DiffuseTap: Data & Crypto... Enough for AI?

Last time on DiffuseTap, **Manu Choudhary, Co-Founder of DeFinity,** talked to us about how using AI for risk management can lead to alpha, the need for readily accessible data in the crypto industry, and the developing use cases for AI in DeFi.

Want to make friends from the Diffuse Fund Ecosystem? Email contact@diffusefunds.com.

DiffuseTap

This networking session is part of our weekly virtual events series. Networking (you'll bump into at least a dozen high caliber fund managers) meets purposeful (you'll tap into brand-new sources of ideas)... straight from your armchair like a boss.

Meet the Speaker

Manu Choudhary is the Co-Founder of <u>DeFinity Markets</u>, CEO of <u>DMALINK</u>, and Non-Executive Director of <u>Axyon AI</u>. With over 20 years in U.K. finance and fintech, he is expert in emerging financial technologies – from electronic trading, algorithmic trading, permissioned and permissionless blockchains, and decentralized finance to CBDCs, quantum computing, and deep learning AI.

LinkedIn: <u>@manuchoudhary</u>

About Diffuse®

We are an alternative fund platform offering differentiated investment products. From digital assets to VC funds and beyond, we identify green field investment opportunities we feel will have market beating returns and turn them into professionally managed funds. For more information, visit <u>www.diffusefunds.com</u>.

KENNY ESTES: Manu, would you mind introducing yourself and what you're up to at Definity?

MANU CHOUDHARY: Sure, thanks Kenny. My name is Manu Choudhary. I am the CEO of DMALink, the co-founder of Definity Markets, and NED at Axyon AI. DMALink is an institutional platform for banks and hedge funds to trade precious metals and FX . Definity Markets is an institutional platform for cryptocurrencies and eventually, wholesale central bank digital currencies. Not retail. I'm not a big fan of those. Axyon AI is a leading European FinTech with expertise in Deep Learning AI for asset managers and trading firms.

Prior to joining the world of fintech and leaving TradFi, I spent seventeen and a half years working in cross-asset derivatives sales. More recently, I previously worked in FX and commodities derivative sales for Lloyds Bank and Barclays investment bank. I've been a big fan of tech pretty much my entire career, and my life changed in 2018 when I discovered what AI truly was.

AYLA KREMB: Amazing. We're going to toss you right in the deep end. As a starting point, data and the availability of data is a bit complicated when it comes to the crypto universe. So, what does the data landscape look like right now in crypto? What's available and what's not available? What's paid, what's free, and what data do you make yourself?

MANU: It's very fragmented. It's not like TradFi where you have awesome data available all the time. It's very difficult to get good quality data. Having said that, one of the big differences that I see between TradFi and crypto is that there's no such thing as on-chain analytics or on-chain data.

For traditional markets, when you do a trade in equities or in fixed income for example, there is a transactional exchange which will be reported either to the regulator or to a repository within 48 hours. You don't necessarily get it in real time, like you do with digital assets. That's one of the big differences.

In terms of where the data set is, it's not that different to FX back in the 90s. You've got big bid offer spreads, you've got a lot of fragmentation in terms of geography and pricing, there are almost a billion venues that you can trade on. You will see a lot of mergers. That's how I think this is going to move forward as well.

Let me give you a bit of an order of magnitude difference. In our FX platform, from a data perspective, we have anywhere between 10,000 to 50,000 pricing updates per currency pair per second. We have 63 currency pairs, plus gold and silver. That is a huge amount of data, and that is many multiples of what is available in the world of crypto as a market full stop, let alone one venue.

I think we have a lot to do between now and the point where we can maximize the availability of data. But even with what we have, there are a lot of very smart techniques you can use to exploit this and take advantage.

KENNY: Okay, let's go there. This is a good framing question from Mr. Sorenson. When I hear AI, I immediately think of data scrubbing, because that seems to be the vast majority of what AI does. It's getting clean data. And then, the actual AI is relatively straightforward on top of that once we have good data. In your words, what is AI? How is that distinct from just data gathering?

MANU: Okay. That's a really valid point. Everyone uses AI. Every time I look at LinkedIn, or Instagram, or Telegram, everyone is developing their new and amazing AI. The reality is that most of it is just automation. That's not AI. Let me give my definition of AI. My definition of AI is the transition from humans telling computers how to act, to computers learning how to act without being told. That is what I believe true AI is, and that is what we use.

In terms of data scrubbing, it depends on what you use. We are fortunate enough that we use the latest evolution of machine learning, which is called deep learning. It's amazing for looking at very rich datasets, but it's also amazingly good at looking at very disparate fractured datasets as well.

You don't necessarily need to do as much scrubbing. And of course, the more you mess with the data, the more you start structuring it. The more you start defining parameters, the more of your human biases are put into that dataset, and that results in slightly jaundiced outcomes.

AYLA: Amazing. As a follow up to that, maybe you could talk about what's possible with AI today versus six months to a year ago, especially in the DeFi side of things. What is your view on that? I think maybe data on FX pairs might be a bit easier to get. How has the availability of unique datasets around DeFi protocols changed?

MANU: There are two things to talk about here. One is getting access to as clean data as possible, and the other is getting proprietary data. But above and beyond both of those, we've got to make money out of them. You've got to be able to execute trades. So, you need to understand quite a lot about the dynamics of liquidity and how that evolves over the course of a day.

There are a lot of factors in terms of taking the dataset and creating something that gives you a signal, and then monetization. There are three separate vectors within that Venn diagram. As the market evolves, the granularity of data becomes bigger, which means you get better data. But you have to be careful with data. More data isn't necessarily a great thing, because what it does is it shortens the predictive timeline.

I'll give you an example. For instance, we did an experiment and we took end of day data, which quite frankly is pretty useless. We said, okay, let's use one target vector. Let's use Bitcoin, for instance. Let's then add a whole bunch of context vectors, which we did. We gathered 126, such as the 10-year yield, the 2-10 spread, the NASDAQ, the S&P, the Euro Stoxx, gold, silver, and so on. And then, we threw that raw data at the AI agents. It came back and gave us a graded response of anomalies.

We built this to help a lot of the banks that were trying to get their head around the crypto market just because it's so much more volatile than the traditional markets. What's interesting about this is that we created a model that 73.3% of the time accurately predicted volatility events 24 hours before they happened. That's remarkable because it was using end of day data.

If you shorten that to, let's say, hourly data, it becomes more accurate, but the prediction or the predictive window decreases to 60 minutes. So, more data is great, but it depends on what you're trying to achieve. If you're a high frequency trader, then yes, absolutely. If you look at Tick Data, you can make predictions 100 milliseconds out into the future. If you're a macro trader, you probably want daily or weekly forecasts. It really depends on what you're trying to achieve.

KENNY: Picking up a question from Eren, what is the actual core strategy? So you get the data, you do the predictive model, and so on. Are you market making around those prices you predict?

MANU: The funny thing is we're not hedge fund traders. We're sales and trading guys. We've worked in markets our entire careers, but we're not running strategies or managing money. What's interesting about this is that we accidentally created a proprietary strategy. I didn't intend to. It just happened when I gave it to my friends who were portfolio managers.

They were trading options and saying "this is amazing! It's got a 4.2 Sharpe ratio", to which I said, "well, that's not really what it's for guys. It's meant to give you a risk matrix." I will post the link in the telegram chat for anyone who wants free access to have a look at it, and any feedback is very welcome. But effectively, what has happened is that we've accidentally created a strategy that generates Alpha.

AYLA: Fascinating. Jared has a question for you. Are you training the predictive model in real time? How does the normalization impact accuracy? Maybe you could get more technical on your algorithms and give us more details.

MANU: Okay. There are a couple of things to explain here. The anomaly detector I'm talking about is not a predictive model. If you imagine a dataset, you tend to have some kind of clustering around a mean. All the detector does is it uses a technique called isolation forest. And just for reference, Google uses random forests when they train their AI agents. What it does is it looks for the distance between that clustering of the outliers, and it gives you a score. The further that data points are away from the cluster, the more likely it was created by a cause that was radically different to the rest of the data.

That is basically what the anomaly score is. It gives you a percentage weighting. It will say, for instance, that there's a 65% chance that Bitcoin will have a volatility event, i.e. move up or down. It's non directional. It's just saying there will be a move in volatility, which is why, again, trading it with options is such a pure way of executing a trade.

When it comes to trading algos, there are several different algo models. We're building a whole bunch of market making algos that are market neutral and that don't really generate alpha. We've also got some alpha generation algos and we have execution algos as well. These are the three cohorts that we're playing with. We use different methodologies. So, when I say AI, we have several dozen AI agents all trained in slightly different ways.

One of the big distinctions I want to make about the way we work and how we train the AI agents, without giving away the secret sauce, is that with most models, working with artificial intelligence relies on competitive methodologies. For instance, Darwinian evolution. We don't use it because we realized that since the 1990s, Darwinian evolution doesn't really work very well. The reason for that is because in the 90s, we had this thing called the Human Genome Project.

Previously, it was thought that one gene equals one protein. Unfortunately, humans have 24,000 genes and 130,000 proteins. So, how does that work? Well, the environment expresses those sequences of genes, required to make the additional proteins, or Lamarckian evolution. What we found is that rather than having the AI Agents working against each other, they work collaboratively and we see how we can tweak the cohort to get the most optimal predictive outcome. Hopefully I answered that question.

KENNY: I think so. Brian actually has a follow up question. You accidentally created a prop trading algo, so kudos to you. And hopefully, that is becoming pretty lucrative for you. But that wasn't your main goal. You were not necessarily looking for alpha returns, but you were more of just predicting volatility and the risk around that. Is that correct?

MANU: Exactly. When we spoke to banks, they were jumping up and down saying, "look, we want to do this but our risk guys won't let us trade because we need to understand what drives this." When we built this, we actually got all 126 vectors. If you want to look at how much of that 70% score is on–chain data, we can tell you specifically what on-chain data (i.e moving assets to or from exchanges), how much of it was related to precious metals, and to equities and so on.

I'm not a big fan of black box models, maybe because I'm not a tech guy. I'm a sales and trading guy. I just want to know exactly what the constituents are. It was purely a risk management strategy. It just so happens that you can make money out of it.

KENNY: Okay, you talked about your institutional clients. when they're looking at your algos, how do they underwrite them? What does that look like? What key things do they want exposure to?

MANU: There are two demographics that we have. Obviously, on the digital asset side we have a small cohort of clients that use our market making algos. But before we dive into that, it's probably worth explaining what the big difference is between regular algos and our algos.

If I put my TradFi hat on, we refer to them typically as quant models. What happens with these models is that when the dynamics of the market change, or there is a fundamental shift in the structure of a market, either they stop working full stop or they stop working as efficiently and effectively. What we have done is used a bunch of AI agents to dynamically modulate those algos.

In terms of retraining the algos and retraining the AI, which would typically take you three to six months, we got that down to a month, and then to a week, and now, we're down to a day. We should be down to about an hour by the end of the year. It won't be real time, but it will be as dynamic as we can have. That's the big fundamental difference. In terms of users, typically we have a couple of crypto projects that use our market making algos for their tokens. On the traditional side, we have a couple of fund managers that use our execution algos for effectively transacting less liquid currency pairs. That includes EM, CEE3 like Czec, HUF and Polish. and the Scandies.

AYLA: Fascinating. One question here is on AI and ethics. In crypto, it might be a little less socially oriented, but let's talk about it more from a risk discovery and risk exploitation perspective. You have insights on a lot of different things that most people can't see because they don't have access to that data, and that bird's eye view that you have. What are your thoughts on AI ethics in crypto?

MANU: The fundamental founding tenant of the businesses that I run was to try and create more open, transparent and more accessible financial markets, be it FX, or precious metals, or crypto. We're big fans of transparency. We're big fans of openness. This is why we have given away a lot of these tools to our token holders. That's so they can have a look at this and have a very unique view on the market.

I think the ethics of it is really about distribution, and how that information or how those outcomes are being distributed. If they've been given away, I think there's no real issue with it because you're trying to increase the overall level of understanding, either within your ecosystem or within the crypto community full stop. That's cool.

Obviously, I've got to put my financial hat on because my CFO will get very upset because we need to generate P&L. So obviously, we have stratifications. Yes, our basic one is available pretty much for free. But we also have ones that are far more accurate and far better, which we principally pitch to institutional clients and charge a fee for. Would I want to give away my hourly signals for free? Probably not. Is it ethical to do that? I think it probably is, because we're giving away the daily signals for free.

KENNY: Okay, you do a lot of AI. If a paradigm shift happens, your signals become invalid. For people who are vetting from the outside looking into an AI type strategy, what are the red flags for them? What would an AI charlatan look like? How can somebody who's not necessarily deep in the weeds come in and try to get a feel for whether this thing is legit or if it's just a bunch of fluff?

MANU: I think it's really about the level of accuracy. I've seen a lot of models that have 90 plus percent accuracy, which is amazing. And then, when you actually deploy them, they're pretty horrific. I think the best way to describe it is the Pepsi Challenge. Get access to signals, watch them for a month or two, and see how they work in real time. I can take the outcome of what we have, I can run a regression analysis, and they will all look amazing. But it's pretty meaningless.

I'm sure there are lots of accounting sleight-of-hand to make things look a little bit more attractive, like adding options into your portfolio of futures to boost your Sharpe ratio. There are lots of things you can do, so you have to be very cautious. Ideally, you want to beta test it. And if people are unwilling to give you free access for a period of time, then perhaps there's something wrong with the strategy.

AYLA: Another question from the audience. How do you separate your ability to analyze risk in different currencies and blockchains, from your ability to analyze other risks such as third party providers that might create a backdrop of theft? You have the insights into the inner workings of the ecosystem. Would you be able to do risk analysis as opposed to trying to chase opportunities as well?

MANU: Yes, absolutely. That's a really interesting question. What you could do, not that we've done this, is you could use the same parameters to look for anomalous data within trading datasets, basically within stock prices, to see if there's any kind of manipulation or fraud. It's not really something that I'm an expert in, but it could be very easily repurposed for real risk management and real risk mitigation.

AYLA: Fascinating. I'm guessing that when we can see some malfeasance bubble up on various social channels and it just starts snowballing, you would probably be able to be an early ticker on that. Are you using any social media data or conversational data for your analysis as well? Or is it more on-chain transaction data?

MANU: No, not yet. We looked at this initially and I think it's a very powerful tool. But when we got to about 70% or 73% accuracy, we thought, "Well, this is pretty awesome. But do we really want to add more vectors?" Here's the funny thing. When we're doing this, we started off with 250 context vectors. Some of them were very, very damaging.

For instance, looking at copper prices, it decreased the prediction to less than 50% of Bitcoin. When we took copper and replaced it with silver, it increased to above 70%. That was amazing. We couldn't really understand why. It doesn't make a lot of sense. What we found was that it's a very iterative process. I think that's the key component to this.

Thank you for downloading this DiffuseTap event transcript.

Sign up for upcoming sessions and check out past features and event transcripts.

Dennis Chookaszian
Corporate Director, CME Group

DiffuseTap: Institutional Grade Governance

Sharing his decades-long expertise on corporate governance, Dennis talked about how to avoid a co-partnership going sour, the problem with overly idealistic CEOs, and the importance of keeping your board in check. Read on

Susan Brazer
CEO & Founder, LionShare Media

DiffuseTap: Media Metaverse 2022

Susan described the 2020 digital media landscape; the evolution of media distribution; how converging, emerging technology points to the metaverse; and the prospect of having an open, decentralized, and free Web 3.0 marketplace. Read on

Raj Mukherjee J.D.
VP/Global Head of Tax, Binance.US

DiffuseTap: Crypto Taxes Decoded with Binance.US

Raj explained the complexities of the US crypto tax landscape, how he built a dynamic tax information system for Coinbase and Binance from scratch, and how investors can profit from crypto without getting caught in a taxation mess. Read on

JOIN US