arthur & harry

[00:00:00] Do you feel like you have enough cash now? Uh, I guess a startup is always fundraising. Do you think enterprises are ready for open source? The most technical saving enterprises are definitely ready for it in order to widen the adoption. There's definitely some tooling to be brought to the market. What are the biggest barriers to Misra today?

We are still bottlenecked by conclude for sure, but that's because we don't have many of it. We have 1.5 K 800, which is. A few percent of our competitors. Can I ask only, was it a mistake for you to not scale that quicker? I mean, you can't really scale that much quicker. You can't raise like 2 billion on the seed round.

I mean, at least you couldn't in 2023, which capacity do you most respect and admire? Uh, we were surprised by recently. It's a competitor landscape.

We are always mixing it up with new intro styles. Let me know what you think of that new intro style and what a show we have for you today on 20VC. Mistral is one of the most exciting AI companies today at the forefront of the foundation model charge. Joining us is Arthur Mensch, co founder and CEO at Mistral, where he's raised over 520 million.

In [00:01:00] funding from the likes of Andreessen Horowitz, General Catalyst, Lightspeed Venture Partners, and Microsoft. And before founding Mistral, Arthur was a research scientist at DeepMind, one of the leading AI institutions in the world. But before we dive into the show with Arthur today We're all trying to grow our businesses here. So let's be real for a second. We all know that your website shouldn't be this static asset. It should be a dynamic part of your [00:02:00] strategy that really drives conversions. That's marketing 1 0 1, but here's a number for you.

Scarlett 2i2 USB-5: 54% of leaders say web updates take too long. That's over half of you listening right now. And that's where web flow comes in. Their visual first platform allows you to build, launch and manage web experiences fast. That means you can set an ambitious marketing goals and your site can rise to the challenge.

Plus web flow allows your marketing team to scale. Without relying on engineering, freeing your dev team to focus on more fulfilling work. Learn why teams like Dropbox, IDEO and orange theory, trust web flow to achieve their most ambitious goals today@webflow.com.

Scarlett 2i2 USB-8: I'm speaking of incredible products that allows your team to do more.

We need to talk about secure frame.

Scarlett 2i2 USB-1: secure frame provides incredible levels of trust your customers through automation, secure frame, empowers businesses to build trust with customers by simplifying information security and compliance through AI and automation.

[00:03:00] Thousands of fast growing businesses, including NASDAQ angel list doodle and Coda trust, secure frame. To expedite that compliance journey for global security and privacy standards such. Such as SOC two ISO 2,701 HIPAA, GDPR, and more backed by top tier investors and corporations such as Google Kleiner Parkins.

The company is among the Forbes list of top a hundred startup employees for 2023 and business insider's list of the 34 most promising AI startups for 2023. Learn more today@secureframe.com. It really is a must.

Scarlett 2i2 USB-3: And finally a company is nothing without its people. And that's why you need remote.com. Remote is the best choice for companies expanding their global footprint, where they don't already have legal entities. So you can effortlessly hire, manage and pay employees from around the world or from one easy to use self-serve platform.

Plus you can streamline global employee management and cut HR costs with remote it's free HR. I S and Hey, even if you are not [00:04:00] looking for full-time employees, Remote has you covered with contractor management, ensuring compliant contracts and on-time payments for global contractors? There's a reason companies like get lab and door dash trust, remote to handle their employees worldwide.

Go to remote.com now to get started and use the promo code 20 VC to get 20% off during your first year Remote opportunity is wherever you are.

You have now arrived at your destination.

Harry Stebbings: Arthur, I am so excited for this. JC introduced us quite a long time ago now. I've known you for a while. I've been wanting to make this happen for a while, so thank you so much for joining me today.

Arthur Mensch: Thank you for having me. Um, it's a pleasure.

Harry Stebbings: Uh, the pleasure is mine, my friend, but I wanna start. What would your, or how would your parents or teachers have described the young Arthur?

I'm just always intrigued by the characteristics and traits of the best founders. How would they have described a nine, 10-year-old Arthur?

Arthur Mensch: I guess I was a bit curious and a bit stubborn, not very nice to my, to my brothers, I was the [00:05:00] healthiest of, of them also.

You should ask them. I don't know. I think they, they have good, uh, good memories, hopefully.

Harry Stebbings: Do you know what, sadly your mother wasn't in our reference list, so we missed that one out. But, uh, you know, JJ JC provided some great commentary, so I do wanna start though, also, you know, growing up, what was your first exposure to ai?

You're a kid in France, how did you first get exposed to AI and machine learning, and what was that first? Passion point,

Arthur Mensch: I think that was Andrew Ang, uh, flying in helicopter, uh, helicopter backward. control problem, which is not easy to solve. And I'm not sure if it was really AI related.

I think he was saying that he was using a neural network to control all of this. But, um, that's the well the first memory of, uh, me being, uh, shown what you could do with, uh, machine learning at the time. that was in 20 20 13, I think

Harry Stebbings: most recently that you spent Yeah.

Two and a half years, three years at DeepMind. Can I ask, what are the biggest takeaways for you from that experience and how did they impact how you think about building this role?

Arthur Mensch: A team of five is, uh, faster than a team of 50. Uh. [00:06:00] Except if you organize the team of 50 to be, uh, 10 teams of five that are sufficiently uncoupled.

one finding that I learned the hard way at DeepMind and the reason why we created the company in a slightly different way in term of organization, of the science team. and also the reason why we knew we, we had a chance to do interesting things, uh, with the smart team.

Harry Stebbings: Can I just ask sufficiently uncoupled, do you not lose efficiency or is there not. Uh, leakage between those silos and it creates actually inefficiency. By having such silos,

Arthur Mensch: you have to share some things. So you share the infrastructure, you share the code base, uh, you share findings. we're doing general purpose models.

In general purpose models, you need to evolve them in different directions. So you need to make them speak different languages. You need to make them be able to code, be able to do mathematics, be able to reason, uh, you need to add multimodality to them. All of these things are loosely coupled. It's useful if, uh, you use the same framework for optimization, for data, for training, but, uh, you don't want to.

Have your team spend their entire day in meetings for [00:07:00] coordination. And it's actually pretty hard to figure out. I think so far we've managed to scale it relatively well, the team is only 25 people, so that's actually not super challenging. Uh, it'll become more, more and more of a challenge.

that's what I remember from DeepMind. It worked very well at the beginning. Gemini was a bit too slow and I think they recovered. Efficiently. Well, since, we, we have optimized the team, to be as fast as possible and to ship as fast as possible.

Harry Stebbings: Was it an easy decision to leave, to start Misra? You know, you're at DeepMind, one of the best institutions in the world for AI with some incredible talent around you.

Was it an easy decision and just take me to that moment when you decided to leave to found or co-found Misra.

Arthur Mensch: So it's not a zero to one decision. It's not a binary decision. You start to think, uh, like I'm 10%, uh, leaning on living, and then it grows. And at some point you cross the threshold. Then you say, okay, well, I guess know that I'm sufficiently decided.

There's no way in which I stay more than a, uh, than a few days. otherwise I wouldn't be counted with my colleagues. And so that's, uh, that's how you, you, you get started. You say there's no turning back to,

Harry Stebbins: what was that point for [00:08:00] you?

Arthur Mensch: that point for me was probably, uh, around March, uh, end of March at the. Uh, last year where, I decided to leave on Friday and I resigned on Monday.

you you can't stay If you have decided to resign. It's, uh, all the way, it's not very fair.

Harry Stebbings: Uh, I, I totally agree with you. Now, I do wanna run this with some chronology.

I spoke to so many of your advisors, investors, and I wanna start with actually the first model. You know, Michelle seven B being one of the most popular. Released, you know, a while ago now. Why did you think it was so popular? What do you think you did? So, right, and what did you learn from that?

Arthur Mensch: I think it served two purposes. So the first was to, uh, show that, uh. There was a lot of, uh, slack in compressing models. And so from a scientific perspective, it was a good finding and, and a good learning from, for, from the community. it also filled a gap, in the efficiency to performance, uh, to the space of models, uh, where there was definitely something missing.

Uh, seven B is the size, uh, that allows to run [00:09:00] efficiently a model on your MacBook or on your smartphone. We made it sufficiently smart so that it was still useful. So there was already seven B models before, but they weren't good enough to do interesting applications. And so by, by targeting this specific space, uh, we talked to the developers immediately.

'cause developers, like the casual developers running on a, gaming GPU or on its MacBook. So it, it created a lot of curiosity and adoption because it was a missing spot in the, in the performance to efficiency space.

Harry Stebbings: When you look at like lessons from that and how it impacts future releases, any that really stand out for you,

Arthur Mensch: I guess?

Uh, it taught us that there was a lot of interest for efficiency, uh, other than scale. And so that's why we continued of targeting very efficient models with the mixed trial, H six seven B and and more recently mixed out ATX 22 B, ensuring that, uh, for certain, uh, costs and for a certain size, we were reaching the top performance of the market.

that has been our, major [00:10:00] motivation for us to target efficiency. Simultaneously scaling toward the larger and larger models.

Harry Stebbings: I spoke to Sarah Grow before the show and she said the core question that I think is, you know, with the focus on efficiency and the efficiency frontier, does scale matter?

Arthur Mensch: Well, scale matters in the sense that, uh, if you spend more training compute, you can make the models more compressed. So you do need to have some compute to compress models. no scale isn't the only recipe, the only uh, ingredient to, uh, the recipe. You need to scale, but you also need to have proper data.

Otherwise you reach, uh, some data quality. You need to have proper techniques, uh, for training. I mean, people call it compute multiplayer, I guess. how do you actually make some efficiency gain that are not costing you compute? Because compute is, is expensive.

And so one of the thing that we do at the Strat is to try and harvest is, uh, compute multiplayers.

Harry Stebbings: Can I ask, in that chasm of efficiency gain without costing more compute, is there much more efficiency we can eke out? Like is there a lot for us to eke out or are we working really at marginal improvements already?

Arthur Mensch: I think it's an open question. I believe there [00:11:00] is, I believe we can make models that are much better for a certain size, but it's, it's as open a question as, uh, can you find, can you make a much better model on the same kind of data by making it bigger and training it for longer? things you need to discover them also on the way you can try and predict, uh, the kind of performance you reach out, uh, you, you will achieve.

At the end of the day, you need to try it out. So that's the. I mean, it's really much a research field. You need to do the research and you need to try things.

Harry Stebbings: So I asked Sam Altman this question

what is the end state for the model landscape? Most people say, ah. It'll become commoditized and actually there'll be 12 players and it'll be a race to the bottom. What is the end state for models in your mind, and how do you think about the commoditization question?

Arthur Mensch: I think the end state is, uh, to have a more more features on developer platforms that allow us to do customization, that tell us to make low latency models that serve a certain purpose, that tell us to evaluate them and to improve them over time. And so the model is only a, like a tiny part. I mean, it's a central part, but it remains a tiny part of an application and what you want [00:12:00] to do across time.

And when you deploy an application, that you expose to users, you want to ensure that it works, ensure that is. That its latency reduces over time ensure that its quality increases over time. And so I think that's the, the end state is, models are effectively going to be. Starting point for any AI application developer, they need to be surrounded by tools, by, lifecycle management, uh, platform basically.

And that's the one thing that, uh, we started to build, like general purpose models are a bit differentiated, but the differentiation that you need to create for, for your application comes from the data you put into it, the user feedback that you gather, and the intelligence that you have to figure out what the application should be doing.

And that is not commoditized at all. There's no. Recipe that, uh, allows to go from a, a general purpose model to a model that is super good and better than all of the others at your specific task. this is a missing piece in the puzzle. That's one of the aspects where we're putting our strength, uh, on the product side.

Harry Stebbings: Sam and Brad said the other day that models just aren't actually that good any [00:13:00] like yet, and they need to improve a lot in quality. What are the largest constraints or bottlenecks on model quality today and what needs to change for them to improve?

Arthur Mensch: I think the data quality is, is a constraint. how do you leverage the entire world knowledge and ensure that the model follows a certain path toward learning more and more complex things? Uh, that's very important part, and I think it has been a neglected part.

There's obviously compute, but, uh, given the amount of data you have. We have at hand, compute is already running into, is no longer the bottleneck. The

bottleneck is more the data at that point, if you look at text to text models. Uh, and so the question is how do you refine the data and how do you feed very high quality data to the model itself, uh, in order to improve it over time?

And I think in, in that setting, it becomes a bit. one bottleneck that is associated to, bringing better model performance is the question of how do you evaluate these performances? You need to have very good evaluation that targets very specific topics. Like you [00:14:00] want the model to be good at, helping diagnosis in, in, in hospital, but in French.

And oftentimes you're a bit out of domain compared to the data you have, and that's where you, you should identify a gap and you should try and fill it out. the pushing the model capabilities become also a question of mapping where they're failing and figuring out ways of improving it. So for instance, they're failing at mathematics, how you improve, their mathematic thinking.

How do you improve the way they demonstrate theorems? the answer to this is very different from the way you answer the question to how do you improve, medical diagnosis in French, for instance.

Harry Stebbings: Will we see large scale generalized models that are able to answer huge swats of very complex problems?

Or do you think we'll see much more vertically specific, smaller, more specific models that are much more vertically aligned?

Arthur Mensch: Yeah, we, we believe that, and actually these vertical models are not going to be out there. They're going to be built by the application makers, because the only way you can.

Make [00:15:00] a low latency model that is super good at a specific task is to, get rid of the general purpose aspect because a general purpose model is a bit bloated. You can think about everything, but if you, you want your model to think thoroughly about a specific topic so that you can. Call it in your AI application while maintaining a good user experience with low latency.

Harry Stebbings: What role do you play in that world? if it's actually in the application layer where you have that specific model creation, where that kind of value accrues, where do you play in that?

Arthur Mensch: It's a very hard job to make a specialized model. Uh, so it's actually very tied to the way you create a pre-trained model.

Uh, and so bringing, uh, the tools that allow to do it, uh, in a foolproof way. So allowing developers to create a customiz model. that are performing very well at their task. Uh, but that doesn't require expert AI knowledge, which is hard to find, uh, is definitely something where we're insisting.

Harry Stebbings: So I'm an investor stay and I'm, I'm pleased that you just said that there will be value accrued at the [00:16:00] application layer.

'cause I look and I worry that bluntly everything is going to. Get steamrolled by some of the players that we mentioned. How do you answer the question of will value accrue at the application layer? And for me as an investor today, Arthur, you know me, how would you advise me

Arthur Mensch: There's two opposing direction. So the first is that the models are getting better and better. So it means that creating a verticalized application, as long as you have the data for it, and a good understanding of the use case you're facing.

It's going to be easier and easier if you have access to the tools that facilitate it. Uh, so that's the first aspect, which would make me think that, uh, the application layer is going to grow thinner and thinner. But then there's also the fact that the moles are, are getting, uh, cheaper and cheaper because we manage to compress them, because we make a lot of improvement on their efficiency.

And so that means that. Effectively this plus the competitive pressure there is on the model layer means that, uh, the price around the model, the dollar per intelligence unit, let's say, is definitely going to reduce. So, there's these two aspects of growing [00:17:00] ability, compressed price, which on one side says that the application layer is going to grow thin, and on the other side says that the model part is going to grow thin.

So for us, the, the approach that we're taking is that, The model part is still going to be, big enough and that we need to build this platform on top of that, uh, because that's where we are going to enable all of the vertical applications. That will be interesting for humanity.

Harry Stebbings: How do you think about that positioning and brand?

Because there are other players who are much more direct in saying, Hey, we we're gonna dominate a lot of different verticals and kind of be afraid. How do you think about that? Enabler, two vertical applications. Or not in that positioning.

Arthur Mensch: We are not a ized uh, company. we started Misra to, uh, bring value to developers and to bring freedom to developers.

So, when we started there was basically, uh, one API out there, soon, two. and. The field of generative AI was starting to look like it would be very centralized around a couple of players. And we took this platform approach where the model that we're making and the technology that [00:18:00] we are making, we are allowing developers to own it, to modify it.

and so bringing freedom to developers and AI application makers is, I think, the best way in, distributing generative AI as widely as possible, which is our objective as a company making AI ubiquitous, bringing frontier AI into everyone's head, is the reason why we started.

we did a good job at it. but, uh, this open source part was, I believe, a good enabler for the community and made people realize that they could build very interesting technology by modifying the models themselves instead of depending on the APIs of a, a couple of providers.

Harry Stebbings: Dude, what do AI developers care about? Everyone kind of gets on Twitter and goes, oh, did you see X's performance this week is better than y performance last week. What do they care about? Efficiency, scale, cost? What drives that usage and decision making?

Arthur Mensch: They care about, um, cost for sure. Um, they care about customization, being able to modify the models that will and.

On that aspect, I think we are only scratching the surface of what can be done. [00:19:00] Like the fine tuning aspect that has been, like the Go-to solution is probably a little too low level from what, uh, we should be doing. They care about being able to deploy anywhere, so they operate in a certain space, in certain cloud.

they, uh, might be operating on-prem. They might have some edge devices to deploy to, and they want to be able to put their technology there and so. They also care about portability, which in turn offer data control. Usually LLMs AI becomes very useful when you connect it to knowledge bases or to, anything that is related to certain business.

In that respect, it becomes a very sensitive part of your application because it sees everything, it sees all of the data you have, and so. enterprises for instance,

do care about ensuring that the Coch data they have, is accessed in something that they com they can completely secure. Uh, and that's the reason why we, uh.

Deployed our platform on Azure and AWS for instance, that is bringing the security layer that they need.

Harry Stebbings: We're gonna get to Enterprise. Can I just ask you, does [00:20:00] brand matter in this segment, you know, when we think about building brand, both in terms of developer adoption brand, corporate brand is brand. A large determinant of adoption in this segment

Arthur Mensch: brand, uh, seems to be critical.

Uh, and this is something that, uh, we have learned on the way people use. Certain models because they're known to be good. Uh, you can't afford to evaluate everything out there. and so having some form of, um, community, uh. Botching is super important. the approach we took with, uh, A PHD uh, distributed models has contributed to what I think has become, well at least a known brand.

and we believe that it's definitely going to be important. Brand is important because trust is important in that domain, and open source brings trust in terms of. Provide, uh, some trusted brand.

Harry Stebbings: You mentioned the word open source there. I'm, I'm gonna get to that. I do just want to touch on that. You mentioned cost.

Also, I wanna touch on cost. How, when and who will make marginal revenue that exceeds marginal cost [00:21:00] in LLM based products?

Arthur Mensch: You should be telling me, you, you, I mean, you are, you are the investor, right? You have your own project.

Harry Stebbings: That means I know nothing. Okay. And I can't

Arthur Mensch: tell you who is doing the most, uh, the most margin at the moment.

That, uh, that it's, it's probably going to evolve over time.

Harry Stebbings: Who, who's doing the most margin at the moment?

Arthur Mensch: NP is at that point, the cloud providers are pretty much at cost. Uh, LLM providers, we are not at cost, hopefully, but, uh, the margin that. Are known to be lower than the typical software margins.

AI application makers, uh, some of them, the one that are most used seems to be doing a pretty good margin. I think it's going to be quite a moving space. as I've said, the, the capacity of models makes the, the cost of making an application, lower and lower. I don't think there's any way in which, uh, the marginal cost, and the margin of, uh, the most important part of, uh, of that technology, which is really the foundational layer.

become zero. 'cause otherwise there's definitely going to be a. a fairness problem.

Harry Stebbings: What do you mean by the fairness problem? Talk to me about that.

Arthur Mensch: Usually the value tends to accrue where [00:22:00] most of the difficult part is and most of the defensibility is. it was for a while, it has been on foundational models.

I think it's obviously evolving with time and there's no mode that isn't disappearing or evolving with time. That will remain the part where most of the innovation will will be made, and where most of the. It is a significant part of, uh, of the accrued value. Will the, well, the value will,

Harry Stebbings: is there actually much of a barrier to creating a foundational model company today?

I know that's a really broad, stupid question in many respects, but you have so many different players now and new ones popping up every day. Is the barrier just reducing day by day?

Arthur Mensch: I don't think it is, to be relevant in that space, uh, is a very hard topic. you need to be, uh, dominating on the cost, efficient on the efficiency performance by it front.

And there isn't, there's only a few companies that, that are currently well positioned. Uh, so you can try and do something, but, but if it's not relevant, if it's strictly dominated by another model or another technology. then you have a problem there's a few barriers that are pretty hard to, to face that you need to, to raise sufficient capital, to have [00:23:00] enough compute and be relevant. You

need to have, uh, people that knows how to train models, which is still scar resource. And then you need to have a good brand, because as you've said, it's highly competitive and this is not something that comes outta thin air.

Uh. I think there's still a lot of defensibility on the market, although there is a lot of noise, which is different.

Harry Stebbings: How quickly does the cost of compute go down, do you think? Because if you look at those things, actually, you said cost of compute, access to talent and brand. If we drastically bring down cost of compute, like many think we will very quickly.

You've got access to talent and brand. Two of those. A more doable,

Arthur Mensch: um, the cost of compute reduces over time just based on hardware cost. it reduces around 30% every two years. Uh, if you show low NV roadmap, the other thing that increase is the efficiency of algorithms.

So if you look at the way we train models from three years, three years ago, and the way we train model, uh, today, I think we have probably made something around a hundred times algorithmic improvement. that's probably where most of the gains were actually made, in the last three years.

Obviously the cost of, of compute does [00:24:00] reduce, but it doesn't reduce faster than the more load so. our bet is more on efficiency, where I think there's a lot of, of, uh, improvement that can still be made

Harry Stebbings: given NVIDIA's prominence there, and Nvidia being the one where the gains are, as you mentioned.

Is bluntly one of the single most important things, not simply the quality of your relationship with the core provider being Azure or being Nvidia, or being one of these players. Is that not the core determinant of success today?

Arthur Mensch: I guess it's an important aspect.

there is a strategic dependency from the AI layer on the collaborators. And on Nvidia the competition is hitting up as well. Yeah. but it's, uh, it's effectively important. It's effectively useful, uh, when you develop a software to also know the hardware provider 'cause they can help you out, uh, in optimizing for the hardware.

It's useful when you are selling, your developer platform to enterprises, to bring that platform through their usual provider, which happens to be a cloud provider. So there's definitely some important. Collaboration to be made there

Harry Stebbings: when like Amazon invest like $2 billion in [00:25:00] anthropic or whatever it was.

Is that not just like a trade where like anthropic then spend $1.8 billion on Amazon and return it back to them? Do you see what I mean? Is it, is it not a bit of a misnomer?

Arthur Mensch: It looks like you're round tripping. Yes. Uh, I don't know about that deal particularly, but, uh, it, uh, it makes sense. Uh, from, from both perspective.

Harry Stebbings: Can I ask how does the unlimited availability of open source L LMS impact the answer to the above being marginal cost and marginal revenue? Does it change much?

Arthur Mensch: It moves the value a little higher than the model itself. Uh, it moved the value to the platform and customization part, which is really, uh, I guess something that we're expecting.

And it's, it accelerated that process.

Harry Stebbings: You started off completely like very open source, very much, uh, open to the community. Now you have small models open and then larger ones closed. Am I right?

Arthur Mensch: We also have large models that are open now. depends on, uh, the threshold for small and large, but, uh, the eight times 22 B is actually relatively large.

Uh, by any standard.

Harry Stebbings: What [00:26:00] was behind the decision then to close some models? Is it just a business case where you need to make money?

Arthur Mensch: opportunistically there was, uh, opportunity to grow the business using that asset as something that we were selling. Uh, it's still the case that we're growing, that our business on top of, uh, commercial models in

particular, it was also a good, uh, way of, uh, cementing some, uh, strategic relationships with cloud providers.

And it's going to be to continue to be the case like we. Still intend to, to be a, a leader in the open source part. and to have some unique assets that we can, uh, license and to have some unique platform that, uh, developers can use.

Harry Stebbings: it's hard when you suddenly have, you know, some closed and you start building an enterprise team.

For you as a founder now, how do you think about that balance between a research team and a sales team and making sure that the two cultures. Come together? Well,

Arthur Mensch: I think some important thing is to create, uh, empathy. So ensure that, uh, the science team also understand, the problems that the users are facing.

It improves the science because at the end of the day, the [00:27:00] general purpose technology we're making, I. It's only general purpose if you identify the use cases. So that comes back to the earlier discussion we had. So ensuring that the science team has some relatively direct exposure to the product and to the business team is actually important to make them understand what, where the model is failing and how it could be improved significantly.

and on the other side, the go-to market team has to understand it. It's a very technical sales, uh, sales motion. 'cause you, you are selling not the product, but you're selling something that is going to product. So you need to. Tell the customer, how these things should be used to actually make something that brings value to the business.

and that only goes through strong enablement of the Go-to market team. So it's, I think it's a challenge. Uh, they all don't operate on the same scale. Uh, the science team has cycles of several months. The go to market team, goes faster, shorter cycles, let's say. but I think so far, uh, we've managed to recruit.

Go to market, people that have some technical interest and uh, and technical people that have some business interest. And I think that's how, that's how you ensure that, uh, [00:28:00] you don't have silos at the end of the day.

Harry Stebbings: One of my worries with, bluntly this space as we move into enterprise is that brand matters so much in terms of enterprise actually, and they

already have existing agreements with Microsoft, and I worry that actually product or model quality doesn't matter as much as distribution.

Microsoft just tack on existing. Your clients with new products, how do you think about that as a core challenge? And am I wrong to be worried about it?

Arthur Mensch: So I think it's true. distribution is very important. Uh, a shortcut to distribution is to create demand through open source models.

Do

Harry Stebbings: you think open source. Is ready for enterprise or do you think enterprises are ready for open source and do they care about it enough?

Arthur Mensch: It depends on the enterprises, but some have been early adopters and are using, uh, a lot of budget into, in production, so for sure they're ready enough, in order to bring them to the next level of putting things into, uh, large scale production, et cetera.

I think they're still lacking some product around like managing, correctly load balancing, customizing the [00:29:00] models because you can do it with the a y solutions, but if you want to make it robust enough and scalable enough, it's actually not easy. And if you want to actually increase the quality of the models, the custom models, the recipe are, are a bit hard to to set.

So the most technical savvy, uh, enterprises are definitely ready for it. And there's a few, there's. Actually many, uh, use cases that are in production using, uh, using open source models. Know in order to widen the adoption. There's definitely some tooling to be brought to the market.

Harry Stebbings: Obviously every enterprise today is sitting in a boardroom going, what's our AI strategy? What do you advise them and what questions should they be asking?

Arthur Mensch: Start thinking about how they're going to change all of their products, uh, using AI as a premise, using the existence of, uh, like very clever agents because you can build very clever agents today.

Assuming that presence and working backward to understand the consequence in terms of organization. not thinking about generative AI as a way to, as a way of increasing productivity in a world processing, but rather at the way to change

completely the way you operate your [00:30:00] core business, uh, which usually involve, taking models and customizing them.

to, uh, create the differentiation that you will need in like five years time when everybody will have adopted the technology in its core business.

Harry Stebbings: So my question too is you are in France. I'm in London. We both know that European enterprises do not move very fast. Most do not even have slack today. My concern is that we drastically overestimate adoption in the.

Near future and maybe underestimated in the 10 year, 20 year future. Do you think that's the case here? And do you worry about the lethargy of a lot of enterprises, especially in Europe, in adoption?

Arthur Mensch: I mean, it's a general, uh, phenomenon in, uh, in tech that you always, uh, Overestimate the speed, but underestimate the impact.

I think it's probably occurring today. It's slightly different, uh, in the sense that there's some, uh, executive support for, uh, pushing generative AI solutions even in Europe. Mm-Hmm. So there's some delay compared to the US market for sure. I wouldn't say it's, uh, it's, it's very significant.

It's one year maximum in term [00:31:00] of, uh, delay. The challenge here is that it's a technology that is, that can take many forms and so trying to focus on some specific thing that you can bring to the market that have AI in it, is a prioritization challenge. And so you need to be very strategic around that.

I don't think this is super easy for enterprises. Generally. It'll become easier once they try out, off the shelf solutions a bit more. Uh, once they realize that there are some developer platforms that allow us to do it, without hiring a very expensive and hard to find AI scientist in house.

and so we expect that this is. Going to accelerate in the, in the coming years.

Harry Stebbings: Do you think that we're still just playing in the experimental budget game, or do you think that we're moving into core budgets as well?

Arthur Mensch: It depends. Uh, It's moving into core budget for customer support, for instance, where, like areas where the application of AI is pretty obvious.

it's definitely moving into core budget. it's also at the experimental stage, uh, in many other functions and for core applications, uh, in the industry, the telecom [00:32:00] industry. And in healthcare, this is still. In the playground, but I, I think it's going to evolve in the next year.

Harry Stebbings: Can I ask, as you build out enterprise, it's another expensive thing to build out on top of compute and talent. It costs real money. And I spoke to Paul at Lightspeed before and he was mentioning to me bluntly how much less capital you've raised compared to a lot of your competitors, most obviously OpenAI and Anthropic.

He said, in a world where capital equals compute equals quality of model, how does MSRA keep up and stay relevant?

Arthur Mensch: So the good thing is that capital is correlated to compute then compute is correlated with quality. Uh, it's not completely dependent on it. And as I've said, there's some strong opportunity for, providing models that are the best of their class, they might be sufficient to actually solve certain use cases.

that's where we're playing. In addition to be playing on the scaling part, because obviously you do need to keep, uh, to stay relevant. You need to, to keep your, technical team, uh, motivated and to keep your technical team motivated. You need to give them, [00:33:00] uh, the experimental bet they need, to make new discoveries and to, to progress science.

And that that is where you need to compute. in addition to. Growing the model, uh, across time. I mean, we're growing our compute. Uh, like every companies, we are convinced that we don't need to grow at the same rate, because, there's a lot of barriers that are not compute related, uh, that are appearing on the way and that we're already seeing.

So we think we, we can, we can scale. And we are also convinced that we, on the efficiency front, we are already very well positioned and we are strengthening that position.

Harry Stebbings: What are the biggest barriers to Misra today?

Arthur Mensch: We've had a few delays with our compute providers for sure.

that has been a barrier, so. the last answer to your question is to be taken with a grain of salt. We are still bottlenecked by compute for sure. Uh, but that's because we don't have many of it. We have 1.5 K 800, which is a few percent I

think, of the capacity of our competitors. Uh, and so that's definitely a bottleneck that is going to improve significantly in the coming months.

Harry Stebbings: Can I ask ly, was it a mistake for you to not scale that quicker? with the benefit of hindsight now, would you wish you'd scaled it quicker? [00:34:00]

Arthur Mensch: You can't really scale that much quicker. 'cause you, you can't raise like two, 2 billion like, uh, on, on the seed round. I mean, at least you couldn't in 2023.

Maybe you can, uh, today, but you can only hire that fast. you can only scale your infrastructure to manage more GPUs that fast. And you can only raise capital that fast. So. there's some acceleration constraints that are pretty hard to fight and that are pretty much the first principles of, uh, of starting a business.

Harry Stebbings: You mentioned about the scaling constraints and, and, and cash. Does it matter where your cash comes from? Does it matter if you have European funded, Saudi funded, US funded? Do you think that matters?

Arthur Mensch: I guess governance matters. what is important for a young company like us is to be under, under the control of, of the founders, uh, because there's a lot of things to be invented and, and the vision, uh, can only be carried by them, We have, very good governance term, a very simple and clean governance that makes us a, a for-profit company. Growing a business to actually push the science frontier. this is something that we're very attached to, being [00:35:00] able to, control the company, uh, leverage, our, funding partners, uh, appropriately to grow in different parts of the world.

In the us, in the eu, it has been critical as well, So it does matter in the sense that, uh, we want to have partners, that are, uh, supportive. and long term, because we are in the field that is fast moving, where we don't know yet exactly where the value will accrue.

And so being flexible, and being smart, is definitely a requirement. Uh, when you raise money,

Harry Stebbings: would you take money from Saudi or China?

Arthur Mensch: Good question. Uh, it's, uh, it, uh, it depends, uh, depends on the term. China is a bit hard for us. It's even hard to operate in China. I mean,

we don't operate in China 'cause you can't really operate in US and China without being like a very, very large corporation.

and so you need to make some choices.

Harry Stebbings: what chance do you think that Europe has in ai? I know it sounds. Deterministic and defeatist. And so you might be going, oh fuck, Harry, shut up. But it's like what challenge do you think Europe has in ai and what does it take for us to stand up as a serious AI industry with [00:36:00] Europe?

Arthur Mensch: I guess the chance it has is that it's a revolution it's changing the way we do software. And so as every revolution, it opens a lot of opportunity for. for new actors, and there's no reason why there shouldn't be an actor that is, that was created in Europe, that could grow pretty fast.

And that's the mission that we, we gave ourselves. we have the talent, capital. It can cause, uh, oceans, uh, without, without too much problem. we have the market, the market is more fragmented than in the US for sure. The ecosystem, the digital native ecosystem is definitely smaller, but it exists and it's growing.

there's local opportunity for business development. on the talent side. we can hire 23, 24 years old people that we can. Onboard in four months and they operate as well as any software engineer in the valley. So people are quite talented here. And so we ma if we manage to keep them and to, uh, to convince them not to go to the us, we have a lot of opportunities

Harry Stebbings: when we look at computers, mobile, cloud, the kind of core technology shifts.

The way that it's worked is Europe has kind of [00:37:00] seeded control to the US and then just taxed US companies for access to our citizens. If one's being defeatist. Is it different now?

Arthur Mensch: I mean, Europe is paying the, the, the price of not, uh, setting up a VC system in the sixties. but setting it up, uh, like 40 years later or even 50 years later, and so, and by,

Harry Stebbings: and by, and by the way, the dirty secret is that the VC ecosystem in Europe is US funded.

Arthur Mensch: yeah, it was, I think it, is it still, uh,

Harry Stebbings: honestly, in large part, yes, there's government institutions which are backfilling it, but largely backfilling it with bad players who aren't very good. But the best providers in Europe are largely US funded by top US institutions.

Arthur Mensch: Okay. I think, yeah, as I've said, it takes time for an ecosystem to build, so you have layers.

Of entrepreneurs and investors that, uh, stack on top of each other. the US has 60 or 70 years of venture capital, uh, investments, I think Europe has only 20 years, I mean, it takes time. It takes an incompressible time to build an ecosystem. it takes also some willpower.

[00:38:00] Uh, and I think now we are seeing that willpower. We are seeing entrepreneurs creating companies. We're seeing this is like you not going to the us. everything is positive. It just takes time. I'm adamant that, uh, we'll manage to do something interesting

Harry Stebbings: on the engineering side.

Do you feel like you have the depth of talent pool to hire from as you scale that you do?

Arthur Mensch: On the engineering side, on the AI side, we do, we have a team in the US though, which is, uh, working on like specific topics. like for senior AI scientists, you find them more in the valley than, uh, than in France for junior AI scientists.

Uh, there's a wealth of talent, uh, in France, in Poland, in the uk. I think one of the strength of the area.

Harry Stebbings: When you were raising money, was it very different speaking to European investors versus US investors?

Arthur Mensch: I guess in the seed round? No, it wasn't that different 'cause it was a seed round, for the series A, which was a bigger round.

It was, uh. European funds weren't structured to, uh, do the kind of deal that we were, uh, proposing. we didn't even have a lot of conversation 'cause they, they just couldn't get their head around the investment that needed to be [00:39:00] made, whereas we were approved new company. Yeah, I think what is lacking, and it's related to.

To the ecosystem part, in Europe are, are growth funds, uh, that are able to, take huge bets, uh, with lots of conviction and that in turn should improve over time, especially if, If we manage to, uh, to use, uh, European wealth and channel it more into that growth funds, than it's today.

Harry Stebbings: I think you have more hope than me on that one. That is not gonna happen. We are not gonna see many more European growth funds be built in the next few years, for sure. Not in the next. Three to five, I

Arthur Mensch: think. Yeah. It, it hinges on a few political decisions.

Harry Stebbings: and I think, I think it hinges on supply of capital and, and belief in a future European ecosystem that can contend with other large ecosystems.

Arthur Mensch: It's a chicken and egg problem, this could be nudged into the right direction if, uh. if politics wants to do it, if a couple of companies show that you cannot actually have, uh, companies that grow fast in Europe, and that's what we're trying to do, I'm not too pessimistic.

I find you too pessimistic.

You should come to France. I think, yeah, you would get more optimistic.

Harry Stebbings: Do you know what? If a [00:40:00] Parisian is telling me that I'm too pessimistic, then shit, I really need to be more optimistic. my, my question to you is like, you know, you just mentioned that the speed of scaling, hardest thing, dude, to scaling with your company at the same speed.

What was the hardest thing about. Yourself scaling as CEO with such speed of scaling of the company,

Arthur Mensch: I mean, we are learning on the job, it's, uh, effectively you have, uh, organizational challenges. How do you, how do you ensure that, uh, 45 people communicate well together?

how do you manage your time in terms of representation? Time. in some of business development time, 'cause we're still at the stage of the company where we get involved a lot in the, in the deal making aspect. And how do you Yeah. Ensure that, uh, you set proper directions and, maintain the team in the state of, uh.

tranquility despite, uh, the amount of noise that there is on the competition, on the competitive side, the fact that the direction is obviously going to be changing over time because there's a lot of uncertainty in, in, in that field. So this is, I think, the hard part. I'm not, I don't think I'm doing it [00:41:00] properly, but uh, we are actively, uh, trying to find sources of information to learn new things, let's say.

Harry Stebbings: If you could call yourself up to like the night before you became CEO and founded Misra and give yourself some advice with the now knowledge that you have. What would you say to yourself, Arthur,

Arthur Mensch: maybe stage a bit more the product development in go to market. Development. we did start the Go-to market motion at the time where we had absolutely nothing to sell.

Uh, it did work out. Uh, it did create some brand awareness despite the absence of anything. I think it might have been slightly simpler, to stage things, maybe a little more, developing the product a little before developing the Go-to market. But since it's such a fast moving field that we did start everything a bit together with.

Uh, some organization that was a bit lacking and now we are, uh, we are solidifying it, uh, on the fly. it has worked out. It hasn't been optimal for sure. In hindsight, you can always give you, I could give me like a few tactical advices on, uh, who to hire or when. Generally, I think the strategy we had one year ago hasn't [00:42:00] changed much.

Uh, we did realize that we need more capital. and that, uh, we could not operate only from Europe and that we needed to go to the US very quickly. Those were findings that we did on, on the way. Uh, I don't think that they would have helped. It would've helped that much to know it, uh, a year ago.

Harry Stebbings: One. Do you feel like you have enough cash now?

Arthur Mensch: Uh, I guess the startup is always fundraising. It's, it's a field where for the years to come, uh, the investment are going to exceed the, the revenue by design because you do need to scale and you do need to stay relevant, uh, as the frontier company.

So effectively there need to be some investment. the revenue is ramping up. So there will be some, some revenue to reinvest. But today and for the years to

come, the speed for developing research should be faster than the speed at which you can, uh, develop your, your go to market.

Before

Harry Stebbings: we do a quick fire, when you look at the landscape today, which competitors did you do you most respect and admire?

Arthur Mensch: I mean, they all delivered, uh, we were surprised by cohere recently. Came up with new good models. [00:43:00] and I think that was a surprise for us. And obviously, uh, open AI and Entropic and my friends at Google are also doing a good job.

So, uh, it's a, it's a competitive landscape, and we respect all of them. We also all work, uh, in the same direction. I, and eventually we've, uh. The same, uh, higher goals. So, uh, it's great, to have respect for, uh, one another.

Harry Stebbings: Is it too late to start one? Now we see like holistic starting now. Is it too late?

Arthur Mensch: Holistic. I know them well. is it too late? I, I wouldn't recommend going into the foundational layer business. I know some, uh, didn't recommend to do that, uh, one year ago, and we did, and it seemed to have, uh, so far, uh, changed a few things. So, I think it would be arrogant for me to say that there's no chance for a new competitor to arrive in Pitus, for sure.

Harry Stebbings: Listen, my friend, I wanna move into a final thing, which is just a quick fire. So I say a short statement, you give me your immediate thoughts. Does that sound okay?

Arthur Mensch: Yeah, let's do that.

Harry Stebbings: So what worries you most in the world today?

Arthur Mensch: Global warming. There's the race of, uh, the planet hitting up [00:44:00] and us finding solutions for it.

Uh, I think AI is part of the, of the solution. brings more control, it brings, uh. Potentially more efficiency in some of the processes, but there's effectively a race for survival. So I think this is something that we should be a bit more aware of.

Harry Stebbings: What have you changed your mind on most in the last 12 months?

Arthur Mensch: I think I've changed my mind on a lot of management premises that I had and that I had never tested the, uh, enroll.

Harry Stebbings: What was the biggest one?

Arthur Mensch: Transparent feedback is actually super useful for a company. And so, operating in a almost fully transparent manner, has helped out, uh, growing at with a breaking

Harry Stebbings: What element has been the most unexpectedly challenging in the scaling of Misra?

Arthur Mensch: The amount of demand that we had to manage, uh, which is, uh, too high for, uh, for what we can handle. The brand success, the fact that people knows us, was a bit unexpected. We knew that it would be noticed. Uh, we had no idea that, uh, people would start, uh, using us, [00:45:00] uh, that fast.

Harry Stebbings: What do you do to calm down? You have a lot going on now, Arthur, and you have a lot of that expectation and cash on your shoulders. What do you do to just,

Arthur Mensch: I run, I cycle, I think my partner will yell at me, but I, I try to take care of my daughter.

Harry Stebbings: Okay. You've recently become a father. What do you know now that you wish you'd known when you first had your daughter?

You know, very recently,

Arthur Mensch: I had no idea that, uh, you needed so much energy to, uh. to care for, uh, for a small children. For a small children.

Harry Stebbings: where do you think AI will take the world in the next 10 years? Like, what does the future of society look like in a world where AI is embedded into everything?

Arthur Mensch: Well, it's changing the way people work, uh, significantly, in the sense that it requires to be, uh, more creative and to bring more value, beyond what can be automated. So it's a very structural change, on the job market, which means that there should be some adaptation that are taken, pretty quickly, uh, in training, in, uh, [00:46:00] education, that people can get a sense of what is going to be expected from them in their daily job, assuming that there's some AI out there.

Harry Stebbings: Do you think the fears of job replacement are grossly over exaggerated?

Arthur Mensch: I think they are. I mean, depends on who you're speaking to. Uh, I think job are going to be displaced for sure. Uh, some will be replaced, uh, some will open up. we're just trying to move humanity to higher level of, of traction, so we cannot talk to machines.

and machines can understand an answer, uh, in a human-like fashion. this is not so much of a bad change compared to what we were doing with. Computers. I think what's happening right now is that probably the, the speed in, in our elevation toward higher abstraction level is probably occurring at a non, at a unmatched uh, rate.

Uh, in the story though, that means that the society adaptation is going to be more challenging. It needs to be anticipated.

Harry Stebbings: Final one for you. We do a show in 20 34, 10 years time. If everything goes right, where's msra? Then

Arthur Mensch: Msra has some, uh, [00:47:00] very relevant models, uh, commercial and open source. Uh, and it has a very strong developer platform that allows to do, everything that you need to create your AI application.

Be a good achievement.

Harry Stebbings: Arthur, listen, I've so enjoyed doing this. Thank you for putting up with me going in many different, fast moving directions. You've been incredibly patient and a brilliant guest, so thank you so much my friend.

Arthur Mensch: Thank you for hosting me.

Scarlett 2i2 USB-5: We're all trying to grow our businesses here. So let's be real for a second. We all know that your website shouldn't be this static asset. It should be a dynamic part of your strategy that really drives conversions. That's marketing 1 0 1, but here's a number for you.

54% of leaders say web updates take too long. That's over half of you listening right now. And that's where web flow comes in. Their visual first platform allows [00:48:00] you to build, launch and manage web experiences fast. That means you can set an ambitious marketing goals and your site can rise to the challenge.

Plus web flow allows your marketing team to scale. Without relying on engineering, freeing your dev team to focus on more fulfilling work. Learn why teams like Dropbox, IDEO and orange theory, trust web flow to achieve their most ambitious goals today@webflow.com.

Scarlett 2i2 USB-8: I'm speaking of incredible products that allows your team to do more.

We need to talk about secure frame.

Scarlett 2i2 USB-1: secure frame provides incredible levels of trust your customers through automation, secure frame, empowers businesses to build trust with customers by simplifying information security and compliance through AI and automation.

Thousands of fast growing businesses, including NASDAQ angel list doodle and Coda trust, secure frame. To expedite that compliance journey for global security and privacy standards such. Such as SOC two ISO 2,701 HIPAA, GDPR, and [00:49:00] more backed by top tier investors and corporations such as Google Kleiner Parkins.

The company is among the Forbes list of top a hundred startup employees for 2023 and business insider's list of the 34 most promising AI startups for 2023. Learn more today@secureframe.com. It really is a must.

Scarlett 2i2 USB-3: And finally a company is nothing without its people. And that's why you need remote.com. Remote is the best choice for companies expanding their global footprint, where they don't already have legal entities. So

you can effortlessly hire, manage and pay employees from around the world or from one easy to use self-serve platform.

Plus you can streamline global employee management and cut HR costs with remote it's free HR. I S and Hey, even if you are not looking for full-time employees, Remote has you covered with contractor management, ensuring compliant contracts and on-time payments for global contractors? There's a reason companies like get lab and door dash trust, remote to handle their employees worldwide.

Go to [00:50:00] remote.com now to get started and use the promo code 20 VC to get 20% off during your first year Remote opportunity is wherever you are.