# [KOBRA Conference 2026 AI-Biotech Special Talk] From Silicon Valley AI to Multi-Omics Medicine - Versatile Smart Assays Transform Cancer Diagnostics and Prognosis Prediction Through Cross-Domain Innovation


**Sunghee Yun**

Co-Founder & CTO @ Erudio Bio, Inc.
Co-Founder & CEO @ Erudio Bio Korea, Inc.
Founder & Leader of Silicon Valley Privacy-Preserving AI Forum
Advisor to KASPA of AI Semiconductor @ KASPA
Member of AI-Korean Medicine Integration Initiative Task Force
Global Leadership Initiative Fellow @ Salzburg Global Seminar
Visiting Professor & Advisory Professor @ Sogang Univ. & DGIST

# About Speaker

- *Co-Founder & CTO @ Erudio Bio, Inc., San Jose & Novato, CA, USA* 2023 ∼
- *Co-Founder & CEO @ Erudio Bio Korea, Inc., Korea* 2025 ∼
- *Leader of Silicon Valley Privacy-Preserving AI Forum (K-PAI), CA, USA* 2024 ∼
- *Advisor to Korean American Semiconductor Professional Alliance (KASPA)* 2026 ∼
- *CGO / Global Managing Partner @ LULUMEDIC, Seoul, Korea* 2025 ∼
- *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 ∼
- *AI-Korean Medicine Integration Initiative Task Force Member @ The Association of Korean Medicine, Seoul, Korea* 2025 ∼
- *Director of AI Semiconductor @ K-BioX, CA, USA* 2025 ∼
- Global Advisory Board Member @ Innovative Future Brain-Inspired Intelligence System Semiconductor of Sogang University, Korea 2020 ∼
- Technology Consultant @ Gerson Lehrman Gruop (GLG), NY, USA 2022 ∼
- Advisor @ CryptoLab, Inc., Seoul, Korea 2025 ∼

- Co-Founder & CTO / Head of Global R&D / Chief Applied Scientist / Senior Fellow @ Gauss Labs, Inc., Palo Alto, CA, USA 2020 $\sim$ 2023
- Senior Applied Scientist @ Amazon.com, Inc., Vancouver, BC, Canada 2017 $\sim$ 2020
- Principal Engineer @ Software R&D Center, Samsung Electronics 2016 $\sim$ 2017
- Principal Engineer @ Strategic Marketing & Sales, Memory Business 2015 $\sim$ 2016
- Principal Engineer @ DT Team, DRAM Development, Samsung 2012 $\sim$ 2015
- Senior Engineer @ CAE Team, Memory Business, Samsung, Korea 2005 $\sim$ 2012
- PhD - Electrical Engineering @ Stanford University, CA, USA 2001 $\sim$ 2004
- Development Engineer @ Voyan, Santa Clara, CA, USA 2000 $\sim$ 2001
- MS - Electrical Engineering @ Stanford University, CA, USA 1998 $\sim$ 1999
- BS - Electrical & Computer Engineering @ Seoul National University 1994 $\sim$ 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 Bio-Medical Landscape for MIT GSW 2026

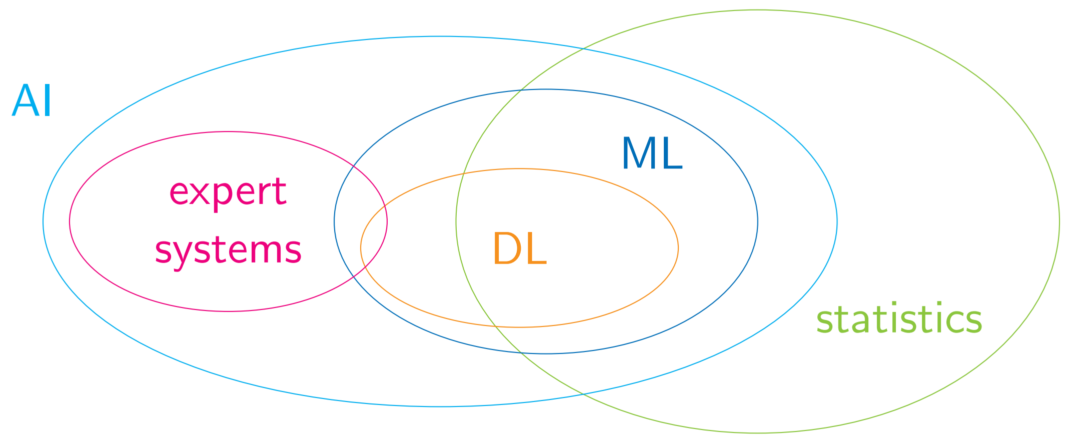
# Artificial Intelligence

# Definition and History

# Definition & relation to other technologies

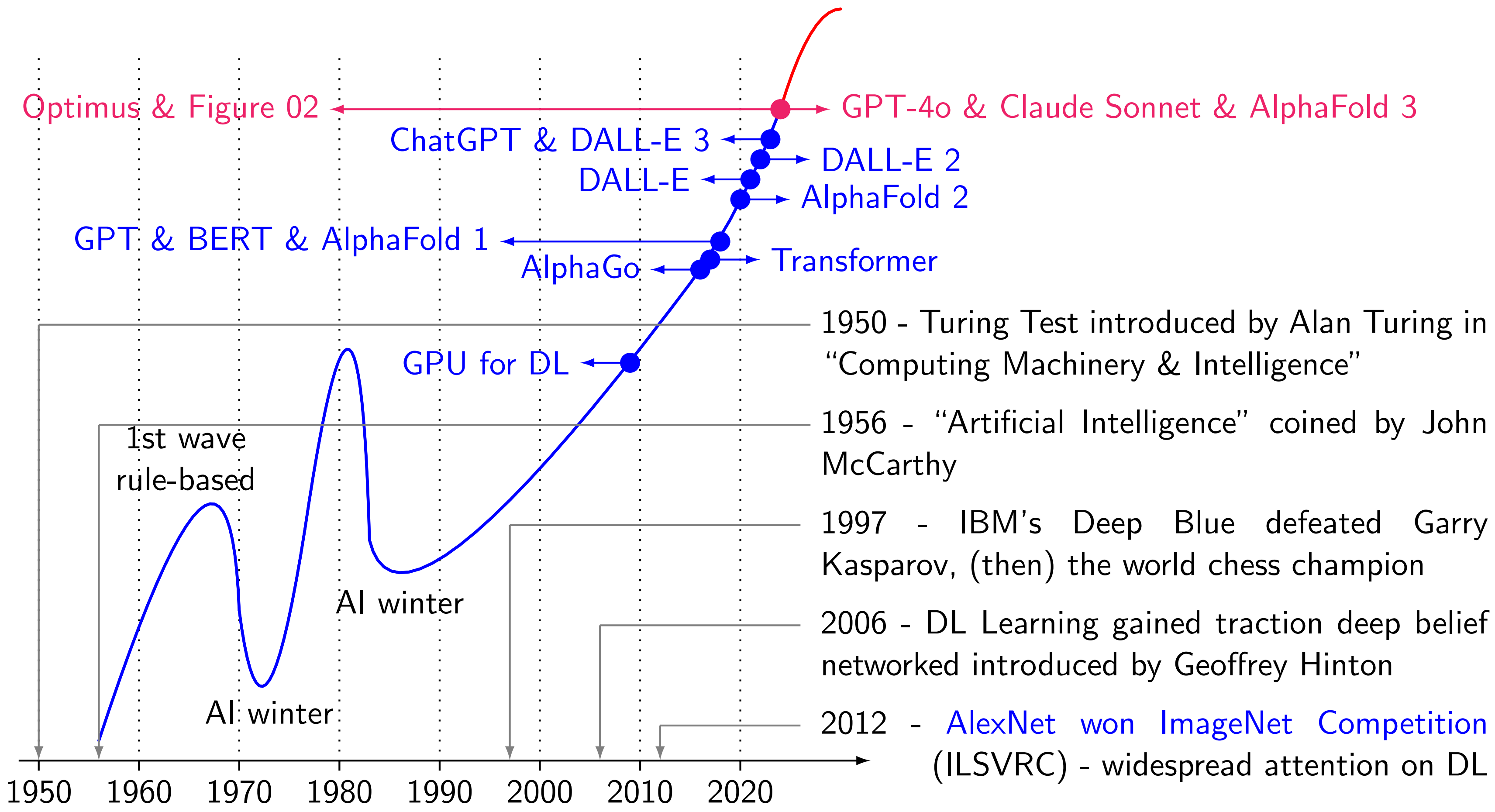
- AI
  - is technology doing tasks requiring human intelligence, such as learning, problem-solving, decision-making & language understanding
  - encompasses *range of technologies, methodologies, applications & products*

- AI, ML, DL, statistics & expert system[1] [HGH+22]



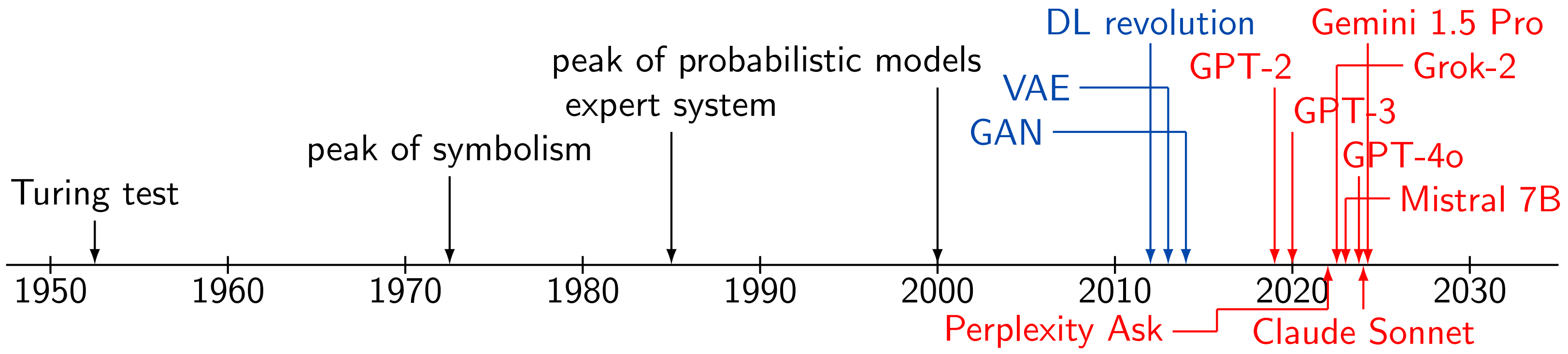

[1] ML: machine learning & DL: deep learning

# History


Optimus & Figure 02
GPT-4o & Claude Sonnet & AlphaFold 3
ChatGPT & DALL-E 3
DALL-E 2
DALL-E
AlphaFold 2
GPT & BERT & AlphaFold 1
AlphaGo
Transformer
GPU for DL
1st wave
rule-based
AI winter
AI winter
1950 - Turing Test introduced by Alan Turing in "Computing Machinery & Intelligence"
1956 - "Artificial Intelligence" coined by John McCarthy
1997 - IBM's Deep Blue defeated Garry Kasparov, (then) the world chess champion
2006 - DL Learning gained traction deep belief networked introduced by Geoffrey Hinton
2012 - AlexNet won ImageNet Competition (ILSVRC) - widespread attention on DL
1950
1960
1970
1980
1990
2000
2010
2020

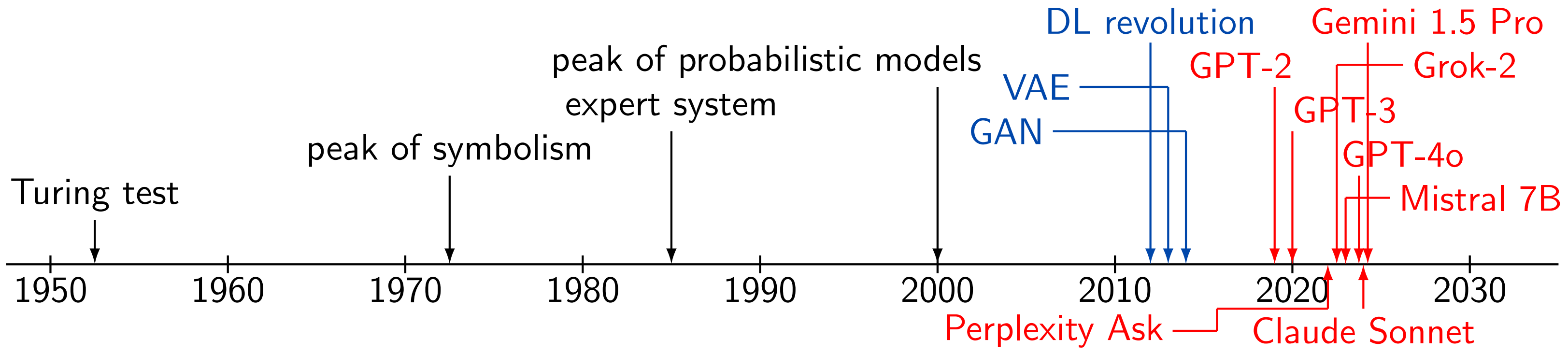
# Birth of AI - early foundations & precursor technologies

- 1950s ∼ 1970s
  - Alan Turing - concept of *"thinking machine" & Turing test* to evaluate machine intelligence (1950s)
  - *symbolists* (as opposed to connectionists) - early AI focused on symbolic reasoning, logic & problem-solving - Dartmouth Conference in 1956 by *John McCarthy, Marvin Minsky, Allen Newell & Herbert A. Simon*
  - precursor technologies - genetic algorithms (GAs), Markov chains & *hidden Markov models (HMMs)* - laying foundation for generative processes (1970s ∼)



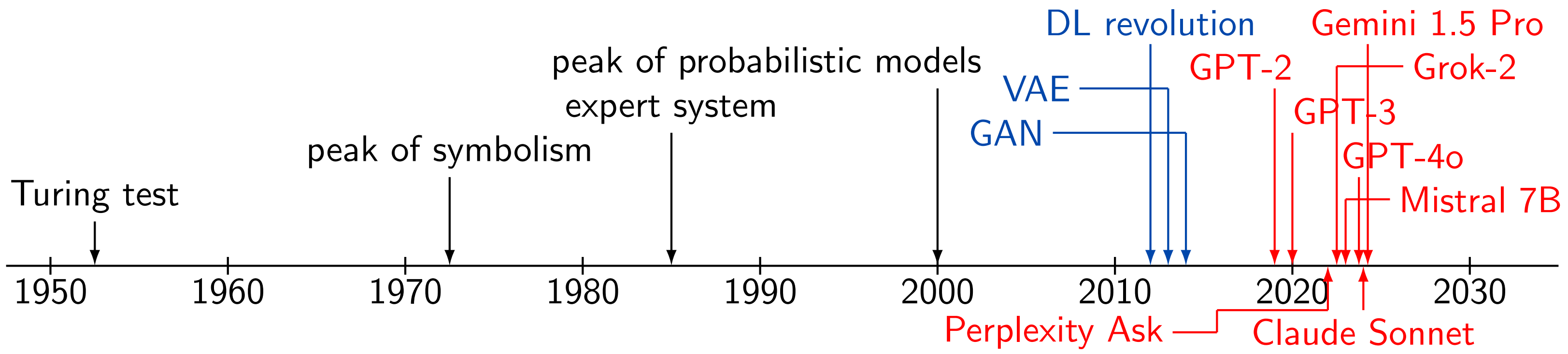
# Rule-based systems & probabilistic models

- 1980s $\sim$ early 2000s
  - *expert systems* (1980s) - AI systems designed to mimic human decision-making in specific domains
  - development of neural networks (NN) w/ backpropagation *training multi-layered networks* - setting stage for way more complex generative models
  - *probabilistic models* (including network models, *i.e.*, Bayesian networks) & Markov models - laying groundwork for data generation & pattern prediction



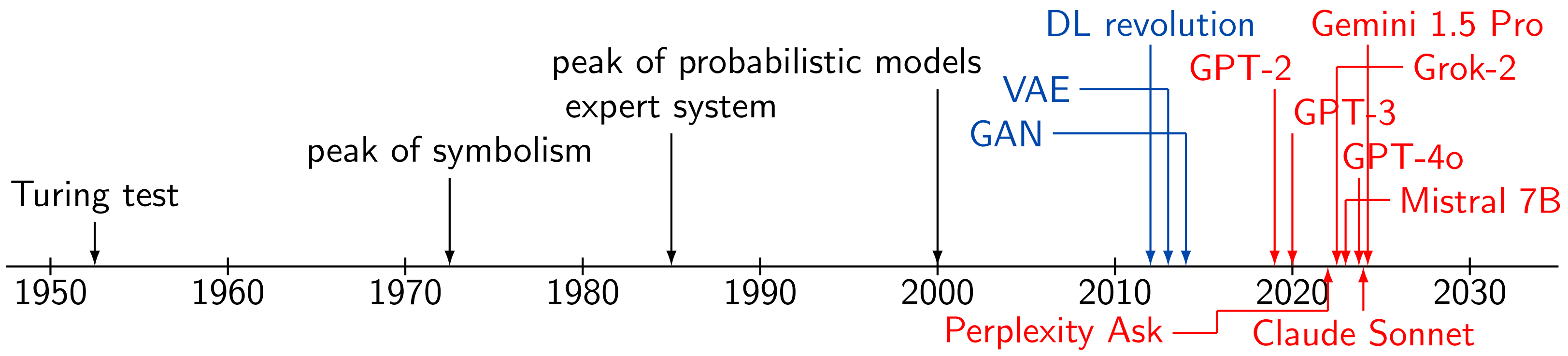
## Rise of deep learning & generative models

- 2010s - breakthrough in genAI
  - *deep learning (DL) revolution* - advances in GPU computing and data availability led to the rapid development of deep neural networks.
  - *variational autoencoder (VAE)* (2013) - by Kingma and Welling - learns mappings between input and latent spaces
  - *generative adversarial network (GAN)* (2014) - by Ian Goodfellow - game-changer in generative modeling where two NNs compete each other to create realistic data
    - widely used in image generation & creative tasks



## Transformer models & multimodal AI

- late 2010s $\sim$ Present
  - Transformer architecture (2017) - by Vaswani et al.
    - *revolutionized NLP,* $e.g.$, LLM & various genAI models
  - GPT series - generative pre-trained transformer
    - GPT-2 (2019) - generating human-like texts - *marking leap in language models*
    - GPT-3 (2020) - 175B params - set *new standards for LLM*
  - multimodal systems - DALL-E & CLIP (2021) - *linking text and visual data*
  - emergence of diffusion models (2020s) - new approach for generating high-quality images - progressively "denoising" random noise (DALL-E 2 & Stable Diffusion)



# Significant AI Achievements - 2014 – 2025

# Deep learning revolution

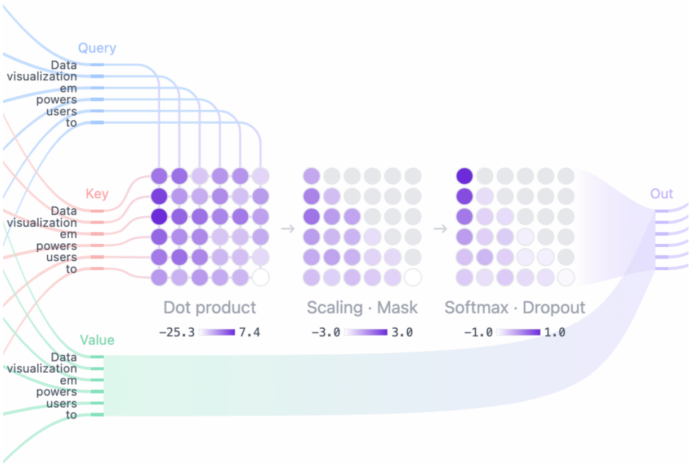
- 2012 – 2015 - DL revolution[2]
  - CNNs demonstrated exceptional performance in image recognition, *e.g.*, *AlexNet's victory in ImageNet competition*
  - widespread adoption of DL learning in CV transforming industries
- 2016 - AlphaGo defeats human Go champion
  - DeepMind's AlphaGo defeated world champion in Go, extremely complex game *believed to be beyond AI's reach*
  - significant milestone in RL - AI's potential in solving complex & strategic problems



[2] CV: computer vision, NN: neural network, CNN: convolutional NN, RL: reinforcement learning

# Transformer changes everything

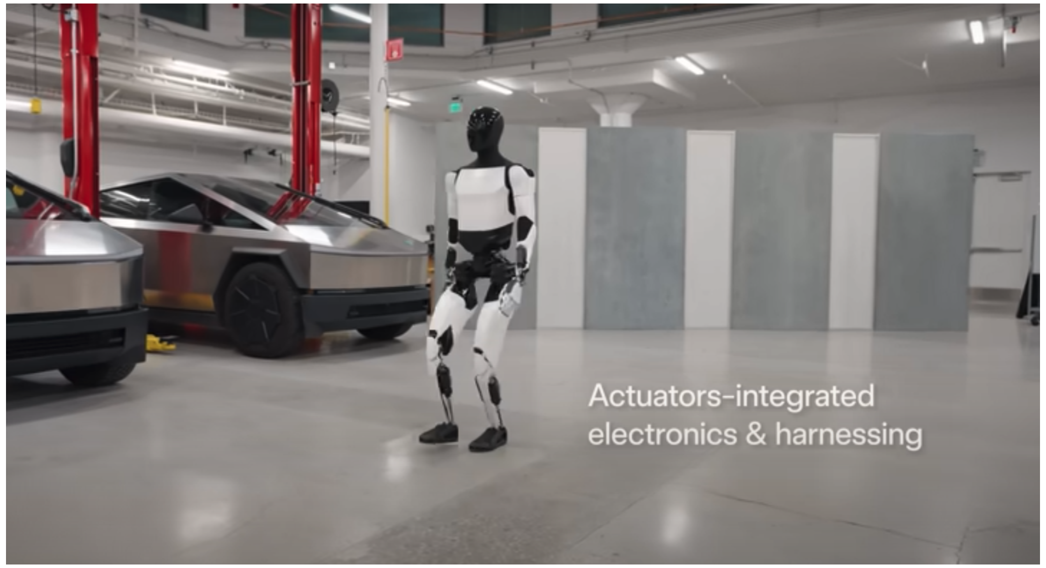
- 2017 – 2018 - Transformers & NLP breakthroughs[3]
  - *Transformer ($e.g.$, BERT & GPT) revolutionized NLP*
  - major advancements in, $e.g.$, machine translation & chatbots
- 2020 - AI in healthcare – AlphaFold & beyond
  - DeepMind's *AlphaFold solves 50-year-old protein folding problem* predicting 3D protein structures with remarkable accuracy
  - accelerates drug discovery and personalized medicine - offering new insights into diseases and potential treatments


[3] NLP: natural language processing, GPT: generative pre-trained transformer

# Lots of breakthroughs in AI technology and applications in 2024

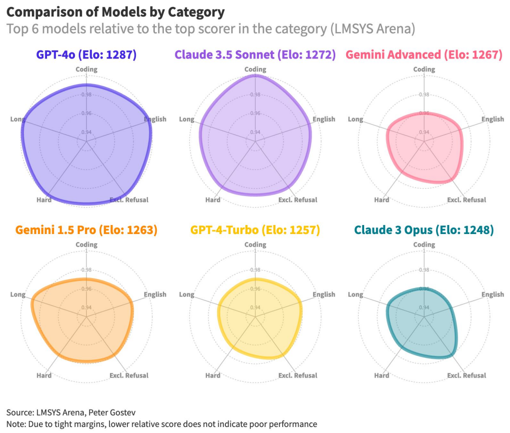
- proliferation of advanced AI models
  - GPT-4o, Claude Sonnet, Claude 3 series, Llama 3, Sora, Gemini
  - *transforming industries* such as content creation, customer service, education, *etc.*
- breakthroughs in specialized AI applications
  - Figure 02, Optimus, AlphaFold 3
  - driving unprecedented advancements in automation, drug discovery, scientific understanding - *profoundly affecting healthcare, manufacturing, scientific research*



# Major AI Breakthroughs in 2025

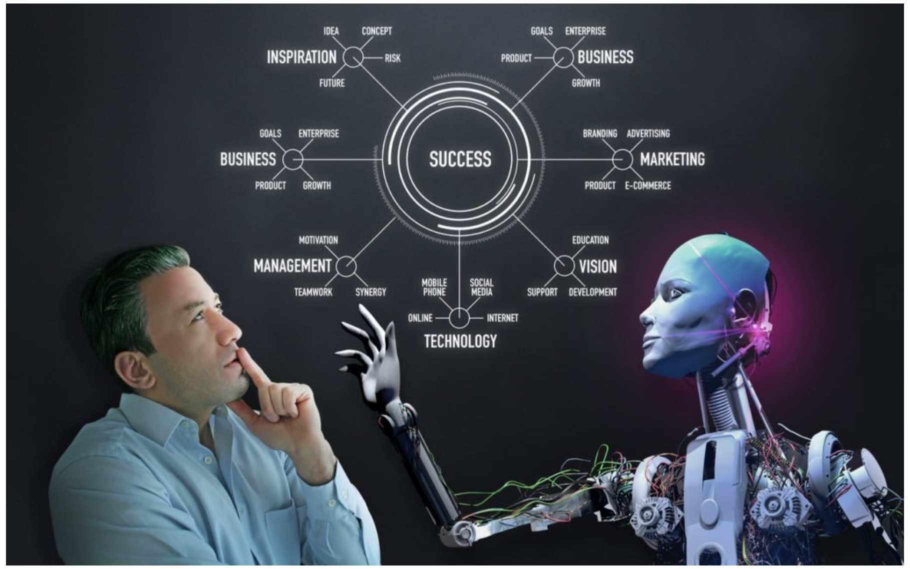
- next-generation foundation models
  - GPT-5 and Claude 4 demonstrate emergent reasoning abilities
  - open-source models achieving parity with leading commercial systems from 2024
- hardware innovations
  - NVIDIA's Blackwell successor architecture delivering 3-4x performance improvement
  - AMD's MI350 accelerators challenging NVIDIA's market dominance
- AI-human collaboration systems
  - seamless multimodal interfaces enabling natural human-AI collaboration
  - AI systems effectively explaining reasoning and recommendations
  - augmented reality interfaces providing real-time AI assistance in professional contexts

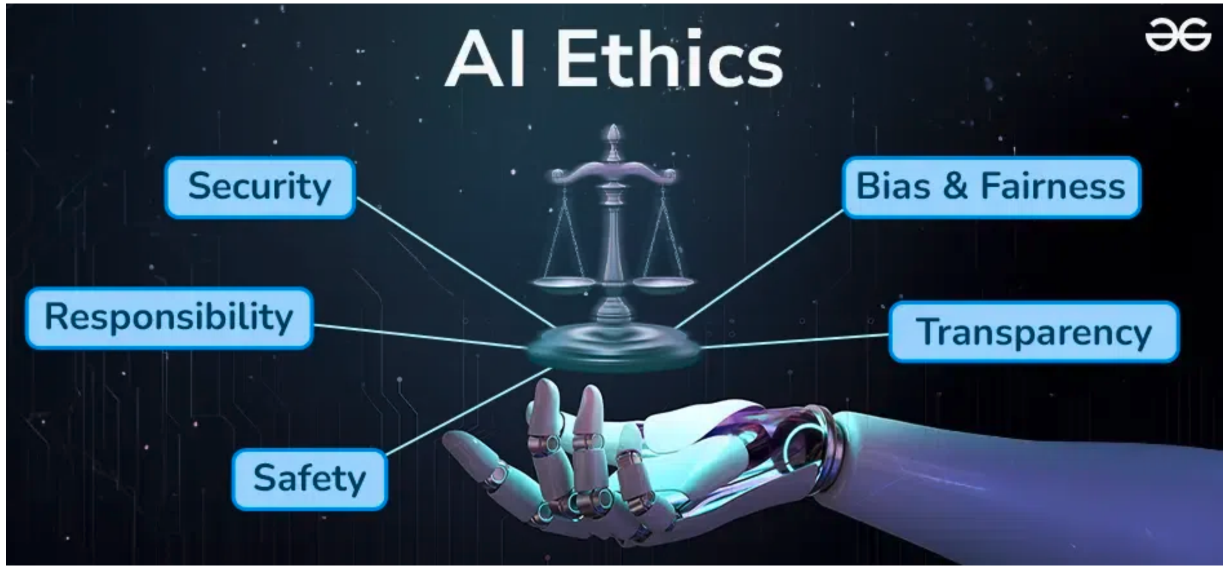

WHO WILL WIN?
V/S
V/S
Claude
ChatGPT
Bard


Comparison of Models by Category
Top 6 models relative to the top scorer in the category (LMSYS Arena)
GPT-4o (Elo: 1287)
Claude 3.5 Sonnet (Elo: 1272)
Gemini Advanced (Elo: 1267)
Gemini 1.5 Pro (Elo: 1263)
GPT-4-Turbo (Elo: 1257)
Claude 3 Opus (Elo: 1248)
Coding
English
Excl. Refusal
Hard
Long
1
0.98
0.96
0.94
Source: LMSYS Arena, Peter Gostev
Note: Due to tight margins, lower relative score does not indicate poor performance

# Transformative impact of AI - reshaping industries, work & society

- accelerating human-AI collaboration
  - not only reshaping industries but *altering how humans interact with technology*
  - AI's role as collaborator and augmentor redefines productivity, creativity, the way we address global challenges, *e.g.*, *sustainability & healthcare*
- AI-driven automation *transforms workforce dynamics* - creating new opportunities while challenging traditional job roles
- *ethical AI considerations* becoming central not only to business strategy, but to society as a whole - *influencing regulations, corporate responsibility & public trust*

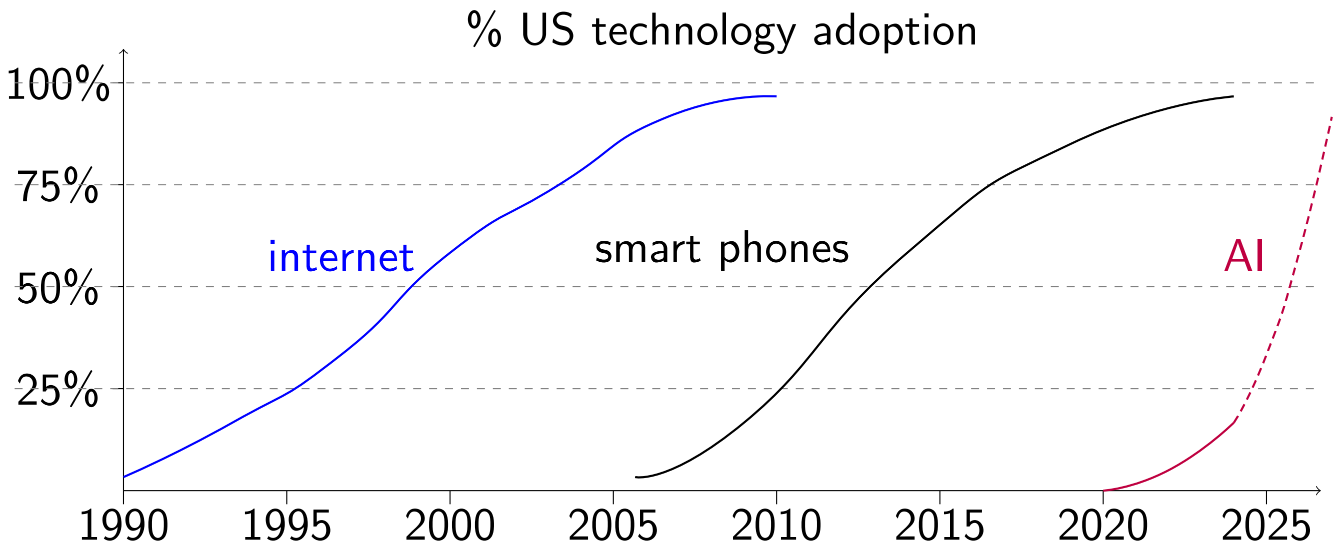


# Measuring AI’s Ascent

# Where are we in AI today?

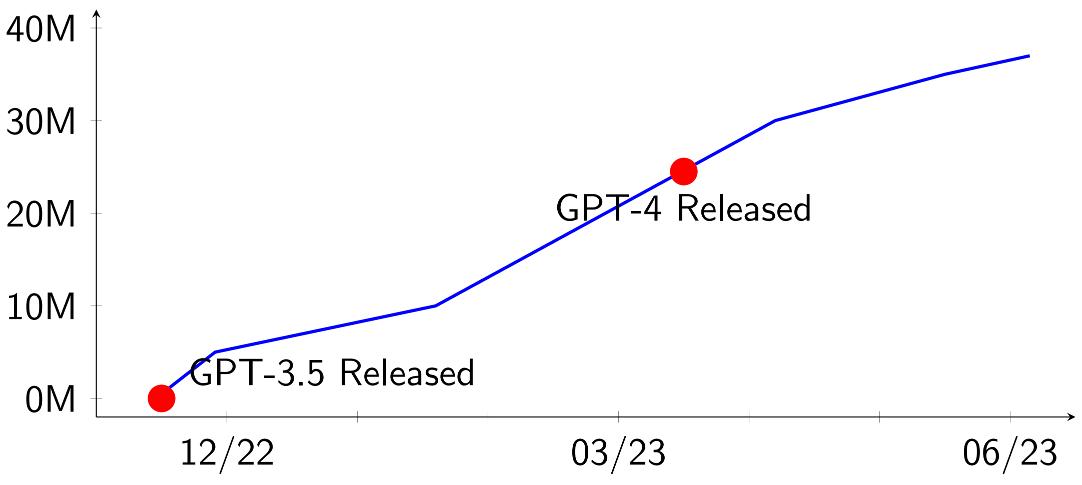
- sunrise phase - currently experiencing dawn of AI era with significant advancements and increasing adoption across various industries
- early adoption - in early stages of AI lifecycle with widespread adoption and innovation across sectors marking significant shift in technology's role in society



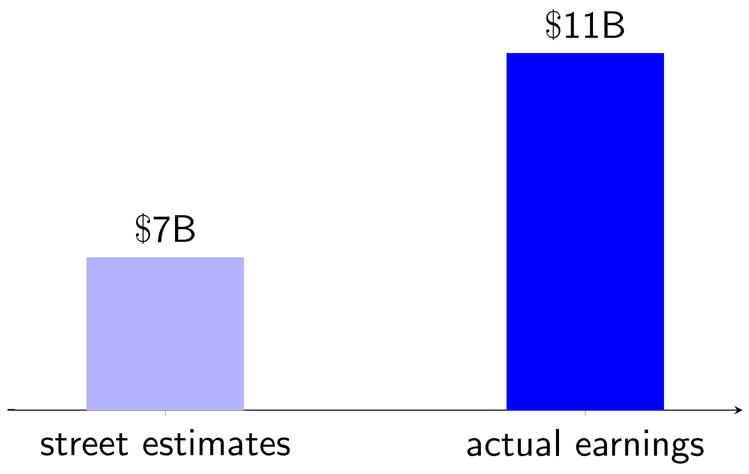
# Explosion of AI ecosystems - ChatGPT & NVIDIA

- took only *5 months for ChatGPT users to reach 35M*
- NVDIA 2023 Q2 earning exceeds market expectation by big margin - \$7B vs \$13.5B
  - surprisingly, *101% year-to-year growth*
  - even more surprisingly *gross margin was 71.2%* - up from 43.5% in previous year[4]

ChatGPT weekly unique visits




NVDA 2023-2Q earning

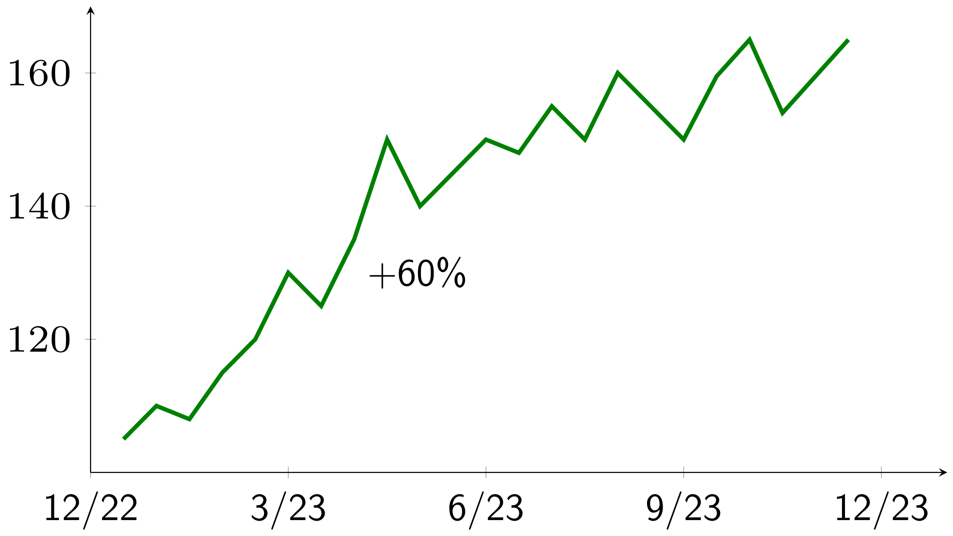



[4] source - Bloomberg

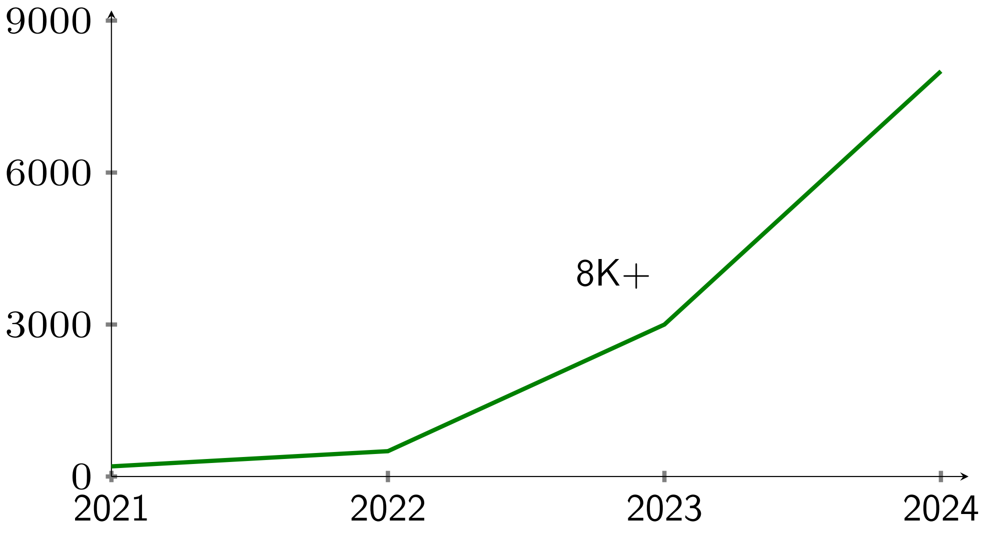
# Explosion of AI ecosystems - AI stock market

- *AI investment surge in 2023 - portfolio performance soars by 60%*
  - AI-focused stocks significantly outpaced traditional market indices
- *over 8,000 new AI applications* developed in last 3 years
  - applications span from healthcare and finance to manufacturing and entertainment

AI basket performance - 2023

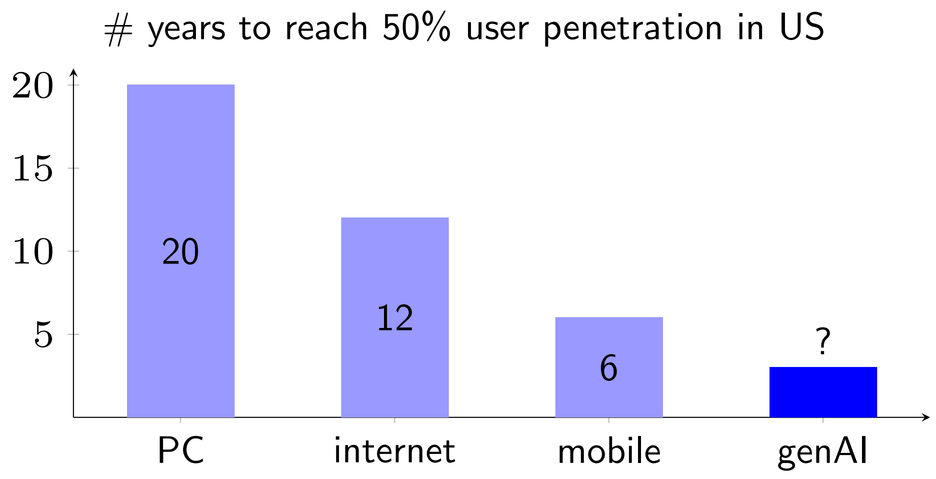


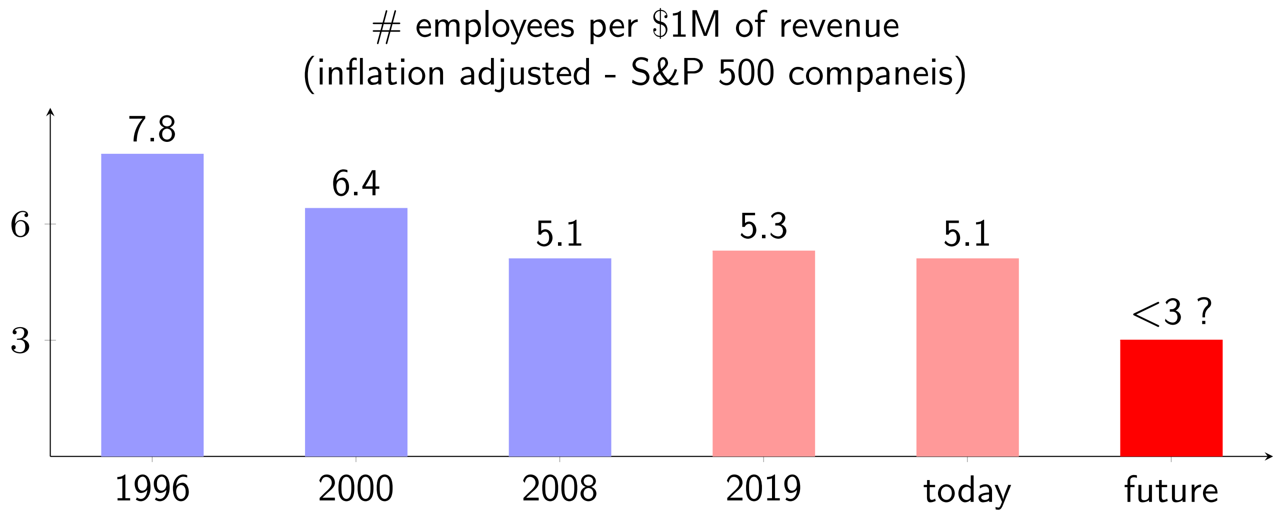

new AI applications over time



# AI's transformative impact - adoption speed & economic potential

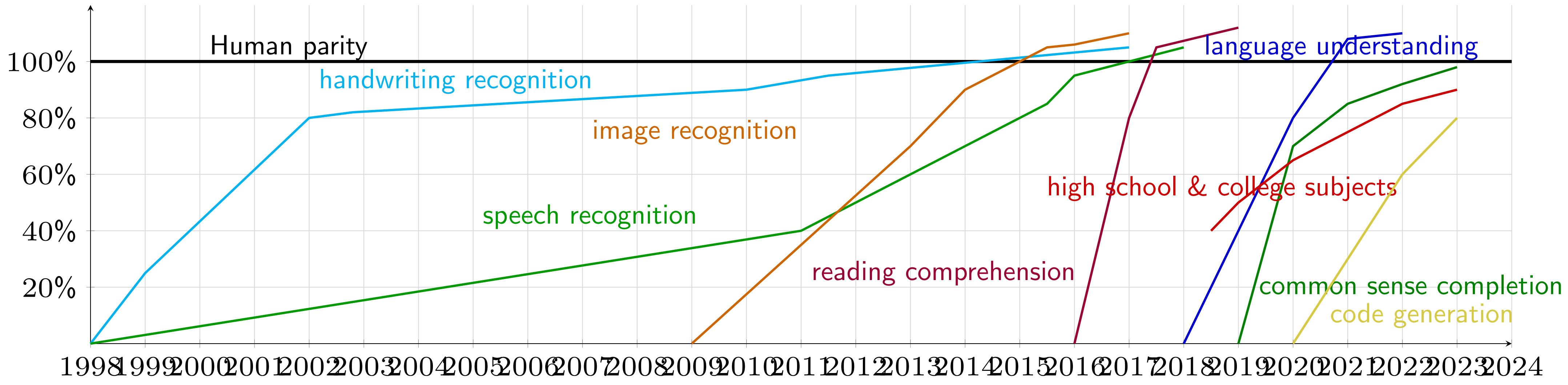
- adoption - has been twice as fast with platform shifts suggesting
  - increasing demand and readiness for new technology improved user experience & accessibility
- AI's potential to drive economy for years to come
  - 35% improvement in productivity driven by introduction of PCs and internet
  - greater gains expected with AI proliferation



# AI getting more & more faster

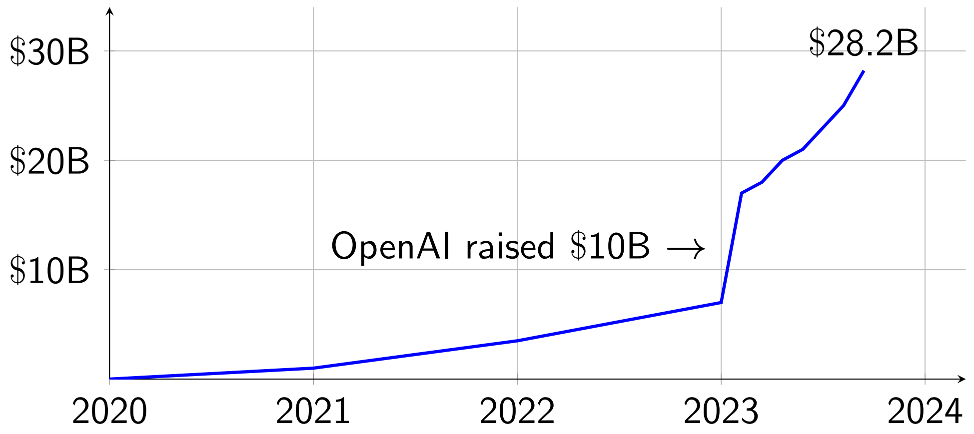
- steep upward slopes of AI capabilities highlight accelerating pace of AI development
  - period of exponential growth with AI potentially mastering new skills and surpassing human capabilities at ever-increasing rate
- closing gap to human parity - some capabilities approaching or arguably reached human parity, while others having still way to go
  - achieving truly human-like capabilities in broad range remains a challenge



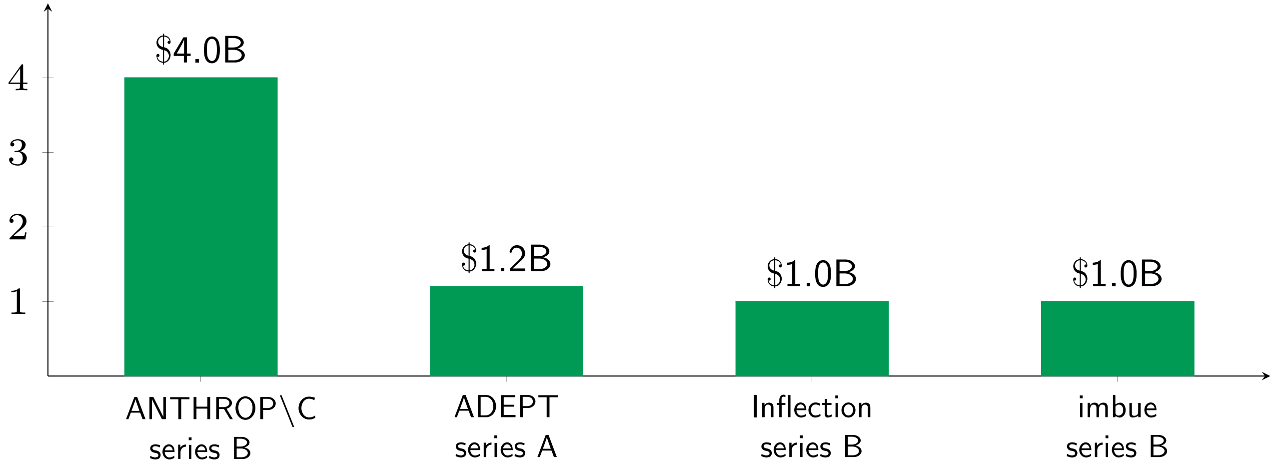
# Massive investment in AI

- *explosive growth* - cumulative funding skyrocketed reaching staggering $28.2B
- OpenAI - significant fundraising (=$10B) fueled rapid growth
- *valuation surge* - substantial valuations even before public products for stella companies
- *fierce competition for capital* among AI startups driving innovation & accelerating development
- massive investment indicates *strong belief in & optimistic outlook for potential of AI* to revolutionize industries & drive economic growth

cumulative funding in private AI companies since 2020

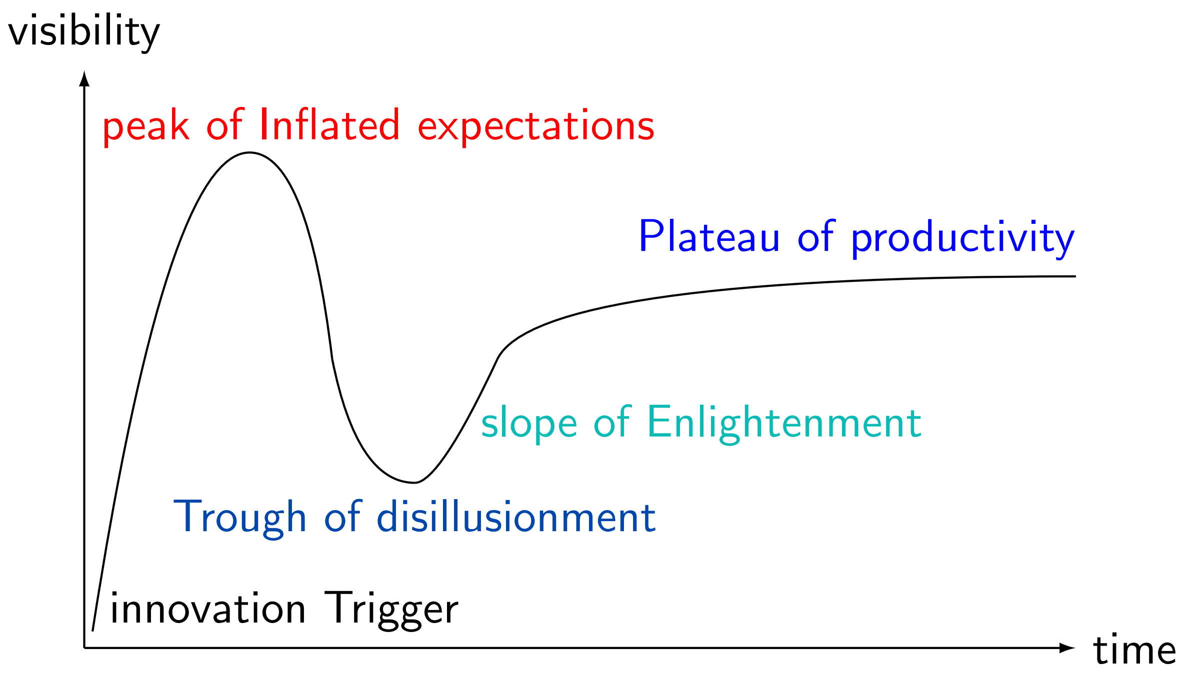



valuation before public product launch



# Is AI hype?

# Technology hype cycle




- innovation trigger - technology breakthrough kicks things off
- peak of inflated expectations - early publicity induces many successes followed by even more
- trough of disillusionment - expectations wane as technology producers shake out or fail
- slope of enlightenment - benefit enterprise, technology better understood, more enterprises fund pilots

# Fiber vs cloud infrastructure

- fiber infrastructure - 1990s
  - Telco Co's raised $1.6T of equity & $600B of debt
  - bandwidth costs decreased 90% within 4 years
  - companies - Covage, NothStart, Telligent, Electric Lightwave, 360 networks, Nextlink, Broadwind, UUNET, NFS Communications, Global Crossing, Level 3 Communications
  - became *public good*
- cloud infrastructure - 2010s
  - entirely new computing paradigm
  - mostly public companeis with data centers
  - *big 4 hyperscalers generate* $150B + annual revenue

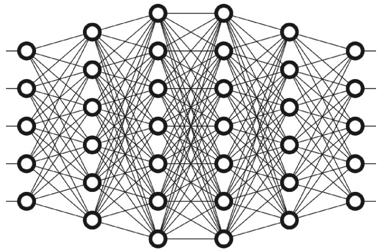
# Yes & No

| characteristics of hype cycles | speaker's views |
|---|---|
| value accrual misaligned with investment | • OpenAI still operating at a loss; business model *still* not clear<br>• gradual value creation across broad range of industries and technologies ($e.g.$, CV, LLMs, RL) unlike fiber optic bubble in 1990s |
| overestimating timeline & capabilities of technology | • self-driving cars delayed for over 15 years, with limited hope for achieving level 5 autonomy<br>• AI, however, has proven useful within a shorter 5-year span, with enterprises eagerly adopting |
| lack of widespread utility due to technology maturity | • AI already providing significant utility across various domains<br>• vs quantum computing remains promising in theory but lacks widespread practical utility |

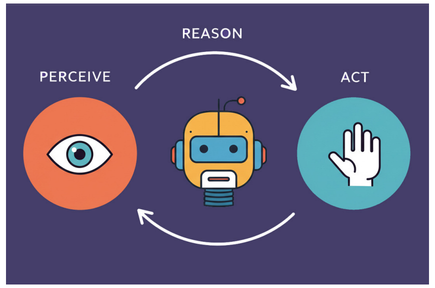
# AI Agents

# 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)

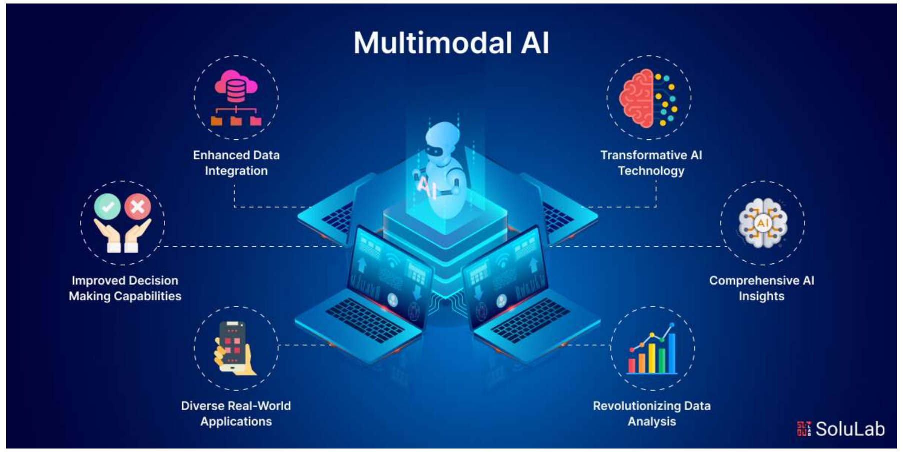



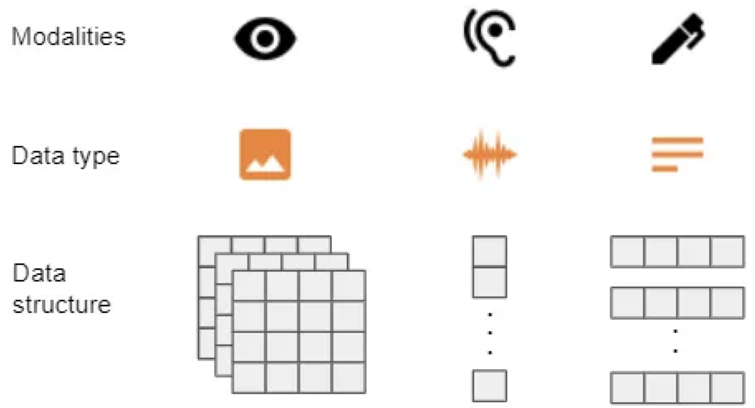




# Multimodal learning

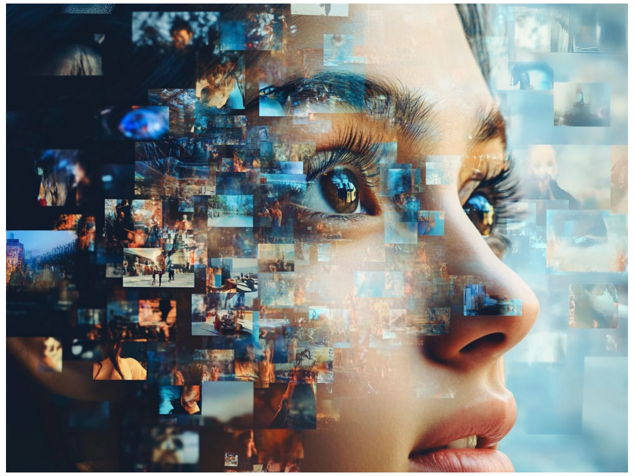
- understand information from multiple modalities, $e.g.$, text, images, audio, video
- representation learning methods
  - combine multiple representations or learn multimodal representations simultaneously
- applications
  - images from text prompt, videos with narration, musics with lyrics
- collaboration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

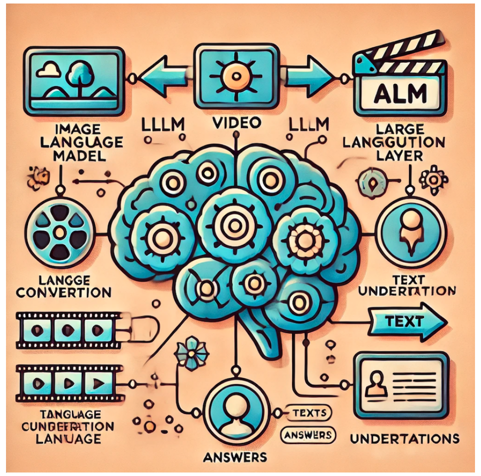


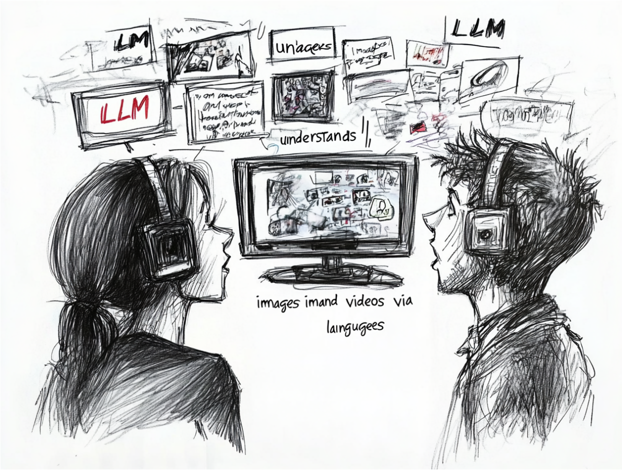



# Implications of success of LLMs

- many researchers change gears towards LLM
  – from computer vision (CV), speach, 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)*

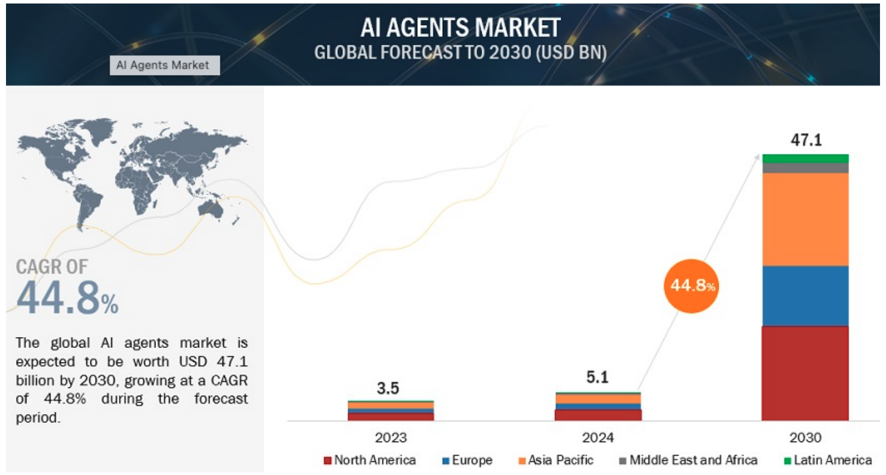


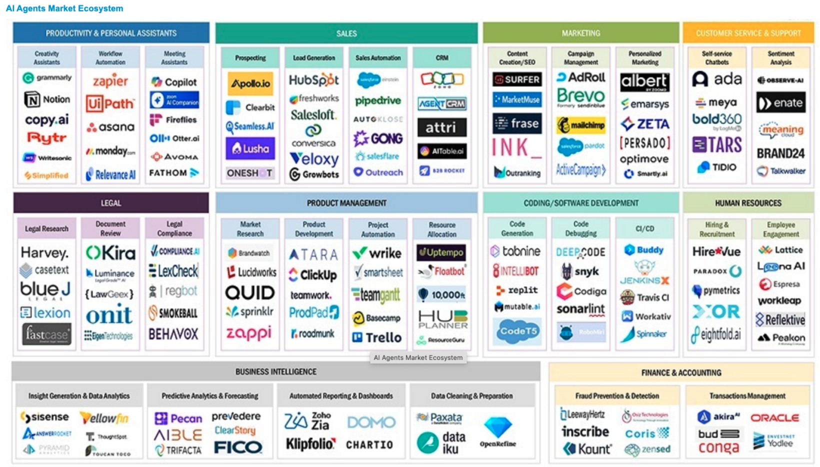


# Multimodal AI (mmAI)

- 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



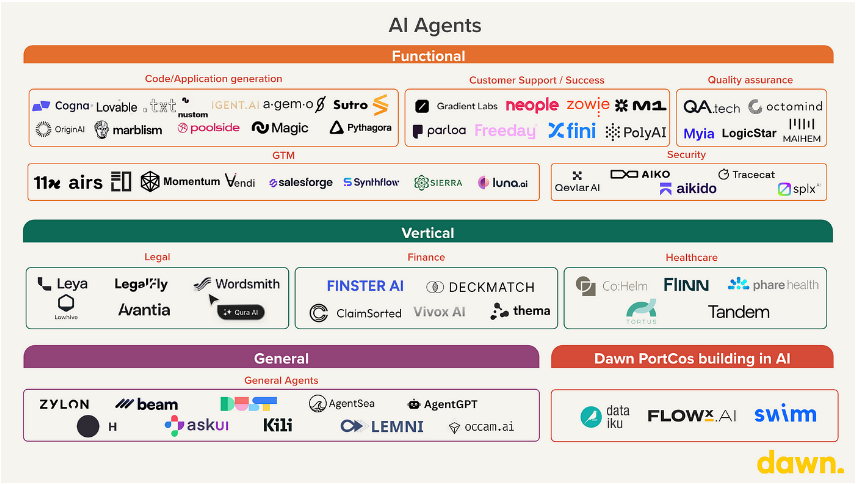



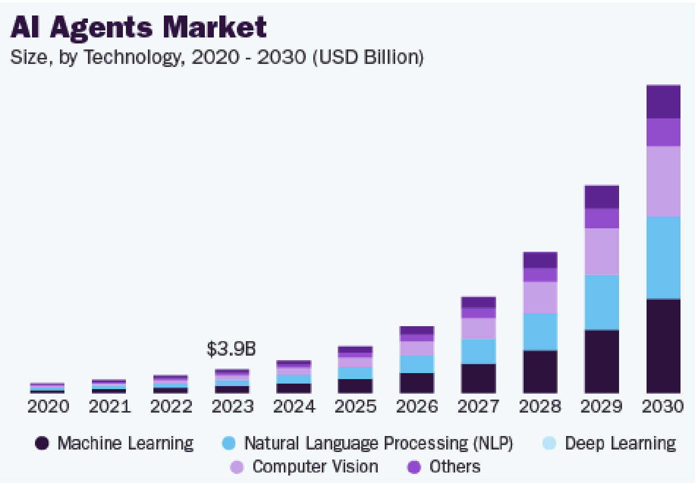
# mmAI Technology

- 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



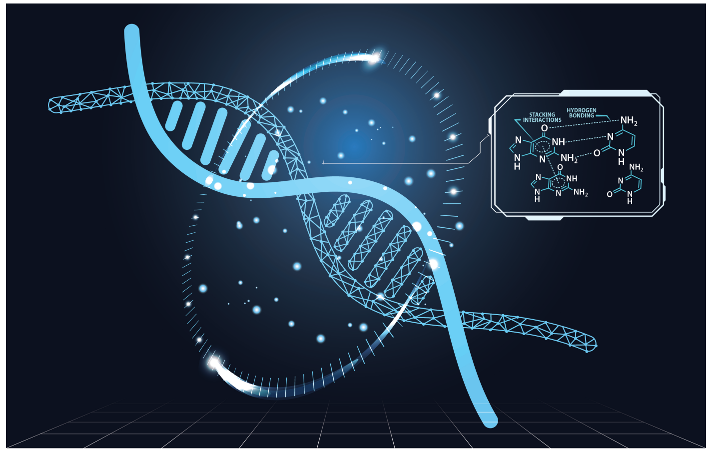
# AI agents - Present & Future

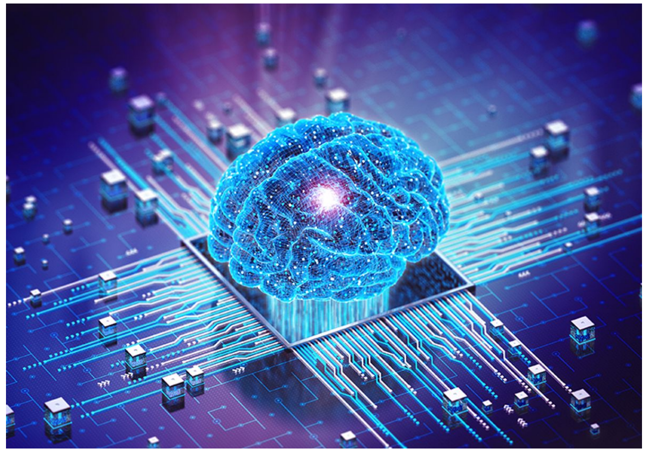
- emerging applications
  - scientific research - agents analyzing & running experiments & generating hypotheses
  - creative collaboration - AI partners in design & art combining multiple mediums
  - environmental monitoring - processing satellite sensor data for climate analysis
  - healthcare - enhanced 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

# 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





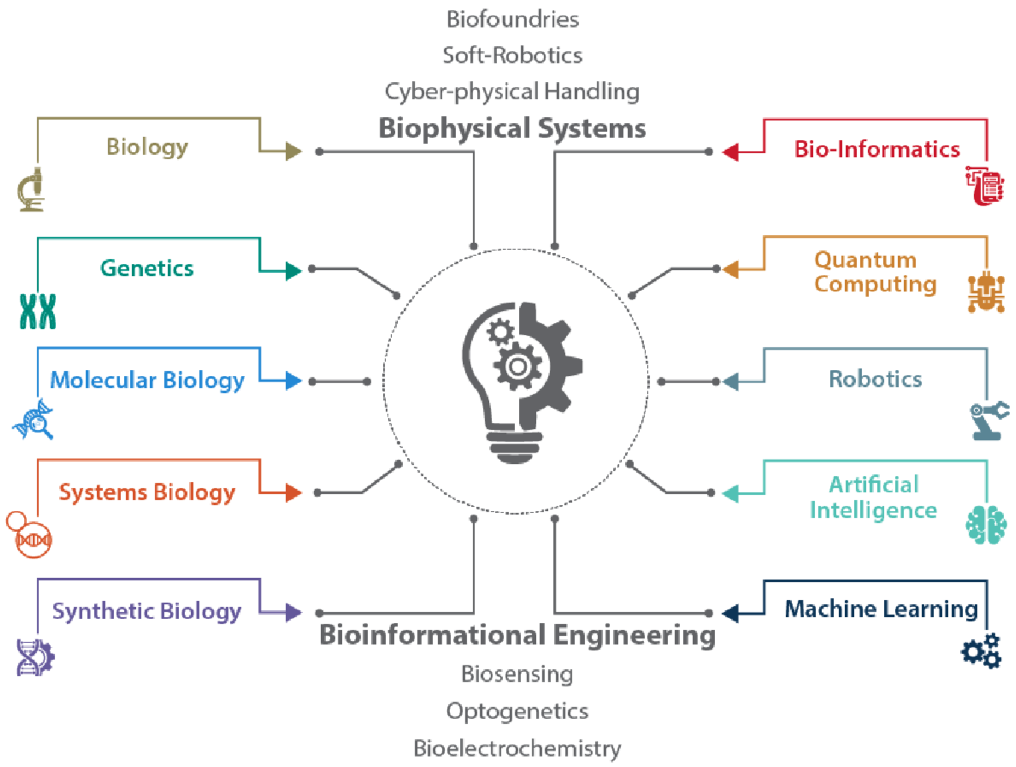
# Biotech

# 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 take 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]
  - *AI*, 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

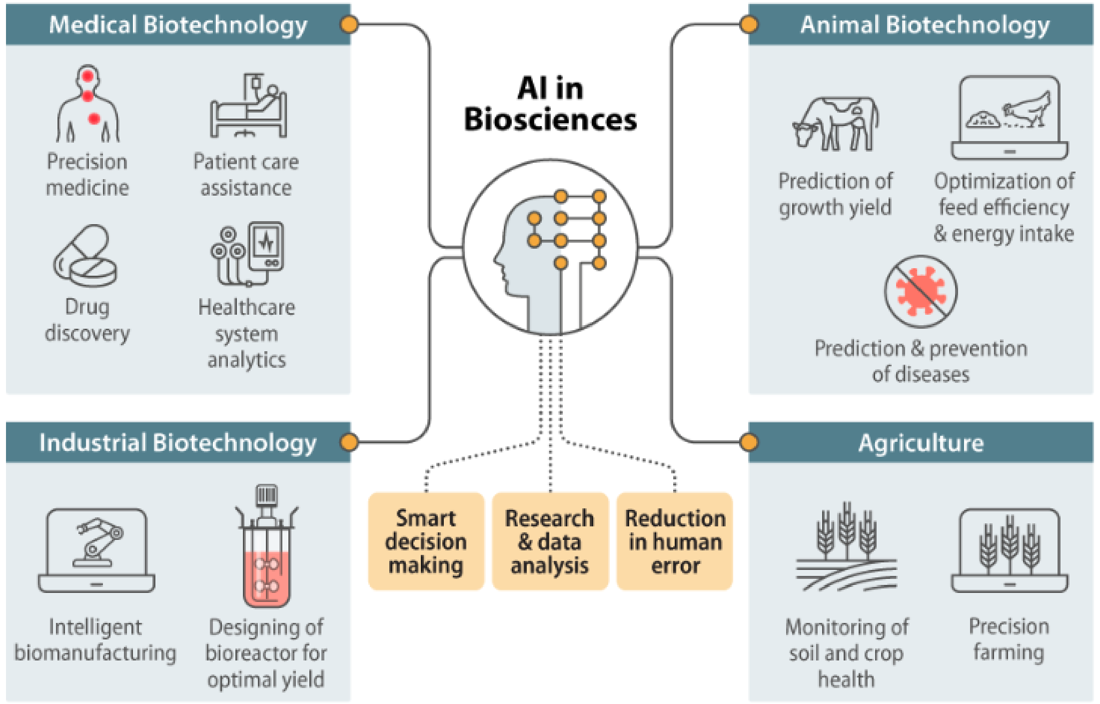
- sciences and technologies enabling biotechnology include (but not limited to)
  - (molecular) biology, genetics, systems biology, synthetic biology, bio-informatics, quantum computing, robotics [DFJ22]



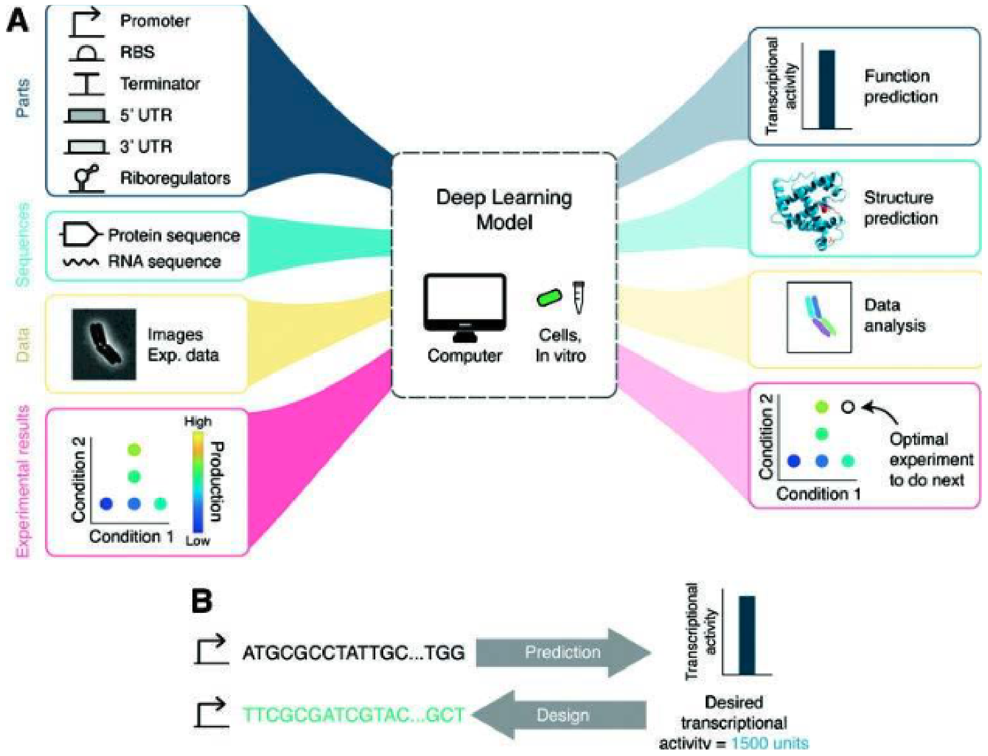
# Convergence of AI and biological design

- AI & biological sciences converging [BKP22]
  - each building upon the other's capabilities for new research and development across multiple areas
- Demis 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, 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)

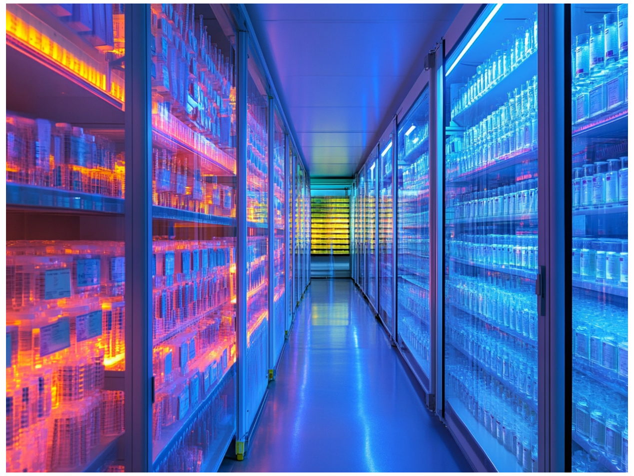
  – process huge amounts of biological data, *e.g.*, genetic sequence data, coming from different biological sources for understanding complex biological systems

    - 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 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



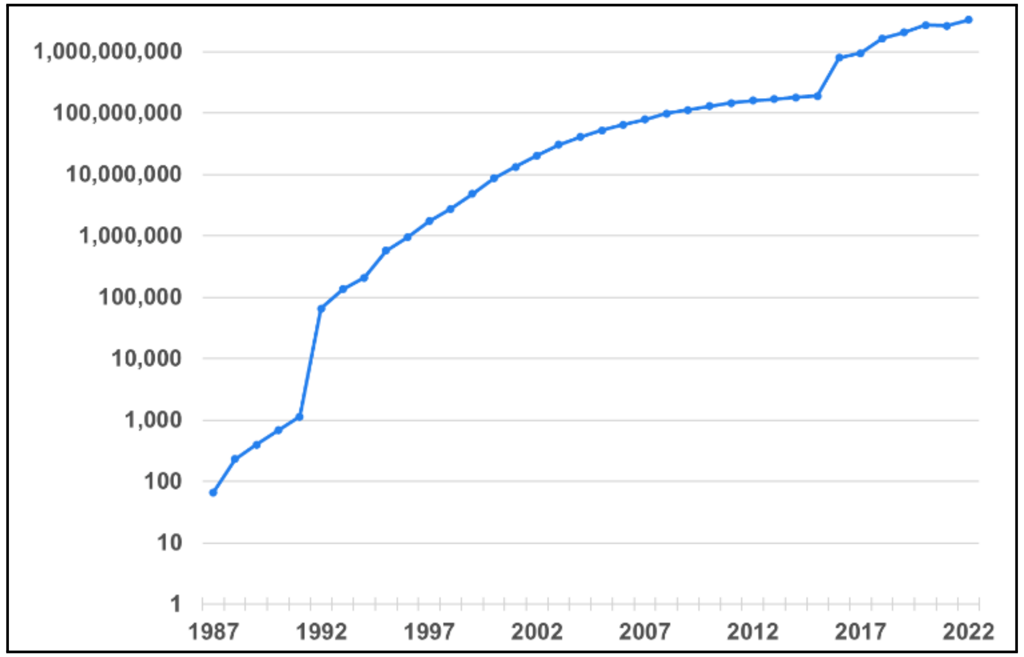
# Rapid growth of biological data

- volume of genetic sequence data grown exponentially as sequencing technology 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 - more than *terabyte of three-dimensional structure data* for biological molecules, *e.g.*, proteins, DNA, RNA
  - proprietary database
    - Gingko Bioworks - 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

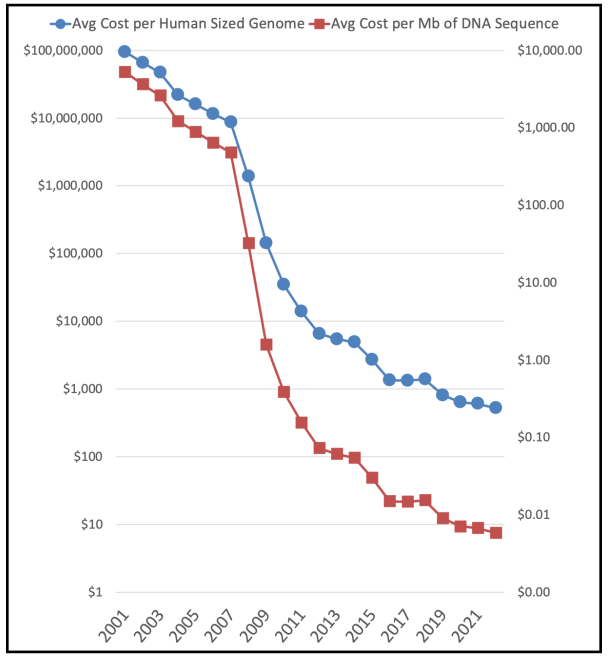
# Volume and sequencing cost of DNA over time

- volume of DNA sequences & DNA sequencing cost
  - data source: National Human Genome Research Institute (NHGRI) [Wet23] & International Nucleotide Sequence Database Collaboration (INSDC)
- *more dramatic than Moore's law!*

# sequences in INSDC




DNA sequencing cost



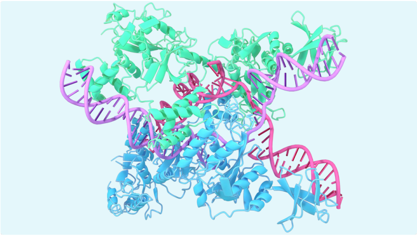
# Bio data availability and bias

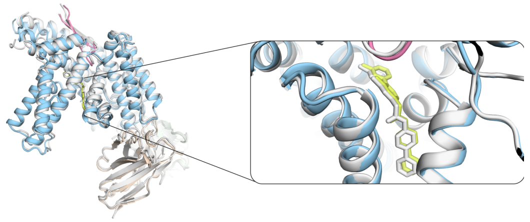
- US National Security Commission on Artificial Intelligence (NSCAI) recommends
  - US fund and prioritize development of a biobank containing *"wide range of high-quality biological and genetic data sets securely accessible by researchers"*
  - establishment of database of broad range of human, animal, 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

# AlphaFold

- solving 50-year-old protein folding problem, *"one of biology's grand challenges"*
  - definition - given amino acid sequence, predict how it folds into a 3D structure
  - proteins fold in microseconds, but predicting computationally nearly impossible
- AlphaFold 1 (2018) - DL + physics-based energy functions → AlphaFold 2 (2020) - attention-based NN solving protein folding "in principle" → AlphaFold 3 (2024) - diffusion-based DL, drug-protein interactions, protein complexes
- AlphaFold protein structure database
  - >200MM protein structures - nearly every known protein, used by >2MM researchers
- Applications & implications
  - drug discovery - target identification, lead optimization, side effect prediction
  - enzyme engineering, agriculture, environmental, vaccine development

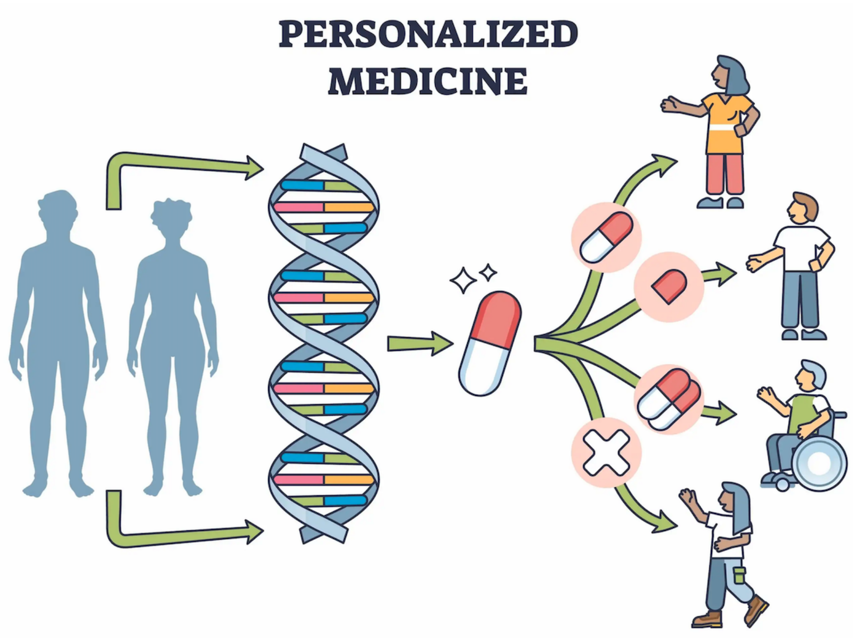


# AlphaGo

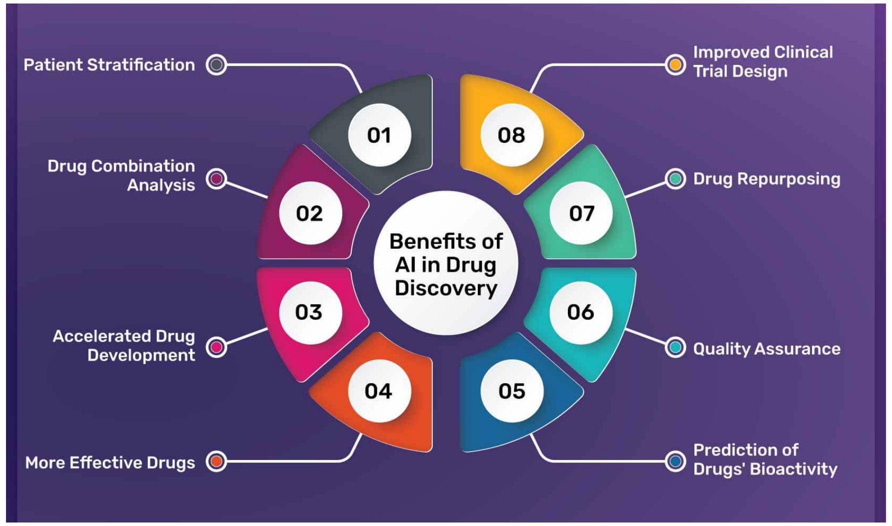
- Go board marked with 19×19 grid - $(19 \times 19)! = 361! \approx 1.44 \times 10^{768}$
  1437923258884890654832362511499863354754907538644755876127282765299227795534389618856841908003141196
  0714137944348905859683839682333043216077138088370565578796691924861827097800358990211005794501073330
  5079262777172275041226808677528136885057526541812043502150623466302643442673632627092764643302557772
  2695595343233942204301825548143785112222186834487969871267194205609533306413935710635197200721473378
  7338269803085351043174203653673779887217565513450041291061650506154496265581102824241428406627054585
  5623101563752892899924857388316647687165212001536218913733713768261861456295440900774337589490771443
  9917299937133680728459000034496420337066440853337001284286412654394495050773954560000000000000000000
  000000000000000000000000000000000000000000000000000000000000000000000
- deep reinforcement learning with Monte Carlo tree search
  - trained on thousands of years of Go game history
  - AlphaGo Zero learns by playing against itself
- development experience, insight, knowledge, know-how transferred to AlphaFold

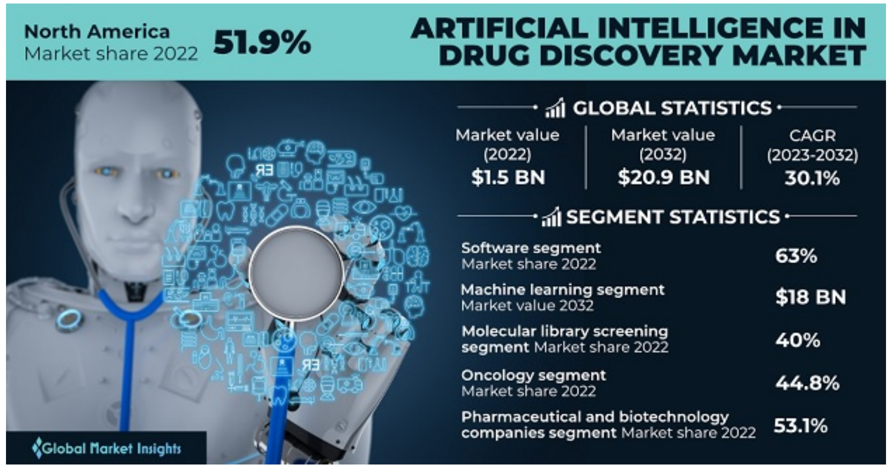

# Personalized medicine

- *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

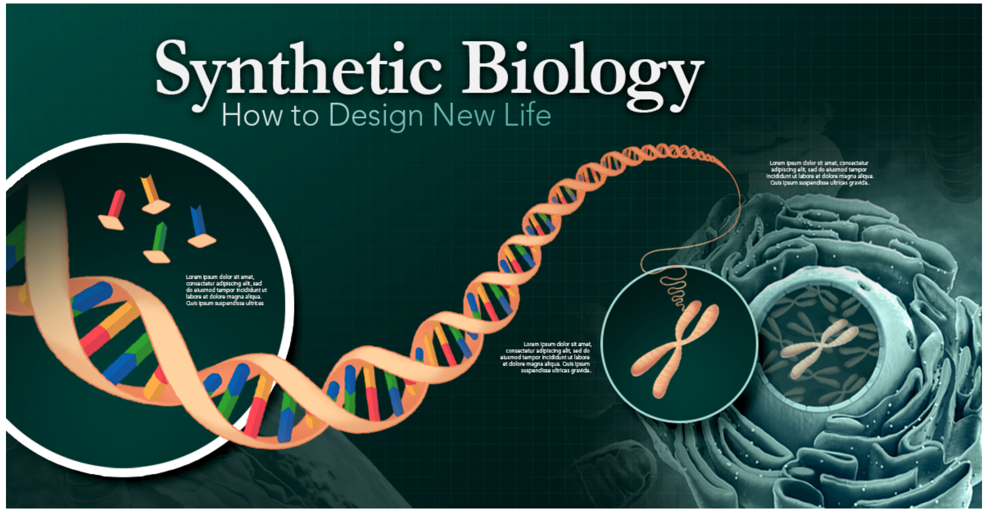


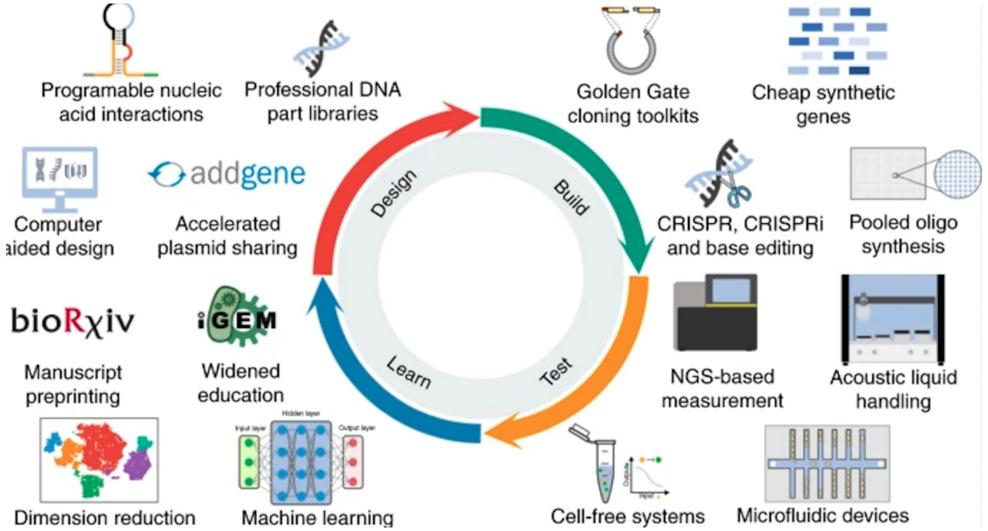




- traditional drug discovery process - time-consuming and costly often taking decades and billions of dollars
- AI streamlines this process by predicting the efficacy and safety of potential compounds with more speed and accuracy
- AI models analyze chemical databases to identify new drug candidates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

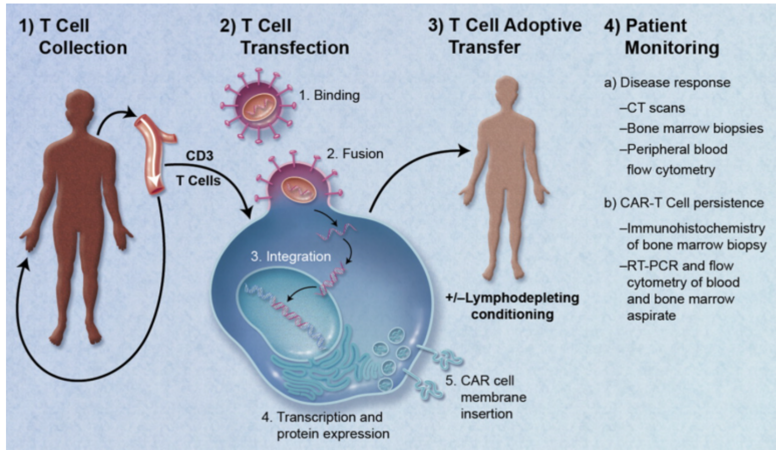
# Synthetic biology

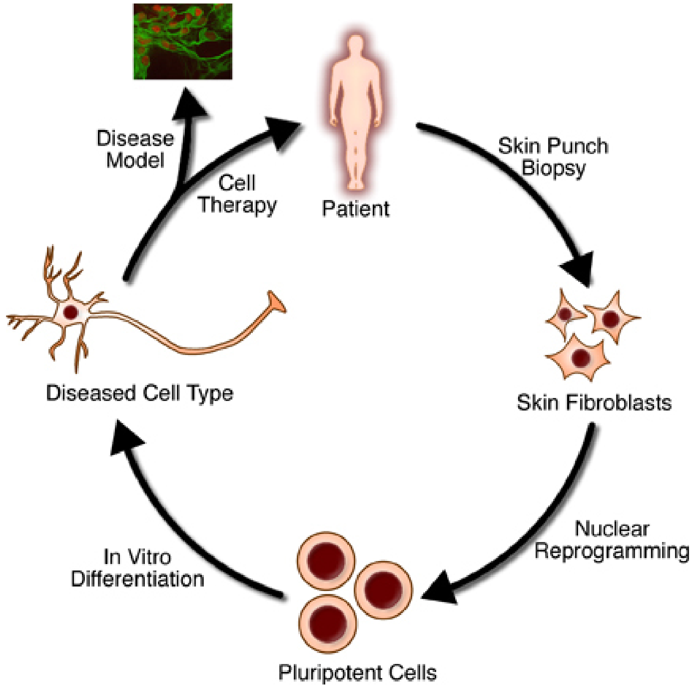
- use AI for gene editing, biomaterial production and synthetic pathways
- combine principles of biology and engineering to design and construct new biological entities
- AI optimizes synthetic biology processes from designing genetic circuits to scaling up production
- company - Ginkgo Bioworks uses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



# Regenerative medicine

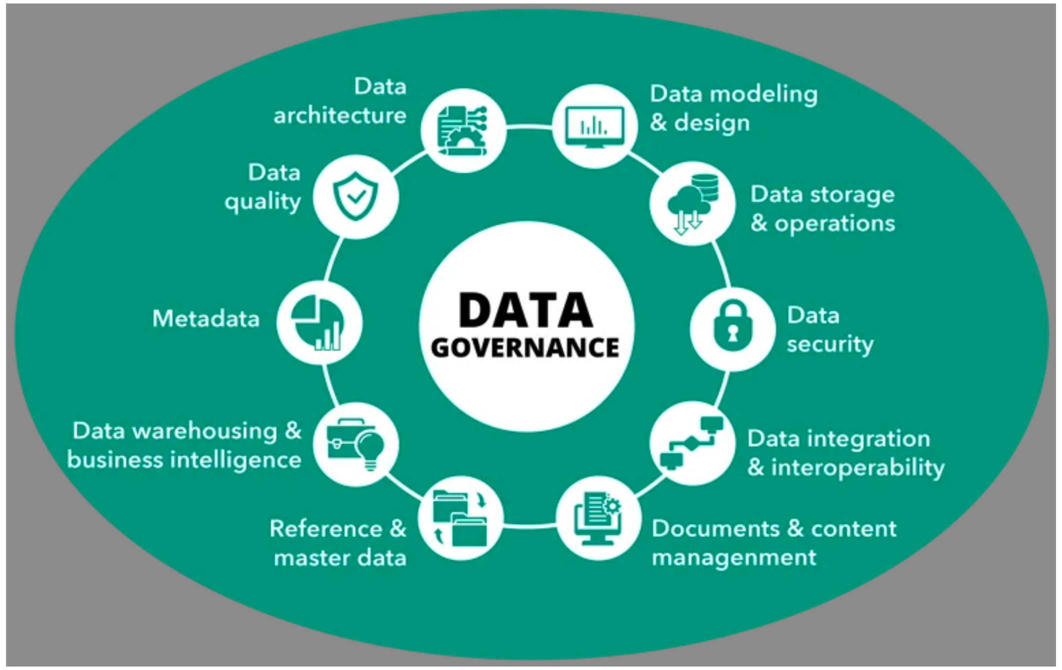



- AI advances development of stem cell therapies & tissue engineering
- AI algorithms assist in identifying optimal cell types, predicting cell behavior & personalized treatments
- particularly for conditions such as neurodegenerative diseases, heart failure and orthopedic injuries
- company - Organovo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

# Bio data integration

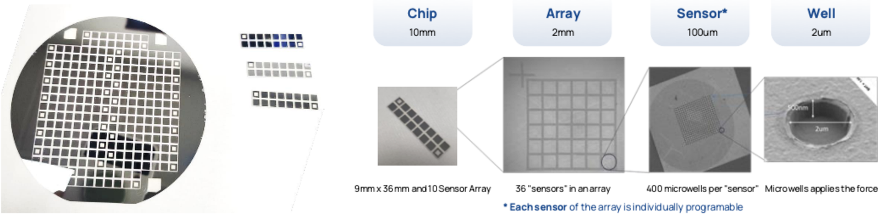
- integration of disparate data sources, including genomic, proteomic & clinical data - one of biggest challenges in biotech & healthcare
- AI delivers meaningful insights *only when* seamless data integration and interoperability realized
- developing platforms facilitating comprehensive, longitudinal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



# 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 ($\sim 10\mu$L)
  - measuring not just presence, but *strength* and *kinetics* of molecular interactions
- *dual business model*
  - diagnostics - multi-cancer biomarker detection with clinical institutions & hospitals (Seoul National University Hospital Bundang, Keimyung University Dongsan Hospital)
  - drug discovery - bioTCAD$^{\mathrm{TM}}$ 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 and expanding pharma partnerships

**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*



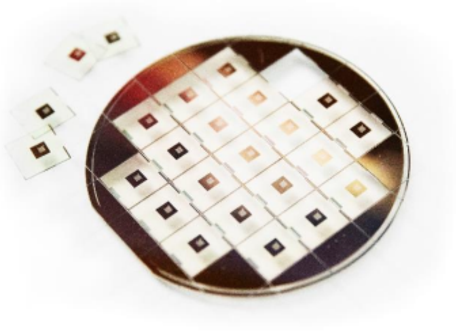

# Erudio Bio starting Revolution

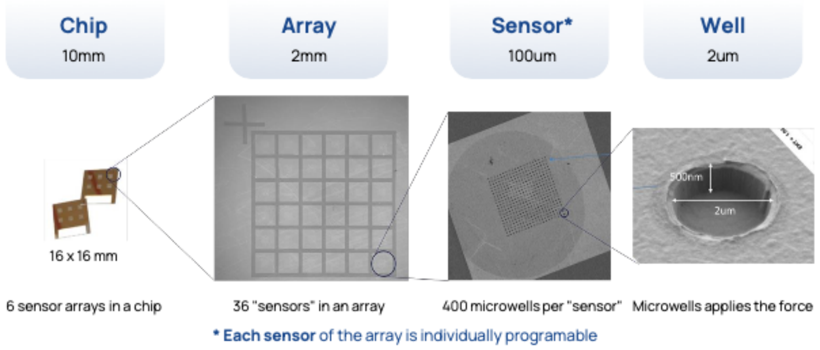
## Erudio Bio starts revolution with Gates Foundation's support

- 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

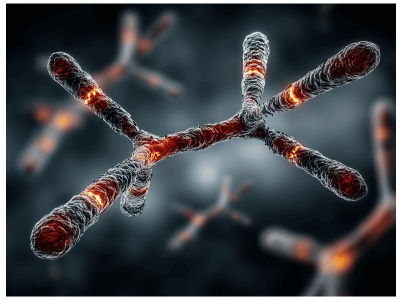




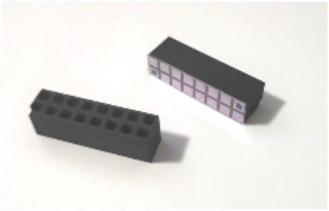

- generates *1000x more data* than the prevailing technology
  - scalable multi-omic microarray sensor
- *21 patents* in US, Canada, China, and Europe
- indicates how good the data is in real time
  - patented "dynamic force spectroscopy" and "powerful Bayesian inference" method provides our data *quality 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 in multiplex formats
    - 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

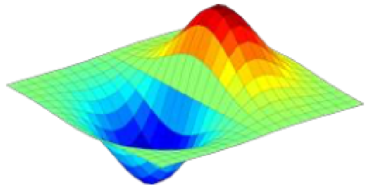
# Erudio Bio's business models



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

# Milestones

# Erudio Bio key milestones

- IP developed at Stanford School of Medicine under mentorship by father of microarrays - Dr. Ronald Davis
- data vetted by Analog Devices ($115B market cap semiconductor company), Harvard Medical School & Massachusetts General Hospital
- commercial partnership with Analog Devices to manufacture at scale
- JDA with Shanghai General Hospital's National Clinical Research Center for Eye Diseases
- JDA with Seoul National University General Hospital (SNUBH) for multiplexed cancer biomarker detections
- partnerships with
  - Keimyung University Dongsan Hospital, KRIBB, KAIST NanoFab, Lulumedic

HARVARD MEDICAL SCHOOL

Stanford MEDICINE

ANALOG DEVICES

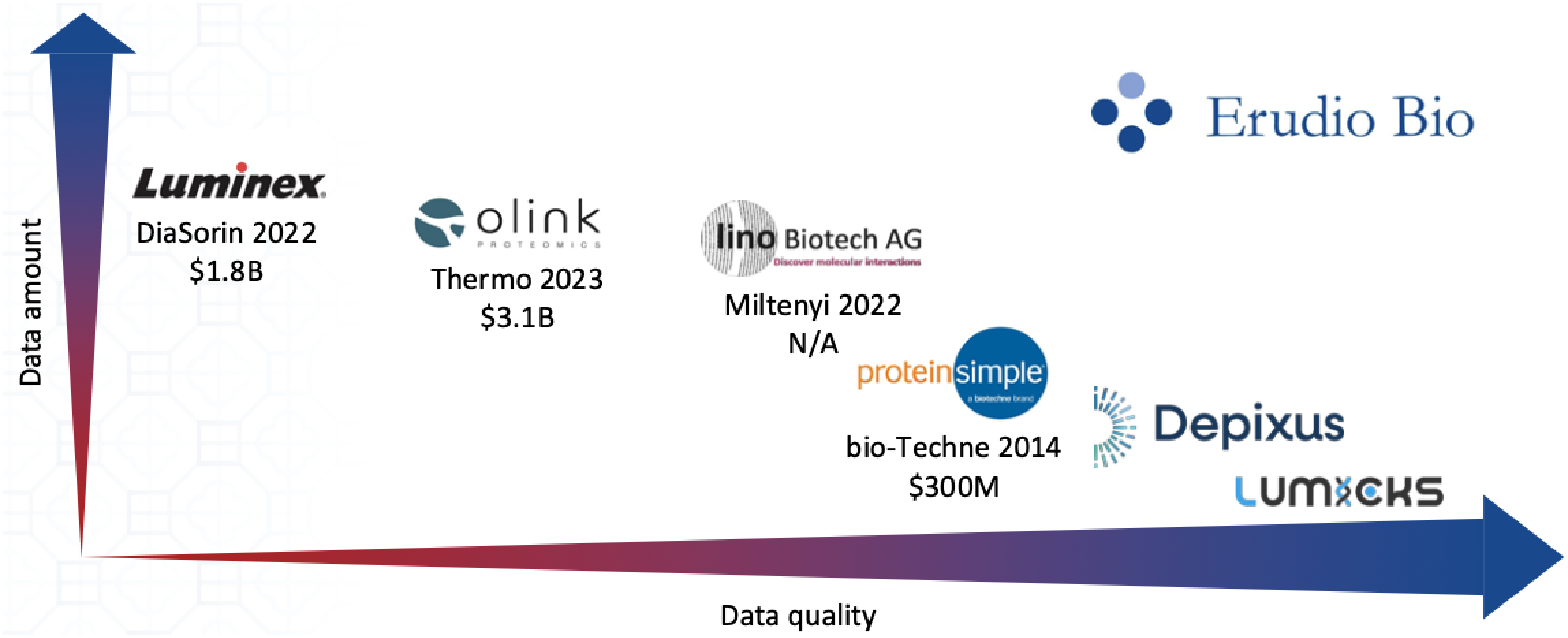
SNUH

계명대학교 동산병원 KEIMYUNG UNIVERSITY DONGSAN HOSPITAL

한국생명공학연구원 Korea Research Institute of Bioscience and Biotechnology

나노종합기술원 NATIONAL NANOFAB CENTER

## Competitive landscape




- Erudio Bio

- *only company* providing both high quality data and large data output

- efficient workflow integration facilitates customer adoption

## Erudio Bio engaging with Customers

# Validating out technology in clinical settings


- joint development agreement signed with *Shanghai General Hospital's National Clinical Research Center for Eye Diseases*
  - co-develop multiplexed diagnostic test for uveitis
  - addresses low sample volume and diagnostic inaccuracies
- globally unique partner, customer so large that it is a market by itself
  - 45 large hospitals with 83M patient visits per year
  - standard of care for smaller hospitals in Shanghai to access additional 280M patients

# Clinical validation to market leadership in Korean preventive care


- market opportunity - *preventive care* is foundation of Korean healthcare
  - $\sim$ 15M health screenings performed in S. Korea testing
  - ideal market segment for Erudio Bio's multi-omic multiplexed VSA platform
  - multi-billion dollar addressable market for multi-cancer early detection
- joint development agreed with *Seoul National University Hospital Bundang (SNUBH)*
  - premier hospital of national importance
  - enabling expansive proactive health assessment for efficient health care system
  - collaboration with target of *multiplexed cansor biomarker medical equipment business*

# Teams

# Team & advisory board

- team
  - Sunghee Yun, Ph.D. (CTO) - AI, optimization, business development, software
  - Kee-Hyun Paik, Ph.D. (CEO) - chip, microfluidics, instrumentation
  - Susanne Baumhueter, Ph.D. - biology, immunology, project management
  - Leon Chen, MBA, CFA (COO) - business development, product, operations
  - Jin Young Huh (CLO) - chief legal officer, business development in Korea
- 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
  - Ronald W. Davis - Director of Stanford Genome Tech Center ($15B+ exits)
  - William J. Greenleaf - Prof. Genetics and Applied Physics, Stanford University

# Gates Foundation Grant

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

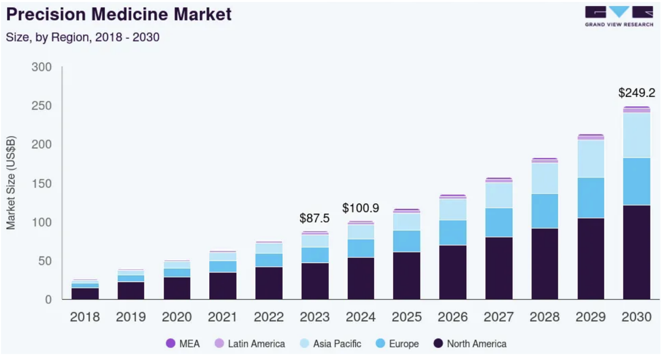
- $1M Grant Award (August 2025)
  - Gates Foundation recognizes Erudio Bio's potential to transform drug development for global health
- mission alignment - democratizing medicine by making preclinical drug design faster, yet reliable & accessible
  - lowering development costs for diseases affecting low- and middle-income countries (LMICs)
  - addressing the 90% clinical trial failure rate that drives up drug costs
- funded project - scale bioTCAD$^{\mathrm{TM}}$ platform to generate ground-truth binding datasets
  - 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

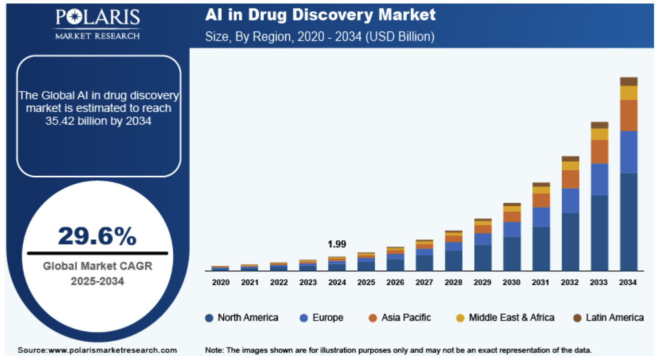
# Appendices

# The Trillion Dollar Opportunities

# The market is real - and enormous

- AI in biotech/pharma market size
  - AI in drug discovery - \$4B (2023) $\rightarrow$ \$50B+ by 2034 (Global Market Insights)
  - AI diagnostics market - \$1.2B (2023) $\rightarrow$ \$5-12B by 2030
  - precision/personalized medicine - \$80B (2023) $\rightarrow$ \$230B by 2030
  - synthetic biology - \$15B (2023) $\rightarrow$ \$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 $\rightarrow$ 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

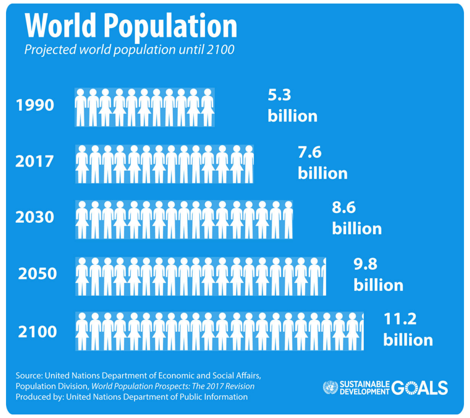


# Macro tailwinds amplifying these opportunities

- 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*

# Cross-domain inevitabilities - The Technical Moat

- core insight
  - fundamental mathematical structures recur across seemingly unrelated domains
  - recognizing these - strategic advantage - optimization $\rightarrow$ biological energy landscapes
  - protein folding is fundamentally an energy minimization problem over high-dimensional conformational space
- information theory $\rightarrow$ 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) 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

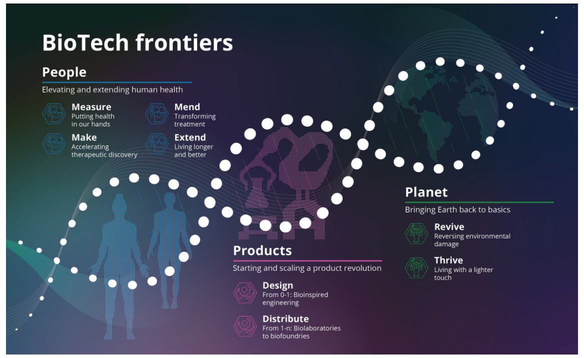
- *dimension 5 - human welfare at the center; not a constraint, but a strategic asset!*
  - mission alignment with patient outcomes unlocks - NIH/Gates/BARDA non-dilutive funding, academic medical center partnerships, favorable regulatory posture, and - increasingly - LP mandates in impact-oriented VC funds

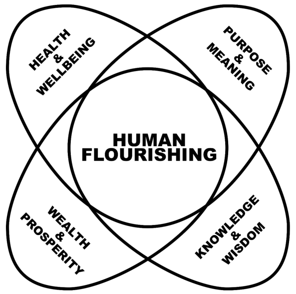
## Investment landscape & white spaces

- 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

# The Rare Entrepreneur This Moment Demands

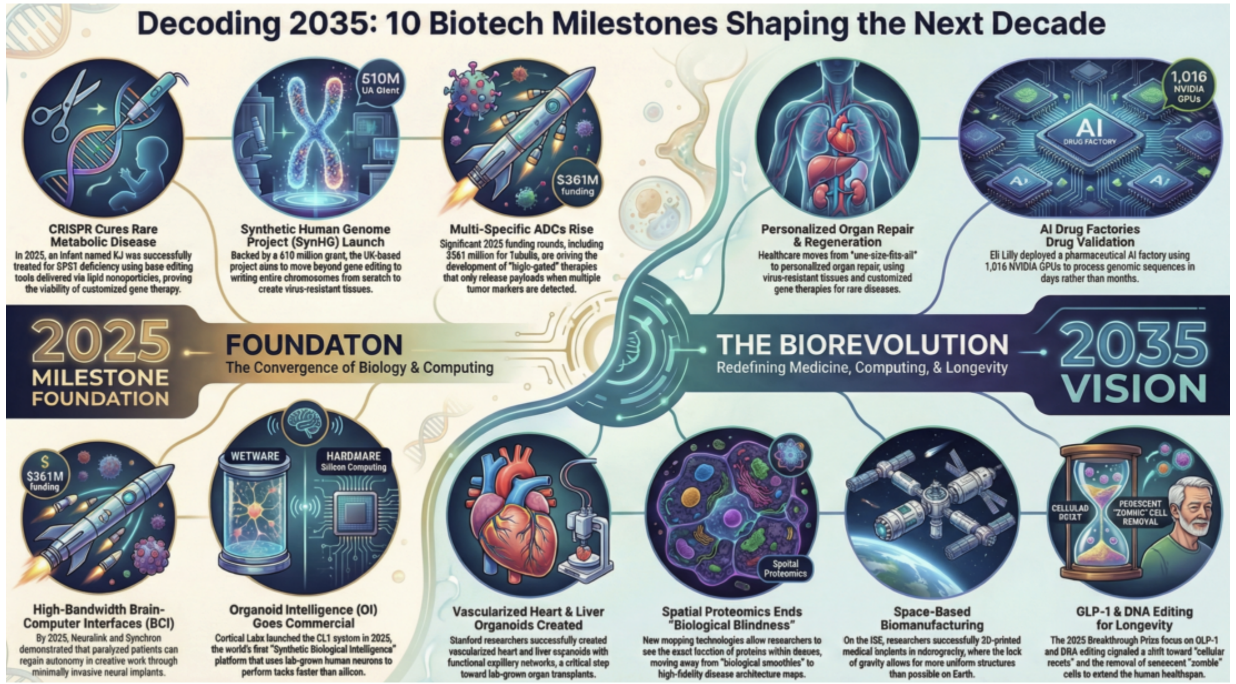
- three simultaneous operating modes – all required
  - *deep technologist* - must understand inner workings well enough to know what AI can and cannot do & *spot when a competitor's technical claim is hollow*
  - *business strategist* - must navigate regulatory complexity, reimbursement strategy, partnership 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



# 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


[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.

[GPAM+14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[HGH+22] Sue Ellen Haupt, David John Gagne, William W. Hsieh, Vladimir Krasnopolsky, Amy McGovern, Caren Marzban, William Moninger, Valliappa Lakshmanan, Philippe Tissot, and John K. Williams. The history and practice of AI in the environmental sciences. *Bulletin of the American Meteorological Society*, 103(5):E1351 – E1370, 2022.

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

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

[KW19] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019.

[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.

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

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

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of 31st Conference on Neural Information Processing Systems (NIPS)*, 2017.

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