

MIT CSAIL Alliances | Mike Stonebraker_Project_4

Welcome to MIT'S Computer Science and Artificial Intelligence Labs Alliance's podcast series. I'm Kara Miller.

[MUSIC PLAYING]

Mike Stonebraker may be a legend in the world of computing, but that does not stop him from thinking constantly about problems people deal with and how businesses can solve them.

I don't get good ideas by sitting alone on a mountain top staring at my navel.

[LAUGHTER]

So I get ideas by talking to people, especially talking to real world users.

Stonebraker has started businesses, advised them, seen them struggle, and he's unfiltered in talking about the headwinds they face, from luring talent--

The big boys are offering enormous salaries. The people who have the most money are winning.

To getting the most out of that talent.

Software engineers have all decided to use ChatGPT, or copilot, or one of those tools. And the ones I talked to report a 30% productivity improvement.

So coming up on today's show, candid advice from a tech legend.

[MUSIC PLAYING]

Mike Stonebraker has been thinking about how to organize things for decades. He helped create the world of data storage that we take for granted today, and he did it both through academic research and by creating a slew of data related companies. For his work, he's been awarded the Turing Award, which is thought of as the Nobel Prize of computing, but Stonebraker is still in the trenches, with a new company, Hopara, which we will talk about, and with plenty of advice for decision makers.

In 2019 in a speech about common business blunders, he advised folks to start thinking about AI, about machine learning, which turned out to be a pretty good call.

My point of view is that ChatGPT is good at certain things and not very good at other things. And there are other ML techniques that are also very good at other things. And so ML is a basket of stuff which collectively will be extremely disruptive in my opinion, and ChatGPT is just one piece of it that's gotten a lot of press recently.

And how does that term, disruptive, manifest? Well, of course, it's going to change the lives of coders, but new tech is not just going to affect tech jobs.

I think another thing where it will be very disruptive is writing marketing PR, which are fairly simple, not technical tasks, and you just ask ChatGPT to do it. Also, my daughter is a high school math teacher, and ChatGPT is awfully good at writing recommendation letters. So I think on simple kinds of tasks, it will be very, very disruptive.

And while lots of folks are jumping on the bandwagon, Stonebraker worries that there are other areas of business where there's not nearly enough disruption.

Essentially, all sizable businesses that I know of are badly siloed. And if I was a CEO, I would be trying very hard to stamp out silos, and that doesn't seem to be happening.

Stonebraker says sometimes that means that people in different parts of a company don't really talk to each other or understand each other. Other times it means that pools of data don't talk to each other, and both can be very bad for business.

My favorite example is Toyota Motor Europe, sells Toyotas in Europe, obviously. And they have something like 250 distributors all over Europe and they're all data siloed. And so if you buy a Toyota in Spain and you move to France, there's total amnesia.

[LAUGHTER]

Nobody's going to help you in France because they have no record of that or something.

That's right.

Yeah, OK.

And so the dream of Toyota is to provide a seamless customer experience, obviously, because people move around. And so to do that, they've got to unwind these data silos, of which there are 250 and 40 different languages.

Wow.

So they're hard at work trying to do that.

Carnival Cruise Lines is actually a big conglomerate of nine different cruise line companies. So its Carnival Holland America, Costco in Europe, and there's nine of them, and they have nine different parts systems. So sometimes the parts are on the docks, sometimes they're on the boat, sometimes they're on in the warehouse. And so Carnival, the conglomerate, wants to share parts across these various silos, and so they have nine parts silos.

And of course, there's no common ID system for parts. So I call them rubber gloves and you call them latex hand protectors, no unique IDs. So in order to achieve business value it doesn't help you with my parts guy talking to your parts guy. We have to share the data in order to get value. So I think the data is more important than the people, typically.

It also sounds, from the way you talk about it, like this is like a huge hurdle in front of companies, like this isn't a little problem, this is a big, enormous problem.

Yep.

Yeah.

We were talking before about the reshuffling of, jobs potential layoffs, that kind of thing. You also have talked about, that there is an arms race underway for talent. On the other side, there's not that many people who really, really understand machine learning, for example. Do you feel-- when you talk to former students, when you talk to colleagues, do you feel like that arms race for talent is underway?

Oh, absolutely.

Yeah.

And I think the big boys are offering enormous salaries, enormous sign on bonuses. And say, the people who have the most money are winning.

Does that worry you in terms of--

Yeah, of course.

I don't know a new generation of--

Because startups, like us, we can't afford to do that.

Right.

And so one of the best students who wrote a bunch of the pilot that came before Hopara, he couldn't afford to come work for the company when we started because-- all kinds of personal reasons. But he needed cash, and that's something startups never have.

Mm-hmm.

And so one of the big boys was willing to offer 2x what we could afford to offer him. So I think this is to the detriment of startups, and seriously to the detriment of, let's say, traditional companies because Google and Amazon and Microsoft buy all the talent, and if you don't want to work for them, well, then next down the line is people like Oracle and IBM. And so people like Citizens Bank are well down that chain, and they have a very difficult time hiring world class talent.

It seems to me then you're tying into something I've heard other people talk about, which is this idea that the incumbents right now are so powerful and so rich, as you indicate, that I don't-- know if this is different from other periods in history, but does this mean that they take all the sunlight and the growth below them really struggles?

I think that's happening. Well, it's happening in the cloud.

OK.

But it's not happening in candy manufacturing, for example, So I think in an important piece of the economy, that's absolutely happening.

And I'll ask you one more question about incumbents, which is, there was recently a piece in The Boston Globe about Jeremy Wertheimer, who came out of MIT, started the travel site ITA, which was then bought by Google. He became part of Google, though he recently left. He indicated he felt like really the West Coast is the epicenter of a lot of exciting AI funding companies.

The numbers do seem to back him up, and I just wonder, we are sitting here in the Boston area-- obviously, a lot has come out of places like MIT, a lot of the basis for these companies. What's your thinking when you talk to venture capitalists and people who are thinking about both coasts?

I think that's-- so let me tell you about Hopara.

Yeah.

The engineering team is all in Brazil.

OK, interesting. Why?

Well, first of all, it's half the price of here.

OK.

And so I knew a really good guy who turned out to be-- who's now the vice president of Engineering at Hopara, and he lives in Rio. And of course, he's Brazilian, doesn't want to move here. And so the current engineering world in high tech is, if you want to hire the best talent, it's wherever they are, and most people aren't interested in moving. So I think West Coast versus East Coast doesn't matter. In other words, you want to hire the best talent, ' it's wherever it is.

And maybe the last few years have even increasingly dispersed that talent pool so that they're not just in these big cities.

Yeah. And I think San Francisco has-- I think it's 50% office space vacancy right now. I mean, it's huge.

Right.

Wow, this is a dark offices. This isn't necessarily unleased offices. And regardless of what the elephants do, they're trying to get people to come back to the office, and it's failing. And so I think off into the future, people are going to live wherever they live. And I think a lot of MIT graduates want to stay near the red line.

[LAUGHTER]

Yes. The red line is a pool.

So I think, from my point of view, the only allure of California is better weather. And to me-- I don't mind winter. I like it.

Well, I'll leave that there. People can write in if they see any other alerts or California, but I hear you. So let's talk a little bit about Hopara and digital twins and how you started down this path. I'd be interested to know.

Sure. So I am a huge fan of Google Maps and Google Earth. So you can start from an image of the Earth, and in 21 clicks, you can get to a plot map on the street where you live. And so basically, it's a detail on demand move and zoom paradigm. And there's no user manual for Google Maps. It's obvious how to do it.

And the only problem with Google Maps is, it only works on geographic maps. And so if you're somebody like Airbnb, who wants to drop business objects onto a Google Maps landscape, that's straightforward. But suppose you want to find 250 H Stata, which I had a really hard time finding--

Building at MIT place--

And so I want a floor plan.

OK. OK.

Right.

And Google Maps can't help you. And so I want a time series of occupancy of 250 H. And so if you want other kinds of objects than maps, then you're out of luck. So we built a prototype here at MIT that allowed you to use any kind of object and drop it on a two dimensional Canvas, and it could represent floor plans. It could represent maps. It could represent time series. It could represent scatter plots. We didn't really care.

And it had nice features. You could have more than one Canvas. So if you were sitting here in 250 H and you wanted to teleport to my home address, you can do that, that's simply another Canvas. And so you--

You can teleport? Now explain to me what you mean by that.

OK. So Gant turns out to have two houses.

OK, Gant, the CEO of Hopara, who's sitting next to you, I should say.

Yes, OK. Yeah.

Has two houses.

Sure.

So one of them is in the South end and one is in Exeter, New Hampshire.

Both lovely places.

And so in Google Maps, you can laboriously click from one place to the other, but you'd like to just jump there.

OK.

And so Hopara has a bunch of nifty stuff that Google Maps doesn't. So it allows much more data diversity and has a bunch of nifty new features. So we had this prototype running. And so I said, maybe we can make a commercial business out of this.

OK.

And so in the MIT-- well, in the CSAIL tradition, the original prototype, which was called KIRIX, is open source, anybody can have it. And we threw it over the wall to a commercial entity and started commercializing, and that's Hopara.

Let me bring Gant Redmon, the CEO, in for a minute here. So I want to ask you about the question of how you commercial-- I mean, it's a really interesting idea, as Mike has outlined, so let's talk about how you commercialize it. Before we get to that, a really critical question, which is the chocolatier in Exeter, New Hampshire.

Sure.

How great is it?

Yeah, got to bring a box to my mom every time I go down to visit her.

Yeah. OK, now to the real important stuff.

OK.

Yeah, so talk about how-- what Mike outlined that was academic, something that was academic originally and came out of MIT has commercial applications.

Well, absolutely. So if you look at Hopara, Hopara is about a new way to see massive amounts of data. So this really follows along with Mike's whole path. He created the database. He created normalization of data. Now it's in a database. It's organized. It's distributed. But now how do you see it? How do you really use it?

Mike was talking about Google Maps, Google Earth. They're super, but they stop at the roof. What we're doing is, we're going down, deep. We're going into the floor plan, into the room, into the device sitting on the bench, whatever you need as deep as you want to go to be able to go, what is that, and how does it relate to the things around them? So then you take this and go, OK, who would really be delighted by this? Whose lives are we going to change?

Right.

And it turns out it's people with lots of stuff, and they want to know what it is, where is it, and what is it saying, and that really ties in with IoT because now everything is sending you massive amounts of data. How do you make sense of it? You make sense of it by putting it in context and allowing people to navigate around and see it to derive those insights.

So give me an example of a kind of company that wants to go beneath the roof of Google Maps and has all this stuff that they are needing to see.

We've found we're really delighting lab managers and people in facilities because labs are full of lots of things, and the data that comes out of those things is really what drives the value of a pharma.

OK. That data has to be good. It has to be collected. And you have to make it easy on the people that are creating it. Those are often scientists. So the scientist wants to go, hey, I need to go use something, and I need these three or four things to accomplish the tests I have to do today. And she needs to go and-- OK, that's a room that has one of the things that I need.

How about a room that has all of them? Where can I go to make my life easy? How can I reserve it? Is it up to date on maintenance? Is it available? That's from a user standpoint. So the lab manager wants to delight the user. You delight the user. The user is more comfortable giving you the data in the form that they need.

Let's say you're the lab manager. We've had lab managers say, you know, I get alerts in the middle of the night. Oh, my gosh, there's a failure. There's this refrigerator with this really, really important sample, is just skyrocketing in temperature. It's getting hot. Something is definitely wrong. What's probably wrong is, somebody left the refrigerator door open. It can be solved, but he needs to solve it quickly. So instead of saying, OK, where is this room? What lab? What are we talking about?

If the screen pops up, it's Hopara, and you go, oh, I'm going to drill right-- oh, I see it's on floor four. It's over in the corner. It's in here. Oh, who's the security guard? Oh, have them go over. Just tell them to shut the door. And they want to make that easy for themselves and for the person solving the solution. So it's really just seeing that in context.

It sounds also-- to the conversation I was having with Mike, like it's bringing together all these different data streams. I mean, you were talking about, what's in a room? But has it been cleaned? Is it booked for today? And normally, those would be-- if you were looking to figure those things out, there are multiple different systems, or you might just have to show up at the room, but multiple different things you'd have to do to understand all those things. Does it have what I need? Is it available? Is it clean?

Well, totally. This actually goes along with-- I'll call it the Stonebraker progression, and that is, if you have different databases and you normalize them, you can actually draw data from all kinds of different places. So you're wanting to see these layers of information in context, but it doesn't have to be from just one database. It can be from multiple databases. So it's a visualization to bring multiple databases together.

Let me ask you one final question about this, which is-- when I think about going beneath the roof, that image of going beneath the roof on Google Maps to the next level of what's in the room--

Sure.

How do you actually build the digital twin? How do you actually create-- I mean, is it a question of taking pictures of everything? It seems complex and hard to do.

That is a great segue, and thanks for the opportunity. Because the problem with digital-- we love digital twins. We love using the term. Problem is, lots of people use the term. Lots of people have a preconception. Often that preconception is some holographic 3D representation. That's great, except not everybody needs a 3D representation of their spectrometer, or that piece of machinery. Maybe 2D is enough.

Right, right. They know what a refrigerator looks like. They just need to know it's there, but they don't need it to be in living color.

Absolutely. It can be in-- I don't want to oversimplify, but really, when we talk to lab managers, they don't use terms like, an agile digital twin. They say, I want to see my stuff on a floor plan. Absolutely. They want to see it real time, and they want to be able to drill in for more detail. That is what we mean by an agile digital twin. We're 3D when you need it, and we're 2D when you don't, and a lot of times, you don't.

When you play this out and you think about the plan a few years down the road, how do you imagine that this gets integrated into more businesses? We talked a little bit about what need do you feel like it's filling, but how does this become something big that a lot of people are using?

I'll tell you, I'm really counting on FOMO.

[LAUGHTER]

OK.

And what I mean by that is, if you can just see these things and derive the insights and you're making it easy, and suddenly, there's a company that adopts it. And then the competitor goes, wow, that looks really good. But even more importantly, the customer says, yeah, if you just look like that over there, that would make me a lot happier. So really, I see a future of, this is just normal. This is an expectation. You want to put data in a way that is easy to navigate and gain insights.

Gant Redmon, thanks so much, CEO of Hopara.

Thank you.

Mike, let me turn back to you. When you think about the trajectory of your career and the things you've worked on and the companies you've started, do they tend to be ideas that came about accidentally and they captivated you, or did you say, my next company, my next big endeavor, I want to be in this space?

The former.

OK.

I mean, I think-- this building is a caldron of interesting, and sometimes not so interesting, ideas.

[LAUGHTER]

You have to extricate the two, the interesting and the not so interesting.

It's a fabulous place to hang out and come up with ideas. I don't get good ideas by sitting alone on a mountaintop staring at my navel.

[LAUGHTER]

So I get ideas by talking to people, especially talking to real world users, talking to venture capitalists. I mean, you talk to people, and the light bulb goes off at some point, and you say, I can build that. So just for instance, I was listening to a talk by Matei Zaharia, who's Mr. Sparky. He wrote the original spark, and he's a founder of Databricks. So he was saying, well, Databricks routinely is managing a million spark subtasks.

And scheduling them is a horrendous problem that he couldn't solve with traditional techniques. So he put all his scheduling data into a Postgres database and was using it to schedule spark subtasks. And he said, it's too slow. And so I said, gee, I can do something about that. So those kind of things that are, by and large, random.

Yeah. It's hearing people's problems and trying to meet their needs. I don't know if you've ever thought about this, but if you were just starting out now and you were a researcher, is there an area that you feel like, this is rarely where I'd concentrate, these are the kinds of things I'd try to address?

[SIGH]

Well, right now if you don't have AIML and generative dot, dot, dot in your portfolio-- so that's what the world is hiring right now. But I think the question you want to ask is, what's the next big thing? That's the current big thing. And I think that's incredibly difficult to predict.

So for example, if you rewind to 1990 and you say, what are going to be the two most important big deals in the 1990s? Well, it's pretty much the fax machine and the internet. And who would have predicted that? So I don't think I have any crystal ball, and so I don't know how to answer your question.

That's OK, despite the fact that in 2019, you pretty much called it with AI machine learning. And last question as you look forward. Is there something that when you think about technology that concerns you the most or a problem that you worry we're not doing a good job of solving? It could be anything. But is there something that you sometimes lie awake at night being like, yeah, this seems like a problem to me.

Climate change.

Climate change.

And I think it's obvious that we're sailing off a cliff. And a lot of people are talking, but not anywhere near enough is happening. The other big issue that worries me is income inequality. It's getting worse and worse. And the 60% of the population is worse off in real dollars than they were 40 years ago.

And so they are, of course getting more and more and more frustrated, and that is what I think is the underlying cause of the political malaise. So I think these are two problems that if we don't address, things are going to get really bad.

I mean, I think people, obviously, think a good deal about climate change and income inequality, but I wonder if when you think about them you think about them at all through the lens of technology, and if you either feel like technology is the solution, technology is making them worse. How do you see those things intersecting with the stuff you think about all the time in your own life?

No, I think technology is making income inequality worse. I think technology may be able to help with climate change. I don't know if there's a magic bullet that will happen soon enough to make a difference. But I mean, I think it could help, but it seems to me the obvious thing is, we've got to put a lot less carbon into the atmosphere immediately, which I think is a political problem.

Yeah. Mike Stonebraker, what a fun conversation. Thanks so much for being here.

Oh, thanks, Kara. This was fun.

Gant, thanks for joining us too.

My pleasure. Thank you.

[MUSIC PLAYING]

And if you're interested in learning more about the CSAIL Alliance's program and the latest research at CSAIL, please visit our website at cap.csail.mit.edu. And catch our podcast series on Spotify, Apple Music, or wherever you listen to your podcasts. I'm Kara Miller. Our show is produced by Matt Purdy and Nate Caldwell with help from Audrey Woods. Tune in next month for a brand new edition of the CSAIL alliances podcast and stay ahead of the curve.

[MUSIC PLAYING]