



## The Deloitte On Cloud Podcast

**Mike Kavis, Managing Director, Chief Cloud Architect, Deloitte Consulting LLP**  
**Gary Arora, Chief Architect of Cloud and AI Solutions, Deloitte Consulting LLP**

**Title:** 2025 tech trends with Mike Kavis and Gary Arora: Agentic AI, governance, sustainability, and talent

**Description:** In this episode, Mike Kavis and Gary Arora discuss what they believe will be 2025's top tech trends, including the rise of agentic AI, the benefits and complexity of sustainable technology, and the growing demand for upskilling talent. They also explore how organizations are tackling governance and standards policies around the growing integration of AI, while leveraging new AI advancements to drive innovation and growth in an increasingly competitive, global environment.

**Duration:** 00:22:44

**Mike Kavis:**

Welcome back to the On Cloud Podcast, where we get real about cloud technology. I'm your host, Mike Kavis, chief architect at Deloitte. I'm excited to introduce a good friend of mine and new co-host, Gary Arora, who's also a chief architect of cloud and AI solutions. And today, we're going to talk about some of the trends and innovations that we're seeing, and I'm going to lean on Gary to give me his insights there, but first, Gary, welcome to the show. Excited to have you as a co-host. So, why don't you just tell everyone a little bit about your background and what your areas of focus are and the types of things we can expect to be talking about.

**Gary Arora:**

Thank you, Mike. Always a pleasure to be here. I'm Gary Arora. I'm a chief architect building cloud and AI solutions for organizations in the financial services and health care domains. And I'm really excited about this episode because we're going to be doing some crystal ball gazing into the technology landscape of 2025 and beyond. So, can't wait to get started.

**Mike Kavis:**

And I've done a lot of these over the years, but for some reason this one feels a lot more exciting than some of the other, maybe because AI so hot and data centers are hot again. There's just so much stuff happening at such a fast pace that we're just going to try to capture a few of them and touch on those. And I know you had done one looking back at last year, as well. So, now we're going to look forward and tell the people what trends we're seeing.

Everyone's talking AI. There's a lot of proof of concepts out there, not a ton of production implementations, yet. I think we're going to start seeing those next year, but one of the hot topics that our listeners may or may not be familiar with is the concept of an agentic AI. So, before we get in and talk about what we expect to see there, why don't you explain to our audience exactly what agentic AI is and why is this kind of so important and the next big thing within AI that everyone's looking at?

**Gary Arora:**

I really think we are in the golden age of AI, especially looking at 2023 and even 2024, and the number one trend on everyone's list for 2025 and beyond is agentic AI. It's one of the most hyped AI technologies at present and for all the right reasons. And your traditional AI takes orders and does a task. So, whether that's predicting the next best action or probability of something happening or not. Those are things you would do with your traditional machine learning models. Agentic AI is different in that it is figuring out what the right action is based on its inherent planning and reasoning ability, and then it's able

to take action on user's behalf. So, the mental sequence of agentic AI is to first understand the ask, think what should be the right steps to meet that ask, and then act on it. So, this understanding and thinking part, these are the novel attributes of agentic AI, and this is what makes it so exciting.

**Mike Kavis:**

That's really exciting, and if you couple that every company has its own processes, its own controls, its own policies, and the ability to put that information in the model because not everyone's workflow is going to be the same and not everyone's thought process the same. So, how does that customization of your cultures, rules, policies, and tendencies, how does that play a part into layering on here with the agentic AI capability?

**Gary Arora:**

That's a major area of research right now. So, agentic AI is able to make complex, layered decisions. So, we call this the OODA reasoning, which is O-O-D-A. So, Observe, Orient, Decide, and then Act reasoning. And so it's really powerful for our complex dynamic multi-variable scenarios that require real-time decisioning because it can breakdown the task and adapt on the fly. Some popular use cases here is, one is self-driving vehicles that can assess and respond to road conditions moment by moment, and it's happening in health care where I work a lot. We are using agentic AI agents that can serve as assistants in health care setting by engaging with the patient daily to support mental health care treatments and things like that. So, there are a lot of these very niche, very tailored-to-the-organization kinds of use cases that are now getting unlocked.

And I think that's the good news, but there is bad news here in that agentic AI is still very complex. I was reading an article on CIO.com that said about 75% of the firms that build aspirational agentic architectures on their own will fail, and that enterprises will face significant hurdles to their agentic ambitions in 2025 because the challenge here, as you just mentioned, is that these architectures are congregated, and they require diverse and multiple models, understanding all the data that exist within an organization, and sometimes the access and the metadata, the labeling of the data is not right. And that makes the training of the agents and the models incredibly difficult.<sup>i</sup>

**Mike Kavis:**

That's the data again, and I've been through several decades of technology change and always there's this data piece. If your data isn't any good, once you get out the other end, regardless of the tech isn't going to be that good. So, remember when big data became a big key in MDM and then we got SOA (service-oriented architecture). And again if your data is no good, we got the cloud. If your data is no good, here we are again. It's like, I've been dealing with this for 40 years. So, how do we finally get the data right or better?

**Gary Arora:**

I think that could become its own episode; how do we get the data right? One of the issues with the data is that the generation of data itself is exponential. It said that about 90% of the world's data was generated in the last two years alone.<sup>ii</sup> So, this is something I was reading a couple of months ago. So, that means 2022 and 2023 that generated more data than any time before. And if you look at not just organizations, but if you look at overall on the Internet, and you break this data down into different categories, video, for example, is half of all global data traffic, where all your social media platforms, all your news platforms are using video as a way to convey their messages. And this is followed by social and gaming as the three categories that make up 76% of all the data that is available on the Internet.<sup>iii</sup>

Within organizations, we are seeing a mix of legacy data from decades ago and then the new data, and they're often in different data sets, often don't have synergies, and the thing that is missing the most is metadata. Do you even understand what data is where and what are the qualities of the data? What is the data lineage? So, this is going to require a data platforming, data engineering work, but you can also just invest in metadata and labeling and identifying the data sets, and use that as a basis of training very specific use cases that these models can furnish.

**Mike Kavis:**

Maybe we can leverage AI to help us with that task because that's a task that's always been really, really hard for organizations. We can get to classifying data, but actually tagging it all and getting it right, it's always been the challenge and I think that's where a lot of data projects die. Maybe this is another opportunity for AI to help there. Thoughts?

**Gary Arora:**

Absolutely, especially with LLMs that can now understand code, including COBOL, and understand datasets. You can just use AI to solve its own problem and take this in sequence, where first the AI needs to figure out the right dataset, label the right dataset, and then the next step would be to get trained on the right dataset.

**Mike Kavis:**

So, speaking of training that kind of leads to our next topic. And again, as we've gone through the decades and gone through the new buzzwords and new technologies, skills shortage always comes up, but this time, it might be a little different because of the pace at which things are moving. So, when it comes to AI and all the things within AI and the machine learning and the amount of compute it takes, and all that, there's a skill shortage out there, and how are your clients looking at this addressing it? And then once we answer that, I'm going to ask you because one of your many responsibilities is on that topic. And I'll let you introduce what you're doing there and talk about how we at Deloitte are addressing that, but first what do you see when you look out at your clients and your peers out in the field? How are they dealing with a shortage in all of this emerging tech that's coming out?

**Gary Arora:**

Absolutely. We are seeing a significant shortage of talent especially in data sciences and AI skill sets, and this is substantiated by so many industry and workforce reports out there. In fact, I'd extend that and conclude that the problem most companies are going to face in executing their digital transformation is not going to be the access to the emerging technology, but it's actually going to be the shortage of workers with the right skills to execute those technologies. And what we have seen is that organizations with more advanced upskilling programs in house, they report significant innovation capabilities.

So, this is a clear difference we are going to see in mature organizations that invest in learning and development platforms, and as you mentioned, one of my other responsibilities is that of the dean at the Deloitte Cloud Institute. This is our in-house training program where everyone who is in our Cloud Engineering practice goes through to get the latest skills and training on cloud technologies. So, they are hands on in application development, setting up the infrastructure, understanding what are the services, when to use those services, when not to use specific services, and only then they get deployed on these engagements to build the cloud platforms. We are seeing a lot of companies starting to do this in house and also leveraging third-party L&D partners to make sure their workforce is ready for the kind of transformation that we're going to be seeing next with AI and GenAI.

**Mike Kavis:**

And I think the point there is it's a combination of the normal third-party training platforms along with custom training platforms to do things the way that your company wants to learn and wants to do. So, talk a little bit about that. I mean, there's so many training platforms out there and people can get lost trying to find the right stuff to do. I think what we do is we kind of bring them the best training platforms, but kind of put a border around here is the type of classes that we think you should be looking at, and then here are the custom ones for our culture that makes sense to learn things the way we want to learn. So, talk about how our institute does that.

**Gary Arora:**

There are a couple of approaches in closing this digital skill gaps. The one where we are seeing the most success is the concept of digital academies. So, this is a combination of bringing the right training and curriculum and topics. So, that's one, but the second part here, and this is more important, is making sure there is a personalization and tailoring of that content to that specific company's technology landscape and culture and existing skill set because this is the part that sometimes gets overlooked and many companies, on paper, may say they have an L&D program where anyone can go and learn in data science and cloud skills, but if the skills that they are learning are completely different than the kind of tooling they have in house, then they won't be able to apply those skills.

So, it's really important that any kind of training program has to be seeped into the company's culture so that, whether it's a financial organization that has high amount of regulations, you need to understand how encryption and tokenization works in your context, how do you classify data, how do you expose data to public cloud, or how do you keep the data secure. And all of this differs based on the specific industry, say, healthcare, or finance, or government, and then based on the context of the solution that you're developing.

So, this bridging of the right topics to learn, but then the right way to apply them to your company's context, that is really the secret sauce here in making sure you have a very effective training program. And that's one of the things we are doing with the Deloitte Cloud Institute. We are taking the success that we have had internally to organizations and our clients, and we are able to launch this program for those clients so that they can get their workforce upskilled on the same kind of pathways that we have done.

**Mike Kavis:**

We thank you for your work in that area. I've been pretty impressed with our education system internally and the amount of time and effort we put to make even nontechnical people tech savvy. It's pretty important. So, moving on to the next topic, it's not the sexiest word in the world, but it's governance, and a lot of my clients are dipping their feet in the water and AI, but they're worried about things like the ethics of AI and lack of governance standards and talking to you previous to this recording, you're seeing some trends in that area. So, what are those trends around AI and governance and some regulations that we can assume to expect in the near future?

**Gary Arora:**

So, you're right that it's perhaps not the coolest topic when anyone thinks of what's next in AI, but it is so important, and 2025 is going to be a really crucial year for AI governance because the European Union AI Act, which was introduced in August 2024, the obligations are set to apply in February 2025, and this is going to have major implications. Now, the Act itself is very broad and it covers, I think, 500 some pages, but there are three major areas, and they are really important from the perspective of data privacy and human manipulation point of view. So, this law would ban, for example, AI systems that monitor employee or student emotions or conduct social scoring or engage in predictive policing based on a person's profile or characteristics.

So, these are some of the genuine risks with AI that we have seen, and this act is really designed to curb some of those possibilities. So, it's also going to prohibit AI systems that manipulate human behavior or exploit people's vulnerabilities. So, it's going to have three different categories. This is the one I just mentioned is the ban use of AI. The second one is going to be around the obligations for high-risk AI systems. So, this is making sure the AI tools that we are creating, it's not going to have a harm on health, safety, human rights, environment, and things like that. And at least for the European residents, they will have a right to submit complaints about high-risk AI systems, and they have to receive explanations about those decisions. So, this is also going to encourage transparency and that's one of the other areas that can lack in some of the AI-based decisioning. And this act would also bring out the development of those systems.

So, we are going to see a lot of important, and perhaps much needed, kind of regulations in this space, and for us in the architecture community what this really means is our design and architecture will also get influenced by these regulations that are coming in if we want our AI systems to be globally available—especially in Europe. Europe is the first one that has started this, but we are seeing similar kind of acts coming through in the US and other markets, as well.

**Mike Kavis:**

I've been hearing the same. I think, it's the same way with all new technology. I remember when cloud first came out. Again, one of the biggest barriers to entry was lack of standards, lack of governance, fear of data out in the wild, stuff like this. It's the same here. I think what's different here, though, is the velocity at which technologies moving and the hype and the amount of companies moving toward AI, that we need to close the gap on governance and standards, and really fast before it just gets really out of control. Are there any standards that are being pushed to the providers of LLMs? What we just talked about was a lot on the output, but anything on their creation or the visibility in what they're creating, is there anything being pushed toward the makers of the LLMs?

**Gary Arora:**

Absolutely. And you're right that the reason that AI governance has been a complex topic is because of the rapid innovation in this space and the absence of universal template standards or certifications. So, one of the ways through which some of these LLM providers and AI providers are getting influenced is the push for AI watermarking and authentication standards. So, think about your deep fakes, your artificial or manipulated images, audio, and video. I mean, we could be recording this podcast in the voice of Morgan Freeman, some of the standards that are getting pushed now will require that to be clearly labeled that this is generated by AI.

So, those are the kind of things we're going to start seeing. So, if you watch it commercial on TV or if you come across a poster on the Web advertising a service, it will have watermarks that will specify this was generated by AI. Even today, if you do online shopping and you're looking at the pictures of a product, let's say you want to buy a toaster, you go on your favorite e-commerce site and you're looking at the pictures, it's a good chance all of those pictures that the toaster is sitting on a nice kitchen shelf with the bread coming out, it's AI generated. And you just don't know if this is exactly what you're going to get when it gets delivered.

**Mike Kavis:**

That's an interesting topic. It's hard to tell what's real and what's not anymore. And we're just going to be dealing with that from this point forward. Pretty amazing stuff. So, we talked a lot about AI and which leads to the next topic. We're going to talk a little bit about green technology, but green tech and carbon footprint and all that stuff has been a topic for a very long time, but I think again AI is accelerating some of the discovery in this area, and a lot of that's because the amount of compute required from these LLMs is like nothing we've ever seen before. So, you already have all this compute to do your normal cloud stuff, and now we're doing LLMs and it's requiring more, which is creating a huge amount of carbon out there. And we talked a little bit before the show, you had some metrics on what we're seeing there. So, talk a little bit about what are some of the new challenges for being green when you throw on all the compute the AI is bringing to us.

**Gary Arora:**

Absolutely. And this one I am personally excited about. Because the topic of sustainable technology is of interest to me, and I've been following it for a couple of years. And it's only going to move up, which to me indicates that organizations are realizing its value and are also realizing that organizations that invest in sustainable tech could unlock the potential to increase operational resilience and financial performance. So, it's not just goodwill here, it's also impacting your bottom line. And I think that's very important. There was a paper published by University of Massachusetts earlier this year that concluded that training a single AI model that uses a transformer or a neural architecture can emit as much carbon as five cars in their lifetimes.<sup>iv</sup>

So, the advancements in AI and machine learning have resulted in a very increased demand for the technologies and that's also resulting in increased carbon footprint. So, reducing the carbon footprint is at the heart of all these approaches that we are seeing in the sustainable technology space because we know that training and running AI systems, they just demand an enormous amount of computing power. So, there are a couple of things that data centers are doing. There are a couple of things that businesses are doing. From a data center perspective, we are seeing a consolidation of servers retiring your zombie server. So, when we talk about technology, this is where we talk about if you're running containers or if you're running Kubernetes, you're able to club multiple applications onto a single server. And this is a stark difference from running an application per server.

So, by consolidating and reducing the number of servers that you're running, you're already reducing your carbon footprint. But there are additional things that are being considered in terms of how the data centers are designed, how the cooling is coming in, and you only leave the cooling to minimal temperatures. What about the automation for lights and security, moving to sustainable suppliers? So, there are a number of things that are being approached and tried by the data centers, but also by the organizations that are using these data centers for their compute power.

**Mike Kavis:**

Well, that's our episode for today and thanks for looking ahead into 2025 with us, and I look forward to being a co-host with you as we progress in the future to all those listening out there. We covered a lot of space here. If there's any topic that we covered, you want to go deeper on, let us know in the comments. We're happy to do that or if there are other topics that we didn't cover, just let us know; we're happy to cover anything that you, the listeners, want to talk about. So, again that's it for our show today. Make sure to like, leave a review, and subscribe. You can also explore many of our previous posts on the tools that you listen to your favorite podcasts on. Feel free to send any questions on topics to us. You can find me at [mikekavis@deloitte.com](mailto:mikekavis@deloitte.com). Gary, what's your e-mail? I'm assuming it's similar.

**Gary Arora:**

Yes, first name last name @deloitte.com.

**Mike Kavis:**

OK. So you know where to find us. Appreciate your listening. We'll see you next time on the Cloud Podcast. Thank you.

**Operator:**

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to [Deloitte.com/about](https://www.deloitte.com/about).

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Visit the On Cloud library  
[www.deloitte.com/us/cloud-podcast](http://www.deloitte.com/us/cloud-podcast)

#### About Deloitte

---

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more about our global network of member firms. Copyright © 2025 Deloitte Development LLC. All rights reserved.

---

<sup>i</sup> <https://www.cio.com/article/3583638/companies-to-shift-ai-goals-in-2025-with-setbacks-inevitable-forrester-predicts.html>

<sup>ii</sup> <https://www.statista.com/statistics/871513/worldwide-data-created/>

<sup>iii</sup> Chart: Video Drives Surge in Mobile Data Traffic | Statista

<sup>iv</sup> <https://www.supermicro.com/en/article/ai-training-5-tips-reduce-environmental-impact#:~:text=They%20found%20that%20the%20process,manufacturing%20of%20the%20car%20itself>