

URL: <https://stvp.stanford.edu/clips/data-is-the-new-code>

Scale AI founder and CEO Alexandr Wang explains his motivation for starting his venture. He explains the essential role of better data in building better AI and machine learning systems.



## Transcript

- So our mission is to accelerate 00:00:04,970 the development of AI applications.. We believe that AI and machine learning is, if not the most important technology of today, one of the most important technologies that's going to enable huge amount of goodness for the world and just enable the world to operate significantly more efficiently and effectively and enable new customer experiences.. And our vision in accomplishing that is building the most data centric infrastructure platform for AI and machine learning.. And the real insight that we have or the thing that powers everything that we do is this thought that data is the new code and that the thing that will dictate the performance of these machine learning systems and these AI systems in the future is actually the datasets and the data that they're trained on much more so than the code that is written to power them.. And so, if we're to boil it down to two a sentence, it's the better data results and better AI.. And we've taken that sort of core idea and use that to build out an infrastructure platform to power a large number of... A large swath of this sort of like AI ecosystem or the sort of like AI use cases out there.. So we originally started with data labeling or data annotation, which is the problem of converting sort of raw data feeds to useful label tag data that we can actually use to train large-scale machine learning systems.. We actually started in autonomous vehicles and then since then we've scaled across a variety of different industries like you mentioned, from e-commerce, to financial services, to the government, to work with large tech platforms and sort of everything in between.. And what we do with these customers is we help them not only with data annotation data labeling, we help them with data management, we help them build out actually algorithm..

And so with some customers, we provide algorithms directly to them for stuff like document automation or e-commerce AI or in the government use cases.. And so we've been able to expand pretty rapidly into a huge number of product areas into a huge number of verticals but again, essentially at the core, it's all been powered by this concept that, better data results in better AI and the most valuable thing that we can do to ensure that we have great AI systems in the future is to build incredible systems, to be built great data sets.. So that's what we do.. - God, that's amazing.. 00:02:24,160 And the fact that you came across the seedling of that insight when you were 19, you weren't even 20, I'm actually curious if we'd just go back now, knowing what ScaleIO has grown to.. When you were 19, what gave you the confidence in this concept that data is the new code that this.... That was the basis, that was the soil for the growth of beautiful AI algorithms in the future.. And what gave you the courage to jump in and do this? - Yeah.. 00:02:57,410 So it's sort of a few things.. So when I was actually I was at MIT, I was studying AI and machine learning and this was the year when Google released TensorFlow and DeepMind released AlphaGo..

So sort of this like big seminal moment for AI and machine learning and it really felt it was this.... I actually remember there was like this reporter who was walking around MIT campus and was like interviewing MIT students to see what they thought about AlphaGo.. And so it was like.... It was very clear moment like felt like AI is actually gonna happen, it's gonna be big.. And then I remember both in some projects as well as like in some school projects as well some side projects.. In one of these side projects, I remember very viscerally it was like, I wanna build a camera inside my fridge that would tell me when my roommates were stealing my food.. And I remember very viscerally it's like, hey, there's all these great neural networks.. They're really, really cool, but at the end of the day, the algorithm is only as good as the data that it's trained on.. And so it was like, hey, this is a.... This is going to be a critical almost pillar of whatever AI looks like in the future..

And I looked around and I realized like, hey, this is a big problem that there aren't actually that many people trying to solve or there aren't that many people focused on solving this problem.. And ultimately the thing that gave me conviction was frankly the I'd seen sort of the success stories in sort of the years prior of platforms like AWS, which enabled everybody to build these large scale internet systems or sort of like websites and large-scale internet platforms.. I had seen the success with platforms like Stripe for it to enable payments and enable sort of like you to build businesses on the internet.. And so ultimately the realization was kind of, hey, you know what AWS has done for the cloud or what Stripe has done for payments, there's an opportunity for a company to do that for AI and unlock this huge amount of potential for the technology by solving one of these critical pillars.. And the pattern recognition was that, hey, data centric AI was going to be just as important as some of these other sort of like foundational pillars.. And so that's really what kind of got me excited at the time.. And honestly speaking, I didn't have all the answers, I didn't know necessarily that this idea was definitely going to be as important as I think we believe it is now or I didn't necessarily know that it was gonna be as exciting as we think it is now but the sort of fundamentals were there for it to be like, hey, this is certainly worth exploration...