

Building better NLP models with TitanML

As part of our quick fire questions series – or QFQs – we spoke to Meryem Arik, COO at TitanML about the power of NLP and deep learning models, research-grade techniques and taking the time to recognise the impact our the team's work.

ChatGPT has been an amazing demonstration of the power of NLP and deep learning models – it has captured the public and business imagination of what can be done with the latest generation of AI.

However, large language models like those behind ChatGPT are very difficult for businesses to deploy into their workflow. Here are the major problems that businesses currently face when trying to deploy NLP in their businesses:

1. Maintaining data sovereignty. OpenAI / APIs are the easy route, it's very easy to design applications using them, however most large organisations will not use them because they lose control of where their data goes, they can't control uptime and they don't own the key IP in the application. So they need to build models in-house.
2. Building in house models is very difficult. Given that you can't use an API you have to build in house – and this involves a lot of difficult infrastructure and academic challenges that requires a high level of expertise and time. We have solved the MLOps challenges involved in training these kinds of models – so the only thing that the team need to do is define the business problem and collect training data, everything else is managed by the TitanML platform.

3. Language models more generally are expensive and slow. Given that you have build a model in house – it is likely to be too slow for your given application and / or too expensive to run at scale. We use compression technology to make these models much much faster and cheaper to inference for the task that you care about – you do not need an understanding of this technology, it just happens as part of the platform.
4. The hardware required to run these models is scarce: You need powerful GPUs to inference and train these language models – these are not in good supply, they are expensive and difficult to get your hands on. We can deploy the same model on much much smaller and cheaper hardware, freeing up availability of the more expensive and powerful hardware when it is really needed.

TitanML is an NLP development and compression platform – we allow businesses to build NLP models that can run on low powered hardware quickly and easily. These models are very highly performant but are small and fast enough to be able to be built and deployed in house.

The USP of the TitanML platform is the focus on deployability – we allow users to make models that are efficient and scalable, making them easy to deploy (rather than just building flashy demos). We have proprietary and state-of-the-art compression technology which our clients are able to benefit from to create very fast and highly accurate models that run on smaller and cheaper hardware.

Our target market is enterprise deep learning teams – we have received significant interest from financial institutions. But we are seeing an expansion of these target industries as the market rapidly grows.

We have opened up parts of the platform for personal and academic use.

How has the business evolved since its launch?

Our business has really come a long way since we started. Initially, we were all about research, with our founding team being academics. But as we've grown, we've evolved alongside the market need to focus on delivering these academic techniques to end users.

While we continue to employ research-grade techniques and engage in cutting-edge research, our primary goal is to help businesses overcome the challenges of deploying AI and deep learning models. This approach sets us apart from others who contribute to the growing number of research papers and GitHub repositories without giving much thought to how it helps the end user with their

business problem.

We've put a lot of effort into making our tools user-friendly and adaptable to different businesses' needs. So, while we've kept our academic roots to be able to promise best-in-class results, we're now all about making AI and deep learning more deployable and useful for everyone.

What is your favourite thing about being a founder?

My favourite thing about being a founder is being able to point to things that exist in the world only because of our team. That ranges from everything from the software that we have built to the products that have become viable because of our technology.

Witnessing the impact our work has on the world and knowing that it exists solely because of our team's dedication and collaboration is a source of immense satisfaction and pride.

Which founders or businesses do you see as being the most inspirational?

My parents – they quit their jobs and built a business bootstrapped while I was in secondary school – it has now grown to around 100 employees without ever taking external funding. I saw the ups and downs of business building every day at the dinner table and it was their achievements that inspired me on this journey. They will forever be my inspiration.

Which other figures in your life inspire you?

My cofounders – they are incredibly impressive all while being endlessly kind, I am very lucky to be able to work with them every day.

What has been your biggest business fail?

No failures, just learnings. The only thing I wish we did differently was start earlier – we've had to make all the mistakes that we've made to get us where

we are.

What are the things you're really good at as a leader?

I don't know about leadership in particular, but one of my strengths that my co-founders often talk about is my relationship building and empathy. I love spending time with clients, partners and our team to really understand their motivations and ways of thinking to figure out how I can support them.

I think that female founders may have a unique advantage in this area, and I think being able to really listen to people is an incredibly powerful skill.

But what is really excellent about our founding team, is Jamie, Fergus and I all have very different strengths – this makes us very well placed as a leadership team to build TitanML to its full potential.

Which areas do you need to improve on?

Delegation, this is something that I am working on – it's not as easy as it seems though...

What's in store for the future of the business?

We are currently in building mode – we are laser focused on building and iterating towards product market fit. This is a really exciting process considering how quickly the AI industry is moving.

Our long term goal is for TitanML to become 'market standard' for how enterprises build AI models. We measure success by how many models in production went through the TitanML process.

What advice would you give to other founders or future founders?

Build a network and constantly be iterating and learning.

And finally, a more personal question! We like to ask everyone we interview about their daily routine and the rules they live by. Is it up at 4am for yoga, or something a little more traditional?

Nothing as wholesome as 4am yoga unfortunately...

My routine changes fairly frequently depending on if I've been travelling and which timezones I'm working with. But I'm in the office pretty much every day - starting at a reasonable time but typically finishing fairly late. Unlike a lot of people my age, I don't really drink but I do go to a lot of industry events which gives me a nice dose of socialising.

I try to go and see my parents and my family dog (Hubble) around once a week to unwind over the weekend - there is nothing like a walk and a pub lunch to clear the head!

Meryem Arik is COO at [TitanML](#).

Article by MERYEM ARIK