[Future of Medicine]
From Pixels to Patients - How Agentic AI Transforms Pathology's Data Bottleneck

Sunghee Yun

Co-Founder & CTO @ Erudio Bio, Inc.
Co-Founder & CEO @ Erudio Bio Korea, Inc.
Leader of Silicon Valley Privacy-Preserving AI Forum (K-PAI)
Director of AI Semiconductor @ K-BioX
Member of AI-Korean Medicine Integration Initiative Task Force @ The Association of Korean Medicine
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 ∼
- *Director of AI Semiconductor @ K-BioX, CA, USA*   2025 ∼
- *AI-Korean Medicine Integration Initiative Task Force Member @ The Association of Korean Medicine*   2025 ∼
- *KFAS-Salzburg Global Leadership Fellow @ Salzburg Global Seminar*   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 Gruop (GLG), NY, USA   2022 ∼
- Chief Business Development Officer @ WeStory.ai, Cupertino, CA, USA   2025 ∼
- Advisor @ CryptoLab, Inc., San Jose, CA, USA   2025 ∼

- Co-Founder & CTO / Head of Global R&D / Chief Applied Scientist / Senior Fellow @ Gauss Labs, Inc., Palo Alto, CA, USA   2020 ∼ 2023

- 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, Memory Business   2015 ∼ 2016

- Principal Engineer @ DT Team, DRAM Development, Samsung   2012 ∼ 2015

- Senior Engineer @ CAE Team, Memory Business, Samsung, Korea   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 & BioTech

Erudio Bio

Transforming Precision Medicine

Erudio Bio - Technology & Foundation

- *Core Innovation* - Versatile Smart Assay (VSA) platform w/ patented force spectroscopy
  - measures *how strongly* molecules bind, not just if they bind
  - generates 1,000x more data from single biological sample vs existing methods
- *Stanford Pedigree* - 16 years of development at Stanford School of Medicine
  - mentored by Dr. Ronald Davis (father of microarrays)
  - validated by Analog Devices ($115B market cap), Stanford & Harvard Medical School
- *Founding Team* - Silicon Valley + Korea bridge
  - Kee-Hyun Paik (CEO): Stanford PhD, 20+ years semiconductor/biotech
  - *Sunghee Yun (CTO)*: Stanford PhD, ex-Gauss Labs CTO, Amazon AI ($200M impact), COE @ Erudio Bio Korea, Inc.
- *Recent Validation* - $1M Gates Foundation Grant (Aug 2025) for bioTCAD platform
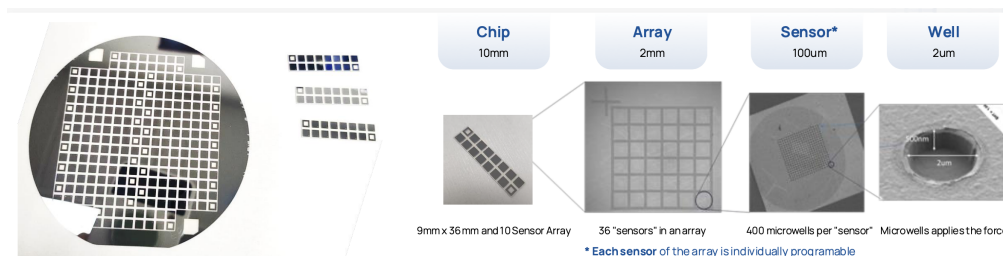
Building the Asia-Pacific Bridgehead with US

Erudio Bio - **Strategic Partnerships & Market Position**

- *China Market* - Shanghai General Hospital National Clinical Research Center
  - 83M annual patient visits / joint development of uveitis multiplexed diagnostic
  - set standard of care for 280M additional patients
- *Korea Market Hub* - Erudio Bio Korea, est. July 2025
  - *SNUBH* - cancer biomarker detection co-development; IRB approval in progress
  - *KAIST NanoFab* - manufacturing partnership for nano-fabrication
  - *KRIBB* - technology development collaboration
  - *Songdo Naeun Hospital* - marketing/sales partnership discussions
  - *Lulumedic* - Korean market entry + future bio/medical data business
- *Unique Position* - Only platform delivering both high data quality AND high data volume
  - Korea as Asia expansion platform
  - Convergence of semiconductors + AI + bio (perfect for Korea's strategic positioning)

AI-BioX ConfEX Grand Summit 2025

Convergence Across All Industries

December 11-13, 2025 — COEX, Seoul

- *World-Class Collaboration*
  - co-hosted by Rutgers, StartX, Johns Hopkins GBI Center, KAIST, K-BioX, IFEZ
  - bringing together scholars, industry leaders, and next-gen talent on international stage
- *My Role* - Three Strategic Theme Tracks
  - AI Semiconductor Track - Korea's Competitive Edge
  - Korean Medicine Track - Heritage Meets Innovation
  - Silicon Valley & AI Track - Cross-Pacific Innovation Bridge
- *Vision* - Creating Korea's "Three Convergences" leadership moment
  - Practical pathways from research to commercialization

K-PAI - Silicon Valley Privacy-Preserving AI Forum

Silicon Valley Privacy-Preserving AI Forum (K-PAI)

- 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)
  - Korean Consulate General, 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 2+ K-PAI Forums to qualify
- 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

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
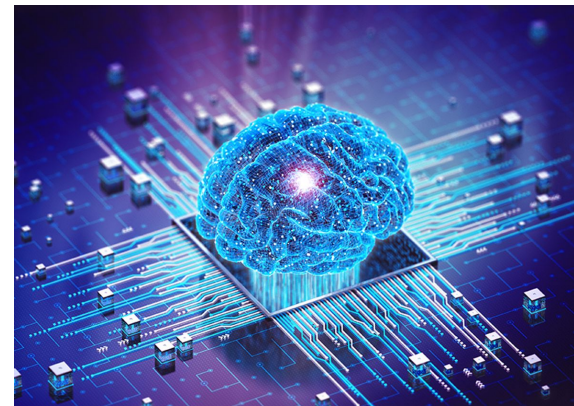
Biotech

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]
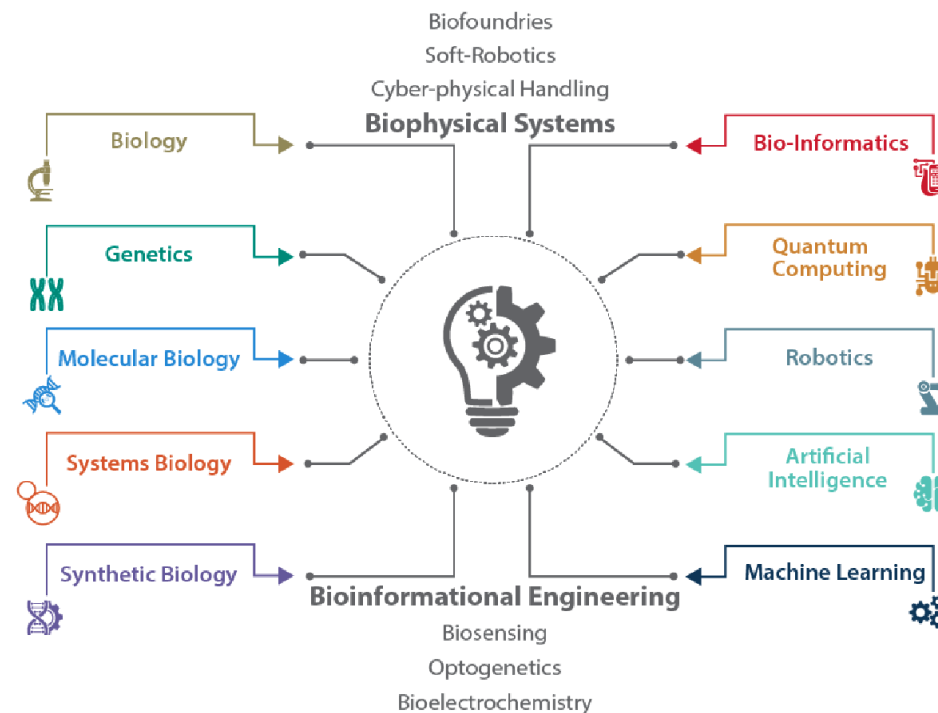
Convergence of AI and biological design

- 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*
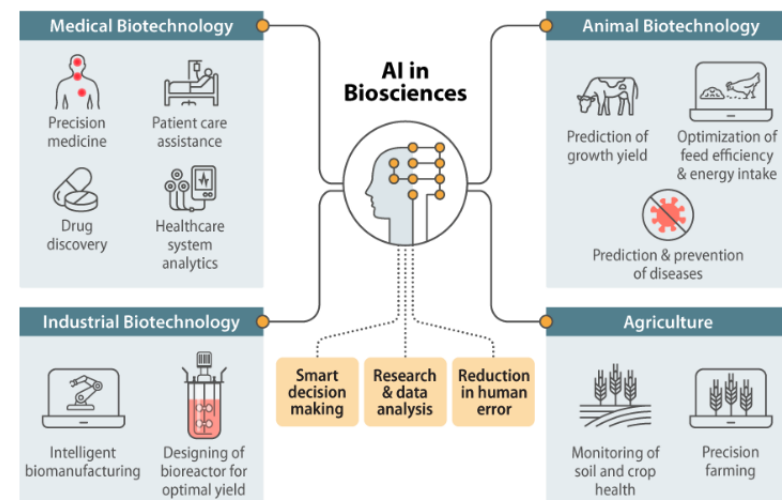
Multi-source genetic sequence data
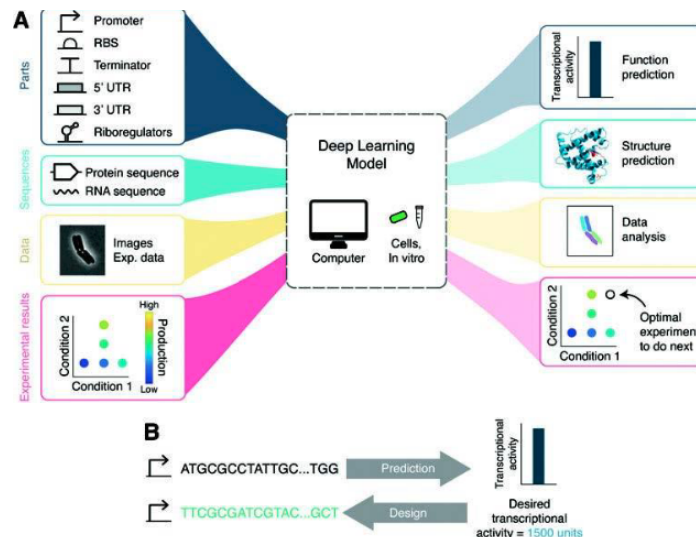
- AI, essential to analyzing exponential growth of genetic sequence data

  "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
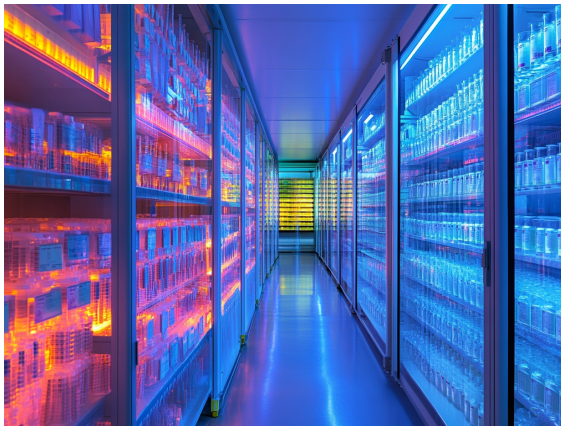
Rapid growth of biological data

- 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
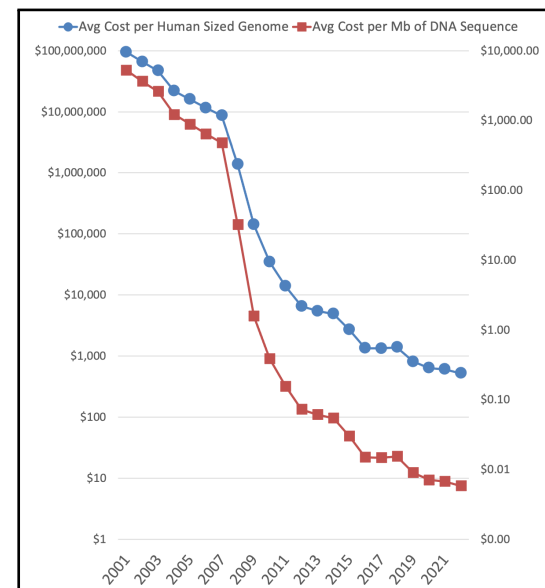
Volume and sequencing cost of DNA over time

- 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!*

<div align="center">
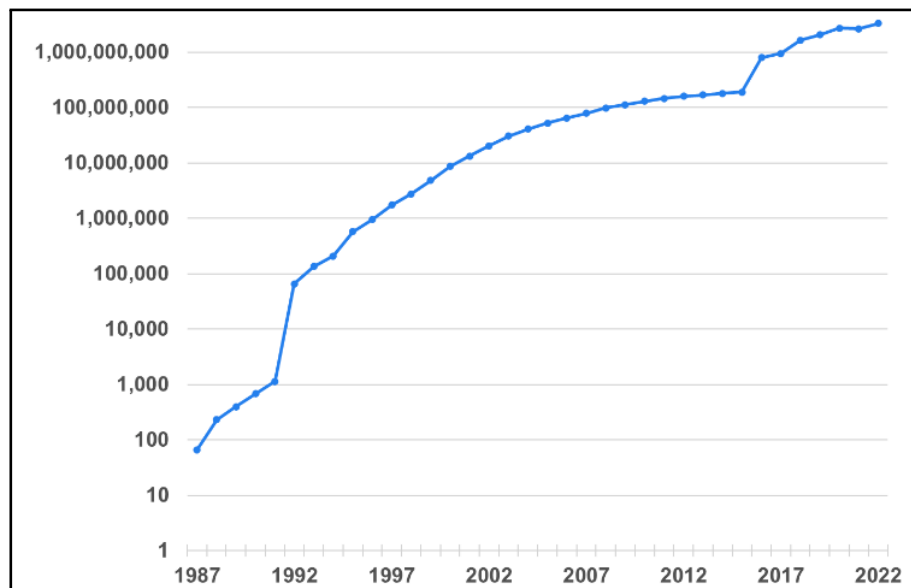
sequences in INSDC   DNA sequencing cost
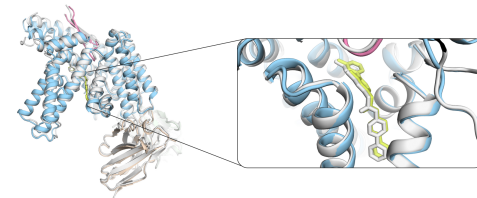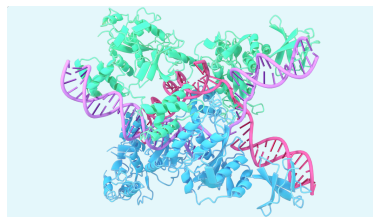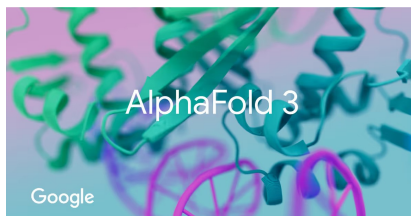
</div>

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 $\rightarrow$ AlphaFold 2 (2020) - attention-based NN solving protein folding "in principle" $\rightarrow$ 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
  0000000000000000000000000000000000000000000000000000000000000000000

- 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.*
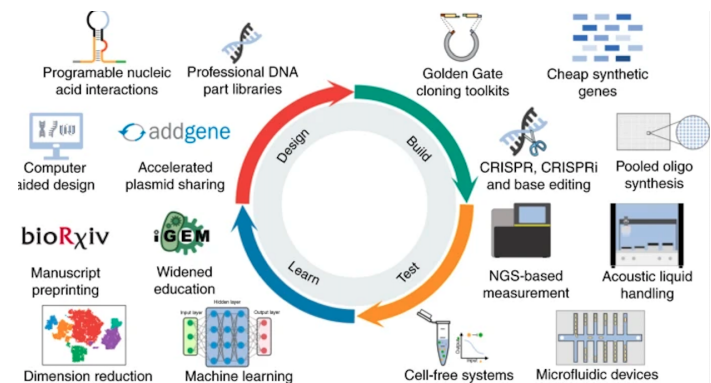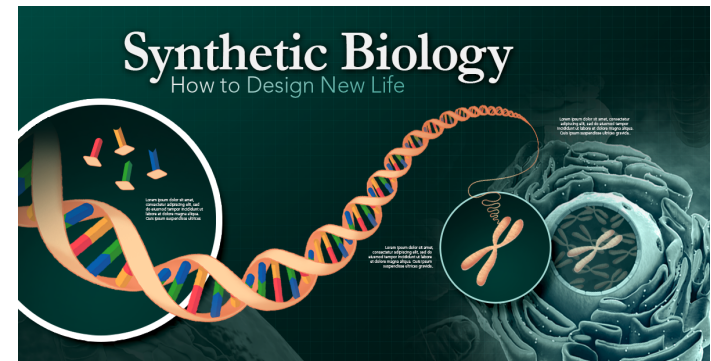

PERSONALIZED MEDICINE
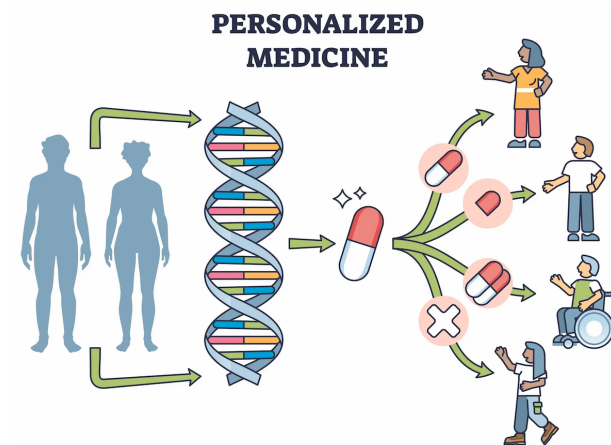
AI-driven drug discovery
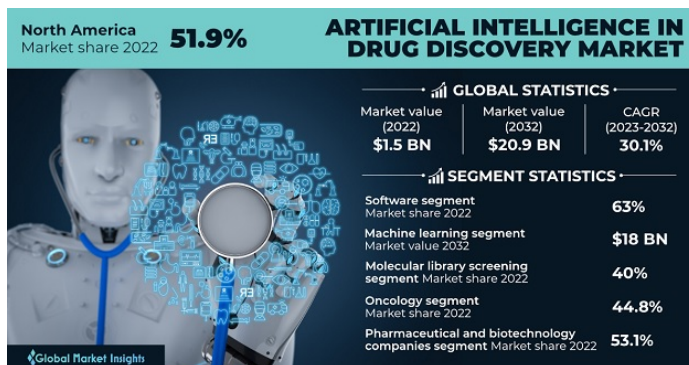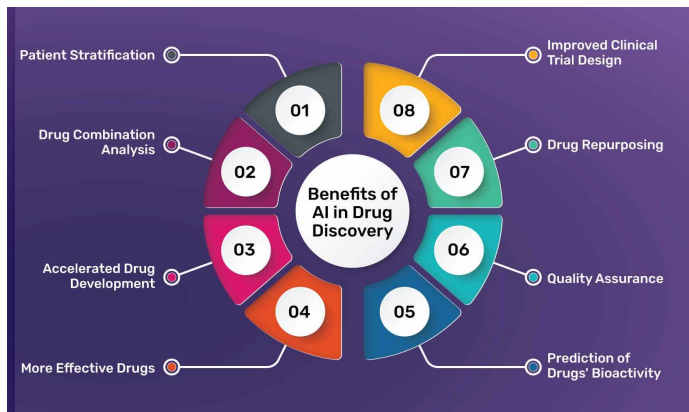
- traditional drug discovery process - time-consuming and costly often taking decades and billions of dollars

- AI streamlines this process by predicting the efficacy and safety of potential compounds with more speed and accuracy

- AI models analyze chemical databases to identify new drug candidates or repurpose existing drugs for new therapeutic uses

- companies - Insilco Medicine, Atomwise.

Synthetic biology

- use AI for gene editing, biomaterial production and synthetic pathways

- combine principles of biology and engineering to design and construct new biological entities

- AI optimizes synthetic biology processes from designing genetic circuits to scaling up production

- company - Ginkgo Bioworks uses AI to design custom microorganisms for applications ranging from pharmaceuticals to industrial chemicals

Regenerative medicine
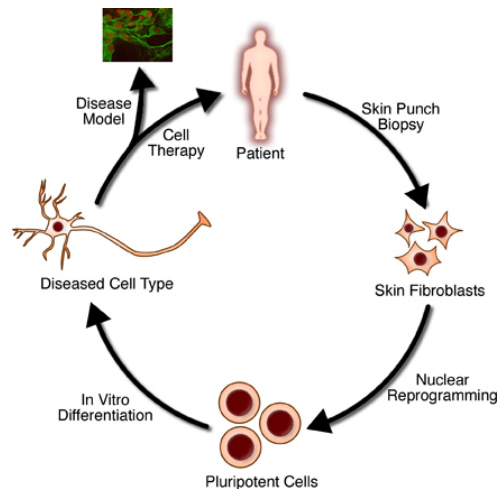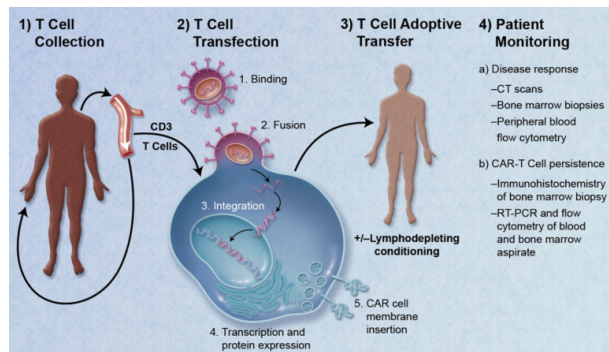
- AI advances development of stem cell therapies & tissue engineering

- AI algorithms assist in identifying optimal cell types, predicting cell behavior & personalized treatments

- particularly for conditions such as neurodegenerative diseases, heart failure and orthopedic injuries

- company - Organovo leverages AI to potentially improve the efficacy and scalability of regenerative therapies, developing next-generation treatments

Bio data integration

- integration of disparate data sources, including genomic, proteomic & clinical data - one of biggest challenges in biotech & healthcare

- AI delivers meaningful insights *only when* seamless data integration and interoperability realized

- developing platforms facilitating comprehensive, longitudinal patient data analysis - vital enablers of AI in biotech

- company - Flatiron Health working on integrating diverse datasets to provide holistic view of patient health
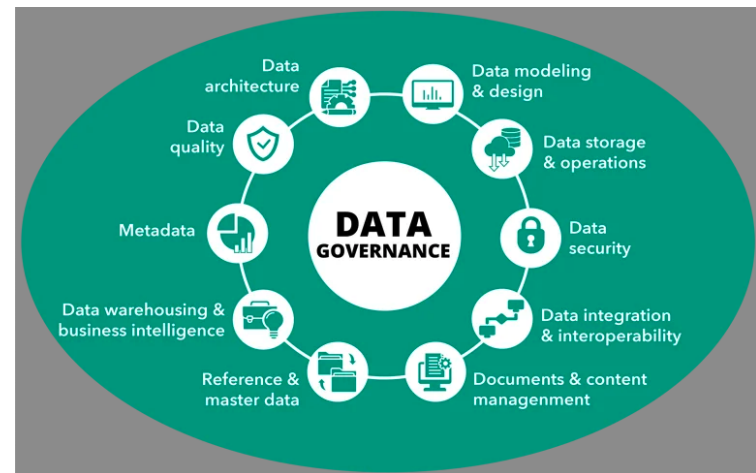
Biotech companies

- Atomwise - small molecule drug discovery

- Cradle - protein design

- Exscientia - precision medicine

- Iktos - small molecule drug discovery and design

- Insilico Medicine - full-stack drug discovery system

- Schrödinger, Inc. - use physics-based models to find best possible molecule

- Absci Corporation - antibody design, creating new from scratch antibodies, $i.e.$, "de novo antibodies", and testing them in laboratories
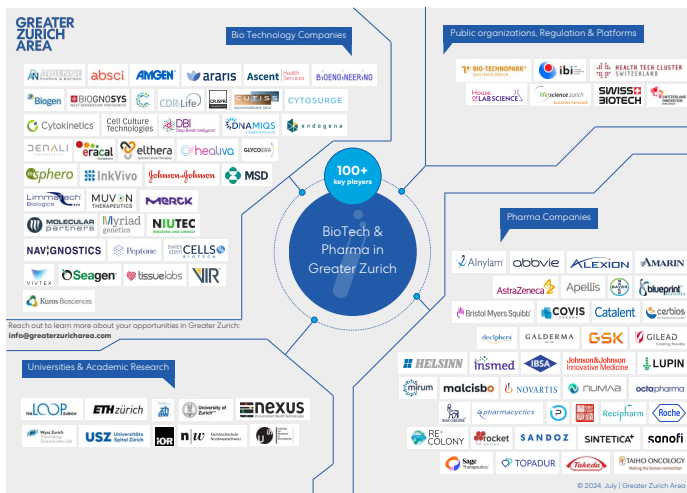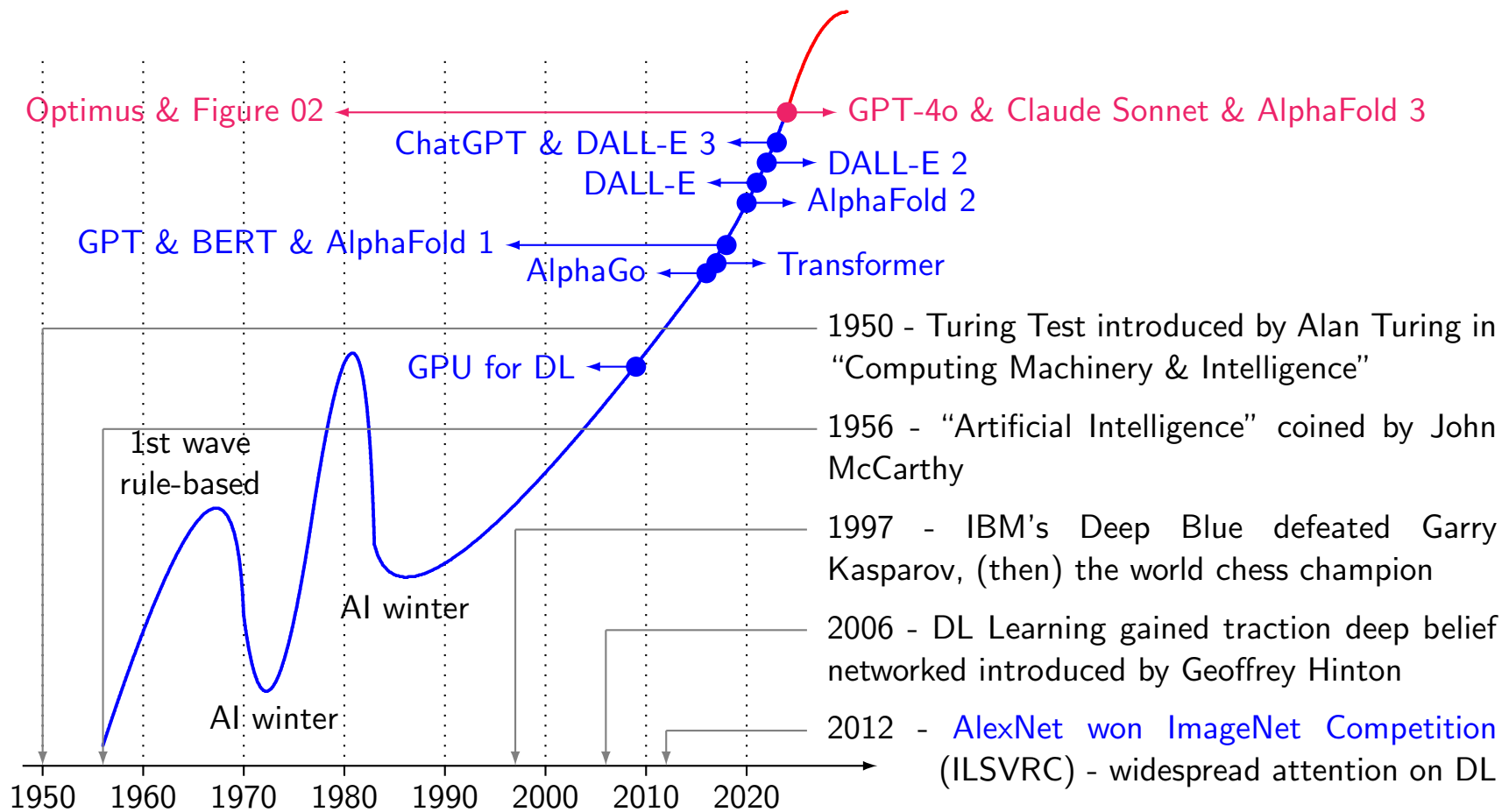
Silicon Valley's Cultural Engine of Innovation and Disruption

My journey from Samsung & Amazon to Gauss Labs & Erudio Bio

- Samsung Semiconductor, Inc.
  - inception into industry from academia, the world's best memory chip maker!
- Amazon.com, Inc.
  - experience so-called Silicon Valley big tech culture and technology
  - set tone for my future career trajectory!
- Gauss Labs, Inc.
  - found & operate AI startup, shaping corporate culture & spearheading R&D as CTO
  - inherent challenges of Korean conglomerate spin-off startup - cultural constraints, over-capitalization, and leadership limitations
- Erudio Bio, Inc.
  - concrete & tangible bio-technology in addition to AI
  - great decisions regarding business development; business models, market fit, go-to-market (GTM) strategies based on lessons learned *in a hard way* ☺

Joining Amazon.com, Inc. at the inflection point of AI

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

1950 - Turing Test introduced by Alan Turing in "Computing Machinery & Intelligence"

GPU for DL ←

1956 - "Artificial Intelligence" coined by John McCarthy

1st wave rule-based

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

AI winter

AI winter

2012 - AlexNet won ImageNet Competition (ILSVRC) - widespread attention on DL

1950   1960   1970   1980   1990   2000   2010   2020

Innovation ecosystem of Silicon Valley

- key characteristics
  - risk-taking culture, *trust* in technology $\rightarrow$ *genuine* respect for engineers and scientists
  - easy access to huge capital - VCs, angel investors alike
  - talent density - engineers, researchers, scientists, entrepreneurs, PMs, TPMs, . . .
  - diversity, "collision density" of ideas
  - ecosystem of collaboration and competition - startups, academia, industry leaders
- what they mean for global big tech
  - set trends in AI, software & hardware (and or hence) product & industry innovation
  - act as testing ground for disruptive ideas
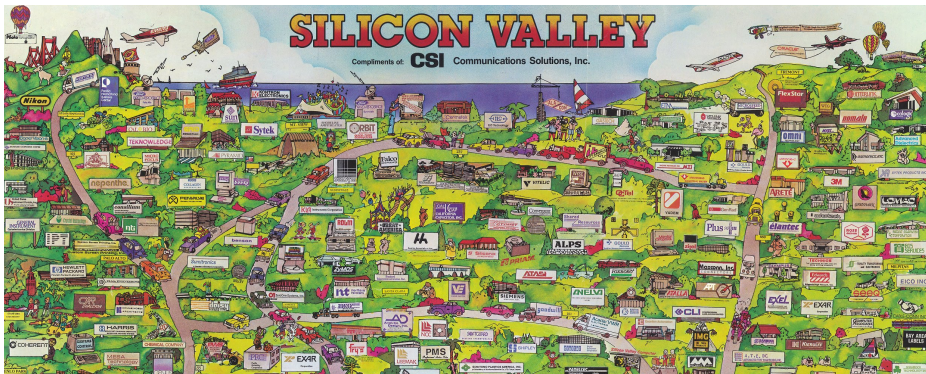
# Case study: Amazon - amazing differentiators of big techs

- Amazon's culture & leadership principles
  - customer obsession as driver of innovation
  - high standards & ownership culture, disagree & commit
  - bias for action and long-term thinking - sounds contradictory?
  - mechanisms like "two-pizza teams" & "Day One" for (or rather despite) scalability
- lessons for Korean corporations
  - applying customer-centric innovation in hardware & AI, $e.g.$, on-device AI
  - balancing agility with long-term R&D
  - *build / adapt / apply on the core strength of Samsung that no other company has!*
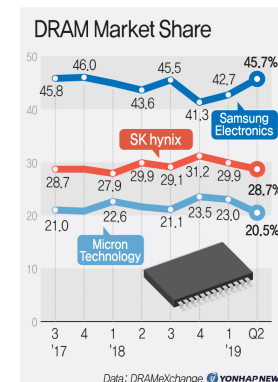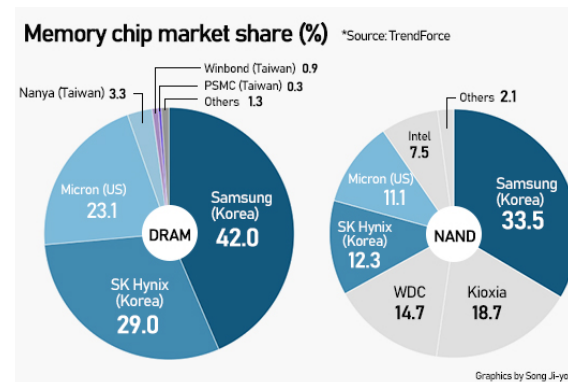
Founding and scaling startups

- challenges
  - competence of and chemistry among co-founders crucial
  - technology & great team are *necessary*, but *not sufficient (at all!)* for success
  - business models, market fit, timing, agility, flexibility for pivoting / perseverance
- insight
  - importance of domain expertise in addition to AI
  - balancing innovation with good business decisions
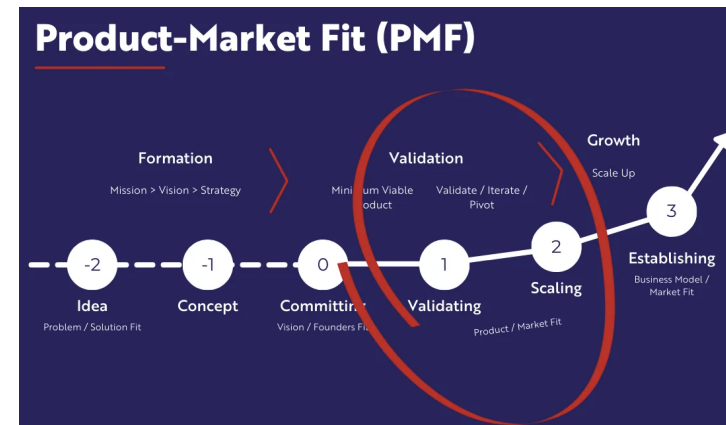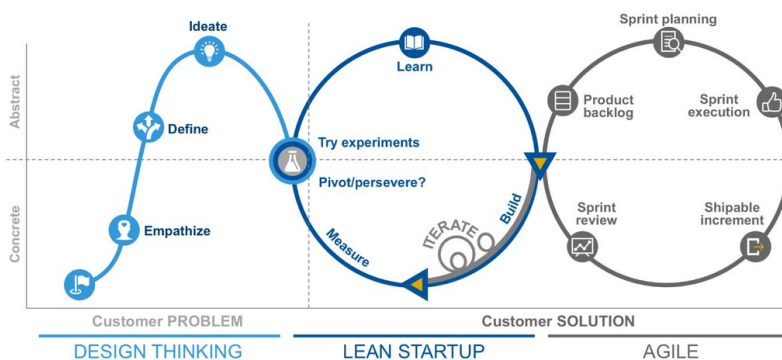
Bridging Silicon Valley & Korea

- cultural differences
    - risk appetite & failure tolerance
    - decision-making speed vs hierarchy
    - innovation vs execution focus
- opportunities for collaboration
    - leveraging Korea's manufacturing expertise with Silicon Valley's software/AI strengths
    - building global teams with diverse perspectives

To be successful . . .

- embrace customer/market-centric mindset in innovation and for business decisions

- balance agility with long-term vision

- foster cross-cultural collaboration for global impact

- ((very) strategically and carefully) leverage AI to solve real-world industrial challenges

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.

[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.

[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.

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

Thank You