



What's Next in Computing?

2016-02-21

The computing industry progresses in two mostly independent cycles: financial and product cycles. There has been a lot of handwringing lately about where we are in the financial cycle. Financial markets get a lot of attention. They tend to fluctuate unpredictably and sometimes wildly. The product cycle by comparison gets relatively little attention, even though it is what actually drives the computing industry forward. We can try to understand and predict the product cycle by studying the past and extrapolating into the future.



1981



1994



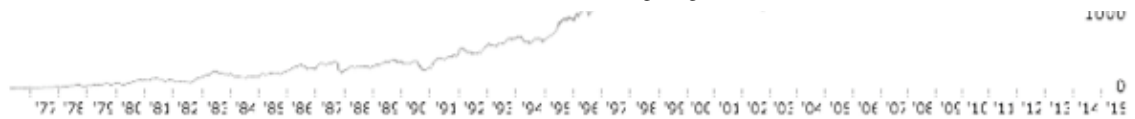
2007

New computing eras have occurred every 10–15 years

Tech product cycles are mutually reinforcing interactions between platforms and applications. New platforms enable new applications, which in turn make the new platforms more valuable, creating a positive feedback loop. Smaller, offshoot tech cycles happen all the time, but every once in a while — historically, about every 10 to 15 years — major new cycles begin that completely reshape the computing landscape.



Financial cycles



Key products and companies



Product cycles

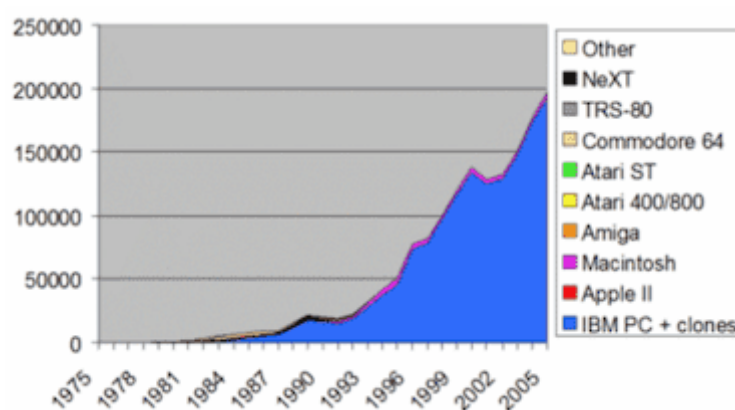


Financial and product cycles evolve mostly independently

The PC enabled entrepreneurs to create word processors, spreadsheets, and many other desktop applications. The internet enabled search engines, e-commerce, e-mail and messaging, social networking, SaaS business applications, and many other services. Smartphones enabled mobile messaging, mobile social networking, and on-demand services like ride sharing. Today, we are in the middle of the mobile era. It is likely that many more mobile innovations are still to come.

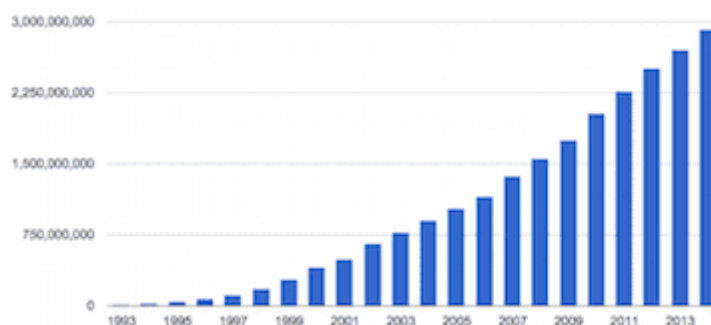
Each product era can be divided into two phases: 1) *the gestation phase*, when the new platform is first introduced but is expensive, incomplete, and/or difficult to use, 2) *the growth phase*, when a new product comes along that solves those problems, kicking off a period of exponential growth.

The Apple II was released in 1977 (and the Altair in 1975), but it was the release of the IBM PC in 1981 that kicked off the PC growth phase.



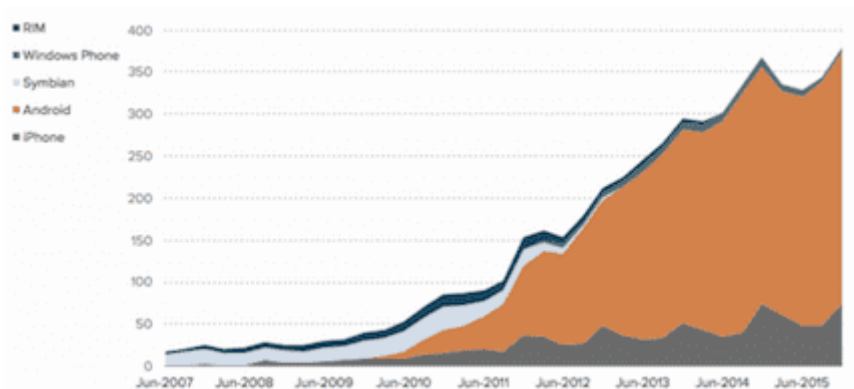
PC sales per year (thousands), source: <http://jeremyreimer.com/m-item.lsp?i=137>

The internet's gestation phase took place in the 80s and early 90s when it was mostly a text-based tool used by academia and government. The release of the Mosaic web browser in 1993 started the growth phase, which has continued ever since.



Worldwide internet users, source: <http://churchm.ag/numbers-internet-use/>

There were feature phones in the 90s and early smartphones like the Sidekick and Blackberry in the early 2000s, but the smartphone growth phase really started in 2007–8 with the release of the iPhone and then Android. Smartphone adoption has since exploded: about 2B people have smartphones today. By 2020, 80% of the global population will have one.



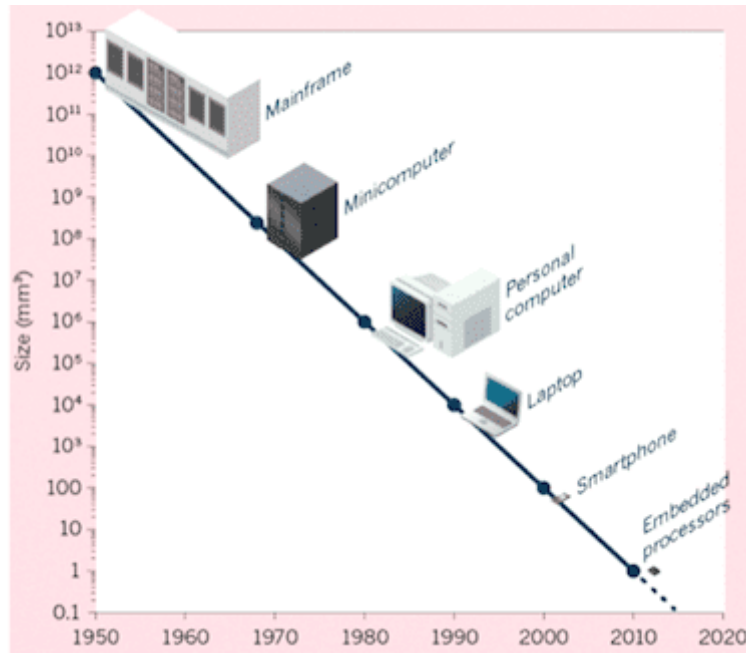
Worldwide smartphone sales per year (millions)

If the 10–15 year pattern repeats itself, the next computing era should enter its growth phase in the next few years. In that scenario, we should already be in the gestation phase. There are a number of important trends in both hardware and software that give us a glimpse into what the next era of computing might be. Here I talk about those trends and then make some suggestions about what the future might look like.

Hardware: small, cheap, and ubiquitous

In the mainframe era, only large organizations could afford a computer.

Minicomputers were affordable by smaller organization, PCs by homes and offices, and smartphones by individuals.



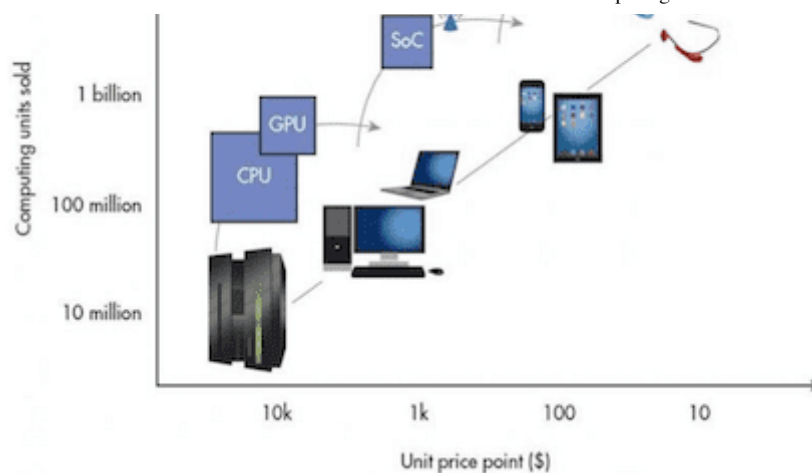
Computers are getting steadily smaller, source: <http://www.nature.com/news/the-chips-are-down-for-moores-law-1.19338>

We are now entering an era in which processors and sensors are getting so small and cheap that there will be many more computers than there are people.

There are two reasons for this. One is the steady progress of the semiconductor industry over the past 50 years (Moore's law). The second is what Chris Anderson calls “the peace dividend of the smartphone war”: the runaway success of smartphones led to massive investments in processors and sensors. If you disassemble a modern drone, VR headset, or IoT devices, you'll find mostly smartphone components.

In the modern semiconductor era, the focus has shifted from standalone CPUs to bundles of specialized chips known as systems-on-a-chip.





Computer prices have been steadily dropping, source: <https://medium.com/@magicsilicon/computing-transitions-22c07b9c457a#.j4cm9m6qu%5C>

Typical systems-on-a-chip bundle energy-efficient ARM CPUs plus specialized chips for graphics processing, communications, power management, video processing, and more.

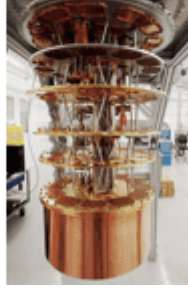


Raspberry Pi Zero: 1 GHz Linux computer for \$5

This new architecture has dropped the price of basic computing systems from about \$100 to about \$10. The Raspberry Pi Zero is a 1 GHz Linux computer that you can buy for \$5. For a similar price you can buy a wifi-enabled microcontroller that runs a version of Python. Soon these chips will cost less than a dollar. It will be cost-effective to embed a computer in almost anything.

Meanwhile, there are still impressive performance improvements happening in high-end processors. Of particular importance are GPUs (graphics processors), the best of which are made by Nvidia. GPUs are useful not only for traditional graphics processing, but also for machine learning algorithms and virtual/augmented reality

devices. Nvidia's roadmap promises significant performance improvements in the coming years.



Google's quantum computer, source: <https://www.technologyreview.com/s/544421/googles-quantum-dream-machine/>

A wildcard technology is quantum computing, which today exists mostly in laboratories but if made commercially viable could lead to orders-of-magnitude performance improvements for certain classes of algorithms in fields like biology and artificial intelligence.

Software: the golden age of AI

There are many exciting things happening in software today. Distributed systems is one good example. As the number of devices has grown exponentially, it has become increasingly important to 1) parallelize tasks across multiple machines 2) communicate and coordinate among devices. Interesting distributed systems technologies include systems like Hadoop and Spark for parallelizing big data problems, and Bitcoin/blockchain for securing data and assets.

But perhaps the most exciting software breakthroughs are happening in artificial intelligence (AI). AI has a long history of hype and disappointment. Alan Turing himself predicted that machines would be able to successfully imitate humans by the year 2000. However, there are good reasons to think that AI might now finally be entering a golden age.

“Machine learning is a core, transformative way by which we’re rethinking everything we’re doing.” — Google CEO, Sundar Pichai

A lot of the excitement in AI has focused on deep learning, a machine learning technique that was popularized by a now famous 2012 Google project that used a giant cluster of computers to learn to identify cats in YouTube videos. Deep learning is a descendent of neural networks, a technology that dates back to the 1940s. It was brought back to life by a combination of factors, including new algorithms, cheap parallel computation, and the widespread availability of large data sets.

ImageNet challenge error rates, source: <http://www.slideshare.net/nervanasys/sd-meetup-12215> (red line = human performance)

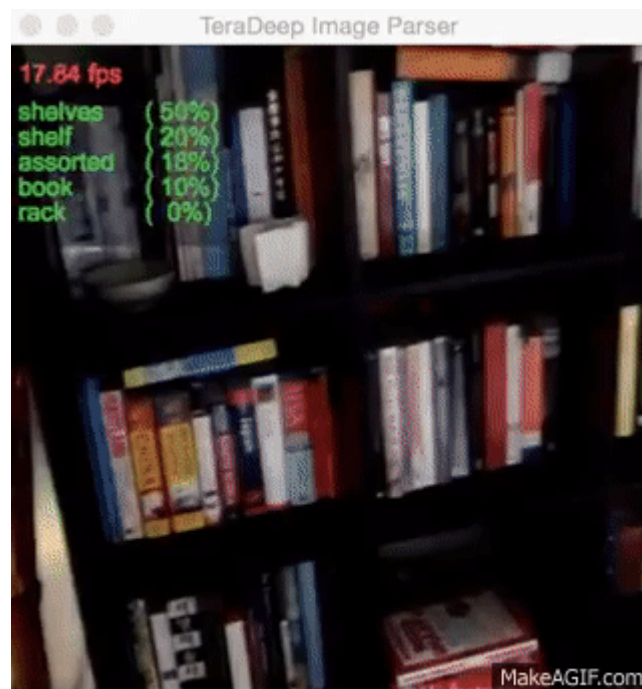
It's tempting to dismiss deep learning as another Silicon Valley buzzword. The excitement, however, is supported by impressive theoretical and real-world results. For example, the error rates for the winners of the ImageNet challenge — a popular machine vision contest — were in the 20–30% range prior to the use of deep learning. Using deep learning, the accuracy of the winning algorithms has steadily improved, and in 2015 surpassed human performance.

Many of the papers, data sets, and software tools related to deep learning have been open sourced. This has had a democratizing effect, allowing individuals and small organizations to build powerful applications. WhatsApp was able to build a global messaging system that served 900M users with just 50 engineers, compared to the thousands of engineers that were needed for prior generations of messaging systems. This “WhatsApp effect” is now happening in AI. Software tools like Theano and TensorFlow, combined with cloud data centers for training, and inexpensive GPUs for deployment, allow small teams of engineers to build state-of-the-art AI systems.

For example, here a solo programmer working on a side project used TensorFlow to colorize black-and-white photos:

Left: black and white. Middle: automatically colorized. Right: true color. source: <http://tinyclouds.org/colorize/>

And here a small startup created a real-time object classifier:



Which of course is reminiscent of a famous scene from a sci-fi movie:



One of the first applications of deep learning released by a big tech company is the search function in Google Photos, which is shockingly smart.



We'll soon see significant upgrades to the intelligence of all sorts of products, including: voice assistants, search engines, chat bots, 3D scanners, language translators, automobiles, drones, medical imaging systems, and much more.

The business plans of the next 10,000 startups are easy to forecast: Take X and add AI. This is a big deal, and now it's here. — Kevin Kelly.

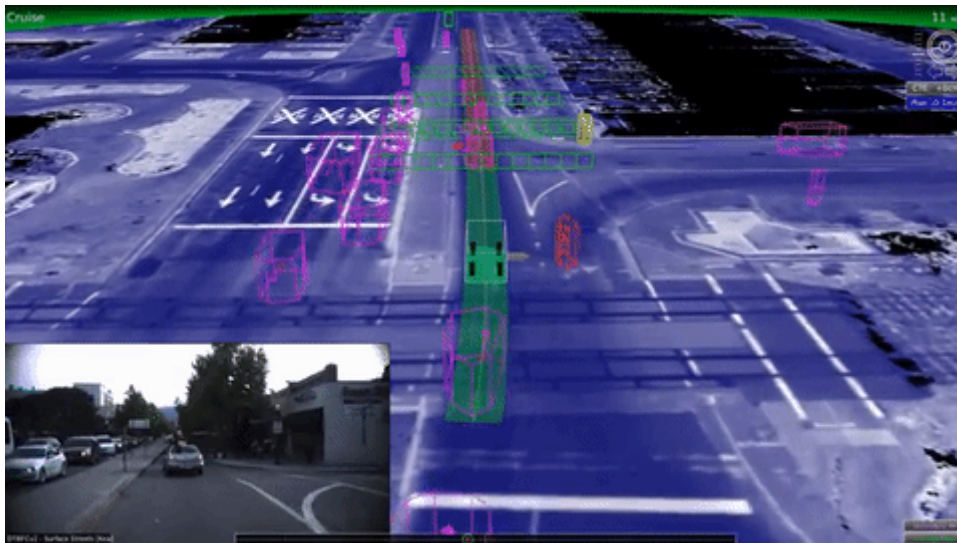
Startups building AI products will need to stay laser focused on specific applications to compete against the big tech companies who have made AI a top priority. AI systems get better as more data is collected, which means it's possible to create a virtuous flywheel of data network effects (more users → more data → better products → more users). The mapping startup Waze used data network effects to produce better maps

than its vastly better capitalized competitors. Successful AI startups will follow a similar strategy.

Software + hardware: the new computers

There are a variety of new computing platforms currently in the gestation phase that will soon get much better — and possibly enter the growth phase — as they incorporate recent advances in hardware and software. Although they are designed and packaged very differently, they share a common theme: they give us new and augmented abilities by embedding a smart virtualization layer on top of the world. Here is a brief overview of some of the new platforms:

Cars. Big tech companies like Google, Apple, Uber, and Tesla are investing significant resources in autonomous cars. Semi-autonomous cars like the Tesla Model S are already publicly available and will improve quickly. Full autonomy will take longer but is probably not more than 5 years away. There already exist fully autonomous cars that are almost as good as human drivers. However, for cultural and regulatory reasons, fully autonomous cars will likely need to be significantly better than human drivers before they are widely permitted.



Expect to see a lot more investment in autonomous cars. In addition to the big tech companies, the big auto makers are starting to take autonomy very seriously. You'll even see some interesting products made by startups. Deep learning software tools have gotten so good that a solo programmer was able to make a semi-autonomous car:



Drones. Today's consumer drones contain modern hardware (mostly smartphone components plus mechanical parts), but relatively simple software. In the near future, we'll see drones that incorporate advanced computer vision and other AI to make them safer, easier to pilot, and more useful. Recreational videography will continue to be popular, but there will also be important commercial use cases. There are tens of millions of dangerous jobs that involve climbing buildings, towers, and other structures that can be performed much more safely and effectively using drones.



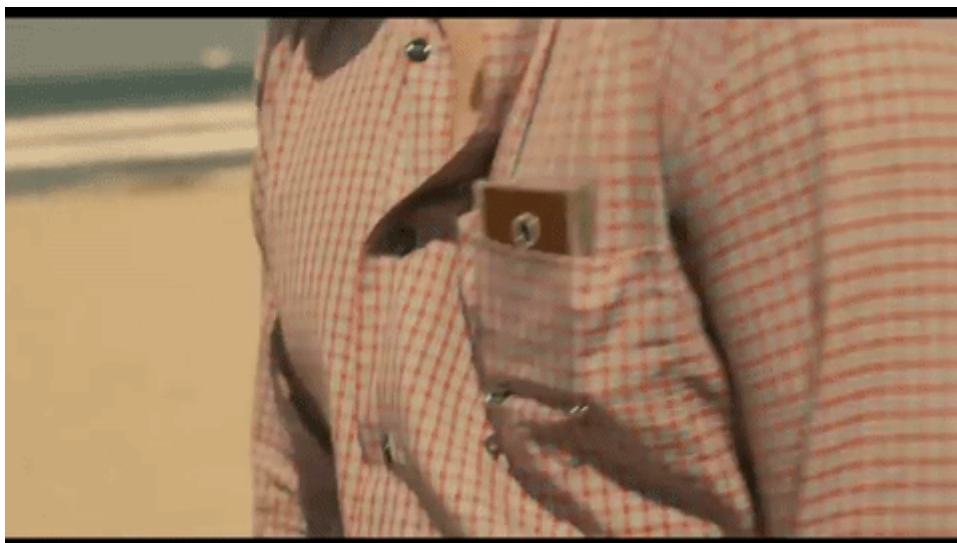
Internet of Things. The obvious use cases for IoT devices are energy savings, security, and convenience. Nest and Dropcam are popular examples of the first two categories. One of the most interesting products in the convenience category is Amazon's Echo.

Three main uses cases for IoT

Most people think Echo is a gimmick until they try it and then they are surprised at how useful it is. It's a great demo of how effective always-on voice can be as a user interface. It will be a while before we have bots with generalized intelligence that can carry on full conversations. But, as Echo shows, voice can succeed today in constrained contexts. Language understanding should improve quickly as recent breakthroughs in deep learning make their way into production devices.

IoT will also be adopted in business contexts. For example, devices with sensors and network connections are extremely useful for monitoring industrial equipment.

Wearables. Today's wearable computers are constrained along multiple dimensions, including battery, communications, and processing. The ones that have succeeded have focused on narrow applications like fitness monitoring. As hardware components continue to improve, wearables will support rich applications the way smartphones do, unlocking a wide range of new applications. As with IoT, voice will probably be the main user interface.



Virtual Reality. 2016 is an exciting year for VR: the launch of the Oculus Rift and HTC/Valve Vive (and, possibly, the Sony Playstation VR), means that comfortable and immersive VR systems will finally be publicly available. VR systems need to be really good to avoid the “uncanny valley” trap. Proper VR requires special screens (high resolution, high refresh rate, low persistence), powerful graphics cards, and the ability to track the precise position of the user (previously released VR systems could only track the rotation of the user’s head). This year, the public will for the first time get to experience what is known as “presence” — when your senses are sufficiently tricked that you feel fully transported into the virtual world.



VR headsets will continue to improve and get more affordable. Major areas of research will include: 1) new tools for creating rendered and/or filmed VR content, 2) machine vision for tracking and scanning directly from phones and headsets, and 3) distributed back-end systems for hosting large virtual environments.



Augmented Reality. AR will likely arrive after VR because AR requires most of what VR requires plus additional new technologies. For example, AR requires advanced, low-latency machine vision in order to convincingly combine real and virtual objects in the same interactive scene.

Real and virtual combined (from *The Kingsmen*)

That said, AR is probably coming sooner than you think. This demo video was shot directly through Magic Leap's AR device:



What's next?

It is possible that the pattern of 10–15 year computing cycles has ended and mobile is the final era. It is also possible the next era won't arrive for a while, or that only a subset of the new computing categories discussed above will end up being important.

I tend to think we are on the cusp of not one but multiple new eras. The “peace dividend of the smartphone war” created a Cambrian explosion of new devices, and developments in software, especially AI, will make those devices smart and useful. Many of the futuristic technologies discussed above exist today, and will be broadly accessible in the near future.

Observers have noted that many of these new devices are in their “awkward adolescence.” That is because they are in their gestation phase. Like PCs in the 70s, the internet in the 80s, and smartphones in the early 2000s, we are seeing pieces of a future that isn't quite here. But the future is coming: markets go up and down, and excitement ebbs and flows, but computing technology marches steadily forward.

[Virtual Reality](#)[Tech Adoption](#)[Self Driving Cars](#)[Computers](#)[Hardware](#)

Next post: [The Internet Economy](#)

Previous post: [If an alien film crew chose to feature our species in a nature documentary, they'd have plenty of...](#)

VIEWS EXPRESSED IN “CONTENT” (INCLUDING POSTS, PODCASTS, VIDEOS) LINKED ON THIS WEBSITE OR POSTED IN SOCIAL MEDIA AND OTHER PLATFORMS (COLLECTIVELY, “CONTENT DISTRIBUTION OUTLETS”) ARE MY OWN AND ARE NOT THE VIEWS OF AH CAPITAL MANAGEMENT, L.L.C. (“A16Z”) OR ITS RESPECTIVE AFFILIATES. AH CAPITAL MANAGEMENT IS AN INVESTMENT ADVISER REGISTERED WITH THE SECURITIES AND EXCHANGE COMMISSION. REGISTRATION AS AN INVESTMENT ADVISER DOES NOT IMPLY ANY SPECIAL SKILL OR TRAINING. THE POSTS ARE NOT DIRECTED TO ANY INVESTORS OR POTENTIAL INVESTORS, AND DO NOT CONSTITUTE AN OFFER TO SELL -- OR A SOLICITATION OF AN OFFER TO BUY -- ANY SECURITIES, AND MAY NOT BE USED OR RELIED UPON IN EVALUATING THE MERITS OF ANY INVESTMENT.

THE CONTENT SHOULD NOT BE CONSTRUED AS OR RELIED UPON IN ANY MANNER AS INVESTMENT, LEGAL, TAX, OR OTHER ADVICE. YOU SHOULD CONSULT YOUR OWN ADVISERS AS TO LEGAL, BUSINESS, TAX, AND OTHER RELATED MATTERS CONCERNING ANY INVESTMENT. ANY PROJECTIONS, ESTIMATES, FORECASTS, TARGETS, PROSPECTS AND/OR OPINIONS EXPRESSED IN THESE MATERIALS ARE SUBJECT TO CHANGE WITHOUT NOTICE AND MAY DIFFER OR BE CONTRARY TO OPINIONS EXPRESSED BY OTHERS. ANY CHARTS PROVIDED HERE ARE FOR INFORMATIONAL PURPOSES ONLY, AND SHOULD NOT BE RELIED UPON WHEN MAKING ANY INVESTMENT DECISION. CERTAIN INFORMATION CONTAINED IN HERE HAS BEEN OBTAINED FROM THIRD-PARTY SOURCES. WHILE TAKEN FROM SOURCES BELIEVED TO BE RELIABLE, I HAVE NOT INDEPENDENTLY VERIFIED SUCH

INFORMATION AND MAKES NO REPRESENTATIONS ABOUT THE ENDURING ACCURACY OF THE INFORMATION OR ITS APPROPRIATENESS FOR A GIVEN SITUATION. THE CONTENT SPEAKS ONLY AS OF THE DATE INDICATED.

UNDER NO CIRCUMSTANCES SHOULD ANY POSTS OR OTHER INFORMATION PROVIDED ON THIS WEBSITE -- OR ON ASSOCIATED CONTENT DISTRIBUTION OUTLETS -- BE CONSTRUED AS AN OFFER SOLICITING THE PURCHASE OR SALE OF ANY SECURITY OR INTEREST IN ANY POOLED INVESTMENT VEHICLE SPONSORED, DISCUSSED, OR MENTIONED BY A16Z PERSONNEL. NOR SHOULD IT BE CONSTRUED AS AN OFFER TO PROVIDE INVESTMENT ADVISORY SERVICES; AN OFFER TO INVEST IN AN A16Z-MANAGED POOLED INVESTMENT VEHICLE WILL BE MADE SEPARATELY AND ONLY BY MEANS OF THE CONFIDENTIAL OFFERING DOCUMENTS OF THE SPECIFIC POOLED INVESTMENT VEHICLES -- WHICH SHOULD BE READ IN THEIR ENTIRETY, AND ONLY TO THOSE WHO, AMONG OTHER REQUIREMENTS, MEET CERTAIN QUALIFICATIONS UNDER FEDERAL SECURITIES LAWS. SUCH INVESTORS, DEFINED AS ACCREDITED INVESTORS AND QUALIFIED PURCHASERS, ARE GENERALLY DEEMED CAPABLE OF EVALUATING THE MERITS AND RISKS OF PROSPECTIVE INVESTMENTS AND FINANCIAL MATTERS. THERE CAN BE NO ASSURANCES THAT A16Z'S INVESTMENT OBJECTIVES WILL BE ACHIEVED OR INVESTMENT STRATEGIES WILL BE SUCCESSFUL. ANY INVESTMENT IN A VEHICLE MANAGED BY A16Z INVOLVES A HIGH DEGREE OF RISK INCLUDING THE RISK THAT THE ENTIRE AMOUNT INVESTED IS LOST. ANY INVESTMENTS OR PORTFOLIO COMPANIES MENTIONED, REFERRED TO, OR DESCRIBED ARE NOT REPRESENTATIVE OF ALL INVESTMENTS IN VEHICLES MANAGED BY A16Z AND THERE CAN BE NO ASSURANCE THAT THE INVESTMENTS WILL BE PROFITABLE OR THAT OTHER INVESTMENTS MADE IN THE FUTURE WILL HAVE SIMILAR CHARACTERISTICS OR RESULTS. A LIST OF INVESTMENTS MADE BY FUNDS MANAGED BY A16Z IS AVAILABLE AT [HTTPS://A16Z.COM/INVESTMENTS/](https://a16z.com/investments/). EXCLUDED FROM THIS LIST ARE INVESTMENTS FOR WHICH THE ISSUER HAS NOT PROVIDED PERMISSION FOR A16Z TO DISCLOSE PUBLICLY AS WELL AS UNANNOUNCED INVESTMENTS IN PUBLICLY TRADED DIGITAL ASSETS. PAST RESULTS OF ANDREESSEN HOROWITZ'S INVESTMENTS, POOLED INVESTMENT VEHICLES, OR INVESTMENT STRATEGIES ARE NOT NECESSARILY INDICATIVE OF FUTURE RESULTS. PLEASE SEE [HTTPS://A16Z.COM/DISCLOSURES](https://a16z.com/disclosures) FOR ADDITIONAL IMPORTANT INFORMATION.