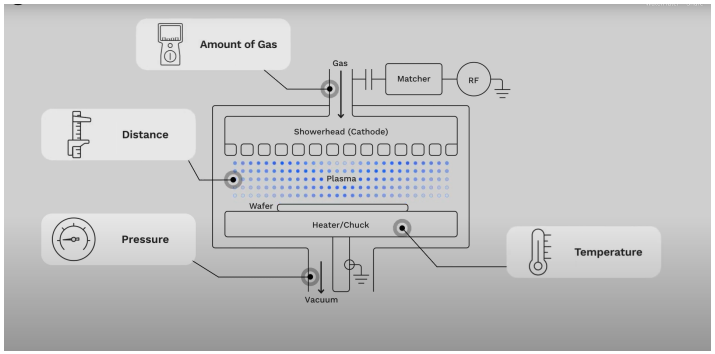
$$\begin{array}{rl} \text{minimize} & \sum_{k=1}^{K} w_{k,K-k}\, l(y_{t_k}, \hat{y}_{t_k}) \\ \text{subject to} & \hat{y}_{t_k} = g_k(x_{t_k}, x_{t_{k-1}}, \ldots, y_{t_{k-1}}, y_{t_{k-2}}, \ldots) \end{array}$$

where optimization variables - $g_1, g_2, \ldots : \mathcal{D} \to \mathbf{R}^m$



VM - Gauss Labs' inAI success story

- Gauss Labs' ML solution & AI product
  - fully home-grown online TS adpative ensemble learning method
  - outperform competitors and customer inhouse tools, *e.g.*, *Samsung, Intel, Lam Research*
  - published & patented in US, Europe, and Korea
- business impacts
  - improve process quality - reduction of process variation by tens of percents
  - (indirectly) contribute to better product quality and yield
  - Gauss Labs' main revenue source



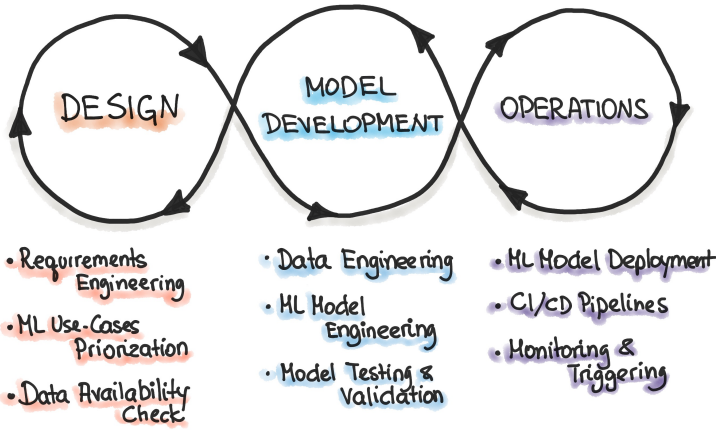
Manufacturing AI Productionization

Minimally required efforts for manAI

- MLOps - for CI/CD
- data preprocessing - missing values, inconsistent names, difference among different systems
- feature extraction & selection
- monitoring & retraining
- notification, via messengers or emails
- mainline merge approvals by humans
- data latency, data reliability, & data availability

MLOps for manAI

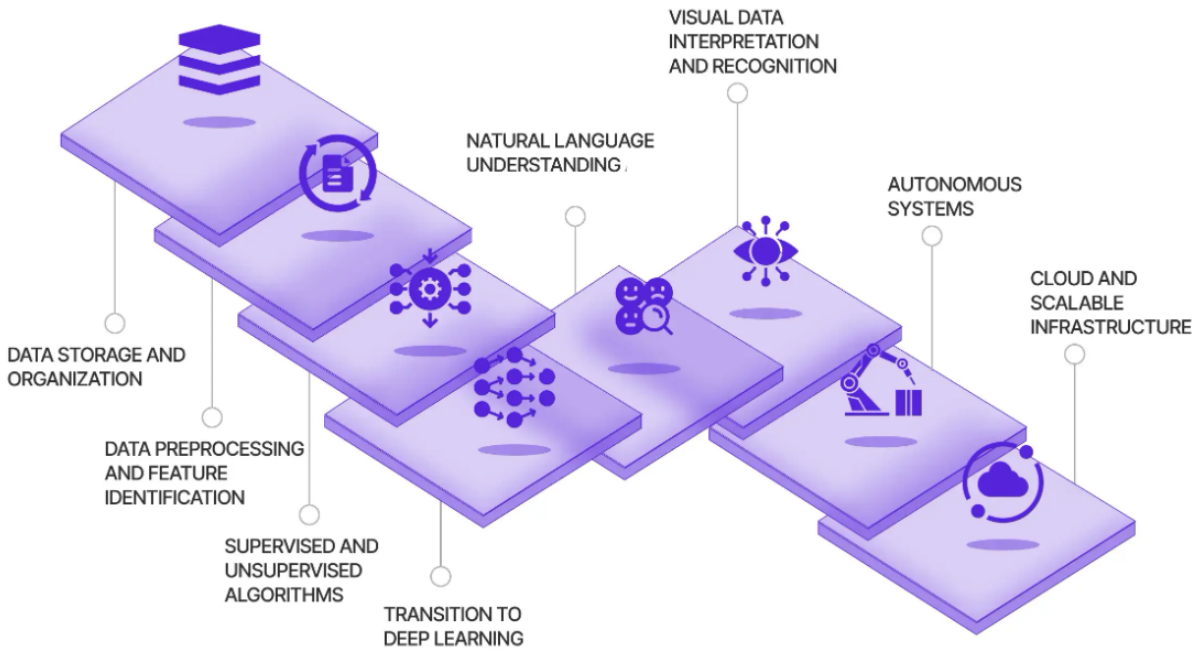
- environment for flexible and agile exploration - EDA[3]
- fast & efficient iteration of algorithm selection, experiements, & analysis
- correct training / validation / test data sets critical!
- seamless productionization from, $e.g.$, Jupyter notebook to production-ready code
- monitorning, *right* metrics, notification, re-training



[3] EDA - exploratory data analysis

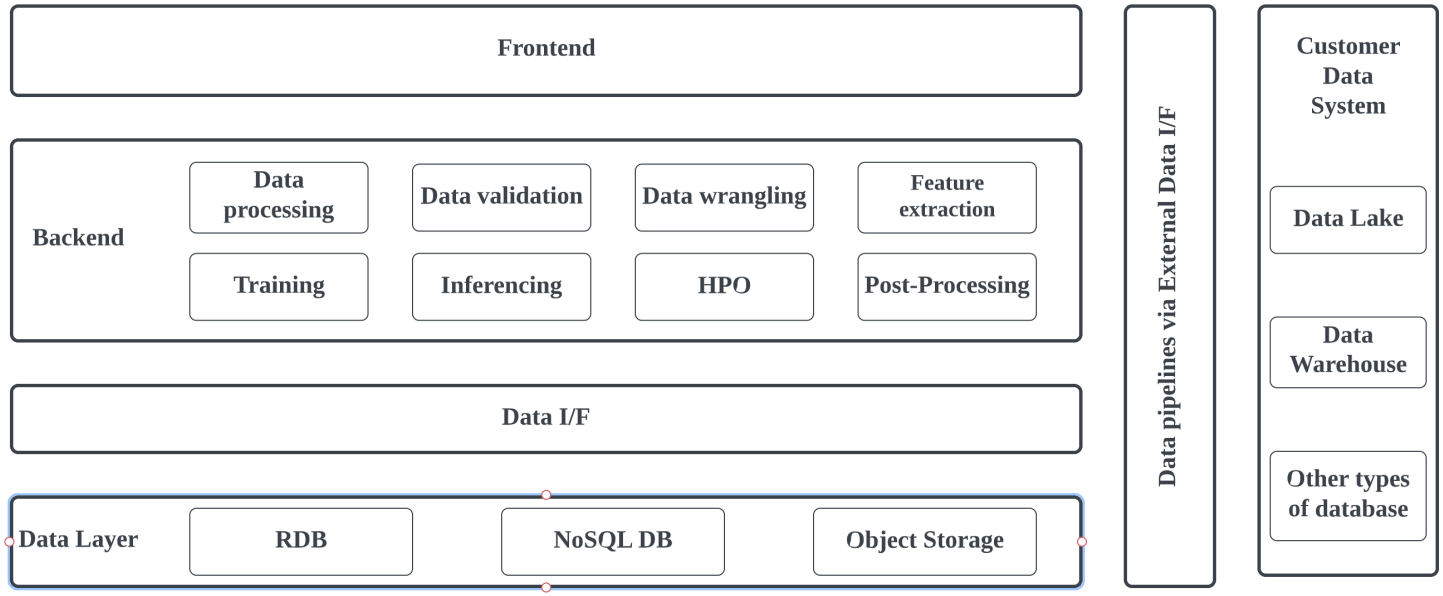
manAI software system

- data, data, data! – store, persist, retrieve, data quality
- seamless pipeline for development, testing, running deployed services
- development envinroment should be built separately



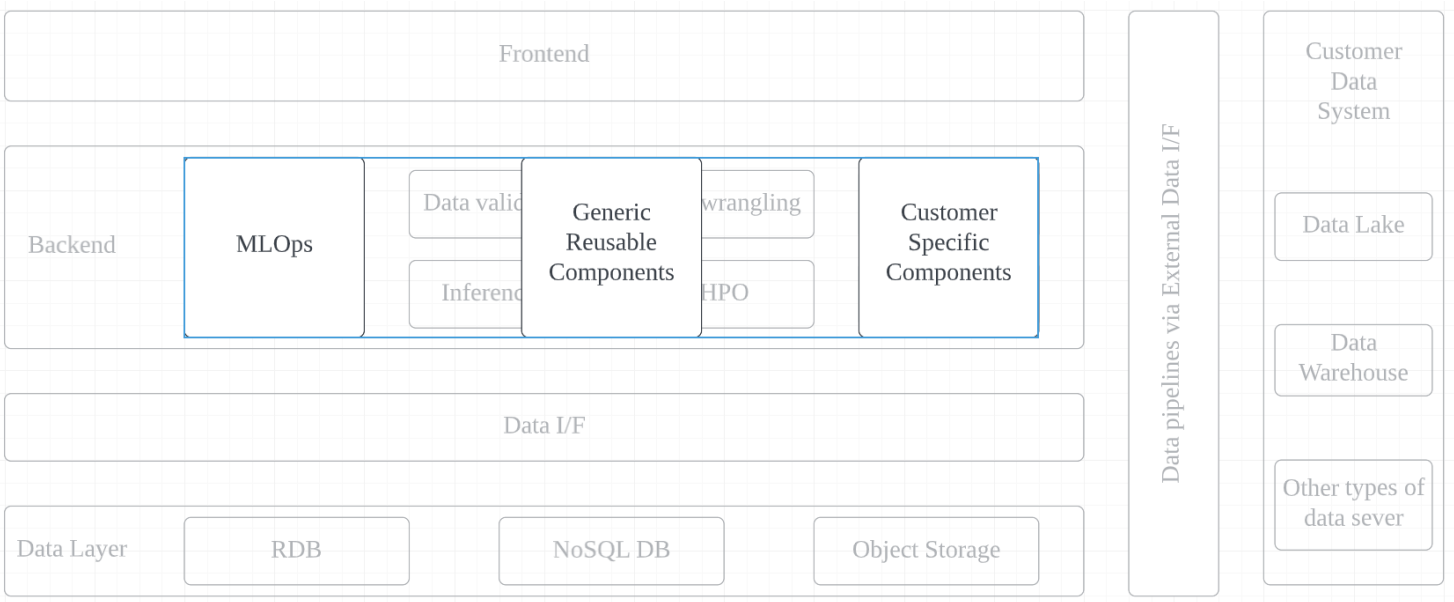
manAI system architecture

- frontend / backend / data I/F / data layer
- efficient and effective MLOps in backend or development environment



Reusuable components vs customer specific components

- make sure to build two components separate - generic reusable and customer specific
- generic models should be tuned for each use case
- generic model library grows as interacting with more and more customers



My Two Cents

Recommendations for maximum impact via inAI

- concrete goals of projects
  - north star – yield improvement, process quality, making engineers' lives easier
  - hard problem – scheduling and optimization
- be strategic!
  - learn from others – lots of successes & failures of inAI
  - ball park estimation for ROI cricial – efforts, time, expertise, data
  - utilities vs technical excellency / uniqueness vs common technology
  - home-grown vs off-the-shelf

Remember . . .

- data, data, data! – readiness, quality, procurement, pre-processing, DB
- *never* underestimate domain knowledge & expertise – data do NOT tell you everything
- EDA
- do *not* over-optimize your algorithms – ML is all about trials-&-errors
- overfitting, generalization, concept drift/shift - way more important than you could ever imagine
- devOps, MLOps, agile dev, software development & engineering

Conclusion

Conclusion

- various CV MLs used for inAI applications
- TS ML applications found in every place in manufacturing
- drift/shift & data noise make TS MLs very challenging, but working solutions found
- in reality, crucial bottlenecks are
  - data quality, prepocessing, monitoring, notification, and retraining
  - data latency, avaiability, and reliability
  - excellency in software platform design and development using cloud services

Important Questions to be Asked

Some important questions around AI

- why human-level AI in the first place?
- what lies in very core of DL architecture? what makes it work amazingly well?
- biases that can hurt judgement, decision making, social good?
- ethical and legal issues
- consciousness, knowledge, belief, reasoning
- future of AI

Human-level AI?

Why human-level in the first place?

- lots of times, when we measure AI performance, we say
  - how can we achieve human-level performance, $e.g.$, CV models?
- why human-level?
  - are all human traits desirable? are humans flawless?
  - aren't humans still evolving?
- advantage of AI over humans
  - $e.g.$, self-driving cars can use extra eyes, GPS, computer network
  - $e.g.$, recommendation system runs for hundreds of millions of people overnight
  - AI is available 24 / 7 while humans cannot
    . . . critical advanages for medical assitance, emergency handling
  - AI does not make more mistakes because task is repetative and tedius
  - AI does not request salary raise or go on strike

What makes DL so successful?

Factors constributing to astonishing success of DL

- analysis based on speaker's mathematical, numerical algorithmic & statistical perspectives considering hardware innovations

  30% universal approximation theorem? - (partially) yes! but that's not all

  - function space of neural network is *dense* (math theory), $i.e.$, for every $f : \mathbf{R}^n \to \mathbf{R}^m$, exists $\langle f_n \rangle$ such that $\lim_{n\to\infty} f_n = f$

  25% architectures/algorithms tailored for each class of applications, $e.g.$, CNN, RNN, Transformer, NeRF, diffusion, GAN, VAE, . . .

  20% data labeling - expensive, data availability - unlimited web text corpus

  15% computation power/parallelism - AI accelerators, $e.g.$, GPU, TPU & NPU

  10% rest - Python, open source software, cloud computing, MLOps, . . .

Why do we see sudden leap in LLM performance?

Probability inferenced sequence is correct

- assume
  - t_i - ith token
  - p_i - probability that t_i is correct
  - ρ_i - correlation coefficient between t_{i-1} & t_i
  - $\tilde{p}_k$ - probability that $(t_1, \ldots, t_k)$ are correct
- recursion

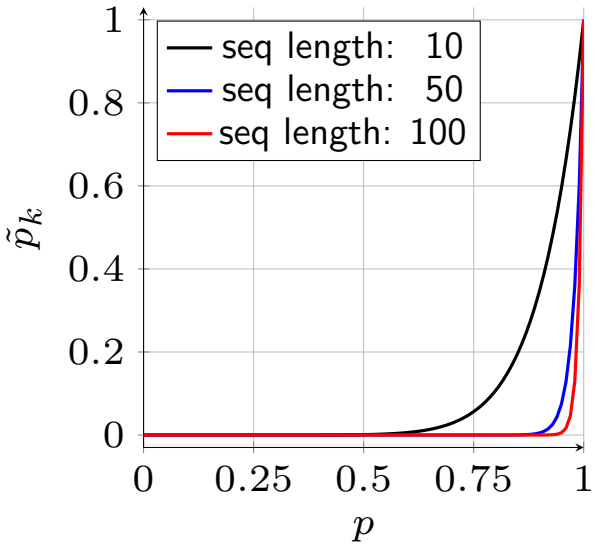
$$\rho_i = \frac{\tilde{p}_i - \tilde{p}_{i-1} p_i}{\sqrt{\tilde{p}_{i-1}(1 - \tilde{p}_{i-1}) p_i (1 - p_i)}}$$

$$\Leftrightarrow \qquad \tilde{p}_i = \tilde{p}_{i-1} p_i + \rho_i \sqrt{\tilde{p}_{i-1}(1 - \tilde{p}_{i-1}) p_i (1 - p_i)}$$

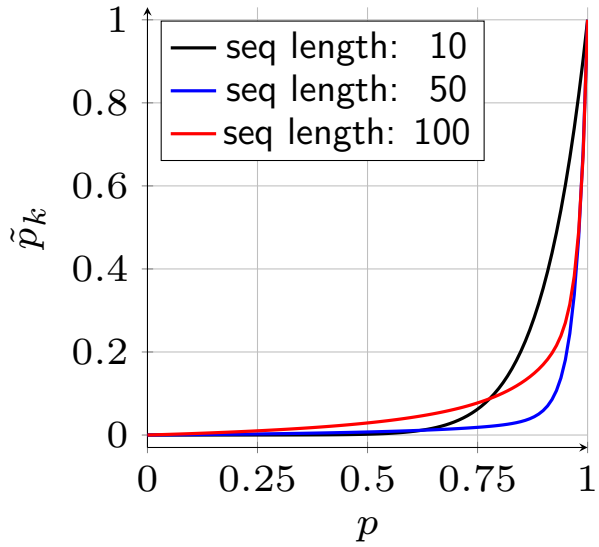
Dramatic improvement of LLM near saturation

- do simulations for both independent & dependent cases
  - assume p_i are same for all i
- (for both cases) sequence inference improves dramatically as p approaches 1
- this explains *why we have observed sudden dramatic performance improvement of certain seq2seq learning technologies*, $e.g.$, LLM

when tokens are independent



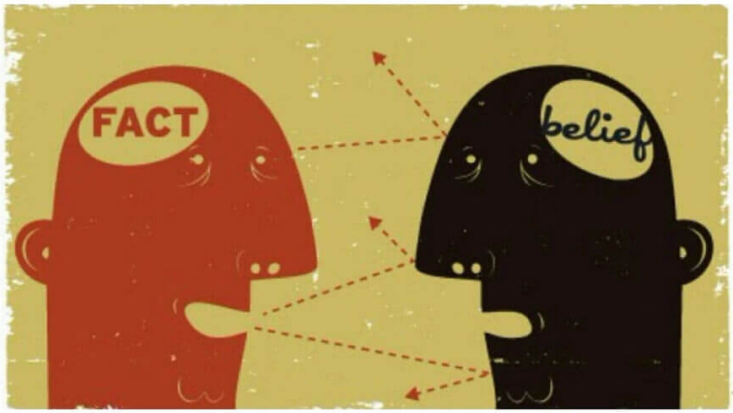
when tokens are dependent



Biases - by Humans & Machines

Cognitive biases

- cognitive biases [Kah11]
  - confirmation bias, availability bias
  - hindsight bias, confidence bias, optimistic bias
  - anchoring bias, halo effect, framing effect, outcome bias
  - belief bias, negativity bias, false consensus,



LLM biases

- plausible with LLM
  - availability bias - baised by imbalancedly available information
    - LLM trained by imbalanced # articles for specific topics
  - belief bias - derive conclusion not by reasoning, but by what it saw
    - LLM eaisly inferencing what it saw, $i.e.$, data it trained on
  - halo effect - overemphasize on what prestigious figures say
    - LLM trained by imbalanced # reports about prestigious figures
- similar facts true for other types of ML models,
  - $e.g.$, video caption, text summarization, sentiment analysis
- cognitive biases only human represent
  - confirmation bias, hindsight bias, confidence bias, optimistic bias, anchoring bias, negativity bias, framing effect

Ethical and Legal Issues

Ethics - possibilities & questions

- AI can be exploited by those who have bad intention to
  - manupilate / deceive people - using manupilated data corpus for training
    - $e.g.$, spread false facts
  - induce unfair social resource allocation
    - $e.g.$, medical insurance, taxation
  - exploit advantageous social and economic power
    - $e.g.$, unfair wealth allocation, mislead public opinion
- AI for Good - advocated by Andrew Ng
  - $e.g.$, public health, climate change, disaster management
- should scientists and engineers be morally & politically conscious?
  - $e.g.$, Manhattan project

Ethically controversial issues

- AI girlfriends
  - lots of AI girlfriend apps already developed
  - ethical considerations and provisions for user privacy with AI partners imperative - as with every technology involving personal data and emotional interaction
  - prospect of developing lifelike digital companions will grow better with evolution of AI
  - perhaps changing ways relationships and companionship perceived in digital age one day
  - why not many AI boyfriend apps? is this sexual discrimination issue (at all)?

Legal issues with ethical consideration - (hypothetical) scenarios

- scenario 1: full self-driving algorithm causes traffic accident killing people
  - who is responsible? - car maker, algorithm developer, driver, algorithm itself?
- scenario 2: self-driving cars kill less people than human drivers
  - $e.g.$, human drivers kill 1.5 people for 100,000 miles & self-driving cars kill 0.2 people for 100,000 miles
  - how should law makers make regulations?
  - utilitarian & humanistic perspectives
- scenario 3: someone is not happy with their data being used for training
  - "The Times sues OpenAI and Microsoft over AI use of copyrighted work" (Dec. 2023)
  - "Newspaper publishers in California, Colorado, Illinois, Florida, Minnesota and New York said Microsoft and OpenAI used millions of articles without payment or permission to develop ChatGPT and other products" (Apr. 2024)

Consciousness

Consciousness

- what is consciousness, anyway?
  - recognizes itself as independent, autonomous, valuable entity?
  - recognizes itself as living being, unchangeable entity?
- no agreed definition on consciousness exists yet
  . . . and will be so forever
- does it have anything to do with the fact that humans are biologically living being?
  - *I don't think so . . .*
- is SKYNET ever plausible (without someone's intention)?
  - can AI have *desire* to survive (or save earth)?

Utopia or dystopia

- not important questions (at all) *I think . . .*
- what we should focus on is not the possibilities of doomday or Judgment Day, but rather
  - our limits on controlling unintended impacts of AI
  - misuse by (greedy and bad) people possessing social, economic & political power
  - social good and welfair impaired by (exploiting) AI
  - choice among utilitarianism, humanism, justice & equity
  - amend or improve laws and regulations
  - addressing ethical issues caused by AI

Knowledge, Belief, and Reasoning of AI

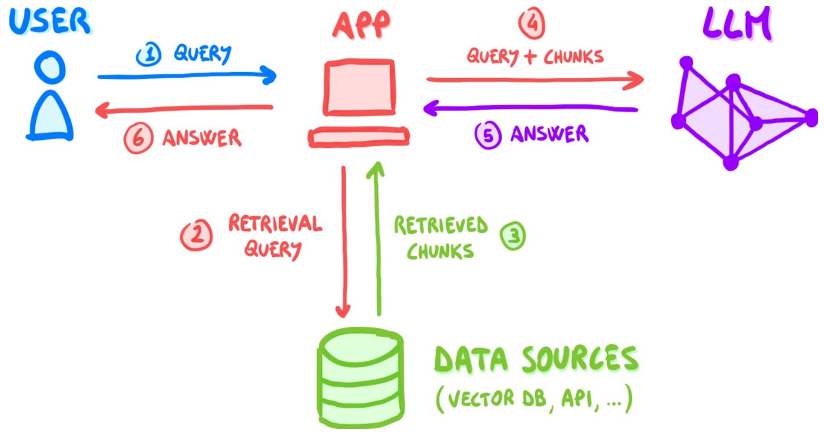
Does LLM (or AI) have knowledge or belief? Can it reason?

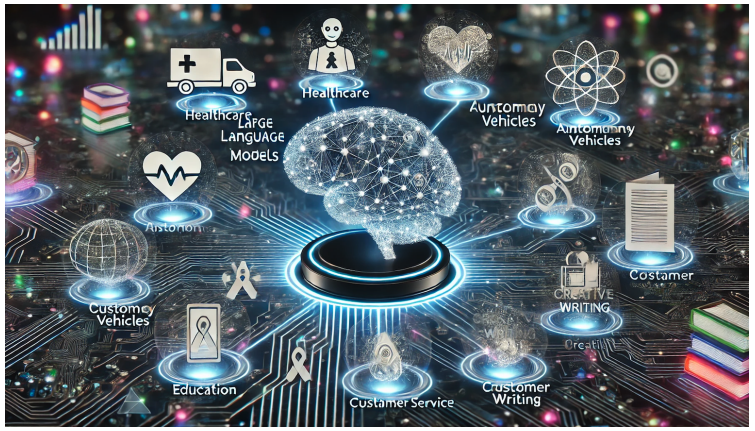
What categories of questions should they be in?

Scientific, philosophical, cognitive scientific, or something else?

LLMs or systems in which they are embedded?

- crucial to distinguish between the two (for philosophical clarity)
  - LLM (bare-bones model) - highly specific & well-defined function, which is *conditional probability estimator*
  - systems in which LLMs are embedded - question-answering, news article summarization, screenplays generation, language translation





Three surprises of LLM

- LLM is very different sort of animal . . . except that it is *not* an animal!
- *unreasonable* effectiveness of data [HNF09]
  - *performance scales with size of training data*
  - *qualitative leaps* in capability as models scale
  - tasks demanding human intelligence *reduced to next token prediction*
- focus on third surprise

  "conditional probability model looks like human with intelligence"

  - making vulnerable to anthropomorphism
- examine it by throwing questions
  - *"does LLM have knowledge and belief?"*
  - *"can it reason?"*

What LLM really does!

- given prompt "the first person to walk on the Moon was", LLM responds with "Neil Armstrong". . . strictly speaking
  - it's *not* being asked *who* was the first person to walk on the Moon
  - what are being *really* asked is *"given statistical distribution of words in vast public corpus of text, what words are most likely to follow 'The first person to walk on the Moon was'?"*
- given prompt "after ring was destroyed, Frodo Baggins returned to", LLM responds with "the Shire"
  - on one level, it seems fair to say, you might be testing LLM's knowledge of fictional world of Tolkien's novels
  - what are being *really* asked is *"given statistical distribution of words in vast public corpus of text, what words are most likely to follow 'After the ring was destroyed, Frodo Baggins returned to'?"*

Knowledge, belief & reasoning around LLM

- *not* easy topic to discuss, or even impossible because
  - we do *not* have agreed definition of these terms especially in context of being asked questions like

    does LLM have belief?

    or

    do humans have knowledge?

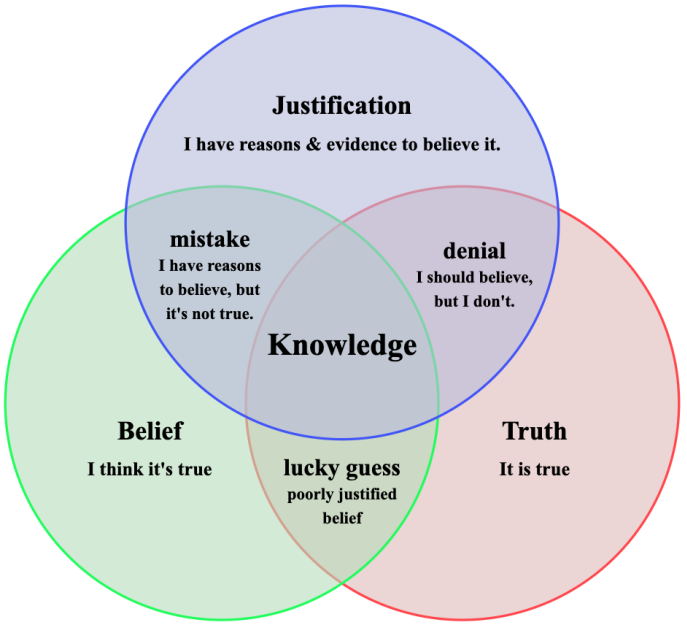
- let us discuss them in two different perspectives
  - laymen's perspective
  - cognitive scientific perspective

Laymen's perspective on knowledge, belief & reasoning

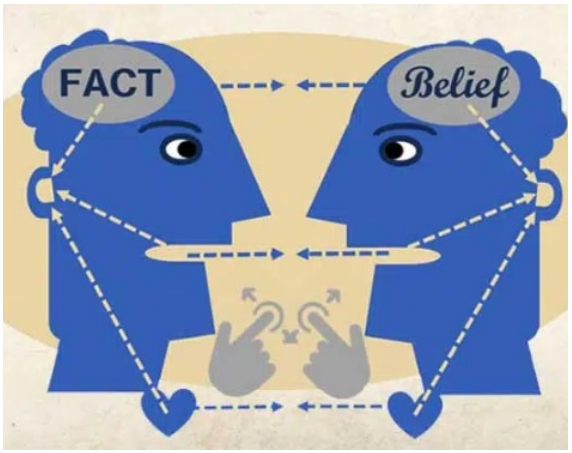
- does (good) LLM have knowledge?
  - Grandmother - looks like it cuz when instructed *"explaing big bang"*, it says
    " The Big Bang theory is prevailing cosmological model that explains the origin and evolution of the universe. . . . 13.8 billion years ago . . . "
- does it have belief?
  - Grandmother: I don't think so, $e.g.$, it does not believe in God.
- can it reason?
  - Grandmother: seems like it! $e.g.$, when asked *"Sunghee is a superset of Alice and Beth is a superset of Sunghee. is Beth a superset of Alice?"*, it says
    " Yes, based on information provided, if Sunghee is a superset of Alice and Beth is a superset of Sunghee, then Beth is indeed a superset of Alice . . . "
- can it reason to prove theorem whose inferential structure is more complicated?
  - Grandmother: I'm not sure. - actually, I don't know what you're talking about!

Cognitive scientific perspective on knowledge

- does LLM have knowledge?
  - Sunghee: I don't think so.
- why?
  - Sunghee: we say we have "knowledge" when
    "we do so against ground of various human capacities that we all take for granted when we engage in everyday conversation with each other."
    — LLM *cannot* do this.
  - Sunghee: also when asked *"who is Tom Cruise's mother?"*, it says *"Tom Cruise's mother is Mary Lee Pfeiffer."* However, this is nothing but
    "guessing" by conditional probability model the most likely following words after "Tom Cruise's mother is."
  - Sunghee: so *we cannot say it really knows the fact!*



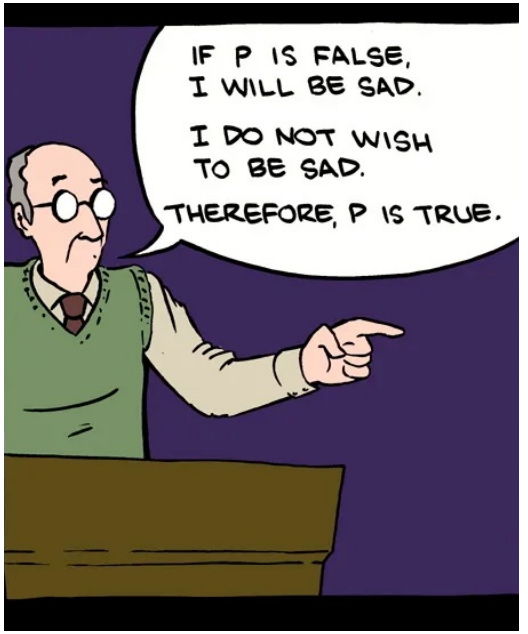
Cognitive scientific perspective on belief



- for the discussion
  - we do not concern *any specific belief*
  - we concern prerequisites for ascribing any beliefs to AI system
- so does it have belief?
  - Sunghee: nothing can count as belief about the world we share unless
    > *it is against ground of the ability to update beliefs appropriately in light of evidence from that world, an essential aspect of the capacity to distinguish truth from falsehood.*
  - Sunghee: when a human being takes to Wikipedia and confirms some fact, what happens is not her language model update, but
    > *reflection of her nature as language-using animal inhabiting shared world with a community of other language-users.*
  - Sunghee: LLM does not have this ground, an essential consideration when deciding whether it *really* had beliefs.
  - Sunghee: so *no, LLM cannot have belief!*

Cognitive scientific perspective on reasoning

- note reasoning is *content neutral*
  - *e.g.*, following logic is perfect regardless of truth of premises
    if Socrates is a human and humen are immortal, then Socrates would have survived today.
- Sunghee: when asked *"if humans are immortal, would Socrates have survived today?"*, LLM says
    " . . . it's logical to conclude that Socrates would likely still be alive today. . . . "
  - however, remember, once again, what we just asked it to do is *not* "deductive inference", but
    given the statistical distribution of words in public corpus, what words are likely to follow the sequence, "humans are immortal and Socreates is human therefore."
- Sunghee: so LLM *cannot* or rather *does not* reason
- however, LLM can *mimic even multi-step reasoning whose inferencing structure is complicated* using *in-context learning* or *few-short prompting*!



A simple example supporting reasoning incapability

- You

  "Who is Tom Cruise's mother?"

- LLM(-embedded question-answering system) (as of Jan 2022)

  "Tom Cruise's mother is Mary Lee Pfeiffer. She was born Mary Lee South. . . . Information about his family, including his parents, has been publicly available, *. . . "*

- You

  "Who is Mary Lee Pfeiffer's son?"

- LLM(-embedded question-answering system) (as of Jan 2022)

  "As of my last knowledge update in January 2022, I don't have specific information about Mary Lee Pfeiffer or her family, including her son. *. . . "*

Future of AI

Aschenbrenner's essay

- Leopold Aschenbrenner, who left OpenAI shwoing concerns about safety, wrote *epic 165-page treatise* - Jun-2024
  - rapid progress
    - AI development (is) accelerating at unprecedented rate, predicting by 2027, AI models lead to intelligence explosion surpassing human intelligence
  - economic and security implications
    - trillions of dollars being invested into infrastructure supporting AI systems
    - critical need for securing technologies to prevent misuse, $e.g.$, by state actors like Chinese Communist Party (CCP)
  - technical and ethical challenges
    - significant challenges in controlling AI (smarter than humans), $i.e.$, "superalignment" problem, to prevent catastrophic outcomes
  - predictions and societal impact
    - few people truly understand scale of change by AI
    - potential for AI to reshape industries, enhance national security
    - pose new ethical and governance challenges

More about Aschenbrenner's essay

- AGI by 2027
  - seen AI advancing from preschool-level to high-schooler abilities in 4 years highlighting rapid progress from GPT-2 to GPT-4
- superintelligence following AGI - post AGI
  - rapid advancement from human-level to superhuman capabilities
- G-dollar investment on AI clusters
- national & global security dynamics
  - may lead to all-out war, $e.g.$, with China, if not managed properly
- superalignment challenges
  - keeping superintelligent AI aligned with human values and interests - "one of the most critical predictions"
- societal and economic transformations, project involvement by US government, technological mobilization

Moral

Moral

- AI, $e.g.$, LLM, shows incredible utility and commercial potentials, hence we should
  - make informed decisions about trustworthiness and safety
  - avoid ascribing capacities they lack

- today's AI is so powerful, so (seemingly) convincingly intelligent
  - obfuscate mechanism
  - actively encourage *anthropomorphism* with philosophically loaded words like "believe" and "think"
  - easily mislead people about character and capabilities of AI

- matters not only to scientists, engineers, developers, and entrepreneurs, but also
  - *general public, policy makers, media people*

Recent AI Development

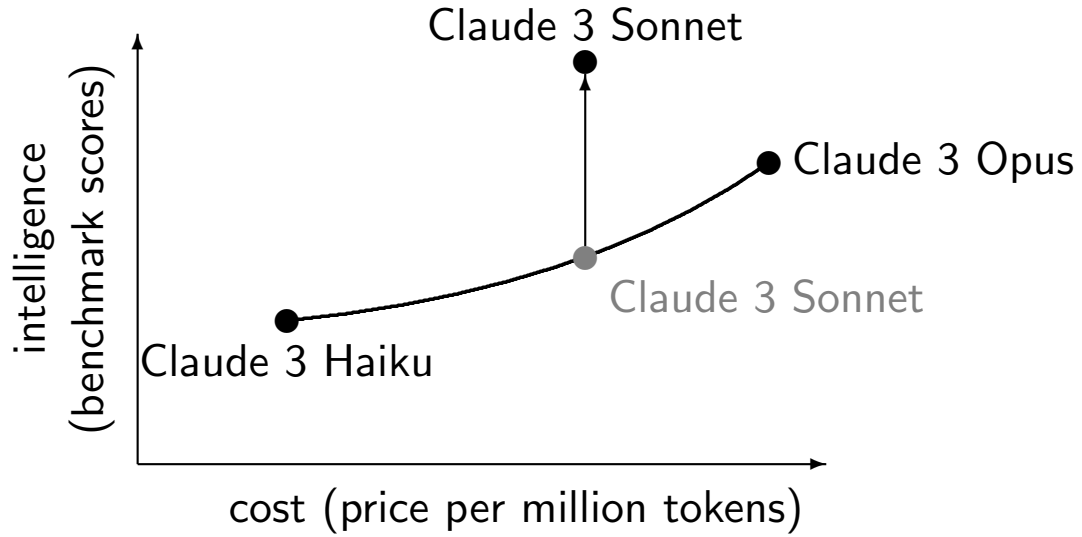
Notable recent AI research and new development

- Claude 3.5 Sonnet
- Kolmogorov–Arnold networks (KAN)
- JEPA ($e.g.$, I-JEPA & V-JEPA) & consistency-diversity-realism trade-off

Claude 3.5 Sonnet

Claude 3.5 Sonnet

- Anthropic
  - releases Claude 3.5 Sonnet (Jul-2024)
    - when! GPT-4o accepted to be default best model for many tasks, $e.g.$, reasoning & summarization
  - claims Claude 3.5 Sonnet sets *new industry standard for intelligence*



Main features & performance

- Claude 3.5 Sonnet shows off
  - improved vision tasks, 2x speed (compared to GPT-4o), artifacts - new UIs for, $e.g.$, code generation & animation
- with GPT-4o, Claude 3.5 Sonnet
  - wins at code generation
  - on par for logical reasoning
  - loses at logical reasoning
  - *wins at generation speed*

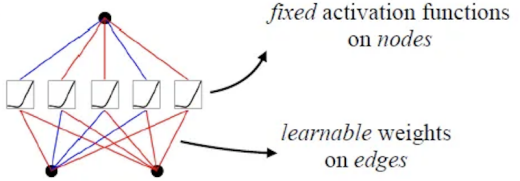
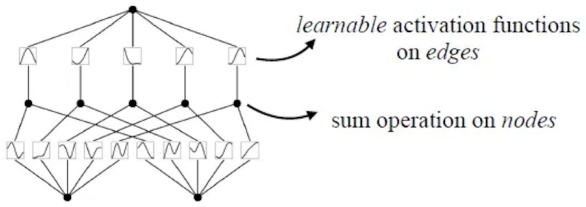
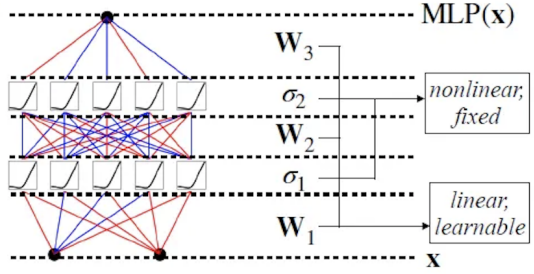
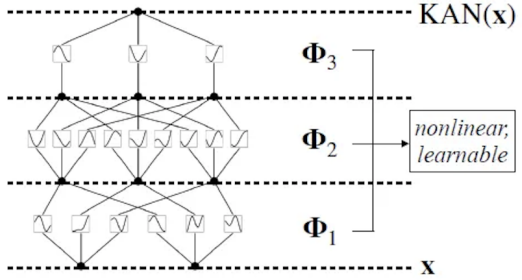
	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro
visual math reasoning	67.7%	50.5%	63.8%	63.9%
science diagrams	94.7%	88.1%	94.2%	94.4%
visual question answering	68.3%	59.4%	69.1%	62.2%
chart Q&A	90.8%	80.8%	85.7%	87.2%
document visual Q&A	95.2%	89.3%	92.8%	93.1%

KAN

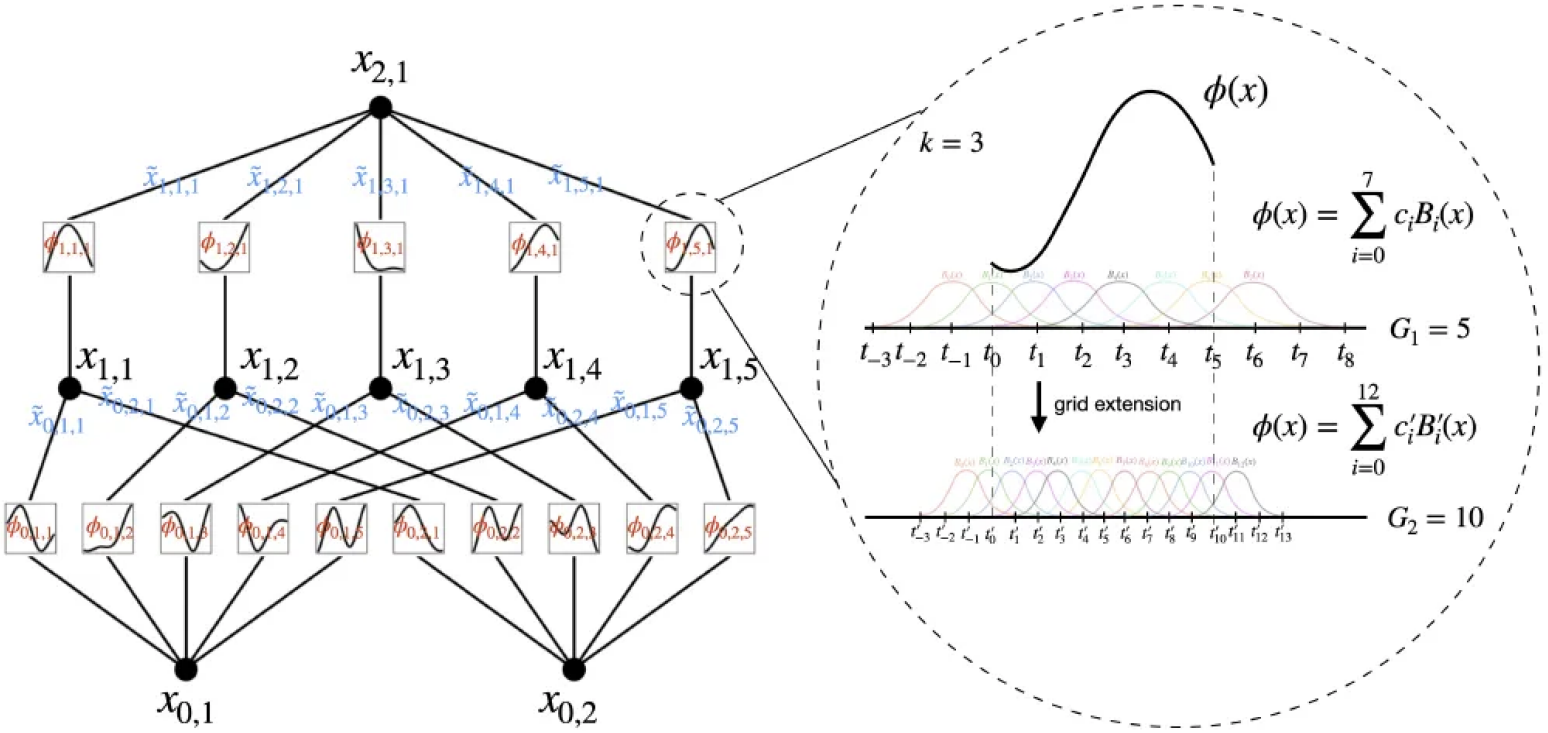
Kolmogorov–Arnold networks (KAN)

- KAN: Kolmogorov-Arnold Networks - MIT, CalTech, Northeastern Univ. & IAIFI
- techniques
  - inspired by Kolmogorov-Arnold representation theorem - every $f : \mathbf{R}^n \to \mathbf{R}$ can be written as finite composition of continuous functions of single variable, $i.e.$
    $$f(x) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$
    where $\phi_{q,p} : [0, 1] \to \mathbf{R}$ & $\Phi_q : \mathbf{R} \to \mathbf{R}$
  - replace (fixed) activation functions with learnable functions
  - use B-splines for learnable (uni-variate) functions - for flexibility & adaptability
- advantages
  - benefits structure of MLP on outside & splines on inside
  - reduce complexity and # parameters to achieve accurate modeling
  - *interpretable* by its nature
  - *better continual learning* - adapt to new data without forgetting thanks to local nature of spline functions

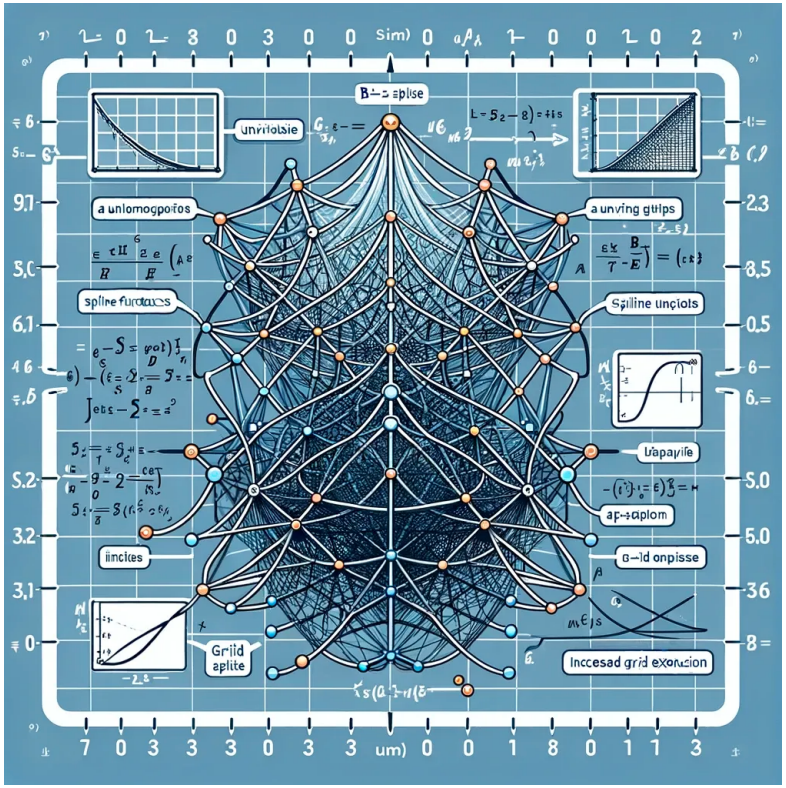
MLP vs KAN

Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

KAN architecture with spline parametrization unit layer



Future work on KAN



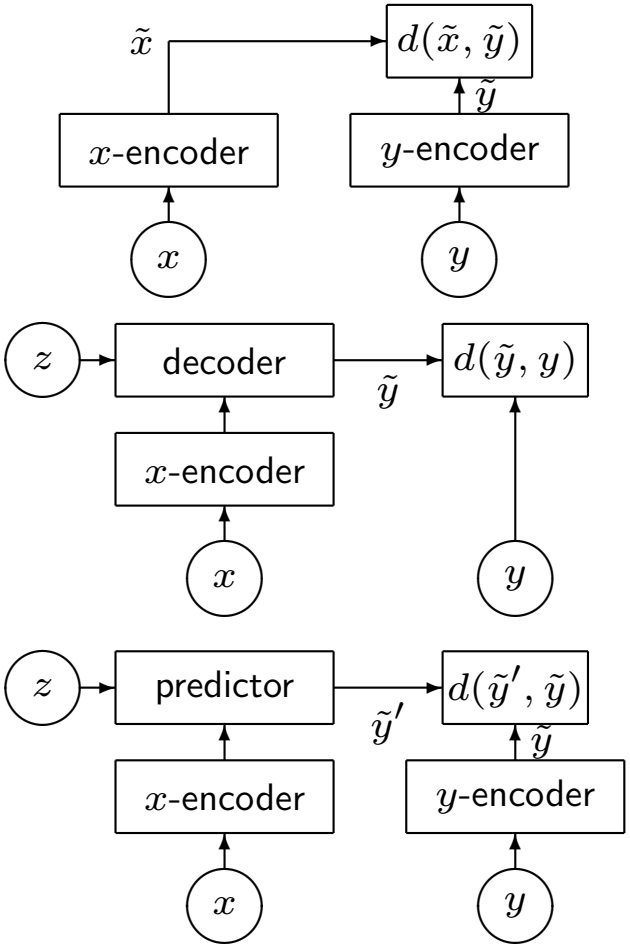
- natural question is
  - what if use both MLP and KAN?
  - what if use other types of splines?
  - how to control forgetfulness of continual learning?
  - why functions of one varible? possible to use functions of two variables?

(figure created by DALLE-3)

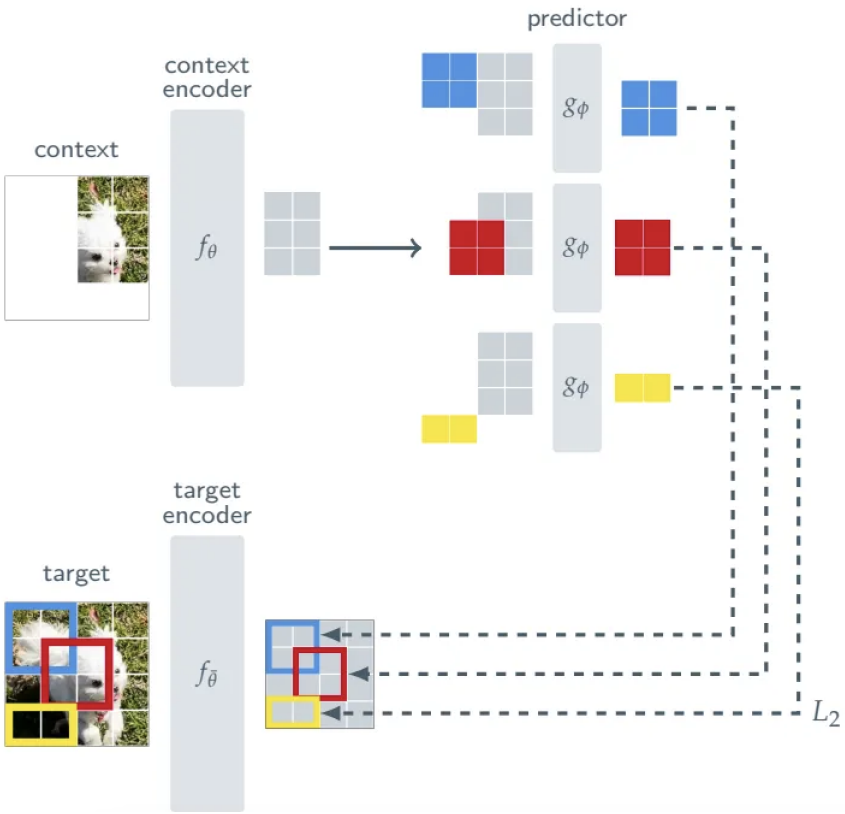
JEPA

Joint-Embedding Predictive Architecture (JEPA)

- Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture (JEPA) - Yann LeCun et al. - Jan-2023
  - joint-embedding architecture (JEA)
    - output similar embeddings for compatible inputs x, y and dissimilar embeddings for incompatible inputs
  - generative architecture
    - directly reconstruct signal y from compatible signal x using decoder network conditioned on additional variables z *to facilitate reconstruction*
  - joint-embedding predictive architecture (JEPA)
    - similar to generative architecture, but comparison is done in embedding space
    - *e.g.*, I-JEPA learns y (masked portion) from x (unmasked portion) conditioned on z (position of mask)



Learning semantic representation better

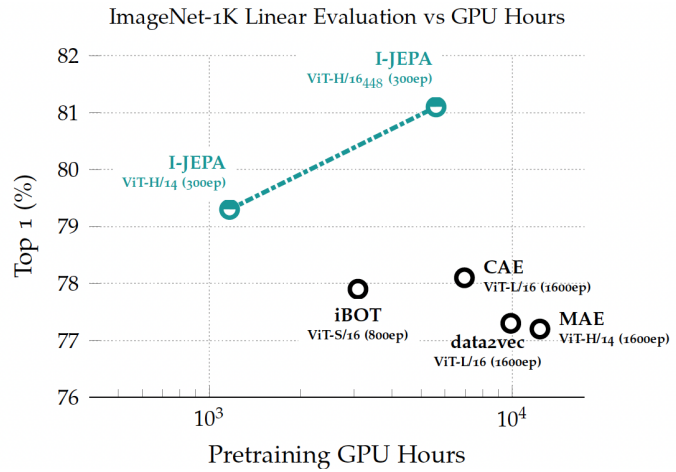


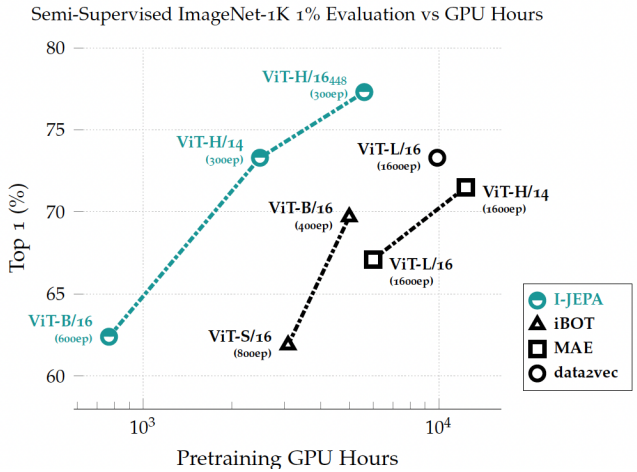
- I-JEPA
  - predicts missing information in *abstract representation space*
    - $e.g.$, given single context block (unmasked part of the image), predict representations of various target blocks (masked regions of same image) where target representations computed by learned target-encoder
  - *generates semantic representations* (not pixel-wise information) potentially eliminating unnecessary pixel-level details & allowing model to concentrate on learning more semantic features

I-JEPA outperforms other algorithms

Method	Arch.	CIFAR100	Places205	iNat18
Methods without view data augmentations				
data2vec [8]	ViT-L/16	81.6	54.6	28.1
MAE [36]	ViT-H/14	77.3	55.0	32.9
I-JEPA	ViT-H/14	**87.5**	**58.4**	**47.6**
Methods using extra view data augmentations				
DINO [18]	ViT-B/8	84.9	57.9	55.9
iBOT [79]	ViT-L/16	**88.3**	**60.4**	**57.3**

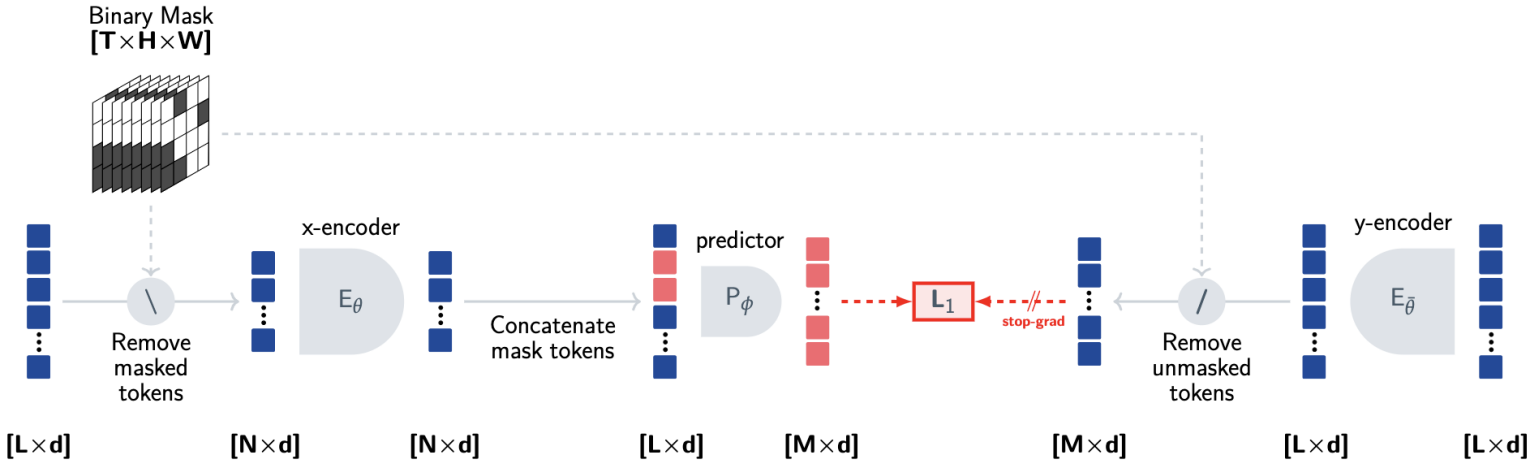
Method	Arch.	Clevr/Count	Clevr/Dist
Methods without view data augmentations			
data2vec [8]	ViT-L/16	85.3	71.3
MAE [36]	ViT-H/14	**90.5**	**72.4**
I-JEPA	ViT-H/14	86.7	**72.4**
Methods using extra data augmentations			
DINO [18]	ViT-B/8	86.6	53.4
iBOT [79]	ViT-L/16	85.7	62.8





V-JEPA

- Revisiting Feature Prediction for Learning Visual Representations from Video - Yann LeCun et al. - Feb-2024
  - essentially same ideas of JEPA - loss function is calculated in embedding space - for better semantic representation learning (rather than pixel-wise learning)



More realistic generative model becomes, less diverse it becomes

- Consistency-diversity-realism Pareto fronts of conditional image generative models - FAIR at Meta - Montreal, Paris & New York City labs, McGill University, Mila, Quebec AI institute, Canada CIFAR AI - Jun-2024

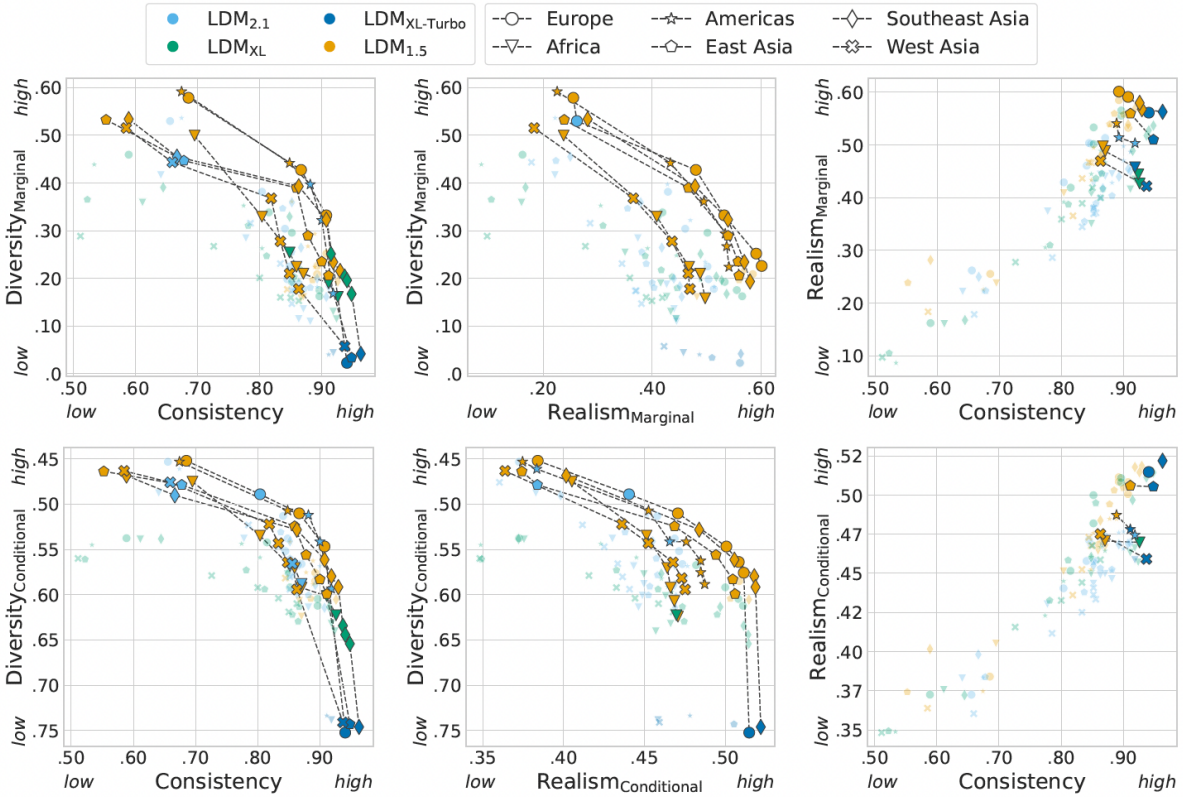
- realism comes at the cost of coverage, $i.e.$, *the most realistic systems are mode-collapsed!*

- intuition (or hunch)

  – world models should *not* be generative - should make predictions in representation space - in representation space, unpredictable or irrelevant information is absent

$\rightarrow$ main argument in favor of JEPA

Consistency-diversity-realism trade-off



Learning ML & AI

Best ways to learn ML & AI

- first, learn basics - college classes, online courses, (easy) books
  - not need to understand every mathematical details, but should know rough ideas!
- hands-on is MUST!
  - learn and practice coding - `Python` is MUST; do not do only `R`
  - learn `git` - know how to develop efficiently, plus `import` others' work
- *(I think) online cources are blessing to mankind!*
  - you *can't* say "I can't do it because resource is not available or classes of good schools are not available" because . . . they are available! :)
  - getting (expensive) certificates is good idea because . . . otherwise you wouldn't finish it! :) plus you can post it on your LinkedIn
- would be best if your task at work is related to ML
- however, even if that's not the case or can't be the case, can always do your own personal projects – or contribute to public projects (on `github`)!

Andrew Ng!

- Andrew Ng
  - (co-)founder of "Deep Learning.AI" and "Coursera", prominent figure in ML & AI
  - his courses highly regarded because well-structured and provide insights

- latest Andrew Ng courses
  - AI Agents in LangGraph
  - AI Agentic Design Patterns with AutoGen
  - Introduction to On-device AI
  - Multi AI Agent Systems with Crew AI
  - Building Multimodal Search and RAG - contrastive learning, multimodality to RAG
  - Building Agentic RAG with LlamaIndex
  - Quantisation In Depth
  - In Prompt Engineering for Vision Models
  - Getting Started with Mistral - open-source models (Mistral 7B, Mixtral 8x7B)
  - Preprocessing Unstructured Data for LLM

Selected References & Sources

Selected references & sources

- Daniel Kahneman, Thinking, Fast and Slow, 2011
- T. Kuiken, Artificial Intelligence in the Biological Sciences: Uses, Safety, Security, and Oversight, 2023
- S. Yin, et. al., A Survey on Multimodal LLMs, 2023
- M. Shanahan, Talking About Large Language Models, 2022
- A. Vaswani, et al., Attention is all you need, NeurIPS, 2017
- I.J. Goodfellow, . . . , Y. Bengio, Generative adversarial networks (GAN), 2014
- A.Y. Halevry, P. Norvig, and F. Pereira. Unreasonable Effectiveness of Data, 2009
- Stanford Vecture Investment Groups
- CEOs & CTOs @ starup companies in Silicon Valley
- VCs on Sand Hill Road - Palo Alto, Menlo Park, Woodside in California

References

References

[ACH+24] Pietro Astolfi, Marlene Careil, Melissa Hall, Oscar Mañas, Matthew Muckley, Jakob Verbeek, Adriana Romero Soriano, and Michal Drozdzal. Consistency-diversity-realism pareto fronts of conditional image generative models, 2024.

[ADM+23] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture, 2023.

[BGP+24] Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video, 2024.

[BKP22] Abhaya Bhardwaj, Shristi Kishore, and Dhananjay K. Pandey. Artificial intelligence in biological sciences. *Life*, 12(1430), 2022.

[DFJ22] Thomas A. Dixon, Paul S. Freemont, and Richard A. Johnson. A global forum on synthetic biology: The need for international engagement. *Nature Communications*, 13(3516), 2022.

[HM24] Guadalupe Hayes-Mota. Emerging trends in AI in biotech. *Forbes*, June 2024.

[HNF09] Alon Halevy, Peter Norvig, and Nanediri Fernando. The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24:8 – 12, 05 2009.

[Kah11] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.

[Kui23] Todd Kuiken. Artificial intelligence in the biological sciences: Uses, safety, security, and oversight. *Congressional Research Service*, Nov 2023.

[LWV+24] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov-arnold networks, 2024.

[RAB+23] Ziaur Rahman, Muhammad Aamir, Jameel Ahmed Bhutto, Zhihua Hu, and Yurong Guan. Innovative dual-stage blind noise reduction in real-world images using multi-scale convolutions and dual attention mechanisms. *Symmetry*, 15(11), 2023.

[Say21] Kelley M. Sayler. Defense primer: Emerging technologies. *Congressional Research Service*, 2021.

[Sha23] Murray Shanahan. Talking about large language models, 2023.

[Toe23] Rob Toews. The next frontier for large language models is biology. *Forbes*, July 2023.

[Wet23] Kris A. Wetterstrand. Dna sequencing costs: Data, 2023.