Achieving Alpha: AI Driven Transformation | Transcript

[00:00 –00:31]

Daniel Angelucci: Hello, everybody. My name's Dan Angelucci. I'm a managing director for Alvarez and Marsal. I lead the digital and technology services organization in Southeast Asia and Australia.
Today I'm talking with one of my good friends, JP Morgenthal, and our plan is to actually discuss AI.

And, while this is a topic that I think everybody has addressed, I think what we would like to do is take a different look at how it is that you can actually enable this in your organization and what some of the challenges are to get you there.

So, but before we get started, JP, please take the time to introduce yourself.

[00:32-01:52]

JP Morgenthal: Thank you and good evening over there.

It's morning here in the U.S. I'm based in the United States in Florida. JP Morgenthal, I am the CEO of a company called **NexAgent Solutions.**

We help companies implement Agentic solutions, right now based on real pragmatic understanding of what this technology does, but also help them with the adoption and overcoming hurdles in making this production level as systems within the U.S.

Prior to that, my history is long and sorted, a serial CTO, entrepreneur. I have, you know, worked for large corporations—expertise in managed services, retail management, distribution, logistics, financial services. And I would say, my, to sum it up real easily, my jam so to speak, the thing I have always done is jump on new technologies, as they're emerging to understand them and help enterprises understand what the state-of-the-art, what the, you know, what is the art of the possible is and also how to adopt and bring these technologies in a, you know, governed manner within the enterprise.

[01:53- 01:56]

Daniel Angelucci: Excellent serial CTOs. No cure for that is there.

[01:57 – 02:21]

JP Morgenthal: There is not you, you know. You and I both suffer from that. Although, you're doing much better than I am. It seems like it. It's a very difficult role to escape because it's so broad and everything else is a box that everyone expects you to be able to fit in. And the truth of the matter is, you know, the skill set just expands outside of the box very quickly.

[02:22]

Daniel Angelucci:100 percent.

[02:23-02:53]

JP Morgenthal: It's marketing, it's technology understanding. For me it's even hands on, right, these technologies I am hands on experienced with. So, you know, having somebody that can actually implement your, you know, a genetic solution, but also explain to the CEO or CIO or COO, how to what, what the right agent design is so that they can improve performance of their organisation is a is a unique capability.

[02:54- 04:25]

Daniel Angelucci: Excellent.

Well, I think one of the things, I wanted to kind of start off with, was a bit of a high-level view, with respect to artificial intelligence. What's been different about it in terms of, as the technology has emerged over the last couple of years. And I think, I think, one of the really big differences is that, you know, sort of the exciting part of some of the technologies that have really emerged prior to AI, very much focused on what the technology team liked, things like cloud or even some of the data and analytics capabilities that preceded AI.
The AI is different. The way that it's sort of emerged vis-à-vis ChatGPT or whatever, has really been about the part of the iceberg that's actually above the ocean. And it's been very interesting in that—it's gotten everybody very, very excited because that part that's above the ocean, that part of the iceberg is extremely interesting and has seemingly endless possibilities.
Unfortunately, I think, what we sort of sometimes neglect is that as technologists, at least, we're still responsible for what it is that goes on underneath the ocean. And that piece of it has been, some, I've had some real challenges in terms of discussions with customers

around how it is that we can do that, how it is that we can actually sort of address the piece that's below the water line with respect to AI.

And I think, you know, the analogy that I was thinking in my head this evening, actually was about, you know sort of when the Internet first came out. I'm old enough to remember when that started right. And you know, the idea was–oh gosh! look, the Internet is Netscape, isn't it? Because I can go run Netscape 2.0 and go to whatever.

[04:25 - 04:26]

JP Morgenthal: It's a good analogy.

Daniel Angelucci: Yeah, it really is.

[04:27-04:43]

JP Morgenthal: …because it hits that above the iceberg. It also was above the iceberg introduction to the world, right? It wasn't, I don't need IT in order to do and own this for me, right! This is something I can go home and dial up my computer and get on and do myself.

[04:44-00: 06:10]

Daniel Angelucci: Yeah**.** Except for the fact that it involved, all of these, you know millions of dollars. And you know, tens of years of technology development that have been going on in universities to make that happen, right. And all the plumbing was still there.

And I mean, I remember learning IP from scratch, right, because that was what you did back then. And that really enables the Internet. And this is very difficult to explain to people who are just using browsers.

And I think, often, we have that same sort of challenge with AI—because it's very difficult to explain to people with the possibilities or the interesting part of AI is because we also have to bring in all these things about how you actually get it done, and that's been the trick.

So, like, just to provide a little bit of structure in the conversation, what I thought, we would try to address, two things, really, like, what are some of the hurdles that are there in order to make AI possible— like what are really sort of the core things that need to be in place in order to actually just even get started or get started well with respect to AI.

But the second piece, and I think, you know, probably even more interesting is, what it is that we can do to make artificial intelligence really valuable for companies. So, it's not just something that they can say in their brochure that they've done, that they have AI enabled services or whatever, but it's actually something that they can monetize, really drive value for their clients and their customers across the board. So that's sort of how I wanted to start the discussion.

And I guess, like, why don't I go ahead, and you know, sort of, just take–think of a little bit about what it is that we can do to make AI possible. And think, the number one piece is….

…Go ahead.

[06:10 – 06: 12]

Daniel Angelucci: Yeah, sure, go.

[06:13 – 08: 09]

JP Morgenthal: It is really because people will just call it AI, and AI is a huge category in and of itself. I mean, I always joke with people that—you know, AI—I've been doing AI since the early 80s. One of the people you forget, or in our age, I don't remember...the book clubs.

The Columbia Book Club, Columbia Tape of the Month Club, right? You get like 50 upfront, but you have to buy like six more throughout the year, right? And so, they had a book club like that for, and you know, a big part of it was a technology section. And I had this series—I actually just got rid of it recently—but it was 4 volumes, each about that thick on artificial intelligence.

And it was all the algorithms. I mean, you had vision, you had natural language processing, you had knowledge base management—you know, it's a very comprehensive space. Neural networks.

This actually, what people call AI now is a very specific subset—that is generative AI. It's based on the ability to translate natural language into a….and generate from a natural language what we equate with an understandable or a human response based on context.

Very, very narrow in actual scope. What it does is huge, but really a very small subset of the entire arsenal that is artificial intelligence. And people should understand that right. We call this, I like to say, we're talking generative AI right now, because that's —it's a very specific technology subset.

[08: 10 – 09:24]

Daniel Angelucci: That's a fair point, right. So, I think, if we specifically narrow the scope down to generative AI and talk about what it is that makes that possible, right? Like, I would argue that the number one thing is going to be data, right? And most organizations that I encounter haven't even begun to address this comprehensively.

I remember actually having this discussion about data and analytics, just more broadly, four or five years ago, saying—like—every data project looks about the same, right?

There's this notion of visibility. There's this notion of insight. And there's this notion of automation. So, first, you get as much of the data as you possibly can so that you can actually sort of say, "I have a comprehensive data set." And then usually you start off with some very preliminary kind of conclusions in terms of the analysis of that data - "Here are some things that we can do." And then eventually, you automate that. And that's where AI and generative AI have really started to live.

The secret to this—or I guess, it's not really a secret—the unfortunate truth about this, I believe, is that there are simply no shortcuts.

Everybody that we bump into, a lot of times like clients have these discussions, and they just want to start off with, you know, at best, insight—often automation—without really discussing visibility in the data. Has that been your experience as well?

[09:25-10:57]

JP Morgenthal: I would say there's one shortcut; I wouldn't call it a shortcut. I would say that in the past, to do the steps you've described and outlined to get to the point where you can gain insight, you had to undergo an extensive data architecture project. You had to do master data management. You had to organize the data in a structured approach.

I would say one of the major changes that I've seen with regard to this, that is good, is that this technology tends to lend itself towards unstructured, operating on unstructured content.

You could throw anything at it—I mean literally anything at it—it's just a stream of content. And it will decipher and incorporate it into its space in a way that, you know, it's nothing more than relationships**.** It's like, "Oh, here's a stream of data. In this stream, X is close to Y, which is close to Z. I'll take that in. I'll remember that—that's in my memory now—that I continually saw that these three pieces of information are close to each other."

Great. I didn't have to go through all that, you know, data organization—structural data organization. But I think the other piece of it that you alluded to, is still needed, which is—you know—garbage in, garbage out. I still have a data quality issue that I need to start with.

[10:58-11:06]

Daniel Angelucci: It's interesting—how is it that you would solve an unstructured data quality problem? And that's not the simplest thing to manage, is it?

[11:07-13:20]

JP Morgenthal: No. And this is where Agentic AI is really interesting, right? Because I can start to organize my processing, you know, I can use the technology itself to teach the technology.

So how would I do data quality? I would create an agent whose whole thing is to focus on the quality of the data. So, I would start by—that's the one where I would put my focus and my education. That's where I would, you know, create the fine-tuning. I'd really focus on that one. I'd be like– "This is appropriate. This is inappropriate." right?
And so I'd look at my data, I'd pull out a lot of samples. I would create a fine-tuning training package that says, "I've gone through N number of samples"—and it's gotta be large, you know, a fairly large sample of the data—and I'd say, "Here are examples. This is good. This is bad."

Now you go and you see what it does. And I'm going through that right now with something really simple and stupid. And I'm, three days, I'm banging my head against the wall trying to train this lightweight LLM. There’s a large language model, and there’s trade-offs—about fine-tuning these things versus speed, how many tokens it actually is using, and how big that token space is versus small token spaces and trying to train them, right?

I can get really quick results, but bad classification.

Or I can give up, you know, another 30 seconds to a minute to let it think about it and look it up and evaluate it, and I get better classification. And of course, then there are trade-offs of hardware—I can throw more hardware at that to bring the length of time down.

So, it, right now is kind of a “pull the lever and see what happens” kind of activity. Which I don't get where you’re getting all these people who have never touched this technology before, that claim to be experts in understanding how these levers work without years of pulling the levers and watching it.

[13:21 –14:47]

Daniel Angelucci: Now, it's a great point. And I think the notion of supervisory agents is going to be really important for another key reason.

So, we experienced this with a client who has essentially PII data that sometimes leaks into what they want to analyze using generative AI. So, it's not, you know, it doesn’t typically happen, but it does occasionally happen.

And it's enough that it merits having an agent who is actually reviewing the information that's going out to the LLM, that's actually, sort of, determining the value of those discussions, to actually edit them and say, “Okay, this is not data that we should be passing on to the next agent. We should actually be restricting this data.” So, we'll remove it from the dataset that's going to be analyzed.

And I thought that was interesting. The problem that we bumped into was that essentially those supervisory agents are, you know, not perfect. And the issue becomes like, okay, so if they're 95 percent okay, if they get it right 95 percent of the times, is that good enough?

And I think this is something that—as we look at this—the client needs to really ask themselves—under what circumstances is 95 percent going to be good enough? And if it’s not, how is it that I can train it for the rest of that to get to a confidence level that is okay?

And that piece is really important for stuff that’s, you know, PII data or you know, certain other kinds of sensitive data you’d want to analyze.

[14:48 – 18:09]

JP Morgenthal: It’s a much larger domain than even that. I think it’s—so, you know, I participated in development of this report called the Agentic Process Automation. It's a representation of robotic process automation or business process automation.

It's the next evolution of those technologies that are all kind of being compressed or merged into one space now, which is this APA, right? And what APA is, is essentially the combination of deterministic workflow or process management using nondeterministic AI as a means of executing tasks. And it’s all—it’s a really complex space because you're expecting a deterministic outcome.

Again, I did this—I’m building this Gmail inbox cleaner. And initially, I was like, “Oh, I’ll let it do the work.” And then I backed away from that. I said, “No, I think I’m just going to have it labelled and classify the emails for me.”

And you begin to learn really quickly, that no matter how much you train it, it’ll, you know—it’s labelling some of the political emails as ads. It’s labelling things that are clearly ads as “for review.”

And you, you know, you’re taking that away and you go, “Okay, now why?” All of a sudden, it’s got to return just one-word answers. I instructed: “Just give me one-word answers. You're either an ad, you're either politics, or you're a review.” Right? I created three buckets—very simple—not even a lot of buckets. And it does about 80 percent accuracy.

Now, for this process, that's fine.

For what you're talking about, it’s not acceptable. I need a perfect and repeatable deterministic outcome. Not gonna get that from AI today. It’s just not there. Now, you’re getting better with the larger language models. The ones with billions—you know, the Claude models, the Gemini models, right? Okay.

But now I’m delegated to using a service provider—an LLM service provider who has spent a lot of money building their model by consuming the entire world. And so I have the issues of that that we probably will talk about in a moment about what's in there. Garbage in, garbage out.
But, they give me speed, and they give me better accuracy because they have billions of parameters that they've put through and a lot more samples, right?

So now you get these trade-offs, but if I'm an organization and I'm looking at this, do I really want to pay Anthropic, you know, millions of dollars a year to use their engine, or do I want to try to get my own engine fine-tuned? And working specifically and narrowly on my domain and problem.

But even still, now I have the issue of, alright, well, I gotta go get some GPUs to speed this thing up, make it usable, and I also have to fine-tune it and watch it.

[18:10 – 18:26]

Daniel Angelucci: I mean, I think it's even more sort of complicated than that in a lot of ways. This is the client issue here is that the agent—the supervisory agent—is the one that's now going to burn all the cycles, right? And you're still one step away from actual business value.

[18:27]

JP Morgenthal: That's right.

[18:28- 19:13]

Daniel Angelucci: And even if you get the information out and it's PII-clean, it still has to go out to another model, which is actually going to provide the business insight.

So, the problem, is that this ends up looking an awful lot like overhead, right? And its overhead that actually just reduces the value of whatever it is that you're trying to achieve, which makes you sort of ask the question—and I think it's been asked quite a bit lately—is like, well, is this really worth it? Is this really going to yield results in these kinds of circumstances?
So I think as much as what makes it possible, I think, understanding the limitations of what's available with respect to generative AI is important so that you can ask the right questions of it and really get the value out of it from a business perspective because it sounds and feels to me like it can be a black hole, right?

[19:14- 19:58]

JP Morgenthal: You raise a great point. Just because I can do this—should I be doing this? I mean, that's the answer, right?
Alright, great, you've given me this new tool and I can use this tool to do all this work on my behalf. I might even be having people who are doing this today, so I don't have to expend labour on it. Now, if you're expending labour on it, I think you immediately get to a labour charge, labour arbitrage calculation. That's fairly simple, right? That's a gimmick. But if you're creating this new architecture, you then have to ask the question you're asking, which is: Okay, I can do this. Should I be doing this? Is this worth it?

[19:59-21:13]

Daniel Angelucci: Yeah. And look, I think that segues really nicely into the second part of our discussion, which is really about what makes it valuable, right?
So if I'm an organization and I'm looking at deploying AI—and generative AI specifically—and I have all of the requirements that we talked about earlier with respect to, you know, understanding data quality and having unstructured data, but still being able to bring it to the right places, it's available. I've got a relatively complete view across that.

That next question then is: Okay, so how do I actually make this something that's business-focused? And is actually going to show consistent ROI, right?
And, I think one of the things we bump into a lot is that there is an architecture associated with training versus inference. There is not a great deal of depth of understanding on this with respect to how businesses are going to go forward with this. So, the view of having the right sort of dataset that we can present to the training models—and then being able to use the inference models closer to the edge in such a way that we can actually return data or return information quickly and make sure that users are getting real value out of this in the short term—that's the piece where we haven't seen a lot of discipline in terms of the emerging architectures. Kind of wondering if you're seeing the same thing?

[21:14-22:38]

JP Morgenthal: I see two primary uses—one much more heavily discussed than the other. The first, which we've hit on a little, is process automation.
I see this technology, attempting to being heavily thrown at process automation, which, it can be useful In the right spaces. I think the opportunity that's more ignored—and I think this is because people just haven't got their head around it—is, I mean, we've been trained to be deterministic in business. Our CRMs, our ERPs, all of our systems—our HR systems—they're all deterministic. If I press this button, this time, in this way, with this data—every single time—I'll get the same result. In no way, shape, or form will that deviate because the outcome is not random; it's based on a program.

Right, you introduce large language models, or nondeterministic outcomes.
Every time you press that button, it could be different. There's actually something called LLM drift—prompt drift—where because of the way you asked the question last time, the next time you might get a different answer. It's Schrödinger's Cat. The minute you press that button, you change the outcome, and you change it for everybody else.

[22:39-22:47]

Daniel Angelucci: I was thinking actually it's more like talking to my children, right?
Every time I ask them a question, I get a slightly different answer.
But yes, yes, I agree.

[22:48-24:14]

JP Morgenthal: Yeah, right. That's a fair one. Children very much are like AI. But the other piece of it is where nondeterministic works well—sentiment analysis. It's okay—a 75 percent chance that this might happen, you know? Financial markets, right? Black Swan-type events, things like that. Fraud—what are the chances, what's the likelihood that this is fraud? What is the likelihood that if I combine these chemicals in this way—which I've never done before—it will fix this problem or heal this person? What are the…those types of solutions.
And it is being applied there, but you don't hear it as much. And I think that's where this technology really thrives right now because you don't have the consistency, you don't have the determinism that you would expect for process automation.

So, yeah, you know, again, everyone's looking at it as a labour arbitrage issue. And "Oh, I can have AI replace..."—no, you can't. Because the only thing they really replace well today is the person in accounting who you ask, "What is the likelihood we're going to get paid?" "Oh, sir, that's an 85 percent likelihood they're going to pay their bill." Right. Okay. Great. You're replaced. Now I can replace you with the bot.

[24:15- 25:28]

Danile Angelucci: Yeah, there has been a lot of talk, particularly, in consulting, about how it is that we would use this. We are asked to process vast amounts of data – a lot of it is non-deterministic in larger sense, right. I think, the ability to come to, you know, likely insights –

not certain insights but likely insights, very quickly, is extremely valuable for my particular profession, right. That's what we need to do in a lot of ways. And, I think you are right. It's a perfect use case.

Because it doesn't really hinge on the fact that you're a 100 percent correct. Presuming that it's unsupervised, right? So that supervised/unsupervised sort of capability along this line sounds kind of dangerous to instil because there is that zone of uncertainty. But the supervised capabilities—of being able to consume that much information and then spit back something that looks like insight—that is extremely valuable. And I guess, I mean, I am curious, 'cause you've obviously done more in the process automation space than I have certainly around this.

So yeah, so what does it mean in process automation to be 85 percent correct? Is there a supervised level that needs to be there in order to make sure it works? Or what are the safety levers that need to be in place to make this valuable?

[25:29- 26:35]

JP Morgenthal: Well, people will say that they can make that work. Again, it wouldn't be for, I wouldn't apply it to a bank transaction, right? Repeatability and consistency must be incorporated into the outcome of the transaction. Sales, marketing—those are areas where I think it can apply. Who is a likely candidate that I should approach? Or who might need my product? Right? Based upon all these other pieces of information—the type of industry they're in, the type of products they produce—what kind of message can I send them that would make sense? I think it does well there.
"Formulate for me a strategy. Formulate for me the messaging that I should include that would make sense to them in their language," about their industry, and how what I'm selling, and my service will help them and benefit them. I think it can do that.

[26:36-27:29]

Daniel Angelucci: Yeah. I mean, I think, it's, one of the strange side effects of generative AI's emergence has been this notion that there is actually an awful lot of business value that's gleaned from non-deterministic analysis, right?
And I think maybe that wasn't something that people would admit to before. I don't know, but the use cases around AI have driven that in such a way that I find it very interesting, 'cause there is real value to be had there, right? And I think it is, how is it that it can be captured.

Well look, so before we, you know, wrap up, I wanted to ask you too. So look, there are a lot of people who are sort of committed to this journey. And maybe they've committed in ways that are not, as we've described, in terms of the discipline that's associated with it.
So how would you recommend people go forward who maybe have sort of started this journey**,** but maybe haven't thought about it in the way we're describing? What would you do for them?

[27:30-30:05]

JP Morgenthal: For me, it always gets back to—what is the business goal, right? I actually have a methodology and a workshop that walks customers through from business goal to POC/MVP in a very short time frame.

And the whole real purpose of that is to take a business goal and break it down into the pieces of the deterministic and non-deterministic components, and then identify, you know: What can I do, or what do I need to do to have this operate in a production-like manner? What kind of security do I need? How do I need to validate the data that I'm putting into this? Can I use a publicly hosted LLM or do I need to do something on my own? If I did something on my own, how would I go about training this? Who can access it? How do I make sure people aren't adding stuff into the LLM?

And that ends up messing up my outcomes.

Right after I train it, after I fine-tune it—what if another process, I'm not even talking about maliciously, what if another process, determines, "Oh, I'll reuse Joe's LLM. He's got one out there." And now all of a sudden—it's like when my wife and I use the same package, and she gets annoyed because I, you know, she comes in and it's like: "I had my project all set up for this, and now I'm seeing these other pieces... these other options."

It's like, "Yes, I had this organized to work for a specific purpose, and now you're adding in all this peripheral information that has nothing to do with that." And it doesn't know the difference. It's trying to accommodate. It has an algorithm. Right? YouTube—something as simple as that. All of a sudden, you have your YouTube. You go to YouTube, after a while, it learns you. It gives you data, it gives you videos that you're interested in.

Now imagine your kids get on there and start watching videos on your account. Right? And the next time you go to watch, it's like, "Why is SpongeBob coming up all of a sudden in my list of videos?" It learns. It's an algorithm. And those algorithms can be modified based on what it's being asked to do. And I think, that's where the risk is involved—or part of the risk is. You know, you have to have controls. And I don't think many of these organizations that started this have any clue what controls look like, in, you know, a generative AI architecture.

[30:06- 30:30]

Daniel Angelucci: So that's a great point. So the controls that we're talking about are around output quality, as well as what is being drawn in. Because effectively, you need to have some sort of feedback loop, But you also need to be able to understand what the limitations are—what's being drawn in to be used across this LLM. Is that right?
And are there others that we would want to sort of think about across that?

[30:31-31:31]

JP Morgenthal: Well, for me, that's the big one in. You need to control who has access to it, what data is being sent to it after you fine-tune it. You only want it consistent for that purpose. And you don't want peripheral, extraneous information introduced.

I would say, that if you control the inputs and fine-tune it a certain way, you can rely to a certain degree on the outputs—which is good, because that requires less maintenance and overhead. But I think, occasionally, you have to do sampling, right? It's like a chip foundry. You get to the point where, "Okay, my sample size can be smaller over time. I'm pretty much sure this process is creating, you know, I have had batches that are 99.8 percent pure."

That's great. But I think, you know, twice a week, I've got to take a sampling and make sure the machine hasn't moved a thousandth of a millimetre and now I'm off.

[31:32]

Daniel Angelucci: Yeah.

[31:33- 32:15]

JP Morgenthal: And so yeah. Initially, you're, it's a curve, right. Initially, you are doing a high degree of testing, a high degree of testing to ensure the outcome you expect is consistent and remaining consistent. And then over time, you can reduce— but you still need to continually sample to make sure there's been no drift. And if you do, you may need to restart and get back to a..here's a great one—no one's taking snapshots.

"I'm going to snapshot this model at this point in time because I may need to restore it later to get my process back in alignment."

[32:16-33:12]

Daniel Angelucci: Yeah, absolutely. Right. We had this discussion about chatbots it seems before, where the problem is, if you don't test and you suddenly find that you have a racist chatbot. It's not just racist once. It's racist to every single person who comes through.
It's not the problem, the same level of problem as having an employee who is misbehaving. This chatbot is misbehaving consistently, which can provide some pretty ugly business results very, very quickly. So yeah, test, test, test.

And I think, you know, we're seeing, I am seeing, the notion of how test-driven development has worked in software development now for years and has started to become prompt-engineering, kind of, a feedback loop. And that piece of it is, I think, really, really valuable—is to get people with that kind of discipline to be looking at how it is they're building prompts. And I hope to see more of that in the future, I guess.

[33:13- 35:24]

JP Morgenthal: Well, and that's where you kind of need LLMs to train LLMs.
Because, you know, if I do an agile process, and the way I do test-driven development, I build my test: "I'm going to send you X, Y, Z. I expect you to send me back 1, 2, 3." I can build a test that makes a call to a function, and then I can look at the outcome, and say, "Great—I expected 1, 2, 3 and I got 1, 2, 3."

Now I send an LLM a prompt. Now, the prompt might come back as different words but mean the same thing. So now I need something that can interpret that and say— Does it mean this? That's a lot harder. My testing now is I'm looking for something, a response that means this. Doesn't matter the exact words—does the response mean XYZ?

So, you need an LLM that can then look at the outcome of an LLM and say, "Yeah, I'd say there's a 95 percent chance it means what you think it means." If you're saying, "The sky is blue," I'm going to say: Yes, it answered, "The sky is blue." It didn't tell you it was yellow. It didn't tell you it was green. It didn't tell you it was orange. It may have said, "sky blue", it may have said "cyan."
Here's a great one—when I was working on with pharmaceuticals. You've got generics, multiple tiers, all with different names, but they all mean the same thing, right? Acetaminophen is acetaminophen—but is it Advil? Is it Tylenol? You know, so depending on what name you get, you then have to say: Look, I'm just asking—do they want to see acetaminophen? I will figure out what tier later, but are they requesting acetaminophen? So, you know, it gets to that kind of degree where, yeah, I'm pretty sure that this order, this sales order, is asking for acetaminophen, they may have had it differently by name. So those type of operations, all require, can get to, you can do it in testing, but it's LLM and LLM working together, right?

[35:25-35:55]

Daniel Angelucci: I find it just amazing, sort of as a parting thought here too—'cause my degree is in philosophy. I studied philosophy of language in school. And after 25 years in the industry, it's finally valuab**e.** Not exactly sure how that happened.
But being able to answer questions about equivalent phrases and what those actually mean—what that term actually means. It's been a very interesting twist for people who are unused to thinking that way. So, it's been kind of fun.

[35:56-36:51]

JP Morgenthal: Right after attention-based models came out and I understood what they were doing, one of the first things that struck me, that I noted, was how—not surprising, 'cause everything's math—but how mathematical all language is.
What they've broken our language down to is the relative distance of one word to another to decipher meaning. Right. What follows, what word is most likely to follow "The cat is..." Well, I've seen, 50 million, I have seen that sentence 50 million times. Here's the probability layout. The next answer is: "brown." Really, what it tells you is that most cats are brown.

[36:52 – 36:56]

Daniel Angelucci: Yeah, I was going with Schrödinger on that one: "The cat is dead," right? That's right.

[36:57- 37:03]

JP Morgenthal: But it is interesting, right? Ultimately, what this tells us how mathematical our universe is.

[37:04- 37:10]

Daniel Angelucci: Absolutely, absolutely. So, we're about out of time. Any parting thoughts, JP, before we sign off?

[37:11-37:50]

JP Morgenthal: My parting thoughts are really just—I enjoyed this. I wish you and I would do more of this. We've chatted over the years, and this topic really...what's interesting is how we're both kind of passionate about making our clients aware of the challenges and hurdles that occur because of this technology. And what goes into really using it in a business value sense. And I think that's important**.** I don't see enough of that.
I do see a lot of pitches on Facebook about how to use AI to make millions. But I think the only people making money there are the ones getting $5 a click.

[37:51-38:03]

Daniel Angelucci: I think AI was generating that too, so it's that much more interesting, right? So anyway, I wish we had more time for discussions. 12 time zones is always a challenge, but we'll have to find a way around that. Thank you very much for your time.

[38:04-38:07]

JP Morgenthal: Exactly. Yeah. And thank you for doing it in the evening for me. Next time, I'll be doing the evening for you.

[38:08-38:13]

Daniel Angelucci: Excellent. Look, thanks, I really appreciate it, and we will speak again soon.