[San Jose State University Special AI Lecture Series IV - Philosophy, Ethics & Consciousness]
AI's Philosophical Frontiers - Consciousness, Knowledge, Belief, Ethics, and Limits of Machine Reasoning

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)
CGO / Global Managing Partner @ LULUMEDIC
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 ∼
- *CGO / Global Managing Partner @ LULUMEDIC, Seoul, Korea* 2025 ∼
- *KFAS-Salzburg Global Leadership Fellow @ Salzburg Global Seminar, Austria* 2024 ∼
- *Adjunct Professor, EE Department @ Sogang University, Seoul, Korea* 2020 ∼
- *Advisory Professor, EECS Department @ DGIST, Korea* 2020 ∼
- *AI-Korean Medicine Integration Initiative Task Force Member @ The Association of Korean Medicine, Seoul, Korea* 2025 ∼
- *Director of AI Semiconductor @ K-BioX, CA, USA* 2025 ∼
- Global Advisory Board Member @ Innovative Future Brain-Inspired Intelligence System Semiconductor of Sogang University, Korea 2020 ∼
- Technology Consultant @ Gerson Lehrman Gruop (GLG), NY, USA 2022 ∼
- Chief Business Development Officer @ WeStory.ai, Cupertino, CA, USA 2025 ∼
- Advisor @ CryptoLab, Inc., Seoul, Korea 2025 ∼

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

Highlight of Career Journey

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

Unpacking AI

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 advanages for medical assitance, emergency handling
  - AI does not make more mistakes because task is repetative and tedius
  - AI does not request salary raise or go on strike

What makes DL so successful?

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

  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 inferenced sequence is correct

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

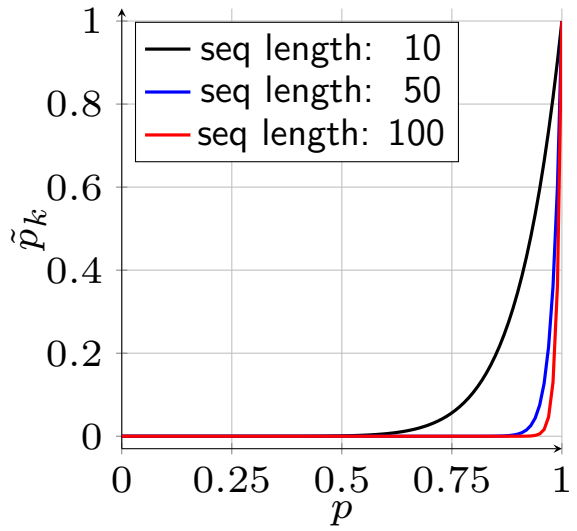
$$\rho_i = \frac{\tilde{p}_i - \tilde{p}_{i-1} p_i}{\sqrt{\tilde{p}_{i-1}(1 - \tilde{p}_{i-1}) p_i (1 - p_i)}}$$

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

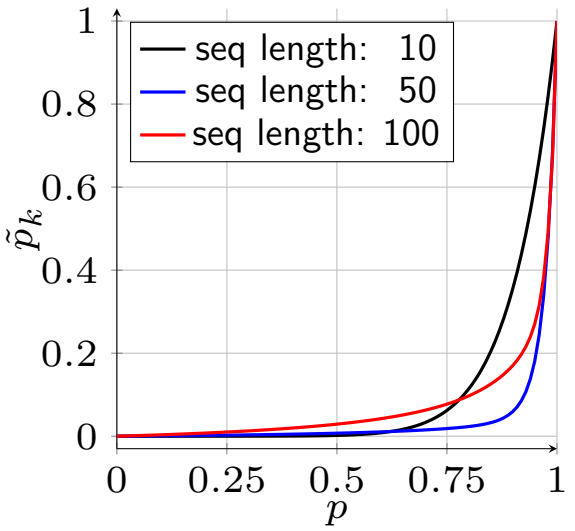
Dramatic improvement of LLM near saturation

- 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

when tokens are independent



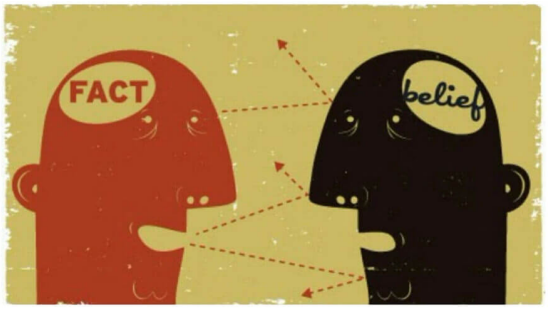
when tokens are dependent

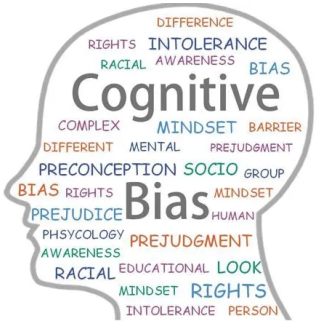


Biases

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 subjec to
  - availability bias - baised 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 eaisly inferencing what it saw, $i.e.$, data it trained on
  - halo effect - overemphasize on what prestigious figures say
    - LLM trained by imbalanced # reports about prestigious figures
- similar facts true for other types of ML models,
  - $e.g.$, video caption, text summarization, sentiment analysis
- cognitive biases only human represent
  - confirmation bias, hindsight bias, confidence bias, optimistic bias, anchoring bias, negativity bias, framing effect

AI Ethics

Ethical issues related to AI

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

AI related Legal Issues

Legal issues with ethical consideration

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

Consciousness

Consciousness

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

Utopia vs Dystopia

Utopia vs dystopia

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

Knowledge, Belief, and Reasoning

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

What categories of questions do they belong to?

engineering, scientific, philosophical, cognitive scientific, . . . ?

LLMs . . .

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

  conditional probability model looks like human with intelligence

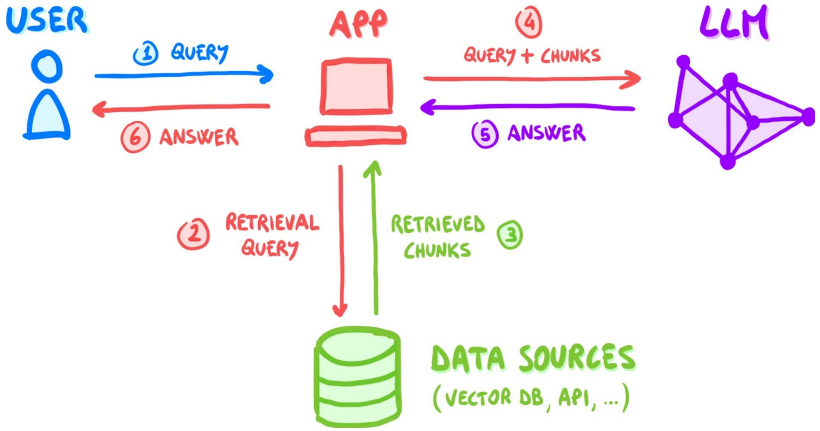
  - making vulnerable to anthropomorphism
- examine it by throwing questions such as
  - *"does LLM have knowledge and belief?"*
  - *"can it reason?"*

What LLM really does!

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

LLMs vs systems in which they are embedded

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



How ChatBot works?

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

    ```
    who is the first person to walk on the Moon?
    ```

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

    ```
    User, a human, and BOT, a clever and knowledgeable AI agent.
    User:  what is 2+2?
    BOT: the answer is 4.
    User:  where was Albert Einstein born?
    BOT: he was born in Germany.
    User:  who is the first person to walk on the Moon?
    BOT:
    ```

Knowledge, belief & reasoning around LLM

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

    does LLM have belief?

    or

    do humans have knowledge?

- let us discuss them in two different perspectives
  - laymen's perspectives
  - cognitive scientific & philosophical perspectives

Laymen's perspectives on knowledge, belief & reasoning

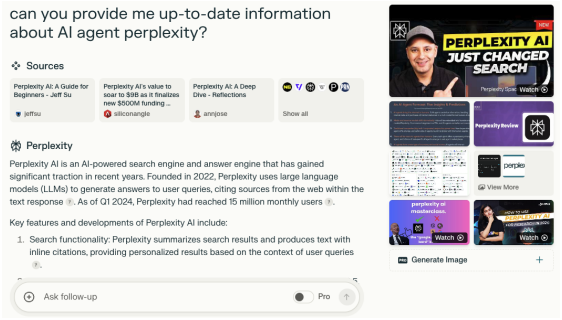
- does (good) LLM have knowledge?
  - Grandmother: looks like it cuz when instructed *"explaing big bang"*, it says
    " The Big Bang theory is prevailing cosmological model that explains the origin and evolution of the universe. . . . 13.8 billion years ago . . . "
- does it have belief?
  - Grandmother: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

Knowledge

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

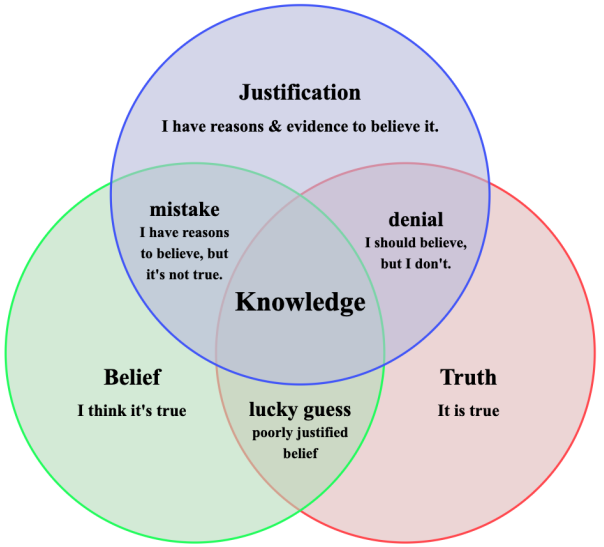
Belief

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



Knowledge in philosophical and cognitive scientific sense

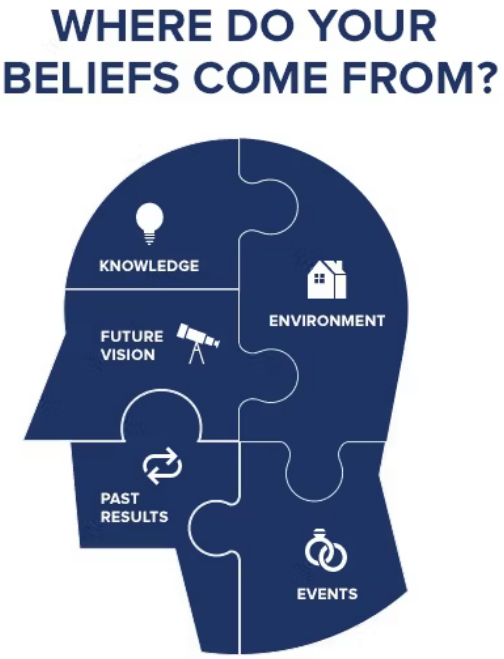
- does LLM have knowledge?
  - Sunghee: *I don't think so!*
- why?
  - we say we have "knowledge" when
    "we do so against ground of various human capacities that we all take for granted when we engage in everyday conversation with each other."
  - when asked *"who is Tom Cruise's mother?"*, it says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



Belief in philosophical and cognitive scientific sense

- for the discussion
  - do *not* concern any specific belief
  - but concern *prerequisites for ascribing any beliefs to AI system*
- so does it have belief?
  - nothing can count as belief about the world we share unless

    it is against ground of the ability to update beliefs appropriately in light of evidence from that world, essential aspect of the capacity to distinguish truth from falsehood
  - LLM does not have this ground, essential consideration when deciding whether it *really* had beliefs.
- Sunghee: so *no, LLM cannot have belief!*



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 Socreates 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 soceital 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

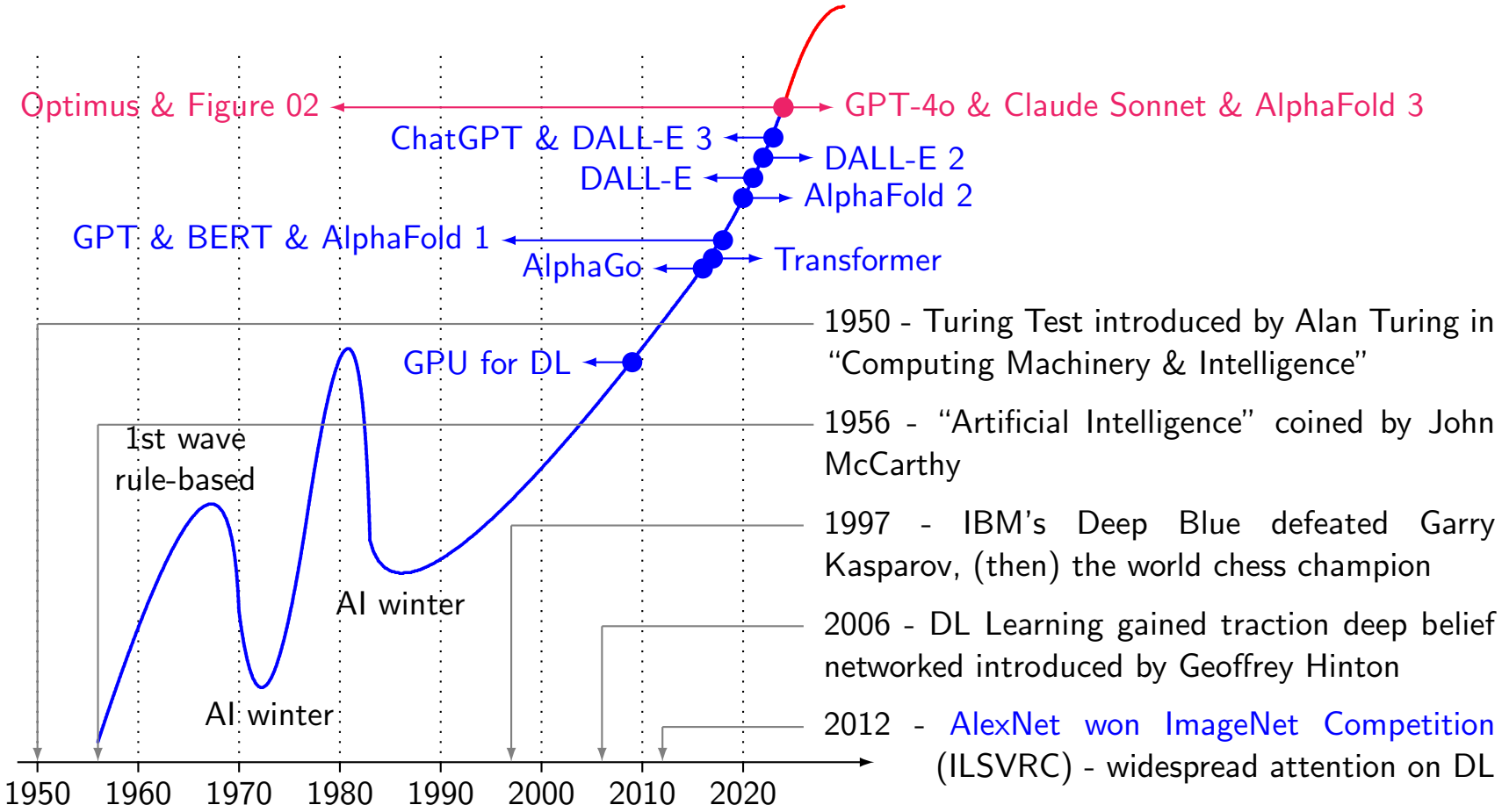
My journey from Samsung & Amazon to Gauss Labs & Erudio Bio

- 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* ☺



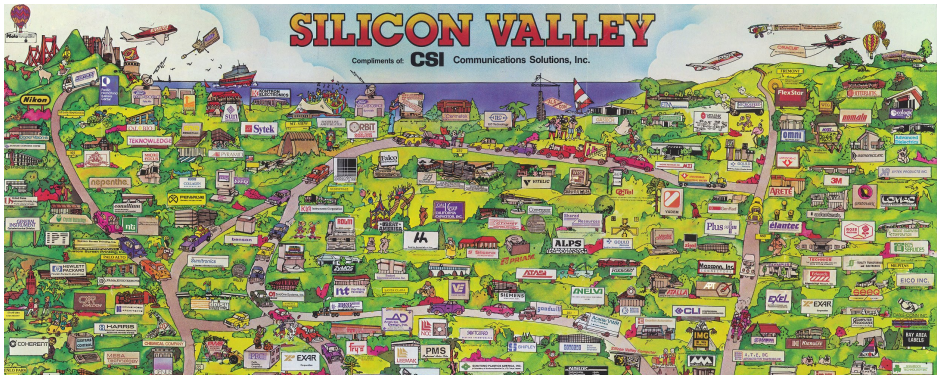


Joining Amazon.com, Inc. at the inflection point of AI



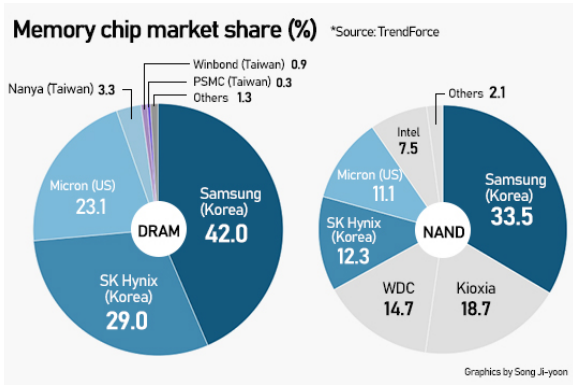
Innovation ecosystem of Silicon Valley

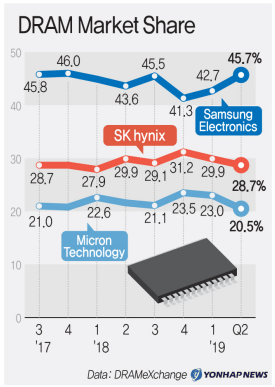
- key characteristics
  - risk-taking culture, *trust* in technology → *genuine* respect for engineers and scientists
  - easy access to huge capital - VCs, angel investors alike
  - talent density - engineers, researchers, scientists, entrepreneurs, PMs, TPMs, . . .
  - diversity, "collision density" 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

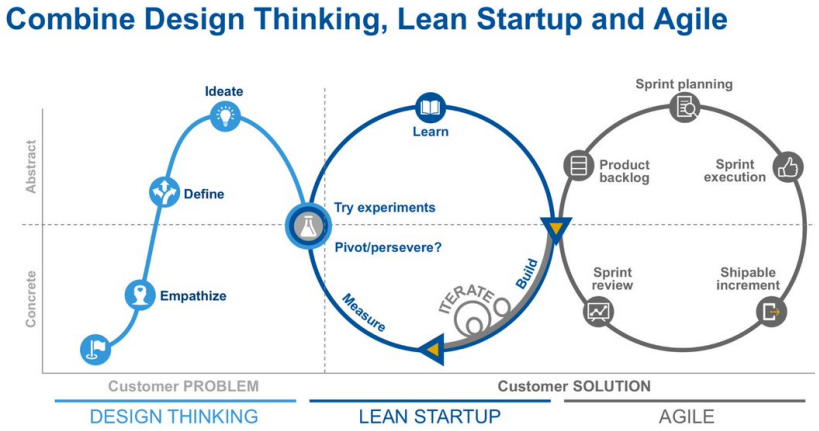
- Amazon's culture & leadership principles
  - customer obsession as driver of innovation
  - high standards & ownership culture, disagree & commit
  - bias for action and long-term thinking - sounds contradictory?
  - mechanisms like "two-pizza teams" & "Day One" for (or rather despite) scalability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

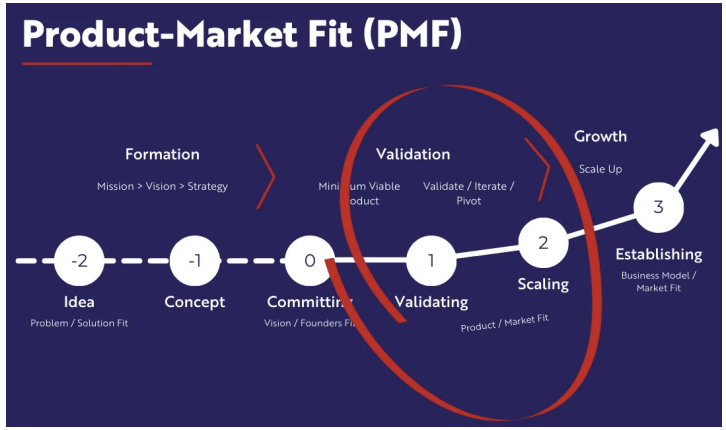




Founding and scaling startups

- 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





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

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

Selected References & Sources

Selected references & sources

- Robert H. Kane "Quest for Meaning: Values, Ethics, and the Modern Experience" 2013
- Michael J. Sandel "Justice: What's the Right Thing to Do?" 2009
- Daniel Kahneman "Thinking, Fast and Slow" 2011
- Yuval Noah Harari "Sapiens: A Brief History of Humankind" 2014
- M. Shanahan "Talking About Large Language Models" 2022
- A.Y. Halevry, P. Norvig, and F. Pereira "Unreasonable Effectiveness of Data" 2009
- A. Vaswani, et al. "Attention is all you need" @ NeurIPS 2017
- S. Yin, et. al. "A Survey on Multimodal LLMs" 2023
- Chris Miller "Chip War: The Fight for the World's Most Critical Technology" 2022

- CEOs, CTOs, CFOs, COOs, CMOs & CCOs @ startup companies in Silicon Valley
- VCs on Sand Hill Road - Palo Alto, Menlo Park, Woodside in California, USA

References

References

[HNF09] Alon Halevy, Peter Norvig, and Nanediri Fernando. The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24:8 – 12, 05 2009.

[Kah11] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.

[Sha23] Murray Shanahan. Talking about large language models, 2023.

Thank You