Enterprise Philosophy and The First Wave of AI

🔲 **stratechery.com**/2024/enterprise-philosophy-and-the-first-wave-of-ai

Tuesday, September 24, 2024

The popular history of technology usually starts with the personal computer, and for good reason: that was the first high tech device that most people ever used. The only thing more impressive than the sheer audacity of "A computer on every desk and in every home" as a corporate goal, is the fact that Microsoft accomplished it, with help from its longtime competitor Apple.

In fact, though, the personal computer wave was the 2nd wave of technology, particularly in terms of large enterprises. The first wave — and arguably the more important wave, in terms of economic impact — was the digitization of back-end offices. These were real jobs that existed:

These are bookkeepers and and tellers at a bank in 1908; fast forward three decades and technology had advanced:

The caption for this 1936 Getty Images photo is fascinating:

> The new system of maintaining checking accounts in the National Safety Bank and Trust Company, of New York, known as the "Checkmaster," was so well received that the bank has had to increase its staff and equipment. Instead of maintaining a mininum balance, the depositor is charged a small service charge for each entry on his statement. To date, the bank has attracted over 30,000 active accounts. Bookkeepers are shown as they post entries on "Checkmaster" accounts.

It's striking how the first response to a process change is to come up with a business model predicated on covering new marginal costs; only later do companies tend to consider the larger picture, like how low-marginal-cost checking accounts might lead to more business for the bank overall, the volume of which can be supported thanks to said new technology.

Jump ahead another three decades and the back office of a bank looked like this:

Now the image is color and all of the workers are women, but what is perhaps surprising is that this image is, despite all of the technological change that had happened to date, particularly in terms of typewriters and calculators, not that different from the first two.

However, this 1970 picture was one of the last images of its kind: by the time this picture was taken, Bank of America, where this picture was taken, was already well on its way to transitioning all of its accounting and bookkeeping to computers; most of corporate America soon followed, with the two primary applications being accounting and enterprise resource planning. Those were jobs that had been primarily done by hand; now they were done by computers, and the hands were no longer needed.

Tech's Two Philosophies

In 2018 I described Tech's Two Philosophies using the four biggest consumer tech companies: Google and Facebook were on one side, and Apple and Microsoft on the other.

In Google's view, computers help you get things done — and save you time — by doing things for you. Duplex was the most impressive example — a computer talking on the phone for you — but the general concept applied to many of Google's other demonstrations, particularly those predicated on AI: Google Photos will not only sort and tag your photos, but now propose specific edits; Google News will find your news for you, and Maps will find you new restaurants and shops in your neighborhood. And, appropriately enough, the keynote closed with a presentation from Waymo, which will drive you…

Zuckerberg, as so often seems to be the case with Facebook, comes across as a somewhat more fervent and definitely more creepy version of Google: not only does Facebook want to do things for you, it wants to do things its chief executive explicitly says would not be done otherwise. The Messianic fervor that seems to have overtaken Zuckerberg in the last year, though, simply means that Facebook has adopted a more extreme version of the same philosophy that guides Google: computers doing things for people.

Google and Facebook's approach made sense given their position as Aggregators, which "attract end users by virtue of their inherent usefulness." Apple and Microsoft, on the other hand, were platforms born in an earlier age of computing:

This is technology's second philosophy, and it is orthogonal to the other: the expectation is not that the computer does your work for you, but rather that the computer enables you to do your work better and more efficiently. And, with this philosophy, comes a different take on responsibility. Pichai, in the opening of Google's keynote, acknowledged that "we feel a deep sense of responsibility to get this right", but inherent in that statement is the centrality of Google generally and the direct culpability of its managers. Nadella, on the other hand, insists that responsibility lies with the tech industry collectively, and all of us who seek to leverage it individually…

This second philosophy, that computers are an aid to humans, not their replacement, is the older of the two; its greatest proponent — prophet, if you will — was Microsoft's greatest rival, and his analogy of choice was, coincidentally enough, about transportation as well. Not a car, but a bicycle:

You can see the outlines of this philosophy in these companies' approaches to AI. Google is the most advanced, thanks to the way it saw the obvious application of AI to its Aggregation products, particularly search and advertising. Facebook, now Meta, has made major strides over the last few years as it has overhauled its recommendation algorithms and advertising products to also be probabilistic, both in response to the rise of TikTok in terms of customer attention, and the severing of their deterministic ad product by Apple's App Tracking Transparency initiative. In both cases their position as Aggregators compelled them to unilaterally go out and give people stuff to look at.

Apple, meanwhile, is leaning heavily into Apple Intelligence, but I think there is a reason its latest ad campaign feels a bit weird, above-and-beyond the fact it is advertising a feature that is not yet available to non-beta customers. Apple is associated with jewel-like devices that you hold in your hand and software that is accessible to normal people; asking your phone to rescue a fish funeral with a slideshow feels at odds with Steve Jobs making a movie on stage during the launch of iMovie:

That right there is a man riding a bike.

Microsoft Copilot

Microsoft is meanwhile — to the extent you count their increasingly fraught partnership with OpenAI — in the lead technically as well. Their initial product focus for AI, however, is decidedly on the side of being a tool to, as their latest motto states, "empower every person and every organization on the planet to achieve more."

CEO Satya Nadella said in the recent pre-recorded keynote announcing Copilot Wave 2:

> You can think of Copilot as the UI for AI. It helps you break down these siloes between your work artifacts, your communications, and your business processes. And we're just getting started. In fact, with scaling laws, as AI becomes more capable and even agentic, the models themselves become more of a commodity, and all the value gets created by you how steer, ground, fine-tune these models with your business data and workflow. And how it composes with the UI layer of human to AI to human interaction becomes critical.
>
> Today we're announcing Wave 2 of Microsoft 365 Copilot. You'll see us evolve Copilot in three major ways: first, it's about bringing the web plus work plus pages together as the new AI system for knowledge work. With Pages we'll show you how Copilot can take any information from the web or your work and turn it into a multiplayer AI-powered canvas. You can ideate with AI and collaborate with other people. It's just magical. Just like the PC birthed office productivity tools as we know them today, and the web makes those canvases collaborative, every platform shift has changed work artifacts in fundamental ways, and Pages is the first new artifact for the AI age.

Notice that Nadella, like most pop historians (including yours truly!), is reaching out to draw a link to the personal computer, but here the relevant personal computer history casts more shadow than light onto Nadella's analogy. The initial wave of personal computers were little more than toys, including the Commodore 64 and TRS-80, sold in your local Radio Shack; the Apple I, released in 1976, was initially sold as a bare circuit board:

Apple Introduces the First Low Cost Microcomputer System with a Video Terminal and 8K Bytes of RAM on a Single PC Card.

The Apple Computer. A truly complete microcomputer system on a single PC board. Based on the MOS Technology 6502 microprocessor, the Apple also has a built-in video terminal and sockets for 8K bytes of on-board RAM memory. With the addition of a keyboard and video monitor, you'll have an extremely powerful computer system that can be used for anything from developing programs to playing games or running BASIC.

Combining the computer, video terminal and dynamic memory on a single board has resulted in a large reduction in chip count, which means more reliability and lowered cost. Since the Apple comes fully assembled, tested & burned-in and has a complete power supply on-board, initial set-up is essentially "hassle free" and you can be running within minutes. At $666.66 (including 4K bytes RAM!) it opens many new possibilities for users and systems manufacturers.

You Don't Need an Expensive Teletype.

Using the built-in video terminal and keyboard interface, you avoid all the expense, noise and maintenance associated with a teletype. And the Apple video terminal is six times faster than a teletype, which means more throughput and less waiting. The Apple connects directly to a video monitor (or home TV with an inexpensive RF modulator) and displays 960 easy to read characters in 24 rows of 40 characters per line with automatic scrolling. The video display section contains its own 1K bytes of memory, so all the RAM memory is available for user programs. And the Keyboard Interface lets you use almost any ASCII-encoded keyboard. The Apple Computer makes it possible for many people with limited budgets to step up to a video terminal as an I/O device for their computer.

No More Switches, No More Lights.

Compared to switches and LED's, a video terminal can display vast amounts of information simultaneously. The Apple video terminal can display the contents of 192 memory locations at once on the screen. And the firmware in PROMS enables you to enter, display and debug programs (all in hex) from the keyboard, rendering a front panel unnecessary. The firmware also allows your programs to print characters on the display, and since you'll be looking at letters and numbers instead of just LED's, the door is open to all kinds of alphanumeric software (i.e., Games and BASIC).

8K Bytes RAM in 16 Chips!

The Apple Computer uses the new 16-pin 4K dynamic memory chips. They are faster and take ¼ the space and power of even the low power 2102's (the memory chip that everyone else uses). That means 8K bytes in sixteen chips. It also means no more 28 amp power supplies.

The system is fully expandable to 65K via an edge connector which carries both the address and data busses, power supplies and all timing signals. All dynamic memory refreshing for both on and off-board memory is done automatically. Also, the Apple Computer can be upgraded to use the 16K chips when they become available. That's 32K bytes on-board RAM in 16 IC's—the equivalent of 256 2102's!

A Little Cassette Board That Works!

Unlike many other cassette boards on the marketplace, ours works every time. It plugs directly into the upright connector on the main board and stands only 2" tall. And since it is very fast (1500 bits per second), you can read or write 4K bytes in about 20 seconds. All timing is done in software, which results in crystal-controlled accuracy and uniformity from unit to unit.

Unlike some other cassette interfaces which require an expensive tape recorder, the Apple Cassette Interface works reliably with almost any audio-grade cassette recorder.
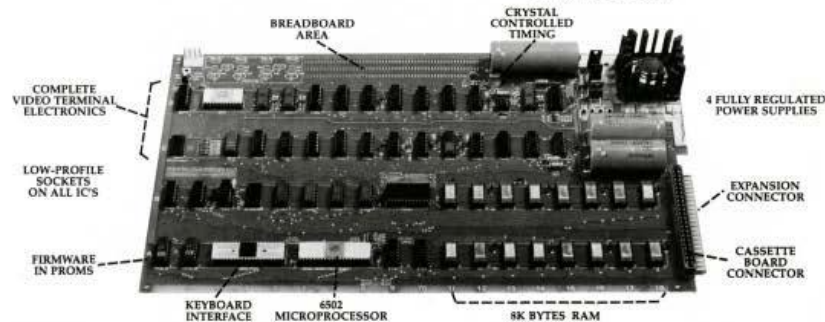
Software:

A tape of APPLE BASIC is included free with the Cassette Interface. Apple Basic features immediate error messages and fast execution, and lets you program in a higher level language immediately and without added cost. Also available now are a dis-assembler and many games, with many software packages, (including a macro assembler) in the works. And since our philosophy is to provide software for our machines free or at minimal cost, you won't be continually paying for access to this growing software library.

The Apple Computer is in stock at almost all major computer stores. (If your local computer store doesn't carry our products, encourage them or write us direct). **Dealer inquiries invited.**

Byte into an Apple $666.66*
* includes 4K bytes RAM

BREADBOARD AREA — CRYSTAL CONTROLLED TIMING — COMPLETE VIDEO TERMINAL ELECTRONICS — 4 FULLY REGULATED POWER SUPPLIES — LOW-PROFILE SOCKETS ON ALL IC'S — EXPANSION CONNECTOR — FIRMWARE IN PROMS — CASSETTE BOARD CONNECTOR — KEYBOARD INTERFACE — 6502 MICROPROCESSOR — 8K BYTES RAM

APPLE Computer Company • 770 Welch Rd., Palo Alto, CA 94304 • (415) 326-4248

OCTOBER 1976   CIRCLE NO. 7 ON INQUIRY CARD   INTERFACE AGE 11

A year later Apple released the Apple II; now there was a case, but you needed to bring your own TV:

Two years later and Apple II had a killer app that would presage the movement of personal computers into the workplace: VisiCalc, the first spreadsheet.

VisiCalc's utility for business was obvious — in fact, it was conceived of by Dan Bricklin while watching a lecture at Harvard Business School. That utility, though, was not about running business critical software like accounting or ERP systems; rather, an employee with an Apple II and VisiCalc could take the initiative to model their business and understand how things worked at a view grounded in a level of calculation that was both too much for one person, yet not sufficient to hire an army of backroom employees, or, increasingly at that point, reserve time on the mainframe.

Notice, though, how this aligned with the Apple and Microsoft philosophy of building tools: tools are meant to be used, but they take volition to maximize their utility. This, I think, is a challenge when it comes to Copilot usage: even before Copilot came out employees with initiative were figuring out how to use other AI tools to do their work more effectively. The idea of Copilot is that you can have an even better AI tool — thanks to the fact it has integrated the information in the "Microsoft Graph" — and make it widely available to your workforce to make that workforce more productive.

To put it another way, the real challenge for Copilot is that it is a change management problem: it's one thing to charge $30/month on a per-seat basis to make an amazing new way to work available to all of your employees; it's another thing entirely — a much more difficult thing — to get all of your employees to change the way they work in order to benefit from your investment, and to make Copilot Pages the "new artifact for the AI age", in line with the spreadsheet in the personal computer age.

Clippy and Copilot

Salesforce CEO Marc Benioff was considerably less generous towards Copilot in last week's Dreamforce keynote. After framing machine learning as "Wave 1" of AI, Benioff said that Copilots were Wave 2, and from Microsoft's perspective it went downhill from there:

> We moved into this Copilot world, but the Copilot world has been kind of a hit-and-miss world. The Copilot world where customers have said to us "Hey, I got these Copilots but they're not exactly performing as we want them to. We don't see how that Copilot world is going to get us to the real vision of artificial intelligence of augmentation of productivity, of better business results that we've been looking for. We just don't see Copilot as that key step for our future." In some ways, they kind of looked at Copilot as the new Microsoft Clippy, and I get that.

The Clippy comparison was mean but not entirely unfair, particularly in the context of users who don't know enough to operate with volition. Former Microsoft executive Steven Sinofsky explained in Hard Core Software:

Why was Microsoft going through all this and making these risky, or even edgy, products? Many seemed puzzled by this at the time. In order to understand that today, one must recognize that using a PC in the early 1990s (and before) was not just difficult, but it was also confusing, frustrating, inscrutable, and by and large entirely inaccessible to most everyone unless you had to learn how to use one for work.

Clippy was to be a replacement for the "Office guru" people consulted when they wanted to do things in Microsoft Office that they knew were possible, but were impossible to discover; Sinofsky admits that a critical error was making Clippy too helpful with simple tasks, like observing "Looks like you're trying to write a letter" when you typed "Dear John" and hit return. Sinofsky reflected:

> The journey of Clippy (in spite of our best efforts that was what the feature came to be called) was one that parallels the PC for me in so many ways. It was not simply a failed feature, or that back-handed compliment of a feature that was simply too early like so many Microsoft features. Rather Clippy represented a final attempt at trying to fix the desktop metaphor for typical or normal people so they could use a computer.

> What everyone came to realize was that the PC was a generational change and that for those growing up with a PC, it was just another arbitrary and random device in life that one just used. As we would learn, kids didn't need different software. They just needed access to a PC. Once they had a PC they would make cooler, faster, and more fun documents with Office than we were. It was kids that loved WordArt and the new graphics in Word and PowerPoint, and they used them easily and more frequently than Boomers or Gen X trying to map typewriters to what a computer could do.

> It was not the complexity that was slowing people down, but the real concern that the wrong thing could undo hours of work. Kids did not have that fear (yet). We needed to worry less about dumbing the software down and more about how more complex things could get done in a way that had far less risk.

This is a critical insight when it comes to AI, Copilot, and the concept of change management: a small subset of Gen Xers and Boomers may have invented the personal computer, but for the rest of their cohort it was something they only used if they had to (resentfully), and only then the narrow set of functionality that was required to do their job. It was the later generations that grew up with the personal computer, and hardly give inserting a table or graphic into a document a second thought (if, in fact, they even know what a "document" is). For a millenial using a personal computer doesn't take volition; it's just a fact of life.

Again, though, computing didn't start with the personal computer, but rather with the replacement of the back office. Or, to put it in rather more dire terms, the initial value in computing wasn't created by helping Boomers do their job more efficiently, but rather by

replacing entire swathes of them completely.

Agents and o1

Benioff implicitly agrees; the Copilot Clippy insult was a preamble to a discussion of agents:

> But it was pushing us, and they were trying to say, what is the next step? And we are now really at that moment. That is why this show is our most important Dreamforce ever. There's no question this is the most exciting Dreamforce and the most important Dreamforce. What you're going to see at this show is technology like you have never seen before…The first time you build and deploy your first autonomous agent for your company that is going to help you to be more productive, to augment your employees, and to get these better business results, you're going to remember that like the first time it was in y our Waymo. This is the 3rd wave of AI. It's agents…

Agents aren't copilots; they are replacements. They do work in place of humans — think call centers and the like, to start — and they have all of the advantages of software: always available, and scalable up-and-down with demand.

> We know that workforces are overwhelmed. They're doing these low-value tasks. They've got kind of a whole different thing post-pandemic. Productivity is at a different place. Capacity is at a different place…we do see that workforces are different, and we realize that 41% of the time seems to be wasted on low value and repetitive tasks, and we want to address that. The customers are expecting more: zero hold times, to be more personal and empathetic, to work with an expert all the time, to instantly schedule things. That's our vision, our dream for these agents…
>
> What if these workforces had no limits at all? Wow. That's kind of a strange thought, but a big one. You start to put all of these things together, and you go, we can kind of build another kind of company. We can build a different kind of technology platforms. We can take the Salesforce technology platform that we already have, and that all of you have invested so much into, the Salesforce Platform, and we can deliver the next capability. The next capability that's going to make our companies more productive. To make our employees more augmented. And just to deliver much better business results. That is what Agentforce is.

This Article isn't about the viability of Agentforce; I'm somewhat skeptical, at least in the short term, for reasons I will get to in a moment. Rather, the key part is the last few sentences: Benioff isn't talking about making *employees* more productive, but rather *companies*; the verb that applies to employees is "augmented", which sounds much nicer than "replaced"; the ultimate goal is stated as well: business results. That right there is tech's third philosophy: improving the bottom line for large enterprises.

Notice how well this framing applies to the mainframe wave of computing: accounting and ERP software made companies more productive and drove positive business results; the employees that were "augmented" were managers who got far more accurate reports much more quickly, while the employees who used to do that work were replaced. Critically, the decision about whether or not make this change did not depend on rank-and-file employees changing how they worked, but for executives to decide to take the plunge.

The Consumerization of IT

When Benioff founded Salesforce in 1999, he came up with a counterintuitive logo:

Of course Salesforce was software; what it was not was SOFTWARE, like that sold by his previous employer Oracle, which at that time meant painful installation and migrations that could take years, and even then would often fail. Salesforce was different: it was a cloud application that you never needed to install or update; you could simply subscribe.

Cloud-based software-as-a-service companies are the norm now, thanks in part to Benioff's vision. And, just as Salesforce started out primarily serving — you guessed it! — sales forces, SaaS applications can focus on individual segments of a company. Indeed, one of the

big trends over the last decade were SaaS applications that grew, at least in the early days, through word-of-mouth and trialing by individuals or team leaders; after all, all you needed to get started was a credit card — and if there was a freemium model, not even that!

This trend was part of a larger one, the so-called "consumerization of IT". Douglas Neal and John Taylor, who first coined the term in 2001, wrote in a 2004 Position Paper:

> Companies must treat users as consumers, encouraging employee responsibility, ownership and trust by providing choice, simplicity and service. The parent/child attitude that many IT departments have traditionally taken toward end users is now obsolete.

This is actually another way of saying what Sinofsky did: enterprise IT customers, i.e. company employees, no longer needed to be taught how to use a computer; they grew up with them, and expected computers to work the same way their consumer devices did. Moreover, the volume of consumer devices meant that innovation would now come from that side of technology, and the best way for enterprises to keep up would be to ideally adopt consumer infrastructure, and barring that, seek to be similarly easy-to-use.

It's possible this is how AI plays out; it is what has happened to date, as large models like those built by OpenAI or Anthropic or Google or Meta are trained on publicly available data, and then are available to be fine-tuned for enterprise-specific use cases. The limitation in this approach, though, is the human one: you need employees who have the volition to use AI with the inherent problems introduced by this approach, including bad data, hallucinations, security concerns, etc. This is manageable as long as a motivated human is in the loop; what seems unlikely to me is any sort of autonomous agent actually operating in a way that makes a company more efficient without an extensive amount of oversight that ends up making the entire endeavor more expensive.

Moreover, in the case of Agentforce specifically, and other agent initiatives more generally, I am unconvinced as to how viable and scalable the infrastructure necessary to manage auto-regressive large language models will end up being. I got into some of the challenges in this Update:

The big challenge for traditional LLMs is that they are path-dependent; while they can consider the puzzle as a whole, as soon as they commit to a particular guess they are locked in, and doomed to failure. This is a fundamental weakness of what are known as "auto-regressive large language models", which to date, is all of them.

To grossly simplify, a large language model generates a token (usually a word, or part of a word) based on all of the tokens that preceded the token being generated; the specific token is the most statistically likely next possible token derived from the model's training (this also gets complicated, as the "temperature" of the output determines what level of randomness goes into choosing from the best possible options; a low temperature chooses the most likely next token, while a higher temperature is more "creative"). The key thing to understand, though, is that this is a serial process: once a token is generated it influences what token is generated next.

The problem with this approach is that it is possible that, in the context of something like a crossword puzzle, the token that is generated is wrong; if that token is wrong, it makes it more likely that the next token is wrong too. And, of course, even if the first token is right, the second token could be wrong anyways, influencing the third token, etc. Ever larger models can reduce the likelihood that a particular token is wrong, but the possibility always exists, which is to say that auto-regressive LLMs inevitably trend towards not just errors but compounding ones.

Note that these problems exist even with specialized prompting like insisting that the LLM "go step-by-step" or "break this problem down into component pieces"; they are still serial output machines that, once they get something wrong, are doomed to deliver an incorrect answer. At the same time, this is also fine for a lot of applications, like writing; where the problem manifests itself is with anything requiring logic or iterative reasoning. In this case, a sufficiently complex crossword puzzle suffices.

That Update was about OpenAI's new o1 model, which I think is a step change in terms of the viability of agents; the example I used in that Update was solving a crossword puzzle, which can't be done in one-shot — but can be done by o1.

o1 is explicitly trained on how to solve problems, and second, o1 is designed to generate multiple problem-solving streams at inference time, choose the best one, and iterate through each step in the process when it realizes it made a mistake. That's why it got the crossword puzzle right — it just took a really long time.

o1 introduces a new vector of potential improvement: while auto-regressive LLMs scaled in quality with training set size (and thus the amount of compute necessary), o1 scales inference. This image is from OpenAI's announcement page:
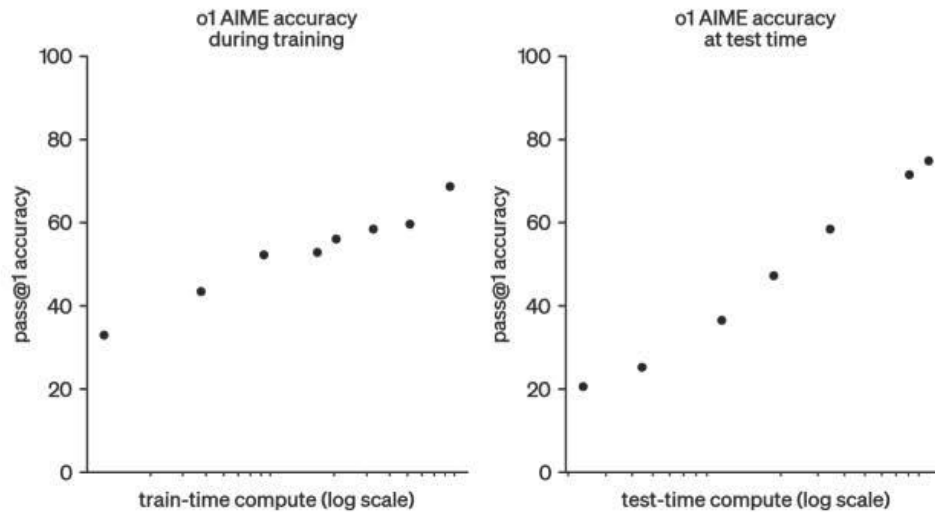
This second image is a potential problem in a Copilot paradigm: sure, a smarter model potentially makes your employees more productive, but those increases in productivity have to be balanced by both greater inference costs and more time spent waiting for the model (o1 is significantly slower than a model like 4o). However, the agent equation, where you are talking about replacing a worker, is dramatically different: there the cost umbrella is absolutely massive, because even the most expensive model is a lot cheaper, above-and-beyond the other benefits like always being available and being scalable in number.

More importantly, scaling compute is exactly what the technology industry is good at. The one common thread from Wave 1 of computing through the PC through SaaS and consumerization of IT, is that problems gated by compute are solved not via premature optimizations but via the progression of processing power. The key challenge is knowing what to scale, and I believe OpenAI has demonstrated the architecture that will benefit from exactly that.

Data and Palantir

That leaves the data piece, and while Benioff bragged about all of the data that Salesforce had, it doesn't have everything, and what it does have is scattered across the phalanx of applications and storage layers that make up the Salesforce Platform. Indeed, Microsoft faces the same problem: while their Copilot vision includes APIs for 3rd-party "agents" — in this case, data from other companies — the reality is that an effective Agent — i.e. a worker replacement — needs access to everything in a way that it can reason over. The ability of large language models to handle unstructured data is revolutionary, but the fact remains that better data still results in better output; explicit step-by-step reasoning data, for example, is a big part of how o1 works.

To that end, the company I am most intrigued by, for what I think will be ==the first wave of AI, is Palantir==. I didn't fully understand the company until <u>this 2023 interview with CTO Shyam Sankar and Head of Global Commercial Ted Mabrey</u>; I suggest reading or listening to the whole thing, but I wanted to call out this exchange in particular:

> **Was there an aha moment where you have this concept — you use this phrase now at the beginning of all your financial reports, which is that you're the operating system for enterprises. Now, obviously this is still the government era, but it's interesting the S-1 uses that line, but it's further down, it's not the lead thing. Was that something that emerged later or was this that, "No, we have to be the interface for everything" idea in place from the beginning?**
>
> **Shyam Sankar**: I think the critical part of it was really realizing that we had built the original product presupposing that our customers had data integrated, that we could focus on the analytics that came subsequent to having your data integrated. I feel like that ==founding trauma was realizing that actually everyone claims that their data is integrated, but it is a complete mess and that actually the much more interesting and valuable part of our business was developing technologies that allowed us to productize data integration==, instead of having it be like a five-year never ending consulting project, so that we could do the thing we actually started our business to do.

That integration looks like this illustration from the company's <u>webpage for Foundry</u>, what they call "The Ontology-Powered Operating System for the Modern Enterprise":
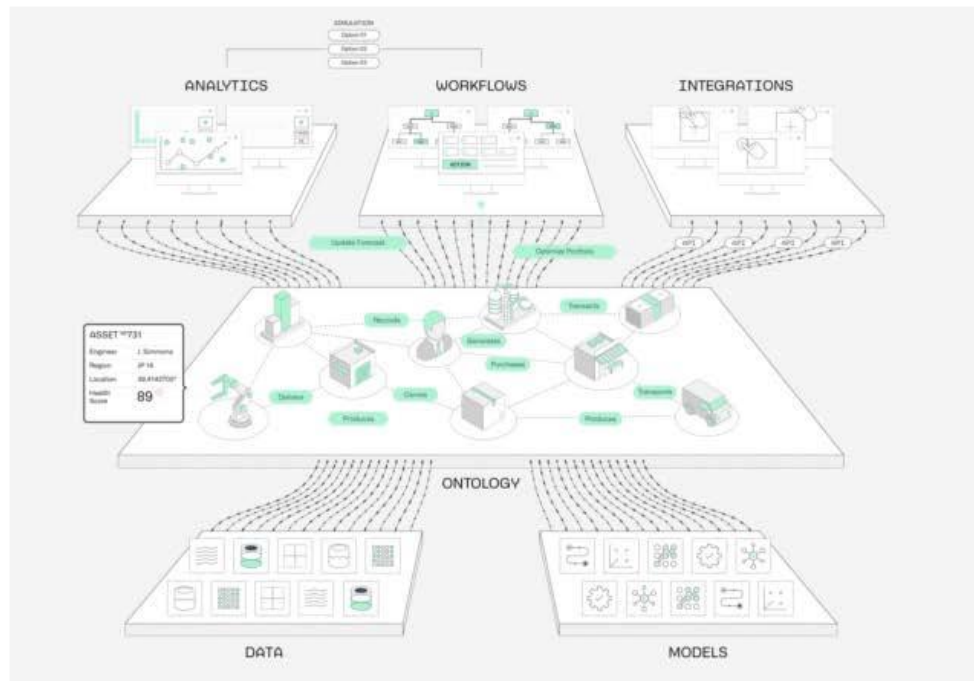
What is notable about this illustration is just how deeply Palantir needs to get into an enterprise's operations to achieve its goals. This isn't a consumery-SaaS application that your team leader puts on their credit card; it is SOFTWARE of the sort that Salesforce sought to move beyond.

If, however, you believe that AI is not just the next step in computing, but rather an entirely new paradigm, then it makes sense that enterprise solutions may be back to the future. We are already seeing that that is the case in terms of user behavior: the relationship of most employees to AI is like the relationship of most corporate employees to PCs in the 1980s; sure, they'll use it if they have to, but they don't want to transform how they work. That will fall on the next generation.

Executives, however, want the benefit of AI now, and I think that benefit will, like the first wave of computing, come from replacing humans, not making them more efficient. And that, by extension, will mean top-down years-long initiatives that are justified by the massive business results that will follow. That also means that go-to-market motions and business models will change: instead of reactive sales from organic growth, successful AI companies will need to go in from the top. And, instead of per-seat licenses, we may end up with something more akin to "seat-replacement" licenses (Salesforce, notably, will charge $2 per call completed by one of its agents). Services and integration teams will also make a

comeback. It's notable that this has been a consistent criticism of Palantir's model, but I think that comes from a viewpoint colored by SaaS; the idea of years-long engagements would be much more familiar to tech executives and investors from forty years ago.

Enterprise Philosophy

Most historically-driven AI analogies usually come from the Internet, and understandably so: that was both an epochal change and also much fresher in our collective memories. My core contention here, however, is that AI truly is a new way of computing, and that means the better analogies are to computing itself. Transformers are the transistor, and mainframes are today's models. The GUI is, arguably, still TBD. *GUI = Graphical user interface (icone, boutons, menu)*

To the extent that is right, then, the biggest opportunity is in top-down enterprise implementations. The enterprise philosophy is older than the two consumer philosophies I wrote about previously: its motivation is not the user, but the buyer, who wants to increase revenue and cut costs, and will be brutally rational about how to achieve that (including running expected value calculations on agents making mistakes). That will be the only way to justify the compute necessary to scale out agentic capabilities, and to do the years of work necessary to get data in a state where humans can be replaced. The bottom line benefits — the essence of enterprise philosophy — will compel just that.

And, by extension, we may be waiting longer than we expect for AI to take over the consumer space, at least at the scale of something like the smartphone or social media. That is good news for the MKBHDs of everything — the users with volition — but for everyone else the biggest payoff will probably be in areas like entertainment and gaming. True consumerization of AI will be left to the next generation who will have never known a world without it.