



RPM—43 | The emergence of Responsible AI with Suzanne Tavill

Michael Venne: [00:00:02] A lot has been said and written about AI, specifically generative AI's potential to bring enormous benefits to the world.

- Businesses stand to realize greater efficiency from streamlined processes and reductions in manual work.
- Healthcare researchers may one day find it easier to diagnose illnesses and come up with novel therapies.
- Customers may find they have a better consumer experience, and
- Financial institutions might be better able to tailor their products and services to their clients

If it's any indication as to just how optimistic the world is on generative AI, OpenAI is now valued at \$80 billion, making it the most valuable tech company in the world after ByteDance and SpaceX.

Today's guest isn't here to refute the potential upside of generative AI, which could pervade every aspect of private markets. Rather, as she describes in one of our recent whitepapers, she makes the case that in order to realize this enormous potential, AI needs to be developed and deployed responsibly. And to this end, because of where they sit, private markets has a significant role to play. Joining me today to discuss the burgeoning field of Responsible AI is StepStone's Head of Responsible Investment, Suzanne Tavill, a partner based in our Sydney office. Suzanne, welcome back to RPM.

Suzanne Tavill: [00:01:19] Thanks, Michael. And it's a pleasure to be here today. And I totally want to reiterate your point. This is absolutely not about debating the upside that is likely to come from AI. It's about ensuring that what we deliver can really achieve this potential. And you know, we're always keen at StepStone to contribute something into the literature and our community. And certainly what we see now is that a lot of the thinking about how AI can be made responsible is still relatively nascent, particularly within the context of private markets. And so hopefully this is our contribution to that discussion.

MV: [00:02:06] And because of its nascency, as you put it, , I think would be helpful to start with some level setting. So what is responsible AI and what is it trying to achieve?

ST: [00:02:16] So Responsible AI is the name of an initiative or approach to try and address the harms that can come from AI and in particular, generative AI. So these harms are occurring because of risks that are evident within the underlying generative AI architecture, the fact that these are autonomous and adaptive systems. So we understand their risks. These risks are causing harm. And Responsible AI is stepping up as a way to institute systems and processes to address those risks and therefore remove or attempt to remove the ultimate harms.

MV: [00:03:11] And why do you believe that private markets are positioned to influence the course and direction of this Responsible AI movement?

ST: [00:03:22] Because we house the venture capital community that is a key funding source for AI businesses. And even more broadly than venture, across private markets, be that infrastructure or real estate or more developed companies under private equity ownership, all of those businesses are going to be key customers of generative AI systems.

So what we have seen is a venture, GP, supporting the growth of a generative AI business, it is far easier to embed good practices earlier on than try and retrofit them once the company has become far more developed and established and even more complex. And so we see that there is an enormous opportunities for GPs to support their underlying assets in driving Responsible AI practices through these businesses. And because the space is so nascent, everyone is really keen to get support.

- Founders are keen to get support.
- Venture GPs are happy to support each other.
- And the broader peer group initiatives that exist in the industry, such as VentureESG, are trying to lean in and support the development of Responsible AI thinking.

Similarly, as I said, being a customer of the system, as we always say, the customer is always right. And if customers start to demand that these generative AI systems are developed in a way

that drives a trustworthy AI system, which is obviously the ultimate objective, you know, a trustworthy system is the antidote to a system that is riddled with all these risks.

MV: [00:05:43] Okay, so we're talking essentially about a feedback loop. I do want to address that in a bit more depth. Before we do, just want to back to basics. because of the name Responsible AI, this may strike some as a corollary to ESG. But as you mention in the paper, this is not the case. What is the relationship between ESG and Responsible AI?

ST: [00:06:08] So Responsible AI has not been birthed out of the ESG movement, but rather it shares a similar DNA in thinking about how to address risks and opportunities. Interestingly, what we can already see is that GPs who have more developed responsible investment practices, they are better prepared at addressing the risks of AI in their underlying portfolio companies—be that as a company that's developing or as a consumer of this technology.

The paper explores how these key risks in generative AI, in particular, coalesce across E, S and G issues. And therefore, it goes without saying that GPs that have thought about how to address E, S and G risks are better prepared at doing the same thing here, within the context of generative AI.

Again, the paper spends quite a lot of time considering the G or the governance issue because when we talk about effective risk management, and in this case, effective risk management around these specific generative AI risks, then governance comes once again to the fore as being a key lever to manage such issues. And, you know, this is an area where private markets generally tends to be strong. And what we have noted is, again, GPs with more developed ESG practices tend to be even stronger in terms of what they can offer their portfolio companies.

MV: [00:08:01] Oh, thank you for that. So as I mentioned at the top of the show, almost everyone can envision how generative AI can make the world a more creative and efficient place. But perhaps fewer people are contemplating the risks and harms. Because of how widespread these risks could be and the myriad ways they could manifest, I'm not suggesting that we list them all, but would it be possible to categorize or classify the risks?

ST: [00:08:29] Yeah. So we think about the risks in terms of five groupings.

1. The first one is that generative AI systems can drive and reinforce bias. Said differently, they can generate discriminatory outcomes. This is something that I think has hit mainstream media. So it can be, the employment company with the CV sorter that has generative AI enabled, but because of the historical data, tends to bias to the white male. But some of these biases are far more subtle.
2. The second big grouping is the risks around data protection. So we know generative AI feeds off enormous amounts of data. Now the issue is how that data is sourced and collected and how within those processes there is a respect or not around copyright, around IP.
3. The third is explainability—being able to explain why the system produces the output that it does. Now, given that these systems are adaptive, it is far from the case that the output of these systems is repeatable. That's something that we really have got in our brains from traditional software systems. You put in one input, you again and again will get the identical output. Not so here. And the question is to explain why the output is what it is, particularly when you think that we're going to be heavily dependent on that output, as you said, for creating drugs, for making decisions around people's financial situations. So this is pretty critical stuff.
4. The next one is human control—how within these systems can humans intervene or where within the system is a point to appeal to a human. So again, think about the person that's just had their loan rejected from a bank that was decided through a generative AI process. How does the person understand why the rejection has happened? Who can the person appeal to in terms of a human? How is that human going to be making a decision or are they going to loop back into another generative AI system?

MV: [00:11:12] Could the data have been biased and there could be discrimination at play compounding risks?

ST: [00:11:17] Exactly. And so that's another thing, is that these are not independent, groupings of risk. Often a number of these intertwine together.

5. Just like the last grouping of risk, which is what we call human centered, which is whether the system is designed to respect human values and human life. And so this becomes particularly important if we think of a lot of government systems; if we think of

a lot of defense systems. This is one that is intertwined with human control, right? Think about a drone system, how can humans intervene to protect human life if the need occurs.

MV: [00:12:03] So now that we've kind of broadly defined the risks, I'd like to shift focus on how we might mitigate said risks. Broadly speaking, as you've laid out in the paper, there are three roles in the AI value creation chain: developers, vendors and end users. Now, on the one hand, if you're attempting to mitigate risks to end users, the onus of responsibility lies with the developers. However, with fund managers, just turning to private markets here—our industry, many, many fund managers are already deploying generative AI to add value to their assets. As intermediaries, they have a critical role to play in both vetting the tech they're implementing, but then taking any feedback back to the source, i.e., the developer. I think in an ideal setting there would be a clear linear progression down the value chain. But you know, as you alluded to earlier, talking about customers providing feedback to the developers, I mean, the reality is a bit more chaotic. What is everyone's role in this more, I guess we'll call it a more realistic dynamic?

ST: [00:13:11] Yeah. So, you know, in the in the paper we draw on the work of NIST, the National Institute of Standards and Technology. This is a US agency under the Department of Commerce. And they really, in my view, have been the first group to come out with a really considered approach to what needs to be done at the coalface to develop Responsible AI systems. As a result, they spend a lot of time looking at the development process of generative AI.

And what you can see in their thinking, which I think is absolutely appropriate given the nature of these systems, is that everyone in the chain has a responsibility—from the developer, through to the vendors, through to the consumers. Their roles obviously are different. But one thing about a cascade of responsibilities is that if it doesn't start at the developer, it becomes extremely difficult for the vendor and the consumer to institute really effective risk-management processes.

And the second thing is that there has to be feedback loops the whole time within a development group, from vendors to developers, from customers to vendors, because as we've said, we're talking about adaptive systems. And one of the key issues has to be, is the system actually doing what we need it to do, and is it delivering it in a way that we believe is trustworthy? And so that is why the feedback loops become really, really so important.

MV: [00:15:41] I mean, that strikes me as something that is going to require like a immense behavioral change. I mean, if you're just thinking about how, you know, as consumers of certain technology, we might take feedback back to the vendor and say, hey, something isn't working for us. And all those cases, right, it's not going to lead to commercial harm, but it's just really inconvenient for us. But it takes a lot of time to sort out those issues. But with generative AI and some of the commercial, you know, I think the big commercial risks that you're going to talk to us about, it's it sounds like that the system is going to need to be much more adaptive, antifragile and responsive. You know, when we're thinking about big commercial harms from generative AI, how could they manifest?

ST: [00:16:24] Yeah. So already today we seeing businesses getting sued for breach of copyright; for producing systems that are creating discriminatory results; for [systems that cause] physical harm to a person. So, we see that there's a real risk of financial loss from not managing these risks, appropriately.

So what I'm saying is rehabilitation is pretty difficult. People, I think, will rather move on and go, what's an alternative and fund the alternative. So I think there's low patience, if you will. And so what does that mean? It means that you need to get this right pretty quickly. This is not a discretionary spend, that people have to do in terms of ramping up these processes and systems. This is actually, required to ensure and protect the value of the business.

MV: [00:18:17] So going back to roles and responsibilities and the value creation, you mentioned at the top that the venture capital community, because they're involved most heavily with funding these companies that are developing these tools, work more closely with the developers. How are you seeing VC firms that work closely with these developers deal with the upstream risks?

ST: [00:18:46] Forward-thinking VC firms are rapidly getting up to speed with the risks that are being presented and thinking about what is best practice that needs to be embedded in their portfolio companies. I think it's a very much sort of a collective thinking that is going on, and certainly we are seeing some groups take a leadership role in that.

Similarly, what we've seen is some forward-thinking VCs have made commitments around Responsible AI. Obviously this is a great start. But as we've seen in other spaces, commitments need to be swiftly followed by action.

So one of the big things that we can see where GPs can step up pretty quickly, is around the governance functions that need to be enhanced to ensure that Responsible AI practices, are being embedded. And this, as I said before, is really a natural extension of the VCs' traditional toolkit. So we think that this is a huge opportunity for the VC community to step up and show how they can drive this best practice.

MV: [00:20:14] We've just addressed the "G." We've also spoken to the "S" some of the societal risks. Let's talk about the "E," specifically the energy sources for data centers are something that GPs and LPs are going to have to contemplate going forward. I read something recently that, according to the International Energy Agency, energy consumption from data centers will double before the end of the decade. And clearly, there's a rub with climate change and the efforts to slow global warming. How should investors think about these cross purposes?

ST: [00:20:52] Yes, things are never simple. So firstly, we have to recognize that AI and all its successors, will require ever greater amounts of computing power. This is our reality and it's the reality for generations to come. So to square the circle, some things really need to become standard practice.

1. Firstly, data centers need to measure their emissions and produce carbon footprint reporting. So all this energy consumption is transparent, so they can start to take action.
2. The second this is, based on that information, they need to drive down their carbon footprint, and they can do this through adopting renewable resources as their energy supply. And they can enhance the data center design, literally the physical design of the data center, so that it is far more energy efficient—how cooling occurs, levels of insulation, etc.
3. Finally, chips embedded in servers have to be continually upgraded, to those that produce less heat and therefore are far more energy efficient.

And ultimately it is the customers of data centers that have to demand and make purchasing decisions based on this type of best practice. They need to see a carbon footprint and they need

to see resultant actions. And what we can see from some of the largest procurers of data centers, is that this is already an approach that is being adopted, but, you know, it does need to become far more widespread.

MV: [00:22:50] Suzanne, this has been excellent. Thanks as always for your time and insights. Take care and hope to see you again soon.

ST: [00:22:57] Thank you, Michael.

MV: [00:23:01] That does it for this episode of RPM. To download a copy of the paper that Suzanne mentioned, head to our website, stepstonegroup.com. RPM is available wherever you normally listen to podcasts.