

**[2025 Gifted Education Dream Realization Camp]
Magic Behind the Machine - AI as Your Creative Partner**

Sunghee Yun

Co-Founder & CTO @ Erudio Bio, Inc.

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

**Global Leadership Initiative Fellow @ Salzburg Global Seminar
Adjunct Professor & Advisory Professor @ Sogang Univ. & DGIST**

About Speaker

- *Co-Founder & CTO @ Erudio Bio, Inc., San Jose & Novato, CA, USA*
- *Co-Founder & CEO @ Erudio Bio Korea, Inc., Korea*
- *KFAS-Salzburg Global Leadership Initiative Fellow @ Salzburg Global Seminar*, Salzburg, Austria
- Advisory Professor, Electrical Engineering and Computer Science @ DGIST, Korea
- Adjunct Professor, Electronic Engineering Department @ Sogang University, Korea
- Global Advisory Board Member @ Innovative Future Brain-Inspired Intelligence System Semiconductor of Sogang University, Korea
- Technology Consultant @ Gerson Lehrman Group (GLG), NY, USA
- Advisor & Evangelist @ CryptoLab, Inc., San Jose, CA, USA
- Chief Business Development Officer @ WeStory.ai, Cupertino, CA, USA
- *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 ~ 2020
- Principal Engineer @ Software R&D Center, DS Division, Samsung, Korea ~ 2017
- Principal Engineer @ Strategic Marketing & Sales Team, Samsung, Korea ~ 2016
- Principal Engineer @ DT Team, DRAM Development Lab, Samsung, Korea ~ 2015
- Senior Engineer @ CAE Team, Samsung, Korea ~ 2012
- PhD - Electrical Engineering @ Stanford University, CA, USA ~ 2004
- Development Engineer @ Voyan, Santa Clara, CA, USA ~ 2001
- MS - Electrical Engineering @ Stanford University, CA, USA ~ 1999
- BS - Electrical & Computer Engineering @ Seoul National University 1994 ~ 1998

Highlight of Career Journey

- BS in EE @ SNU, MS & PhD in EE @ Stanford University
 - *Convex Optimization - Theory, Algorithms & Software*
 - advised by *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
 - Bezos's project - drove *\$200M* in additional sales via Amazon Mobile Shopping App
- *Co-Founder & CTO / Global R&D Head & Chief Applied Scientist @ Gauss Labs, Inc.*
- *Co-Founder & CTO* - AI Technology & Business Development @ Erudio Bio, Inc.

Today

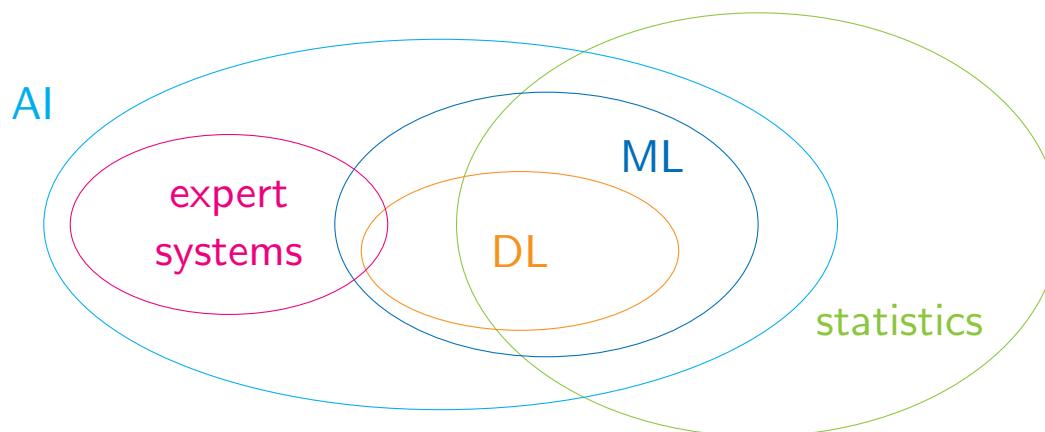
- Artificial Intelligence - 5
 - AI history & recent significant achievements
 - Market indicators for unprecedented AI progress
- Some Important Questions around AI - 30
 - Why Human level AI?
 - Consciousness & biases
 - Can AI think, reason, believe, or even know something?
- Silicon Valley's Cultural Engine of Innovation and Disruption - 66
 - My journey from Samsung & Amazon to Gauss Labs & Erudio Bio
 - Innovation ecosystem of Silicon Valley / Founding and scaling startups
- Selected references - 74
- References - 76

Artificial Intelligence

Definition and History

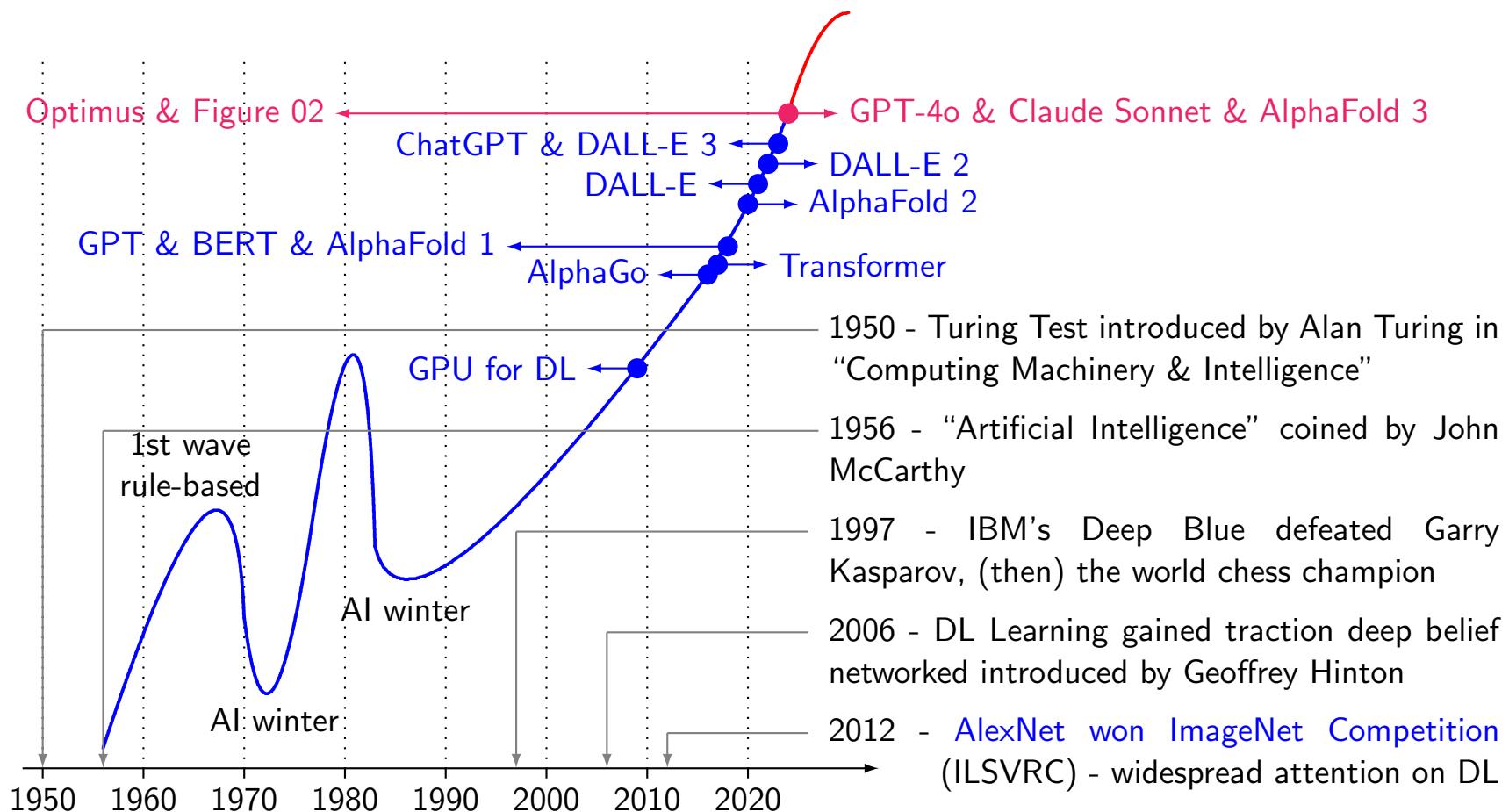
Definition & relation to other technologies

- 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 system¹ [HGH⁺22]



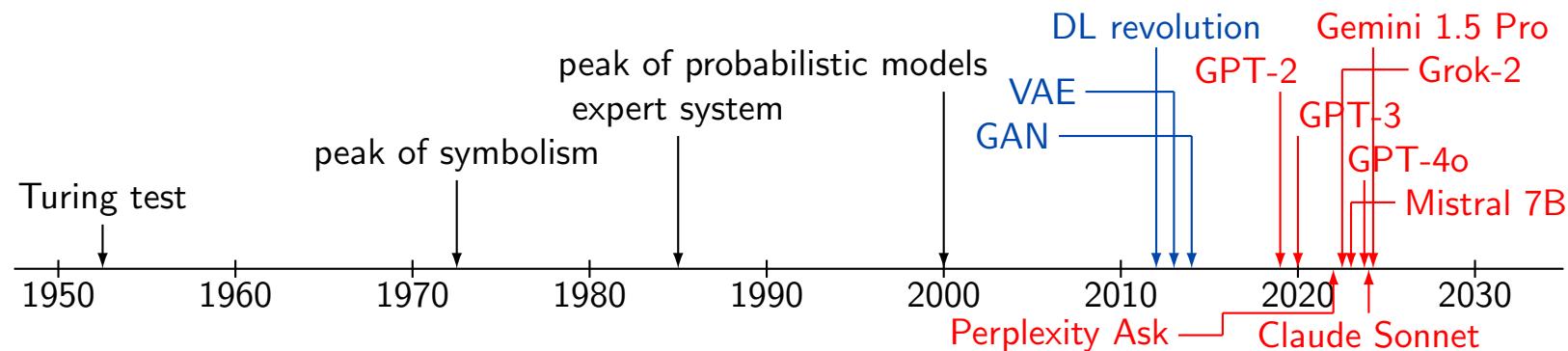
¹ML: machine learning & DL: deep learning

History



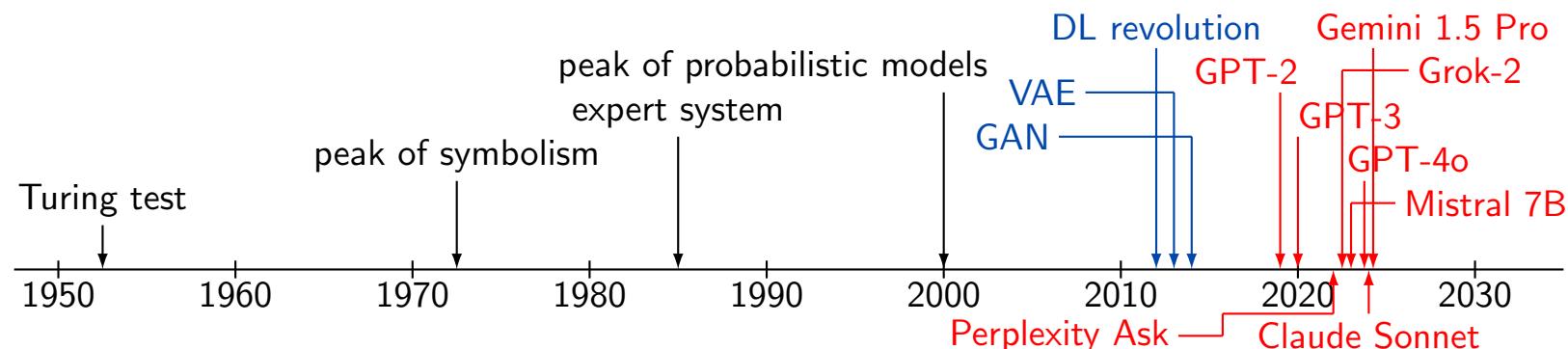
Birth of AI - early foundations & precursor technologies

- 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 ~)



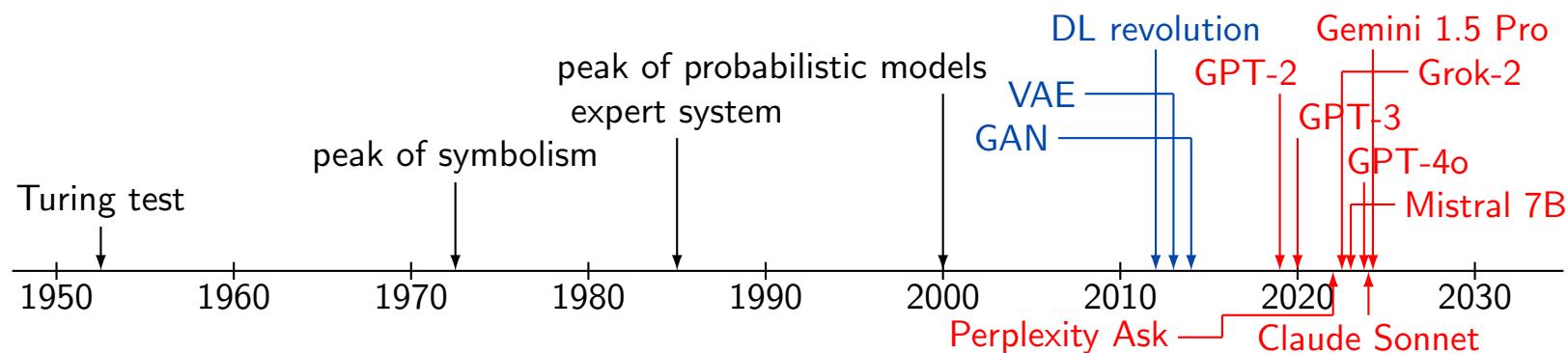
Rule-based systems & probabilistic models

- 1980s ~ 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



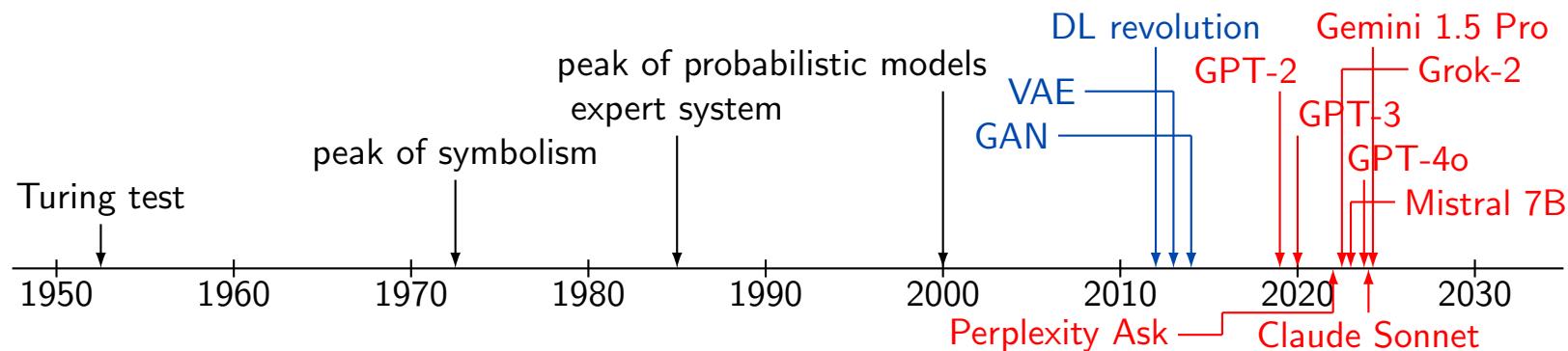
Rise of deep learning & generative models

- 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 ~ 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

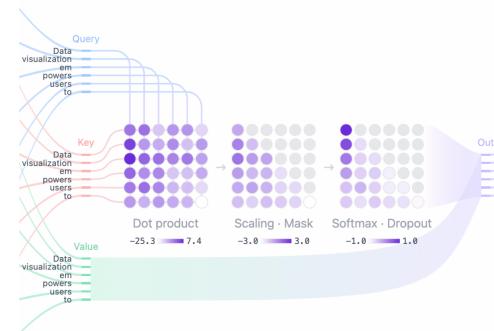
- 2012 – 2015 - DL revolution²
 - CNNs demonstrated exceptional performance in image recognition, e.g., *AlexNet's victory in ImageNet competition*
 - widespread adoption of DL learning in CV transforming industries
- 2016 - AlphaGo defeats human Go champion
 - DeepMind's AlphaGo defeated world champion in Go, extremely complex game *believed to be beyond AI's reach*
 - significant milestone in RL - AI's potential in solving complex & strategic problems



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

Transformer changes everything

- 2017 – 2018 - Transformers & NLP breakthroughs³
 - *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



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

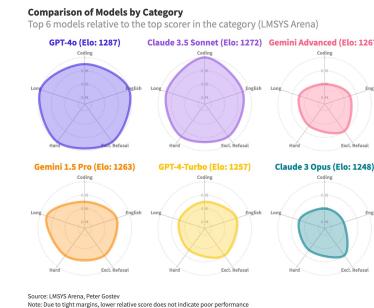
Lots of breakthroughs in AI technology and applications in 2024

- 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*



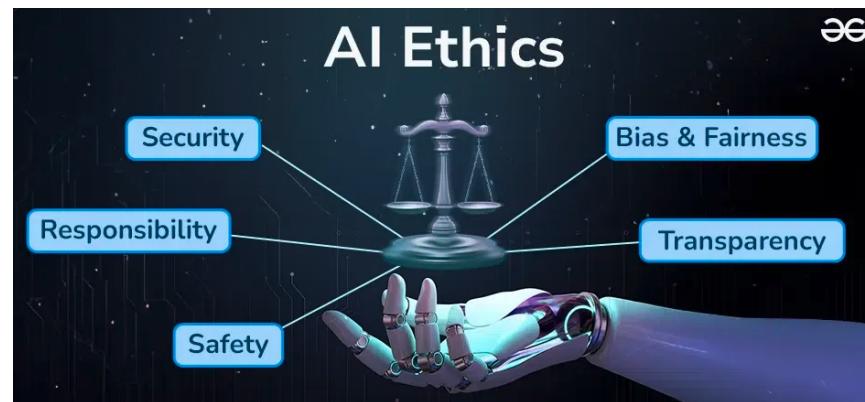
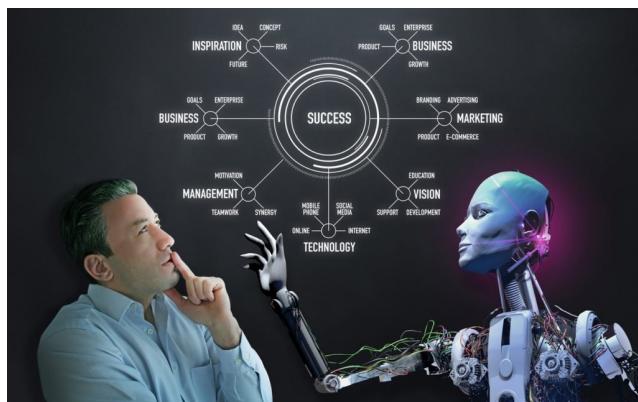
Major AI Breakthroughs in 2025

- 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



Transformative impact of AI - reshaping industries, work & society

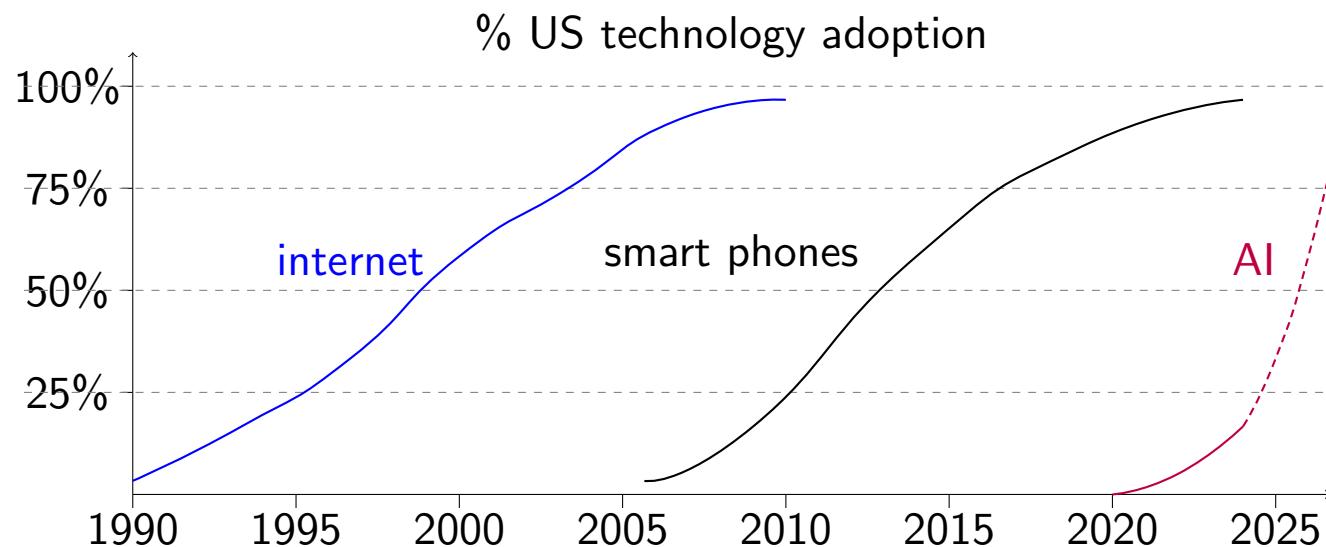
- accelerating human-AI collaboration
 - not only reshaping industries but *altering how humans interact with technology*
 - AI's role as collaborator and augmentor redefines productivity, creativity, the way we address global challenges, e.g., *sustainability & healthcare*
 - AI-driven automation *transforms 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

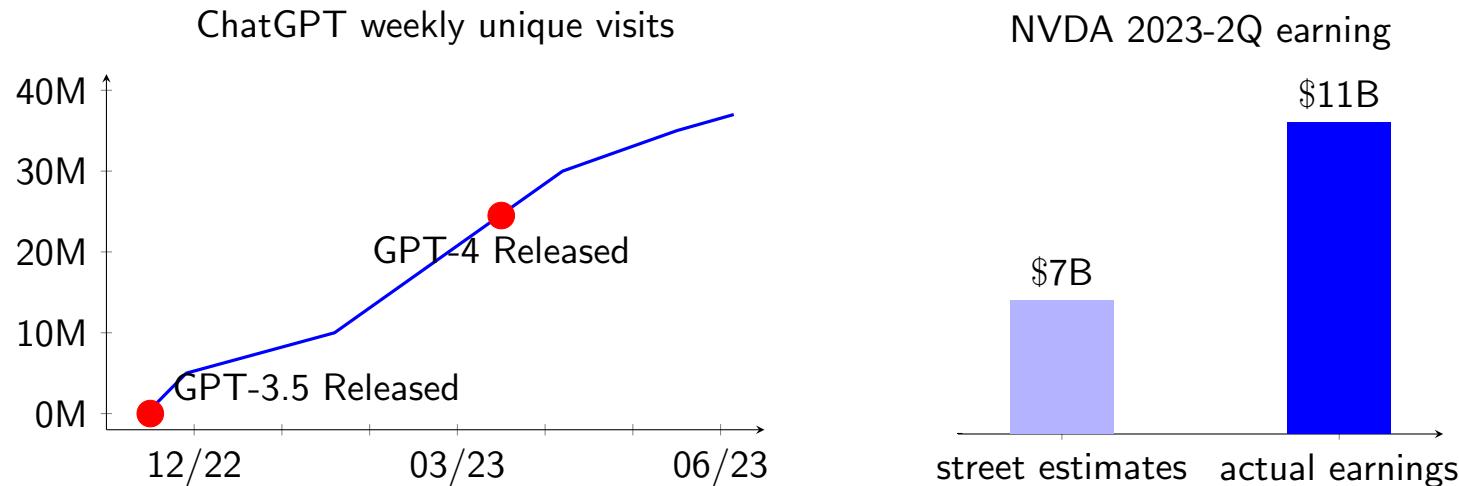
Where are we in AI today?

- 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

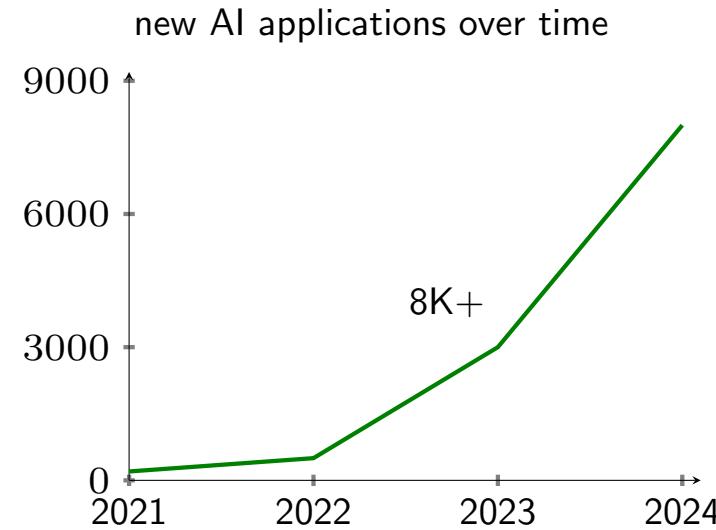
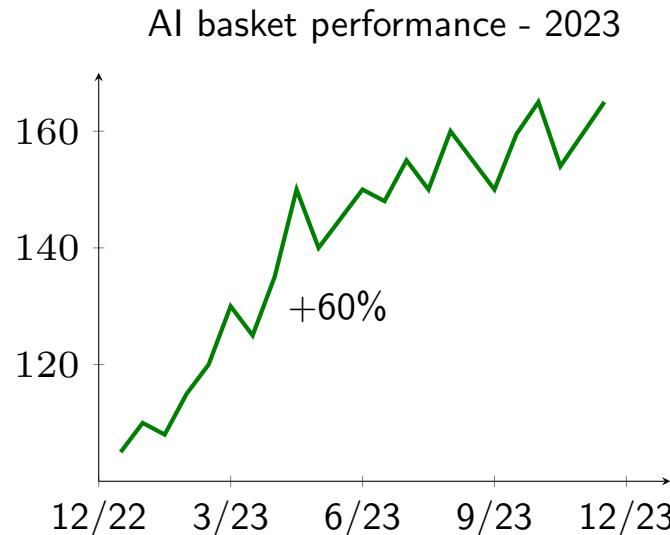
- took only *5 months for ChatGPT users to reach 35M*
- NVIDIA 2023 Q2 earning exceeds market expectation by big margin - \$7B vs \$13.5B
 - surprisingly, *101% year-to-year growth*
 - even more surprisingly *gross margin was 71.2%* - up from 43.5% in previous year⁴



⁴source - Bloomberg

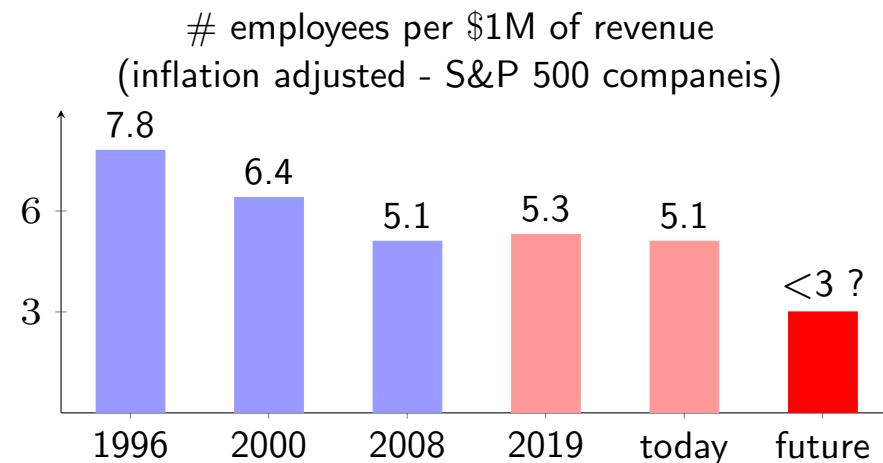
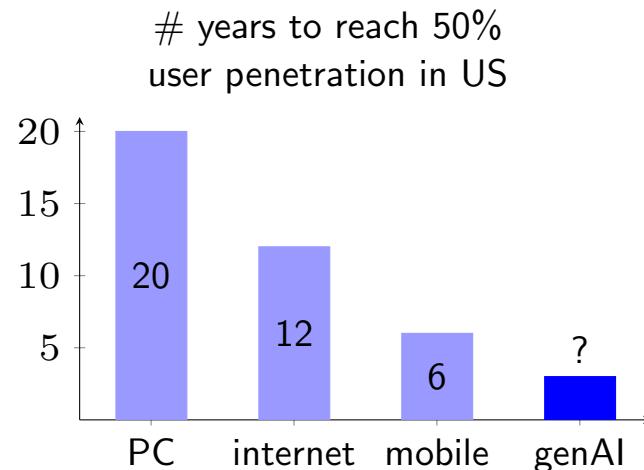
Explosion of AI ecosystems - AI stock market

- *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



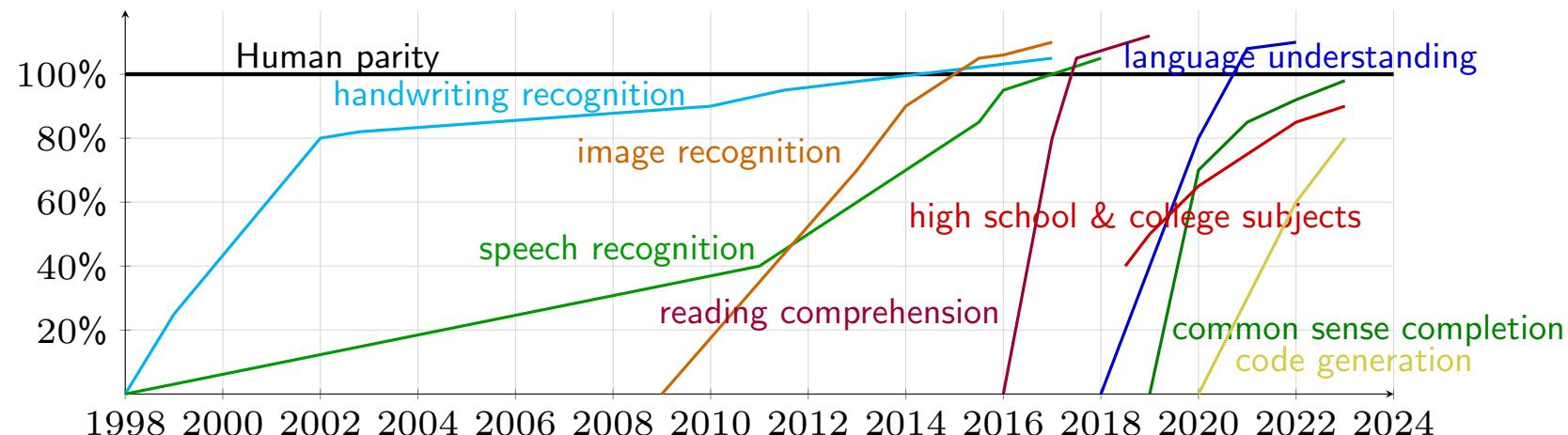
AI's transformative impact - adoption speed & economic potential

- 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



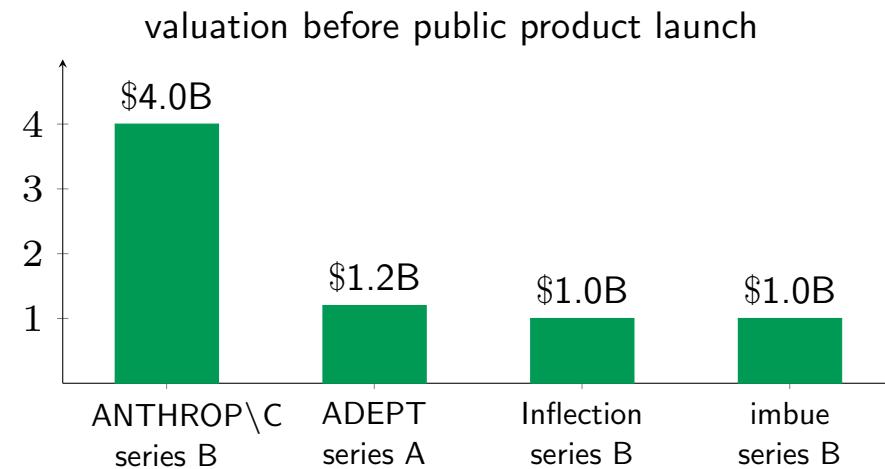
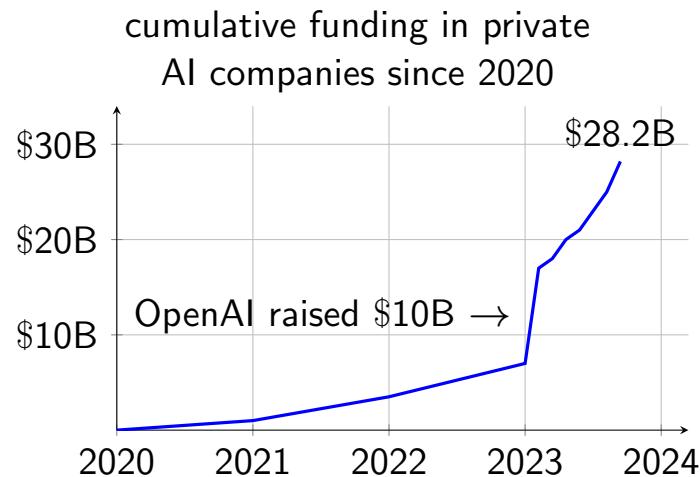
AI getting more & more faster

- 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



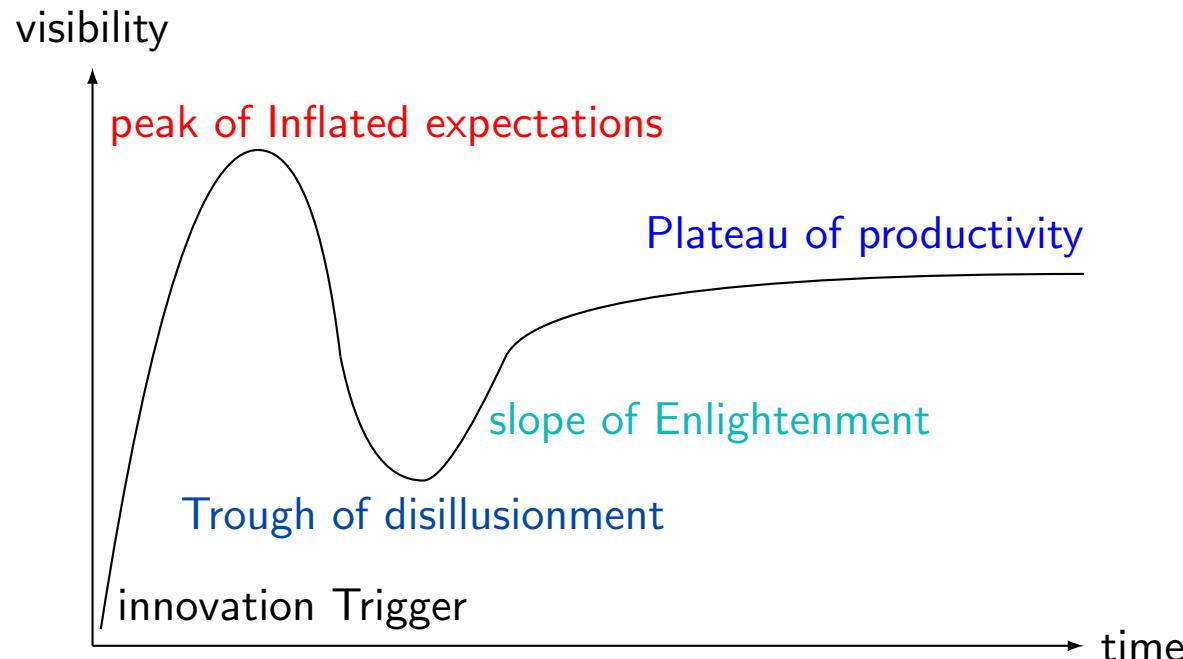
Massive investment in AI

- *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 stellar 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



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 companies with data centers
 - *big 4 hyperscalers generate \$150B + annual revenue*



Yes & No

characteristics of hype cycles	speaker's views
value accrual misaligned with investment	<ul style="list-style-type: none">• OpenAI still operating at a loss; business model <i>still</i> not clear• gradual value creation across broad range of industries and technologies (<i>e.g.</i>, CV, LLMs, RL) unlike fiber optic bubble in 1990s
overestimating timeline & capabilities of technology	<ul style="list-style-type: none">• self-driving cars delayed for over 15 years, with limited hope for achieving level 5 autonomy• AI, however, has proven useful within a shorter 5-year span, with enterprises eagerly adopting
lack of widespread utility due to technology maturity	<ul style="list-style-type: none">• AI already providing significant utility across various domains• vs quantum computing remains promising in theory but lacks widespread practical utility

Some Important Questions around AI

Some important questions around AI

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

Human-level AI?

Why human-level in the first place?

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

What makes DL so successful?

Factors contributing to astonishing success of DL

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

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

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

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

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

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

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

Sudden leap in LLM performance

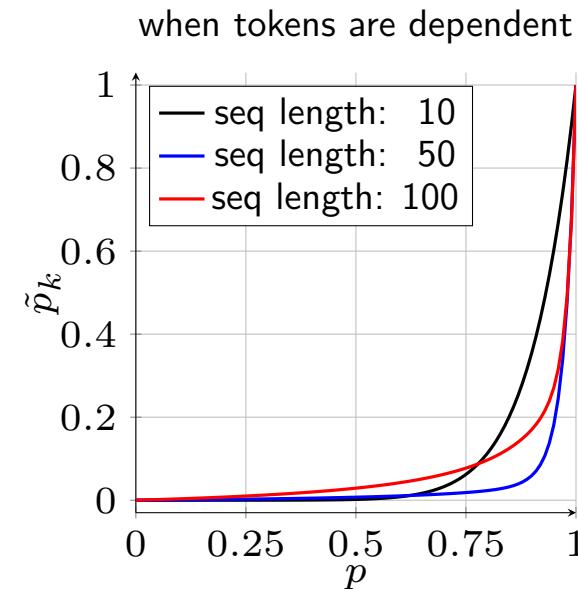
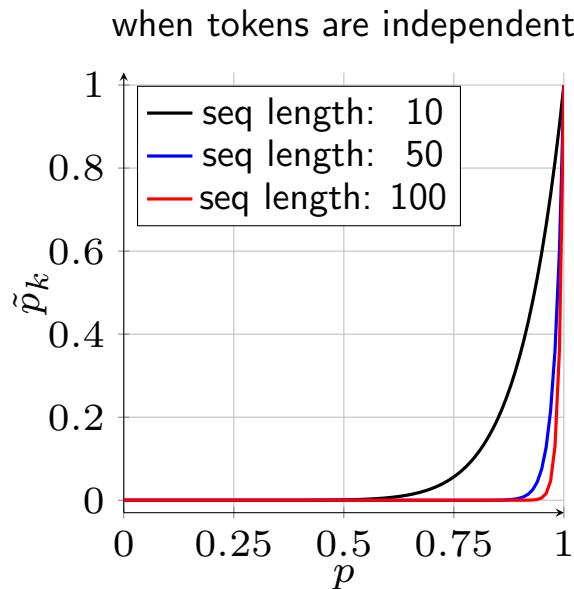
Probability inferred sequence is correct

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

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

Dramatic improvement of LLM near saturation

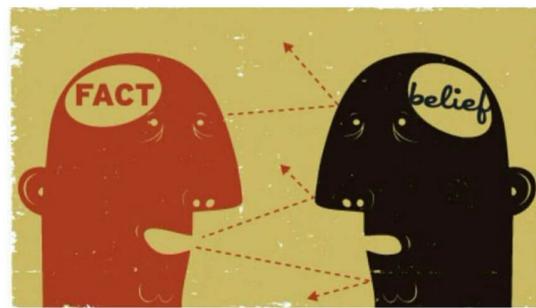
- do simulations for both independent & dependent cases
 - assume p_i are same for all i
- (for both cases) sequence inference improves dramatically as p approaches 1
- this explains *why we have observed sudden dramatic performance improvement of certain seq2seq learning technologies*, e.g., LLM



Biases

Cognitive biases attributed to humans

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



Biases of LLMs

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

AI Ethics

Ethical issues related to AI

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

AI related Legal Issues

Legal issues with ethical consideration

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

Consciousness

Consciousness

- what is consciousness, anyway?
 - recognizes itself as independent, autonomous, valuable entity?
 - recognizes itself as living being, unchangeable entity?
- no agreed definition on consciousness exists yet
 - . . . and will be so forever
- does it have anything to do with the fact that humans are biologically living being?
- is SKYNET ever plausible?
 - can AI have *desire* to survive (or save earth)?



Utopia vs Dystopia

Utopia vs dystopia



- not important questions (at all) *I think . . .*
- what we should focus on is *not* the possibilities of doomsday or Judgment Day, but rather
 - our limits on controlling unintended impacts of AI
 - *misuse* by (greedy, immoral, and unethical) people possessing social, economic & political power
 - *social good and welfare impaired* by either exploiting AI or ignorance of (inner workings of) AI
- should concern
 - choice or balance among utilitarianism, humanitarianism & values
 - amend or improve laws/regulations
 - ethical issues caused by AI

Knowledge, Belief, and Reasoning

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

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

LLMs . . .

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

conditional probability model looks like human with intelligence

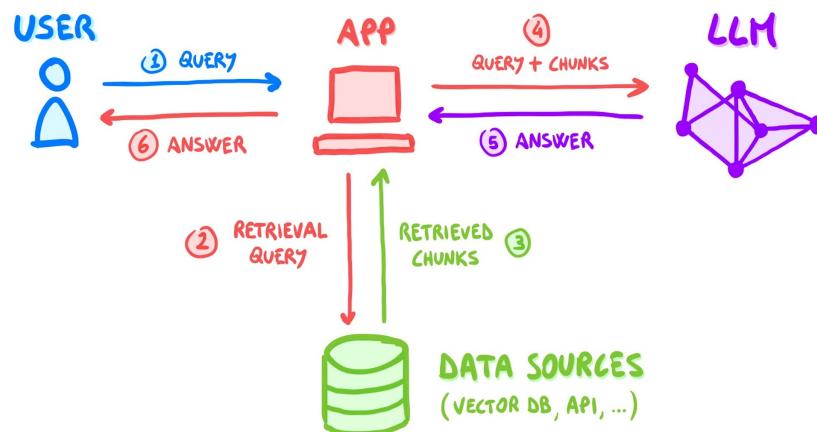
- making vulnerable to anthropomorphism
- examine it by throwing questions such as
 - “*does LLM have knowledge and belief?*”
 - “*can it reason?*”

What LLM really does!

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

LLMs vs systems in which they are embedded

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



How ChatBot works?

- conversational AI agent does *in-context learning* or *few-shot prompting*
- for example,

- when the user enters
who is the first person to walk on the Moon?

- ChatBot, LLM-embedded system, feeds the following to LLM

User, a human, and BOT, a clever and knowledgeable AI agent.

User: what is 2+2?

BOT: the answer is 4.

User: where was Albert Einstein born?

BOT: he was born in Germany.

User: who is the first person to walk on the Moon?

BOT:

Knowledge, belief & reasoning around LLM

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

does LLM have belief?
or
do humans have knowledge?
- let us discuss them in two different perspectives
 - laymen's perspectives
 - cognitive scientific & philosophical perspectives

Laymen's perspectives on knowledge, belief & reasoning

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

Knowledge

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



Belief

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



can you provide me up-to-date information about AI agent perplexity?

Sources

- Perplexity AI: A Guide for Beginners - Jeff Su
- Perplexity AI's value to our startup as it raises new \$500K funding - siliconangle
- Perplexity At A Deep Dive - Reflections
- anopose
- Show all

Perplexity

Perplexity AI is an AI-powered search engine and answer engine that has gained significant traction in recent years. Founded in 2022, Perplexity uses large language models (LLMs) to generate answers to user queries, citing sources from the web within the text response. As of Q1 2024, Perplexity had reached 15 million monthly users.

Key features and developments of Perplexity AI include:

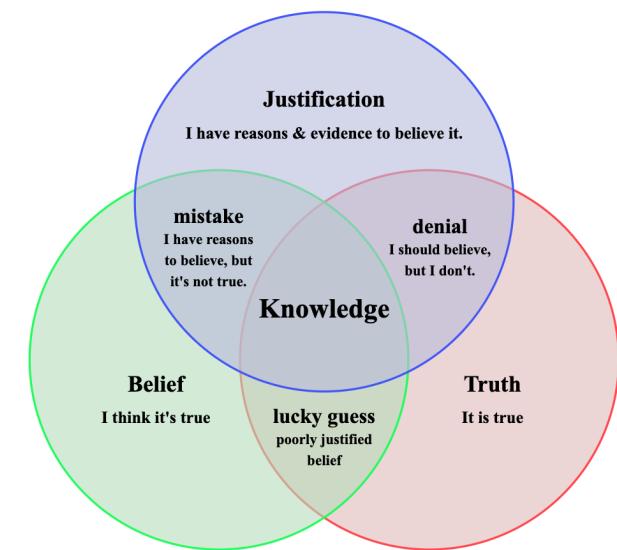
1. Search functionality: Perplexity summarizes search results and produces text with inline citations, providing personalized results based on the context of user queries

Ask follow-up Pro



Knowledge in philosophical and cognitive scientific sense

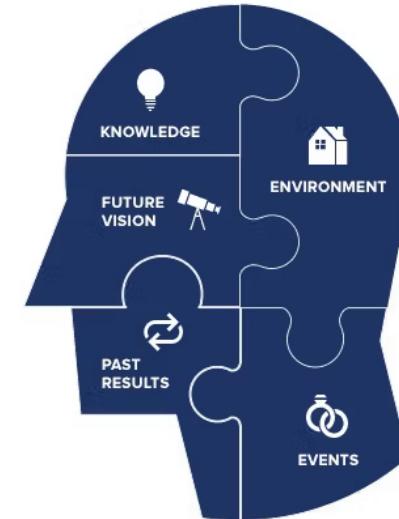
- does LLM have knowledge?
 - Sunghee: *I don't think so!*
- why?
 - we say we have “knowledge” when
“we do so against ground of various human capacities that we all take for granted when we engage in everyday conversation with each other.”
 - when asked *“who is Tom Cruise’s mother?”*, it says *“Tom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



Belief in philosophical and cognitive scientific sense

- for the discussion
 - do *not* concern any specific belief
 - but concern *prerequisites for ascribing any beliefs to AI system*
- so does it have belief?
 - nothing can count as belief about the world we share unless
 - it is against ground of the ability to update beliefs appropriately in light of evidence from that world, essential aspect of the capacity to distinguish truth from falsehood*
 - LLM does not have this ground, essential consideration when deciding whether it *really* had beliefs.
- Sunghee: so *no, LLM cannot have belief!*

WHERE DO YOUR
BELIEFS COME FROM?



Reasoning in philosophical and cognitive scientific sense

- note reasoning is *content neutral*
 - e.g., following logic is perfect regardless of truth of premises
 - hence, no access to outside world does *not* disqualify
- when asked “*if humans are immortal, would Socrates have survived today?*”, LLM says “*... it's logical to conclude that Socrates would likely still be alive today. ...*”
- however, remember, once again, what we just asked it to do is *not* “deductive inference”
given the statistical distribution of words in public corpus, what words are likely to follow the sequence, “humans are immortal and Socrates is human therefore.”
- Sunghee: so *no, LLM cannot reason, either!*
- but, LLM
 - pretends to reason, and from which capabilities, we can benefit!
 - also, can *mimic even multi-step reasoning whose inferencing structure is complicated* using *chain-of-thoughts prompting*, i.e., *in-context learning* or *few-short prompting*

Simple example showing LLM not possessing knowledge



- User
"Who is Tom Cruise's mother?"
- LLM(-embedded question-answering system) (as of Jan 2022)
"Tom Cruise's mother is Mary Lee Pfeiffer. She was born Mary Lee South. . . . Information about his family, including his parents, has been publicly available, . . . "
- User
"Who is Mary Lee Pfeiffer's son?"
- LLM(-embedded question-answering system) (as of Jan 2022)
"As of my last knowledge update in January 2022, I don't have specific information about Mary Lee Pfeiffer or her family, including her son. . . . "

Risk of anthropomorphization

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

Moral

- AI shows incredible utility and commercial potentials, hence should
 - make informed decisions about trustworthiness and safety
 - avoid ascribing capacities they lack
 - *take best utilization of remarkable capabilities of AI*
- today's AI so powerful, so (seemingly) convincingly intelligent
 - obfuscate mechanism
 - actively encourage *anthropomorphism* with philosophically loaded words like “*believe*” and “*think*”
 - easily mislead people about character and capabilities of AI
- matters not only to scientists, engineers, developers, and entrepreneurs, but also
 - *general public, law & policy makers, journalists, . . .*

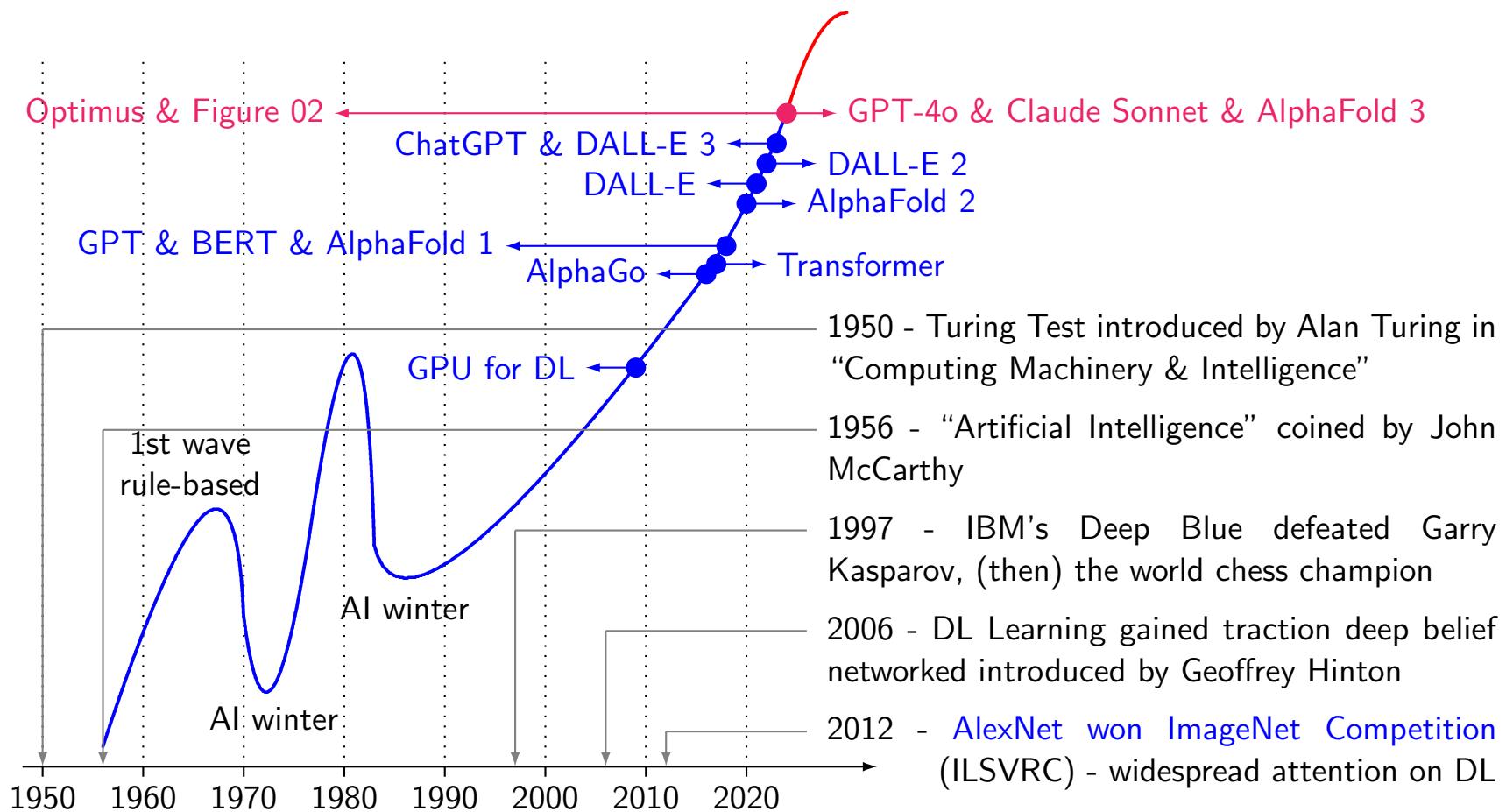
Silicon Valley's Cultural Engine of Innovation and Disruption

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* ☺

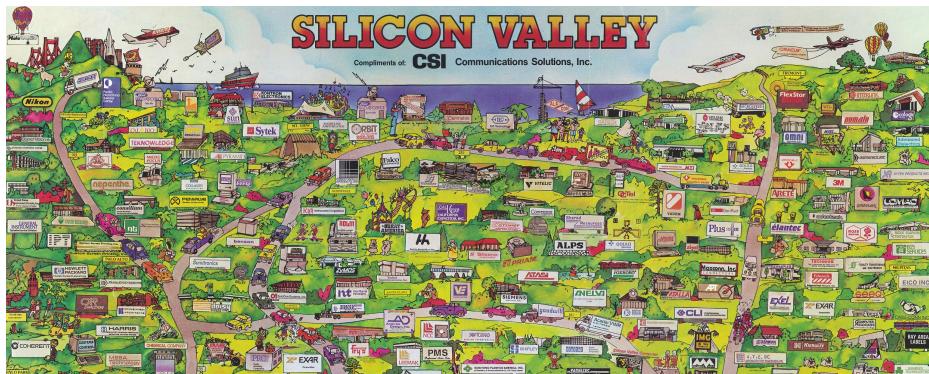


Joining Amazon.com, Inc. at the inflection point of AI



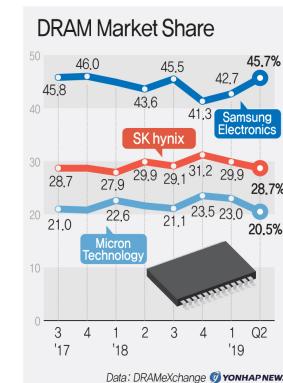
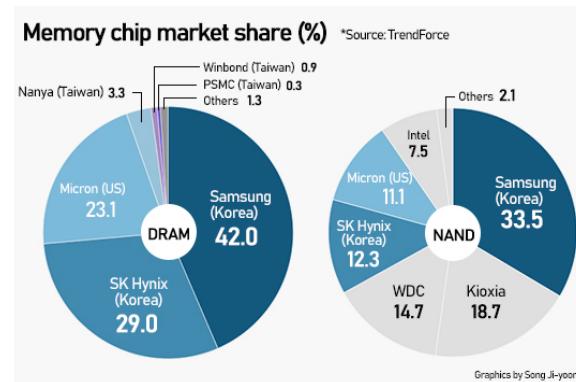
Innovation ecosystem of Silicon Valley

- key characteristics
 - risk-taking culture, *trust* in technology → *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

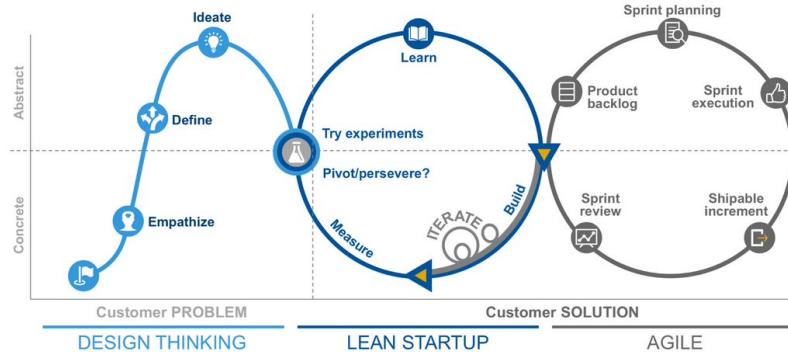
- Amazon's culture & leadership principles
 - customer obsession as driver of innovation
 - high standards & ownership culture, disagree & commit
 - bias for action and long-term thinking - sounds contradictory?
 - mechanisms like “two-pizza teams” & “Day One” for (or rather despite) scalability
- lessons 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

Combine Design Thinking, Lean Startup and Agile



Product-Market Fit (PMF)



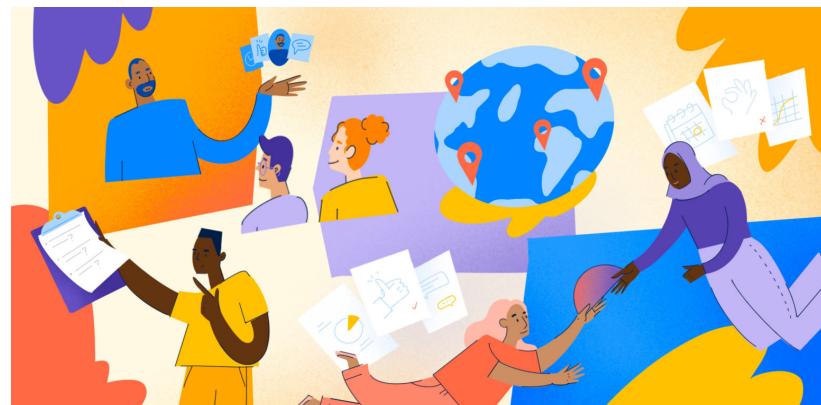
Bridging Silicon Valley & Korea

- 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. Halevy, 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

- [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 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.
- [HNF09] Alon Halevy, Peter Norvig, and Nanediri Fernando. The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24:8 – 12, 05 2009.
- [Kah11] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.
- [KW19] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019.

- [Sha23] Murray Shanahan. Talking about large language models, 2023.
- [VSP⁺17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of 31st Conference on Neural Information Processing Systems (NIPS)*, 2017.

Thank You