

**AI for Fusion Biweekly Seminar:
Industrial AI & Biotechnology - Technology, Market, and
Future**

Sunghee Yun

Co-founder / CTO - AI Technology & Product Strategy

Erudio Bio, Inc.

Today

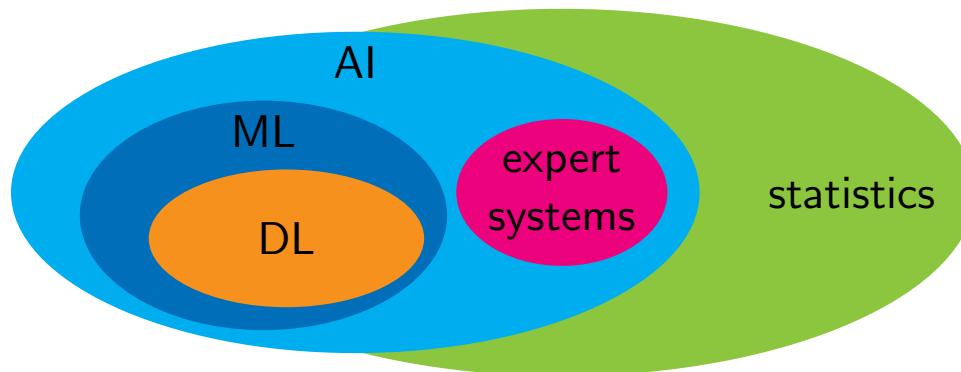
- industrial AI
 - why industrial AI?
 - computer vision (CV) and time-series (TS) AI in manufacturing
 - challenges for manufacturing AI
 - industrial AI success story - virtual metrology
- biotechnology
 - AI & bio
 - biotechnology - multidisciplinary field
 - bio data and processing cost
 - emerging trends in biotech
- AI market
 - AI hyper cycle
 - AI products, AI market outlook
 - AI & global economy

AI & Bio

AI in Bio

AI

- AI technologies, methodologies & applications used throughout biological sciences and biology R&D
- AI is *not* one thing - encompass range of technologies, methodologies & applications
- ML, DL & NN [HGH⁺22]
 - ML - umbrella term for training computer systems to solve complex problems, using large amount of data, learning algorithms, statistical models
 - DL (subset of ML) - using techniques such as multi-layer neural networks (NN) for computer vision (CV), natural language processing (NLP), time-series predictions to improve the performance and



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.

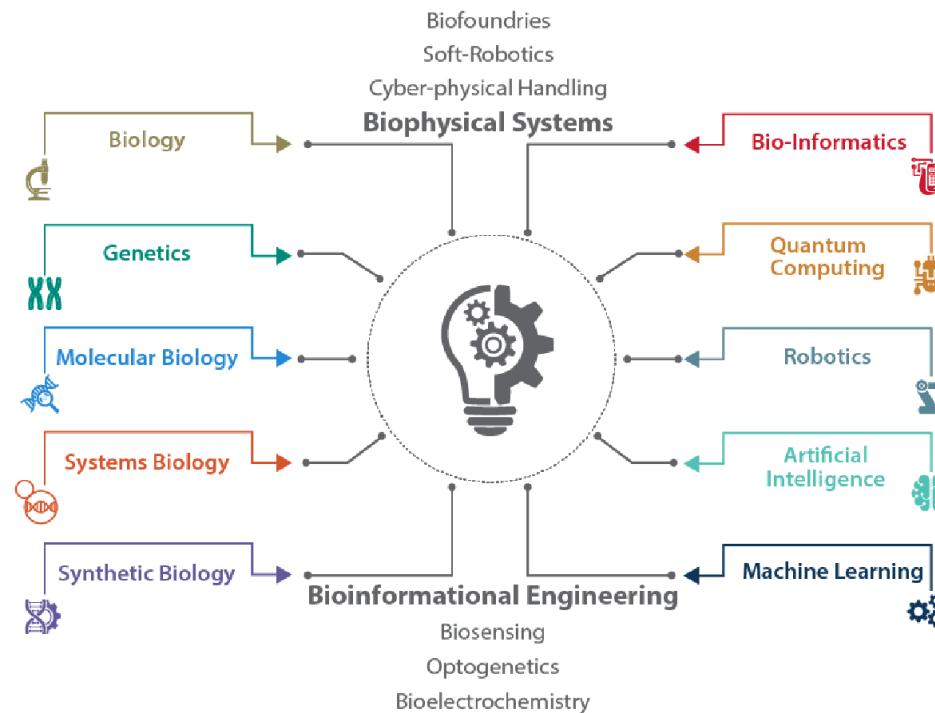
Biotechnology

Biotechnology

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

biotechnology - multidisciplinary field

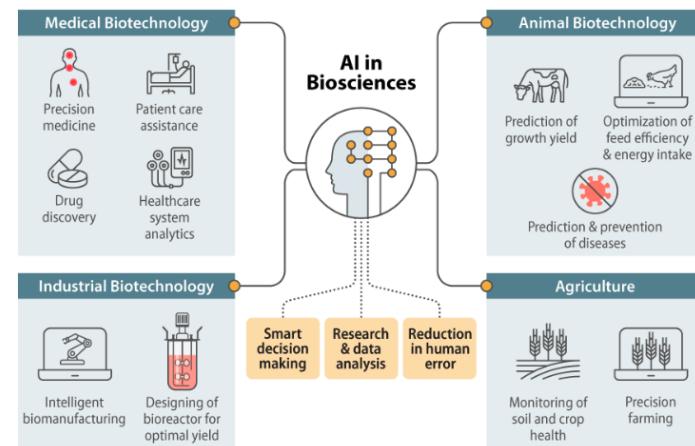
- sciences and technologies enabling biotechnology include, but not limited to,
 - (molecular) biology, genetics, systems biology, synthetic biology, bio-informatics, quantum computing, robotics [DFJ22]



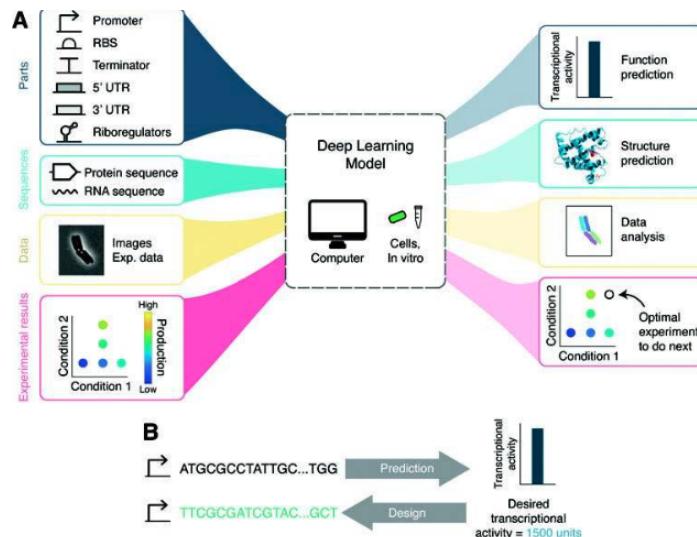
Convergence of AI and biological design

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

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



Multi-source genetic sequence data



- AI is essential to analyzing exponential growth of genetic sequence data
 - "AI will be essential to fully understanding how genetic code interacts with biological processes"
 - US National Security Commission on Artificial Intelligence (NSCAI)
- process huge amounts of biological data, e.g., genetic sequence data, coming from different biological sources for understanding complex biological systems
 - sequence data, molecular structure data, image data, time-series, omics data
- e.g., analyze genomic data sets to determine the genetic basis of particular trait and potentially uncover genetic markers linked with that trait

Quality & quantity of biological data

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

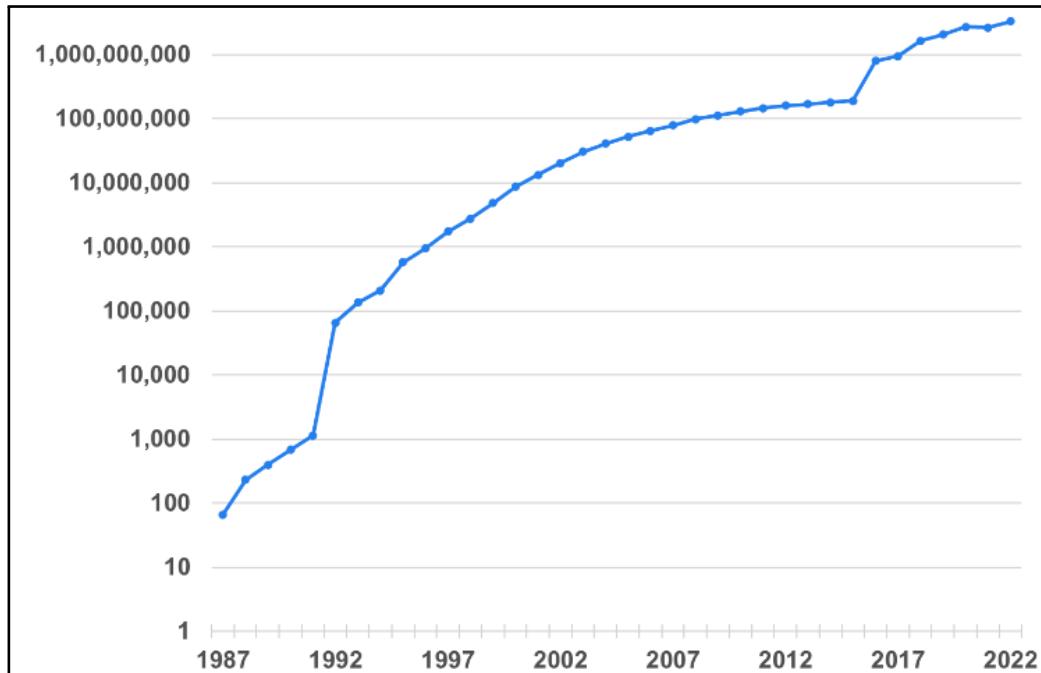
Rapid growth of biological data

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

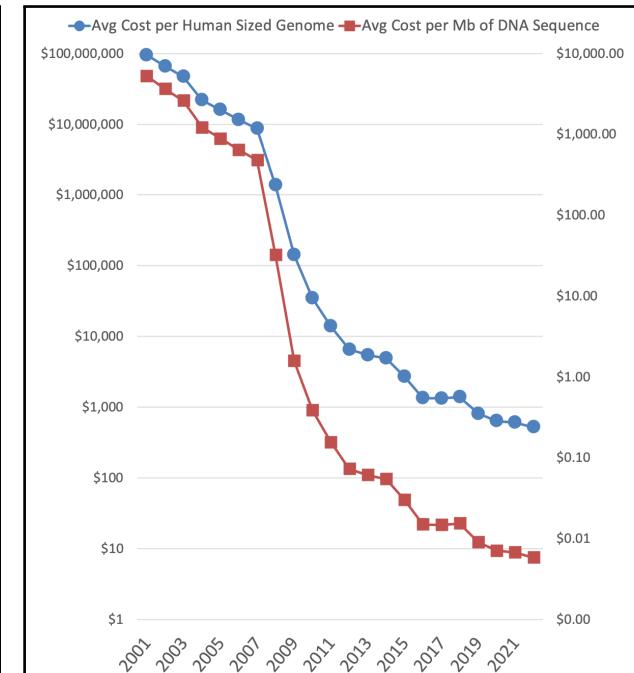
Volume and sequencing cost of DNA over time

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

sequences in INSDC



DNA sequencing cost



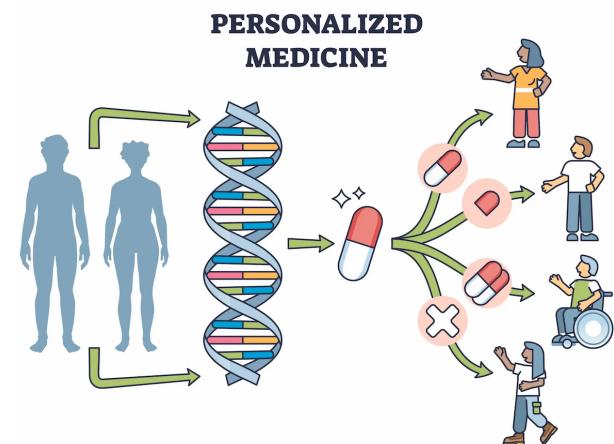
Bio data availability and bias

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

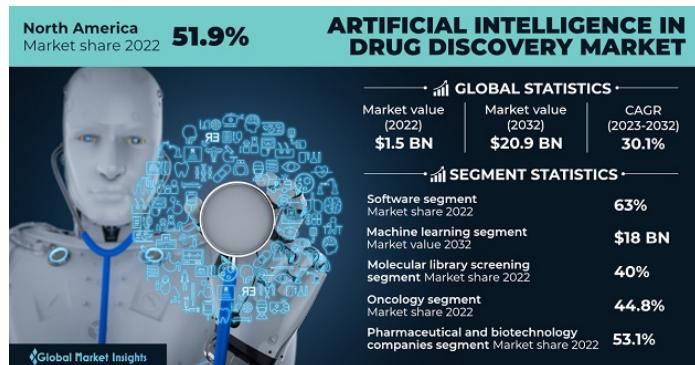
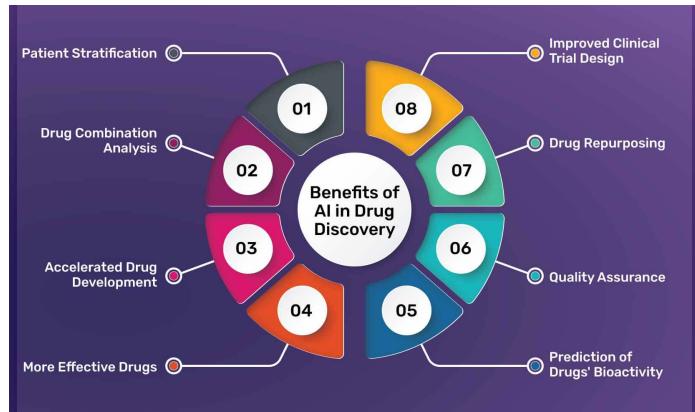
Emerging Trends in Biotech

Personalized medicine

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



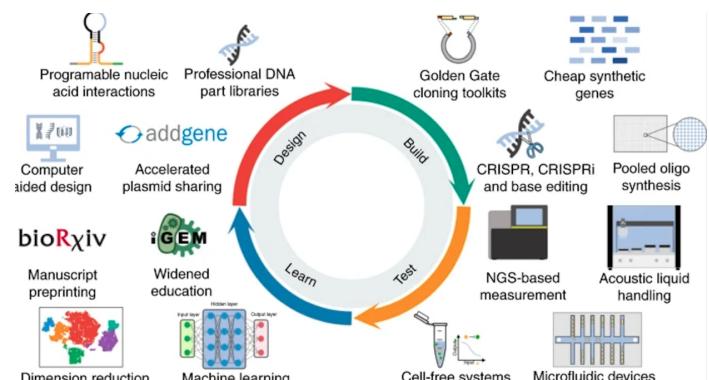
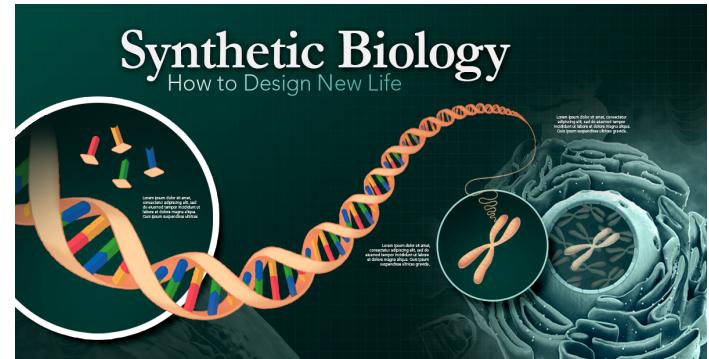
AI-driven drug discovery



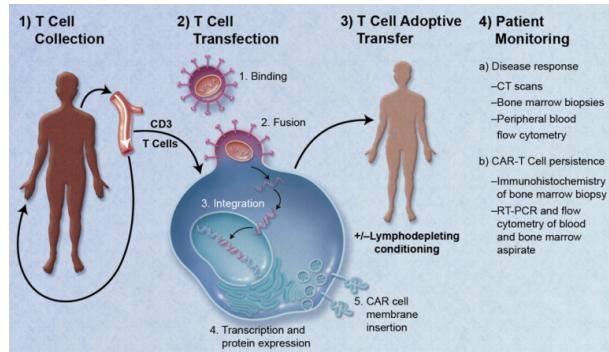
- traditional drug discovery process - time-consuming and costly often taking decades and billions of dollars
- AI streamlines this process by predicting the efficacy and safety of potential compounds with more speed and accuracy
- AI models analyze chemical databases to identify new drug candidates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

Synthetic biology

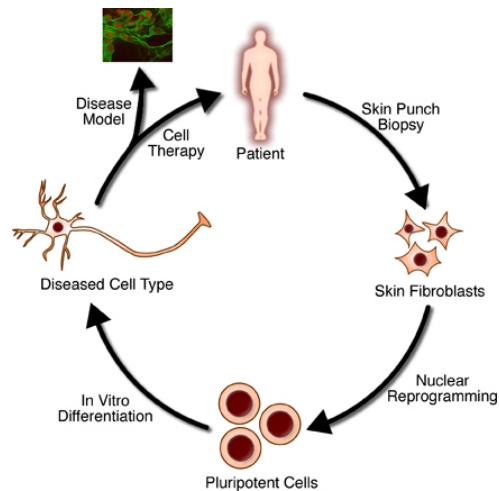
- use AI for gene editing, biomaterial production and synthetic pathways
- combine principles of biology and engineering to design and construct new biological entities
- AI optimizes synthetic biology processes from designing genetic circuits to scaling up production
- company - Ginkgo Bioworks uses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



Regenerative medicine

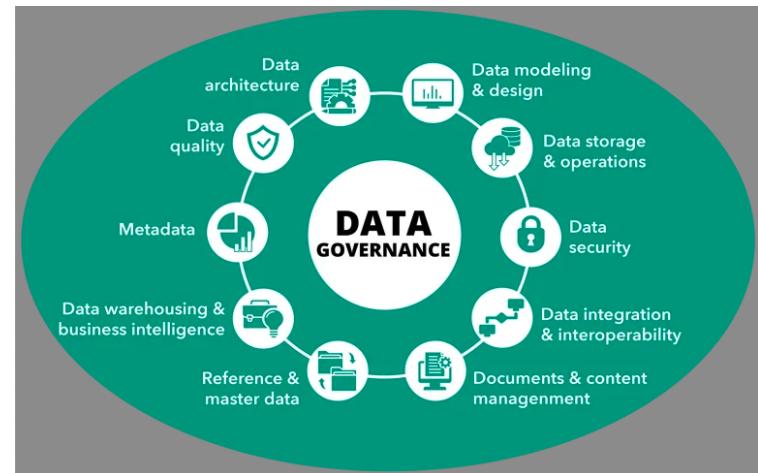


- AI advances development of stem cell therapies & tissue engineering
- AI algorithms assist in identifying optimal cell types, predicting cell behavior & personalized treatments
- particularly for conditions such as neurodegenerative diseases, heart failure and orthopedic injuries
- company - Organovo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

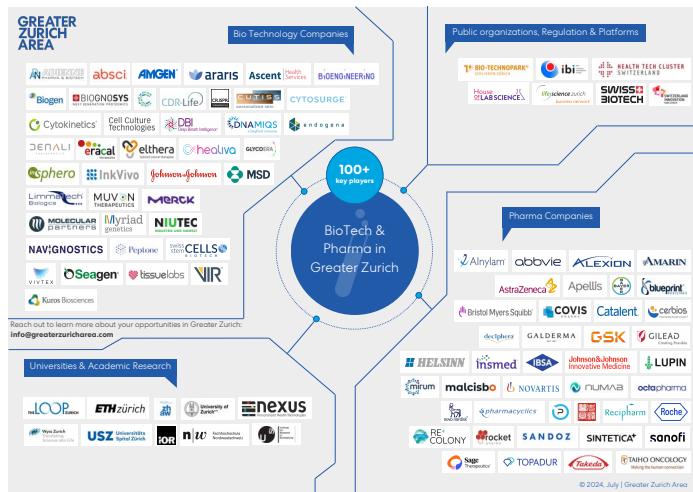


Bio data integration

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



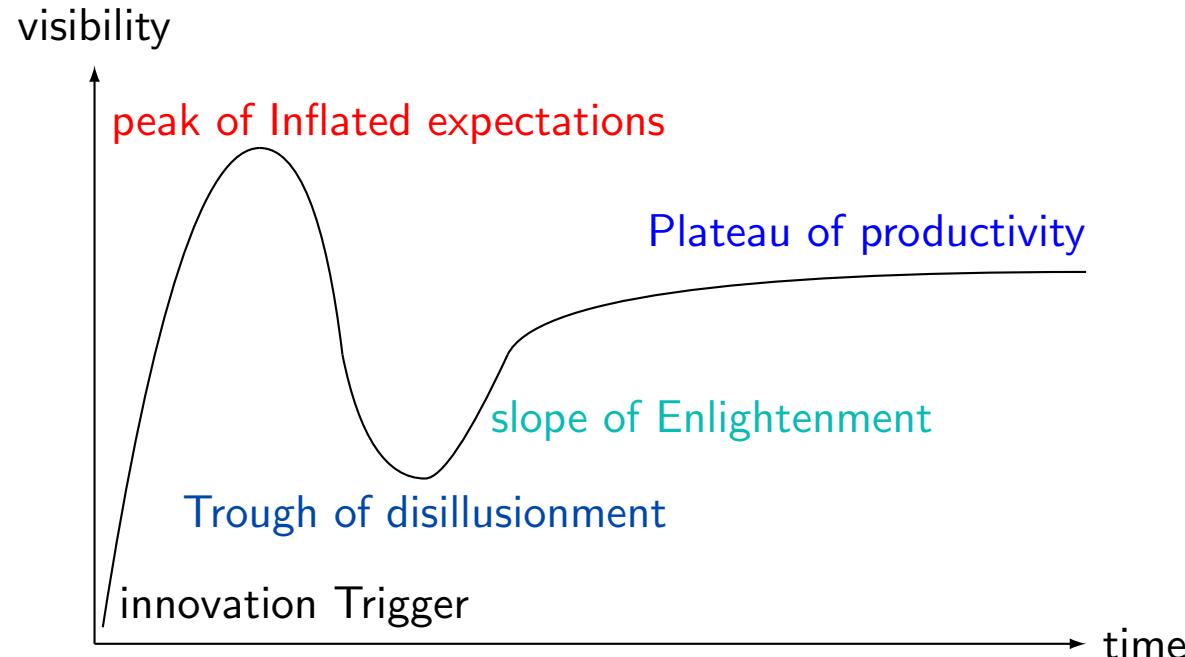
Biotech companies



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

AI Market

AI 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

genAI products

- DALL-E (OpenAI)
 - trained on a diverse range of images
 - *generate unique and detailed images based on textual descriptions*
 - understanding context and relationships between words

- Midjourney
 - let people *create imaginative artistic images*
 - can interactively guide the generative process, providing high-level directions



genAI products

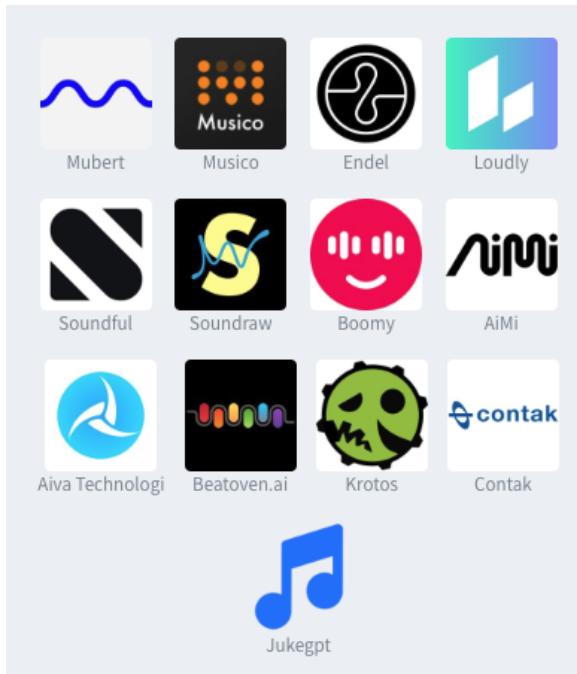


- Dream Studio
 - enables people to create music
 - *analyze patterns in music data and generates novel compositions based on input and style*
 - *allows musicians to explore new ideas and enhance their creative processes*
 - offer open-source free version
- Runway
 - provide range of generative AI tools for creative professionals
 - *realistic images, manipulate photos, create 3D models, automate filmmaking, . . .*
 - “artificial intelligence brings automation at every scale, introducing dramatic changes in how we create”

AI products

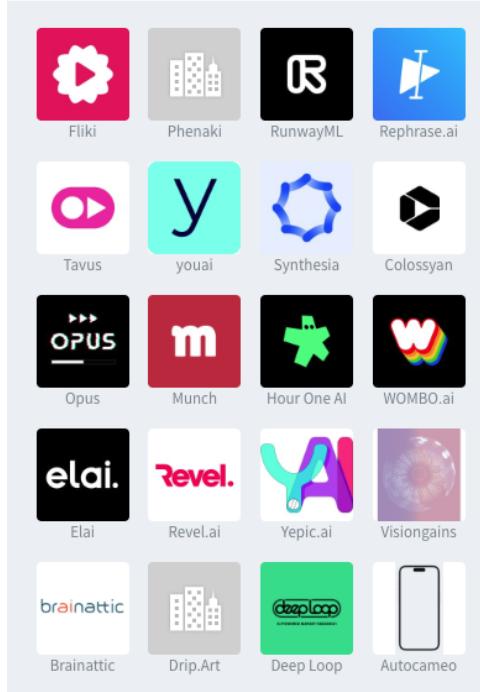
Audio: music generation

Combined funding \$ 61M



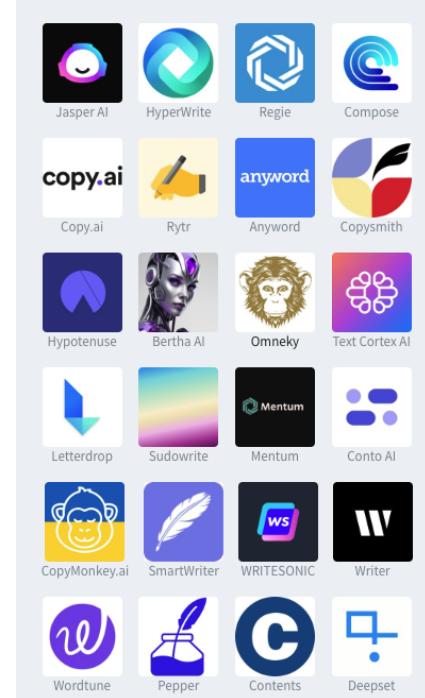
Video

Combined funding \$ 428M



Text: copy & writing

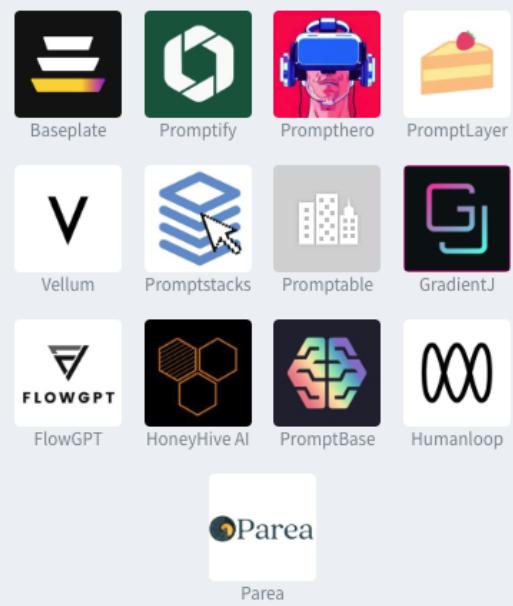
Combined funding \$ 863M



AI products

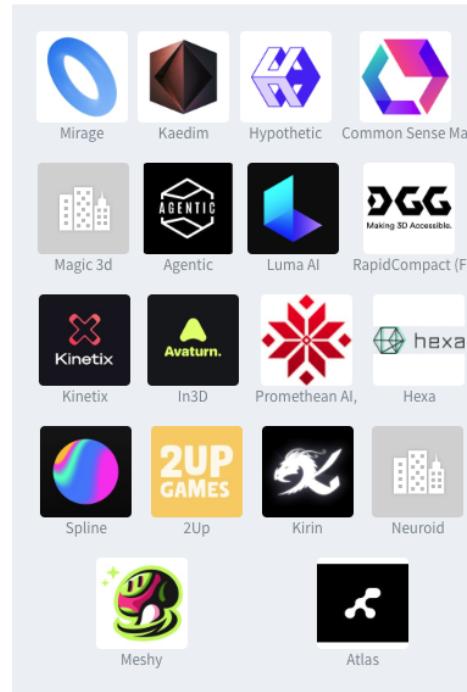
LLMs tools: Prompt Engineering and Management

Combined funding \$ 7.5M



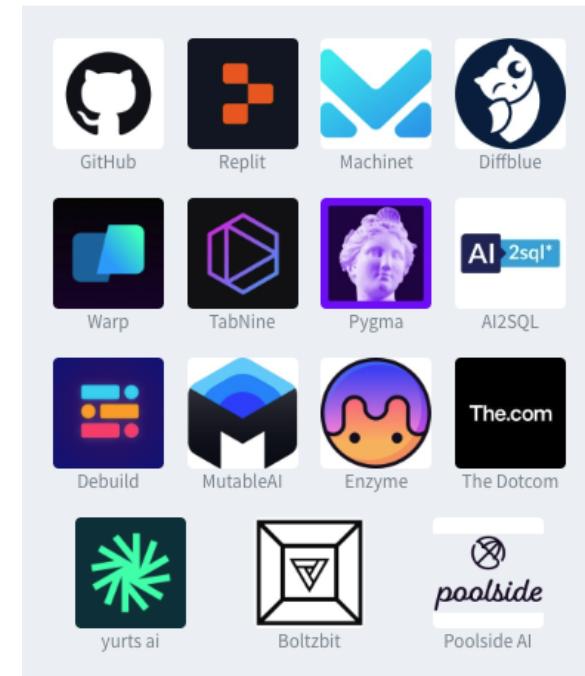
Gaming & design: 3d assets & worlds

Combined funding \$ 117M



Code: code generation

Combined funding \$ 828M



AI companies

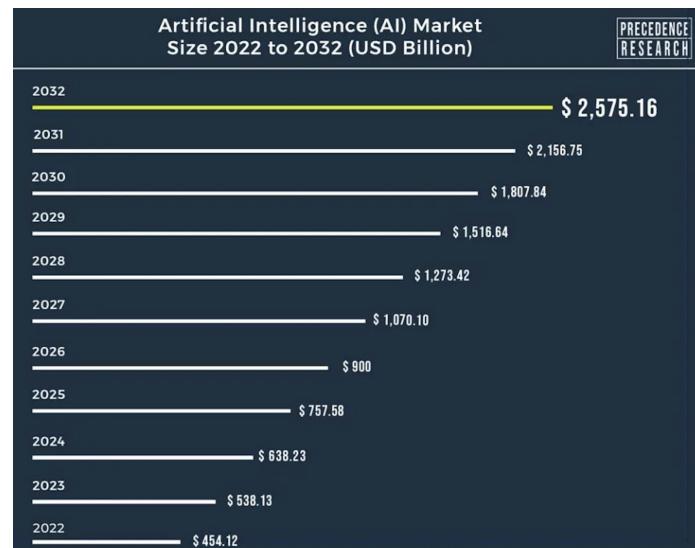
- big tech companies
 - OpenAI, Microsoft, Google, Meta - foundation models
- small(er) players
 - Figure AI, Mistral AI
- AI hardware companies - benefiting from LLM and genAI market dominance
 - Nvidia, AMD, Samsung, SK hynix, Micron, Intel, TSMC (AI processors & memory chips)
- *tiny fraction of Silicon Valley startups gets majority of total funding*
 - Anthropic - \$3.5B - large-scale AI systems - Claude
 - AssemblyAI - \$58M - speech AI
 - Hugging Face - \$400M - AI model/data platforms
 - Inflection AI - \$1.5B - conversational AI - Pi

Opportunities among big tech's domination

- OpenAI/Microsoft, Meta, Google's races for foundation models heated up!
- no small players can compete with rare exceptions, *e.g.*, Mistral AI
- hyperscalers stand strong - AWS, Azure, and Google Cloud
- *speaker's proposals for strategies*
 - accurately (or roughly) predict how far & up to where big players will reach
 - target niche markets
 - focus on (creative) downstream applications of LLMs and/or genAIs

AI market outlook in 2024

- global AI market expected to reach *USD 0.5T by 2024* (IDC @ Mar-2023) & expected to reach around *USD 2.5T by 2032* (Precedence Research @ Dec-2023) [P.R23]
 - was valued at USD 454B in 2022, expanding at *double-digit CAGR of 19%* from 2023 to 2032
- *AI funding soars to USD 17.9B for Q3 in 2023 in Silicon Valley while rest of tech slumps* (PitchBook data, Bloomberg @ Oct-2023) [Blo23]
 - multibillion-dollar investment in AI startups almost commonplace in Silicon Valley
 - genAI dazzles users and investors with photo-realistic images & human-sounding text
- genAI software sales could surge *18,647% by 2032*



Productivity, inflation & jobs

- Federal Reserve probes AI's impact on productivity, inflation & jobs - Jul-2024
 - feds acknowledging significant AI investments
 - Jerome Powell emphasizes uncertainties on whether AI will eliminate, augment, or create jobs - stating it's too early to predict
 - Powell acknowledges limited influence of central banks like the Fed on AI's technological shifts
 - fed actively researching various AI forms beyond genAI to understand potential economic impacts
 - IMF predicts AI (could) impact up to 60% of jobs in advanced economies potentially lowering labor demand and wages in sectors like finance and insurance

AI & global economy

- five ways AI is transforming global economy
 - reshape job markets, creating new roles while rendering some obsolete
 - enhance productivity across industries
 - contribute to global economy by optimizing processes and innovation
 - *may widen economic disparities if not managed inclusively*
 - *governments* has to develop policies to address AI's economic and social impacts



References

References

- [BKP22] Abhaya Bhardwaj, Shristi Kishore, and Dhananjay K. Pandey. Artificial intelligence in biological sciences. *Life*, 12(1430), 2022.
- [Blo23] Bloomberg. Bloomberg, 2023.
- [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.
- [HGH⁺22] Sue Ellen Haupt, David John Gagne, William W. Hsieh, Vladimir Krasnopolksky, 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.

- [P.R23] P.R. Precedence research, 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