Welcome to the OpenAI podcast, the podcast that opens up the world of AI in a quick and concise manner.

Tune in daily to hear the latest news and breakthroughs in the rapidly evolving world of artificial intelligence.

If you've been following the podcast for a while, you'll know that over the last six months I've been working on a stealth AI startup.

Of the hundreds of projects I've covered, this is the one that I believe has the greatest potential.

So today I'm excited to announce AIBOX.

AIBOX is a no-code AI app building platform paired with the App Store for AI that lets you monetize your AI tools.

The platform lets you build apps by linking together AI models like chatGPT, mid-journey and 11Labs, eventually will integrate with software like Gmail, Trello and Salesforce so you can use AI to automate every function in your organization.

To get notified when we launch and be one of the first to build on the platform, you can join the wait list at AIBOX.AI, the link is in the show notes.

We are currently raising a seed round of funding.

If you're an investor that is focused on disruptive tech, I'd love to tell you more about the platform.

You can reach out to me at jaden at AIBOX.AI, I'll leave that email in the show notes. Welcome to the AI Chat podcast.

I'm your host, Jaden Schaefer.

Today on the podcast, we have the pleasure of being joined by Ari Kaplan, who is a leading influencer in analytics, artificial intelligence and data science.

With a career that has touched everything from Major League Baseball to Fortune 500 companies, he's known as the real money ball guy.

He has spearheaded transformative AI and analytics projects and is the co-author of five bestselling books in these fields with roles that are ranging from a Databricks head of evangelism to leading the Chicago clubs analytics department, Ari's diverse experience positions him as a pivotal voice in the AI community.

Welcome to the show today, Ari.

Hey, thanks.

Great to be here.

I loved listening to a lot of your podcasts and appreciate everything you do.

Super exciting, very eclectic group of people and always inspiring.

Yes.

Yeah.

Very eclectic.

Super excited to have you on the show.

As I said, something I'd love to kind of kick this off with is asking you a little bit about your background and your journey in tech specifically.

Did you always know that you are going to kind of be in tech because it's always something you're interested in AI now today?

Or is this something that you kind of discovered as you went through?

Tell us a little bit about Walker.

So you're kind of background and journey.

Sure.

No, I'm a naturally curious type of person.

So I love exploring new things and really in technology and innovation.

So I find like through my journey, every like three or four years, there's some like major shift, your generative AI right now, more traditional AI, you know, four or five years ago, lake house technology, database, you know, and you keep rewinding.

So it's always really been in technology and I think kind of drives with the human fascination of taking the real world and trying to quantify, digitize it, whether it's old school chemistry from 500 years ago to the physics.

How can you take something in the real world and automate it and repeat it through technology? Yes.

Very cool.

That's incredible.

I think curiosity has led a lot of us into this space.

So much is changing.

So much is evolving.

So that's really, really cool.

Something I want to like double, double click on from your past that seemed really interesting to me for those that have seen them, the movie Moneyball, they'll know what, what you're talking about here.

But tell us a little bit about your role, how you got into major league baseball and how your role in analytics kind of played in that whole thing.

And maybe, you know, explain for people that haven't seen the movie.

Sure.

No, it's been a real honor that most of my career has been around sports analytics.

And you know, when, when I, the journey's been, you know, pretty unique when I started, I was one of the first four people that have been known to work in some capacity is what you might call like a data engineer or data analyst.

Okay.

So I was just a student at Caltech.

If you've seen the Big Bang Theory, that's kind of the school.

I resonate, it feels like it was based on my life a little bit, characters would spot on and how they behaved.

But yeah, I was an undergraduate and had this summer undergraduate research fellowship to just show that there are statistics that were out there that are better than were commonly used and was able to explain that and communicate that in simple terms.

And Fred Clair, who was the general manager of the Dodgers had heard me starting to get into the news media, you know, it is Hollywood LA area after all and said he thought it might help him identify or trade or, you know, work better with his team.

So he invited me up and did some work immediately with some of, you know, future Hall of Fame

players, Orle Hirschizer, Kirk Gibson.

So that was like way before Moneyball, but always had to keep reinventing from coming up with better ways to analyze players.

And mind you, these are like significant decisions.

Right.

If you get one player different, it's playoffs or not, but then merging into scouting technology, how do you quantify what scouts are saying so that it's like everyone's talking on the same scale and people like who are non-technical.

Maybe we can talk about, you know, that democratization aspect, but back then it was the democratization

of being able to query data without having to know sequel or information.

And then Moneyball was like a great movie, my opinion in book about how this like really changed the whole industry, how they were first to move in some regards, and that helped get them, you know, good players with lower cost.

And now fast forward, the movie's been an inspiration to a generation.

It's been over 20 years.

And every team went from zero to one data scientist analyst to 30, 40, 50 for each team. And now went from an industry of about four people to, you know, tens of thousands of people.

If you include like vendors and the people that work for teams all around the world. So that was like the movie changed my life.

And then it, the, the great thing actually happened after the movie, after every team implemented analytics, it changed the game so much that they just last year had to change the rules to like limit defense shifting.

Since it got quote, guote, boring, it worked too well.

That's so funny.

Yeah.

I mean, I feel like with any kind of arbitrage like that, whoever gets it at the very beginning, so kudos to yourself for kind of being on the cutting edge of that sees the major benefit. And when everyone adopts the same strategy, it's, it's a little trickier, but that's awesome. And yeah, I think it really kind of goes to show the importance of focusing on data with these players obviously beyond just, you know, their skills or what you think they can add to the team really looking at taking an analytical approach.

It's awesome in sports, but that also applies to so many different areas.

And so I think, you know, that whole movie and movement kind of helped raise awareness for that.

I think for a lot of different industries as well, so very, very exciting.

So currently you are the head of evangelism at Databricks.

Tell us a little bit.

I would be curious, like, how did you, how did you come into that?

How did you meet Databricks?

What's your, what's your story with them?

Yeah, great story.

And if you haven't heard the title evangelist, it's kind of a new, a newish, newer trend that tech companies are doing.

It's not like preaching religion specifically, but it's, you know, like a brand ambassador or somebody who can, like going on here with you, Jayden, people who've never heard of us explain it or explain it in simple terms, sometimes things get a bit technical.

So kind of like, what do you do and why?

And I had that role a couple of times before my prior job, I created and led evangelism at DataRobot, which is auto machine learning.

And then I was the head of the user community worldwide for Oracle, a big database company that when we acquired MySQL Java, people know of, people soft.

But yeah, with, with Databricks, you know, a quick growing company.

I'll explain who we are in a minute, but, you know, after a certain size, when you're certain growth, you know, you need to expand and build out your brand.

And so a role came up and rewind all the way back to that Oracle mention.

But Rick Schultz, who's our chief marketing officer, he and I go back to our Oracle days.

So there was that good personality, you know, that relationship, which is so important.

And so, yeah, that came about.

And now it's, you know, a great opportunity.

You know, Databricks, if you haven't heard of us, we created what's called the Lakehouse technology, and that joins the old school, like structured database data warehousing, you know, like Oracle, the equal server of the world, and unstructured data that they call data lake.

So machine learning, video, audio, you know, PDF things.

You might have heard of the data swamp.

It's like ungoverned, just like pop data into files.

Yeah.

So the Lakehouse solves the problem of having one unified environment where you have all that data in one place.

And it's based on open source technology.

Our founders actually created Apache Spark and offload Delta Lake, which gets hundreds of millions of downloads a year blows me away.

Crazy.

But because it's, yeah, it's incredible.

It performs typically, you know, many times faster.

You know, everyone is different, but multiple times faster than traditional data lake data warehouse.

And it's oftentimes 10, 20, 50, 80 percent less expensive.

And it's simpler.

So usually you have to pick one of the three less expensive, faster or simpler, but you get the benefit of all three.

So that's in a nutshell, Databricks, you know, we did make news a couple of weeks ago when we raised an additional round, which funding, which makes us the third largest private tech company in the world behind SpaceX and Chime.

So incredible.

Yeah.

And congratulations.

It's, you know, and to your, to your and the team's credit, looking at a lot of the moves Databricks is doing, I've been super, super impressed.

And I think it was just earlier this year, right?

There are a big merger happened between Databricks and Mosaic ML.

So I see Databricks really doing some, some incredible moves in the space.

What are some of the things that you're most excited about that Databricks is currently working on or implementing?

Yeah, there's so many exciting things.

And, you know, our core lake house technology, you know, what we've been doing up to this date, it's now 1.5 billion a year in revenue and it's focused on data and AI, everything from the data engineering, you know, how do you get these pipelines and these flows to the machine learning to the ML ops to operationalizing it.

And so that's all exciting, what we're doing at the core.

And, you know, what we're, what we have now in building the whole marketplace is exciting.

But what we call Unity catalog, which is everything from the governance, you know, how do you do audit trails?

How do you understand what version of your data is accurate?

What version of your data is from five years ago?

You know, you want to understand that to the transparency.

So understanding that we talk, everyone's talking about LLMs and we don't know if it hallucinates, but having that governance to understand where your data comes from, maybe have it privatized.

So things that are private to your company, don't get out in the world.

That is getting more and more important.

So that's our Unity catalog.

And then you had mentioned Mosaic ML.

That was the large acquisition just a couple of months ago.

And building large language models is like all the attention these days.

I think it'll get back, you know, a little bit where we'll be large language models and more traditional predictive modeling and classification.

But that's super exciting.

The use cases that are coming out and that are going to come out that we haven't even thought of is super exciting.

And then like looking like even more forward, Databricks, other companies, we are looking to leverage AI within our products.

So like as you're a coder, you're writing Python or, you know, would pick your language.

It helps, you know, used to just be autofill the name of the table you're looking up. Right.

AI can like help suggest comments and do quality assurance, synthetic data generation.

So it's like a code assist, a co-pilot as you code, which is really exciting.

And what we're able to do now is great.

And one year from now, two years, like the world of software development is going to be like advanced, well, well ahead of what it is now.

Yeah, I see a lot of a lot of changes and shifts.

And it's kind of interesting because I feel like when you look at a lot of

different players in the space, Databricks is definitely one that is

is going to be a really big winner in this whole AI space.

And for good reason, they're doing some really cool things, making some cool acquisitions.

I believe you just raised, you know, \$500 million at a \$43 billion valuation.

So congratulations to the whole team on that.

Really exciting to see, you know, the market's confidence in essentially your strategy and your plan.

Something I'd be curious to ask you about is, you know, what do you think sets Databricks apart from perhaps competitors or other people that may try to do some of the similar things? What do you think is, you know, the secret sauce and some of the uniqueness of Databricks?

Yeah, you know, great question.

And, you know, we've had a lot of success creating the Lakehouse.

So, you know, there are and there will be other companies coming into the space.

You know, all different things.

Number one is we were built like the whole company.

We created Lakehouse and the foundation was open source.

So as an effect, like we're not taking existing software and trying to tweak it and back fit it.

So as a result, you know, we have companies running their own benchmarks against, you know, anything out there and benchmarks, you know, by and large are like way, way faster, way, way less expensive.

So people like, especially in this economic environment, we have to kind of watch your bottom line migrating.

The other thing is the scalability.

So like, dated basis, you know, level work.

I used to work there as the head of the user group, but they, you know, you only scale to like the billions of records.

And, you know, it's kind of wild, but now data sets are hundreds of billions.

We have companies that have trillions of records, you know,

Grammarly five billion new, you know, grammar things people are entering every single day.

So that's, you know, hundreds of billions a year.

The level of data just doesn't scale.

So if your level of data goes up 10 fold, you know, it's more like the

cost consumption is more linear with us.

And it's more exponential with other solutions.

The secret sauce is that since the lake house has everything in one platform

with other platforms, you have to copy data, migrate data.

You also have to have like two or more governance, like different user names, different passwords.

The lineage of some data is here and some over there.

It's just like inconsistent.

But, you know, really you avoid by avoiding copying of data everywhere.

I mean, that, that's a huge deal when you get large data sets.

Yeah, that definitely makes a big difference.

Something I'd be curious about is like, how does the lake house platform compare in terms of like price and performance with, you know, like traditional cloud data warehouses or something?

Yeah, you know, great question.

There's, you know, more neutral like Gartner, Forrester, MIT report that have done their own benchmarks.

Customer, you know, customers do their own benchmarks as well. And it, you know, all depends on the size of the files and the size of the data, but we, we're, we are winning accounts from pretty much everywhere else because of those three factors.

So it's, it's not a marketing trick.

It's like in real life, people are like ditching, you know, what you might call legacy systems and moving to what we call the modern data stack for all those three reasons.

Yeah, yeah, I definitely see that talking to people in the space and then something else that I have heard, but I'd love to get your, you know, pick your, get your thoughts on it.

But I'm wondering if you can elaborate a little bit about how lake house platform is accelerating the machine learning like life cycle. Yeah, yeah.

So machine learning life cycle, something I've done.

I've been a hands on data scientist, you know, beyond being someone like an evangelist that talks about it, like with the Chicago Cubs created all of those analytics from scratch.

So the challenges I faced, you know, having been a practitioner are greatly accelerated through a data brick.

So there is, you know, all different steps.

There's ingesting the data that's called ELT or ETL, depending on what you're doing, you're, you know, loading data, you're transforming data. So it's like this whole process flow and Databricks has like really easy to use capabilities that you could give rules.

Like when do you reject something super helpful when you're doing like real time streaming, like you're ingesting social media, you're ingesting, you know, claims or you're ingesting, you know, sales or something like that, as things happening real time, what are the business rules

to process it and transform it.

But then once you have that machine learning, you know, the, there's traditional machine learning on structured data only, but the more types of data you can put into a machine learning model, oftentimes that gives you better insights, it's more fine tuned to the reality, to the complexities, innuendos of real life.

So Databricks does a great job taking, it's called multimodal, some structured data like sales, some unstructured data like transcript of a complaint call, social media images and make a better forecast in there. So that's making the forecast.

And then once you have the forecast and you like productionalize it, it's kind of one of the final steps.

The model, is it perfect?

Maybe not, but is it good enough to productionalize?

Maybe put it into production, which used to be a really hard task.

It would take you days, now it's like click of a button.

And then once you do that and you're scaled to have hundreds of models or thousands of models, it's this whole new sub industry called ML Ops. How do you have a dashboard of hundreds or thousands of models and see where the data is drifting?

Know, when am I going to refresh the data?

Where has a model failed?

Like where has the data stopped ingesting?

So now you have like dashboards and alerts for hundreds or thousands of models to just understand who's using which models, what assets are being used and who, yeah.

Yeah, that's incredible.

That's very cool.

Yeah.

I think that's a big part of the solution to what was otherwise a very large problem to grapple with for a lot of people.

So I think that's definitely one reason why Databricks is excelling.

In a lot of ways, I definitely hear a lot of positive things about it.

Something I would love to ask you, just kind of from your own perspective and in your own background and everything you've seen in AI machine learning.

I'm wondering what's a piece of advice you feel like you could give to, you know, aspiring machine learning data scientists, AI experts, people in this field, what's a piece of advice you feel like you could give?

Yeah.

Great question.

And I do teach some college courses on data science.

So love giving advice, you know, all depends on where your journey is.

But, you know, don't, I guess my big advice is, you know, be a self-learner.

Things are changing so quickly.

You know, there's so many different websites and YouTube videos.

You're a podcast to listen to.

So like always be learning and then like be humble enough to ask questions when you don't know, you know, you get stuck.

A lot of people are worried about like their own, like it looks bad to ask questions.

It's OK to fail as long as you're learning and it's OK to not know something.

You know, like at my company, I tell people, how does this work?

How does that not work?

How do I do something?

And so it's a great culture, but, you know, even even myself, I've, you know,

had had a I'm not starting out.

I've had a career, but yeah, the advice of people starting out

and anywhere along the way is, you know, just that keep learning,

keep asking questions and don't worry about being vulnerable and failing.

I love it.

That is definitely impactful advice.

I mean, that's good advice that applies to anyone in so many different industries and so many different areas that people are looking to to go into or to learn about.

But yeah, especially, I think, an AI that's really, really critical today,

something I would love to ask you about as well as, you know, what are some

misconceptions about AI that you encounter in your role as an evangelist

and a thought leader in this space, you know, some things that people are just getting wrong? Yeah.

Well, I would say like we're in the phase now of gen AI, which is not even a year old.

So there's like a lot of

like a lot of misconceptions and a lot of people

you are focusing on what things can and can't do.

So I think like the main thing is people who just see gen AI as like an enhanced

chatbot and think that that's the only like use case that you could do

or try to back fit some public publicly generated LLM

like chat GPT and like use it in their company.

Sometimes that works, but, you know, you a lot of companies are starting to realize

that you don't need Taylor Swift in your database.

You need something based on your own data for, you know, for certain use cases.

So yeah, I think that's a misconception that you could use publicly available general like general gen AI models.

And it's like a cure all for, you know, for what your business needs.

And most cases, it's not.

Yeah, I think that's a really, really important, definitely a misconception

that a lot of people have in the space.

But yeah, that's that's some great advice, I think there for sure.

Something I'd love to ask you a little bit about, of course, you are a best selling

author, you've written a number of really incredibly well received books.

I'm wondering if you could give us a little brief overview or summary or pitch for, you know, your books and some that you feel like are really relevant specifically today for people to read.

Yeah, well, I've gone from specific to general, but they're all like technology.

So like how to do Oracle how to was one.

Baseball hacks is like every baseball team has that on their shelf.

It's like how to write our and Python code to like scrape data off the web and visualize it.

That that was a big one.

One that's out of date now, but that was at the time, my bestseller was the first book on Windows 2000.

And it was wild.

It became such a bestseller since Microsoft always comes out with their own books, but their book got delayed by like a month or two.

Oh, veah.

Barnes and Noble, like it was the only one on the shelf if you needed to learn it.

So yet it made like top 20, like beating out some for at least that window,

like Oprah book club and everything.

So that was wild.

I went to Russia and I saw it like translated without my permission

in Russian on the street.

It was pretty, pretty fun.

And I saw I went to Japan, saw it there on the street.

Oh, that's funny.

When you make something that people need, I'm just going to do whatever it takes,

I guess that's hilarious.

That's awesome.

I'm wondering one other thing that I looked into and I saw about you is that

I know you have some patented mobile technology.

I'm wondering if you can talk a little bit about that and maybe some of its implications for the future of AI and analytics.

Yeah, you know, great research theory.

You know, I was talking about that like every four years reinvent yourself.

So one of the the ones I didn't mention was mobile technology.

And I worked at U.S.

Robotics that made what was called the the Palm Pilot.

And people may not remember that, but it was like personal digital assistant

that started the whole like mobile craze.

Motorola came out with something then BlackBerry.

They just made a movie on how they epically failed to capture it.

And then a couple of Microsoft tried to get into it.

You know, Apple, Samsung and others are there.

But at the time, it was the first business software that worked on a mobile device.

So think like you can have a notebook on your computer writing code.

You could do if there are hardcore people on this podcast, like an SSH,

like a telnet into a server and write command lines from your mobile device.

So how do you manage an Oracle or a Teradata database?

People used to be on call and they would get paged and have to come back to the office into their home, which could take an hour.

It could take 10 minutes.

But here you could be out and about and respond immediately.

Type in a command, restart or what have you.

So that that was wild.

It was a company expand beyond.

We raised the largest series A in Illinois in 2001, got a bunch of patents there and, you know, got later acquired.

But it was really exciting to be there, you know, at the very forefront of the mobile revolution for business.

And now all of us, you know, Databricks, but, you know, everyone here we call,

like we just had a conference and our theme was Generation AI.

So we all, whether you're Databricks or not, we're all part of this,

the first generation of shaping what is Gen AI and how do we explore the possibilities, make it the best we can.

And while at the same time, you know, limiting it in the right way, governing it in the right way.

So that digital mobile database and now Gen AI is the latest.

Very cool.

Yeah.

It's incredible to see some of the technological advancements and shifts that we're seeing today.

That's cool that you've been kind of at the forefront of each of these.

And now you're kind of seeing this new kind of wave of AI.

It's been incredible to have you on the podcast today as we're wrapping up.

I'd love to ask you, you know, one last question, which is based on your unique perspective, I believe you have right now in this space,

what are some of the biggest changes and advancements you see coming down the pipe in like, let's say the next three years in the AI space that you're particularly interested and excited about?

Yeah.

Well, yeah, it's been fantastic coming on here.

And, you know, three years used to seem like a small time horizon,

but now it's like, I have to put on my futurist hat.

I used to say five to 10 years.

Now I'm like three to five.

Yeah.

Just in the last, you know, since last November, it's incredible.

You know, there's going to be hundreds of companies coming out that like don't exist now that are going to come out with incredible technology.

So, you know, the main conceptual areas, you know, Gen AI, I'm just starting to see some really incredible use cases like beyond your traditional ask a question and text, and it comes back with like resources with functions, you know, everything dealing with like real time video, things that are going to make interactions with humans much better, humans interacting with each other much better.

The other thing I see rapidly changing is the world of software development itself. Yeah.

So like I have kids that are in college, you're going to go to college and I'm like, what, what are you going to learn that will still be applicable?

Probably the concepts, but a lot of the first level, like easy, boring parts of software development are, they're all going to be automated.

So everyone's going to, if they're able to be elevated to do more and more complex.

So that is exciting, but I just see this onrush of data that's out there.

So you're one other thing I do want to pull back with the Databricks is we have this credible marketplace, which is sharing of data, either open or closed.

So the reason I pulled that back is to push forward that there's going to be huge market places since it's very costly to make your own, you know, trillion record data set.

So I see companies where they can and should sharing data.

So there's not like 50 companies scraping the same data that will be shared.

But then every company will augment that with their own proprietary, like customer data.

So I think that's going to be exciting.

This entire like global marketplace of data, of videos, of social media.

And as that grows, the use cases are going to keep growing.

And I want to be blown away.

I want to see use cases that we're not even thinking about now start happening in three years. Yep. 100%.

I think, I think at the rate of progress we're making right now, that's not very far away.

Listen, Ari, it's been incredible to have you on the podcast.

Really appreciate all of your insights you've shared and your unique perspective and background. If people want to get in contact with you or learn more about Databricks and what you guys are

building over there, what's the best way for them to do that?

Yeah, well, databricks.com is great.

We, you know, have all sorts of like hands on labs and demo, a demo hub, a demo center with like videos of everything.

I love it if people want to follow me or connect with me on LinkedIn.

Just search for Ari Kaplan.

I'm not the lawyer who's popular there.

I'm the guy from Databricks.

But I post my insights.

I try to make it fun.

And there's that way you can kind of learn and see where I travel around the world.

And, you know, hopefully I give all different nuggets of wisdom as I see and talk to people.

So I love to connect on LinkedIn.

Very cool.

I'll end up for the listener.

I'll leave a link to Databricks in the show notes.

You can go over there and check out some of the really cool things they're doing.

Ari, again, thank you so much for coming on the show to the listener.

Thank you so much for tuning in to the AI chat podcast.

Make sure to rate us wherever you get your podcasts and have a fantastic rest of your day.

If you are looking for an innovative and creative community of people using chatGPT,

you need to join our chatGPT creators community.

I'll drop a link in the description to this podcast.

We'd love to see you there where we share tips and tricks of what is working in chatGPT.

It's a lot easier than a podcast as you can see screenshots.

You can share and comment on things that are currently working.

So if this sounds interesting to you, check out the link in the comment.

We'd love to have you in the community.

Thanks for joining me on the OpenAI podcast.

It would mean the world to me if you would rate this podcast wherever you listen to your podcasts.

And I'll see you tomorrow.