**CRAIG:** Hi, I’m Craig Smith and this is Eye on AI.

I heard from some far afield listeners after my appeal last week, but none from Nepal, Bangladesh or Kazakhstan. Come on people, send me some love. I’m wanting.

This week I’m doing something different: our guests are the hosts of another popular AI podcast, AI Today, produced by the team at Cognilytica. And they are having me on their podcast at the same time, so if you want to hear me opine instead of just asking questions, go over to AI Today and listen in. I encourage all my listeners to subscribe to AI Today, which is an excellent podcast, and leave a review for them on whatever podcast platform you use. The likes and reviews honestly make a big difference in attracting new listeners.

Ron Schmelzer and Kathleen Walch founded Cognilytica xxx, and have grown it into a premier research, advisory, and education firm focused on advanced big data analytics, cognitive technologies, and evolving areas of artificial Intelligence and machine learning. They track tens of thousands of technology vendors and provide services to enterprises of all sizes. I am a great admirer and turn to their reports often.

We talked about the rise of MLOps, data labeling, unsupervised learning and what it will take to get us to human-level artificial general intelligence. Both Ron and Kathleen had some surprising insights. I hope you find the episode as informative as I did.

**CRAIG:** One of the reasons I'm interested in talking to you guys is you have tremendous coverage and very deep coverage of what's happening in the ML space.

I tend to cherry pick depending on my interests, primarily with basic research. But you guys are much more complete.

Can you introduce yourselves to our listeners?

**RON:** Well, first of all, I want to thank everybody who's listening to us. We're thrilled to be on this podcast. I'm Ron Schmelzer, I'm a managing partner and principal analyst here at Cognilytica and we're a research advisory and education firm focused on the universe of AI and machine learning.

We've been adding more on automation as well. That's very hot area and even more advanced analytics, maybe Kathleen, you want to introduce yourself as well.

**KATHLEEN:** Yeah. I'm so excited to be here as well. I'm Kathleen Walch. I'm also a managing partner and principal analyst at Cognilytica. As Ron mentioned, we're an AI focused research advisory and education firm.

We track many vendors in the space over 20,000, actually. So, you know, we, we have a good pulse on what's going on, what markets seem to be expanding, which ones are contracting, uh, you know, there's lots of deals happening as well. So, we, we definitely have a pulse on all of it. And then in addition to that, we also have our own podcast called the AI Today podcast. We're excited to always be guests on podcasts and for, Craig's listeners out there, he will be on our podcast as a swap.

So, we encourage you to check that out. If you're interested in hearing what he has to say, when we interview him on the AI today, pod.

**CRAIG:** Wow, you guys are good. Okay.

Much, much better than me.

I I've looked sporadically and machine learning platforms and I'm interested in the tooling behind machine learning and how machine learning models get built. You guys track over a hundred vendors involved in the space from toolkits and notebooks to machine learning as a service. my first question is, is the market scratch rated and without picking winners, what products do you see rising to the top?

**RON:** Yeah. It's actually good that you're asking us this question right now in like in, in the general timing of the world, because here we are -for those who are listening to the podcast is August of 2021, people might be listening to this podcast a year from now.

So, this will all seem really quaint to those in the future. But, um, the market's actually in the midst of a consolidation phase, we're actually starting to see a lot of acquisition activity and we do track, over a hundred vendors in the machine learning platform space about 72 of which meet the minimum threshold of viability.

There's lots of startups in the space. We love startups. We have an affinity companies of all sizes, but, when we're looking at companies who are buying products and services, we tend to look at those companies that have either at least 10 customers or have at least 10 million in revenue, or at least $10 million in venture capital 10, if they have like two customers and no venture funding raised and a little bit of revenue, then we're like, just grow a little bit more, we'd love you to just grow a little bit more. So that's about 72 companies, at least that are in that genre. Of course, all the cloud vendors, you know, are in that space , the major cloud vendors, Microsoft, IBM, Google, Amazon, uh, and a few others.

So those are what we call the cloud SAS machine learning as a service vendors basically. And then there's a whole other category of pure play machine learning platform vendors. So, you may be familiar with a bunch of others that are kind of trying to.

Pull together, all the components of what's required to put machine learning and advanced analytics solutions into play. And increasingly what they're doing is they're growing through both building out their product suites and through acquisition. Especially DataRobot has been on, uh, uh, tear lately, Dataiku as well has been, been really growing and they raised a very significant round recently.

But the answer is that this market is actually starting to condense, right?

**KATHLEEN:** Right. Yeah. We're definitely seeing that. So, we've been tracking markets for a number of years and we have seen, like Ron mentioned, this market in particular as, as one of the markets that is contracting.

There's still a lot of companies in this space, but as Ron mentioned, there's a few that in particular that are really acquiring a lot of other companies and we're starting to see, see it shrink. In addition, a few companies did go under somewhat surprisingly to us.

But in general, most of them are getting acquired.

**CRAIG:** Yeah. Within that, ML ops, is, is a subset of that platform and tooling space. It's actually growing from what I can see. I'm constantly getting contacted by people with a new startup. it's from what I can see increasingly crowded and frankly confusing because there's so much overlap.

**RON:** You're right. To be confused by the way, because it is confusing even to those of us that spend all of our time in the market.

Actually, this connects to the previous question, because the idea is like, what does it really take to put a machine learning model of any type into production?

And it turns out there's a lot of pieces. You know, one of those pieces is just dealing with data, getting data prepared. So, there's a whole data preparation space. Then we may talk about this later, which is if you have the data to make certain kinds of machine learning work like supervised learning, you have to tell the computer, what stuff is it?

Can't look at it to know there's a cat in there. Right? You have to say, there's a cat in this image and that's what you're doing, you're labeling the image to say cat. And in many cases doing even more specific things, uh, in the image, if you spent trying to do more sophisticated things, right, then you have the whole, you know, building the model, training the model.

And there was this whole movement towards trying to simplify that because data scientists are, are a little hard to come by. Um, not all organizations can afford them, but also a lot of things we're trying to do, don't really require sort of building everything from scratch and, you know, going right down to the lowest levels of the algorithm.

Some companies pioneered this area of auto ML, which is trying to automate a lot of these aspects and that really has helped. And there's a lot of stuff there. And then once the machine learning model is sort of produced and tested and evaluated, which are whole other sets of steps. You have to get it out there, the model, and you have to also manage it because these machine learning models, they evolve over time and the whole space of ML ops is really about trying to manage that life cycle of the machine learning model once it's kind of out there models, drift, data drift. And so, there's a, there's a bunch of, these vendors who are tackling this problem of managing the machine learning model and operation.

That's the whole ML ops space.

Kathleen, we could talk about some of the vendors in the space and kind of how we see it sorting out as a starting point.

**KATHLEEN:** we do see it growing as well.

There's a, a lot of companies in this space. I think that from a market standpoint, people need to Educate themselves on what particular vendors are really saying. As an analyst firm, we need to cut through the marketing spin and say, okay, what can this product really do?

What does this company really do? And they, as a whole, in this space, don't always do a good job of helping people really understand that. as Ron mentioned, we have covered this space and we have produced a report on it and then our new website has updated continual coverage on it. So rather than producing one large report that can get stale quite quickly, we will continually feed that.

We've found that in these spaces, especially some of these hot spaces with all the acquisition that's going on and with fundraising that's going on, companies that maybe didn't meet our minimum threshold quickly will with these ridiculous rounds of funding that they're getting. And also, companies can get acquired.

We've seen a lot of that happening in this space, even if it was a rather large company.

**CRAIG:** So, the overall tooling and platform space is contracting there's a consolidation

**RON:** consolidation is probably the right word. Contracting, people might think that the market is contracting like spending dollars. That's not what we mean. Right. It's sort of the number of vendors.

**CRAIG:** Number of vendors. That's what I meant, too., but within that, the More granular ML ops space is growing. And I just I'll name a couple of companies that I'm familiar with.

There's a company called Verta, it's built around a versioning tool called ModelDB. and they say that they're an end to end ML ops platform. But when you look at them, they pick up during the build process and then they carry through to deployment and monitoring. But then there's a company co determined to AI.

It's actually now open source. They also say they are end to end. They start after the data prep, largely in, in choosing architectures and hyper parameters, though, to deployment. Are developers using these products, sequentially? Why don't They join forces and bolt there's products together and then have closer to a true end to end solution. Is part of the consolidation that's going on doing just that, taking these pieces, sticking them together, or is there value in having discreet platforms and tools so you can pick and choose from?

**RON:** That is a classic question for all software markets, it's always the question of is the product a feature.

That's what venture capitalists will ask. If you're trying to raise money, like, I love what you're doing, but is this a feature meaning that this is something that will inevitably just be part of something else. Or is this a market? Like as in it'll clearly be on its own for a long time . ML ops is maturing. It's a space that people didn't really talk about it at all maybe three or four years ago. Now we're realizing how important it is. It's sort of like dev ops. We didn't really talk about much the year, 2000, but now with agile and all that sort of stuff and continuous integration,

now, everybody talks about dev ops and it's his own universe of stuff. The thing about ML ops is that some of the vendors have been acquired. So Algorithmia, one of the big players, was just acquired by DataRobot in a pretty substantial deal. they also acquired another company before that called parallel M, which was sort of in that genre.

Kathleen alluded to a company that actually went bust , Dot Science, but those people kind of went off into working with some of the other vendors that are in the spaces like Pachyderm, Tecton, uh, I think we're tracking about 40 companies in this space, 39 40 in the space right now. there's a lot of people who do believe that ML ops is a feature of, of what will be sort of a more complete product, you know, right now a lot of the cloud vendors haven't added too much in the way of ML ops. So, if you look at what's on the AWS SageMaker platform or the Azure platform or Google Cloud or something like that.

In that sort of universe, they're going to probably add that stuff. They're going to probably either add it through acquisition there, they may pick up some of these companies probably most likely Microsoft and is known to be acquisitive Google as well.

It's hard to see that this is going to be necessarily a long-term market. A lot of people believe that ML ops is a philosophy and it's a way of doing things and that these tools can help. So, I would say maybe two or three years from now, we might not be talking about a separate ML ops market category.

**CRAIG:** So, ultimately, we're going to end up with a handful of true end to end platforms where you prep the data, if it's supervised learning, you label the data, you do a neural architecture search, you do a hyper parameter search. You build the model, you run the training iterations, and then you have a compiler that spits it out into the world.

**RON:** In our ML platforms report, we actually talked about there's really five major platforms of how you build a machine learning.

If I said to you, Hey, you need to build a machine learning model and you have to go out there and choose a way to build it. You actually have five discreet, five separate ways. If you want to think of choices, they're not, alternatives. They're just, they're just different ways.

You know, one thing you could do is you can build your whole thing in the cloud.

You could say all my data's in the cloud, it doesn't make sense for me to take the data out of my cloud and bring it to somewhere else and then build a model and push the model back in the cloud. You're saying, well, if I live in the cloud, I should probably build my machine learning model in the cloud.

Those platforms are just going to continue to expand in functionality. It's going to make more sense. If you're on an Azure or AWS person, you're going to probably stay there. And you're going to do those things to the extent that the data's in the cloud, because if the data's not in the cloud, then actually that you can make a solid argument.

Well, why should I put my data in the cloud, especially if I have issues with that, but maybe I've got exabytes of data moving data to the cloud. It's just painful. Right. And then we have all these other issues of, dealing with, with data that might be private and confidential and all that sort of stuff and, and secure.

Number two is you could use an analytics tool.

There are tools that have been around for decades, SAS, SPSS, which is actually now part of IBM, MATLAB, and they're focused on the needs of the data analytics. And if you're a data scientist, who's been living in a data world and you need to build a model and you're used to the environment of analytics.

You're probably going to stick with those tools. So, you might build a model in R, with RPython, whatever inside of one of those environments. Right? So that's actually a second choice.

The third choice is that you might decide that you are living in the open source world.

If you're coming out of like being a Python person and you're used to using scikit, or all these things and you want to continue in that ecosystem.

DataRobot, they kind of build on top of the open source ecosystem and they allow you to sort of stay in that world. You don't have to move your data to a cloud environment. And also, you could be vendor neutral. Different people may be working in different environments. Hybrid environments can use the same tool to build machine learning models and in the same environment, and then push those models into different environments. You could take that model and push it into Azure or AWS . So, it's a third choice.

And then the fourth choice you can make is that instead of using machine learning platform, you can sort of pull them together yourself, out of the, out of the pieces.

A lot of researchers and academicians and if you go to code academy, they're just like, here are the tools you pull it together.

There are reasons for doing each of those things. A lot of it has to do with where the data is, a lot has to do with what your expertise is, a lot of it has to do with where these models are going to end up being used and who's developing them, or you can have citizen data scientists, you can have experts. So, I think we're going to continue to see quite a bit of diversity in product here.

**CRAIG:** I noticed you have published a report on label box, which is a company I'm familiar with. Those guys were on the podcast. Label box is a platform for labeling data. They're not a service company. They're a platform and you can invite labeling teams on the platform and manage them through the platform.

I also had Alex Ratner on , one of the founders of Snorkel, which just raised a lot of money and they do programmatic labeling. And from what I understand, it's kind of brute force, computation that may be is not as precise, but it's good enough, for very large projects. And then I've also had Da he's a former colleague from the New York times, who has a company called the AI Reverie, which does synthetic data.

And then there, there are the BPOs in India. So . how do you see labeling developing? Particularly hand labeled programmatic and synthetic, will they all survive?

Is the market going to continue to grow so that there is space for all of them? Or do you think that it'll go to hand labeled data, and then depending on the size of the project, the rest is programmatic or synthetic.

**RON:** That was another six part question. I'll take some of it.

**KATHLEEN:** Yeah, I can start.

Well, at Cognilytica, we have been covering the data labeling space for many years. In fact, before, most people in the market were actually covering it or really understanding what it was about. So, we have a great pulse on that space. We have seen from the beginning of when we initially produced our report, I think our first report was back in maybe 2018 or 2019.

When we, when we really started covering this space. We've seen a lot of different companies come into the market. So, we cover a wide variety of vendors. I think we have something between 60 or 70 vendors that we're tracking in the data labeling space in particular, what we've seen is that companies from back in the earlier stages to now have started to specialize in what they do.

So, label box in particular has just raised a pretty significant round. And we have seen the growth in this market, tremendous growth in this market. But it's starting to show signs of definite leaders in this space and people and companies that can raise money and companies that are struggling maybe will merge, maybe will get acquired. Maybe will go under in the next 12 to 24 months.

**RON:** So , we divide up the university sort of the second part of your question. There are a bunch of folks who are providing the labor pool to do the labeling. So, for those who are on the podcast are not familiar with this idea of labeling.

The idea is that if you want to train a machine to do something that requires recognition or even conversation things like that, have this little chicken and egg problem. Like the machine doesn't understand what it is. So, it can't just learn on its own. Somebody has to tell the machine kind of what the data is all about so that it can say, oh, you know, here's a thousand pictures of cats.

This is a cat. This is a cat. This is a cat. This is a cat, you just take an image, which is just a bunch of pixels, a machine doesn't understand what it is, just a grid of pixels. And you say this particular grid of pixels is a cat. This particular a grid of pixels is a cat. And actually, if you think about it that way, it's a better way of understanding, because if I gave you a stack of a thousand images of lots of pixel and say all of these with these pixels, they're all cats now learn something, find the pattern, which is basically what supervised learning is.

Find the pattern in those pixels, where if I give you a new image that had a bunch of pixels, you could tell me with what likelihood that image is a cat or not, which is basically what a machine learning system is doing with all that. But it can get more sophisticated in that.

That's just like the whole image. What if I want to say, draw the box around where the cat is inside this image or track the cat as it's moving in this series of frames in this video from one side to the other, or basically determined what the intention is. It's sort of cat jumping or person reaching on a shelf, or, you know, somebody is scratching, you know, now you're starting to get more complicated, right.

Until we, the machines can figure out how to learn from themselves, which is definitely a holy grail of artificial Intel. We are getting better at doing things like requiring less training data, learning from other examples, doing what's called zero shot learning, and one shot learning.

We've actually recorded some podcasts on that subject. Until then you need to have this labor pool. Now, some of the companies in the space that they provide the labor pool and you're right, a lot of them originate from traditional business process outsourcing firms or firms who have been doing other things with data.

So, what you might see as there are a lot of companies in the space that used to do, for example, content moderation for organizations like Facebook, where they have to go in and say, this is not good. You know, bad content, NSFW, whatever it is not acceptable. And you know, as much as Facebook and other organizations say they have an AI algorithm for better, for worse, there's really just armies of people who are, who are doing that. Maybe they're validating what some of the algorithms do or they're not really using an algorithm. And it's really tough work, honestly. And , that work tends to go of course, to areas of low labor costs.

In the very beginning, people used crowdsourcing for basic labeling tasks. Amazon was one of the first to really pioneer, this, they had a thing called the Mechanical Turk, which is a reference to the chess playing bot mechanical Turk, and actually way back in 2005 or six, uh, mechanical Turk came out and you can do these little tasks for pennies, right.

And it started for a very logical reason when Amazon started selling music they had images and they needed to get the stuff into their database. Right. Somebody had to basically type in the name of the album and the name of all the people on it. OCR as good as it was, was just not working on images because first of all, crazy fonts, all sorts of stuff going on.

It was just easier to say. Go on here, get a bunch of CD images and just type in who the artist is and type in the other metadata . And that's actually how it started. They're like, Hey, typical Amazon fashion. They're like, well, if we needed other people, we need it. Why are we spending money on it, when we could be making money on it? Let's make this a service and mechanical Turk for, for many years. And things like it. We're sort of dominant in the, labeling space, but then people realize that it's kind of not so great to rely on random people to do things, especially when you needed certain tasks, like, go in and label all of the street signs in Thailand so that I can train my autonomous vehicle to recognize stop and stuff like that.

Well, you need to know Thai. You need to know what the stop signs look like in Thailand, which may be different and that required more specialized labor pools. Of course, we have issues of confidential data. People are trying to label medical imagery. A lot of that medical imagery has protections and things like that.

Of course, let's not talk about military intelligence data and all this stuff . And that's where these new companies came out. So, organizations like Allegion, which we've been tracking very, very much. They have some very powerful technology for doing video annotation.

Scale, very heavy in the autonomous vehicle space, you can imagine who their clients are.

They also got a pretty substantial deal, uh, with the governments they have some connections. Say the former chief technology officer for the United States, Michael Kratsios now works for scale. It's a fact, I'm just saying it, um, so these companies are growing, somebody is going to go public. As mentioned, figure eights they used to be called CrowdFlower. They got acquired by Appen. But then they spun off their federal arm because Appen itself is based in Australia and you can't get federal contracting money, especially for certain areas, if you're not headquartered in the U S so they've spun out federal figure eight federal, which is here.

**CRAIG:** A side note on Mechanical Turk, Fei-Fei Li used Mechanical Turk to label the original image net data set that Geoff Hinton used to validate deep learning.

And so, the reason we both have podcasts is because of mechanical Turk or at least mechanical Turk played a role,

I wanted to restate the question, you have a view on the hand labeled programmatic or synthetic going forward. What's the future?

**KATHLEEN:** What we have seen in the market is that it depends heavily on the use case as well, but in general, a lot of that easy, low hanging fruit has already been labeled. Synthetic data we've really started to see use cases for this and say, okay , maybe I do want synthetic data for a variety of different reasons, maybe for privacy reasons. Or we had an example with Amazon, with their Amazon Go stores, where they needed to have a variety of different poses from people, people of different heights and children as well.

And they wanted them to stand at a store, put their hand on the top shelf, put their hand on the middle shelf, the bottom, maybe a toddler like, you know, trying to put something in the cart. To have people do that would be incredibly difficult and take a very long time. So, they used synthetic data to help train their systems.

So, for examples like that, that can be incredibly useful. And we're starting to see synthetic data being used more as well with different applications. So, it depends on what exactly you're trying to do with where things are going to go. We can't say, you know, synthetic data is the wave of the future and that's what everybody will be using going forward.

But I think that people are starting to see more use cases for it.

**CRAIG:** And then there's this whole question as to whether or not unsupervised learning will eventually overtake supervised learning and you won't need label data at all.

**RON:** That is certainly one of the, the goals of, of a lot of research. Because if you think about our brains if you put water in a cup and you turned it upside down, the water spills out, you know, maybe that's what you do when you're a toddler.

You're like, Ooh, look at all these cool things I can do. No one needs to teach. Like, well, what would happen if I put water in a shoe? And it turned kind, kind of can guess

unfortunately, machines do need that specificity. They can't generalize the ways that we want them to what we call machine reasoning, which is that a cup is a kind of vessel, right. And a vessel as an opening on one side, and if you tip the vessel over, whatever's on the inside, comes out. We get that in like an instant.

If we could figure out how to do some of these generalized learning things apply knowledge from one domain to another, certainly we're trying then it would lessen the need and the demand for labeling, which is a solution to the problem that we have to still teach it.

We're not going to have this on the short term, like in the next few years. Uh, so in the meantime, if we're trying to accomplish short-term goals with machine learning, we're going to have to do, the labeling. And as I mentioned, there's increasingly specialization now among the firms.

Companies like label box are more of a platform provider , what they're calling labeling ops, right? Allegion is focusing much more on video cases and more sophisticated cases, which will require more specific domain knowledge.

Some focus on natural language processing, , our friends out in India, Desi crew, you know, that's like their thing. It's a nice space. We were not expecting the space to be so hot. What we did is we divided them universe of vendors into what we call this layer cake. It's like, okay, you have all these infrastructure vendors on the bottom. And then you have some enabling technology kind of one level above it, like computer vision and natural language processing.

And then above that, we have these horizontal applications that we can build, like chat bots, which is an application of NLP. And then on top of that, we can do like domain specific stuff, like financial services chat bot, which you build on top of a chat bot, which is built on top of NLP, which is built on top of the data.

And what we decided is like, okay, we're going to start covering the market from the bottom up because it's a little bit of an inverted pyramid. I a lot of it vendors on the top, not as many, not as many, right. The bottom. And actually, we had started with, and we're like, okay, let's start with data prep and data labeling and.

Wow. Okay. That got hot unexpected. We're like, wait, we haven't even gotten to the top yet. Okay. All right.

**CRAIG:** One of the questions I have intellectually is what happens. So, all of this labeled data, it doesn't go away. It's, it's stored somewhere, and it's piling up as supervised learning makes its way through the economy in silos because it's all privately held.

But at some point, as algorithms become more general, and can feed across those silos, it seems that that that will provide some, food for an eventual, I don't want to say AGI, but a more general AI system that can cross, domains. And, different kinds of data. Have you thought at all about what becomes of all that label data?

I mean, you're basically encoding human knowledge in the data and that's valuable .

**KATHLEEN:** Data in general is very valuable and companies are holding that close, so they will not share that. And that is partially what is fueling all of this. company A needs to label this set of data company B needs to label a very similar set and company C needs to do the same thing.

They will not share that because that is their secret sauce, whatever helping to differentiate them. And then as far as maybe one day, if we do share that, without machine reasoning, we really can't. Move to that next level that we need. We always, we talk about the DIKUW pyramid a lot. So, data is at the bottom and information, knowledge, understanding and wisdom. We can't get to that understanding level, let alone the wisdom level. So, the wisdom is really, you know, a human level, which if we're talking about AGI would be at that W level, we can't get to that K level until we get machine reasoning.

So, we're very far away. I know when we talk to experts in the, in the field, you can get a wide range of responses to that question. Some people think we're a decade away, maybe 50 years, 250, maybe we'll never get there. And right now, we think until we get to that machine reasoning, we will not get there.

**CRAIG:** Although what I'm saying is you don't have to get to, to reasoning. Right now, an AI algorithm cannot operate on natural language data, and then operate on pixelated images, right.

They are two different domains. But eventually one would hope you could develop algorithms that could move from one kind of data to another. And eventually, too, I would think these silos of data will be open because after 10 or 20 years they're not as valuable as they are right now to the company.

What happens, do people start dumping their data in this massive repository and then researchers can have their time with it or what?

**RON:** Well, a couple of things, first of all, if you think about sort of like how your brain works or any animal brain, or even an insect brain, it doesn't have one locus.

It actually has multiple areas. Like you have a part of your brain that processes speech, you have another part of your brain that processes image data, and . You have another part of your brain that handles moving. And then of course the more sophisticated things like, you know, planning and all that sort of stuff.

And we know that because we've had we've, you have people that have actually brain injuries where a part of that brain may be disabled. They've lost the ability to say speak, but they can see, or they've lost. They've lost the ability to see or process or something like that. But you'll see people have strokes.

So, who, some people like don't like acts and changes, like there's this weird sort of phenomenon develop a strange accent. It's unusual. Right. But it shows that there's specialization. Right. We might think. Okay, well that would, that might imply that maybe we're not going to have a single algorithm. Maybe it will be.

Algorithms that are really optimized for different things, but we do need sort of like the brain. So, if the brain is not a single algorithm, maybe AI is not a single algorithm. Maybe it's something that coordinates between the different centers of learning and knowledge and somehow synthesize it.

That might be machine reasoning. Machine reasoning might not actually be an algorithm. Machine reasoning might be a way of coordinating between multiple algorithms and knowing how, which is the hard part, knowing it's like, oh, I'm seeing something. I'm hearing something, these two things meet, relate to each other.

You just bit into an apple. And I can know what that is. That's what the reasoning is. I know what that is. Right. Or some, some other synthesis right.

**CRAIG:** My last question is about, auto ML.

You mentioned a couple of times you have reports on AutoML. There are emerging these standalone, no code auto ML platforms. There's one. I'm familiar with, I'm actually trying to find somebody with horse racing data, can validate it with me to see if we can predict an outcome in a horse race.

I don't code. And although I'm trying to learn I'm in my sixties, I don't have any real expectations to be able to build things, but just as, windows, abstracted from, dos commands, these no code platforms are increasingly abstracting from the code base.

do you have a view of how that's going to develop? To me, that's pretty exciting. There's a very finite number of people in the world who can code but everybody could use AI or, 80% of adults could use AI if it were simple enough.

**KATHLEEN:** So, I think kind of two parts to this question. So first with auto ML and how we've seen the rise of this. Auto ML has allowed individuals at organizations who may be very knowledgeable, take a data scientist, for example, to just do their job better, where they don't need to be doing everything manually anymore.

They need to have some understanding of what is going on you do need to interpret the results that come out but I think that it's allowing people to do their jobs quicker. Better, maybe you need less people because of that. And then taking that one step farther with low-code and no-code tools. we have seen that companies are starting to deploy that for various use cases, and it's allowing people that aren't as technical.

So maybe they're not the data scientists, they're a business analyst or a sales rep or somebody in a specific role. That's not very technical, a customer service agent to go in and use some of that for their benefit. So, they can go in, they don't need to be technical and they can go in with, you know, low code, do drag and drop and get the results that they need.

So, both of them are really shifting the market, different users for each.

**RON:** It's certainly one of the goals of, of some of these platforms to democratize and to put auto ML in the hands of people who are not data scientists, not data engineers. and you know, Hey, if you talk to Microsoft, even they'll tell you where's Excel going, you know, there's their whole power BI platform there.

Process automation into the whole platform. And if you could basically just drag a column down and say predict, right? So, you could say horse races, right? So here are all the results. And then you have the name of the horse here.

And there are times or something and then their results or whatever other, you have to know what the factors are that have predictive value. That's where you get stuck. Because even if I can give you the tools, I can't say, I don't know what actually predicts a win.

Actually, if you knew that you probably would not be talking to me on this.

**CRAIG:** Well, except that's, that's the power of ML is that the more data you throw at it, the more accurate predictions and who knows what in the data is. , creating those predictions.

**RON:** Certainly, that's a vision. They would love for you to be able to have that spreadsheet and basically say, okay, predict the winner. I'll put an extra row here and then the value for win or . Loss, maybe the probability of win you predict that value based on everything. I don't think we're many years away from that we may actually be a couple of years away from that.

**CRAIG:** Yeah, well actually the platform that I'm fooling around with is called Akkio. And I've got a guy who, who built a predictive model for the racing form. This is just a fun project. not that I expect to make money in the horse race. It's more to see the, the power , of these platforms, this particular platform.

The biggest use case from what I can tell are sales teams that have an overwhelming number of leads and don't have time to chase them all down and need a way of ranking . So, my thought was, well, Hey, if you could do it with sales leads, why couldn't you do it with horse it's so we'll see.

But, it doesn't seem very far from the day when anybody who has a bunch of data can go to a no-code platform do something recognize a pattern or make a prediction.

**RON:** It's good for tabular numeric data. I remember being very confused about auto ML because I'm like, oh, I can just drag a bunch of images in there, but not exactly. Now turns out that Google has and Microsoft, and I think Amazon now, but Google had early this cloud vision auto ML thing where they had already pre train the network on just being able to recognize things.

Right. But let's say you want to build a cloud identifier, there's actually the example they get, where you take a picture of the sky and I'll tell you what kind of cloud it is. Physical clouds, not the internet. And it'll say cumulonimbus or whatever, they actually show that even though the model has not been trained specifically to do cloud recognition, you can you just put in like a very small number of images?

I think it was something like 30 or 40. And then the system does its little, uh, transfer learning thing. And now you have a very specific model for it. And actually, in one of our classes, I have to note that, one of the things that we do, one of the big things that we do for Cognilytica is we spend time training organizations, surprisingly large ones.

You might be thinking that they know what they're doing. The answer is they usually don't on how to run AI and machine learning projects. And usually they get tripped up because they're so enamored with the building, the model part that they sort of forget about the needing to collect and clean data, which is the not fun part.

Sometimes they didn't even know what data they need to clean. They don't know what access to have to it. And they're doing things in the wrong order and they wasted all this time and money. And there's a methodology by the way out there that optimizes for this. It's something called CPM AI, the cognitive project management for AI methodology, which is itself based on a 20 year old methodology called crisp DM.

CRISP-DM is old. It was good though, but you know, needed to be advanced. So, we advanced it understood as well. And, um, one of the things it teaches you is to basically think big, start small and iterate often. And this is actually one of those places that AutoML can help, which is that if you know what the problem is, you start small enough.

You could start with a small dataset. Maybe you're trying to build some insurance application that's trying to categorize the damage to roofs. And then you'd be like, oh, I need like a million images. And like, whoa, let's slow down. First of all, can you even recognize what the roof looks like? Start with AutoML.

Prove that you can do that. Or once that may be the thing, there's a bunch of things you need to do in between. But that methodology, which is like learning how to do things. You go through 20 hours of instruction. It's like 2,500 bucks. It's like, you know, you wait, trust me, spend more money doing it wrong than, than learning more money, more time.

A little plug there. If you go to courses.cognilytica.com, you can go through that training yourself self-paced and all that,

**CRAIG:** maybe I'll do it. Yeah.

**RON:** We'd love to have you and all your listeners.

**CRAIG:** Yeah. Yeah. Okay. This has been fascinating for me. You guys are very eloquent.

**KATHLEEN:** Well, thanks so much, Craig, for having us on, we really enjoyed this podcast and, we encourage your listeners to check out the AI today podcast that Ron and I are hosts of and in particular, the interview that we'll be doing with Craig.

**RON:** So, thank you all very much for, for having us on. We really do appreciate it. Fantastic conversation.

**CRAIG:** That’s it for this week’s podcast. I want to thank Kathleen and Ron for their time. As usual, you can find a transcript of this episode on our website, eye-on.ai. Check our Cognilytica at [www.cognilytica.com](http://www.cognilytica.com) and be sure to subscribe to the AI Today podcast, available on all the major podcast platforms.

Remember, the singularity may not be near, but AI is about to change your world. So, pay attention.