Paul Goydan: AI cost has gone from the server room to the boardroom in the last two years. If we simply go after AI cost the same way we go after commodity IT spend, we don't get the value in the business.

Vlad Lukic: AI, especially generative AI, is not like an Ozempic. It is more like a gym membership, right? It will cost you to go there. It will cost you to actually exercise. You will need to dedicate time, but for awesome results.

Georgie Frost: Welcome to The So What from BCG, the podcast that explores the big ideas shaping business, the economy, and society. I'm Georgie Frost. AI was supposed to help businesses become more efficient. Instead, for many, it's adding a whole new layer of cost and putting pressure on the bottom line.

So how do leaders decide where to cut back, where to spend more, and crucially, how to pay for it? Joining me today are Vlad Lukic, BCG's global leader for tech and digital practice, and Paul Goydan, global leader of BCG's AI cost advantage practice. Vlad, Paul, welcome.

Paul, as I said at the top, with AI comes this great promise of greater efficiency, doesn't it? But for many businesses, the economics just don't quite seem to be adding up yet. Why is that?

Paul Goydan: I think first of all, we love to experiment and pilot. When they don't work out, we don't always stop paying for those. A lot of companies also went into token maxing to get their employees to test and try AI, and that led to exorbitant bills, but low value.

Georgie Frost: Vlad, isn't that just simply the cost of adopting a new technology or are companies spending money in the wrong places?

Vlad Lukic: No, it's a little bit of both. We are in the stage where we are, which is you could argue a lot of those costs, as Paul nicely said, are in the early stages. I would put it in the bucket of learning and development where we need to get our hands on keyboards to get a sense of what can be done. But as we mature and progress in the deployment of AI, the companies that really unlock the value are the ones that tie... They're really thoughtful about where they spend the money on AI. They're very clear in articulating where it's adding new capabilities or new services to clients, and it's making a difference, and they're managing it as such.

Georgie Frost: When does encouraging adoption become encouraging waste?

Vlad Lukic: Yeah, that's a great push. And each company will be in a different phase. If you're really trying to encourage your employees to get to learn what's going on and just experiment, you need to be token maxing. But then you need to very quickly, once you get to a level of understanding and usage, quickly start pivoting in the direction of managing and being very explicit on where is the AI adding into the process and it's making you more efficient versus it's doing redundant work or it's giving you a tool that is doing extra stuff that's not needed.

Paul Goydan: Value-maxing is in. I think part of that is when you need to learn how to drive, you have to get behind the wheel, but you don't need to drive a Ferrari or a Maserati. A frontier model costs 30% more than a good enough model. Yet when we come into corporations, we see more often than not, very basic rudimentary tasks are being handed over to frontier models. It's like giving your teenage driver a brand new Ferrari.

Georgie Frost: Which is never a good idea. Vlad, you spoke about the overspending, underspending. Can you give me some examples?

Vlad Lukic: I'm seeing, for example, in operations companies that are really good operators in lean where they feel like, "I don't need more technology in here. I'm really running a lean process," so therefore they don't spend enough. But the mindset they have is one to two, 3% improvement.

With AI, you can be thinking 30, 40, 50. And so therefore they underspend into those core workflows that they're really good at, thinking that they got it versus if they spent more there and really rethought them and got a step change performance improvement, it would be a lot of money that they're creating.

Georgie Frost: Paul, I just want to ask you, the AI spending has become a boardroom issue remarkably quickly. Just explain to us why this is no longer just an IT conversation.

Paul Goydan: AI gives back to the shareholders, makes companies better by impacting the P&L, revenue, customer engagement, and making companies more efficient. None of that touches the CIO's budget. If we simply go after AI cost the same way we go after commodity IT spend, we don't get the value in the business. A rough rule of thumb we have at BCG is for every dollar of value you see, you should be spending about 20 cents more on the tech stack and then 80% is the efficiency gain.

Georgie Frost: Vlad, we know that businesses are being told to spend less, but then at the same time, spend more because obviously you need to invest in this AI and the new technologies. So where should organizations be cutting costs and where should they be investing more? And importantly, how do you figure that out?

Vlad Lukic: You have to do both, right? And one of the things that I like to use, a few of my colleagues came up with a really cool analogy that AI, especially generative AI, is not like an Ozempic. It is more like a gym membership. It will cost you to go there. It will cost you to actually exercise. You will need to dedicate time, but for awesome results. But it's persistency of the motion and presence in it.

Now, as it gets to AI specifically, there's a lot of waste. When we come into companies on the tech side, we usually find 20 to 30% of the dollars being wasted. You can imagine the tech world has been upside down over the last two years, charging for licenses to charging to consumption, to charging to tokens. The procurement practices have not evolved as quickly.

So there is usually, I call it sweating the stack opportunity where you can free up the cash that you can then deploy to buy some of the new technology. As you use the new technology, you also need to link it directly to business outcomes and free up resources there. So as Paul nicely said, you either need to gain some efficiencies or you need to generate some momentum on the top line.

Georgie Frost: You anticipated my next question, which is where is this money coming from? But you're thinking it can come from efficiencies elsewhere.

Vlad Lukic: It has to, right? It cannot just come from the tech side. It needs to come from the new business outcomes. That can be either on the efficiency and the cost side, or that efficiency translated into higher commercial momentum that allows you to charge and get better margin into the business.

Paul Goydan: A great example of what Vlad was just illustrating is one of our clients who was seeing a tremendous return on AI in the marketing business. That AI cost became part of the marketing budget. And the CMO had to justify why they're spending more money on marketing through the traditional metrics of better customer engagement, higher sales, higher revenue per customer engagement. It's no different in any other part of the business. The cost of AI needs to sit with the end business user of the AI.

Georgie Frost: So Vlad, walk me through this. How exactly should businesses actually be thinking about the cost of AI?

Vlad Lukic: We can categorize pretty much all the AI spent either as CapEx, which means we've used it to build some sort of infrastructure within the company. It should be categorized that way. Then we have part of it that should be OpEx, which is new set of tools that are part of me just delivering my daily jobs in a specific way. And I should categorize it as OpEx.

And then there are components of it that are directly tied to the services or products that we provide. So it becomes a COGS. IT can enable this service and access to this technology, but if the companies categorize it this way, it's fairly easy to then link it into the business ownership that then ties it to business outcomes.

Georgie Frost: Paul, does the cost categorization that Vlad was just talking about, does that change who's kind of responsible for the cost and ultimately for deciding whether it's delivering value?

Paul Goydan: I think it does. The business owner needs to be accountable for the cost of deploying AI. The CIO's role is to help you be efficient in the technical choices and the AI application you make. But the ultimate ROI sits with the business, just like any other investment.

Georgie Frost: What should leaders actually be measuring? How do they know that it's paying off?

Vlad Lukic: They should measure the outcome they set out to achieve and value creation. I'll give you a silly example. With a client, they tried to measure, or the initial objective was let's measure, can we do this task shorter than it took us before? And it used to take them ten days. Now they were doing it in a day.

But technically they had a solution that could do it in a day. In reality, the business processes were never changed and the experience of the customer didn't change. It still took ten days to get a response back. And they actually had an extra cost. And there was no forcing function to have them rethink the process because the metric they were looking for is, can I do this faster with technology?

Once they've realized the mistake, they said, "Okay, we're actually going to measure an outcome," which is that the customer gets a response in a shorter amount of time and that it costs us less. Once they did that change or in the metrics they were looking at, they rethought the whole process.

They eliminated a bunch of committees. They eliminated a lot of steps in between and managed to bring the process down to one day so you could get your response much sooner. And they also managed to optimize the cost, as Paul mentioned, on how you manage the context windows, et cetera. So they did it cheaper and much faster, but they had to change the metric they were looking at.

Georgie Frost: Would you say that's a typical example? Do you think most companies get it?

Vlad Lukic: Listen, there's a lot of smart people in a lot of companies. We are seeing the profit numbers across the industries being record high. So companies can do this and when they focus on it, they get it done. The gap right now in the market is just the experience of doing that with the AI in the mix as a new tool, if you will.

And what I'm seeing, the ones that persist with the motion quickly feed the learnings that I just shared back into their workflows and how they manage it. It's not rocket science, but it just requires that persistence and reflection and feedback loop. And the good news is that a lot of companies are figuring it out and they're actually fine-tuning their motions.

Georgie Frost: Paul, you both speak to leaders all the time. How much do you sense that a lot of the spending is being driven by a clear business strategy and how much is I suppose a bit of fear of being left behind?

Paul Goydan: FOMO drives a lot of spend. Every executive I talk to has been asked by their board, "Are you applying AI enough? Are you consuming enough AI?" If we flip the question, like Vlad said, to, "How much have your earnings improved because of AI?" It changes the ask. I absolutely think there's a broad distribution and there are too many companies who are misapplying AI and not getting the return they could.

Georgie Frost: When you say misapplying AI, explain that to me. What are the biggest mistakes?

Paul Goydan: AI, in some cases, is no different than any other application of technology. If you go in with a solution looking for a problem, you're guaranteed to spend money. You're not guaranteed to get a return. If you have a laser-like hypothesis of what AI will do better, faster, smarter in your business and how that hits the bottom line before your fingers ever touch a keyboard, you'll be ten times more successful.

Georgie Frost: Can you give me an example?

Paul Goydan: One of the most recent examples I had was applying AI in a manufacturing setting to reduce waste. The first question I asked is, "What percent of your cost of goods sold goes to waste?" And they said less than 1%. And when you do the math on 100% improvement of 1% of your cost, you couldn't pay for the AI. That was an application of AI to a problem with incredibly low value.

Georgie Frost: Vlad, Paul there spoke about FOMO, the fear of missing out and anxieties from leaders. Does that resonate with you?

Vlad Lukic: Yes, to a degree. And it is real. And I just want to make sure that it's not seen as a silly, I need to quickly copy and I need to be doing the same thing as well, kind of high school behavior, FOMO. It really is coming from really good examples out there where AI has created value and executives feeling, "Whoa, it's real."

So it's coming from, there's enough evidence to show that when it applied the right way, it can really generate tremendous value and it can create a really strong competitive advantage. And therefore it's in that context that they're seeking to make sure that they replicate something like that and to make sure their competitors don't get to that advantage before they do.

Paul Goydan: Georgie, I think FOMO is fear of market obsolescence. CEOs and boards are afraid their company won't be relevant in five or ten years because of AI. It's not missing out because you're not a cool kid. It's actually that your company could be left behind in a way that it becomes entirely obsolete.

Vlad Lukic: Yes. Yes. And it's visceral, it's real, and it for sure is driving a lot of this behavior, which is actually, I would say, very logical.

Georgie Frost: Without veering into the realms of psychology, which I'm certainly not qualified to talk about, but how do you get over that? Because it's absolutely real. It is impacting leaders' decisions. How do they get past that?

Vlad Lukic: I don't think they need to get past it. It is actually the right driver. The question is how do you manage? How do you channel that energy where it's going to create value? But Paul, I don't know if you would agree with that or would you steer differently?

Paul Goydan: I was thinking of an ostrich analogy of you can't put your head in the sand and take flight, but otherwise I think you hit the nail on the head.

Georgie Frost: Beautiful. Vlad, what are you hearing from leaders? What are they telling you? Why do they ask for your help?

Vlad Lukic: Because they're struggling with a lot of things that we mentioned, which is, okay, I'm getting pressure from the board to push this stuff in. I'm getting pressure from my employees who are saying, "Give me these tools." I've never faced it before as a business leader. I didn't learn about it in school.

So I don't have an intuition in my gut that tells me I should lean in or I should not lean in. How far? How should I even sanity check? So it's the ignorance meeting extreme demand and then the pressure and push from a lot of the tech companies. So they're feeling pressure from all of these areas. And fairly easy way out of it is, you know what? We'll put some budget aside. Let's start experimenting and let's just start testing and building intuition around it.

The trick is to then at some point move from those experiments into lessons learned that are feeding into the actual interventions and how you manage this. And the companies that manage that transition the faster are actually the ones that are winning. The ones that have leaned in and have figured out these things are growing one and a half to two times faster than their competitors.

Their stock price, total shareholder return is three to four times higher than those that have not figured it out. Their EBIT is one and a half to two times higher than the others. So it's totally doable.

Georgie Frost: Paul, what sense do you get from those leaders that are going to win in this space?

Paul Goydan: The leaders that are actually seeing return on AI are the ones that can, in an elevator, articulate what they're trying to change in their business and how it drives earnings. That laser-like linkage and understanding of the technology I think is going to set market leaders apart.

Georgie Frost: Laser-like articulation, Paul, why is it so hard?

Paul Goydan: It requires both a deep understanding of what the technology can and can't do, and deep operational insight on your own business. Those two skill sets are rarely overlapping with the modern cutting-edge technology.

Georgie Frost: Vlad, one trend we're seeing is employees signing up to AI tools themselves rather than waiting for them to be approved by IT. What challenges does that create for businesses?

Vlad Lukic: On the good side, it generates enthusiasm and it generates momentum, et cetera. The negative side is it exposes a lot of company data to non-enterprise-grade tools. So your data can leak out the way it shouldn't. Two, it creates a much larger cyber-attack surface for bad actors to access the company.

Three, it can start generating data that is very inconsistent because people are using different things and feeding it back into the corporate environments in a very inconsistent way. So the list becomes very, very, very long and therefore companies need to stay very close to that reality, which is very true, and channel that energy into the enterprise grade solutions that the employees can get access to.

Georgie Frost: Is that something that you would, I don't want to say the word ban, but is it that serious? Because some of those things that you said suggest to me that this is a serious problem.

Vlad Lukic: The question is, can you really ban it given how easily accessible it is? You can now take your phone and take a photo of this and put it in your personal tool and it can start doing things. So it is really hard to ban. I think the key thing is to acknowledge it, educate the employees on the dangers of doing it in an irresponsible way, and then create effective alternatives to that within your corporate environment. So you've got to attack it from a few different ways. You cannot ignore it.

Georgie Frost: We've covered the "so what." Now, the "now what?" What are the next immediate steps leaders need to take to ensure that their companies are the leaders of the future? Paul?

Paul Goydan: All of our clients need to start to understand the technology and the billing mechanisms and treat it just like they do any other large cost category. And there's certainly a lot of efficiency to be had across the board. Model selection is a whole new capability. Likewise, the size of your context window and how much you use AI in the same prompt grows exponentially to the amount of usage.

How you think about that in business applications is a whole new capability. We have a whole new practice emerging, Georgie, that didn't exist six or eight months ago, and it's helping clients understand how to cost effectively apply AI.

Vlad Lukic: The fascinating thing now is things we can do now were impossible to be done even just two, three weeks ago. So not treating it as a project, but rather as an ongoing muscle that you need to be exercising is the way I would think about it.

Georgie Frost: How do you then think about cost? Same buckets, same compartmentalization?

Vlad Lukic: Same buckets and same focus and acknowledging that how you get charged will evolve very quickly based on the different entrants. Some of them will try to charge you based on outcomes, some based on usage, some based on some sort of subscription. There will be a lot of different models. You still need to diligently think about that, right? And that is part of building that intuition of how to anticipate those moves and being part of that ecosystem.

Georgie Frost: Vlad, Paul, thank you so much. And thank you for listening. If you'd like to find out more about this subject, you can find links to Vlad and Paul's research in the show notes.