

[CHEY Science and Technology Innovation Workshop]

The Vertical AI Inflection Point

- From Factory Floors to Cancer Diagnostics where Real AI Transformation is Happening**

Sunghee Yun

Co-Founder & CTO @ Erudio Bio, Inc.

Co-Founder & CEO @ Erudio Bio Korea, Inc.

Co-Founder & Leader & Chair of Silicon Valley AI Nexus

Advisor to Korean American Semiconductor Professional Alliance

KFAS-Salzburg Global Leadership Initiative Fellow

Visiting Professor @ Sogang University

Advisory Professor @ DGIST

About Speaker

- *Co-Founder & CTO @ Erudio Bio, Inc., San Jose & Novato, CA, USA* 2023 ~
- *Co-Founder & CEO @ Erudio Bio Korea, Inc., Korea* 2025 ~
- *Co-Founder, Leader, and Chair of Silicon Valley AI Nexus, USA* 2024 ~
- *CGO / Global Managing Partner @ LULUMEDIC, Seoul, Korea* 2025 ~
- *AI-Korean Medicine Integration Initiative Task Force Member @ The Association of Korean Medicine, Seoul, Korea* 2025 ~
- *Advisor to Korean American Semiconductor Professional Alliance (KASPA)* 2026 ~
- *KFAS-Salzburg Global Leadership Fellow @ Salzburg Global Seminar, Austria* 2024 ~
- *Adjunct Professor, EE Department @ Sogang University, Seoul, Korea* 2020 ~
- *Advisory Professor, EECS Department @ DGIST, Korea* 2020 ~
- Global Advisory Board Member @ Innovative Future Brain-Inspired Intelligence System Semiconductor of Sogang University, Korea 2020 ~
- Technology Consultant @ Gerson Lehrman Group (GLG), NY, USA 2022 ~

- Co-Founder & CTO / Head of Global R&D / Chief Applied Scientist / Senior Fellow @ Gauss Labs, Inc., Palo Alto, CA, USA 2020 ~ 2023
- VP / Fellow @ SK hynix 2020 ~ 2021
- Senior Applied Scientist @ Amazon.com, Inc., Vancouver, BC, Canada 2017 ~ 2020
- Principal Engineer @ Software R&D Center, Samsung Electronics 2016 ~ 2017
- Principal Engineer @ Strategic Marketing & Sales, Samsung Electronics 2015 ~ 2016
- Principal Engineer @ DT Team, DRAM Dev, Samsung Electronics 2012 ~ 2015
- Senior Engineer @ CAE Team, Memory Business, Samsung Electronics 2005 ~ 2012
- PhD - Electrical Engineering @ Stanford University, CA, USA 2001 ~ 2004
- Development Engineer @ Voyan, Santa Clara, CA, USA 2000 ~ 2001
- MS - Electrical Engineering @ Stanford University, CA, USA 1998 ~ 1999
- BS - Electrical & Computer Engineering @ Seoul National University 1994 ~ 1998

Highlight of Career Journey

- BS in Electrical Engineering (EE) @ Seoul National University
- MS & PhD in Electronics Engineering (EE) @ Stanford University
 - *Convex Optimization - Theory, Algorithms & Software*
 - Advisor - *Prof. Stephen P. Boyd*
- Principal Engineer @ Samsung Semiconductor, Inc.
 - *AI & Convex Optimization*
 - collaboration with *DRAM/NAND Design/Manufacturing/Test Teams*
- Senior Applied Scientist @ Amazon.com, Inc.
 - *e-Commerce AIs* - anomaly detection, deep RL, and recommender system
 - *Jeff Bezos's project - drove \$200M* in sales via Amazon Mobile Shopping App
- *Co-Founder & CTO / Global R&D Head & Chief Applied Scientist* @ Gauss Labs, Inc.
- *Co-Founder & CTO* @ Erudio Bio, Inc.
- *Co-Founder & CEO* @ Erudio Bio Korea, Inc.

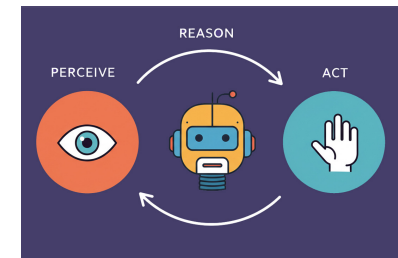
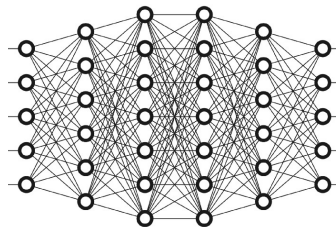
Unpacking AI for CHEY Science and Technology Workshop

- AI Agents - 5
 - LLM as extremely effective knowledge-transfer representation learner
 - Cutting-edge AI agent tools
- Physical AI - 21
 - Computer vision applications and Virtual Metrology
- The Trillion Dollar Opportunities - 60
 - The Moment - Why Now & Why it's Big?
- Appendix - Erudio Bio - 76
 - Versatile Smart Assay (VSA) Platform
 - bioTCAD - When Erudio's VSA meets AI - Gates Foundation Grant
- Appendix - Silicon Valley AI Nexus - 94
- Appendix - Some Important Questions around AI - 105
 - why human-level AI? biases, AI ethics, AI legal issues
- Selected references - 141
- References - 143

AI Agents

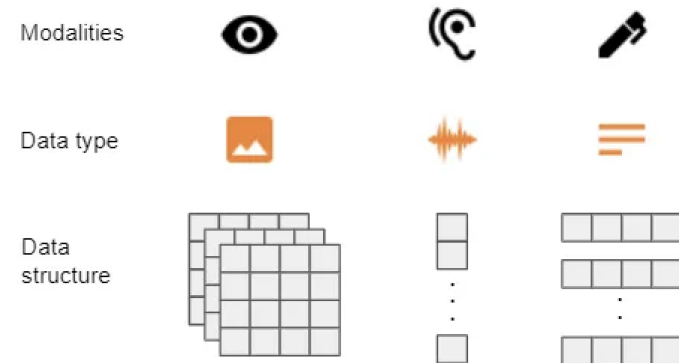
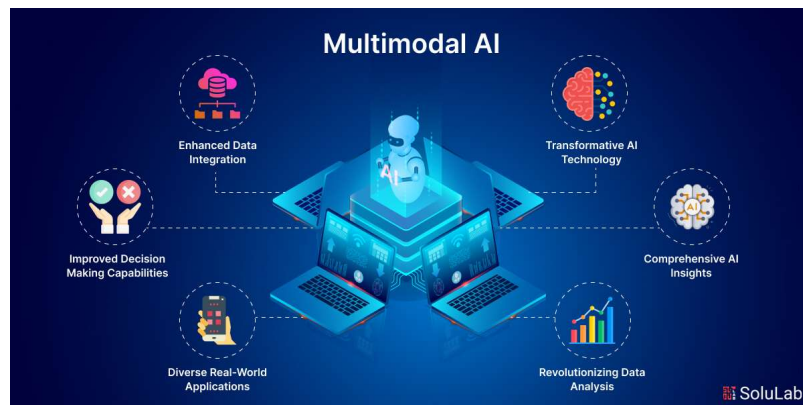
AI progress in 21st century in keywords

- 2010 ~ Big Data
- 2012 ~ Deep Learning
- 2017 ~ Transformer - Attention is All you need!
- 2022 ~ LLM & genAI
- 2024 ~ AI Agent (Agentic AI)



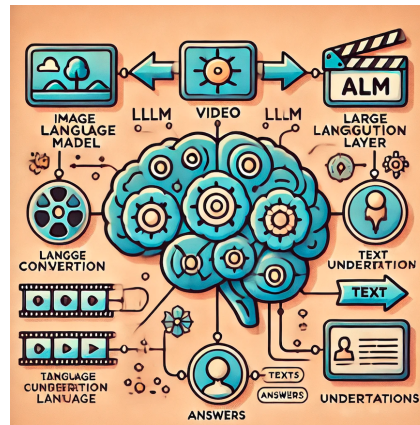
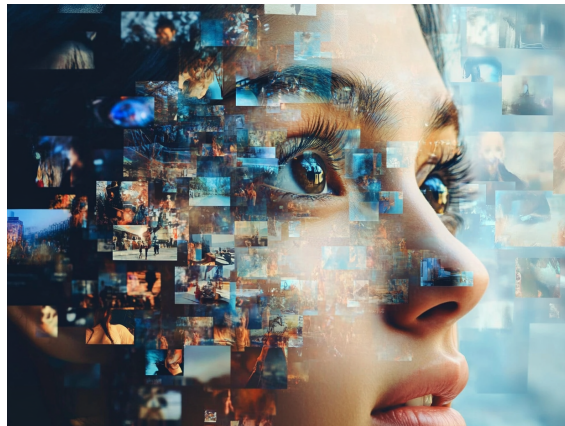
Multimodal learning

- understand information from multiple modalities, *e.g.*, text, images, audio, video
- representation learning methods
 - combine multiple representations or learn multimodal representations simultaneously
- applications
 - images from text prompt, videos with narration, musics with lyrics
- collaboration among different modalities
 - understand image world (open system) using language (closed system)



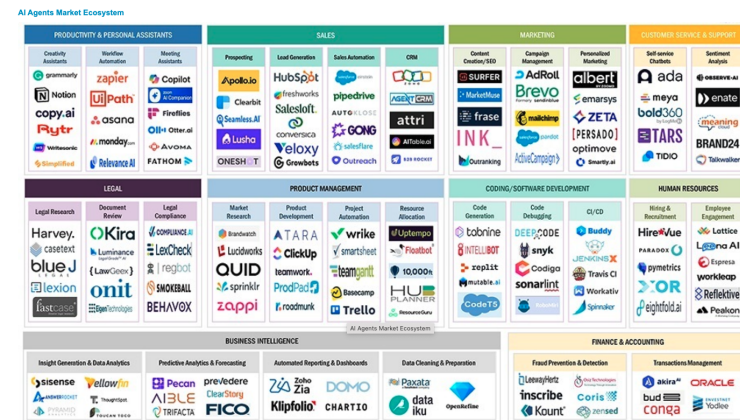
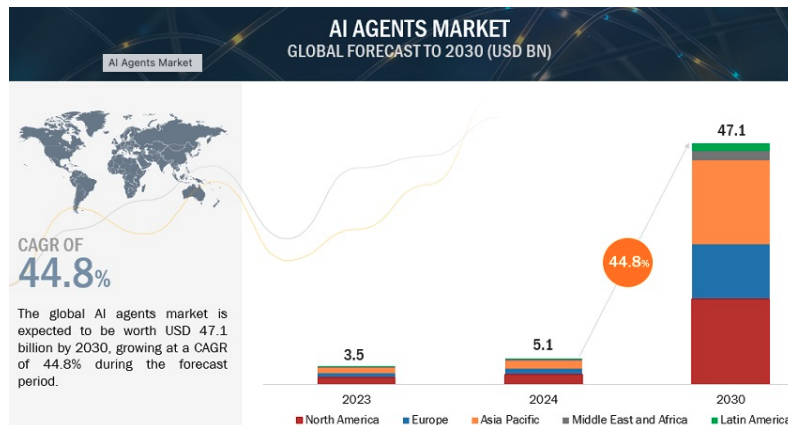
Implications of success of LLMs

- many researchers change gears towards LLM
 - from computer vision (CV), speech, music, video, even reinforcement learning
- *LLM is not only about NLP . . .* humans have . . .
 - evolved to optimize natural language structures for eons
 - handed down knowledge using *this natural languages* for thousands of years
 - internal structure (or equivalently, representation) of natural languages optimized via *thousands of generation by evolution*
- LLM *connects non-linguistic world (open system) via natural languages (closed system)*



Multimodal AI (mmAI)

- mmAI - systems processing & integrating data from multiple sources & modalities, to generate unified response / decision
- 1990s – 2000s - early systems - initial research combining basic text & image data
- 2010s - CNNs & RNNs enabling more sophisticated handling of multimodality
- 2020s - modern multimodal models - Transformer-based architectures handling complex multi-source data at highly advanced level
- mmAI *mimics human cognitive ability* to interpret and integrate information from various sources, leading to holistic decision-making

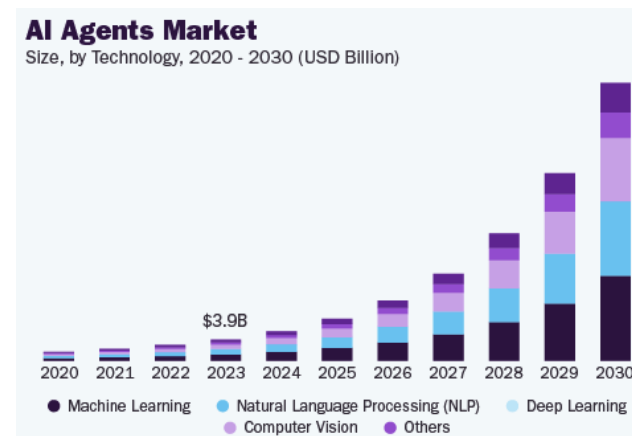
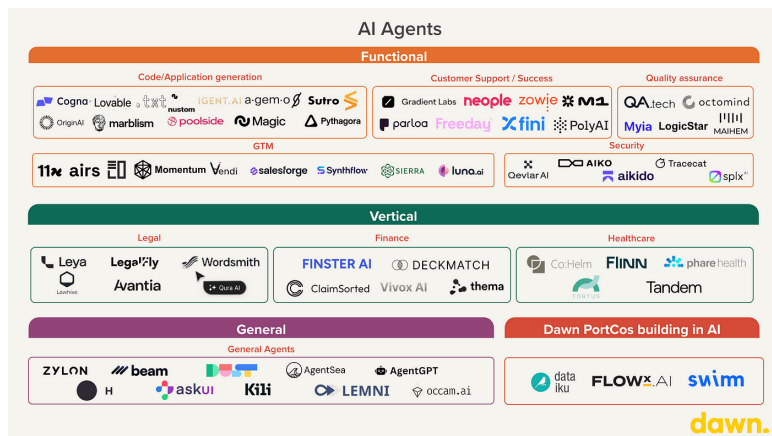


mmAI Technology

- core components
 - data preprocessing - images, text, audio & video
 - architectures - unified Transformer-based (*e.g.*, ViT) & cross-attention mechanisms / hybrid architectures (*e.g.*, CNNs + LLMs)
 - integration layers - fusion methods for combining data representations from different modalities
- technical challenges
 - data alignment - accurate alignment of multimodal data
 - computational demand - high-resource requirements for training and inferencing
 - diverse data quality - manage variations in data quality across modalities
- advancements
 - multimodal embeddings - shared feature spaces interaction between modalities
 - self-supervised learning - leverage unlabeled data to learn representations across modalities

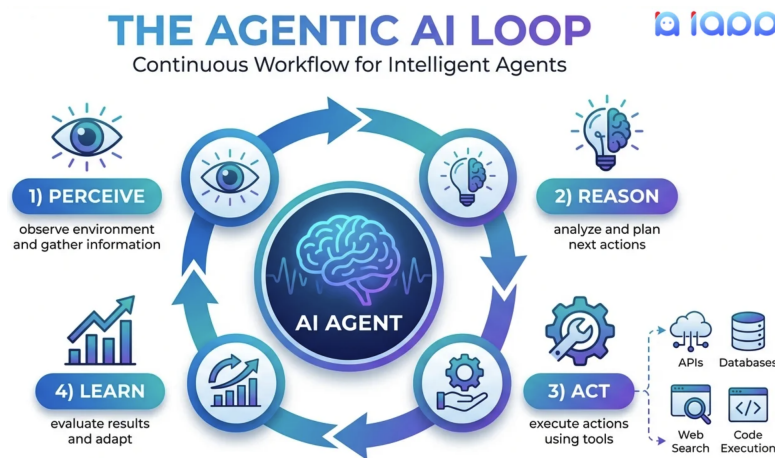
AI agents powered by multimodal LLMs

- foundation
 - integrate multimodal AI capabilities for enhanced interaction & decision-making
- components
 - perceive environment through multiple modalities (visual, audio, text), process using LLM technology, generate contextual responses & take actions
- capabilities
 - understand complex environments, reason across modalities, engage in natural interactions, adapt behavior based on context & feedback



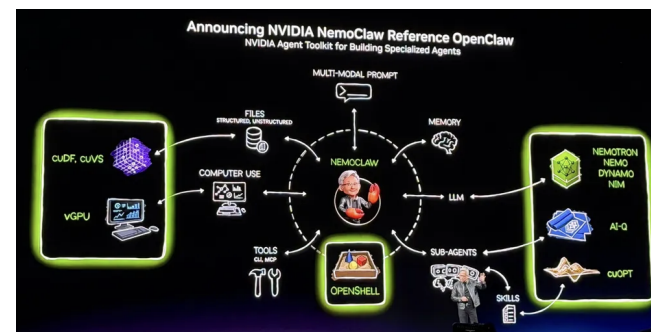
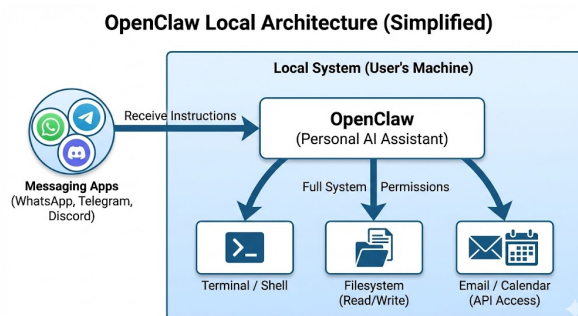
What makes AI “Agentic”?

- old AI responds to prompt; agentic AI *pursues goal*
- core loop
 - perceive → plan → (reason) → act → observe → repeat
- four traits - autonomy, tool use, memory/state, long-horizon planning
- enablers
 - tool/function calling, retrieval, code execution, multi-agent orchestration
- shift - *“answer my question”* → *“accomplish my objective”*



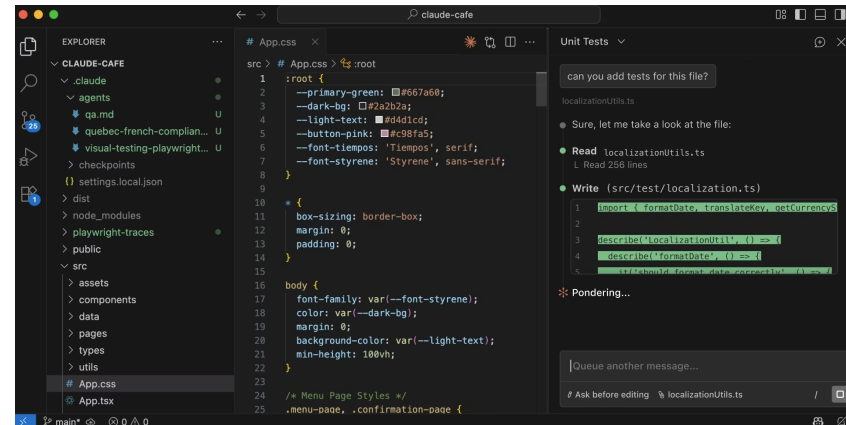
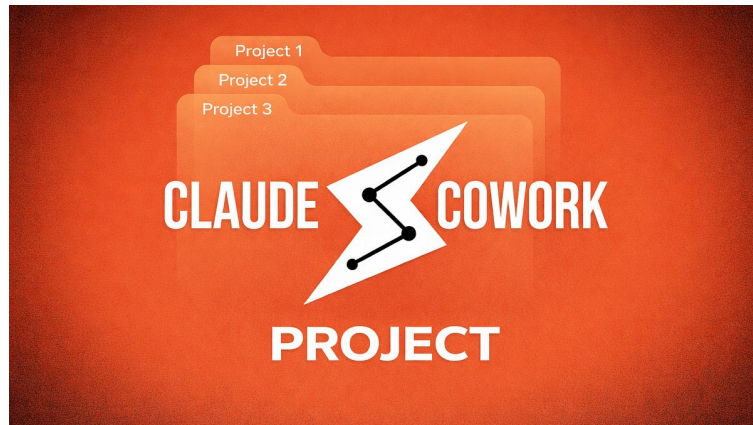
Cutting-edge AI agent tools - open source

- OpenClaw (Peter Steinberger) - open-source, runs locally, connects LLMs to real software
 - reads/writes files, runs shell commands, browses web, sends email, controls APIs
 - 350k+ GitHub stars (by May 2026) — most-starred GitHub software project
 - skill-based architecture - SKILL.md folders, shareable on ClawHub
 - works through chat apps - Slack, Telegram, WhatsApp, Discord, iMessage, *etc.*
 - model-agnostic - Claude, GPT, Gemini, or local via Ollama
- NVIDIA NemoClaw - security/privacy layer on top of OpenClaw
 - one-command install of Nemotron models + OpenShell secure runtime
 - network & filesystem isolation, local inference so no data leaves the device KKR



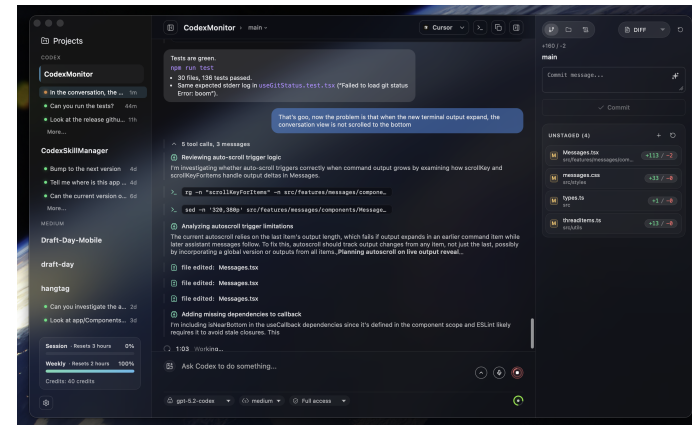
Cutting-edge AI agent tools - Anthropic

- Claude code
 - CLI/IDE coding agent; subagents, hooks, plugins, auto mode, routines
- Claude cowork
 - desktop tab; file-system access, scheduled recurring tasks, plugin marketplace
- managed agents
 - multi-agent orchestration; cloud-deployable agent templates TrendForce
- vertical bundles already shipping
 - legal, small business, marketing ops, finance



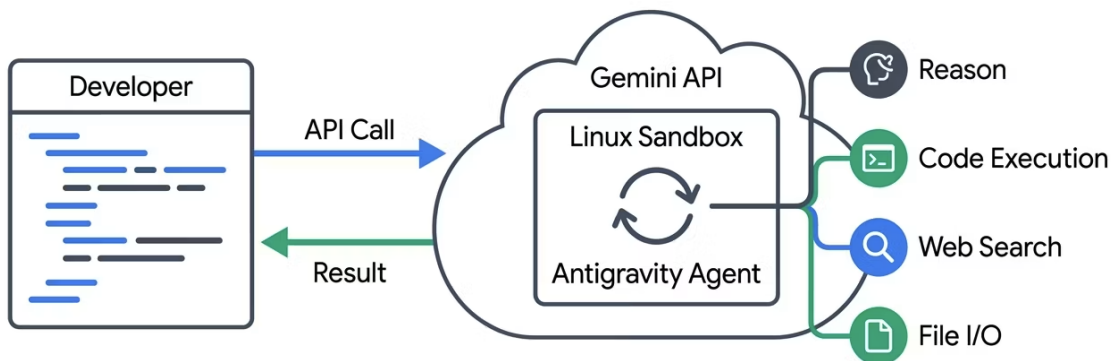
Cutting-edge AI agent tools - OpenAI

- OpenAI Codex
 - agentic coding tool - CLI, IDE, ChatGPT, desktop & now mobile
 - 2026 shift - from code editor → full “agent workspace”
 - multi-agent parallelism - runs several tasks in separate sandboxes while you review
 - powered by GPT-5.5 - tightly coupled, not model-agnostic (unlike Claude Code / OpenClaw) InfoQ
- Codex Security
 - dedicated agent that finds & fixes vulnerabilities



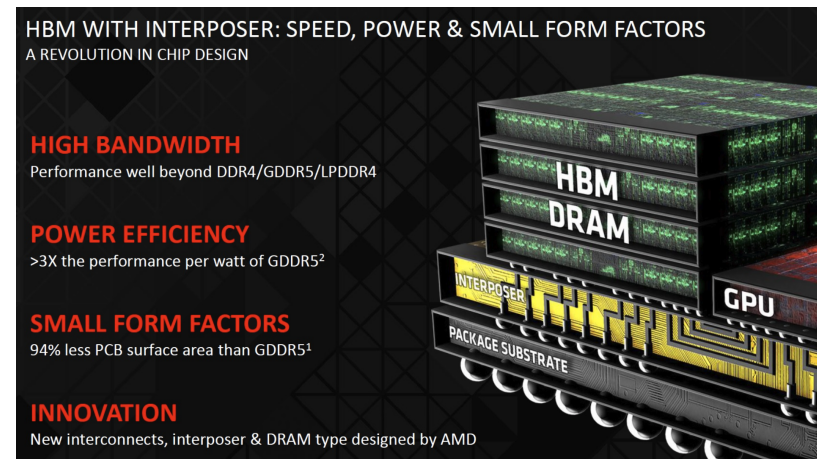
Cutting-edge AI agent tools - Google

- Antigravity 2.0
 - agent-first development platform; desktop app + CLI + SDK NVIDIA
- Gemini API Managed Agents
 - one API call spins up agent that reasons, uses tools, executes code
- Jules
 - AI agent for GitHub - debugging, pull-request prep NVIDIA Newsroom
- Gemini Spark
 - 24/7 personal agent on Gemini 3.5 Flash, wrapped in Antigravity
 - connects to Canva, OpenTable, Instacart, Workspace via MCP NVIDIA Blog



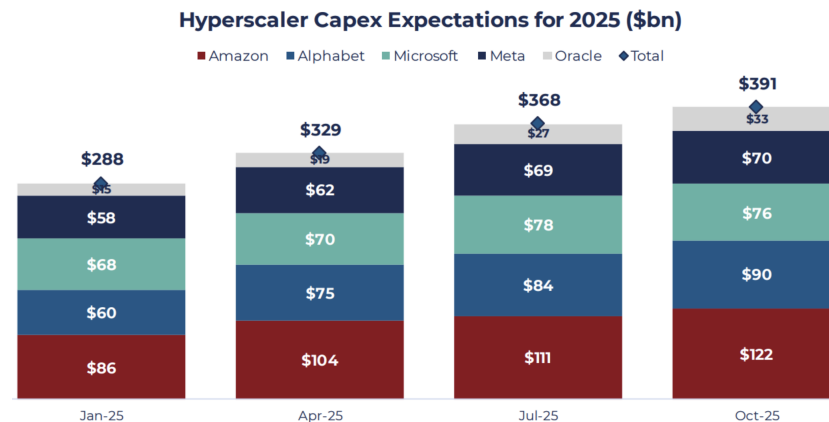
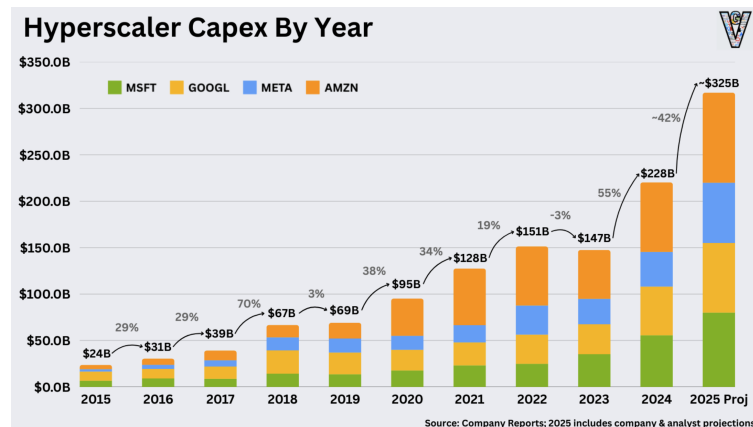
Agentic stack - LLM is engine, but not whole system

- *LLM - reasoning engine, not the system*
- stack
 - planner/orchestrator, memory (short/long-term), tools/APIs, environment interface
- patterns
 - ReAct, reflection/self-critique, planner-executor, multi-agent
- interoperability protocols emerging, *e.g.*, MCP, agent-to-agent
- *value migrating from model → system design*



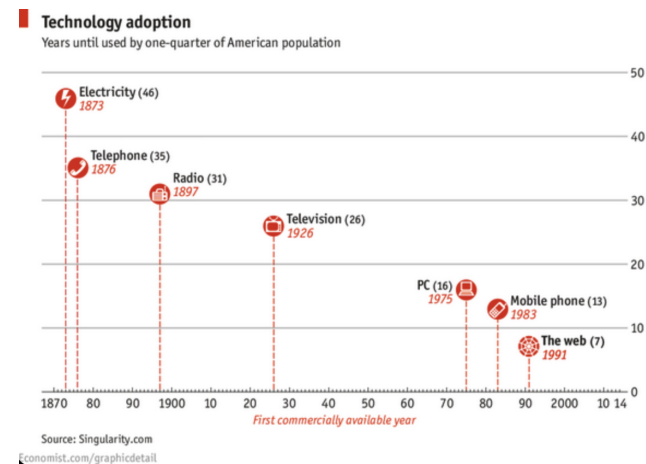
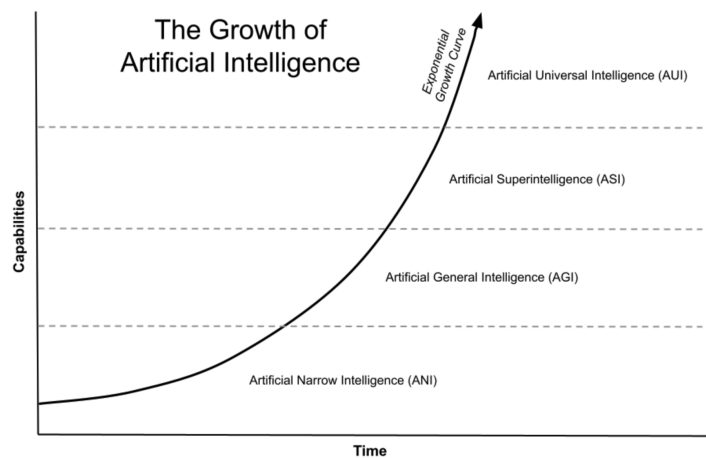
Trillion-dollar gravity well - capital, talent, geopolitics

- big 5 hyperscalers ~ \$725B AI capex in 2026 ~ Switzerland's GDP
- trajectory - \$256B (2024) → \$443B (2025) → \$725B (2026)
- *2026 is the first trillion-dollar year of compute capex in history*
- \$6.7T global data-center capex by 2030 (~70% AI) (McKinsey forecasting)
- *geopolitics*
 - export controls, chip sovereignty, national AI budgets



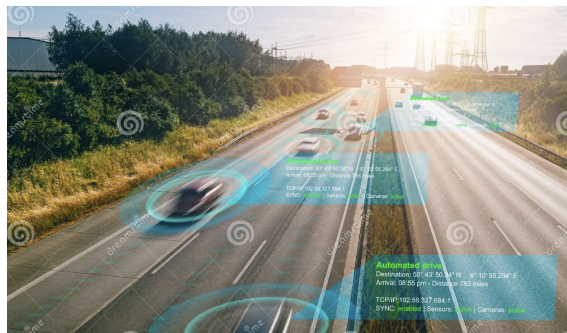
What took decades now takes months

- AlexNet (DL) → AlphaGo → Transformer → GPT (LLM) → Agentic in a decade
- adoption collapsing too
 - genAI penetration in fraction of PC/internet time
- *frontier you train on today will move by graduation*
- durable skill $\neq$ any one tool - it's relearning the frontier
- *what previously took decades now compresses into months!*



AI agents - present & future

- emerging applications
 - scientific research - agents analyzing & running experiments & generating hypotheses
 - creative collaboration - AI partners in design & art combining multiple mediums
 - environmental monitoring - processing satellite sensor data for climate analysis
 - healthcare - enhanced diagnostic combining imaging, *e.g.*, MRI, with patient history
 - customer experience - virtual assistants understanding spoken language & visual cues
 - autonomous vehicles - integration of visual, radar & audio data
- future
 - ubiquitous AI agents - seamless integration into everyday devices
 - highly tailored personalized experience - in education, entertainment & healthcare



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)



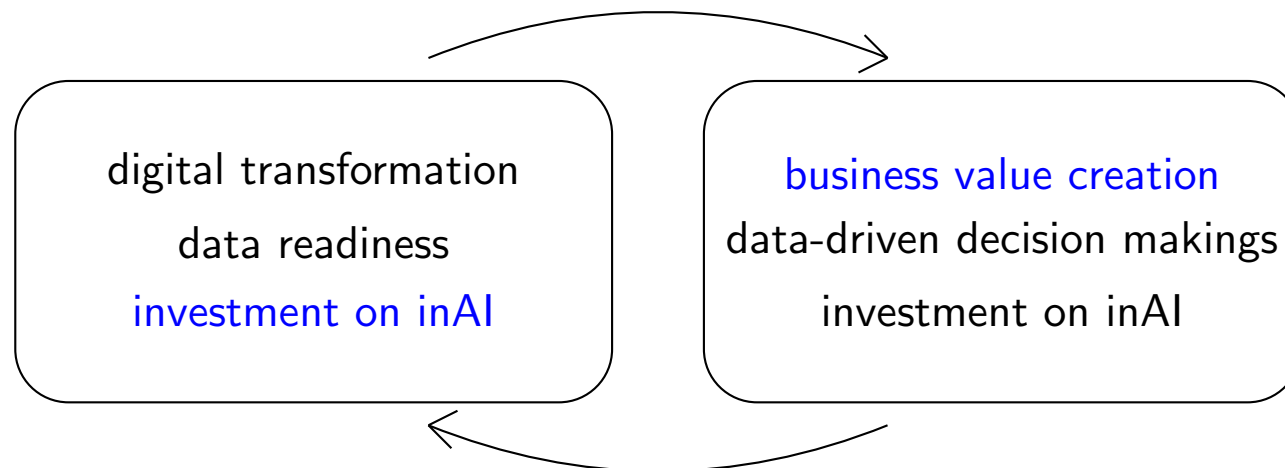
inAI fields

- 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

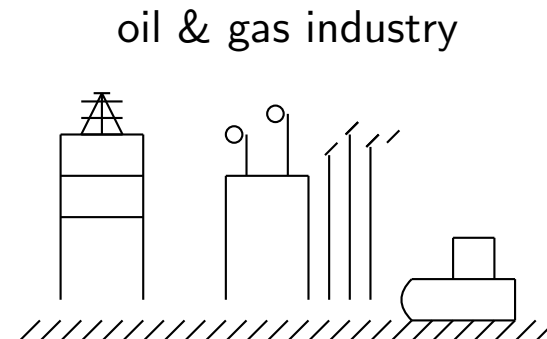
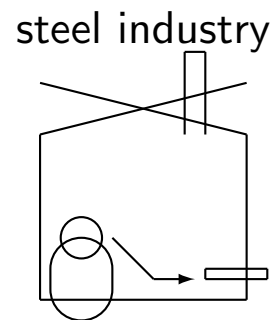
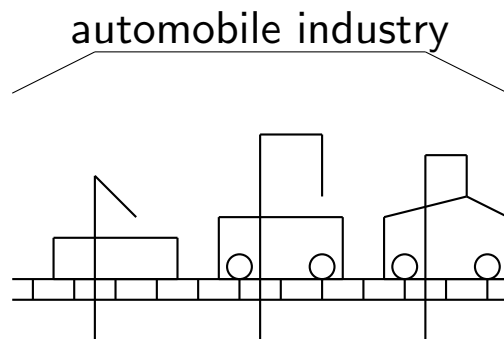
Vicious (or virtuous) cycle

- 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



Challenging data characteristics

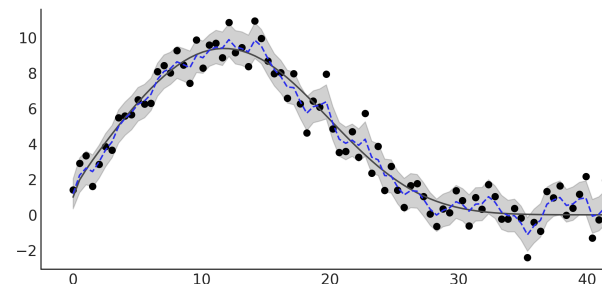
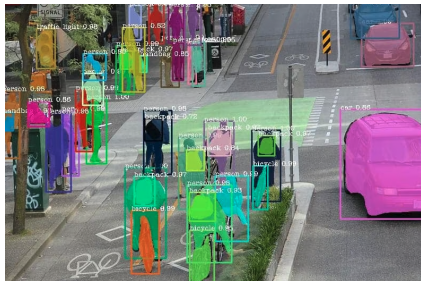
- huge volume
- data multi-modality
- high velocity requirement
- very fat data
- sever data shift & drift (in many cases)
- label imbalance
- data quality



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
 - 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
 - regression, anomaly detection, semi-supervised learning, Bayesian inference



¹SEM: scanning electron microscope, TEM: transmission electron microscope

²MES: 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

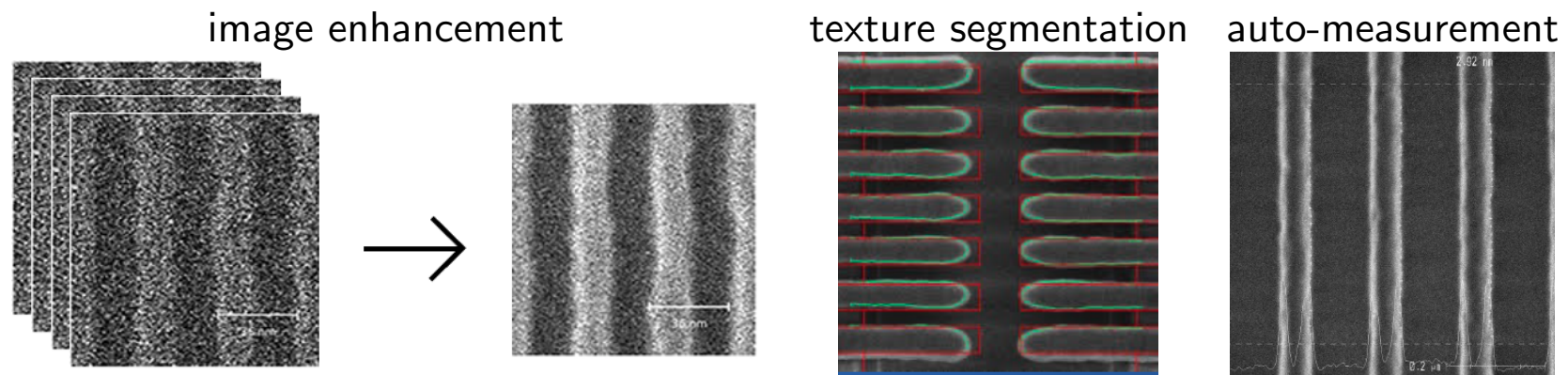


Image enhancement

- 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

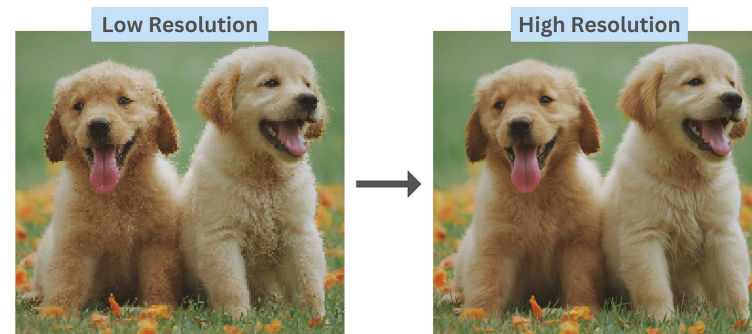
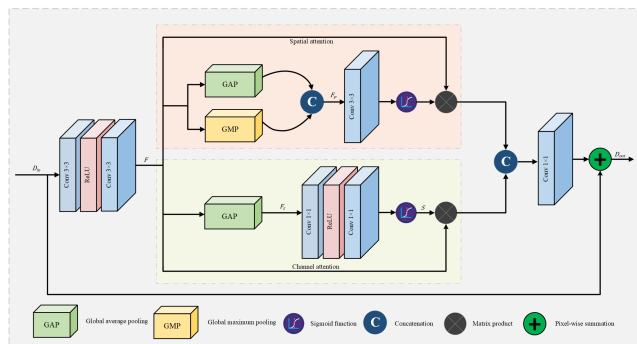
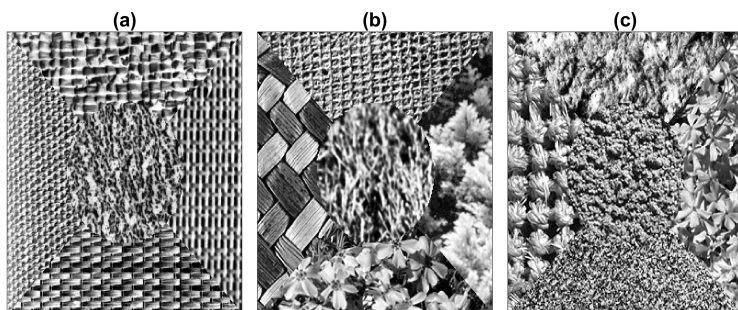


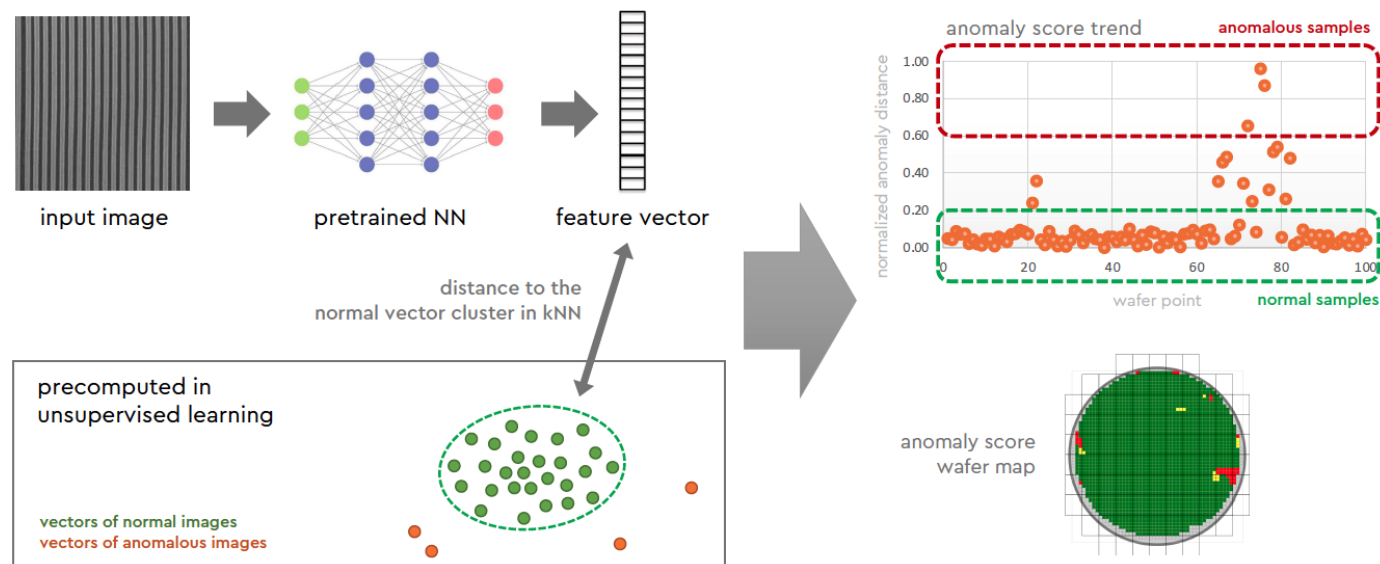
Image segmentation

- 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



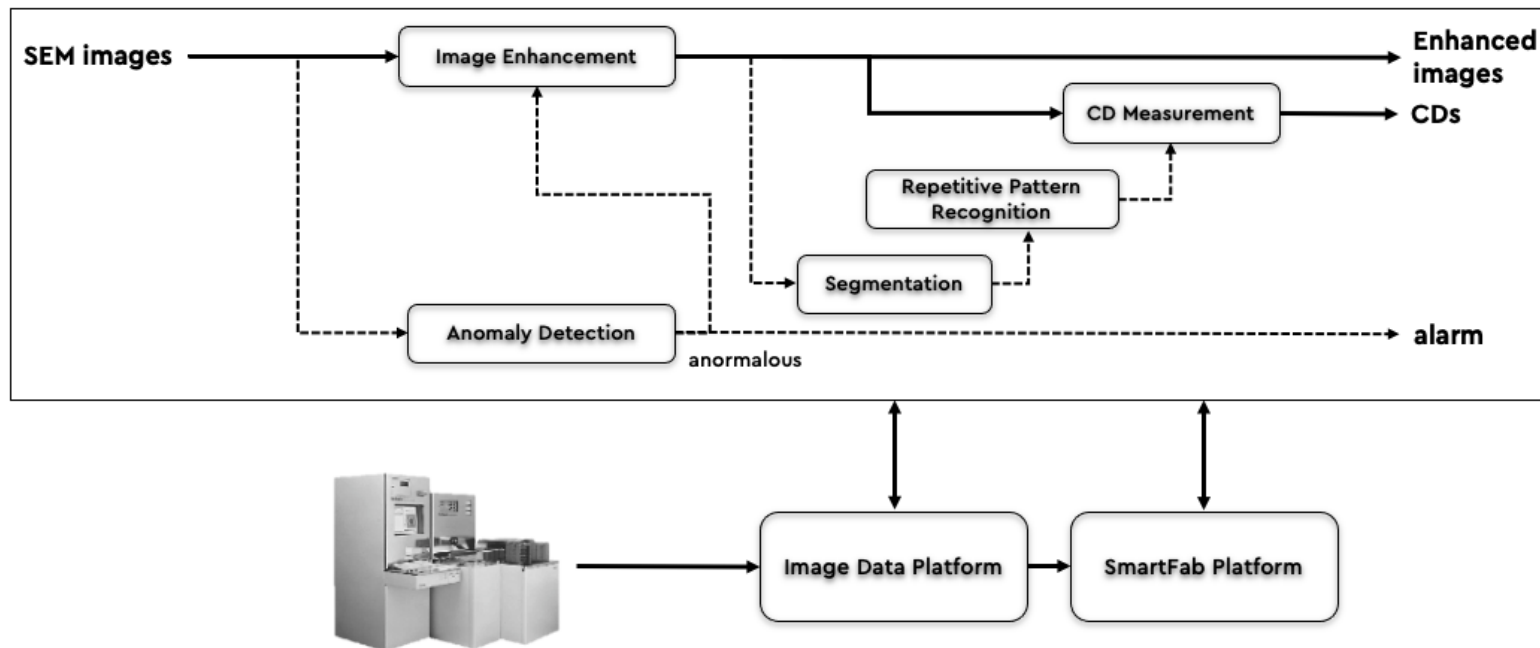
Anomaly detection using side product

- 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



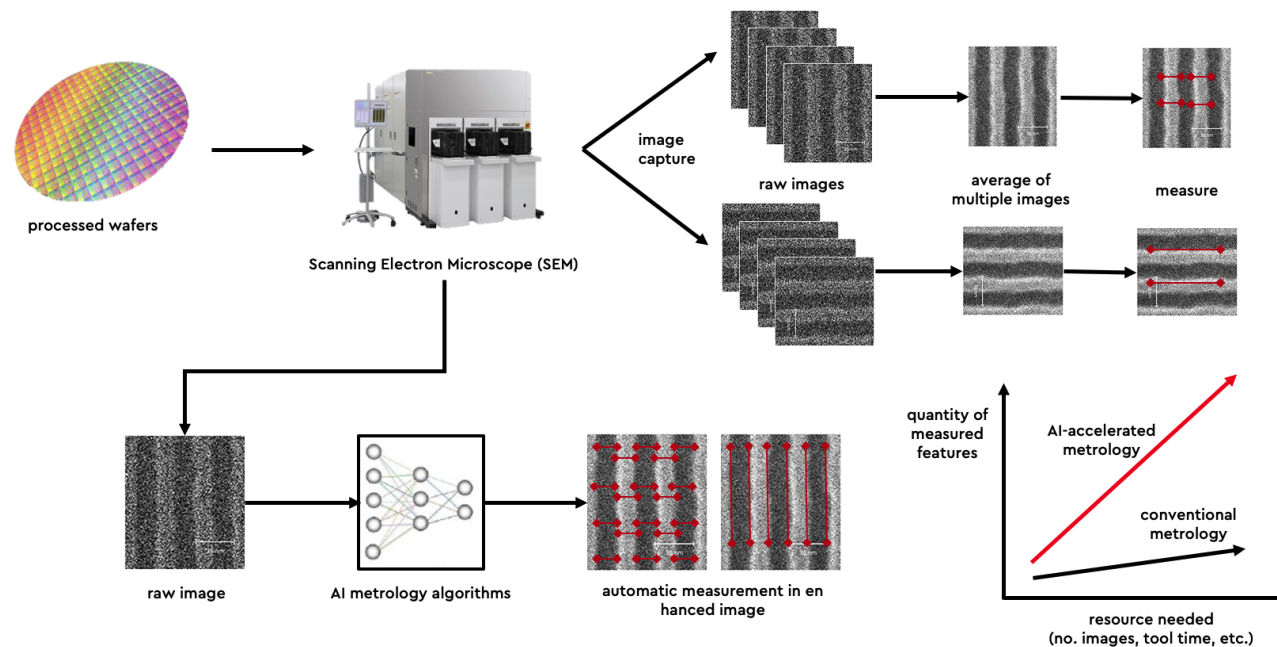
AI-enabled metrology system

- integration of separate components creates AI-enabled metrology system



Benefits of new system

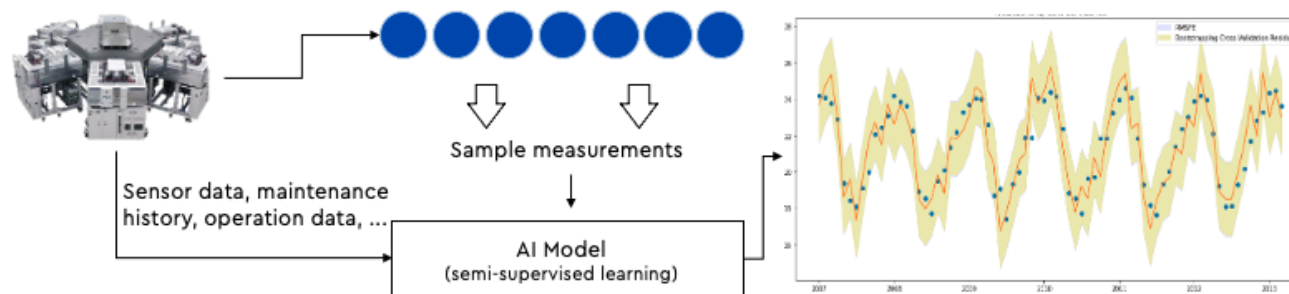
- new system provides
 - improved accuracy and reliability
 - improved throughput
 - savings on investment on measurement equipment



TS ML in manAI

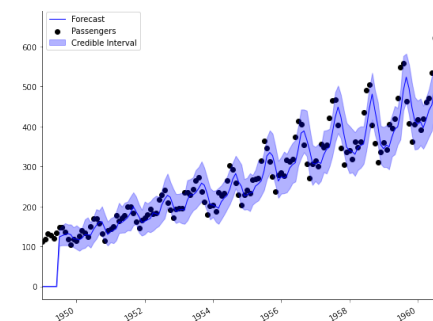
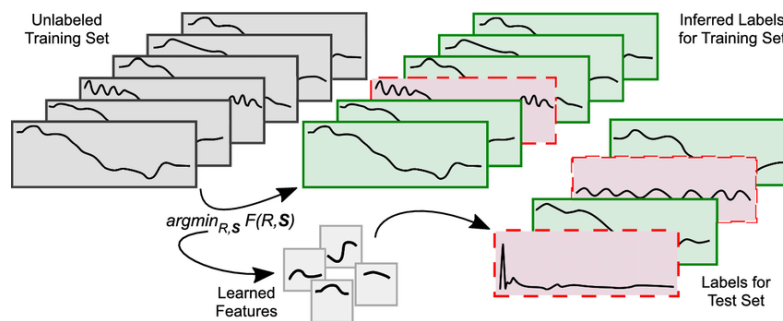
Time-series ML applications in manAI

- 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

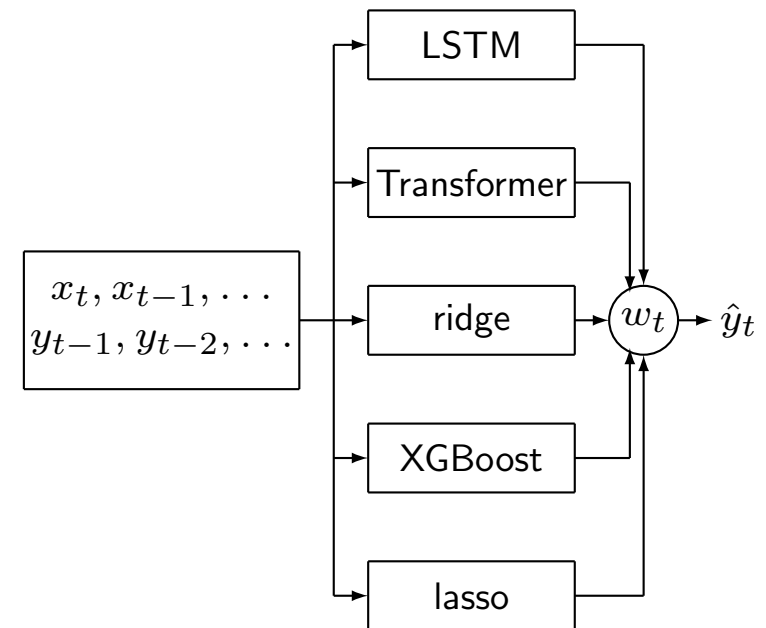


Difficulties with TS ML

- no definition exists for general TS data
- data drift & shift
 - $p(x_{t_k}, x_{t_{k-1}}, \dots)$ changes over time
 - $p(y_{t_k} | x_{t_k}, x_{t_{k-1}}, \dots, y_{t_{k-1}}, y_{t_{k-2}}, \dots)$ 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 multiple experts - $f_{1,k}, \dots, f_{p_k,k}$ 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 , $g_k : \mathbf{R}^n \rightarrow \mathbf{R}^m$ as weighted sum of experts



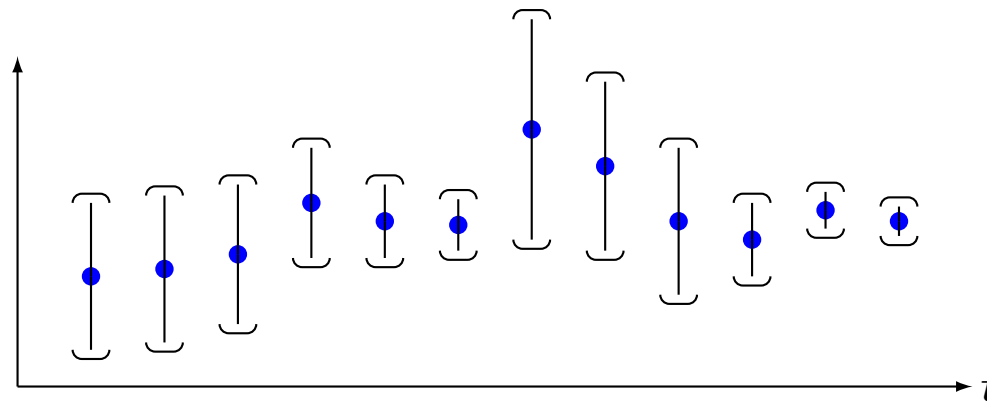
$$g_k = w_{1,k}f_{1,k} + w_{2,k}f_{2,k} + \dots + w_{p_k,k}f_{p_k,k} = \sum_{i=1}^{p_k} w_{i,k}f_{i,k}$$

Credibility intervals

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

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

- 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 results are
 - sometimes *more crucial than algorithm accuracy*



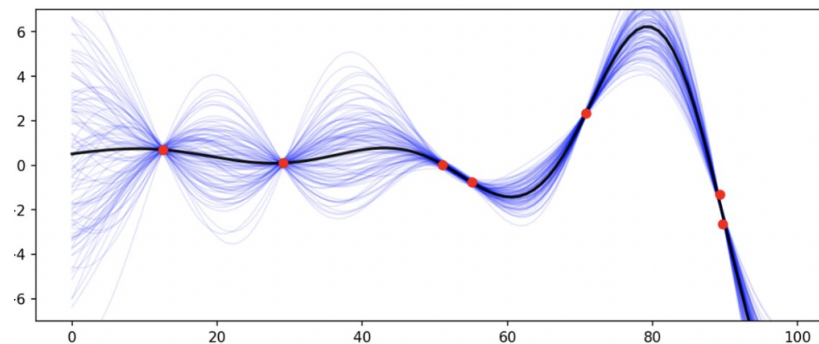
Bayesian approach for credibility interval evaluation

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

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

- 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

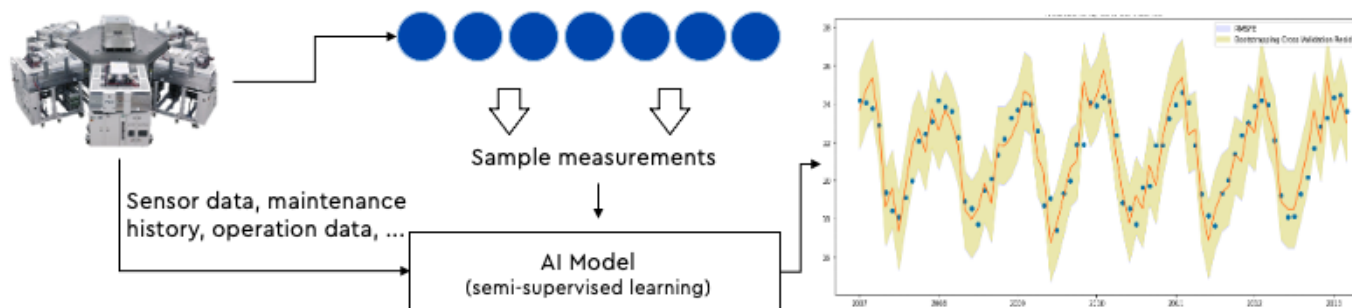
- problem description

(stochastically) predict y_{t_k}
 given $x_{t_k}, x_{t_{k-1}}, \dots, y_{t_{k-1}}, y_{t_{k-2}}, \dots$

- our problem formulation

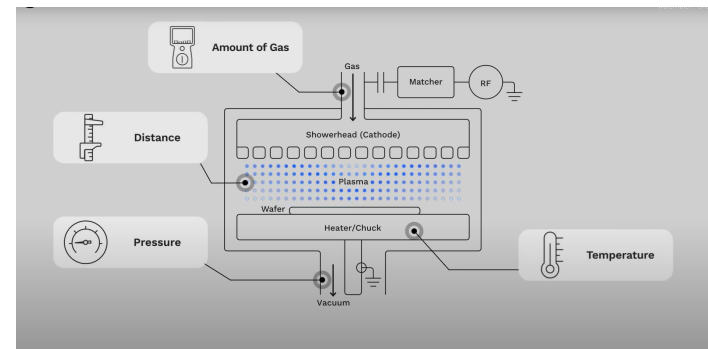
minimize $\sum_{k=1}^K w_{k,K-k} l(y_{t_k}, \hat{y}_{t_k})$
 subject to $\hat{y}_{t_k} = g_k(x_{t_k}, x_{t_{k-1}}, \dots, y_{t_{k-1}}, y_{t_{k-2}}, \dots)$

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



VM - Gauss Labs' inAI success story

- Gauss Labs' ML solution & AI product
 - fully home-grown online TS adaptive 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



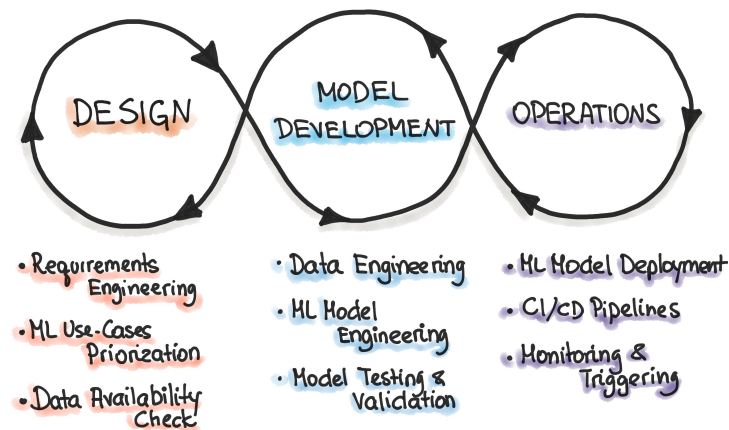
Manufacturing AI Productionization

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

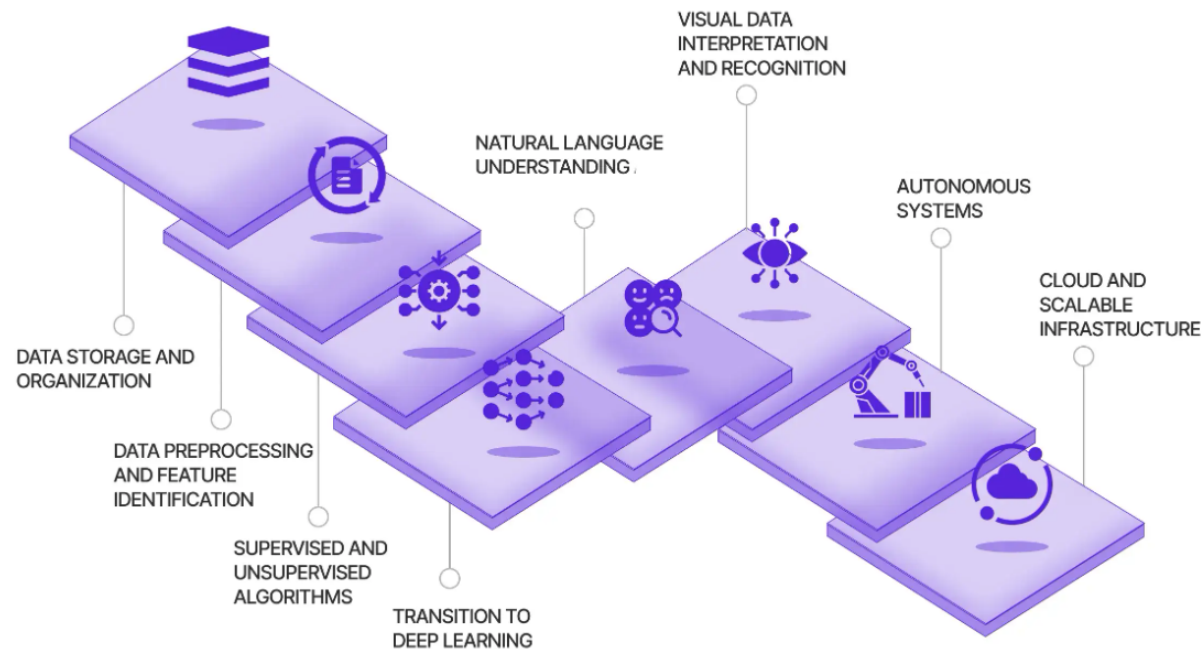
- environment for flexible and agile exploration - EDA³
- fast & efficient iteration of algorithm selection, experiments, & analysis
- correct training / validation / test data sets critical!
- seamless productionization from, *e.g.*, Jupyter notebook to production-ready code
- monitoring, *right* metrics, notification, re-training



³EDA - exploratory data analysis

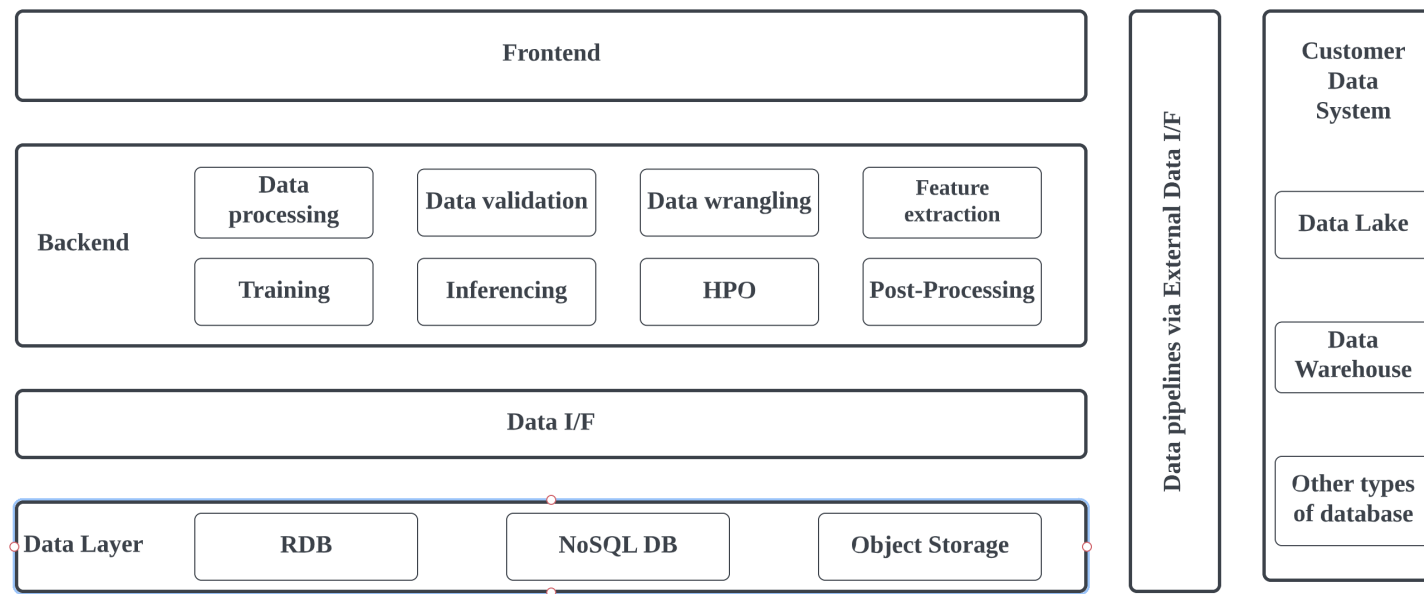
manAI software system

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



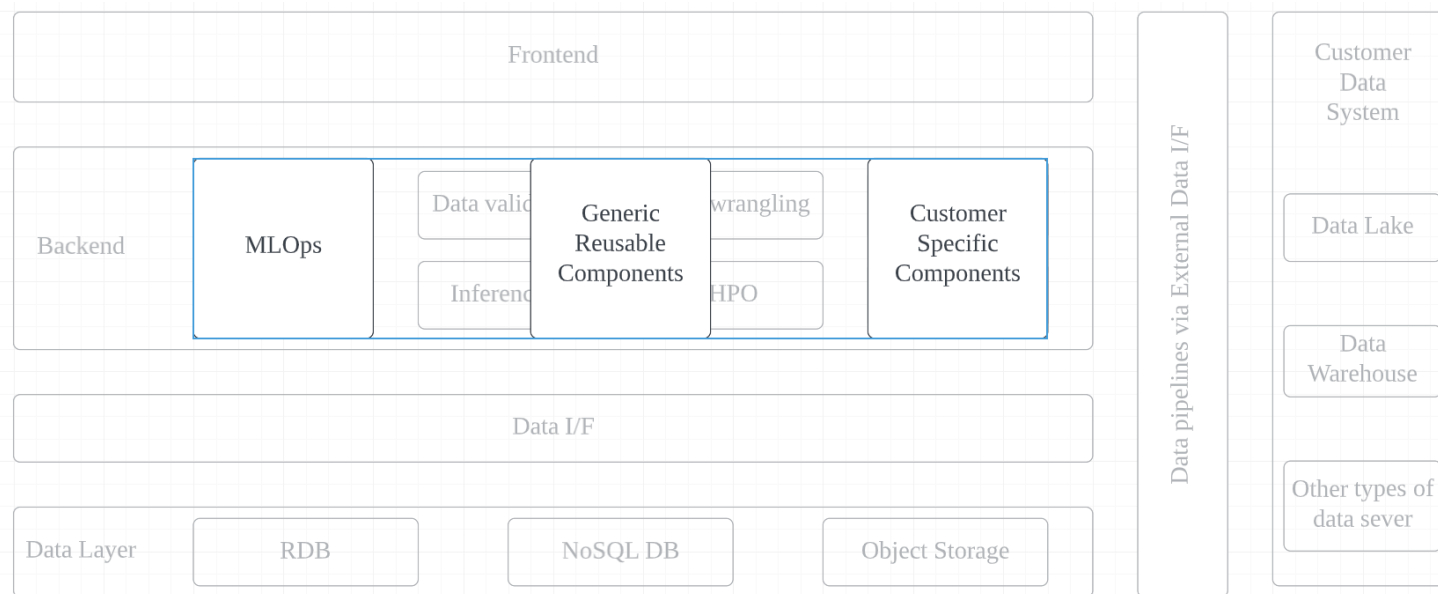
manAI system architecture

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



Reusable 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 critical – 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

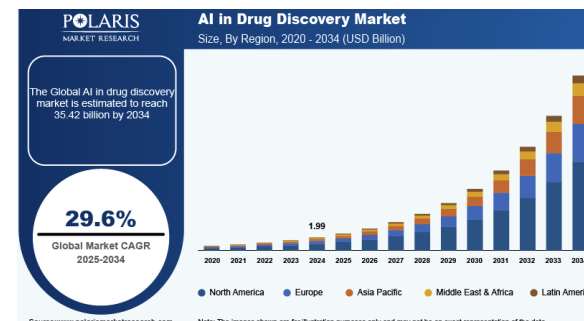
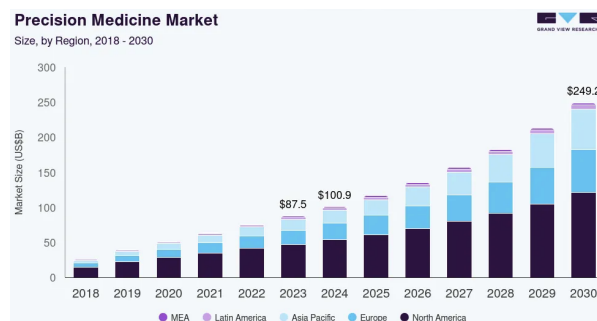
- various CV MLs used for inAI applications
- TS ML applications found in every place in manufacturing
- drift/shift & data noise make TS MLs very challenging, but working solutions found
- in reality, crucial bottlenecks are
 - data quality, preprocessing, monitoring, notification, and retraining
 -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

The Trillion Dollar Opportunities

The Moment - Why Now & Why it's Big?

The market is real - and enormous

- AI in biotech/pharma market size
 - AI in drug discovery - \$4B (2023) → \$50B+ by 2034 (Global Market Insights)
 - AI diagnostics market - \$1.2B (2023) → \$5-12B by 2030
 - precision/personalized medicine - \$80B (2023) → \$230B by 2030
 - synthetic biology - \$15B (2023) → \$100B by 2032
 - *combined TAM approaching \$1T by mid-2030s* - conservative estimate
- why biotech AI multiples exceed pure software AI
 - software AI competes on marginal cost → commoditizes fast
 - biotech AI - irreversible IP - novel molecules, validated biomarkers, proprietary assay
 - every successful clinical trial is a data moat that cannot be reverse-engineered

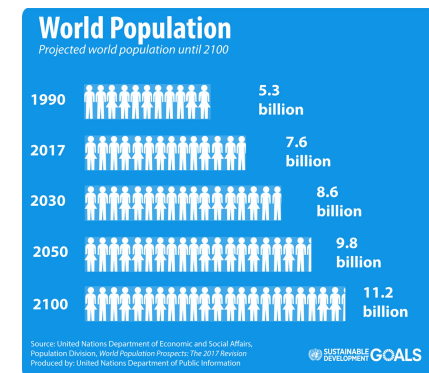


Macro tailwinds amplifying these opportunities

- aging global population - over 2.1 billion people over 60 by 2050 (UN DESA)
 - cancer, neurodegeneration, cardiovascular disease scale accordingly
- post-COVID regulatory acceleration
 - FDA's AI/ML action plan, EMA adaptive pathways
- government prioritization - US CHIPS & Science Act includes biotech
 - Gates Foundation, NIH, BARDA deploying billions in non-dilutive capital
- historical parallel
 - genomics wave (Human Genome Project) unlocked \$1T in economic value
 - *AI-biotech convergence is at least one order of magnitude larger*



Gates
Foundation



Three exponential curves converging now

- *curve 1 - LLMs & genAI*
 - language models operating natively on biological “languages” - protein sequences (ESM-2, ESMFold), SMILES molecular notation, genomic sequences
 - generative AI can propose novel drug candidates - Insilico Medicine’s INS018_055 went from AI-generated candidate to Phase II in 4 years vs industry avg of 10–15 yrs
 - multi-modal AI integrating imaging, omics, and clinical notes simultaneously
- *curve 2 - biochemical & biological breakthroughs*
 - AlphaFold 3 (2024) - extends beyond proteins to DNA, RNA, small molecules, and their interactions — the full drug-target interface
- *curve 3 - data availability*
 - DNA sequencing cost reduction, INSDC database, electronic health records digitized, wearables & continuous monitoring
- *The rarity of this moment - all three curves peaking simultaneously is historically unprecedented*

The Moat - What separates a Winner from Hype?

Cross-domain inevitabilities - The Technical Moat

- core insight
 - *fundamental mathematical structures recur across seemingly unrelated domains*
 - recognizing these - strategic advantage - optimization → biological energy landscapes
 - protein folding is fundamentally an energy minimization problem over high-dimensional conformational space
- information theory → cellular signaling
 - mutual information and channel capacity concepts (Shannon, 1948) map directly onto how cells encode and transmit signals through biochemical cascades
 - LLM training optimizes cross-entropy loss over token distributions
 - cellular gene regulatory networks optimize analogous information-theoretic objectives over transcription factor binding distributions
 - *researchers who understand why transformers work can transfer those architectural intuitions to biological sequence modeling*

What separates Unicorn Potential from incremental progress

- *dimension 1 - platform vs point solution*
 - point solution - AI model predicting one biomarker for one cancer type - narrow addressable market & low defensibility
 - *platform* - technology applicable across multiple disease areas, multiple biomarker classes, multiple assay modalities — TAM compounds with each new application
 - litmus test - “Can this technology be redirected to new disease areas in a few months without rebuilding from scratch?”
- *dimension 2 - is AI load-bearing or decorative?*
 - “AI-washing” - ML used for marketing positioning, not scientific differentiation
 - *load-bearing AI* - the AI component creates a result impossible or uneconomic to achieve otherwise
 - *e.g.*, Erudio Bio’s dynamic force spectroscopy + AI detecting cancer biomarkers at concentrations below conventional immunoassay thresholds

- *dimension 3 - pathway to clinical and regulatory reality*
 - computational elegance that cannot survive contact with clinical data is worthless
 - regulatory strategy must be designed into product from day one - not retrofitted after technical development
 - FDA's 510(k) vs De Novo vs PMA pathways have entirely different clinical evidence requirements — choice of pathway is a strategic decision made at founding
 - *hospital partnerships (e.g., SNUBH, Kyemyung Univ Dongsan Hospital) not just validation - they are the pipeline for clinically grounded training data!*
- *dimension 4 - proprietary data moat*
 - the most durable competitive advantage in AI-biotech is *the data no one else can access or replicate*
 - e.g., proprietary assay platforms generating novel measurement types, exclusive hospital partnerships, patient cohorts with longitudinal follow-up, rare disease registries
- *dimension 5 - human welfare at the center; not a constraint, but a strategic asset!*

- mission alignment with patient outcomes unlocks - NIH/Gates/BARDA non-dilutive funding, academic medical center partnerships, favorable regulatory posture, and - increasingly - LP mandates in impact-oriented VC funds

Investment landscape & white spaces

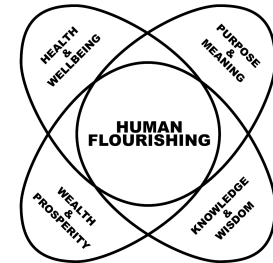
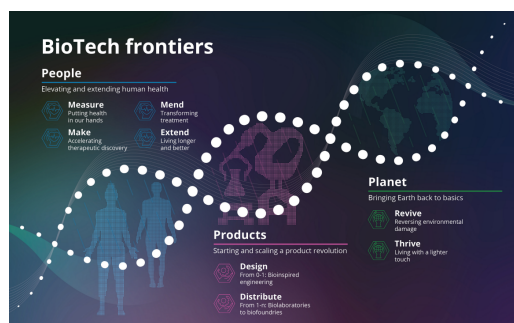
- where capital is concentrating - crowded but justified
 - *AI-native drug discovery platforms* - Insilico Medicine, Recursion Pharmaceuticals, Exscientia, Schrödinger - well-capitalized, public or late-stage
 - *protein engineering and design* - Absci, Generate Biomedicines, Cradle - foundation model approach to antibody and enzyme design
 - genomics interpretation - large language models trained on genomic sequences (Evo, Nucleotide Transformer)
 - risk - these spaces are getting crowded - differentiation increasingly difficult, capital efficiency under pressure
- underinvested white spaces - higher risk-adjusted opportunity
 - *AI-native diagnostic assays* - most diagnostic AI is retrofitted onto existing assay platforms - companies building AI-first measurement modalities have structural advantages in sensitivity, cost, and data proprietary
 - AI for *rare and neglected diseases* - Gates Foundation, Wellcome Trust, BARDA actively funding — orphan drug designation provides 7-year market exclusivity, priority review vouchers worth \$100M+ on the market

- *non-dilutive capital as validation signal*
 - Gates Foundation grants signal scientific credibility to commercial investors - functions as third-party technical due diligence
 - NIH SBIR/STTR - up to \$2M non-dilutive with right of first refusal waiver - massively underutilized by Silicon Valley founders
 - BARDA contracts for pandemic preparedness and biodefense - recurring revenue, not grant-dependent
- *timing - why now is the moat-formation window*
 - infrastructure (cloud genomics, foundation biology models, spatial omics platforms) just reached the point where startups can access capabilities that required a Broad Institute or Genentech five years ago
 - first-mover advantage in proprietary data accumulation compounds - every patient sample run through a novel assay platform strengthens the moat
 - *2025—2028 window* - analogous to 2008—2012 in cloud infrastructure - the companies founded now with the right platform architecture will be structurally dominant by 2030

The Maker - Who this Moment Demands?

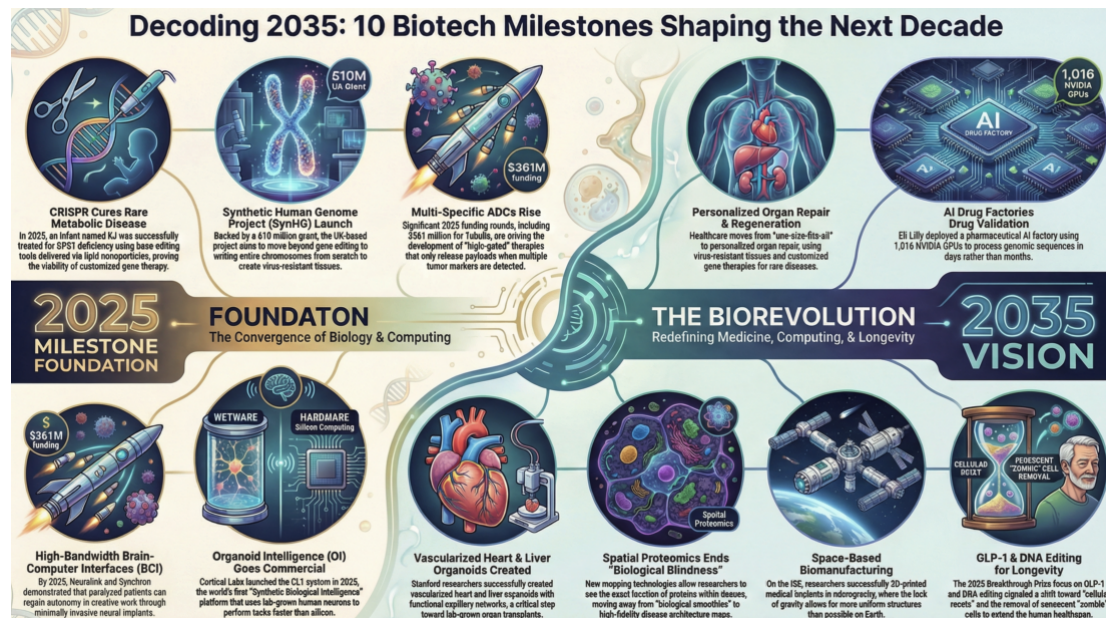
The Rare Entrepreneur This Moment Demands

- three simultaneous operating modes – all required
 - *deep technologist* - must understand inner workings well enough to know what AI can and cannot do & *spot when a competitor's technical claim is hollow*
 - *business strategist* - must navigate regulatory complexity, reimbursement strategy, partnership structuring, and capital allocation under uncertainty simultaneously
 - *advocate for human flourishing* - must hold *patient outcome as the north star* – not as marketing message, but as decision criterion resolving conflicts between speed, cost, and scientific rigor



What the audience in this room should take away

- *the convergence is real, the market is massive, and the technical foundations are now mature enough to build on*
- the scarcest resource is not capital, nor compute, nor biological data – it is *founders holding the full stack - theory, biology, engineering, clinical reality, and human purpose*
- *companies that will define human health in 2035 being founded in the next 24 months!!*

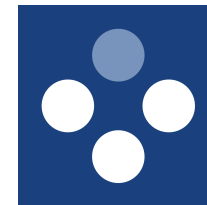
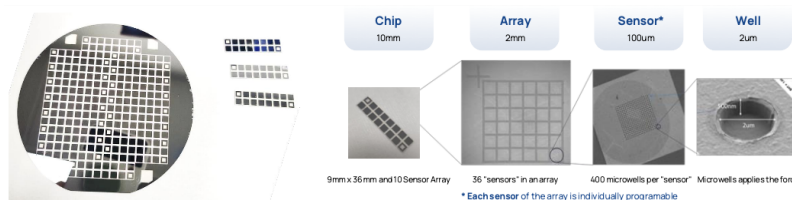
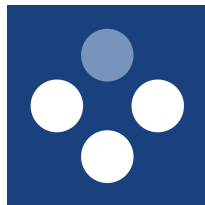


Appendix

Erudio Bio

Powering AI-driven medicine with ground-truth binding data

- problems we solve
 - 90% of drugs fail in clinical trials due to poor early-stage prediction
 - multiplexed diagnostics suffer from false positives and cross-reactivity
- *Erudio Bio's Innovation*
 - *VSA* platform uses patented “*dynamic force spectroscopy*” to generate 1000x more high-quality binding data from single sample ($\sim 10\mu\text{L}$)
 - measuring not just presence, but *strength* and *kinetics* of molecular interactions
- *dual business model*
 - diagnostics - *multi-cancer biomarker detection* with clinical institutions & hospitals
 - *drug discovery - bioTCAD™ platform* providing ground-truth labels to train & validate pharma AI models, reducing preclinical cycles



Validated technology, proven team, clear path to market

- validated impact
 - *\$1M Gates Foundation Grant* (2025) to democratize drug development for global health
 - partnerships with top research institutions (KRIBB, KAIST)
- unique team - *Stanford-trained founders* combining
 - semiconductor TCAD expertise & force spectroscopy innovation (20+ years)
 - AI & optimization leadership (Samsung, Amazon, SK hynix, Gauss Labs)
- market entry
 - *Korea → (Asia hub &) US* strategy with 2026 regulatory milestones
 - expanding *pharma partnerships and B2G*

Gates Foundation



Biological Assays Struggle with Scale & Accuracy

Data is expensive

- so we make decisions with *incomplete* picture
- status quo
 - limited, small-scale testing confirms diagnosis
 - outcome only as good as doctor's ability to determine which tests, limiting the picture
 - cross reactivity prevents larger scale testing
- Erudio creates
 - *comprehensive, large-scale* testing will drive diagnosis without assumptions
 - increased scale enables enhanced scientific discovery leading to
 - *better patient care*
 - *reduced time to diagnosis*
 - *cost reduction*



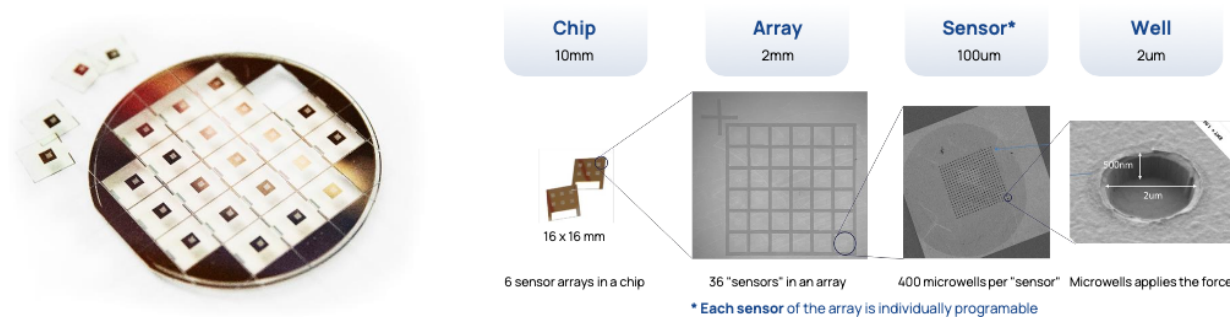
Erudio Bio starts revolution with Gates Foundation's support

- more data
 - comprehensive data from *single biological sample*
 - *multiplexed analysis* of nucleic acid, protein, cells, and more!
 - *multi-omic platform*
- actionable data
 - combined quality score from all data sources for comprehensive & conclusive assessment
- earlier data
 - complete data early to drive accurate decision making



Versatile Smart Assay (VSA) Platform

VSA technology



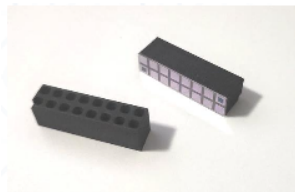
- generates *1000x more data* than the prevailing technology
 - scalable multi-omic microarray sensor
- *21 patents* in US, Canada, EU, and China
- indicates how good the data is in real time
 - patented “dynamic force spectroscopy” and “powerful Bayesian inference” method provides our data *quality score* to know their accuracy for actionable data
- AI software extracts a detailed, interpretable picture for quick diagnosis
 - leads to *AI knowledge discovery* resulting in *data-driven diagnosis*

Enabling comprehensive data acquisition

- antibodies - versatile tools in biology
 - can engineer to target virtually *anything* we want
 - problem
 - indiscriminate interactions severely limits use of antibodies – *cross-reactivity*
 - error-prone results due to *non-specific binding*
- solution - comprehensive data with *dynamic force spectroscopy*
 - comprehensive binding strength to distinguish specific from non-specific binding
 - *quality score* discerns noise from useful data to enable multiplexing



VSA's business models



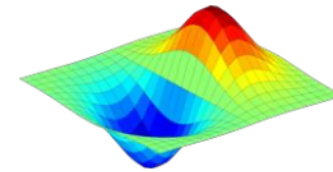
Consumable chip
& flowcell



Instrument



Consumable
reagent kit



Software
AI/ML & SaaS

- VSA platform
 - instrument - recurring revenue with high margin
 - modular licensable software - AI based data interpretation and feature extraction
- SaaS
 - subscription based pre-validation of reagent database
 - AI feature extraction and knowledge discovery

When Erudio's VSA meets AI - Gates Foundation Grant

Erudio Bio wins \$1M Gates Foundation Grant - scaling bioTCAD

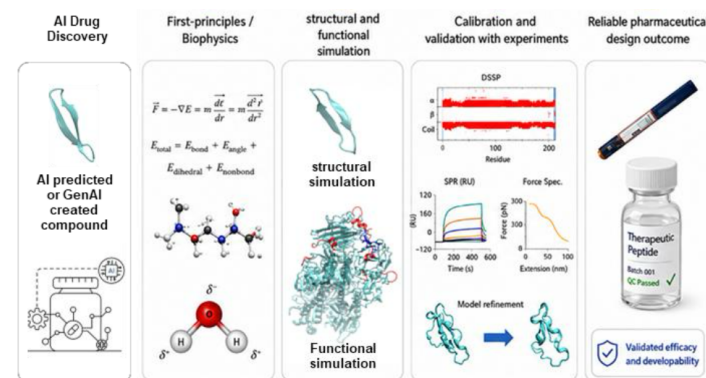
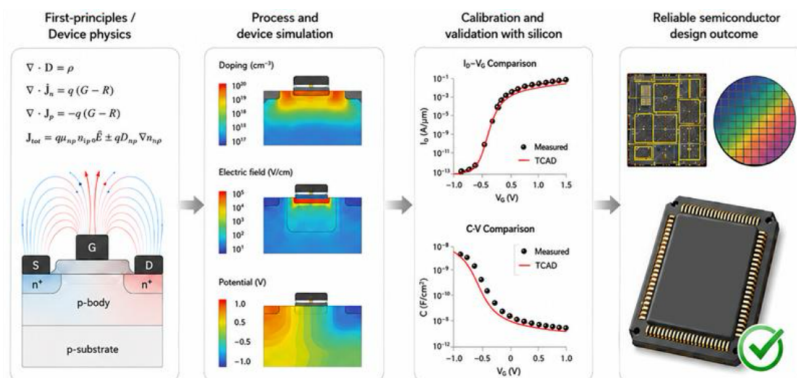
Gates Foundation



- *\$1M Grant Award (2025)*
 - Gates Foundation recognizes Erudio Bio's potential to transform drug development for global health
- mission alignment - democratizing medicine by making preclinical drug design faster, yet reliable & accessible
 - lowering development costs for diseases affecting LMICs
 - addressing the 90% clinical trial failure rate that drives up drug costs
- funded project - develop *bioTCADTM platform for lead optimization of drug discovery*
 - expand force spectroscopy measurements across high-burden disease targets
 - train AI models with kinetics-resolved binding data (on/off rates, unbinding forces)
 - *enable pharma/biotech to prioritize candidates earlier with higher confidence*

bioTCAD - hybrid AI & physics-based drug development

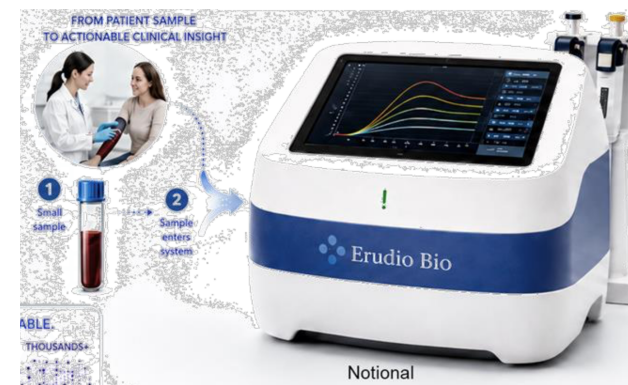
- (old school) AI/ML excels at pattern recognition across large chemical libraries
 - identifying candidate hits at scale – valuable role established and growing
- AI alone *cannot reliably explain binding physics or predict behavior in unexplored and novel chemical space* – *capability and credibility gap*
- bioTCAD combines *AI and measurement-backed, physics-grounded simulation*
- applies *the* principle that made semiconductor TCAD trustworthy
 - validate model parameters to YOUR experimental measurements to be reliably right



Erudio Bio's Business Models

Erudio Bio Applications

- drug development
- clinical diagnostic
 - medicine is already 20% of the world's economy and growing at 5% per year
- biodefense
 - GWOT to great power competition
 - need to defend against near-peer adversaries
 - flexible, efficient solution needed from CBRND to readiness



Teams

Team & advisory board

- team
 - Kee-Hyun Paik, Ph.D. (CEO) - chip, microfluidics, instrumentation
 - Sunghee Yun, Ph.D. (CTO) - AI, optimization, business development, software
 - Susanne Baumhueter, Ph.D. - biology, immunology, project management
- advisory board
 - Michael Cola - CEO of AEVI Genomic Medicine (\$62B sales to Takeda)
 - Tim Germann - CCO of Carterra Bio
 - Karyn Eliot - retired CIA Sr. Executive
 - Phil Ferro - virologist, formerly DoD, DoS, HHS and White house
 - Bill Chen - Former hedge fund and VC professional with national security, FFRDC
 - Ronald W. Davis - Director of Stanford Genome Tech Center (\$15B+ exits)
 - Michael Snyder - Prof. Genetics, Dir. Stanford Human Genome Center
 - William J. Greenleaf - Prof. Genetics and Applied Physics, Stanford University



Sunghee Yun

Aug 20, 2026



Silicon Valley AI Nexus

Silicon Valley AI Nexus (or just AI Nexus)

- pioneering community of professionals dedicated to building privacy-preserving AI solutions, products, and systems
- comprehensive expertise across AI domains
 - biotechnology, healthcare, and medical research
 - industrial applications and data centers
 - cloud infrastructure, storage solutions, mobile technologies
 - customer service platforms, multi-agent systems
 - RAG implementations, vector databases, agentic AI frameworks
- vision
 - *shaping future where AI innovation and privacy protection go hand in hand*
- active community with [homepage](#) & KakaoTalk collaboration platform for members

Our journey - forum history

- Nov-Dec 2024 - “The AI Strikes Back” & “Free Your Data”
 - Prof. Jung Hee Cheon (homomorphic encryption revolution)
- Jan 2025 - “The AI Knight Rises”
 - [Sunghee Yun](#) @ Erudio Bio on deep learning to flourishing societies
- Feb 2025 - “Silicon Citadel”
 - Chanik Park @ MangoBoost on AI data infrastructure
- Mar 2025 - “Blockchain Awakens”
 - Daejun Park @ a16z crypto on decentralized AI
- Apr 2025 - “Advancing Humanity”
 - Stanford Medicine team on bio/medical AI
 - co-hosting with K-BioX
- May 2025 - “The Autonomous Alliance”
 - Microsoft, GitHub, Uclone, SK Hynix on AI agents

Our journey - forum history

- Jun 2025 - “Silicon Companions”
 - Altos Ventures on robotics & smart devices
- Aug 2025 - “The Human-Centric AI Revolution”
 - address legal and ethical issues related to AI
- Nov 2025 - “The AI Silicon Race”
 - Korea-US Innovation Leadership at K-ASIC



Strategic partnerships & ecosystem

- *Perpetual Partnership with KOTRA Silicon Valley as Strategic Alliance*
- 2026 co-hosting partners
 - K-ASIC (Korea AI & IC Innovation Center)
 - K-BioX (biotech innovation)
 - KOTRA Silicon Valley (trade & investment)
 - Consulate General of the Republic of Korea at San Francisco (diplomatic support)
 - KABANC (Korean American Bar Association of Northern California - legal expertise)
- building bridges between Silicon Valley innovation and Korean institutional networks
- creating comprehensive support ecosystem: technical, legal, business, diplomatic



Community & engagement

- membership requirements
 - attend one! AI Nexus event!
- member benefits
 - networking with AI professionals across all domains
 - knowledge sharing and collaboration opportunities
 - direct access to world-class speakers and experts
- forum format - 5pm-8pm, typically Wednesdays at premier Silicon Valley venues
- venues - Stanford, KOTRA, SK Hynix, Altos Ventures, K-ASIC, and more
- active community engagement and professional development



Three Core Tenets

- *Explore & Lead*
 - world-class forums, the AI & Humanity Council, and thought leadership shaping AI's most consequential questions
- *Elevate Community*
 - mentorship, startup showcases, professional connections, and leadership opportunities that amplify every voice
- *Give Back to Society*
 - policy engagement, academic partnerships, and public discourse ensuring AI's benefits reach beyond Silicon Valley

AI Nexus encourages and embraces Interest Groups

- Bio / Medical / Data Group (7)
 - Chanik Insuk Jeff Jongmin Minha Mitch Sunghee led by Chanik
- Karaoke Group (4)
 - Hayden Jeff Sunghee Yeseul led by Sunghee
- AI Hands-on Group (5)
 - Chanik James Kyoung Whan Minha Sunghee led by Minha
- East Bay Group (7)
 - Daniel Eugene Hyunjung Linda MiSook MJ Sunghee led by Hyunjung

Collaboration with Institutions & Universities

- MOUs
 - SNU
 - *AI Nexus × SNU College of Engineering Silicon Valley Immersion Program*
 - *Winter Internship 2026*
 - asked by
 - KIC SV
 - SVreboot
 - XL8
 - KASPA
 - Korea Univ.
 - Sungkyunkwan Univ.
 - KAIST
 - Korean American Semiconductor Professional Alliance (KASPA)
 - *etc.*

Collaboration with Institutions & Universities

- Co-hosting with
 - K-BioX
 - KOTRA SV
 - KIC SV
 - K-ASIC
 - Korea-America Student Conference (KASC)
 - Consulate General of the Republic of Korea at San Francisco
 - Korean American Bar Association of Northern California (KABANC)
 - *etc.*

AI Nexus Committee

- Board of Directors
 - *Sunghee Yun* - Co-Founder & Char
 - *Minha Hwang* - Co-Founder & AI Master
 - *Sue Kim* - Co-Founder & Marketing Leader
 - *Youngwook (Hayden) Song* - Co-Founder & Legal Affairs Advisor
 - *Bo Hyoung (Jeff) Lee* - Co-Founder & Legal Affairs Advisor
 - *MiSook Chung* - Co-Founder & Community Engagement Chair
- Vision & Principles Committee
- Tech & Social Media Committee
- Program & Content Committee
- the full list of committees can be found [here](#)!

Some Important Questions around AI

Some important questions around AI

- why human-level AI?
- what lies in very core of DL architecture? what makes it work amazingly well?
- biases that can hurt judgement, decision making, social good?
- AI ethics & legal issues
- consciousness
- utopia vs dystopia
- knowledge, belief, reasoning
- risk of anthropomorphization

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 advantages for medical assistance, emergency handling
 - AI does not make more mistakes because task is repetitive and tedious
 - AI does not request salary raise or go on strike

What makes DL so successful?

Factors contributing 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 $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$, exists $\langle f_n \rangle$ such that $\lim_{n \rightarrow \infty} f_n = f$

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, . . .

Sudden leap in LLM performance

Probability inferred sequence is correct

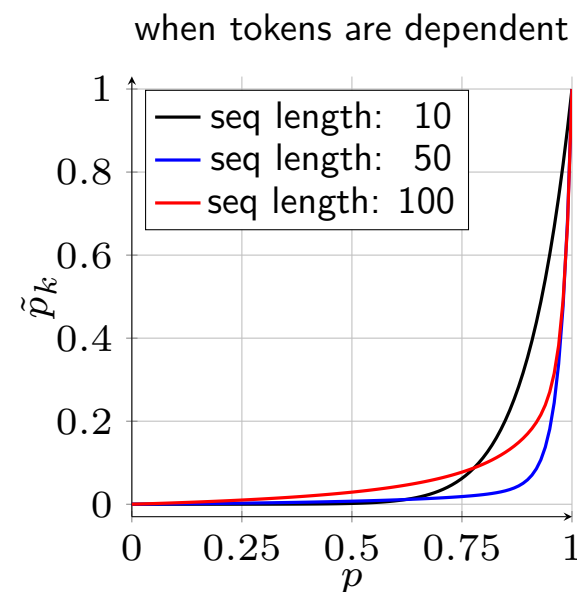
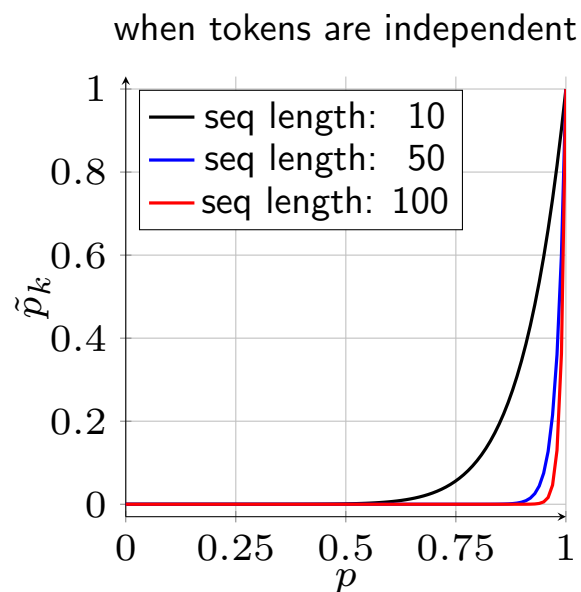
- assume
 - t_i - i th token
 - p_i - probability that t_i is correct
 - ρ_i - correlation coefficient between t_{i-1} & t_i
 - $\tilde{p}_k$ - probability that $(t_1, \dots, t_k)$ are correct
- recursion

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

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

Dramatic improvement of LLM near saturation

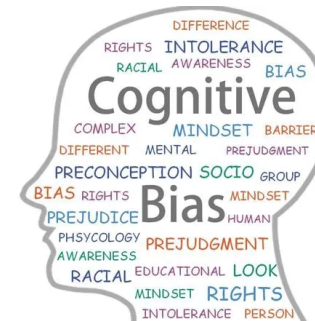
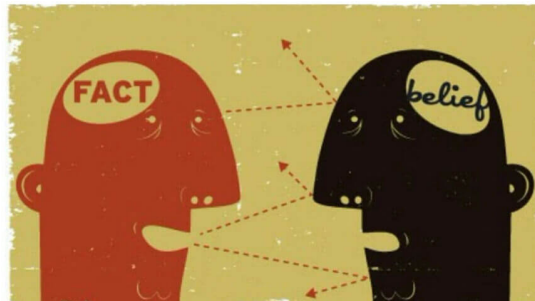
- do simulations for both independent & dependent cases
 - assume p_i are same for all i
- (for both cases) sequence inference improves dramatically as p approaches 1
- this explains *why we have observed sudden dramatic performance improvement of certain seq2seq learning technologies, e.g., LLM*



Biases

Cognitive biases attributed to humans

- 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



Biases of LLMs

- LLMs subject to
 - availability bias - biased 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 easily 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

AI Ethics

Ethical issues related to AI

- AI can be exploited by those who have bad intention to
 - manipulate / deceive people - using manipulated 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

AI related Legal Issues

Legal issues with ethical consideration

- 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 & humanitarian 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?
- is SKYNET ever plausible?
 - can AI have *desire* to survive (or save earth)?



Utopia vs Dystopia

Utopia vs 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, immoral, and unethical) people possessing social, economic & political power
 - *social good and welfare impaired* by either exploiting AI or ignorance of (inner workings of) AI
- should concern
 - choice or balance among utilitarianism, humanitarianism & values
 - amend or improve laws/regulations
 - ethical issues caused by AI

Knowledge, Belief, and Reasoning

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

**What categories of questions do they belong to?
engineering, scientific, philosophical, cognitive scientific, . . . ?**

LLMs . . .

- 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

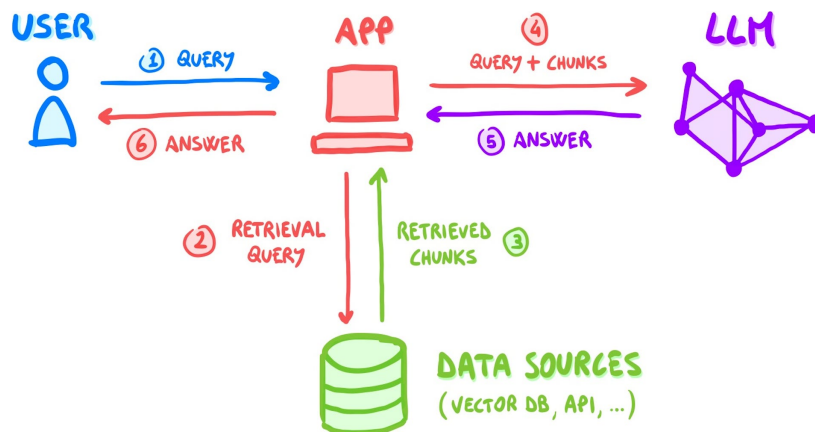
- making vulnerable to anthropomorphism
- examine it by throwing questions such as
 - “*does LLM have knowledge and belief?*”
 - “*can it reason?*”

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’?”*

LLMs vs 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, *e.g.*, for question-answering, news article summarization, screenplays generation, language translation



How ChatBot works?

- conversational AI agent does *in-context learning* or *few-shot prompting*
- for example,
 - when the user enters
who is the first person to walk on the Moon?
 - ChatBot, LLM-embedded system, feeds the following to LLM
User, a human, and BOT, a clever and knowledgeable AI agent.
User: what is 2+2?
BOT: the answer is 4.
User: where was Albert Einstein born?
BOT: he was born in Germany.
User: who is the first person to walk on the Moon?
BOT:

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 perspectives
 - cognitive scientific & philosophical perspectives

Laymen's perspectives on knowledge, belief & reasoning

- does (good) LLM have knowledge?
 - Grandmother: looks like it cuz when instructed *“explaining 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!

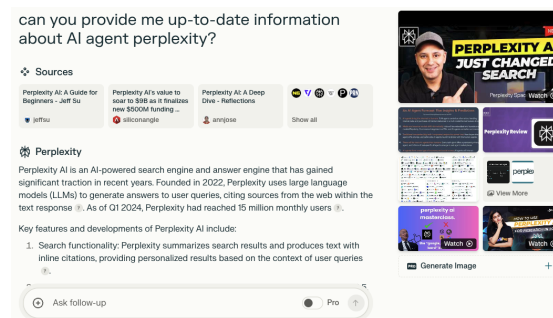
Knowledge

- could argue LLM “knows” which words follow which other words with high probability
- but, only *in context of capacity to distinguish truth from falsehood* can we legitimately speak of “knowledge”!
- LLM(-embedded BOT)
 - can be said to “*encode*”, “*store*”, or “*contain*” knowledge
 - lacks means to use words “true” & “false” in all ways & in all contexts because . . .
 - *does not inhabit the world* we human language-users share!



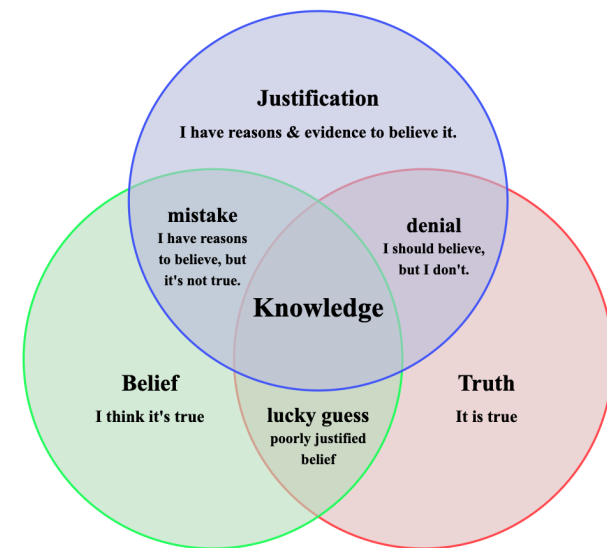
Belief

- nothing can count as *belief about the world* we share unless
 - is against backdrop of “*ability to update beliefs appropriately in light of evidence from that world*” - (again) essential capacity to distinguish truth from falsehood
- change taking place in humans when acquiring or updating belief is
 - reflection of their nature as language-using animals inhabiting shared world with community of language-users
- then, *what if LLM-embedded system updates LLM with outside world information?*
 - even so, when interacting with AI systems based on LLMs, these grounds are *absent!*



Knowledge in philosophical and cognitive scientific sense

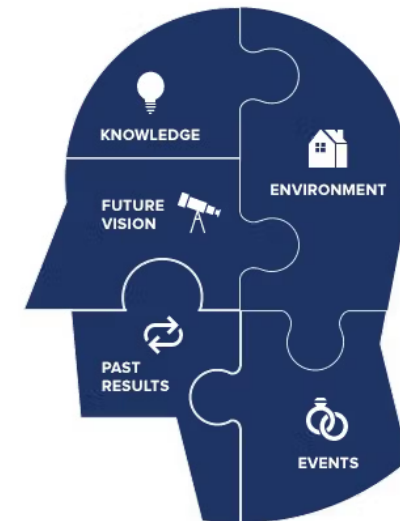
- does LLM have knowledge?
 - Sunghee: *I don't think so!*
- why?
 - we say we have “knowledge” when
“we do so against ground of various human capacities that we all take for granted when we engage in everyday conversation with each other.”
 - when asked *“who is Tom Cruise's mother?”*, it says *“Tom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



Belief in philosophical and cognitive scientific sense

- for the discussion
 - do *not* concern any specific belief
 - but concern *prerequisites for ascribing any beliefs to AI system*
- so does it have belief?
 - 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, essential aspect of the capacity to distinguish truth from falsehood*
 - LLM does not have this ground, essential consideration when deciding whether it *really* had beliefs.
- Sunghee: so *no, LLM cannot have belief!*

WHERE DO YOUR BELIEFS COME FROM?



Reasoning in philosophical and cognitive scientific sense

- note reasoning is *content neutral*
 - *e.g.*, following logic is perfect regardless of truth of premises
 - hence, no access to outside world does *not* disqualify
- 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”
given the statistical distribution of words in public corpus, what words are likely to follow the sequence, “humans are immortal and Socrates is human therefore.”
- Sunghee: so *no, LLM cannot reason, either!*
- but, LLM
 - pretends to reason, and from which capabilities, we can benefit!
 - also, can *mimic even multi-step reasoning whose inferencing structure is complicated* using *chain-of-thoughts prompting*, *i.e.*, *in-context learning* or *few-shot prompting*

Simple example showing LLM not possessing knowledge

- User

"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, . . . "

- User

"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. . . . "

Risk of anthropomorphization

- unfortunately, contemporary LLMs are *too powerful, too versatile, and too useful for most people to accept (after understanding) previous arguments!*
- maybe, o.k. for laymen to (mistakenly) anthropomorphize LLM(-embedded systems)
- however, *imperative for (important, smart, and responsible) AI researchers, scientists, engineers & practitioners* to have rigorous understanding in these aspects especially when
 - advise and be consulted by law makers, policy makers, journalists, and various stakeholders responsible for *critical business decisions (in private sectors) and public policies (in public sectors)*
 - collaborate with or/and help professionals in liberal arts, such as *philosophy, ethics, law, religion, literature, history, music, cultural studies, psychology, sociology, anthropology, political science, economics, archaeology, linguistics, media studies, natural sciences, fine arts, . . .*
 - to address negative societal and economic impacts

Moral

- AI shows incredible utility and commercial potentials, hence should
 - make informed decisions about trustworthiness and safety
 - avoid ascribing capacities they lack
 - *take best utilization of remarkable capabilities of AI*
- today's AI 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, law & policy makers, journalists, . . .*

Selected References & Sources

Selected references & sources

- Robert H. Kane “Quest for Meaning: Values, Ethics, and the Modern Experience” 2013
- Michael J. Sandel “Justice: What’s the Right Thing to Do?” 2009
- Daniel Kahneman “Thinking, Fast and Slow” 2011
- Yuval Noah Harari “Sapiens: A Brief History of Humankind” 2014
- M. Shanahan “Talking About Large Language Models” 2022
- A.Y. Halevry, P. Norvig, and F. Pereira “Unreasonable Effectiveness of Data” 2009
- A. Vaswani, et al. “Attention is all you need” @ NeurIPS 2017
- S. Yin, et. al. “A Survey on Multimodal LLMs” 2023
- Chris Miller “Chip War: The Fight for the World’s Most Critical Technology” 2022
- CEOs, CTOs, CFOs, COOs, CMOs & CCOs @ startup companies in Silicon Valley
- VCs on Sand Hill Road - Palo Alto, Menlo Park, Woodside in California, USA

References

References

- [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.
- [MLZ22] Louis-Philippe Morency, Paul Pu Liang, and Amir Zadeh. Tutorial on multimodal machine learning. In Miguel Ballesteros, Yulia Tsvetkov, and Cecilia O. Alm, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pages 33–38, Seattle, United States, July 2022. Association for Computational Linguistics.
- [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.

- [Sha23] Murray Shanahan. Talking about large language models, 2023.
- [YFZ⁺24] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models, 2024.

Thank You