

## **Market Abuse Surveillance TechSprint (July 2024) video Transcript.**

### **Team 7. b-next**

#### **Delegate 1**

My father told me that when he started using radar in World War Two, it wasn't very good. You basically could see a BLOB on a very low resolution screen. But radar got really good really quickly. Yeah, distance got better, picking things apart one from the other, even underwater, even from space.

The granularity got better, the accuracy got amazing, to the point where radar today can draw you a 3D map of the sky above, so you feel you know what you're getting into. That is, until somebody invents something specifically designed to evade radar. Now, contrary to what Donald Trump thinks, that thing is not invisible to radar. That thing is designed to defract radar waves and blend against the background so that you don't know what you don't know.

We called our project Early Scam, and then in the spirit of AI, we asked ChatGPT to shoehorn an acronym into that, and we came up with Enhanced Aberrant Risk Learning and Use Surveillance Comprehensive Analysis Network. My marketing manager told me that she could put that into a single German word came up with 85 characters. I didn't learn it.

My name is Benjamin Levy. I am the Chief Growth Officer for BNEX, a 30 year old, 30 year old, 30 year young compliance specialist in Northern Germany.

On the line with me hopefully is Sandra Sakarevich from the Ukraine, our head of product. If Plan A goes well, he'll be answering your technical questions. If Plan B goes. If Plan A doesn't go so well, you get the monkey, not the organ.

Growing, cheating or competing is something that I borrowed from your government.

Thank you. Very nicely done. We live in a rule, in a rules based system. Some of the people in this room wrote the rules. Our problem is that the batteries don't live in a rule space system. They can bring along any AI model optimised for profit and for being hard to detect. And we are left with a serious problem of finding a snake in the grass.

If you've been on the Internet anytime in the last five years, you've seen this picture. There's a snake there. Can you see it? It's hiding amongst the grass and the orange peels and candy wrappers.

It's right there. Whoops. So what did we do?

We started by comparing the behaviour of a trader to himself. So you and you a week ago, or you and you a month ago, or you and you a second ago. You and your evil twin, if you want to call it that. That caused a little bit of time travel, and time is going to be an important factor that Andrew will address in just a few minutes, but we had to take that into account.

The second thing we did is compare you to your neighbours. Oops, sorry, can you still hear me OK because I almost broke this.

Thank you.

To compare you to your neighbours, to your reference group traders that behave in a similar manner and at some point you do not behave in a similar matter to them. And this is a good moment to flag you. What did we do?

Let me bring in the organ grinder to describe what we did. Andre, are you on the line?

## **Delegate 2**

I hope so.

## **Delegate 1**

Hello, Benjamin.

## **Delegate 2**

I hope you can hear me. Well.

## **Delegate 1**

We can hear you just fine. Go ahead.

## **Delegate 2**

Excellent. So first of all, hello to everyone. I'm really happy to be here today.

So let's see what we did. But before that, of course we are dealing with a huge amounts of data. And I believe in nowadays it is only about AI to deal with it.

But before you can do anything with data to detect anomalies or something, you should do standardisation or normalisation, classification, data mining, regression and so on and so on. So this all should be in place before you go to in some, before you go to solve some concrete problems like anomalies, problems that we try to identify.

Next one please.

But let's maybe talk about more interesting parts. That is how we did it before that. As mentioned by many colleagues today we all are clear that we are dealing with a multi dimensional non non linear time series data.

What does it mean? That actually means that connection of one data point, like a clear match of one data point to another, doesn't exist at all. So it is always about identifying how one things, like on this chart that you can see with the two data sets how one event actually is linked to many other events and this linkage constantly changes.

So how would you analyse this data? How would you normalise this data?

Of course after all data mining stuff and so one you come to that normalisation and our approach here is to use a dynamic time warping algorithm that really was built for this kind of purposes when two different data sets would be measured in a kind of matrix matrixes and this matrixes that can can later be on used by AI.

In our case, this was about Kanye's neighbourhood model in an unsupervised model that is a very important that no human interaction actually, but AI to define what is what are the normalities, let's call it like this. And for human last step would be simply to put a range of anomalies by a person till that's pretty much all human interaction you have to do.

Benjamin, back to you.

## **Delegate 1**

So this is what data would look. This is not a picture we extracted from the FDA, from the FCA that was not allowed. This is what your data would look if we ran this model on your data and that tells you nothing. The difference between data and intelligence is showing these events that look to be part of the normal lowest regression and are not.

Now are they orange peels? Are they candy wrappers? Are they snakes in the grass? That's what's next.

So what is next?

Well, obviously increase the relevance, minimise the false positives, adding additional AI models, adding near real time processing. That's just a question of adding oomph to the servers.

That is that is it And to create a human readable readout.

So this can be followed with enforcement.