



fintechjunkie @fintechjunkie

Nov 15, 2023 · 33 tweets · [fintechjunkie/status/1724771585796510133](https://twitter.com/fintechjunkie/status/1724771585796510133)

AI is undeniably going to change the world but 99% of AI companies will fizzle and die.

I have real, actionable advice about how to build durable AI companies but understanding my perspective requires a primer on how I think about AI.

Trust me, you'll like it: 📚👉



Artificial Intelligence (AI) has become the focus of the investing community and business world over the past 12-24 months but it isn't a new discipline. Dating back to the 1950s, AI is focused on creating machines capable of mimicking human intelligence.

There are many forms of AI, but the sub-categories of Generative AI and Large Language Models have captured the world's imagination because they feel "human" and "magical".

Together, they form the basis for AI learning machines that are starting to produce human-like outputs.

Fintechjunkie's Taxonomy for Learning Machines

There are two types of learning machines: Carbon based learning machines (people) and Silicon based learning machines (computers).

How these machines go from tabula rasa to complete mastery of a specific skill differs massively.

Carbon Based LMs (CBLMs)

Step 1: Learning to Learn

CBLMs spend the first 18 years of their lives learning the basics of how the world works and how to learn new skills. This isn't cheap.

The estimated cost of raising a child from birth to age 18 in the U.S. is \$237,482.

Step 2: Learning a Skill

CBLMs spend anywhere from a few months learning a trade to 6+ years for skills that require advanced degrees (i.e. - doctors, scientists, engineers).

Costs can easily cross into the hundreds of thousands of dollars depending on how much on-the-job training and additional schooling is required.

And programs that aim to teach complex skills have high dropout rates which implies that not all CBLMs can learn all skills.

Step 3: Mastering a Skill

Over time, CBLMs have trouble finding work if they can't perform their learned skills proficiently.

And some CBLMs self-select out because they aren't excited to use their skills (only 46% of college grads work in their original field of study).

Most CBLMs are able to improve their skills with repetition. But studies show that only 20-25% of skill mastery is correlated with practice with the rest correlated to factors that are highly dependent on the specific skill being mastered (i.e. – genetics, coordination, IQ).

Cost to Run the Machine

Low-skill CBLMs cost \$0.50 a minute in compute + support costs to run (labor costs plus overhead).

Mid-skill CBLMs cost \$1-\$2 a minute in compute + support costs to run.

High-skill CBLMs can cost \$10+ a minute in compute + support costs to run.

And as CBLMs gain experience, the cost to “run the machine” (labor costs plus overhead) increases, sometimes dramatically.

Almost all CBLMs increase their running costs over time with some CBLMs increasing their cost by 10X+ before they’re retired and replaced.

Silicon Based Learning Machines (SBLMs)

Step 1: Learning to Learn

Infinite SBLMs can be spun up that are "born" already preloaded with the ability to learn. Their "parent" is typically a fully trained large language model (LLMs) that passes on its knowledge to its children.

Blank slate SBLMs aren’t expensive to create but renting or buying the right chips and compute to train the parent LLMs can be expensive (many millions of dollars).

And choosing the right LLM to use as a template will impact the quality of the SBLM once completed.

Step 2: Learning a Skill

SBLMs are trained on data that defines the inputs and outputs of the skill being learned. Most training involves expert CBLMs (people) that help refine and improve the SBLMs’ skills. The cost to train a SBLM are coming down rapidly.

The time it takes to train a parent LLM is currently measured in months with estimates of 3-6 months for today’s well-known LLMs.

But the time to train a SBLM can take a fraction of this time, and once trained, a SBLM can be cloned instantaneously for infinite use.

Step 3: Mastering a Skill

Once a SBLM has been trained, it will perform at the same skill level until adjusted. Some SBLMs can learn based on feedback from users, but many SBLMs will only improve if their parent LLM is improved or they are retrained using additional data.

Cost to Run the Machine

SBLMs are massively cheaper to run than CBLMs as measured by cost-per-desired-output. Everything is measured in “tokens used” and a typical long chat might consume a few thousand tokens that cost \$0.05-\$0.10 per thousand.

And text is only one form of output. Image-based outputs like logos, schematics and illustrations currently cost \$0.05-\$0.10 per image.

Video is a newer format, but it’s already estimated that the cost to produce an animated feature will be cut by 90%+ in the next few years.

And repetitive tasks can be copied such that they re-use outputs (tokens) instead of consuming new compute. These shortcuts reduce the cost of running the machine as it's trained which is a stark contrast to CBLMs that increase the cost to run as they gain experience.

Example - Graphic design work needed to create a logo for a new company

CBLM:

Interacts with client to get specs and then promises 8-10 sample logos to deliver for review in a few days. Compute time 480 minutes of work billed at \$2 a minute.

CBLM meets with their client to review and receive feedback and then makes revisions over the next few days using 60 minutes of compute billed at \$2 a minute.

Total compute cost \$1,080. Delivery time: Days.

SBLM:

Interacts with client to get specs and generates 8-10 sample logos in 2 minutes. Compute cost \$0.50 in tokens for gathering specs and \$1.00 in tokens to generate images. Process continues with no delays.

SBLM gathers feedback on favorite logo and generates 3-4 refined logos based on revised preferences. Compute cost \$0.25 in tokens for refined specs and \$0.50 in tokens to generate images.

Total compute cost \$2.25. Delivery time: Minutes.

Example - Customer service

CBLM:

Interacts with customer to authenticate who they are (2 min), asks questions to understand their problem (5 min), pulls up policy docs and explains the proposed resolution to the customer (10 min), executes solution and documents system (3 min).

Total compute time 20 minutes at a loaded cost of \$1 a minute = \$20.

Loaded costs are inclusive of compute costs, benefits, management overhead, real estate, etc.

Accuracy: High for simple questions with an increasing error rate as customer requests become more complex.

SBLM:

Interacts with customer to authenticate who they are, asks questions to understand their problem, pulls up policy docs and explains the proposed resolution to the customer, executes solution and documents system.

Total compute cost of \$0.50.

Accuracy: Near 100%
Interesting Random Factoids

Each CBLM bears the full costs of learning while an infinite number of identical SBLMs can be cloned at no additional cost once a SBLM has been trained.

Each CBLM has a different capacity to learn any given skill while SBLMs can learn any skill.

CBLMs eventually need to be retired and replaced while SBLMs have an infinite lifespan and can be retained at any time.

CBLMs aren't good at adjusting historically learned skills with new information/feedback loops while SBLMs are designed to be "lifelong learning machines".

The costs for CBLMs will go up with inflation while the costs of SBLMs will come down as compute costs are reduced and models are refined.

Today there aren't many problems that can be solved end-to-end by SBLMs but this number is projected to grow rapidly.

Where we are today

CBLMs have existed for thousands of years and SBLMs trained using LLM technology have existed for a few years.

CBLMs have been trained over millennia to solve many classes of problems proficiently. Every generation builds upon the learnings of the past.

SBLMs require massive data sets to train on and for many problems these data sets either have yet to be assembled or a LLM has yet to be trained that makes use of certain large data sets.

As a result, for many problems, a hybrid model is being employed to complete tasks.

These hybrid models employ SBLMs and CBLMs to work together. The SBLM gets 90% of the work done quickly and then the CBLM completes the task. This works for skills that SBLMs can't yet master or skills that require a CBLM for final signoff.

This general framework will hopefully serve as the foundation for future threads about AI and how it will impact the startup and investing worlds.

I'm keen on sharing where ephemeral value will be created by AI startups and what it will take to build durable value. Stay tuned!

...