Peter Thiel's CS183: Startup - Class 17 - Deep Thought

He is an essay version of class notes from Class 17 of CS183: Startup. Errors and omissions are mine.

Three guests joined the class for a conversation after Peter's remarks:

1. D. Scott Brown, co-founder of Vicarious
2. Eric Jonas, CEO of Prior Knowledge
3. Bob McGrew, Director of Engineering at Palantir

Credit for good stuff goes to them and Peter. I have tried to be accurate. But note that this is not a transcript of the conversation.

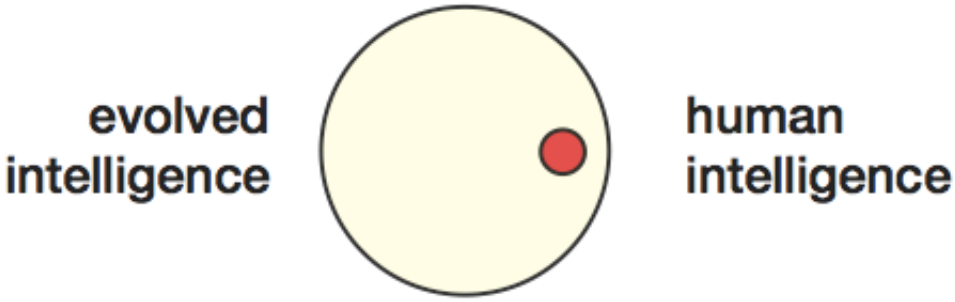
Class 17 Notes Essay—Deep Thought

I. The Hugeness of AI

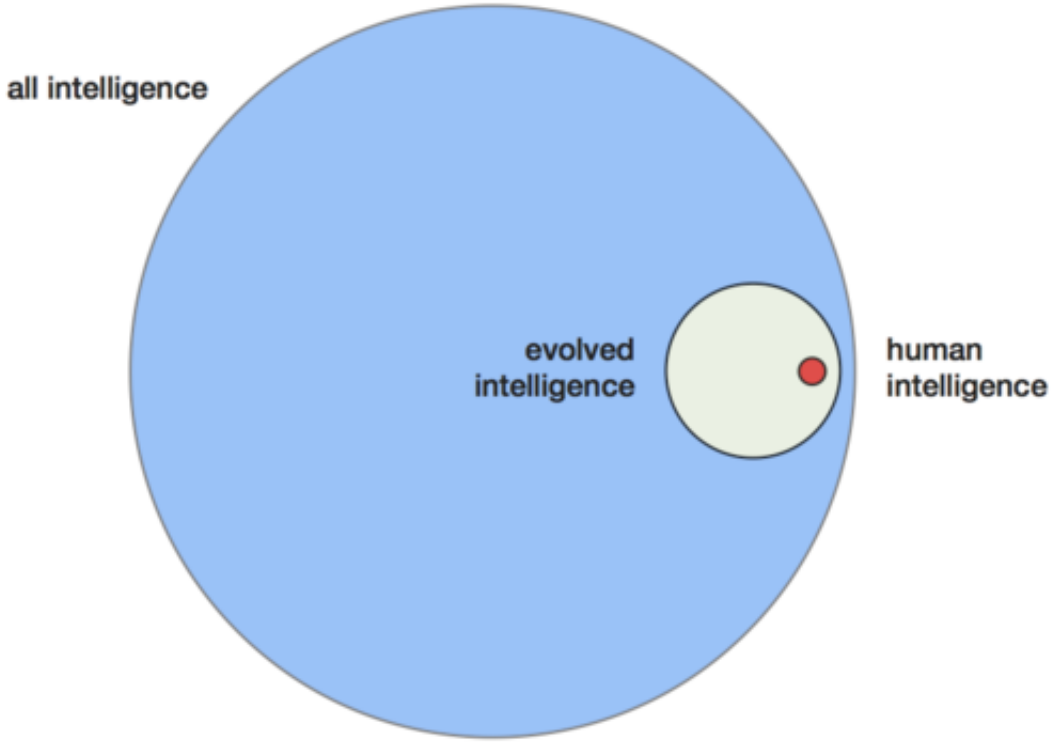
On the surface, we tend to think of people as a very diverse set. People have a wide range of different abilities, interests, characteristics, and intelligence. Some people are good, while others are bad. It really varies.

By contrast, we tend to view computers as being very alike. All computers are more or less the same black box. One way of thinking about the range of possible artificial intelligences is to reverse this standard framework. Arguably it should be the other way around; there is a much *larger* range of potential AI than there is a range of different people.

There are many ways that intelligence can be described and organized. Not all involve human intelligence. Even accounting for the vast diversity among all different people, human intelligence is probably only a tiny dot relative to all evolved forms of intelligence; imagine all the aliens in all planets of the universe that might or could exist.



But AI has much larger range than all naturally possible things. AI is not limited to evolution; it can involve things that are built. Evolution produces birds and flight. But evolution cannot produce supersonic birds with titanium wings. The straightforward process of natural selection involves gradual iteration in ecosystems. AI is not similarly limited. The range of potential AI is thus much larger than the range of alien intelligence, which in turn is broader than the range of human intelligence.



So AI is a very large space—so large that people's normal intuitions about its size are often off base by orders of magnitude.

One of the big questions in AI is exactly how smart it can possibly get. Imagine an intelligence spectrum with 3 data points: a mouse, a moron, and Einstein. Where would AI fall on that scale?

We tend to think of AI as being marginally smarter than an Einstein. But it is not a priori clear why the scale can't actually go up much, much higher than that. The bias is toward conceiving of things that are fathomable. But why is that more realistic than a superhuman intelligence so smart that it's hard to fathom? It might be easier for a mouse to understand the relativity than it is for us to actually understand how an AI supercomputer thinks.

A future with artificial intelligence would be so unrecognizable that it would unlike any other future. A biotech future would involve people functioning better, but still in recognizably human way. A retrofuture would involve things that have been tried before and resurrected. But AI has the possibility of being radically different and radically strange.

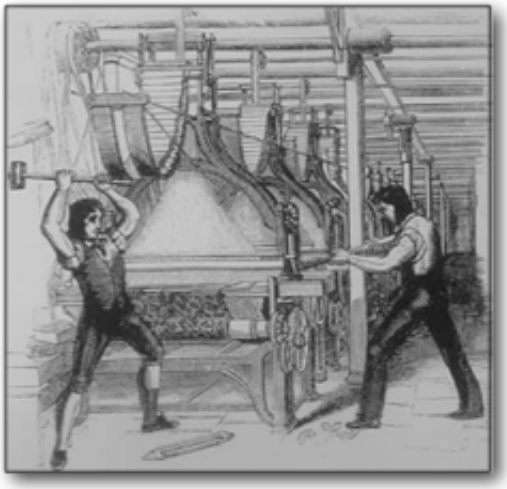
There is a weird set of theological parallels you could map out. God may have been to the Middle Ages what AI will become to us. Will the AI *be* god? Will it be all-powerful? Will it love us? These seem like incomprehensible questions. But they may still be worth asking.

II. The Strangeness of AI

The Turing test is the classic, decades-old test for AI that asks whether you can build a machine that behaves as intelligently as a human does. It focuses on the subset of human behavior that is intelligent. Recently the popular concern has shifted from intelligent computers to empathetic computers. People today seem more interested in whether computers can understand our feelings than whether they are actually smart. It doesn't matter how intelligent it is in more classic domains; if the computer does not find human eye movement emotionally provocative, it is, like Vulcans, still somehow inferior to people.

The history of technology is largely a history of technology displacing people. The plow, the printing press, the cotton gin all put people out of business. Machines were developed to do things more efficiently. But while displacing people is bad, there's the countervailing sense that these machines are good. The fundamental question is whether AI actually replaces people or not. The effect of displacement is the strange, almost political question that seems inextricably linked with the future of AI.

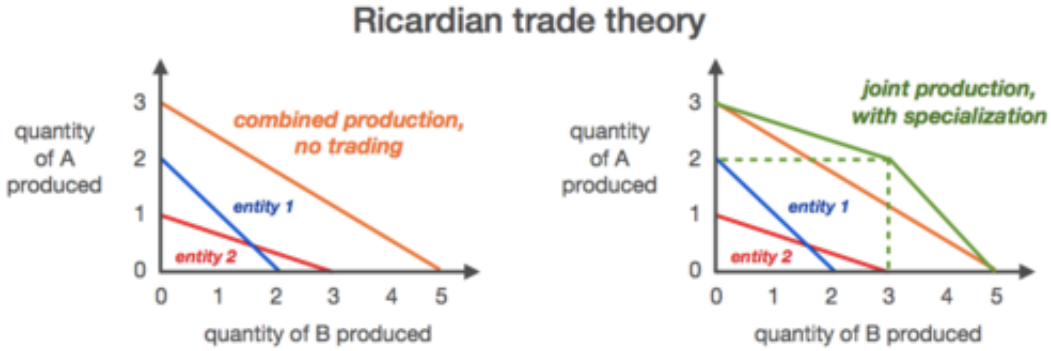
There are two basic paradigms. The Luddite paradigm is that machines are bad, and you should destroy them before they destroy you. This looks something like textile workers destroying factory cotton mills, lest the machines take over the cotton processing. The Ricardo paradigm, by contrast, holds that technology is fundamentally good. This is economist David Ricardo's gains from trade insight; while technology displaces people, it also frees them up to do more.



Luddites

David Ricardo

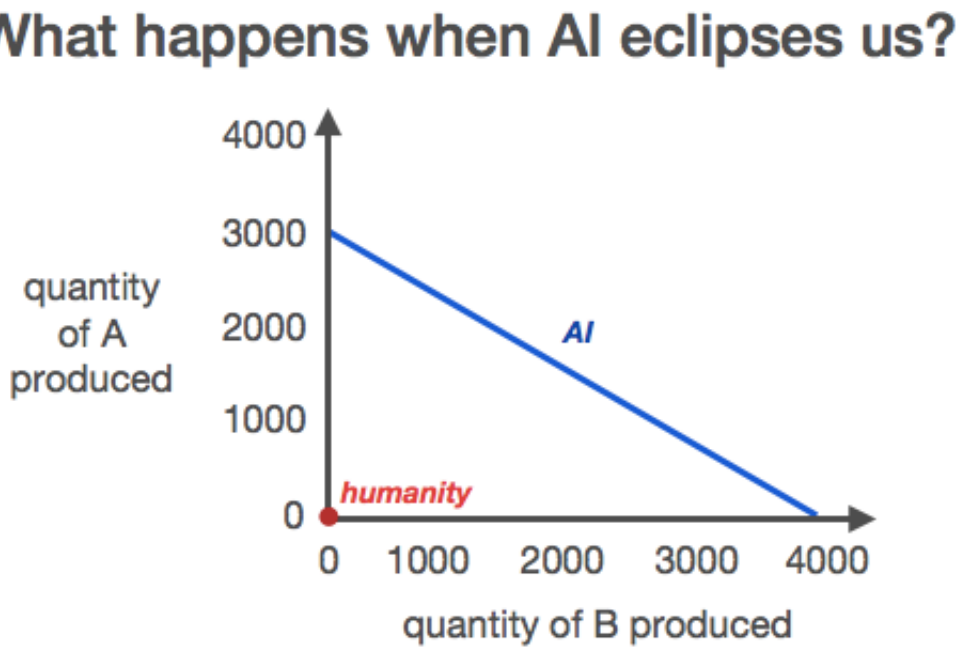
Ricardian trade theory would say that if China can make cheaper cars than can be made in the U.S., it is good for us to buy cars from China. Yes, some people in Detroit lose their jobs. But they can be retrained. And local disturbances notwithstanding, total value can be maximized.



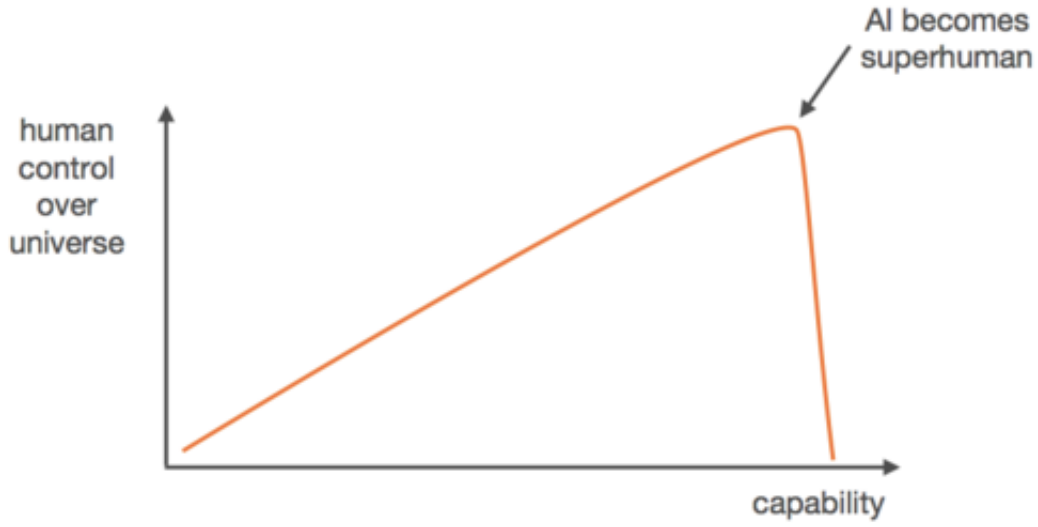
The charts above illustrate the basic theory. With no trade, you get less production. With joint production and specialization, you expand the frontier. More value is created. This trade framework is one way to think about technology. Some cotton artisans lose their jobs. But the price of shirts from the cotton factory falls quite a bit. So the artisans who find other jobs are now doing something more efficient and can afford more clothes at the same time.

The question is whether AI ends up being just another version of something you trade with. That would be straight Ricardo. There's a natural division of labor. Humans are good at some things. Computers are good at other things. Since they are each quite different from each other, the expected gains from trade are large. So they trade and realize those gains. In this scenario AI is not substitute for humans, but rather a compliment to them.

But this depends on the relative magnitudes of advantage. The above scenario plays out if the AI is marginally better. But things may be different if the AI is in fact dramatically better. What if it can do 3000x what humans can do across everything? Would it even make sense for the AI to trade with us at all? Humans, after all, don't trade with monkeys or mice. So even though the Ricardo theory is sound economic intuition, in extreme cases there may be something to be said for the Luddite perspective.



This can be reframed as a battle over control. How much control do humans have over the universe? As AI becomes stronger, we get more and more control. But then AI hits an inflection point where it goes superhuman, and we lose control altogether. That is qualitatively different from most technology, which gives people more control over the world with no end. There is no cliff with most technology. So while computers can give us a great deal of control, and help us overcome chance and uncertainty, it may be possible to go too far. We may end up creating a supercomputer in the cloud that calls itself Zeus and throws down lightning bolts at people.



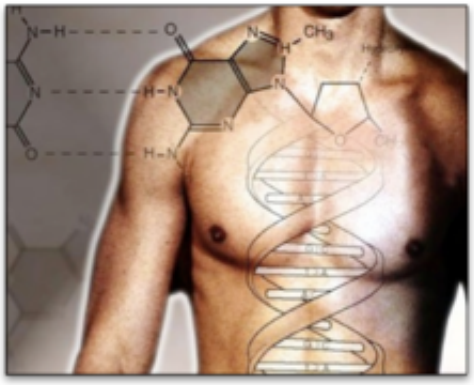
III. The Opportunity of AI

Hugeness and strangeness are interesting questions. But whether and how one can make money with AI may be even more interesting. So how big is the AI opportunity?

A. Is It Too Early for AI?

Everything we've talked about in class remains important. The timing question is particularly important here. It might still be too early for AI. There's a reasonable case to be made there. We know that futures fail quite often. Supersonic airplanes of the '70s failed; they were too noisy and people complained. Handheld iPad-like devices from the '90s and smart phones from '99 failed. Siri is probably still a bit too early today. So whether the timing is right for AI is very hard to know ex ante.

But we can try to make the case for AI by comparing it to things like biotech. If you had a choice between doing AI and the biotech 2.0 stuff we covered last class, the conventional view would be that the biotech angle is the right one to pick. Arguably the bioinformatics revolution is being or will soon be applied to humans, whereas actual application of AI is much future out. But the conventional view isn't always right.



biotechnology

artificial intelligence

B. Unanimity and Skepticism

Last week in Santa Clara there was an event called "5 Top VCs, 10 Tech Trends." Each VC on the panel made 2 predictions about technology in the next 5 years. The audience voted on whether they agreed with each prediction. One of my predictions was that biology would become an information science. When the audience voted, it was a sea of green. 100% agreed with that prediction. There wasn't a single dissenter. Perhaps that should make us nervous. Unanimity in crowds can be very disconcerting. Maybe it's worth questioning the biotech-as-info-science thesis a little bit more.

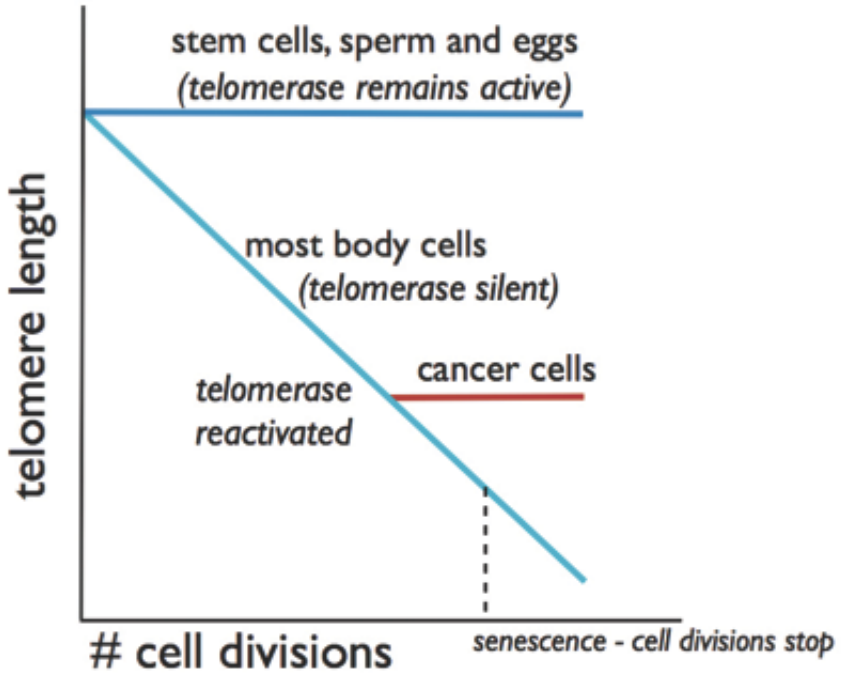
The single idea that people thought was the worst was that all cars would go electric. 92% of the audience voted against that happening. There are many reasons to be bearish on electric cars. But now there is one less.

The closest thing to AI that was discussed was whether Moore's law would continue to accelerate. The audience was split 50-50 on that. If it can accelerate—if it can more than double every 18th months going forward—it would seem like you'd get something like AI in just a few years. Yet most people thought AI was much further away than biotech 2.0.

C. (Hidden) Limits

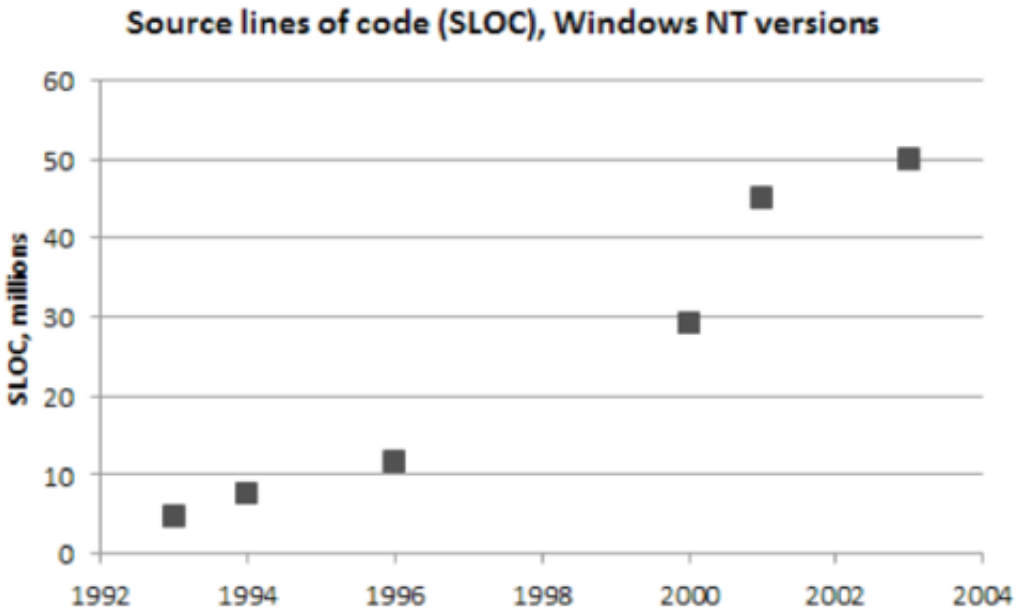
One way to compare biotech and AI is to think about whether there are serious—and maybe even hidden—limits in each one. The biotech revolution narrative is that we're going to figure out how to reverse and

cure all sorts of maladies, so if you just live to *x*, you can stay alive forever. It's a good narrative. But it's also plausible that there are invisible barriers lurking beneath the surface. It's possible, for example, that various systems in the human body act against one another to reach equilibrium. Telomerase helps cells split unbounded. This is important because you stop growing and start to age when cells don't split. So one line of thinking is that you should drink red wine and do whatever else you can to keep telomerase going.

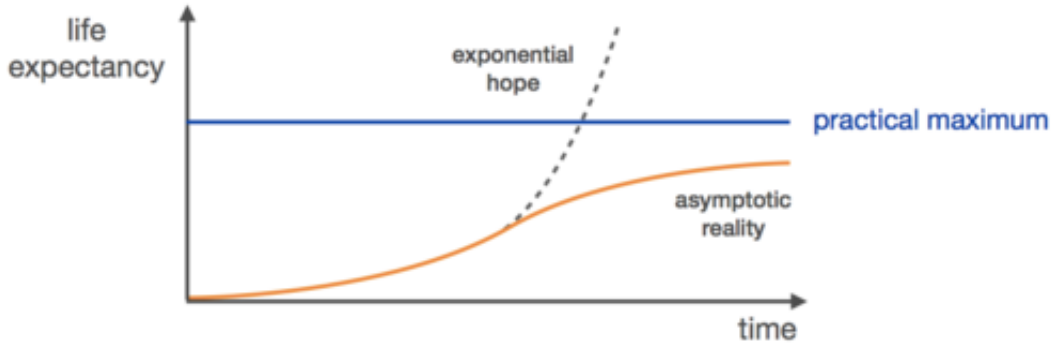


The challenge is that unbounded cell splitting starts to look a lot like cancer at some point. So it's possible that aging and cancer have the effect of cancelling each other out. If people didn't age, they would just die of cancer. But if you shut down telomerase sooner, you just age faster. Fix one problem and you create another. It's not clear what the right balance is, whether such barriers can be overcome, or, really, whether these barriers even exist.

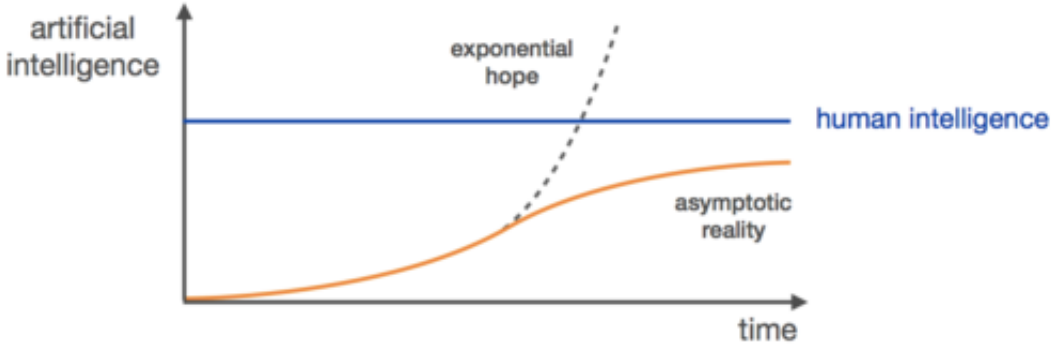
A leading candidate for an invisible barrier in AI is the complexity of the code. The might be some limit where the software becomes too complicated as you produce more and more lines of code. Past a certain point, there is so much to keep track of that no one knows what's going on. Debugging becomes difficult or impossible. Something like this could be said to have happened to Microsoft Windows over a number of decades. It used to be elegant. Maybe it has been or can be improved a bit. But maybe there are serious hidden limits too it. In theory, you add more lines of code to make things better. But maybe they will just make things worse.



The fundamental tension is exponential hope versus asymptotic reality. The optimistic view is the exponential case. We can argue for that, but it's sort of unknown. The question is whether and when asymptotic reality sets in.



and the AI version:



D. AI Pulls Ahead

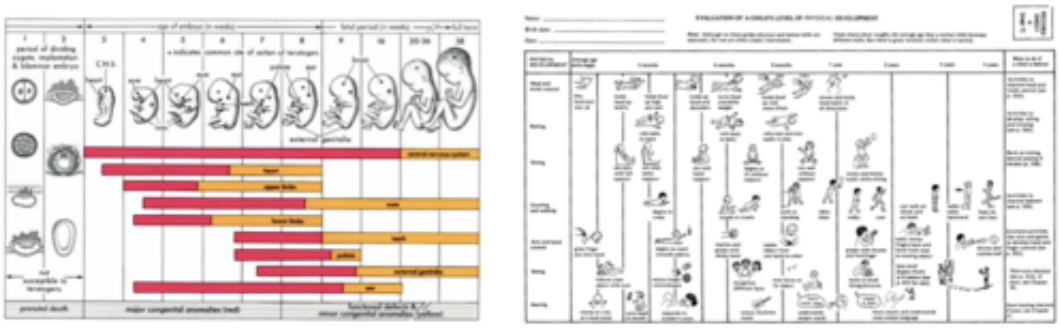
There are many parallels between doing new things in biotech and AI. But there are three distinct

advantages to focusing on AI:

1. Engineering freedom
2. Regulatory freedom
3. Underexplored (contrarian)

Engineering freedom has to do with the fact that biotech and AI are fundamentally very different. Biology developed in nature. Sometimes people describe biological processes as blueprints. But it's much more accurate to describe them as a recipe. Biology is a set of instructions. You add food and water and bake for 9 months. There is a whole series of constructions like this. If the cake turns out to have gotten messed up, it's very hard to know how to fix it simply by looking at the cookbook.

biotech: a recipe



This isn't a perfect analogy. But directionally, AI is much more of a true blueprint. Unlike recipe-based biotech, AI is much less dependent on a precise sequence of steps. You have more engineering freedom to tackle things in different ways. There is much less freedom in changing a biological recipe than there is in designing a blueprint from scratch.

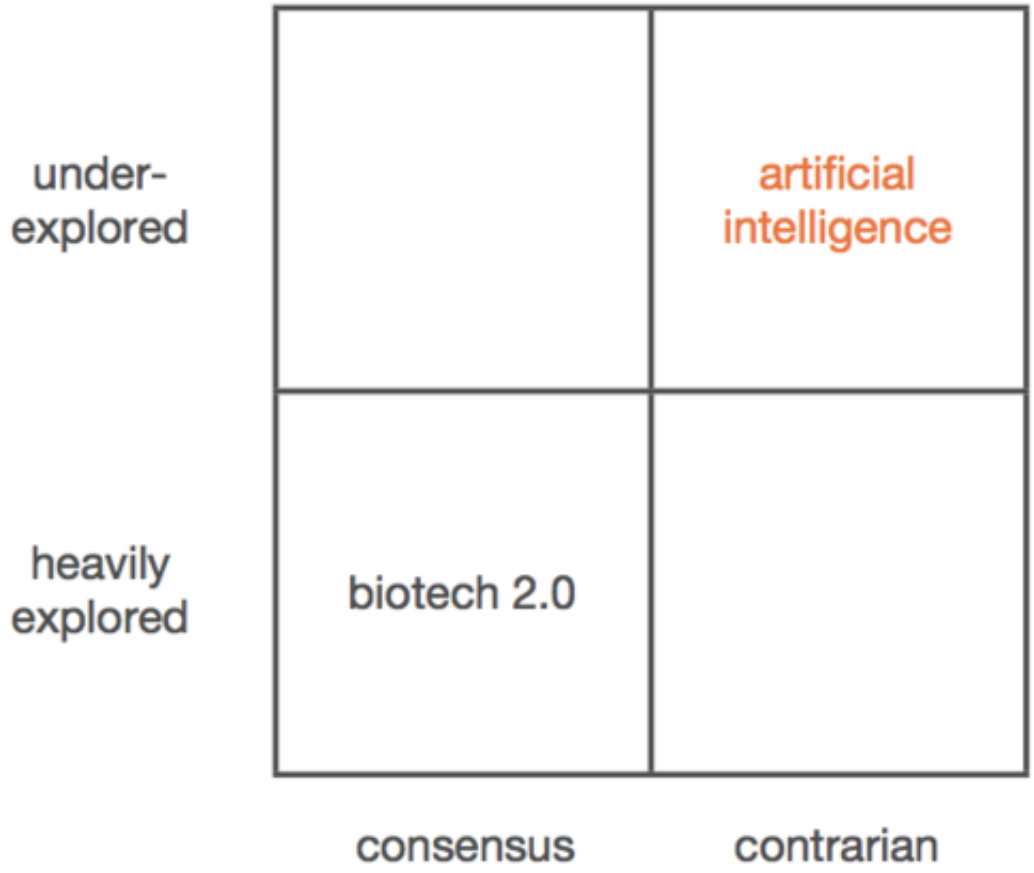
AI: a blueprint

On the regulatory side, the radical difference is that biotech very heavily regulated. It takes 10 years and costs $1.3 billion to develop a new drug. There are lots of precautionary principles at work. There are 4,000 people at the FDA.

AI, by contrast, is an unregulated frontier. You can launch just as quickly as you can build software. It might cost you $1 million, or millions. But it won't cost $1 billion. You can work from your basement. If you try to synthesize Ebola or smallpox in your basement, you could get in all sorts of trouble. But if you just

want to hack away at AI in your basement, that's cool. Nobody will come after you. Maybe it's just that politicians and bureaucrats are weird and have no imagination. Maybe the legislature simply has no mind for AI-kind of things. Whatever the reason, you're free to work on it.

AI is also underexplored relative to biotech. Picture a 2x2 matrix; on one axis you have underexplored vs. heavily explored. On the other you have consensus vs. contrarian. Biotech 2.0 would fall in the heavily explored, consensus quadrant, which, of course, is the worst quadrant. It is the new thing. The audience in Santa Clara last week was 100% bullish on it. AI, by contrast, falls in the underexplored, contrarian quadrant. People have been talking about AI for decades. It hasn't happened yet. Many people have thus become quite pessimistic about it, and have shifted focus. That could be very good for people who do want to focus on AI.

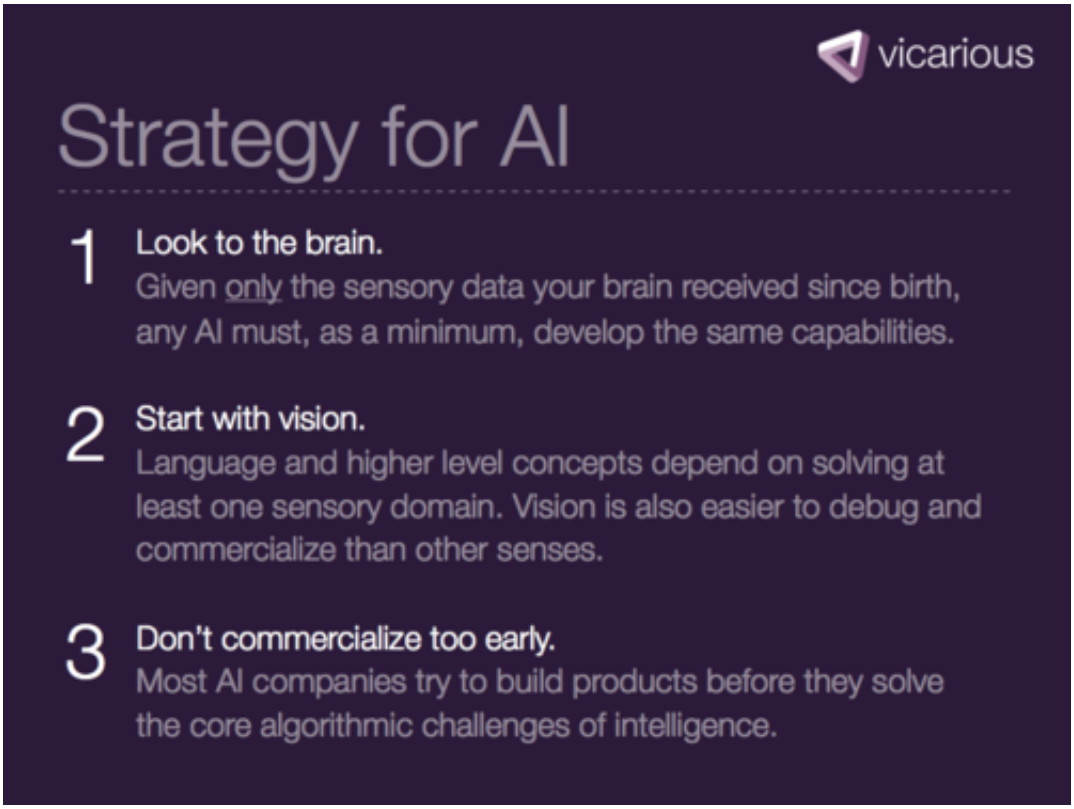


PayPal, at Luke Nosek's urging, became the first company in the history of the world that had cryogenics as part of the employee benefits package. There was a Tupperware-style party where the cryogenics company representatives made the rounds trying to get people to sign up at $50k for neuro or $120k for full body. Things were going well until they couldn't print out the policies because they couldn't get their dot matrix printer to work. So maybe the way get biotech to work well is actually to push harder on the AI front.

IV. Tackling AI

We have people from three different companies that are doing AI-related things here to talk with us today. Two of these companies—Vicarious and Prior Knowledge—are pretty early stage. The third, Palantir, is a bit later.

Vicarious is trying to build AI by develop algorithms that use the underlying principles of the human brain. They believe that higher-level concepts are derived from grounded experiences in the world, and thus creating AI requires first solving a human sensory modality. So their first step is building a vision system that understands images like humans do. That alone would have various commercial applications—e.g. image search, robotics, medical diagnostics—but the long-term plan is to go beyond vision and build generally intelligent machines.



Prior Knowledge is taking a different approach to building AI. Their goal is less to emulate brain function and more to try to come up with different ways to process large amounts of data. They apply a variety of Bayesian probabilistic techniques to identifying patterns and ascertaining causation in large data sets. In a sense, it's the opposite of simulating human brains; intelligent machines should process massive amounts of data in advanced mathematical ways that are quite different from how most people analyze things in everyday life.

Prior Knowledge invented a box that finds causes in data

p(k) prior knowledge, inc.

How?

Everyone knows Bayes' rule

- Universal mathematical rule for learning the causes of data
- Widely held to be intractable for real problems

PK makes it work

- Deep tech lets it scale larger than anyone believes
- Mathematical simplicity makes the *interface* simple too
- A simple interface puts the tech *everywhere*

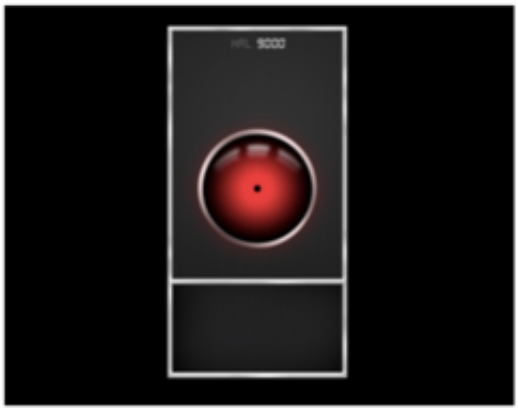
(Know Python or Ruby? Try it for yourself. http://dev.priorknowledge.com)

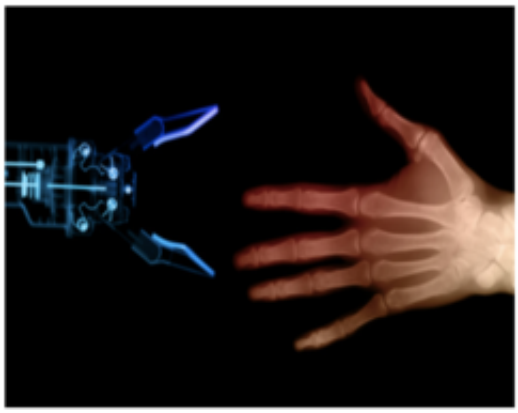
Palantir

The big insight at Palantir is that the best way to stop terrorists isn't regression analysis, where you look at what they've done in the past to try to predict what they're going to do next. A better approach is more game theoretic. Palantir's framework is not fundamentally about AI, but rather about intelligence*augmentation*.It falls very squarely within the Ricardo gains from trade paradigm. The key is to find the right balance between human and computer. This is a very similar to the anti-fraud techniques that PayPal developed. Humans couldn't solve the fraud problem because there were millions of transactions going on. Computers couldn't solve the problem because the fraud patterns changed. But having the computer do the hardcore computation and the humans do the final analysis, while a weaker form of AI, turns out to be optimal in these cases.

Stop thinking
Artificial Intelligence

Start thinking
Intelligence Augmentation





So let's talk with D. Scott Brown from Vicarious, Eric Jonas from Prior Knowledge, and Bob McGrew from Palantir.

V. Perspectives

Peter Thiel: The obvious question for Vicarious and Prior Knowledge is: why is*now* the time to be doing strong AI as opposed to 10-15 years from now?

Eric Jonas: Traditionally, there hasn't been a real need for strong AI. Now there is. We now we have tons more data than we've ever had before. So first, from a practical perspective, all this data demands that we do something with it. Second, AWS means that you no longer need to build your own server farms to chew through terabytes of data. So we think that a confluence of need and computing availability makes Bayesian data crunching make sense.

Scott Brown: If current trajectories hold, in 14 years the world's fastest supercomputer will do more operations per second than the number of neurons in the brains of all living people. What will we do with all that power? We don't really know. So perhaps people should spend the next 13 years figuring out what algorithms to run. A supercomputer the size of the moon doesn't do any good on it's own. It can't be intelligent if it's not doing anything. So one answer to the timing question is simply that we can see where things are going and we have the time to work on them now. The inevitability of computational power is a big driver. Also, very few people are working on strong AI. For the most part, academics aren't because their incentive structure is so weird. They have perverse incentive to make only marginally better things. And most private companies aren't working on it because they're trying to make money now. There aren't many people who want to do a 10-year Manhattan project for strong AI, where the only incentives are to have measurable milestones between today and when computers can think.

Peter Thiel: Why do you think that human brain emulation is the right approach?

Scott Brown: To clarify, we're not really doing emulation. If you're building an airplane, you can't succeed by making a thing that has feathers and poops. Rather, you look at principles of flight. You study wings,

aerodynamics, lift, etc., and you build something that reflects those principles. Similarly, we look at the principles of the human brain. There are hierarchies, sparsely distributed representations, etc.—all kinds of things that represent constraints in the search space. And we build systems that incorporate those elements.

Peter Thiel: Without trying to start a fistfight, we'll ask Bob: Why is the correct intelligence *augmentation*, not strong AI?

Bob McGrew: Most successes in AI haven't been things that pass Turing tests. They've been solutions to discrete problems. The self-driving car, for instance, is really cool. But it's not generally intelligent. Other successes, in things like translation or image processing, have involved enabling people to specify increasingly complex models for the world and then having computers optimize them. In other words, the big successes have all come from gains from trade. People are better than computers at some things, and vice versa.

Intelligence augmentation works because it focuses on *conceptual understanding*. If there is no existing model for a problem, you have to come up with a concept. Computers are really bad at that. It'd be a terrible idea to build an AI that just finds terrorists. You'd have to make a machine think like a terrorist. We're probably 20 years away from that. But computers *are* good at data processing and pattern matching. And people are good at developing conceptual understandings. Put those pieces together and you get the augmentation approach, where gains from trade let you solve problems vertical by vertical.

Peter Thiel: How do you think about the time horizon for strong AI? Being 5-7 years away from getting there is one thing. But 15-20 years or beyond is quite another.

Eric Jonas: It's tricky. Finding the right balance between company and research endeavor isn't always straightforward. But our goal is simply to build machines that find things in data that humans can't find. It's a 5-year goal. There are compounding returns if we build these Bayesian systems so that they fit together. The Linux kernel is 30 million lines of code. But people can build an android app on top of that without messing with those 30 million lines. So we're focusing on making sure that what we're building now can be useable for the big problems that people will tackle 15 years from now.

Peter Thiel: AI is very different from most Silicon Valley companies doing web or mobile apps. Since engineers seem to gravitate toward those kind of startups, how do you go about recruiting?

Scott Brown: We ask people what they care about. Most people want to make an impact. They may not know what the best way to do it is, but they want to do it. So we point out that it's hard to do something more important than building strong AI. Then, if they're pretty interested, we ask them how they conceive of strong AI. What incremental test would something have to pass in order to bea stepping stone towards AI? They come up with a few tests. And then we compare their standards to our roadmap and what we've already completed. From there, it becomes very clear that Vicarious is where you should be if you're serious about building intelligent machines.

Question from the audience: Even if you succeed, what happens after you develop AI? What's your

protection from competition?

Scott Brown: Part of it is about about process. What enabled the Wright brothers to build the airplane wasn't some secret formula that they come up with all of a sudden. It was rigorous adherence to doing carefully controlled experiments. They started small and built a kite. They figured out kite mechanics. Then they moved onto engineless gliders. And once they understood control mechanisms, they moved on. At the end of the process, they had a thing that flies. So the key is understanding why each piece is necessary at each stage, and then ultimately, how they fit together. Since the quality comes from process behind the outcome, the outcome will be hard to duplicate. Copying the Wright brothers' kite or our vision system doesn't tell you what experiments to run next to turn it into an airplane or thinking computer.

Peter Thiel: Let's pose the secrecy questions. Are there other people who are working on this too? If so, how many, and if not, how do you know?

Eric Jonas: The community and class of algorithms we're using is fairly well defined, so we think we have a good sense of the competitive and technological landscape. There are probably something like 200—so, to be conservative, let's say 2000—people out there with the skills and enthusiasm to be able to execute what we're going after. But are they all tackling the exact same problems we are, and in the same way? That seems really unlikely.

Certainly there is some value to the first mover advantage and defensible IP in AI contexts. But, looking ahead 20 years from now, there is no a priori reason to think that other countries around world will respect U.S. IP law as they develop and catch up. Once you know something is possible—once someone makes great headway in AI—the search space contracts dramatically. Competition is going to be a fact of life. The process angle that Scott mentioned is good. The thesis is that you can stay ahead if you build the best systems and understand them better than anyone else.

Peter Thiel: Let's talk more about avoiding competition. It's probably a bad idea to open a pizza restaurant in Palo Alto, even if you're the first one. Others will come and it will be too competitive. So what's the strategy?

Scott Brown: Network effects could offer a serious advantage. Say you develop great image recognition software. If you're the first and the best, you can become the AWS of image recognition. You create an entrenching feedback loop; everyone will be on your system, and that system will improve because everyone's on it.

Eric Jonas: And while AWS certainly has competitors, they're mostly noise. AWS has been able to out-innovate them at every step. It's an escape velocity argument, where a sustainable lead builds on itself. We're playing the same game with data and algorithms.

Scott Brown: And you keep improving while other people copy you. Suppose you build a good vision system. By the time other people copy your V1, you've been applying your algorithms to hearing and language systems. And not only do you have more data than they have, but you've incorporated new things into an improved V1.

Peter Thiel: Shifting gears to the key existential question in AI: how dangerous is this technology?

Eric Jonas: I spend a lot less time worrying about dangers of the underlying tech and more about when we're going to be cash flow positive. Which is why I plan on naming my kid John Connor Jonas…

More seriously, we do know that computational complexity bounds what AI can do. It's an interesting question. Suppose we could, in a Robin Hansonian sense, emulate a human in a box. What unique threat does that pose? That intelligence wouldn't care about human welfare, so it's potentially malevolent. But there might be serious limits to that. Being Bayesian is in some sense the right way to reason in uncertainty. To the extent that I'm worried about this, I'm worried about it for the next generation, and not so much for us right now.

Scott Brown: We think of intelligence as being orthogonal to moral intuition. An AI might be able to make accurate predictions but not judge whether things are good or bad. It could just be an oracle that can reason about facts. In that case, it's the same as every technology ever; it's an inherently neutral tool that is as good or as bad as the person using it. We think about ethics a lot, but not in a way the popular machine ethicists tend to write about it. People often seem to conflate having intelligence with having volition. Intelligence without volition is just information.

Peter Thiel: So you're both thinking it will all fundamentally work out.

Scott Brown: Yes, but not in a wishful thinking way. We need to treat our work with the reverence you'd give to building bombs or super-viruses. At the same time, I don't think hard takeoff scenarios like Skynet are likely. We'll start with big gains in a few areas, society will adjust, and the process will repeat.

Eric Jonas: And there is no reason to believe that the AI we build will be able to build great AI. Maybe that will be true. But it's not necessarily true, in an a priori sense. Ultimately, these are interesting questions. But the people who spend too much time on them may well not be the people who end up actually building AI.

Bob McGrew: We view the dangers of technology a little differently at Palantir, since we're doing intelligence augmentation around sensitive data, not trying to build strong AI. Certainly computers can be dangerous even if they're not full-blown artificially intelligent. So we work with civil liberty advocates and privacy lawyers to help us build in safeguards. It's very important to find the right balances.

Question from the audience: Do we actually know enough about the brain to emulate it?

Eric Jonas: We understand surprisingly little about the brain. We know about how people solve problems. Humans are very good at intuiting patterns from small amount of data. Sometimes the process seems irrational, but it may actually be quite rational. But we don't know much about the nuts and bolts of neural systems. We know that various functions are happening, just not how they work. So people take different approaches. We take a different approach, but maybe what we know is indeed enough to pursue an emulation strategy. That's one coin to flip.

Scott Brown: Like I said earlier, we think emulation is the wrong approach. The Wright brothers didn't

need detailed models of bird physiology to build the airplane. Instead, we ask: what statistical irregularities would evolution have taken advantage of in designing the brain? If you look at me, you'll notice that the pixels that make up my body are not moving at random over your visual field. They tend to stay together over time. There's also a hierarchy, where when I move my face, my eyes and nose move with it. Seeing this spatial and temporal hierarchy to sensory data provides a good hint about what computations we should expect the brain to be doing. And lo and behold, when you look at the brain, you see a spatial and temporal hierarchy that mirrors the data of the world. Putting these ideas together in a rigorous mathematical way and testing how it applies to real-world data is how we're trying to build AI. So the neurophysiology is very helpful, but in a general sense.

Question from the audience: How much of a good vision system will actually translate over to language, hearing, etc.? If it were so easy to solve one vertical and just apply it to others, wouldn't it have been done by now? Is there reason to think there's low overhead in other verticals?

Scott Brown: It depends on whether you think there's a common cortical circuit. There is good experimental support for it being a single circuit, whether incoming data is auditory or visual. One recent experiment involved rewiring ferrets' brains to basically connect their optic nerves with the auditory processing regions instead of visual regions. The ferrets were able to see normally. There are a lot of experiments demonstrating related findings, which lends support to the notion of a common algorithm that we call "intelligence." Certainly there are adjustments to be made for specific sensory types, but we think these will be tweaks to that master algorithm, and not some fundamentally different mechanism.

Eric Jonas: My co founder Beau was in that ferret lab at MIT. There does appear to be enough homogeneity across cortical areas and underlying patterns in time series data. We understand the world not because we have perfect algorithms, but also because tremendous exposure to data helps. The overarching goal—for all of us, probably—is to learn all the prior knowledge about the world in order to use it. It's reasonable to think that some things will map over to other verticals. The products are different; obviously building a camera doesn't help advance speech therapy. But there may be lots over overlap in the underlying approach.

Peter Thiel: Is there a fear that you are developing technology that is looking for a problem to solve? The concern would be that AI sounds like a science project that may not have applications at this point.

Eric Jonas: We think there are so many opportunities and applications for understanding data better. Finding the right balance between building core technology and focusing on products is always a problem that founding teams have to solve. We do of course need to keep an eye on the business requirement of identifying particular verticals and building products for particular applications. The key is to get in sync with the board and investors about the long-run vision and various goals along the way.

Scott Brown: We started Vicarious because we wanted to solve AI. We thought through the steps someone would need to take to actually build AI. It turns out that many of those steps are quite commercially valuable themselves. Take unrestricted object recognition, for instance. If we can just achieve that milestone, that alone would be tremendously valuable. We could productize that and go from there. So the question becomes whether you can sell the vision and raise the money to build towards the

first milestone, instead of asking for a blank check to do vague experiments leading to a binary outcome 15 years down the road.

Bob McGrew: You have to be tenacious. There's probably no low-hanging fruit anymore. If strong AI is the high- (or maybe even impossible) hanging fruit, Palantir's intelligence augmentation is medium-hanging fruit. And it took us three years before we had a paying customer.

Peter Thiel: Here's a question for Bob and Palantir. The dominant paradigm that people generally default to is either 100% human or 100% computer. People frame them as antagonistic. How do you convince the academic people or Google people who are focused on pushing out the frontier of what computers can do that the human-computer collaborative Palantir paradigm is better?

Bob McGrew: The simple way to do it is to talk about specific problem. Deep Blue beat Kasparov in 1997. Computers can now play better chess than we can. Fine. But what is the best *entity* that plays chess? It turns out that it's not a computer. Decent human players paired with computers actually beat humans and computers playing alone. Granted, chess is a weak-AI in that it's well specified. But if human-computer symbiosis is best in chess, surely it's applicable in other contexts as well. Data analysis is such a context. So we write programs to help analysts do what computers alone can't do and what they can't do without computers.

Eric Jonas: And look at mechanical turk. Crowdsourcing intelligent tasks in narrowly restricted domains—even simple filtering tasks, like "this this is spam, this is not"—shows the increasingly blurring line between computers and people.

Bob McGrew: In this sense, Crowdflower is Palantir's dark twin; they're focusing on how to use *humans* to make *computers* better.

Question from the audience: What are the principles that Palantir thinks about when building its software?

Bob McGrew: There is no one big idea. We have several different verticals. In each, we look carefully at what analysts need to do. Instead of trying to replace the analyst, we ask what it is that they aren't very good at. How could software supplement what they are doing? Typically, that involves building software that processes lots of data, identifies and remembers patterns, etc.

Question from the audience: How do balance training your systems vs. making them full-featured at the outset? Babies understand facial expressions really well, but no baby can understand calculus.

Scott Brown: This is exactly the sort of distinction we use to help us decide what knowledge should be encoded in our algorithms and what should be learned. If we can't justify a particular addition in terms of what could be plausible for real humans, we don't add it.

Peter Thiel: When there is a long history of activity that yields only small advances in a field, there's a

sense that things may actually just be much harder than people think. The usual example is the War on Cancer; we're 40 years closer to winning it, and yet victory is perhaps farther away than ever. People in the '80s thought that AI was just around the corner. There seems to be a long history of undelivered expectations. How do we know this isn't the case with AI?

Eric Jonas: On one hand, it can be done. There's an easy proof of concept; All it takes to create a human-level general intelligence is a couple of beers and a careless attitude toward birth control. On the other hand, we don't really know for sure whether or when strong AI will be solved. We're making what we think is the best bet.

Peter Thiel: So this is inherently a statistical argument? It's like waiting for your luggage at the airport: The probability of your bag showing up goes up with each passing minute. Until, at some point, your luggage still hasn't shown up, and that probability goes way down.

Eric Jonas: AI is perceived to have a lot of baggage. Pitching AI to VCs is pretty difficult. Those VCs are precisely the people who expected AI to have come much easier than it has. In 1972 a bunch of people at MIT thought they would all just get together and solve AI over a summer. Of course, that didn't happen. But it's amazing how confident they were that they could do it—and they were hacking on PDP-10 mainframes! Now we know how incredibly complex everything is. So this is why we are tackling smaller domains. Gone are the days where people think they can just gather some friends and build an AI this summer.

Scott Brown: If we applied the baggage argument to airplanes in 1900, we'd say "People have been trying to build flying machines for hundreds of years and it's never worked." Even right before it *did* happen, many of the smartest people in the field were saying that heavier than air flying machines were physically impossible.

Eric Jonas: Unlike things like speed of light travel or radical life extension, we at least have proofs of possibility.

Question from the audience: Do you focus more on the big picture goal or on targeted milestones?

Eric Jonas: It's always got to be both. It's "we are building this incredible technology" and then "here's what it enables." Milestones are key. Ask what you know that no one else does, and make a plan to get there. As Aaron Levie at Box says, you should always be able to explain why *now* is the right time to do whatever it is you're doing. Technology is worthless without good timing and vice versa.

Scott Brown: Bold claims also require extraordinary proof. If you're pitching a time machine, you'd need to be able to show incremental progress before anyone would believe you. Maybe your investor demo is sending a shoe back in time. That'd be great. You can show that prototype, and explain to investors what will be required to make the machine work on more valuable problems.

It's worth noting that, if you're pitching a revolutionary technology as opposed to an incremental one, it is much better to find VCs who can think through the tech themselves. When Trilogy was trying to raise their

first round, the VCs had professors evaluate their approach to the configurator problem. Trilogy's strategy was too different from the status quo, and the professors told the VCs that it would never work. That was an expensive mistake for those VCs. When there's contrarian knowledge involved, you want investors who have the ability to think through these things on their own.

Peter Thiel: The longest-lasting Silicon Valley startup that failed was probably Xanadu, who tried from 1963 to 1992 to connect all the computers in the world. It ran out of money and died. And then Netscape came the very next year and ushered in the Internet.

And then there's the probably apocryphal story about Columbus on the voyage to the New World. Everybody thought that the world was much smaller than it actually was and that they were going to China. When they were sailing for what seemed like too long without hitting China, the crew wanted to turn back. Columbus convinced them to postpone mutiny for 3 more days, and then they finally landed on the new continent.

Eric Jonas: Which pretty much makes North America the biggest pivot ever.