Transcripts: Requirement interview round

Expert 1 at Hawk

Date: May 16th, 2024

Time: 14:00

Place: Microsoft Teams

Part 1: 00:00:00 – 00:27:21

Part 2: 00:00:00 – 00:23:49

Interviewer First of all, I would like to start with your background as an expert in this field. The first question to you would be, could you share details about your hands-on experience with fraud detection, particularly for debit card transactions? What have you experienced there or what are you doing in this area? #00:00:07#

Interviewee Yes. So, as a data scientist, I generally have more than 10 years of experience in different areas. But specifically for fraud or AML, I've been working in this field since I joined Hawk about 3.5 years ago. In terms of specific hands-on experience, I've built numerous AI models, both for detecting fraud and for auto-closing cases to improve workflows. This includes both fraud and money laundering areas for different customers. I can't give you the exact number right now, but it's certainly a double-digit number of customers for sure. #00:00:30#

Interviewer So your experience with fraud detection is primarily with AI. You hadn't worked on fraud detection before you started working on AI models for detection. Is that correct? #00:01:29#

Interviewee That's correct. In my previous jobs, I was not working on fraud detection models. Only since I've been working at Hawk. #00:01:44#

Interviewer Thank you very much. So, let's talk about the state-of-the-art and the challenges in AI-based fraud detection models. What can you tell me about the current best practices in AI fraud detection for debit card transactions, particularly from a user-centric perspective? What is the state of the art currently? #00:01:48#

Interviewee That's a very broad question, I think. But, I mean, there are different aspects to it. So, I would say one big aspect is, of course, the model performance itself and achieving the highest model performance. It really depends on the exact use case and how the data looks and is labelled, because, in the end, the model can only be as good as the labels provided. From a state-of-the-art perspective, the models we are using for fraud detection include a lot of XGBoost, but also more complex models. For instance, we have started adopting a generative AI model where we treat transactions as words. We teach the model a different language, not English, but the language of transactions. The model is then fed tokens representing different attributes of the transaction. This works very well for fraud because fraud often happens in very specific moments. A fraud attack can occur when something changes rapidly in a short time frame, or there is an unusual sequence of transactions. For example, someone logs in with a device ID never observed before or from a country never observed before, and then a transaction is made to a new account with an unusually high amount. This language-based model is very well equipped to address this problem because it processes each individual transaction without losing resolution. It's not like with the XGBoost models that we also work on, where we feed aggregates into the model, like the volume of the last day or something similar. With those models, you lose resolution on the individual transaction level. With this language-based model, we don't lose that resolution. That's why I would consider it state-of-the-art on the modelling side. Another important area is explainability. Because I believe that in our space, not only for fraud models but also for money laundering models and all types of compliance work, it has to be very well explainable. The model really has to give exact reasons for why it made a decision. Only then can we validate the model properly, ensure it really works, and understand how it works. Additionally, it makes the investigation for analysts much faster because they know what aspect to focus on first, rather than just getting an abstract score without any explanation, which could mean anything. I think these are two key areas. The next one, which I alluded to already, is the importance of having a very good long-term process in place that allows you to capture the right labels. For fraud, for instance, it would be good to label not only whether it was fraud or not, but also what type of fraud it was. This way, you can build more specific models and actually ingest more knowledge into the models. #00:02:13#

Interviewer When we talk about recent technological advancements that have significantly impacted the effectiveness of fraud detection systems, I guess one part you already explained is the large natural language processing models. These have recently come up and are really good models to tackle this. But are there any other recent technological advancements that are helping in fraud detection? #00:06:51#

Interviewee I mentioned the generative thing, but what I meant was actually not a language model. We adopt technology like the transformer architecture, but we apply it not on languages, but on transactions. But others have done this in other fields as well. For instance, in bioinformatics, it's very common to apply transformer models to DNA sequences. The same can be done with transactions because, in the end, all of these things, whether it's language, DNA sequences, or transactions, have some commonalities. One is that they are all sequential data. You know, in a way, it's all time series data. Even language can be considered that way because you have one word after the other. That's why the same technology that works very well on the language side also often works very well on the financial transaction side. So we are adopting this technology in a completely different area, and there are other reasons why it works well. All the cornerstones of why it works well for language also apply to finance. For instance, the attention mechanism helps the model separate out the noise from the actual interesting signal in the transactions. Another thing is positional encodings, which help process large amounts of data very quickly, which is super important for us. All the reasons why it works well for language also apply to transactions. I would say this is really something different from natural language processing, it's just adopting the same technologies. To my knowledge, no one else is really doing that apart from us. Then you have the language side that you mentioned, which is also absolutely true. Analyzing the text in the transaction, for instance, the usage text, is super important and useful. We do that as well. For instance, we recently built a model for APP fraud detection, targeting things like elderly fraud. So, for APP fraud, it encompasses various schemes where you trick someone into sending you money by pretending to be someone else. This umbrella term covers all these types of fraud where the fraudster deceives the victim into believing they are someone else. In all of these examples, we use language models to analyze the text. For instance, we use language models to see if the usage text has anything to do with an elderly fraud. In that area, language models are another technological advancement. The last one I would mention is making use of external data sources. This involves combining data from different sources, not only using the data provided by our customers and the KYC data, but also integrating it with databases that provide information about company shareholder structures, relationships between people and companies, and similar details. Integrating these external data sources can be very helpful in detecting fraud schemes. #00:07:30#

[inaudible due to connection interruption]

Interviewer Now I lost you, but it's back. Maybe we should turn off the camera. Maybe that will help? #00:12:14#

Interviewee Yeah, let's turn off the camera. Did you hear me the whole time, or should I repeat something? #00:12:22#

Interviewer Just checking, the last thing I heard was the use of external databases, but then you were gone. #00:12:35#

Interviewee Exactly. Using external data sources for instance, that contain huge datasets for many different companies, you can find out how the shareholder structure of a company looks. You can see if a certain person is affiliated with different companies at the same time. This can be very relevant from a fraud perspective because if you notice that someone is conducting business with their own company all the time and trying to obfuscate that, it could be a relevant fraud signal. #00:12:43#

Interviewer Can you also explain how it works? How does the training process look like when a customer like me comes to you and says they need a model? How does it work? What are you doing then? Maybe also from the learning of the models, what are you doing? What is the process behind it? #00:13:30#

Interviewee The very first step is that we want to understand from the customer what their pain points are at the moment, like where is the need for a fraud detection model, and what specific challenges they are facing. What are the biggest gaps from a risk perspective that they are currently having with their rule set? Ideally, when we build an AI model, it should cover the general purpose, but it should also address the specific issues or gaps they are currently facing. Understanding this is the first step. Additionally, understanding the business model of the customer and the types of fraud that could occur for them is crucial because it varies based on the business model. For example, you would look for different things for a merchant acquirer than you would for an issuer. The next step is understanding the data. There could be data fields that are obscure or harder to interpret because they might use internal codes for some field values. We need to understand what all these things mean. Understanding the data is the next step. Then, we conduct an exploratory data analysis, where we closely examine the data. We look at how different data fields are distributed, how complete the data fields are, and if there's missing data. We would develop strategies to treat the missing data during this phase. After that, we start with the model itself. Training the model is broken down into different steps, almost like a standardized process that we have at Hawk. The first step would be feature extraction. We have a feature store at Hawk with many different fraud signals registered. We can select and modify these features. We define the features in the feature extraction step, run feature extraction, and then run the actual model training. Next, we do model validation. We look at the key

metrics of the model, conduct historical back testing, and examine the types of alerts the model produces on a sample basis. We also perform bias testing to see if the model has any biases built in. Based on the validation outcome, we most likely go back to the feature extraction step and repeat the process until we achieve a satisfactory result. Once we have a satisfactory result, we present it to the customer. We aim to involve the customer as early as possible to ensure the model aligns with their expectations. If the customer is satisfied, we then deploy the model. That would be the last step. #00:14:07#

Interviewer Very interesting. Thank you. What are the most significant challenges facing AI-based fraud detection systems today? I read a lot about issues like you mentioned already – missing data or missing values – but also things like class imbalance, concept drift, and changing customer behavior. I think it's a big problem when clients don't behave as usual or there aren't enough transactions with their card. Real-time processing is also a big challenge. How do you face those challenges? Are there even other challenges you face, and what technology or methods are you using to address these problems? #00:18:02#

Interviewee I mean class imbalance on the fraud side in particular can be very daunting, I agree. This is certainly an issue that we have had in the past as well. First of all, a lot of the typical metrics that you would normally use in machine learning break down a little bit if the class imbalance is too strong. For instance, ROC AUC doesn't really work well. So, you first have to define a proper framework in which you want to estimate and judge the model. What we do is usually contrast the "cost" the model incurs for the customer versus the benefit. By cost, I mean the number of alerts it produces. Every alert is a cost for the customer, because if it's a false alert, they will have to work on it and close it, which takes time. On the other hand, we can estimate how many fraud cases we can catch with the model and also how many we can catch that rules can't catch. This allows us to estimate the benefit from a monetary perspective for the customer. When we contrast these two things, we get a good understanding of the model's usefulness. The benefit should outweigh the costs significantly for the business case to be viable. For the training itself, class imbalance is also an issue, but there are many ways to address it. It depends on the model, but many models offer ways to give higher weight to the underrepresented class in the objective function during training. Another approach is oversampling the underrepresented class or under-sampling the overrepresented class in the training data. We use all these techniques and cross-validate to see what works best. That's on the imbalance side. You mentioned another good point that we also had faced. I forgot now what it was. #00:18:49#

Interviewer Concept drift or changing customer behavior, and not having enough data for a customer because they are a low-frequency user. #00:21:34#

Interviewee Yeah, exactly. I mean, that is actually indeed a very interesting one because if you have an account with not many transactions, then a lot of the features we compute might become a little bit degenerate. Let me give you an extreme example to get the point across. Imagine we have a feature in the model where we check the count of declined transactions versus the overall number of transactions. A feature that tells you something like 50% of all the transactions of this account in the last month were declined. Such a feature will completely degenerate if you have just a single transaction in the account and that single transaction is declined, giving you a 100% decline rate. That's not good, obviously. This is particularly problematic for anomaly detection models because those will almost automatically generate an anomaly if you don't have any safeguards in place. What we do for anomaly detection models is set conditions where we say the account needs at least 10 transactions or something like that to make a prediction. These extra conditions help suppress false positives. For supervised models, it's probably a little less of an issue because you can feed extra features into the model that help it understand it's in a degenerate situation. For instance, in the decline example, you could also feed the total number of transactions in the last 30 days into the model. If the training data shows that a super low number of transactions often leads to false positives, the model will learn that from the training data. However, if the training data does not contain these samples, it's still a problem, and we would apply extra conditions to only evaluate the model if there are at least 10 transactions or something like that. Regarding the other things you mentioned, like changing customer behavior and concept drift, changing customer behavior is actually something that should be alerted for because if an account changes its behavior, that's a cause for suspicion. It might indicate an account takeover or something similar. So, I rather think this is not so much an issue. For concept drift, I think you mean when something significant happens, like the COVID-19 pandemic, and suddenly all transactions look different. Is that what you mean by concept drift, or did you mean something else? #00:21:50#

Interviewer Yes, exactly. #00:25:15#

Interviewee The point is, we monitor. We have monitoring in place in our application where we track key metrics of the models. For each model, we check how many alerts it produces per day and similar metrics. If this changes drastically, we generate an alert within our system. It's not something the customer sees, but it alerts the developers. We also monitor the data going into the model, computing the average feature values over a day, seven days, or whatever is appropriate. If this starts to change a lot, we generate an alert. So, we have monitoring in place to detect concept drift. If it happens, though it hasn't happened to us so far, we would need to talk to the customer and decide what to do. Most likely, we would deactivate

the model, retrain it with new data points collected after the shift, and then bring it back online. #00:21:50#

Interviewer And maybe a follow-up question there. What happens if, for example, you have a big company like us, providing debit cards to many employees who use them for their expenses? They are globally distributed and using the cards everywhere. How do you handle that challenge? Isn't it a big problem to manage all these transactions from employees spending and traveling around? #00:26:42#

Interviewee Potentially, yes. I would need to understand a little bit better how exactly it looks, but it might be a challenge. #00:27:21#

[Recording interrupted, new recording starts]

Interviewee Do you mean a global company where a single person travels a lot with their credit card, or are you referring to the fact that this company has many credit cards in different countries? What exactly is the scenario? #00:00:04#

Interviewer Let's take an example of an employee who needs to travel around the world, like a sales representative. You have many employees doing that, traveling frequently. They are now in France, then in Germany, using the card not always in Switzerland, where the company is based, but everywhere, sometimes just for two days in Germany or even in Dubai or somewhere else. This is different from a private person who mostly uses their debit card in Switzerland. #00:00:19#

Interviewee Right. Of course, this means that all the features based on cross-border signals, like the credit card being used abroad, will become less useful to a certain extent. However, I still think we could do a lot even in that case. Let's imagine we want to detect an account takeover. The signals we would look for include the credit card being used in another country, a very high amount being used, a different device ID logging in, and so on. As you say, the fact that it's now used in another country might not be suspicious in itself, but if it happens along with other factors, like higher-than-usual amounts, new services or goods being purchased, or a new device ID that has never occurred before, it becomes more significant. If these compounding factors are present, the cross-border feature combined with other signals can still be useful. On its own, it might not be useful, but in combination with other factors, it can still provide valuable insights. #00:00:19#

Interviewer Thank you very much. Now, moving on to the next part, let's discuss the desired system characteristics and goals. Can you explain how it works? What does the process look like when I, as a user, make a transaction with my debit card? What happens afterward, and what is the desired process behind it? Also, is it best to combine a rule-based approach or expert approach with AI, or should it be solely AI-based? What is the best system or process behind it? #00:02:42#

Interviewee From my perspective, the best approach would be a pure AI-based monitoring system, with the exception of some rules that are very specific and targeted. For instance, if you want an alert for a transaction in a high-risk country, you can and should use a rule for that. But for other, fuzzier situations, I personally believe that an AI-only approach is best. However, most of our customers are not ready for a fully AI-based system. Most still want to have rules, which is why we offer it and why most of our customers still use it. In that case, we have a rule-based system complemented with an auto-close model. This means that a lot of the false positive alerts created by the rules are automatically closed and do not create noise for the customer. On top of the auto-close model and the rules, we have another model that opens alerts for situations that go beyond what the rules can cover. This approach gets close to the quality of a pure AI-based system but still includes the rule component. Most of our customers end up with this kind of system because they are very used to having rules. Convincing people to just use AI is a process. I think there was another part to your question, right? Could you remind me what it was? #00:03:22#

Interviewer Sorry, I combined them. My other question was, how does the process look? As a debit card user, I pay with my card. What happens then? My information goes to my provider, and then they send the data to your model? How does it work in the background? What is the process behind it? #00:05:24#

Interviewee Exactly. So, you have a core system where all the transactions are processed and integrated into other systems, like ours. When a transaction occurs, it hits your core system, which then triggers a call to us. We receive the transaction data and respond with a decision on whether to block the transaction or not. An API call is made to us, and we send back a response that contains a flag indicating whether the transaction should be blocked. It will be blocked if any of our rules or AI models consider the transaction suspicious. When this is integrated into your overall system landscape, you take our response and, if necessary, stop the transaction. At this point, someone needs to review the alert and decide if it was a true suspicion or a false positive. If it's a false positive, they would close it as such. We then generate a callback that allows you to unblock the transaction in your core system, letting it go through and continue processing. So, that's a bit of how the process looks like. #00:05:45#

Interviewer From your expert perspective, what key features or capabilities should an ideal AI-based fraud detection system include to ensure it is user-friendly and also effective? Are there any metrics that are you taking or that are really important that you take in mind or what are the key features or capabilities of such a system? #00:07:30#

Interviewee I mentioned already the explainability, right? That's one thing. I would also add transparency during the model training time because explainability is what you get during model execution time. But you also want transparency regarding how the model was trained, what principles were used, what features were used, and all of that. There must be a proper audit trail of how the model was generated and proper documentation surrounding that. I would also say what becomes more and more important, and I think it's particularly important in a fraud context, is the whole topic of fairness. Is your model discriminatory? For fraud, where you are blocking transactions, the model really should be fair because it affects people's lives. If they cannot make a transaction and have to wait every time for 10 minutes before they can proceed, it's significant. The model's fairness should be checked to ensure there is no discrimination. Other than that, of course, the model quality must be as high as possible. This means that the cost consideration must be as positive as possible. This depends on the risk appetite of the customer, whether they prioritize recall over precision or the other way around. Is it more important to catch as much fraud as possible, or is it more important to have fewer false positives? This has to be decided together with them. Usually, we give them all of this information, how many alerts we are generating for this threshold, how many of them are false positives, how many fraud instances we can detect, and all of that. They can then make the decision in the end. But of course, the model should be as accurate as possible. So that's clear. At the moment, nothing else comes to mind, but maybe I forgot something. But, at the moment, nothing else comes to mind. #00:07:58#

Interviewer You cannot say, OK, this metric is the best to use to evaluate such models. I mean, in the end, as you explained, it depends on the client, on the customer, what his goal is, or as you said, does he want to have a big risk appetite or is it more like he wants to go on the safer side? So, it depends on the case, the customer, and the model which metric is best to use, right? If I correctly understood. #00:10:35#

Interviewee Exactly. If you have a customer with a very small team of analysts who work on the alerts, they will naturally prioritize having a low number of false positives and might accept that some fraud instances can't be detected in that case. Conversely, if you have a customer with many analysts to work on alerts, it might be very different. In that case, they will most likely want to minimize the number of fraud instances that we cannot detect because every fraud instance we cannot detect is a cost for them in the end. So, it depends on these exact circumstances. We give the decision to customers to make, providing them with as much information as possible. However, as I mentioned at the very beginning of the call today, when we evaluate models internally to see whether it is moving in the right direction, we set up such a validation framework ourselves. We check this to understand how big the customer size is, for instance, and how many alerts they can work on. Then, we determine the number of alerts we can maximally generate for them. #00:11:05#

Interviewer Thank you very much. And also, we already talked about the regulatory and ethical considerations, especially about ethical considerations. Maybe to make a follow-up question, you talked a lot about explainability. How do you ensure explainability in such a system? Is it important that when an alert is created, you present that this decision was especially made based on certain data or features, and based on that, you give a particular percentage that this could be a 90% case? How do you ensure explainability to the compliance user? #00:12:35#

Interviewee So how we do this is when an alert is generated, we provide the reasons as well as to why the alert was generated. The system tells the user things like, this is an alert because the user has never logged in with that device before. It's the first time this device has occurred. It also indicates if the transaction was made abroad while 10 minutes before it was still made in Switzerland, suggesting a potential account takeover. It will note if the amount is unusually high and similar factors. We present this in human-readable text, providing a prioritized list of reasons why it considers the alert to be fraudulent. Additionally, where applicable, we visualize the flow of funds. For instance, in a more complex pattern like a money mule scheme, we would show how the money flows from one point to another, possibly through different money mules, to illustrate connections. We are also experimenting with even more advanced approaches, such as interactive explainability. In this method, users can change certain aspects of the transaction and see how the model outcome would be affected. This helps in understanding the model's decision-making process better. #00:13:23#

Interviewer And also, how do regulatory frameworks influence the development of such systems? Are there any at the moment, or is it still like a gray area? There is nothing regulated, and that's the reason why customers are still not that trusting in such systems and rely on rule-based approaches? What do you think about that? #00:15:26#

Interviewee I think that this, first of all, depends a lot on the country. Really a lot. I'm also not the best person to give a super good overview of all the different regulatory landscapes in all the different countries. I have a quite good understanding of Germany because we have recently interacted with them quite a lot. But for

Switzerland, I am currently not the best person to give insights. However, generally, I think that there's not really a regulatory obstacle that you could not solve. On the contrary, most regulators are opening up more and more to the concepts of AI. In many jurisdictions, there's absolutely no problem. In some jurisdictions, some types of models are maybe a bit more complex and more difficult to sell to the regulator, but I don't think that there's anything insurmountable. I think all of these are problems that can be solved. There's no fundamental limitation put in place by any regulator. I also think the reason why customers don't want to move away completely from rules, preferring a hybrid approach, is not regulatory-driven. It's more driven by their past experience and what they are used to. It's easier for them to accept having control over the rules when using an AI system. #00:15:56#

Interviewer That makes complete sense. So, now to the last question. What advancements or features do you anticipate will be important for future fraud detection technologies, particularly those that improve user interaction and system clarity? Are there any emerging trends or innovations on the way? #00:17:52#

Interviewee Ah, that's a good question. I'm not 100% sure, but here are a few thoughts. I mentioned the integration of external data. The whole entity resolution topic is key, like merging entity resolution with transaction monitoring to not only use it as a tool for investigating alerts but also to produce signals. This helps in identifying overlaps in certain fields, like if someone is using different identities to obfuscate their activities. This closer integration of entity resolution and transaction monitoring is something that comes to mind. We talked about technological advancements earlier, and everything mentioned there applies here too. However, you're focusing more on user experience, right? Combining AML (anti-money laundering) with fraud detection is relevant because these are interconnected. Instead of having these in two different silos within a bank, they should be much closer together. Fraud activities are often followed by money laundering to clean the obtained funds, so bringing these functions and the people working on them together would help a lot. This will also impact the systems, usability, how things are displayed, and how they are connected. Another aspect on the usability side that comes to mind is the development of AI assistants for case investigations. This could be like a chatbot where users can ask questions, query, and get responses. It could summarize alerts, show previous transactions, create plots on demand, and more. This area of AI assistance in investigations is definitely an emerging topic, and some competitors are actively working on it as well. #00:18:16#

Interviewer Perfect. Thank you very much. #00:21:59#

Interviewee No worries. So, it's worth mentioning that at Hawk, we don't believe so much in the whole aspect of assistance and chatbots in the compliance context because of the issue of hallucination, where these systems sometimes make up facts. If investigators start to rely heavily on such an assistant, there's a certain risk. If you want to use this, the system must ask the analyst at every step to fact-check everything. Otherwise, you might become liable for the hallucinations that your model is producing. It's a bit of a dicey area, so we think this should be treated with a lot of caution. At Hawk, we invest a lot in GenAI, but we want to use it to improve detection capabilities. As I mentioned earlier, using the same technology but applying it to transactions or analyzing texts in transactions, like in the app fraud example, we think that's a more promising path than relying on assistants. #00:22:08#

Interviewer Interesting. Perfect. #00:23:27#

Expert 2 at Hawk

Date: May 23rd, 2024

Time: 11:00

Place: Microsoft Teams

00:00:00 – 00:37:34

Interviewer My first part is especially about your role and your background in this area. The first question to you would be, could you share any details about your hands-on experience with fraud detection, particularly for debit card transactions? What is your background in this area? #00:11#

Interviewee My background is generally in computer science, specializing in data engineering topics. I realized I really enjoyed working with data, so I transitioned into a data scientist role at Hawk, where I've been working for five years. I've had the chance to work with massive datasets of transactions, focusing on fraud detection and anti-money laundering patterns. Regarding fraud detection at Hawk, I've had the opportunity to work on various types of fraud, such as merchant fraud, predicting merchant fraud, chargebacks, and card fraud involving stolen cards. Recently, I've been working on APP fraud, specifically authorized push payment fraud. My experience includes both supervised approaches, where we have confirmed fraud labels, and unsupervised approaches, where we don't have predefined labels and need to identify patterns that indicate fraud. Fraud occurs so rarely proportionally, so when we discuss ratios, the infrequency is a significant factor. If we have the chance to obtain fraud labels, then our approach can be completely different and more effective. #00:55#

Interviewer And regarding AI, what has been your experience with it? Have you had any touchpoints with AI since you joined Hawk or even before? #02:23#

Interviewee Well, during my studies, we broadly covered AI, but it was at HAWK where I specialized in it, focusing mostly on transaction datasets. This includes profiling the behavior of accounts involved in those transactions. As I mentioned before, I've worked with both supervised and unsupervised models to predict fraud in live transactions. Detecting fraud in real-time, at the authorization level, adds complexity because you need to prevent financial damage before it occurs. In my experience, supervised methodologies have worked the best, as we can foresee which patterns we want to train for and know how they should look. On the other hand, unsupervised methods usually involve anomaly detection. However, data anomalies might just be pure data anomalies and not represent any sort of threat. For instance, it could be someone purchasing a super expensive item from a very unusual shop, but it might be entirely normal. #02:30#

Interviewer My follow-up question is when we talk about the state of the art and challenges in AI-based fraud detection, what are the current best practices in AI-based fraud detection for card transactions or debit card transactions? And you already mentioned supervised learning, right? #03:49#

Interviewee It doesn't need to be exactly the state of the art. You can achieve good results with supervised methods. With supervised methods, you can tune the model to look exactly for what you want. However, with supervised methods, you have to add some human bias when choosing which risk factors to consider. There's always a chance that you might miss some risk factors. Regarding the guidelines and best practices, the most important aspect is data quality. It's crucial to understand the transaction language of the dataset. Different data providers have different schemas and transaction languages. For instance, in card payments, it's usually a two-step payment flow: first, there's an authorization, and then there's a capture where the money is actually transacted. If you don't consider this, you might count the amounts twice when aggregating for a volume. Understanding what the transaction language means is essential to interpret the data correctly. Feature engineering is also important. You shouldn't just think about the transaction in isolation but consider the whole context. For example, an account might have a specific profile, such as a student or an employee. You need to take into account the transaction history, the relationships with other counterparties, and additional context like POS information, IP addresses, locations, and channels. This holistic view helps in building a more accurate model. #04:18#

Interviewer Can you describe the operation workflow of a typical AI-based fault detection system in card transactions? You also mentioned different sources of data, such as point of sales. Could you explain where you typically gather the data from? As a client, when I use my debit card, what exactly happens in the background? #07:02#

Interviewee For debit cards, we often face a moment during transactions, like at a supermarket, where it takes about five seconds, and you're uncertain whether the transaction will be accepted. We've all experienced that

anxiety, perhaps wondering if we'll have to leave our groceries behind. When you pay, the transaction is immediately sent to the card acquirer, also known as the merchant processor or part of the payment processing flow. Before funds are transferred, several checks occur, including for email discrepancies or fraud, treating it as a live transaction that must respond quickly, typically within a second. Although I'm not sure of the exact TPS, it's about 1000 transactions per second, necessitating rapid processing. In that short period, we build the profile, collect transaction history, and gather all metadata related to the point of sale, among other things. Once we have the data, it's fed to the model for prediction. If a transaction is deemed risky, perhaps because it appears the card was stolen, it gets declined. This is a simplified explanation of the process. But what happens earlier? The predictive process involves extensive training. For example, if a company like amnis provides a dataset, we first import and analyse this data to understand transaction types and identify risk factors, often in close collaboration with the data provider. We then set metrics, like precision or recall, depending on the priority, and enter an iterative loop where we train the model, review our findings as if we were fraud investigators, and adjust based on feedback from the data provider. Once we have a satisfactory model, it's deployed and starts handling live transactions. #07:24#

Interviewer Thank you very much, really interesting. You spoke about fraud prevention, but I'm also curious about the distinction between prevention and detection. Is there a difference? For instance, if a payment goes through successfully, is there another check in the process to detect anomalies, or is your focus primarily on prevention? #10:31#

Interviewee Could you repeat the question? I'm sorry, I didn't catch it. #11:05#

Interviewer You've talked about real-time prevention, but I'm curious about what happens after a payment is either successful or declined. Is there a process where transactions are reviewed again to check for any signs of fraud? #11:07#

Interviewee It really depends on the setup. If you want real-time action, then you focus on prevention. However, if detection after the transaction is the goal, it doesn't have to be done live, and you aren't constrained by performance concerns. This allows for the use of a more complex model that could run on a batch basis, perhaps daily. Detection is generally more accurate not just because you can employ more sophisticated models, but also because you have the full context of the situation. For example, in cases of APP fraud involving social engineering, the pattern often includes the counterparty ceasing communication after the last transaction. This sudden stop in the transaction history can be a clear indicator that something went wrong, like an abrupt end to communication, providing additional context for the detection model. #11:28#

Interviewer It really depends on what the client wants. Some clients may prioritize predictive and preventive measures, while others might focus more on detection. It varies with the setup. Additionally, it's possible to combine these approaches. Clients may have different needs and preferences, which can be accommodated by tailoring the system accordingly. #12:49#

Interviewee Exactly, it's possible to combine approaches. You could have two models running concurrently, one in real-time that blocks transactions and another that operates on batches, which doesn't aim to block transactions immediately. Instead, it can make decisions an hour later, and that would still be effective. #13:01#

Interviewer Which technological advancements have most significantly impacted the effectiveness of fraud detection systems in the last 2 to 3 years? #13:21#

Interviewee Of course, AI is a significant driver in advancements, but if we delve deeper, it's really about the enhancements in data mining that have empowered AI further over the past few years. Data mining has greatly improved by providing richer context to transactions, enriching the datasets that were previously limited in scope. This enrichment includes deeper insights into transactional contexts and user profiles, which in turn enhances the predictive capabilities of AI models. Additionally, the development of Large Language Models specifically tailored for transaction monitoring is a noteworthy advancement. These models can analyze vast amounts of transactional data in real-time, detecting anomalies by understanding the full context of each transaction. For example, if a model learns that someone frequently travels between certain locations and suddenly there's a transaction that deviates significantly in amount, timing, or geographic location, it can flag this as an anomaly due to the context provided by previous transaction patterns. This capability to integrate and analyze complex patterns autonomously represents a major leap forward in fraud detection technology. #13:36#

Interviewer Someone mentioned the LLM part, which was really interesting. When discussing the challenges in developing these models, I've read that issues like class imbalance, concept drift, changing customer behaviors, and limited transaction data as well as real-time processing. What do you think are the most significant challenges you've faced in your experience, and how do you address these challenges? #15:51#

Interviewee Class imbalance is definitely a major issue. For instance, if you have a million transactions and only 1'000 of them are fraudulent, those fraudulent ones will likely be so diverse that it becomes difficult for both humans and models to generalize what's happening in each case due to the vast difference in volume. Additionally, there is an inherent bias in the industry related to fraud detection. Often, we may be victims of fraud without realizing it, meaning many fraudulent transactions might never be reported or caught, especially if they are minor amounts that people choose to overlook. Another significant challenge is the issue of falsely identifying transactions as fraudulent, which can occur when people forget about legitimate transactions and mistakenly report them. This misclassification complicates the accuracy of fraud detection systems. Furthermore, data quality is crucial. Understanding the language of transactions and all the data fields is essential, but often the information provided by customers or systems about these fields is not entirely accurate. This could be due to system quality flaws that produce incorrect data types or values, which in turn pollute the model and complicate the detection process. Addressing these challenges involves continuously refining data quality and improving model training to handle class imbalance more effectively. This might include using techniques like synthetic data generation to better represent minority classes or employing advanced analytics to better understand and adjust for concept drift and changing customer behaviors. #16:35#

Interviewer To sum it up, it is indeed crucial to ensure good data quality and thoroughly understand the data presented to effectively address these challenges. First, you need to comprehend what the data is about and ensure its quality. #18:48#

Interviewee Additionally, it's important to anticipate the presence of mislabeled samples. In the case of supervised learning, you might encounter samples that are thought to be non-fraudulent but are actually fraudulent, and vice versa, simply because they are incorrectly labelled. This possibility must also be accounted for. #18:59#

Interviewer How do you cope with unpredictable customer behaviors, especially concerning debit card use? For example, a Swiss customer typically makes payments in Switzerland but suddenly travels and makes purchases abroad. How do you manage such variations in behavior? #19:23#

Interviewee As mentioned, understanding typical behavior patterns, such as a Swiss customer regularly purchasing in Switzerland, is crucial. However, when considering areas like Basel, where crossing into Germany or France for cheaper groceries or dining out is common, models can learn these cross-border behaviors. The models are trained to recognize that people living near borders might frequently cross into neighboring countries, and this is a normal pattern. For example, it's typical for Germans to travel to Mallorca, and such regular travel patterns can be incorporated into the model's understanding. This way, the system can distinguish between usual and unusual spending behaviors, adapting to recognized patterns like restaurant spending in foreign transactions, and improving its accuracy in fraud detection. #19:46#

Interviewer From your perspective, what key features or capabilities should an ideal AI-based fraud detection system include to ensure effectiveness? Should it combine both preventive and detective aspects, or also incorporate rule-based approaches? What do you think are the most critical features or capabilities? #20:46#

Interviewee I mean, definitely combining a rule-based approach is super useful, particularly in cases that would be almost impossible to detect by a model. For example, if there's a new payment type that the model doesn't support yet, we can quickly add a rule for this case, such as triggering a maximum of two transactions per month. Another very useful feature that I would 100% recommend is systematic feedback gathering. This means that not only do investigators work on cases and conduct their investigations, but the system also collects feedback. Whether the investigator accepts the case or needs more information, we gather all this feedback. This feedback can be explicit, like direct input from the investigator, or implicit, based on their actions within the system. This allows us to retrain the models continuously. So, combining preventive and detective aspects is crucial, and incorporating a rule-based approach adds flexibility and quick adaptability. Besides the quality checks and so on, it's mandatory to have comprehensive feedback mechanisms. We often see that investigators use external sources like Google to get more information about a transaction or person. This indicates they are accessing information our system doesn't have. So, why not integrate Google search information into the model? Of course, this needs to be done thoughtfully, possibly through a provider. Ideally, the operator shouldn't need to leave the platform during the investigation because all the necessary information would already be there, and hence it can be checked by the model as well. This would streamline the process and improve the accuracy and efficiency of the fraud detection system. #21:24#

Interviewer This is what I want to talk about, especially how the user interacts with the system. How can you ensure user-friendliness so that a compliance officer can effectively work with such a system? You mentioned that having all information within the system would prevent the need to switch to Google or another platform, which is great. But what else makes a system user-centric or user-friendly? What are the key aspects to consider? #23:25#

Interviewee Yes, well, I'm not an expert in UX since my focus is on data engineering, but I can say that reducing the amount of clicks is important. It's not just about counting the clicks but minimizing the steps needed to assess whether a transaction is fraudulent. If you can have all the information available in one view or reduce the number of steps required to get the information, that's ideal. It increases efficiency and is better for the user. #24:03#

Interviewer Because for the user, it's crucial to quickly decide whether a transaction is fraudulent or not. This means the way data is presented is very important, what data is shown and how it's explained. Explaining the decisions made by the system is also a key factor. #24:47#

Interviewee Exactly, expandability is crucial. Imagine being an investigator who has to do a lot of clicks and open many windows to get the information needed. This can bias the investigator because they might see a 100 euro transaction and assume it's fine, simply because humans are naturally inclined to avoid extra work. If they don't want to check the full profile or click around, they might make assumptions. Therefore, having clear explainability in one place helps avoid these biases. #25:17#

Interviewer When we talk about explainability, how can you effectively ensure that such systems are understandable for users, especially compliance officers who may not be technically adept? You mentioned reducing the steps involved. How can this be achieved? #25:57#

Interviewee So, a theoretical approach for explainability is the use of SHAP values. These values indicate the importance of each risk factor or feature in the prediction. For example, if we're talking about phishing, and the call is from a different country, SHAP values would highlight how relevant the country is in pointing out the risk. #26:12#

Interviewer Yes, exactly. For instance, if a Swiss customer usually pays in Switzerland and suddenly, within 10 minutes, there's a transaction in China, the SHAP values would highlight the country risk. The system would explain the alert by showing that the rapid change in transaction location is unusual and risky. This kind of explainability helps users understand why an alert was triggered. #26:46#

Interviewee Exactly. #27:12#

Interviewer When we talk about metrics, there are indeed many different ones to consider when training models. Do you compare models based on specific metrics, or does it depend on what the client wants or the use case? For instance, do you use certain metrics for specific use cases? Is there a general approach, or is it really case-dependent? #27:17#

Interviewee I would say it's case-dependent. For the same use case, dataset, and objective, we can compare the same metrics. However, in general, if I tell you we reached a true positive ratio of 90% or a precision of 99%, it sounds impressive. But even 1% of cases being misclassified can have a dramatic impact. It always depends on the customer and the use case, whether it's fraud detection or something else. For instance, if we are detecting fraud live, a high number of false positives, even if it's only 1% or 0.01%, could mean that a million transactions are blocked at grocery stores, causing significant inconvenience and customer dissatisfaction. So, while achieving high precision is important, it's equally crucial to minimize false positives to avoid such issues. #27:43#

Interviewer But the most challenging aspect is understanding how much risk a customer is willing to take. It depends on whether they are more risk-averse or willing to take on more risk. Some clients may prioritize increasing accuracy, while others might focus on speed or efficiency. #28:43#

Interviewee Exactly. #28:58#

Interviewer So it really depends on the case, right? #29:03#

Interviewee Exactly. #29:06#

Interviewer What they want, OK. We've already discussed a lot about the difference between detecting and preventing fraud, but I want to dig a bit deeper. Are there different strategies for detecting and preventing fraud? Is there a significant difference? You mentioned that one involves real-time processing and the speed required, and there's a bigger risk in preventing since blocking too many transactions can upset clients. But are there different strategies here? Can you use the same models, or is there a big difference in how you approach these two tactics? #29:06#

Interviewee I wouldn't say there's a huge difference in the models themselves, rather, it's about the information the model can access. For instance, in detection, the model can see the future in a sense because it has access to a day's worth of transactions. If the first transaction is legitimate but the second one is malicious, the model can use the context of subsequent transactions to identify the malicious one. So, the key difference lies in the data fed into the model for training, not necessarily in the model itself. I wouldn't differentiate between prevention and detection based on the type of model used. However, if a model is very complex

and slow, I might avoid it for real-time prevention due to timing constraints. Essentially, the distinction comes from the data and risk factors considered in training the model, rather than the model's architecture. #29:46#

Interviewer So maybe in prevention, you need a baseline of data to begin with, while detection starts with a model to gather all the necessary data. Once you have enough data and the model has learned sufficiently, you can move towards a prevention approach. Is that the usual method? #30:57#

Interviewee Exactly. Detection can happen offline, allowing you to train your model to detect fraud. Once trained, you can use this model for prevention, provided the model for detection doesn't rely on information that won't be available in real-time. If the model trained for detection uses data that isn't accessible during live operations, it won't function effectively when deployed for prevention. Ensuring that the same data is available for both detection and prevention is crucial for the model to work correctly in real-time. #31:16#

Interviewer I mean, it depends on the customer. For example, if a company like ours doesn't have much transaction data because we only started with debit cards last year, it makes more sense to start with a detection model. In contrast, a big bank with years of transaction data from their customers can more easily start with a prevention model because they have extensive historical data to work with. Since we only have about six or seven months of data, starting with detection is more logical for us. Is that correct? #31:55#

Interviewee Well, yes, I would say so. For them, if they have lots of labelled data, they could do both detection and prevention simultaneously. They can detect fraud based on historical data and use that same data to prevent future fraud effectively. #32:35#

Interviewer Beyond detecting fraudulent transactions, what additional goals should these systems aim to achieve to enhance user trust and system usability? First of all, we talked about explainability, which is definitely a crucial goal for such a system. Additionally, how quickly a user can decide if it's really fraud is important. It's essential for the system to allow users to quickly assess the situation based on the data presented. But besides those factors, what else do you think is really important for such a system? For example, should the user be able to easily understand and say, this is fraud because of this reason? What other key aspects should be considered to make the system more user-friendly and effective? #32:53#

Interviewee Well, for the system, it's crucial that the information is clearly specified when working on cases. This ties back to what I mentioned in some previous questions. You shouldn't have missing information that might lead you to incorrectly determine a transaction is not fraudulent when it actually might be. All data should be properly integrated. For instance, if the system logs admin activity, and in the future, you include mobile payments, it's important to track how often and in what patterns users log in. This information is crucial for the model. If I suddenly log in from another country at 3:00 AM, that's suspicious and valuable context. Traditional systems often exclude such details, focusing only on transactions. But if we enrich transaction data with login information or other relevant details, it significantly improves the accuracy and reliability of fraud detection. #33:36#

Interviewer it's really important not to just focus on transaction data but to include all the surrounding data. This includes device ID, login frequency, and other relevant information. It's a combination of all these different data points that ensures the system is effective in the end. #34:48#

Interviewee Exactly, yeah. #34:58#

Interviewer When we look to the future, what advancements or features do you anticipate will be important for future fraud detection technology? Which advancements do you hope to see in this area? #35:08#

Interviewee I mean, there's still a lot to mature. Ideally, in a future scenario, we would anonymize transactions and other data, and the model would learn from all this information without needing human intervention. The model could automatically consider cross-border transactions, frequency, volume, and other factors, avoiding human bias. For future directions, cross-institution checks would be important. For example, if amnis sends money and a transaction is made, it would be beneficial to track where the funds are going or the history of the received amount. If these transactions occur on different platforms, it's currently difficult to see the complete picture. In the future, if everything is connected via blockchains, we might have publicly accessible records that allow us to trace transactions across different platforms. So, enhanced collaboration between payment platforms and banks is essential. #35:23#

Interviewer Do you want to add anything or do you think I've forgotten to ask something that's really important to consider in this area? Maybe there's something crucial I didn't cover. #36:44#

Interviewee We covered a lot of important points today. We talked about various aspects of fraud detection, including AI and out-of-the-box thinking regarding data integration. We discussed future advancements and the

importance of comprehensive data for improving these systems. If I think of anything else later, I can always reach out to you. But overall, I think we covered most of the critical points. #36:56#

Interviewer Perfect, thank you. #37:34#

Expert 3 at Hawk

Date: May 27th, 2024

Time: 13:30

Place: Microsoft Teams

Part 1: 00:00:00 – 00:36:30

Interviewer The first part of my question is about you as experts and your background. Just a short introduction. Could you share details about your hands-on experience with fraud detection, particularly for card transactions? #0:03#

Interviewee After my bachelor's studies, I had the chance to work with a bank as soon as I started back at home in Pakistan. Although my role was very limited in terms of fraud detection, I still gained firsthand experience with how bank transactions actually work and how the banking system operates. After a few months, I came to Germany to pursue my master's, focusing on mathematics and machine learning. For my master's thesis, I worked on explainable AI. In my thesis, I used causality correlation and similar techniques to work with accountability, particularly through anomaly detection. I generated different kinds of anomalies and used explainable AI to analyze them. Fraud can be considered a type of anomaly, so this experience was relevant. I felt confident about my interview for a position because of my work on explainable AI and anomaly detection. I gained significant knowledge in explainable AI and had a basic understanding of anomalies from a mathematical perspective. After working further in this field, I started to develop a good understanding of fraud detection, learning which features to look for. Of course, it also depends on the client and the business case, as well as the available data. We compute a vast set of features to identify isolated points in the space that indicate fraudulent or suspicious transactions. #0:27#

Interviewer So, your experience regarding AI and fraud detection started when you began working for Hawk is that correct? Especially when we talk about AI-based systems for these applications? #2:49#

Interviewee Exactly. #3:00#

Interviewer Thank you very much. You already started talking a little bit about the state of the art in AI-based fraud detection. Can you tell me about the current best practices in AI-based fraud detection for debit card transactions? What do you consider the best in your expert opinion? #3:02#

Interviewee There are many companies that claim to use AI but don't really have it. Even if they do, the difficult part is not creating the AI system, the challenging part is obtaining the right features for the system. It's a hefty process. You have to go back in time, behind the current transaction, and compute not just normal features but very complicated ones as well, which consider payment types, payment means, and other factors. Some companies don't do that, but the state of the art in AI-based fraud detection is, in my opinion, either a tree-based approach or something similar. For example, isolation forests have a big advantage because they don't require labels for training. This is important since you usually don't have labels for your data on fraud or money laundering unless you have an agent providing feedback. Once you have feedback, you would rather use a deep learning technique, which is very good at handling new and unseen data. #3:34#

Interviewer There is also a lot of talk about large language models, which also sound very promising. #4:58#

Interviewee Yes, we are already working on that. We started developing it, and the first version took about a month as a proof of concept. Unfortunately, I was on vacation during that time and couldn't work on it, but when I came back, I saw the results, and it outperformed other models. I'm not sure if you're familiar with AI algorithms, but XGBoost is one of them, and it is already quite good. The large language models outperformed XGBoost, which we used to use and were quite happy with. In the future, we plan to continue using large language models. If you're curious about how we use them, we treat each transaction as a word. Just as a large language model tries to predict the next word, we try to predict the next transaction and determine how different it is from the expected one. #5:09#

Interviewer You mentioned that it's really part of the best practice to focus on data handling. What kind of data do we have? How do we understand this data? If I understand correctly, the first step is to understand the data and extract the features from it. The algorithms and models are important, but understanding the data comes first. Is that correct? #5:59#

Interviewee Exactly. #6:35#

Interviewer Could you describe the operational workflow of a typical AI-based fraud detection system? For example, as a user, if I have a debit card and I make a payment somewhere, what happens in the background? Can you explain that process? #6:38#

Interviewee So, you're talking about the model being deployed and how it operates in real-time for the bank. Once we train the model, it is deployed to the system, let's say through a UI. When you make a transaction, depending on the client, it could be processed as a real-time transaction or via batch transactions. Let's consider it's coming in real-time. When you make the transaction, the details are sent to our system. These details include various fields such as payment type, payment means, e.g., debit transaction, amount, time, counterparty information, and possibly the counterparty bank identifier. Sometimes this information is available, and sometimes it's not, depending on the client. Once we have this data, we consider the features the model was trained on. For example, if we used ten features to train the model and one of the features was the volume of transactions sent to the same counterparty this month compared to all other transactions, the system would go back 30 days. It would accumulate the volume for just this counterparty and divide it by the total volume of transactions you made in one month. If the percentage is close to 100%, it's suspicious because you're making transactions to just one shop. One interesting case we encountered involved a person trying to launder money by making high-volume transactions to different accounts, all from the same bank identifier. In this case, we grouped by counterparty bank identifier rather than just the counterparty account ID. This helped us detect that although the transactions were to different accounts, they were all linked to the same bank branch. We can extract very complex features as well. For instance, we might subtract or multiply using debit and credit amounts, divide by the total volume, and if the result is close to zero, it indicates that the person is acting as a mule. If you need more details, I can provide further examples. #7:00#

Interviewer Could you also talk about the process of training these models or creating these models? For example, as a customer, if I have my goals and I come to you, what happens next? How do you develop the model for me? #10:02#

Interviewee First of all, the customer tells us their specific goals and areas of focus. For example, they might want to focus on counterparty behavior or high cross-border transactions. We incorporate these preferences into our features, but we also ensure that the data supports these focus areas. For instance, if a customer wants to monitor cross-border transactions but doesn't consistently provide country information for 60% of transactions, we'll advise them that using this data might add noise rather than clarity. Similarly, if they want to check for refund transactions but these only appear for 2-3% of customers, we will explain that such low-frequency events might always appear anomalous in an unsupervised model. In the training process, we extract a lot of features, such as cross-border transactions, changes in transaction volume, and other relevant behaviors. For example, if a customer's average transaction volume suddenly increases from €50 to €1500, especially to high-risk countries like Afghanistan or Iran, we would flag this as suspicious. We analyze these features using 2D plots with time on the X-axis and feature values on the Y-axis. The feature distribution should be skewed in some way to identify anomalies effectively. We filter and select the most relevant features based on their expected behavior and how the model responds to them. Next, we use explainable AI plots, such as SHAP values, to understand how each feature impacts the anomaly score. For example, if the pass-through value is low, close to zero, indicating the person is not keeping any money over a certain period, the anomaly score should be high. If the feature doesn't behave as expected, we re-engineer or exclude it. Once we are satisfied with the feature behavior, we train the model on these selected features and evaluate its performance. We look at cases with high, borderline, and low anomaly scores to ensure the model is functioning well. If everything checks out, we present our findings to the customer, showing examples of high anomaly cases. If the customer is satisfied or prefers not to see the details, we deploy the model as requested. #10:18#

Interviewer Thank you very much. In my research, I've read a lot about the challenges in training and developing these models, such as class imbalance, concept drift, changes in customer behavior, and limited transaction data. What can you tell me about these challenges? What do you think are the most significant hurdles, and how can you cope with them? #14:42#

Interviewee Class imbalance is a significant issue. For example, refund transactions typically make up only 2-3% of transactions, which creates a class imbalance. This is particularly challenging for unsupervised learning methods like isolation forests. In these cases, most data are unlabeled because for rule-based systems, an agent can label data by validating whether the rule worked. However, for anomaly detection in money laundering or fraud, we generally lack labels. Isolation forests work by creating branches and measuring the length of these branches to determine if a transaction is anomalous. A short branch length indicates normalcy, while a long branch length indicates an anomaly. With imbalanced classes, like a small percentage of refund transactions, the model might automatically classify these as anomalies due to their rarity. In such cases, if a client has specific needs, like dealing with crypto transactions which are few, we train a separate model just for those transactions. This approach is also applied to differentiate between B2C and B2B transactions because their behaviors differ significantly. B2B transactions generally have higher volume and frequency, so we use a different model for them. We set up a pipeline where

transactions are directed to the appropriate model based on their type. Limited data is another challenge. For instance, my first project with amnis involved very little initial data, which was troubling. In such situations, we can consider generating synthetic data points, but this has drawbacks, especially with transactional systems. Initially, I trained the model with the available data and retrained it once more data was available. Handling these challenges involves a combination of creating specific models for different transaction types, using pipelines to direct data appropriately, and sometimes relying on synthetic data generation or waiting for more data to improve model accuracy. #15:20#

Interviewer Looking at changing customer behavior, how can you face the challenge when the customer doesn't behave as predicted? For example, I might always pay in Switzerland for many years, but now I start using the card worldwide. How can you face the challenge of adapting to such changes in customer behavior? #18:10#

Interviewee For example, there is a feature we normally use that checks the number of cross-border transactions made in the last month compared to the last week or the last three months compared to the last month. If you weren't making many cross-border transactions and suddenly start making a lot, it will raise the anomaly score. However, that would just be one feature contributing to the anomaly score. Other features like round amounts, transactions with the same counterparty, pass-through amounts, or unusually high amounts would also be considered. If you move to another country and make transactions there, and the amounts are significantly higher than usual, this will likely raise the anomaly score. The UI will also share the reasons for raising the anomaly score, providing an explanation based on a compilation of multiple features. #18:42#

Interviewer Then you have two features. We look at not only the country but also other features like the timing of the transactions. #20:09#

Interviewee Exactly. For one of the clients in the United States, there was no country information available. Instead, we had longitude and latitude data. We implemented a very cool feature where we calculated the Manhattan distance between the longitude and latitude coordinates and divided it by the time elapsed since the previous transaction. If this value was significantly higher than usual, it indicated something suspicious. #20:12#

Interviewer Now let's talk about the desired capabilities. We've discussed the goals and characteristics of such systems. From your expert perspective, what key features or capabilities should an ideal AI-based fraud detection system include to ensure it is effective? I mean, sure, it should find fraudulent activities, but what other key features and capabilities should such a system have? #20:47#

Interviewee I think having proper explanations is as important as finding fraudulent activities. There are infinite features we can use, but they need to be translatable to plain language. For example, we can't just say that a holding party made a high volume of transactions; we need to explain what we mean by high volume. For instance, in the pass-through feature, if the value is close to 0, say 0.1, meaning only 10% of the volume remained in the account in the last few days while everything else moved out, we would explain it as follows, the anomaly score is high because 70% of the weightage is given to this feature, which indicates a pass-through ratio of 10%. A ratio less than 20% is considered suspicious. This way, an agent who might not have studied finance or money laundering would still understand the significance of the feature. We make sure every feature is quantifiable. For example, we might say, this holding party made 70% of their transactions last month to just this counterparty, clearly explaining why this is suspicious. Apart from that, finding good features is crucial. If we are talking specifically about fraud detection and we have labelled data, we also need to consider metrics like true positives and false positives to evaluate the effectiveness of the system. #21:17#

Interviewer And could we also discuss how a user, like a compliance officer, interacts with such a system? We've mentioned that explainability is a big part of it, indicating which feature decided that this is now an alert. Are there other aspects you think are important in representing the data or deciding which data to show? How do you design such a system to be user-centric and really useful for compliance users? Are there other features you think are necessary or important? #23:04#

Interviewee From a compliance perspective? #23:42#

Interviewer How should such a system be designed from your perspective? You're not a compliance officer, but when you work with such systems, what do you think is important in the design? #23:49#

Interviewee As a developer, I would definitely emphasize the importance of the pipeline and the compatibility and connectivity of the pipeline steps. It's crucial that the training platform and the deployed live transaction system are integrated well enough. For example, our training platform and the deployed model system are closely aligned. The training model extracts features by going back 30 days into the database, and the deployed model system does the same for live transactions. This connectivity ensures consistency and accuracy in feature extraction. However, I think that was not your question. #24:06#

Interviewer I just want to know, maybe also I'll get my question better formulated. From previous interviews, I heard about providing not only what the AI found but also additional information. For example, in a web-based approach where the compliance officer has their tool and reviews alerts, it would be great to include all the information from the Internet or other databases. This would help the compliance officer make a decision. Is that what you mean? How can we incorporate stuff like that? #24:50#

Interviewee You mean more data points? Stuff like that? #25:32#

Interviewer How can you make the system more effective for the user? When a compliance officer has an alert, how can you help them make decisions more easily or effectively? Sorry, the question wasn't well-formulated before. #25:36#

Interviewee No, it's OK. So, more UI features are definitely good. For example, let's consider a case where multiple transactions are made from the same IP address. One of our clients had a situation where there were no online transactions, only ATM transactions or bank withdrawals. There was information about the IP address whenever a customer logged in to check their account or change some information on the website. We found that multiple accounts, maybe 10 or 12, were logged in at some point from the same IP address. When we raised this to the bank, they suggested it might be a library or an airport. However, the transactions were being made from the same ATM, so it was probably not an airport or library. If the UI could group multiple accounts using the same IP address or the same ATM, that would be helpful. Right now, we only group by account IDs or customer IDs, but adding this feature for compliance would be significant. We already do something similar for fund flows, showing a tree structure with arrows for amounts coming in from multiple accounts and going out to one account. This could be improved further. #25:49#

Interviewer When developing such models, which metrics are decisive in the development and used to measure effectiveness? I read a lot about metrics like true positive rate, accuracy, or ROC curves in research, but can you say in general which is the best score? Or does it really depend on the specific case or what the customer wants? Can you tell me more about that? #27:48#

Interviewee Yeah, of course. Actually, I've never worked with supervised models at HAWK, I've only focused on unsupervised models. In unsupervised learning, it depends on the customer quite a lot. For example, a customer might have specific requirements that influence how we adjust the model. While these adjustments might not achieve the best AUC score or accuracy metrics typical in supervised learning, they meet the customer's needs. In unsupervised models, we don't use metrics like true positive rate or ROC curves. Instead, we analyze the cases the model has predicted, examine the anomaly score plots, and based on that, decide the proper threshold. #28:24#

Interviewer So, you cannot really say there is a best metric to use in our development. It really depends on the case and what the customer wants. If the customer says they want more speed. #29:20#

Interviewee Yeah, of course. There is no single best metric. #29:26#

Interviewer So, OK, we check how fast the model is. If the customer wants more accuracy, we adjust accordingly. #29:30#

Interviewee Of course. Some clients don't have enough budget, so they might say that our fraud or anomaly detection system opens an alert and then the agent has to check and go through them to see if it's a good alert or not, similar to a rule-based system. There are clients who can't hire many agents, so they might request a maximum of 10 or 20 alerts in a day that an agent can check. So, for example, if our model, at a 70% threshold, gives out 50 cases based on historical statistics, we have to adjust the threshold so that, on average, it only gives out 10 cases daily. The threshold, therefore, varies a lot from customer to customer. #29:37#

Interviewer And also, let me talk about fraud detection. It consists of two parts, right? We have the prevention part that pauses the transaction before it is approved or declined, and the detection part afterward. I think it depends on what the clients want. Some customers might want a prevention system to block the transaction beforehand, while others, like in our case, might not have much data like a big bank. So, in our case, it makes more sense to use the prevention approach first. Can you tell me how the development of such models differs between prevention and detection? Or can you say it's kind of the same model that doesn't care if it's used for prevention or detection? #30:32#

Interviewee No, it does not matter. We just have to configure the system to either block the transaction or alert about the transaction. It doesn't affect how we train the model. #31:30#

Interviewer But I guess in the end, as I said, it depends on the customer. For us, with not that much data, it makes no sense to implement a prevention model because probably every second transaction would be blocked,

right? So, it depends on the customer. A big bank like Citigroup can use a prevention model because they have enough data and can take the risk. #31:42#

Interviewee Exactly. #31:55#

Interviewer And so, it depends. We can do a prevention model or a detection model, but in the end, it really depends on the customer's needs and data availability. #32:07#

Interviewee And, for AI, blocking doesn't normally make sense because the model can be wrong, right? There can be false positives. For rules, it does make sense for blocking. For example, if the rule is about a transaction going to high-risk countries like Russia or if the transaction was made to the same counterparty for the seventh time in a day, it would be blocked. So, for rules, it makes more sense because it's very quantifiable. #32:09#

Interviewer Do you also think the optimal fraud detection system should not only use AI but also include a rule-based approach? So, it's a combination of both? Or would you say it should be just AI and let the agents decide in the end? #32:44#

Interviewee I would say there should be a couple of good rules. Not a lot, because that would just increase false positives, but a couple of good rules should be in place in case the model leaves out something. In the end, the model is just a system with its own logic, and you can't trust it completely. So, a combination of both is ideal. #32:57#

Interviewer One follow-up question. You also talked about challenges. Is bias a big problem in training such models? Do you have to take care of bias? Can you tell me more about that? #33:19#

Interviewee Yeah, it goes in a very close direction towards class imbalance. Bias usually occurs when you have labelled data, but it can also happen with unsupervised learning. As I said, I've mostly worked with unsupervised models. For example, another type of bias can be seen in B2C versus B2B transactions. If most of the transactions are B2B, the model becomes biased towards high-frequency or high-volume transactions with large customers and transactions with the same counterparty. For instance, if a steel contractor is only selling to Mercedes, the model might see this as normal. However, for B2C, the model doesn't see such patterns. There are many different counterparties, with many small transactions from various shops. To combat this bias, we train different models for B2C and B2B transactions. #33:39#

Interviewer Now for a little outlook. What advancements or features do you anticipate will be important for future fraud detection technology? What do you hope to see in the future for such systems? #34:39#

Interviewee I think model governance will be quite significant. There will likely be regulations from the EU that we have to follow, and some features might be forbidden. I had a discussion with another employee a few days ago where I asked if we had a feature based on race, would we use it. Statistically, if we use country as a feature, why not race? It's kind of the same thing. He said we should be able to use it because the model is not biased, it's just using statistics. However, I pointed out that if the model declines a transaction because the person is black, it wouldn't look good on the case manager. These are the types of issues I think will lead to more regulations. On the success side, I think large language models will prove quite useful in the future. #34:58#

Interviewer Is there anything you want to add that I might have missed? Especially in this area, is there something important that I forgot to ask you? #36:11#

Interviewee I don't think so. I think you pretty much covered all of it. #36:24#

User 1 at amnis

Date: May 15th, 2024

Time: 16:00

Place: Microsoft Teams

00:00:00 – 00:59:08

Interviewer The first question I want to ask you in the first section of my interview is especially about you as a compliance officer. What is your background and your role? Could you please describe your role and responsibilities related to fraud detection, especially regarding debit cards or card transactions? #00:00:05#

Interviewee Technically, my experience and background from previous jobs and nowadays with amnis regarding fraud, it's really another department. I have always been part of the compliance group, which usually closely cooperates with the fraud department, the money laundering department, financial crime monitoring, etc., within the banking sector. It's just a different area. However, the root cause is the same, right? The bad guys are trying to jeopardize the customers, and we need to protect them. So, technically, it has many similarities. However, the fraud setup is pretty simple because the initiation or the clear root cause of stealing money from cardholders is just for the usual criminal benefit. There is no hidden or complicated background behind it. Of course, we can talk about organized crime, groups, and money laundering. I would say money laundering over the card business is a little bit different than the typical fraud-related stealing of money. The root cause problems are different, right? So, going back to my background, especially at Citibank, I gained a lot of experience with regards to cards. Back then, cards were considered very innovative and a new product in the 2000s and later on. So, of course, the criminals became more and more active. Nowadays, it's a completely different story because it used to be about skimming ATM machines. Also, technologies like embossed papers and embossed card printouts, which were widely used in the United States, for example, are no longer common. All those factors are technically gone. Now, it's really about the online world and online fraud schemes. Mainly, I would say, right? So from that perspective, my experience so far has not been deeply dedicated to or experienced with such modern online money-stealing practices. It was really the old school ones mainly. #00:00:25#

Interviewer But still, with your long experience in the industry working as a compliance officer at many different big banks, you have gained experience regarding that topic, right? #00:04:17#

Interviewee Exactly. So, to make it really short and targeted, yes. #00:04:25#

Interviewer And specifically, what has been your experience with AI-based fraud detection systems? Have you had any touchpoints before, or is this your first time? #00:04:38#

Interviewee No, it's my very first time. The development of AI for this type of specific monitoring purposes is relatively new, within the last two years max. There's no clear market practice yet, it's really been developed in the last year. Before, it was all hard rule-based systems with set parameters like counting machines, frequency rules, and volume rules. It's still a pretty basic and similar setup. #00:04:47#

Interviewer So, OK. That's the follow-up question I want to ask you. As you said, AI is now coming up. It wasn't around for 10 years already. It is now being adopted, and big companies are starting to use AI for fraud detection. But, as you said, for the past three or four years, it was really rule-based, expert-based decisions, right? #00:04:52#

Interviewee Yeah, exactly. #00:05:55#

Interviewer Now, if you talk about the current fraud detection systems and their challenges, what fraud detection systems are currently used at your organization and how do they function? Additionally, what systems or approaches are used across the industry for fraud detection? As I understand it, you said it is still the common or standard approach currently, still hard rule-based. Is that correct? #00:06:04#

Interviewee Yeah, that's correct. Nowadays, the methodology behind AI is technically pretty simple, right? Because it usually needs history. If you have a history, you are good to go, right? Because you can give the machine the data in terms of transactions and everything, you can pinpoint the bad transactions in that history. The machine has the benefit of learning the patterns you, as a human, identify as problematic. In the current world, such patterns are the key to everything because criminals will always evolve ahead of the compliance team or analysts trying to protect the company and the customer. So, yeah, historical data is key, as well as the use of human experience. However, the machine is much better at this because it can

learn much quicker than a human being, who needs many years to reach a quality level that can be used for prevention purposes. #00:06:38#

Interviewer So, to sum up, the standard approach to fraud detection, including what we are using here, is still rule-based. As you mentioned, it's a rule-based system with hard rules related to volume, varieties, and frequency, but also the rules behind those parameters. #00:08:31#

Interviewee Exactly. #00:08:40#

Interviewer So, can you explain it? #00:08:54#

Interviewee Yeah, it's still there. We are now waiting for the green light because we already have the history. We are prepared, and the provider is now ready. We are just waiting for the slot to enable this, especially on the cards, right? As you can imagine, card transactions are absolutely common on a daily basis. You pay small amounts for Starbucks, or whatever, every day. So, the activity is really frequent with rather low amounts, and being able to identify fraud within this large number of overall transactions is like searching for a needle in a haystack. Something really, really complicated. Technically, now we have to figure out how to identify the fraud, right? #00:08:54#

Interviewer Maybe you can also describe how it works in that context. How does the system currently operate when a client pays with a debit card? So we get the information. What happens next? What is the process behind it? Is it rule-based, or how does it work? #00:10:15#

Interviewee Yeah, sure. #00:10:32#

Interviewer Can you briefly explain that to me? #00:10:34#

Interviewee There are certain card-related rules defined at country levels. For example, if you pay locally or if you go abroad, all those parameters are predefined in terms of thresholds, right? So usually. Depending on where you are residentially, you should fit into a predefined group of typical cardholder usage patterns. If you cross a border and go abroad, the system and monitoring rules should spot it and at a minimum, give us certain alerts to double-check if your activity as a cardholder is legitimate. In the end, we do this to protect you and your money. This is a service for you. It's a little different than monitoring for financial crime and money laundering. It's really a prevention measure. The rule setup should spot and identify borderline transactions that could fit into fraudulent activity. When there's a breach of the rule, such as frequency or volume, it creates an alert. We have a list of card-related rules defined, and every special frequency or volume rule should create an alert. As an analyst, when a card alert is created and assigned as a fraud typology, it should receive closer and quicker attention. We need to resolve the fraud quickly because time matters a lot in these situations. Even though, of course, if the money is gone, it's usually too late. But technically, we want to prevent other transactions, protect other balances, etc., and clearly act to protect the customer in the end. We are responsible every day for resolving such categories of alerts, and as of now, it's purely manual work. It can be perceived as annoying and repetitive because cardholders often have frequently driven methodologies like buying Starbucks every morning, cigarettes, or taking an Uber for travel. So technically, those transactions and even the timelines of when those transactions happen play a significant role in terms of the quality and the alert result distribution. That's exactly the point where we expect huge help with AI, right? #00:10:37#

Interviewer So maybe if I could ask you, the current system works more for detection than prevention. Is that correct? #00:14:55#

Interviewee Detection, definitely. #00:15:02#

Interviewer It's complete detection, so we do not stop the transaction for the client. #00:15:04#

Interviewee It's complete detection, exactly. #00:15:06#

Interviewer Based on some rules, the user can decide the limits of the cards, and if these limits are exceeded, the transaction is stopped, obviously. #00:15:11#

Interviewee Right, this is preventive. #00:15:12#

Interviewer But our current system works more on the detection side rather than prevention, right? #00:15:22#

Interviewee More detection, exactly. #00:15:28#

Interviewer Is that typical across the industry? #00:15:29#

Interviewee No, usually, if you compare it with large corporate banks that have lots of support teams and typically a 24/7 fraud monitoring team, they are able to handle any fraud activity at any time. They usually have

much more restrictive and blocking card functionalities in place. Considering the time zones and other factors, they are usually blocking transactions for customer protection because, in the end, it's about safeguarding customers' money. And if you have such support, meaning a 24/7 help desk to call the customer, imagine you are in LA, with a 9-hour time zone difference. If you want to pay for a hotel at 2 AM, we are sleeping here, right? That's exactly the specialty associated with card transactions that requires that attention. With that regard, of course, it has an ultimate impact on resources, prices, and everything. In our setup, it's clearly impossible to go this way, so we really follow the self-service model. if you as a customer define, analyze, and identify anything unusual, the first thing you can do is block the card on your own and then let us know. Immediately report the fraudulent activity, and we will know by the next morning. If it happens during the night or off-hours, we can help you with it afterward. To be absolutely clear, we have certain transactional daily limits. If the customer breaches such a maximum daily limit, the card is clearly blocked until the next day. These transactional limits are in place, but they do not ultimately prevent situations where there is still a limit remaining. However, if the criminal or fraudster attempts to skim the card and steal the money, that's… #00:15:32#

Interviewer Which brings up another question: is there a difference between debit cards and credit cards? #00:18:50#

Interviewee Yeah, definitely. I think so. #00:18:58#

Interviewer Do you think they are the same, or is it important to differentiate between debit cards and credit cards? #00:19:00#

Interviewee It is important to differentiate between debit cards and credit cards. If someone steals money from your credit card, which is technically a loan card, they are stealing money from your lender. It's a credit from the bank. The bank giving you a credit card is essentially providing you with a loan, which is part of their marketing strategy. As an end user, you are spending different resources, not your own money, but borrowed funds. The responsibility for misuse and fraud differs significantly. A notable example, though not fraud-related, is if you purchase an airline ticket and the airline company goes bankrupt. If you pay with a debit card, the money is gone. You lost everything, and you need to call for an insolvency and try to get the money back after the insolvency proceedings. Instead, if you pay by credit card, you can be refunded quickly because it's a different type of resource, and you, as a customer, are much more protected. The same applies to fraud attempts. #00:19:09#

Interviewer What do you think are the current strengths and weaknesses of the fraud detection systems in use generally, not only here at amnis? What are their weaknesses and strengths? #00:20:58#

Interviewee Given the fact that our counterparty is the fraudster, they continuously develop ways to bypass all possible preventive measures and keep attempting to steal money. It's really difficult to address this because it's a never-ending fight. We will never completely eliminate fraud. Specifically, even on the market and on the competitors' side, the weakness is probably the modern, very fast-evolving methods of card payments. This rapid evolution gives criminals the opportunity to be faster than the measures we implement to secure the money. So, the evolution of modern technologies also brings the possibility to steal money. #00:21:19#

Interviewer So, to sum it up, the weakness, if I understood you correctly, is that the rule-based approach currently used is too rigid and cannot adapt quickly to the modern age or the fast-evolving tactics of fraudsters and changes in payment methods or data. It cannot handle these challenges effectively. #00:22:52#

Interviewee Yeah, it's clearly impossible. #00:23:20#

Interviewer To sum it up, the rule-based approach cannot keep up with the rapidly evolving tactics of fraudsters and changes in payment methods. It's too rigid to handle these modern challenges effectively. #00:23:23#

Interviewee Even the new types of payments need to be considered. We want to satisfy our customers, allowing them to pay correctly, swiftly, and smoothly. The negative aspect of the fraudster's world is to catch those opportunities to steal money. So, the only possible way is to keep an eye on transaction activity and try to spot fraudulent activity as soon as possible. The weakness is that we are using our limited and valuable capacity on non-risky or clearly non-problematic alerts. That's just a fact. Because every single cardholder is different. Some people pay twice a day, and some pay 20 times a day or even more. But still, not all transactions are fraudulent or need to be handled as such. Unfortunately, the situation is that, sooner or later, the volume of fraudulent transactions increases. You probably have researched the global number of fraudulent transactions, and it's a huge amount. So, technically, even if we look at us as a small card issuer not having serious experience with fraud activities until now, it will come sooner or later. #00:23:26#

Interviewer To sum up, the weakness of the current system is that it needs much more time than compliance officers have. It takes more resources from you because it creates many more alerts than necessary, which means it produces higher costs and requires more time to manage. #00:25:47#

Interviewee Yeah, it wastes our effort on normal alerts. That's not wanted, and that's exactly the point. We want to initiate AI to help us with this large number of alerts, transactions, and detection methodologies. So, that's exactly what we are aiming for. #00:26:09#

Interviewer And if we're talking about challenges, what are the primary challenges you have encountered with AI-based fraud detection systems? Have you experienced any while working with them? #00:26:47#

Interviewee What I can say, at least from the AML transactional monitoring point of view, is that so far, so good. I'm impressed, to say the least, right? That's why I really expect the same from our specific set. Now, I don't really distinguish much between the card fraud-related alerts or AML-related alerts for card transactions. It's, of course, one problem in the end, right? The only distinction is that fraud is immediately spotted, like customer dissatisfaction, because if someone steals your money, you are annoyed and upset, and that's just something you want to avoid. To just keep the customer satisfied instead of focusing on monitoring AML with cards, where the customer may pretend everything is normal while attempting to launder money using the card. Right? That's something really, really different. Different strategy, so to say. #00:27:20#

Interviewer Thank you. Now, I want to move on to the next section. This is especially about designing user engagement. How do you interact with such systems, and also about system explainability? So my first question there will be, how are users like you interacting with the fraud detection system, like a platform? Do we have data represented? How do you interact with such systems? How, in your daily work, or in how does this interaction with that system look like? #00:28:49#

Interviewee There is no difference between the modus operandi with the AML and the fraud-related alerts, they both have the potential blocking functionality. If I'm not mistaken, right? Because we have not yet detected anything that would require immediately blocking the transaction on behalf of the given customer to clearly spot something suspicious to be considered in the fraud category. Then yeah, the monitoring system is a separate platform where we can spot and identify the problem or the transaction being considered as fraud. Then immediately we need to go to our system and double-check all the transactions. What we can see is what's being processed, what's being rejected, for example, so the collaboration with the IT team to spot the attempted transactions. Also, the attempts of attempted transactions, because in our platform, we can only see the executed ones, right? So, to protest transactions, and sometimes it's really necessary to go deeper with regards to the IT. Well, set up right to also spot those. And that's something that requires immediate and very fast collaboration among the teams. That's just like usual practice #00:29:17#

Interviewer It is like a platform where you have your transactions and also the ones that were marked as fraudulent or the alerts. You can see there all the necessary information regarding the debit card transaction #00:31:20#

Interviewee Yes. #00:31:35#

Interviewer I guess where it was used, how much was used, and I guess also what I got from the previous info is, and also see why the alert was created, right? There are different data points shown to you and also why which rule was broken that led to the alert creation, right? #00:31:36#

Interviewee Exactly. #00:31:47#

Interviewer This is as far as I can understand it. #00:31:57#

Interviewee Exactly, exactly. That's exactly what you just said from the transactional point of view. Like the detail how much, what currency, where. What's the name of the settler or the merchant, right? That domicile sits in which country? The transactional details, meaning the reference number and etcetera. Those codings are pretty similar, so technically there is not much different than any other transaction. So those are core raw data for every single transaction we use. So that's pretty, pretty, pretty similar. #00:31:58#

Interviewer Maybe if we talk about it, what is presented to you, or I want to go a little bit deeper on those alerts or how the system works. What are the key metrics or outputs you focus on when using the system? How do you decide, or what metric is important for you to determine if an alert is a fraud or not? Can you explain a little more, a little bit deeper on that? What metrics do you use, or what outputs do you work with? #00:32:46#

Interviewee OK, the output in the monitoring system is that a breach of a specific rule, which is considered fraud-related, immediately highlights the given alert as a potential fraud alert to be reconciled or resolved. However, there can also be completely different card alerts that are not generated solely because of fraud. A rule breach can ultimately lead to a closer look at the cardholder's activity. That's the job of a human being nowadays, to determine whether it's possible that if you pay at 2:00 PM in Zurich, you could buy a coffee at 4:00 PM in Bangkok. Technically, those are pretty simple checks. Simple rules to follow can help spot potential fraudulent transactions. These rules operate based on specific breaches, which notify

you about jurisdictional or country-based differences. You can immediately spot something that is clearly impossible, like being in Zurich one day and the same day being in Dubai and back in Zurich. #00:33:23#

Interviewer I mean, with the current rule-based approach, it doesn't show you a probability, like saying this transaction has a 90% chance of being fraudulent. It simply indicates that there was a rule breach. #00:35:30#

Interviewee No. #00:35:42#

Interviewer You have to look at that transaction, and based on the information provided, you will compare it. #00:35:45#

Interviewee Yes, exactly. #00:35:47#

Interviewer Is it possible or not? Based on that, you decide if it's fraud or not. That's correct. #00:35:51#

Interviewee Yes, exactly. Usually, the most commonly used fraudulent typology or rule breach is activity outside of the cardholder's country. So if you are from Liechtenstein or Switzerland and you travel abroad, we are immediately notified about the fact. Many of our employees with cards generate an extreme number of fraud preventive alerts just because of this factor. In the end, it makes sense. If I go back to the weaknesses, larger banks with dedicated fraud prevention departments have the possibility to allow cardholders to notify them in advance. For example,, hey bank, tomorrow I'm leaving and traveling to Vietnam. This notification helps avoid any problems and saves a lot of preventive measures. Currently, for us, we can see when a person appears in Dubai, Vietnam, or the Philippines, and we immediately need to handle this. #00:35:58#

Interviewer But the problem is, what about online transactions? If you pay for something online, that's a different issue, right? #00:37:44#

Interviewee Yeah, that's a different piece, definitely. Even if you are sitting in Zurich and you pay somewhere globally, if the settlement of the card transaction is in the Philippines, it will immediately create an alert. But technically, it can be like noise because you are visiting or buying something from a site in the Philippines. Still, the residency country risk rule makes sense as a preventive measure to handle it. Again, these are examples of where AI can be used to get rid of such problematic patterns which are technically good transactions. We need to get rid of unnecessary and normal transactions that create alerts, and that's it. #00:37:51#

Interviewer Maybe also talk about when an alert is created. How important is the explainability of the system's decisions to you and other stakeholders? You have to show others why we decided that way, right? So how important is explainability for you as a user? #00:39:04#

Interviewee Definitely. But that's something that is not specifically law-required. It's something that, as a compliance officer, you need to set as part of your DNA. This is a core functionality or core setup, you need to be able to justify and correctly explain why something is flagged. You clearly cannot just say, hey, Valentin paid in the Philippines. You need to say, Valentin paid in the Philippines because he's there on annual leave for two weeks. That's the proper explanation. So technically, this is something really, really basic and common. That's why I'm saying I don't distinguish much between the resolution of a fraud alert and a money laundering or financial crime alert. The ability to explain the feasibility of such a transaction is clearly key and must be here, no matter what the root cause of the alert creation. #00:39:23#

Interviewer OK, thank you very much. Now we are going to talk about your requirements or expectations for such a system. From your perspective, what should be the primary goals of an AI-based fraud detection system, specifically for debit card transactions? What are the primary goals? #00:41:06#

Interviewee The primary goal is to detect fraudulent transactions. Simple as that. #00:41:37#

Interviewer Also, regarding the interaction with that system, what do you expect from it? Should it be explainable? Should it be fast? What are your expectations? #00:41:44#

Interviewee It must be fast and really, really precise in the detection because that matters. Technically, the system should learn from the patterns and specifics of each cardholder. As I said, each cardholder works differently. However, we can debate on that because, for instance, if a company provides employees with a card for expenses, it's probably much easier to handle. You would expect payments for petrol, hotels, accommodations, air tickets, and so on. And all those restaurant bill transactions. Technically, it should be really easy to manage, and you probably wouldn't spot something like eBay purchases on such a card. Back to the system requirements: at the moment, every bank and every card-settling monitoring person suffers and struggles with the detection of noise transactions. If AI could help us get rid of 50% of unnecessary alerts, it would mean that the remaining alerts would be of reasonable and better quality. This way, we wouldn't flag a payment for Starbucks just because you always purchase coffee at Costa

and never at Starbucks. That's something that can be funny and can be spotted once a day and resolved easily. But of course, the principle of how the fraudster will misuse your card is probably a different topic. I would like to get rid of the noise alerts because it will free up our hands to focus more on transactions that could actually be considered fraud. That's just it. #00:41:55#

Interviewer And then talking about requirements, what do you think are the must-have features of an effective AI-based fraud detection system? First of all, it needs to be able to learn from customer behaviour, right? Understanding how the customer behaves is an essential feature that an AI-based model should have. Is that correct? #00:45:00#

Interviewee Exactly. #00:45:17#

Interviewer Maybe also take into account the information provided by the client or customer on how they intend to use the card. What do you think? What other features must the AI or the fraud detection system have? #00:45:26#

Interviewee Yeah, all those components are important. You need to tell the machine how you plan to use the card, including limits, frequency, etc. However, the reality and the negative factor here is that it never fits perfectly. If you ask someone how frequently they will use their card, they might say three times, but then they use it ten times, and that would be absolutely OK. It's really hard to make that setup reliable and correct. And that's nothing wrong, because usually you just never know how frequently you're going to use the card. So, but the AI, especially in my conversation with a lady from an Israeli AI company, she said they don't care about what the client tells them. They take the data, run it through the system, and the machine decides on its own. It analyses the activity and tells you, hey, I have a problem with these transactions. It highlights them as problematic. You then review the alert and might decide, no, machine, this doesn't make sense. Even though you considered a million variances, I'm still OK with that transaction or set of transactions you consider suspicious. The system then accepts your decision and learns from it. This feedback loop is the second key component. It's critical to take the effort and time invested in the analysis by the human side and incorporate it into the system's learning process. That's the key factor. #00:45:40#

Interviewer And also, if we talk about an effective AI system, it must be explainable. It must be easy for you to see why the AI model made its decision. In the end, you need to understand why one set of transactions was flagged as fraud, why another was not, or why an alert was auto-closed. It needs to be explainable and traceable, so you can see why the system made that decision. Correct? #00:48:31#

Interviewee Exactly. And that's that. #00:48:49#

Interviewer And which data was the focus for that decision? On which data was the decision based? Correct? #00:48:57#

Interviewee Exactly. And that exactly makes the difference between good AI providers and bad AI providers, right? Imagine you're the auditor or regulator with no clue about this, no technical background. This is all about common sense. You need to show them easily and explain how the machine is set up, how the result was gained, and why. If you make that readable or understandable to a human being, you've achieved something significant. That's my feeling from the current setup from a regulatory point of view. There is a real misunderstanding or a lack of understanding of the setup and miscommunication because of it. No one from the given AI teams is clearly able to explain at the level of the auditor or even an analyst why the machine decided a particular result is OK or not OK. #00:49:07#

Interviewer I'd like to dig a little deeper here and ask you what specific objectives you aim to achieve with your fraud detection system. What is more important for you, reducing the false positive rates, improving detection accuracy, enhancing speed, or achieving higher regulatory compliance? What is the specific objective for you, or what are you aiming for with such a system? #00:50:58#

Interviewee Or all of them, right? All of them are important. Each chapter you mentioned has a specific and deep contribution. Everyone wants to be aligned with regulatory obligations. However, technically, we can say that it's never been clearly set. Regulators don't know exactly how it should be done. They just want us to closely monitor transactions in the right way, to spot it, identify it, block it, and act accordingly. They don't care how we do that. They just make it hard for us if we have a fraud in front of us and we don't see it. Clients call saying, hey, my card was stolen, my card was skimmed, and I lost my money, and you did nothing. So, technically, it's still a preventive measure. We need to protect the money of our customers. Full stop. So all the things around it are just pieces of the puzzle that need to be delivered. No one wants to waste time with noise alerts that are absolutely stupid and unnecessary. That's clear. Technically, the aim is to increase the quality of specific alerts generated, to clearly focus on the appropriate alert. That's what's wanted and needed. The core setup needs to enable us to quickly address alerts. The speed is crucial here. I don't feel there is a problem with the speed of alert generation; rather,

it's the quality of the alerts. We need to be faster in concluding, resolving, blocking, preventing, and protecting the customer's funds. That's the focus regarding fraud detection. #00:51:34#

Interviewer To sum it up, the ideal or successful fraud detection system should learn from behaviors, be explainable, and be traceable. It should be transparent about which data the decision is based on. It should be easy for you to understand and trace why it made a particular decision. The main goal is not only one objective, like reducing false positives, but a mix of everything. It's about balance, increasing accuracy, and efficiency. #00:54:30#

Interviewee Yes. #00:55:15#

Interviewer It's about reducing false positive rates and achieving higher regulatory compliance. It's really about finding that balance. The key is to reduce false alerts and automate more so that you have more time to spend on real cases and other important tasks. #00:55:20#

Interviewee Yeah. #00:55:40#

Interviewer And also, we are still talking here about detection and not prevention, right? #00:55:40#

Interviewee Yeah, exactly.

Interviewer In the future, is it more important to rely on prevention or detection? What do you think? #00:55:49#

Interviewee Prevention is the key. If you conclude and probably spend a lot of effort on the analysis of fraud prevention measures around the globe, the weakest link in fraud is still the customer. We can have better monitoring systems, AI, and everything else, but if the customer's discipline is weak, it all falls apart. So, it's really a piece of the puzzle. Technically, the education of the cardholder is crucial to ensure security and fraud awareness. This is one of the key parameters that must be in place. As far as I know, it's not happening effectively. We can see that from the numerous cases where customers fall victim to fraud. Educating cardholders to be careful and aware is essential. #00:55:56#

Interviewer For the last question, when we talk about future directions and innovations, what advancements or features do you hope to see in future fraud detection technologies? What do you expect in the future? What would be your wish? #00:57:25#

Interviewee Yes, I still want to protect the customer and ensure criminals cannot take their money. That's the very basic and core principle of fraud monitoring. However, I don't want to end with a pessimistic note. Criminals have always been faster than the police or us, and the modern world makes it even worse. The evolution of payment instruments and cards, including virtual cards, makes it vulnerable. On the other hand, I believe that the steps we have in place and those that will come in the future will help. I still believe that the future should be promising. #00:57:45#

User 2 at amnis

Date: May 15th, 2024

Time: 15:00

Place: Microsoft Teams

00:00:00 – 00:44:27

Interviewer Let us start with the first questions. These are especially about you and your background. Could you describe your role and responsibilities related to fraud detection, specifically for debit cards? What is your experience or what are your responsibilities in fraud detection in general? #00:00:04#

Interviewee I'm a junior compliance officer. I joined amnis in August 2022. My role regarding debit cards and fraud detection is to monitor, identify, and resolve various cases related to card transactional monitoring. This also includes the detection of fraudulent activity and the resolution of such activity. #00:00:27#

Interviewer Can you describe your previous experience? Was it also in fraud detection or just in compliance? #00:01:03#

Interviewee No. It was mostly regarding the onboarding and KYC part within the bank. #00:01:10#

Interviewer Thank you. Can you go further, especially regarding your experience with AI-based systems for fraud detection? Have you had any experience, or is this the first time you are dealing with AI fraud detection systems?

Interviewee AI fraud detection systems were a new thing for me at amnis. As I said before, in my previous job, it was regarding KYC, document collection, etc. So, really, my first steps in this were here at amnis with Hawk. My general view on this is that it can be very helpful if set up properly, which is an important "if." I believe we're going to get to that later. #00:01:17#

Interviewer Thank you for that. Now, we are going to talk about the current fraud detection systems and the challenges you face with them. My next follow-up question is, what fraud detection systems are currently used at our organization, and how do they function? Additionally, based on your experience, what are the common systems or approaches used across the industry for fraud detection? Could you tell me more about that? #00:02:05#

Interviewee What is important to mention is that, as I mentioned, this is the first time I have been really introduced to AI for the past two years here. It's been the first real interaction with artificial intelligence within compliance. Throughout this whole time, we still use the same provider. I can see the improvements that have been made over the past two years, but I really have nothing to compare it to in terms of my personal experience, which I think is important to mention. However, I am trying to widen my horizons. I have attended a few webinars to see how AI can be used to our advantage and to help me find other ideas and possibilities for AI. When it comes to the other part of the question, which is what detection systems we are using within our organization, the name that was mentioned before, HAWK, is based in Germany. This is the provider that we've been using not just for card and fraud detection in general, but also for transactional monitoring and all screening. #00:02:49#

Interviewer How does it function? Do we send them all the information, or how does it work? Can you provide a process of how you interact with Hawk? How does the system currently implemented work? #00:04:30#

Interviewee So, for fraud detection in terms of card activity right now, the AI did not go through the proper AI maintenance, so to say. The AI itself does not have what I consider one of the most important features, which is the learning ability. Hawk, in terms of card fraud monitoring right now, can identify possible fraudulent attempts based on a preset configuration staged by our team and the Hawk team. However, it does not currently learn by collecting data and using it as intended. I hope this will change in the future. To give you a proper example, if there is a customer who is making ATM cash withdrawals, specific rules are set up to trigger an alert related to fraudulent activity. That means, let's say a card which was issued in Germany suddenly makes a withdrawal in a high-risk jurisdiction, such as Morocco. In this case, there is a card limit set up, as in other countries. For now, the limit is €1'000 or the equivalent. So if the customer wanted to withdraw more, they would have to make three separate withdrawal attempts, resulting in three ATM withdrawals within a period of 60 seconds. This specific rule doesn't trigger an alert on its own, but it does trigger the alert for high-velocity ATM usage, which is one of the many rules we are using right now. As soon as this transaction happens, a notification is sent as an alert case through the HAWK platform, notifying us about the activity. It's then up to me to look into the case and resolve

it. If I determine the case is a false fraudulent attempt, meaning it was really the customer, and we confirm this with them, perhaps they are with a travel agency, etc., I would close the alert. Ideally, next time, the AI should be able to identify that this is expected behavior for this customer. However, this feature is currently not included. This means every time a customer triggers this rule by withdrawing in a high-risk jurisdiction multiple times, an alert will be created and require manual resolution from one of our officers. #00:04:43#

Interviewer To sum it up, if I understood you correctly, the current fraud system still uses a rule-based approach. This is, I think, mostly common across the industry. You have rules set by the compliance team, and the alerts are triggered according to these rules, not by AI. #00:08:27#

Interviewee Exactly. #00:08:44#

Interviewer It mostly works now by a ruleset set by amnis. Is that correct? #00:09:02#

Interviewee Yes, that is correct. #00:09:06#

Interviewer And if there is, as you said, the example before, maybe it's the amount or whether it's the jurisdiction where the card was used, there are certain rules. If those rules are broken, an alert is created on an online platform where you have your alerts. Then you have to work on it or clarify if it's really a fraud case or not. Is that correct? And were those rules defined by you, by compliance, or also by HAWK? #00:09:07#

Interviewee Exactly, that's correct. We have been given a list of rules to choose from, and the amnis compliance team has specifically chosen which rules, thresholds, and limits to use. So it was set up by us. #00:09:39#

Interviewer Are they official rule sets? Just a follow-up question. Can I see what is common or what kind of rules are typically used if I wanted to search for them? #00:10:06#

Interviewee It's really tricky because we asked Hawk to give us suggestions on which rules to focus on, but it's impossible to get a straight answer because they don't want to be held accountable. So, it was like reinventing the wheel. I believe there is a general idea of what the rules should look like, but I don't think there would be a list somewhere like there is for sanctioned entities. I don't think there is a similar list of suggested thresholds or limits established by an organization like OFAC or FATF or anything like that. it's really up to each organization using an AI system to set up those rules on their own. #00:10:23#

Interviewer OK, perfect. Thank you very much. So, to sum it up, the current system is rule-based with no AI in use, making it quite rigid and not flexible. This leads to my next question, what are the strengths and weaknesses of the current fraud detection system, especially since it is rule-based and does not learn from the behavior of the client? Are there any other weaknesses? #00:11:44#

Interviewee I would say that most issues are created when it comes to rules related to countries. High-risk jurisdiction-related rules, whether we are talking about payments or cash withdrawals, trigger an alert for every single such transaction. Which, of course, makes sense to a certain degree. But this covers 50% of the current card-related transactional monitoring, not just fraud-related, when it comes to our debit card usage in foreign countries, specifically in high-risk jurisdictions. This is a very specific example, but I believe that most of these cons can be resolved with the implementation of artificial intelligence. #00:12:24#

Interviewer So again, the weakness of the current system is that it's a rule set that does not adapt to the behavior changes of the client, and it's really inflexible. Companies operate around the clock and use it everywhere, unlike private individuals who live in one country. The rule set is not suitable for such clients, right? #00:13:51#

Interviewee Without risking exposure to true hits or not being alerted when it really comes to actual fraudulent activity. #00:14:37#

Interviewer Do you think this current system has some strengths compared to other systems, or do you think it only has weaknesses? #00:14:46#

Interviewee I would say that thanks to those weaknesses, we are, and I don't want to jinx it, but we are actually bulletproof. If you take a look at the current trends in terms of fraudulent activity and popular ways to scam people, we've got all of those covered. Thanks to the limits and thresholds and the current rule set being so stiff, as you mentioned, and actually a little bit strict, strict is not the right word, but it's the closest to explain what I'm trying to say, it actually ensures we are alerted in time when it comes to fraudulent activity. #00:14:59#

Interviewer But that, in conclusion, I would say this creates much more alerts than it should, right? #00:15:55#

Interviewee Definitely. #00:16:02#

Interviewer It means also in the end you have much more work because, OK, you have a good framework that is really safe, but in the end, as you said, it doesn't adapt. So it creates many more alerts that lead to much more manual work, right? #00:16:03#

Interviewee That's true. #00:16:21#

Interviewer What are the primary challenges you have encountered with AI-based fraud detection for debit card transactions? Are you now working on AI-based fraud detection? What kind of challenges have you encountered there? I'm asking specifically about AI-based fraud detection systems, not rule-based ones. #00:16:21#

Interviewee Well, if we're talking specifically about card-related fraud, there's really nothing to share because everything is triggered based on the rule setup. My experience with AI in this case is limited because we have not been able to turn on the AI for this specific group of alerts. Therefore, I don't have that experience. On the other hand, in other cases, in other groups of alerts related to non-card activity but standard transactional monitoring, the AI is implemented there, and it's improving with each transaction. We have been using it for a while now, and I would say that each month, if I did a review, the results would be better and better. Of course, another important thing is that as the company grows, there are more transactions, more activity, and more customers coming in every single day. Without AI implementation, it would be impossible to maintain the alerts, and we would be overdue with all of them. In the case of standard transactional monitoring, the AI is helpful. There are still cases where you wonder why it created an alert for a certain transaction that we've dealt with many times before. Sometimes, it happens that you think, "Wow, it's interesting that this got the attention of the AI, and it makes sense." In that case, I hope the feeling and feedback will be the same for card transactions and card fraudulent activity when it comes to AI. I hope so. I believe in that. #00:16:44#

Interviewer Thank you very much. Then let's go to the next one. Let's talk about the design, user engagement, and system explainability. How do you interact as a user with these kinds of systems? How are users typically interacting with the fraud detection system? To sum it up, you have an online platform where you see all the alerts. Is that correct? Do you see all the card transactions that went through and the alerts that were created? #00:19:20#

Interviewee That is correct. You have access to data for all the transactions initiated by a card issued by us, and it's separated into multiple smaller groups. These groups include transactions that triggered an alert, previously closed alerts, and previously closed cases. You also have the option to view which alerts have been auto-closed by the AI. Right now, this field would be empty, but in the future, we hope that the majority of the alerts will be visible in the auto-closed category. Right now, how it works is that the alerts for card activity are together with the standard transactional monitoring of those cards and also fraudulent activity. We have the same overview of card transactions for both standard monitoring and fraud attempts. What we are trying to accomplish is to separate fraud attempts from standard monitoring. Since the AI is not enabled right now, we have these together. Some of the rules might seem similar because the rule setup for fraudulent activity is slightly different from the rule setup in standard transactional monitoring. The rules for transactional monitoring were actually taken over from the non-card activity group of payments. In certain cases, it doesn't really make sense for the alerts that are created. For example, the most frequently triggered rule, whatever the specific rule is, relates to the frequency of transactions. This means the customer sends a certain amount of money so frequently that it triggers an alert. This rule makes sense for standard payments but not as much for card payments. The same customer can pay €10 for parking and, at the same time, pay hundreds of thousands to suppliers. The difference in terms of volume and frequency is not as relevant for card transactions. I believe everyone has a good idea of what a standard card transaction volume should be. While some might pay tens of thousands via card, it can happen, but it's not really the majority of customers. In the future, I hope that we will be able to clearly separate fraud monitoring so that it's its own category accessible to us. That's what I wanted to say. #00:19:57#

Interviewer And what are the key metrics or outputs users like you focus on when using these systems? How are those metrics used to make decisions? Are there any metrics in use currently, or what is the output of the system, and what do you see? #00:24:03#

Interviewee In terms of resolution of the alert, it comes down to the manual resolution of the alert. As of right now, all the results and previous resolved cases are accessible to us. If looked into properly, there might be a way to review the transactions and how they were closed to identify which ones are causing the most problems. But it's really individual with each of those alerts. I can tell you what's causing the most noise, meaning what usually triggers unnecessary attention, and I can differentiate that from a real possibility of a fraudulent attempt. However, there is no specific guideline to follow that would help resolve the alert. I'm not 100% sure this is the answer you were looking for. #00:24:30#

Interviewer Because, with a rule-based system, it's quite difficult to have specific metrics, right? You just have the system that alerts you when something is triggered, and you cannot compare it easily. It's probably not possible for you or for compliance to say how many alerts were generated that were not really alerts. I

mean, is it possible at the moment to see that metric or to see how many alerts were generated and how many of them were not fraudulent? #00:26:07#

Interviewee I believe there is a way to export the data from the platform. Again, I don't want to give you a false answer. I'm not 100% sure about that. Each compliance officer has different user access levels. So, for me, it would be completely different than for one of the senior compliance officers. I'm sure it's possible within the Hawk platform in general to export such data, but I'm not sure how exactly that works because it's not really my expertise. It's not something I would typically do. I don't want to say that I'm not interested, but it doesn't intervene with the manual work that needs to be done in terms of closing the alert. #00:26:41#

Interviewer You know, when we talk about alerts, how important is the explainability of the system's decisions to you and other stakeholders? Could you also explain how this impacts trust and usability? If an alert is generated, how important is it to understand why this alert was triggered? #00:27:32#

Interviewee It's very detailed. Since it's data-driven, all the alerts come with a brief explanation of what has happened, why the alert was triggered, and what limits have been breached. It compares the activity that triggered the alert to the previous activity of the customer, and it's all very well and in detail described in the alert detail. #00:28:03#

Interviewer So explainability is really important in your case, right? #00:28:48#

Interviewee Yes. #00:28:52#

Interviewer And, I believe explainability for auto-close, when it is possible, is crucial. It's important that the user has an overview of why it was auto-closed or triggered and based on which data, right? Explainability and traceability for why something happens or was decided are really important for compliance. Is that correct? #00:28:53#

Interviewee I agree. Just as you said, for improvement in the future, this is necessary. #00:29:23#

Interviewer Now, moving to the next section, let's talk about requirements and expectations, especially the expectations you as a user have for such systems. From your perspective, what should be the primary goals of an AI-based fraud detection system, especially for debit card transactions? What do you think the primary goals must be from your perspective? #00:29:33#

Interviewee Well, alerting users about fraudulent activity is the primary goal, right? That's why the system is created. I would say it's important to find a balance somewhere in the middle. AI and anti-fraud systems that create alerts shouldn't create no alerts at all, as that would be strange and not the expected result. In a perfect world, in a utopia, of course, that's how it would be, but it's impossible. On the other hand, it's also important not to create unnecessary noise. I think this balance is the most important part. Right now, in our case specifically, we are leaning towards the second part. With the rule setup and thresholds being what they are, there are probably more alerts than I would anticipate. So, the primary goal should be to identify all the fraud attempts and deal with them appropriately. #00:30:03#

Interviewer And when you talk about features, what are the must-have features of an effective AI-based fraud detection system? This can include how it works and how the information or interaction with the system should be. What are the main features you think such a system must have? #00:31:49#

Interviewee In this case, for this question specifically, I don't think I'm the best person to answer. As I said, my experiences are only with one AI system, so I don't have anything to compare it to. #00:32:12#

Interviewer What do you think generally? Based on what you have learned during your workdays, from your gut feeling or your point of view, what features do you think are really needed for an effective AI-based fraud detection system? #00:32:32#

Interviewee I need to understand why I'm looking at the data that I'm looking at. What are the possibilities or what to do with the data that I have received and what the outcome should be? It's the same as with every other field. So, I'm getting data, and I need to clearly understand what data I'm getting, why I'm getting it. Then there's really solving the problem and the resolution. I believe that the current system we have enables all of those three. #00:32:47#

Interviewer It is especially important for you to understand why this alert is created, what you have to do, and based on the data that was provided. As you said, first you need to understand what kind of data is presented and how you can decide based on this data, right? It's really about how this data is presented to you, how understandable it is. Also, I would say a must-have feature is that it assists you in decision-making. That's also one of the more important goals, right? It must help you decide. It should not create too many alerts, but not too few either. As you said, if there are no alerts, something is wrong. If it's too many, it's too

much work. It needs to balance. The data must be clearly presented and it must help you decide correctly if it is fraud or not. #00:33:42#

Interviewee Yes, agreed. #00:34:41#

Interviewer And maybe if you talk about how regulatory requirements influence your approach to fraud detection. Are there many laws and regulations on how you have to approach these kinds of issues, or is it more about protecting your clients? Are there any regulatory requirements that influence your approach? #00:34:43#

Interviewee I would say the importance of protecting the customer and ourselves is the number one priority. When it comes to regulatory requirements or limitations in terms of monitoring, I'm trying to think of other examples. The only thing I can think of is really the country-based approach. I don't think there are specifically set regulations for what is the maximum or minimum limit in terms of volume and amount of transactions that a card can make before it should be triggered as an attempt for fraudulent activity. I don't think there's anything like that, something that would be unified for all financial institutions. It can be based on the specific laws of specific countries, but it's not really something that influences us and the work that we are dealing with. So I would say the only relatable explanation here would be that there are certain high-risk jurisdictions in the world where the regulators actually say if a transaction is in this or that country, it needs to be closely reviewed. But otherwise, I don't think there are thresholds or limits for other kinds of transactions. #00:35:07#

Interviewer Good, thank you. And then if you talk about the specific goals and desired outcomes now, a little bit when we look into the future, what specific objectives do you aim to achieve with your fraud detection system? For example, we have talked a little bit about this already, but I want to know from you: Is it reducing false positives to make it more efficient, so that it's not generating more alerts than it needs to? Is it improving detection accuracy, enhancing speed, or achieving higher regulatory compliance? What would you say is more important, efficiency, detection accuracy, or speed? What do you think is the most important aspect of such a system? #00:37:02#

Interviewee But before I answer this question, I actually have something to add to the previous one. Do you want me to say it immediately or do we do it after the answer? It doesn't matter. #00:38:01#

Interviewer If you have something to add, you can just say it. #00:38:10#

Interviewee I just wanted to add that there are certain examples where the limit for a transaction actually makes a difference and is regulated. There are scenarios where transactions under a certain threshold are not reviewed as in detail and do not fall under certain regulatory rules and instances. I don't think this translates to cards. I think it would only apply to wire transactions. For example, if a transaction is slightly less than $10'000 and you know that the country of the issuer of the transaction has a regulatory limit for $10'000 for further review, then this is something that would be worth looking into. But I don't see a way how that would translate to cards because the POS can be in any country. In all of those spectrums, I think that when it comes to fraudulent activity, speed has probably slightly bigger value in this case. From a practical point of view, if you get your card skimmed, for example, and someone has your card details and starts doing fraudulent transactions, speed is probably more important than in other scenarios, such as standard wire transactions. Your goal is to stop the fraudulent activity as soon as possible from its initiation point. I'm not saying that faster is better or more important than the other aspects. Just from a practical point of view, if you think about it, that's my assumption. #00:38:12#

Interviewer So, it's all about real-time processing, right? The data comes in real time, and how fast can the system decide? Fraudulent activity must be detected quickly. But as you said, in the end, it's a balance. Everything must be balanced. #00:40:52#

Interviewee Exactly. Talking about speed, I would not say that it's more important than, for example, detection accuracy. If we created alerts really fast and blocked many cards due to the possibility of transactions being fraudulent, but 99% of them are false positives, that wouldn't be effective. So, balance is important. #00:41:09#

Interviewer Perfect. Now, the last question.: When you look to the future, what advancements or features do you hope to see in future fraud detection technology? What do you expect in the future? #00:41:56#

Interviewee I believe that with the way technologies have been improving in the last two decades, how fast everything is advancing, it's remarkable. Five years ago, I didn't even know of the possibilities that AI could be helpful to mankind in general. And now, it can create funny pictures based on what you write down, right? The advancement is unstoppable. I am very curious to see what it will give us in the future. In the end, I would hope for better results, specifically compliance-wise and transactional monitoring-wise, for fraudulent activity. I don't think the permanent termination of fraudsters is on the menu. I don't think it's possible because as fast as we are, or the good guys are improving, the bad guys also have the possibilities to use AI against us. I think it's like a double-edged sword. I don't think I'm worthy of giving an

expectation because it might be promising, but on the other hand, it also might not be. And that's like a little bit of a how to say. I can't say for sure. I don't think we're going to have AI in our everyday lives, at least not us yet. Maybe our children one day, but yeah, we'll see. To answer the question specifically, I don't have a particular thing that I think could be developed. I believe there are greater minds working on such things, or even the AI itself is working on it already, right? We will see. #00:42:13#

User 3 at amnis

Date: May 17th, 2024

Time: 15:00

Place: Microsoft Teams

00:00:00 – 00:38:36

Interviewer Ich würde gerne mit dir über dich als Person und deinen Hintergrund im Bereich Fraud Detection oder Betrugserkennung, speziell im Bereich Debitkarten, sprechen. Was sind deine Erfahrungen in diesem Bereich? #0:19#

Interviewee Ich bin verantworte den Compliance-Bereich. In diesem Bereich ist auch die Betrugserkennung angesiedelt. Die Firma hat kürzlich neue Debitkarten eingeführt, und deshalb stellte sich die Frage nach Fraud Management im Rahmen des Operation Risk Managements der Firma. Ich bin relativ neu in diesem Bereich und habe daher wenig Erfahrung mit Fällen von Betrugserkennung. #00:42#

Interviewer Hattest du auch schon Berührungspunkte mit AI-basierten Betrugserkennungsmodellen davor, oder ist das das erste Mal, dass du damit zu tun hast? #01:35#

Interviewee Ja, wir haben auch ein AML-System, Anti-Money Laundering, im Einsatz, das unter anderem Künstliche Intelligenz-Modelle einsetzt, insbesondere um falsche Positivmeldungen zu reduzieren. #01:46#

Interviewer Danke vielmals für den Hintergrund. Jetzt wollen wir vor allem darüber sprechen, was der aktuelle Stand der Dinge ist und was der Standard in der Industrie ist. Dabei geht es vor allem darum zu erfahren, welches System aktuell im Einsatz ist oder welche Systeme generell in der Industrie verwendet werden und wie diese funktionieren, insbesondere bei der Erkennung von Betrug bei Debitkarten Transaktionen. #02:06#

Interviewee Ja, ich meine, die Geldwäschebekämpfung wurde in den letzten 20 Jahren sehr stark vorangetrieben. Ein Treiber sind zum Beispiel Sanktionslisten, die den Drogenhandel, Terrorismus usw. bekämpfen sollen. Mit der Implementierung strenger Regeln für Finanzinstitute arbeiten diese oft regelbasiert. Das bedeutet, dass bestimmte Volumina oder Transaktionen für bestimmte Jurisdiktionen überprüft werden. Auf der einen Seite gibt es strikte Vorgaben, auf der anderen Seite gibt es im Moment einen sehr starken Fokus auf die Einhaltung dieser Sanktionsregelungen. Das führt dazu, dass Finanzdienstleister dieses Risiko möglichst minimieren wollen, was wir als relativ kleines Unternehmen stark spüren. Bei den Debitkarten geht es bei der Betrugsprävention vor allem darum, den Kunden und die Reputation der Firma zu schützen. Es geht weniger darum, die Firma selbst zu schützen, sondern vielmehr darum, die Regeln einzuhalten. Hierbei sprechen wir von Massnahmen, die präventiv wirken sollen, wie z.B. die Sperrung der Karte, um finanziellen Schaden zu verhindern. #02:38#

Interviewer Und wie du gesagt hast, wird in der Industrie derzeit mehrheitlich immer noch ein regelbasierter Ansatz verwendet. #04:24#

Interviewee Meine Einschätzung, ja. Was ich im Markt höre, ist, dass in den letzten paar Jahren das Potenzial von neuronalen Netzwerken durchgedrungen ist. Ich meine, dass die meisten Unternehmen auch an solchen Lösungen arbeiten. #04:34#

Interviewer Und wenn du sagst, wir reden hier von Betrugserkennung, sprechen wir dann über die Erkennung nach einer Transaktion oder über Prävention? Was ist bei amnis der Fall? Vor allem, was möchte amnis mit einem neuen System eigentlich erreichen und wie arbeitet amnis hier? Geht es hauptsächlich um die Detektion, also die Erkennung, nachdem die Transaktion erfolgt ist, oder darum, Transaktionen bereits im Vorfeld zu blockieren? #05:01#

Interviewee Wie gesagt, wir sind relativ neu in diesem Geschäft. Wenn du mich nach dem Ziel fragst, dann sage ich klar, es geht sowohl um Prävention als auch um Erkennung. Wenn man die Entwicklung und das Training eines Modells betrachtet, merkt man schnell, dass diese viel Training und Rahmenbedingungen benötigen, um gut zu funktionieren. Wir befinden uns momentan in einem Status, in dem wir vor allem Betrugserkennung betreiben. Das wird aber sicher in Zukunft ausgebaut, wenn die Modelle sehr stabil sind und wir gewisse technische Hürden überwunden haben. Man muss sich vorstellen, dass eine Transaktion innerhalb eines sehr kleinen Zeitrahmens überprüft wird, um dem Finanzintermediär die Möglichkeit zu geben zu sagen, ja, diese Karte ist gedeckt und die Auszahlung darf erfolgen. Wenn dann noch ein tiefergehender Check durchgeführt wird, muss das alles sehr schnell und automatisch ablaufen.

Ich persönlich rechne sicher noch mit etwa einem halben Jahr Vorbereitungszeit, bevor wir daran denken können, solche präventiven oder ex-ante-Checks durchzuführen. #05:30#

Interviewer Was denkst du, sind aktuell die Stärken und Schwächen des regelbasierten Ansatzes im Vergleich zu den neueren Methoden? #07:02#

Interviewee Also, bei dem, was alle machen, ist der regelbasierte Ansatz statisch. Je nachdem, welchen Risikoappetit eine Firma hat, ergeben sich unterschiedliche Herausforderungen. Wenn ich sehr restriktiv bin und streng kontrollieren möchte, um meine Risiken im Griff zu haben, erhalte ich zwangsläufig sehr viele falsch positive Ergebnisse. Diese müssen dann von Compliance-Mitarbeitern überprüft werden. Dabei weiss man, dass wenn ein Mitarbeiter 1'000 Transaktionen überprüft und nur eine davon tatsächlich positiv ist, er möglicherweise die restlichen 999 Transaktionen weniger gründlich überprüft. Das ist ein grosser Nachteil des regelbasierten Ansatzes. Andererseits, wenn ich risikofreudiger bin und alles durchlaufen lasse, habe ich die Risiken nicht im Griff, was meiner Ansicht nach ebenfalls riskant ist. Der Hauptnachteil des regelbasierten Ansatzes ist, dass strikte Regeln zwangsläufig eine hohe Anzahl von falsch positiven Ergebnissen erzeugen und sich nicht anpassen. Hier sehe ich den Hauptvorteil von Artificial Intelligence. Man kann den normalen Bereich für einen Kunden dynamisch anpassen. Wenn eine Transaktion unter den gleichen Bedingungen zwei- oder dreimal als gut bewertet wird, kann die Software lernen, dass sie wahrscheinlich gut ist. Erst wenn es deutlich grössere Beträge oder andere ungewöhnliche Situationen gibt, kann die AI eingreifen und die Transaktionen genauer prüfen. #07:13#

Interviewer Und das kostet natürlich auch Ressourcen. Wenn viele Warnmeldungen generiert werden, muss man natürlich auch ein Compliance-Team dahinter haben, das sich diese ansieht. Das führt zu zusätzlichen Kosten und Arbeitszeit, die eigentlich anders verwendet werden könnte. #09:23#

Interviewee Eben mit dem zusätzlichen Nachteil, dass dies nicht unbedingt zur besten Erkennung führt. Bei vielen falsch positiven Ergebnissen besteht das Risiko, dass die Mitarbeiter irgendwann nicht mehr alles richtig überprüfen. Ich weiss von verschiedenen anderen Firmen, dass sie genau diese Probleme haben. #09:40#

Interviewer Ich habe noch eine Frage, die mir gerade eingefallen ist. Amnis bietet ja vor allem Debitkarten an. #10:05#

Interviewee Ja. #10:13#

Interviewer Also bietet amnis Debitkarten für ihre Kunden an. Was ich mich frage, ist, ob es einen Unterschied gibt, ob man ein Betrugserkennungsmodell für Debitkarten oder für Kreditkarten hat. Behandelt man diese zwei Zahlungsarten gleich, oder gibt es Unterschiede? #10:14#

Interviewee Das ist jetzt nicht reflektiert, aber ich glaube, dass bei der Kreditkarte noch etwas grössere Risiken bestehen, weil sie auch offline eingesetzt werden kann. Viele junge Leute kennen das nicht mehr, aber früher gab es dieses 'Ratzgerät', mit dem man eine garantierte Zahlung offline auslösen konnte. Diese Zahlungen kommen natürlich erst mit einer gewissen Verzögerung ins System, sodass eine präventive Kontrolle dort nicht möglich ist. Bei Kreditkarten ist das Risiko daher höher, aber sonst ist das Einsatzgebiet gleich. Der einzige Unterschied ist, dass die Kreditkarte nicht gedeckt sein muss, also dass der Kreditcheck gegen ein Limit geht, anstelle von einem Kontostand. Ich sehe höchstens in den Offline-Transaktionen einen Unterschied. #10:41#

Interviewer Das ist jetzt wirklich nur eine Überlegung, die mir gerade in den Kopf gekommen ist. Bei der Debitkarte geht das Geld direkt vom Konto weg, es wird also sofort abgebucht. Bei der Kreditkarte nimmst du im Grunde genommen einen Kredit auf. Gibt es da einen Unterschied bei der Haftung? Bei amnis ist das Geld beispielsweise sofort weg. Gibt es bei Kreditkarten den Anspruch, dass die Zahlung gedeckt ist oder dass das Geld zurückgefordert werden kann? Das würde mich jetzt noch interessieren. #11:49#

Interviewee Grundsätzlich ist auch bei der Kreditkarte das Geld weg, sobald die Transaktion autorisiert ist. Das hat mit der Funktionsweise des Kreditsystems zu tun. In beiden Fällen kann der Kunde jedoch die Transaktion beanstanden, und der Anbieter muss den Fall dann prüfen und lösen. Die Kreditkarte birgt für den Anbieter ein etwas höheres Risiko, weil, wenn die Belastung der Kreditkartenrechnung monatlich erfolgt und der Kunde dann Widerspruch einlegt, dieser möglicherweise mehr Einfluss hat. Grundsätzlich sind die Haftungen aber ähnlich. Die Haftung der Anbieter beschränkt sich hauptsächlich auf technische Aspekte. Wenn eine Transaktion autorisiert und korrekt durchgeführt wurde, liegt die Verantwortung beim Kunden. Natürlich können gewisse Betrugsfälle entschädigt werden. Diese Aspekte unterstreichen die Notwendigkeit eines guten Überwachungssystems, das tiefgreifend wirkt. Beispielsweise wenn eine Transaktion in Zürich und zwei Stunden später in São Paulo erfolgt, sollte das System die Karte automatisch blockieren, da dies mit einer physischen Karte nicht möglich ist. #12:20#

Interviewer Danke dir. Jetzt zum nächsten Punkt. Hier geht es darum, was du von einem solchen System erwartest und was dir dabei wichtig ist. Damit möchte ich mit der ersten Frage starten: Wie sieht der Benutzerfluss mit diesem System aus? Also, wie interagieren die Benutzer typischerweise mit

Betrugserkennungssystemen? Welche Aspekte erleichtern diese Interaktion? Wie sieht die allgemeine Interaktion mit solchen Betrugserkennungssystemen aus? #14:12#

Interviewee Ich kann da einfach aus meinen Erfahrungen mit meinem Compliance-Team sprechen. Für uns ist ein System mit einem benutzerfreundlichen Interface wichtig, bei dem die Künstliche Intelligenz im Hintergrund arbeitet und nicht prominent dargestellt wird. Das System sollte so aufgebaut sein, dass es grundsätzlich regelbasiert arbeitet, aber durch die Künstliche Intelligenz falsch positive Ergebnisse erkennt und diese automatisch filtert, sodass sie dem Compliance-Mitarbeiter gar nicht erst angezeigt werden. Das hat zur Folge, dass der Compliance-Mitarbeiter nur das sieht, was er gewohnt ist, Regeln, die Hits generieren. Die Anzahl der Hits wird jedoch bereits durch das Modell reduziert. Die Interaktion für den Compliance Officer bleibt somit wie gewohnt. Unsere Mitarbeiter sind keine Mathematiker, sondern sehr genaue Personen, die die Alarme gewissenhaft bearbeiten. Wenn sie feststellen, dass bestimmte Alarme regelmässig auftauchen, sollten sie Feedback an das Design-Team geben, um das Modell weiterzuentwickeln. Das Modell lernt natürlich auch dazu. Ich sehe, dass selbst Mitarbeiter, die keine Mathematiker sind, mit unserem System von HAWK gut zurechtkommen, da es relativ transparent gelöst ist. Die Entscheidungen des Modells sind nachvollziehbar, und auch falsch positive Ergebnisse werden in der Liste angezeigt. Das System funktioniert gut, obwohl es bei grösseren Datenmengen möglicherweise komplexer wird. Ich bin sicher, dass sich das System in Zukunft weiterentwickeln muss. Wir stehen noch am Anfang und es gibt viel Raum für Verbesserungen und Anpassungen. #14:53#

Interviewer Du hast ja auch schon die Erklärbarkeit angesprochen, die bei solchen Modellen immer sehr wichtig ist. Worauf achtet man dabei speziell? Was ist dir wichtig? Beispielsweise, auf welchen Daten die Entscheidung getroffen wurde oder welche Regeln gebrochen wurden? Was möchtest du wirklich sehen, um nachvollziehen zu können, warum eine Entscheidung getroffen wurde? Was ist für dich wichtig, um die Erklärbarkeit solcher Systeme sicherzustellen oder hinreichend gegeben zu sehen? #18:24#

Interviewee Das ist eine schwierige Frage. Wir sprechen hier von einem Teil des Operational Risk Managements, also von operationellen und rechtlichen Risiken sowie vom Kundenschutz. Insbesondere bei Debitkarten ist es wichtig, dass das Modell praktische Auswirkungen hat. Wenn ein Betrug nicht erkannt wird, entsteht ein finanzieller Schaden, den entweder wir oder der Kunde tragen. Andererseits, wenn das System Karten blockiert, die eigentlich nicht blockiert werden sollten, wird sich der Kunde beschweren. In diesem Bereich ist das Ergebnis entscheidend. Das Modell darf meiner Ansicht nach im Bereich des Kartenbetrug-Monitorings auch intransparent arbeiten, solange das Ziel erreicht wird. Das Ziel ist wichtiger als der Weg. Das unterscheidet sich etwas von anderen Bereichen, wie beispielsweise dem Geldwäsche-Monitoring, wo ein Regulator nachvollziehen möchte, warum eine bestimmte Transaktion nicht als verdächtig markiert wurde. Ich spreche hier natürlich aus einer unternehmerischen Perspektive. Es kann sein, dass Compliance hier eine andere Sichtweise hat, auch im Bereich Betrug. Aus meiner Sicht ist das Wichtigste, dass Betrug vermieden oder erkannt wird und dass das System, wenn es präventiv arbeitet, den Betrieb nicht unnötig unterbricht. Eine solche Unterbrechung würde nur Mehraufwand verursachen, wie das Entsperren von Karten und das Beruhigen von Kunden. #19:07#

Interviewer Das war eigentlich meine nächste Frage gewesen. Was sollte das Hauptziel eines solchen KI-basierten Betrugserkennungssystems sein? Du hast bereits einige Punkte angesprochen. Es geht vor allem darum, nicht nur die Detektion, also die Erkennung, sondern auch um die Prävention. #21:47#

Interviewee Ja. #22:09#

Interviewer In diesem Sinne sollte das System auch Explainability, also Erklärbarkeit, bieten. Es muss nachvollziehbar sein, warum Entscheidungen getroffen wurden. Gleichzeitig muss es effektiv sein, das heisst, es darf nicht zu viele Fehlalarme generieren, aber auch nicht zu wenige. Zu viele Fehlalarme führen zu unnötigem Arbeitsaufwand, während zu wenige Alarme grössere finanzielle Schäden für den Kunden oder für amnis verursachen können. #22:09#

Interviewee Ja. #22:37#

Interviewer Was denkst du noch? Gibt es noch weitere Ziele, die sehr wichtig sind, wenn es um solche Systeme geht? Zum Beispiel, was ist dir wichtiger, die Verringerung von Fehlalarmen oder die Verbesserung der Erkennungsgenauigkeit? Oder ist eher die Geschwindigkeit wichtiger? Viele sagen, es ist immer eine Balance von allem, also es muss eine ausgewogene Kombination dieser Faktoren sein. #22:37#

Interviewee Es ist ganz klar eine Balance. Die Geschwindigkeit ist entscheidend, besonders wenn wir von Prävention sprechen. Diese muss hoch sein, sonst funktioniert das System nicht. Gleichzeitig wollen wir unseren Kunden ein zuverlässiges System bieten. Es darf nicht zu zufälligen Kartensperrungen kommen, aber es muss auch in der Lage sein, Betrugsversuche effektiv zu erkennen und zu verhindern, um finanziellen Schaden zu minimieren. Das bedeutet, es ist eine Kalibrierung zwischen diesen beiden Aspekten. Hier spielt auch der Risikoappetit eine Rolle, wobei der Kundenschutz natürlich im Vordergrund steht. Unsere finanziellen Risiken müssen ebenfalls gemanagt werden. Man will dort keine Risiken eingehen, sondern alles möglichst exakt machen. #23:01#

Interviewer Aber kann man generell sagen, dass das Ziel wirklich ist, die Geschwindigkeit zu erhöhen? Oder ist es wirklich ein Kompromiss, also wie man Kosten und Nutzen abwägt? Es ist ein Abwägen, wie stark man sich auf verschiedene Aspekte fokussiert. Wo kann man etwas nachlassen und wo muss man sich mehr konzentrieren? Es ist letztlich eine Balance zwischen allen Faktoren, oder? #24:25#

Interviewee Ja, vielleicht müssen wir das Themengebiet etwas ausweiten. Wenn wir von Debitkarten sprechen, gibt es bestimmte Muster, die darauf hinweisen, dass eine Karte gestohlen wurde und jemand versucht, Geld vom Konto abzuheben. Das bedeutet häufige Transaktionen unterhalb der Freigrenze, eine hohe Frequenz von Zahlungen in einem engen geografischen Raum oder auch online. Solche Muster kann ein gutes System erkennen. Ein gut aufgesetztes System wird an einem bestimmten Punkt zum Beispiel fordern, dass ein PIN-Code eingegeben wird oder überprüfen, ob der Karteninhaber tatsächlich über die Karte verfügt. Es gibt viele Varianten, wie ein solches System Risiken managen kann. Ich sehe das natürlich auch in einem weiteren Kontext. Wir sind Anbieter einer Online-Plattform, auf der unsere Kunden relativ viel Geld direkt verwalten können. Meiner Meinung nach ist es auch wichtig, dass das System Muster erkennt, wie zum Beispiel von welchen Geräten und Standorten sich Kunden normalerweise einloggen. Die Software sollte erkennen, wenn der Kunde ein neues iPhone kauft oder sich auf einer Reise befindet und entsprechend reagieren. Der Vorteil von Künstlicher Intelligenz liegt vor allem darin, dass das System dynamisch Entscheidungen treffen kann. Wenn der Operator diese Entscheidungen bestätigt, kann das System lernen, dass das Muster passt. Geschwindigkeit ist vor allem dann wichtig, wenn man präventiv arbeitet, z.B. beim Sanktions-Screening oder bei Kartenabhebungen. Entscheidungen müssen schnell getroffen werden, z.B. ob ein PIN verlangt wird oder die Karte blockiert wird. Diese Entscheidungen müssen sehr schnell und ohne manuelle Interaktion getroffen werden. Das Modell muss sehr präzise sein. Zusammengefasst kann man sagen, dass ein präventiv arbeitendes System schnell und präzise sein muss. Bei der detektiven Arbeit sind diese Faktoren nicht so wichtig oder nicht so essenziell. #24:53#

Interviewer Genau, dass hast du gut zusammengefasst, was das Ziel eines solchen KI-basierten Systems ist. Es geht darum, dass es lernt, wie sich der Kunde verhält und basierend auf diesen Aktionen Entscheidungen trifft. Im Gegensatz zu regelbasierten Systemen hat KI die Möglichkeit, zu lernen und dadurch bessere und präzisere Entscheidungen zu treffen. Ich glaube, das ist auch ein wichtiger Aspekt, wenn man solche Systeme implementiert. #28:15#

Interviewee Absolut. #29:02#

Interviewer Ein Hauptziel von amnis und anderen Anbietern, die solche Systeme integrieren, ist es, die Möglichkeit des Lernens zu nutzen. Was denkst du, ist der beste Ansatz in der Zukunft? Wird es eine Kombination aus regelbasierten und KI-Modellen sein, oder nur KI-Modelle? Werden die regelbasierten Ansätze weiterhin existieren, oder wird man sich in Zukunft vor allem auf die KI verlassen? #28:54#

Interviewee Das ist eine spannende Frage. Im Bereich der Geldwäsche sind die Regeln stark durch bestimmte Regulatoren oder Richtlinien getrieben, wie etwa durch die Financial Action Task Force, FATF. Es geht oft um die Sanktionierung von Ländern, Personen oder Firmen, und das bleibt meiner Meinung nach immer ein regelbasiertes Thema. Auch bei hohen Beträgen kommt man nicht um Regeln herum. Es bietet sich jedoch an, diese Regeln dynamischer zu gestalten, als es derzeit oft der Fall ist. Bei HAWK nutzen wir bereits Künstliche Intelligenz zur Reduktion von falsch positiven Ergebnissen, was ich als eleganten Zwischenschritt sehe. Vermutlich wird dies dazu führen, dass man sich stärker auf Modelle fokussiert, die auf Regeln basieren, aber durch KI dynamischer und anpassungsfähiger sind. In unserem spezifischen Fall mit KMUs und Zahlungsverkehrsfirmen wäre es sinnvoll, ein System zu haben, das auf dynamischen Regeln basiert. Zum Beispiel könnte ein Händler, der Bäckereigeräte importiert und verkauft, von einem System profitieren, das basierend auf der Webseite der Firma ein normales Muster erkennt und Abweichungen automatisch anzeigt. Das wäre besonders nützlich, wenn der Händler regelmässig bei bestimmten Herstellern einkauft. Ich glaube, dass in Zukunft regelbasierte und KI-basierte Modelle kombiniert werden, um die Stärken beider Ansätze zu nutzen. Für Compliance Officers wäre es wichtig, dass diese Systeme transparent und nachvollziehbar sind, um deren Entscheidungen zu unterstützen. Rein KI-basierte Modelle könnten in der Zukunft dominanter werden, aber wahrscheinlich in Kombination mit grundlegenden regelbasierten Prinzipien. #29:28#

Interviewer Was erhoffe ich mir in Zukunft von solchen Systemen? Was erwarte ich, was noch kommen wird? Und was hoffe ich, was in Zukunft, vielleicht in fünf Jahren, möglich sein wird, was jetzt noch nicht realisierbar ist? #33:00#

Interviewee Im Moment werden diese Modelle von Programmier-Teams, Mathematikern und Technikern entwickelt, was oft zu einer gewissen Distanz zwischen deren Verständnis und den praktischen Bedürfnissen des Business führt. Ich würde es begrüssen, wenn ein Compliance Officer in einem benutzerfreundlichen Interface die Modelle optimieren und steuern könnte. Dies würde die Kommunikation zwischen IT und Business verbessern und effizienter gestalten. Im konkreten Fall von HAWK wäre es wünschenswert, dass der Compliance Officer die Möglichkeit hat, das Modell anzupassen und zu optimieren, ohne tiefere technische Kenntnisse zu benötigen. Das würde die Transparenz erhöhen und die praktische Anwendung verbessern. Ich bin mir bewusst, dass dies im Moment noch Wunschdenken ist, da die Künstliche

Intelligenz noch nicht so weit entwickelt ist, um dies voll zu unterstützen. Vielleicht werden wir irgendwann Modelle haben, die eng mit der realen Welt und den Business-Impulsen integriert sind und sich dynamisch anpassen können. Ein weiteres Hindernis ist die Risikoaversion der Anbieter solcher KI-Modelle, insbesondere im Bereich Geldwäsche. Diese müssen sicherstellen, dass ihre Maschinen keine falschen Entscheidungen treffen, da sie sonst haftbar gemacht werden könnten. Diese Haftungsrisiken bremsen meiner Meinung nach den Einsatz solcher Technologien. Zusammengefasst hoffe ich, dass wir in Zukunft benutzerfreundlichere und anpassbare KI-Modelle haben, die Compliance Officers direkt steuern können. Zudem hoffe ich auf eine grössere Akzeptanz und weniger Risikoaversion, um die Innovation in diesem Bereich voranzutreiben. #33:24#

Interviewer Gibt es sonst noch etwas, das du zu diesem Thema sagen möchtest oder habe ich etwas vergessen zu fragen, was in diesem Kontext wichtig wäre zu berücksichtigen? #36:13#

Interviewee Ich glaube, der ökonomische Standpunkt ist auch sehr wichtig. Warum sind solche Modelle interessant? Wir haben das am Anfang kurz angesprochen. Ich erwarte mir vom Einsatz dieser Modelle, dass unser Business skalieren kann, ohne dass die Compliance-Abteilung im gleichen Masse wachsen muss. Das ist einer der Haupttreiber, der mich in diese Richtung bewegt. Ich habe gesehen, dass bei einem regelbasierten Ansatz irgendwann viele Mitarbeiter hauptsächlich damit beschäftigt sind, Alarme zu überprüfen, ohne wirklich einen Beitrag zu leisten. Das geschieht oft nur, um den regulatorischen Anforderungen gerecht zu werden. Aus meiner Sicht macht das keinen Sinn. Ich unterstütze diese Regeln grundsätzlich, vor allem im Bereich Terrorismusfinanzierung und dergleichen, aber ich glaube auch, dass man dies mit Augenmass tun muss und nicht unnötigen Aufwand generieren sollte. #36:27#

Interviewer Okay, also die Skalierbarkeit ist noch ein wichtiger Punkt. Das System und das Modell sollten einfach zu skalieren, zu implementieren und zu nutzen sein. #37:52#

Interviewee Und ich erwarte eigentlich, dass wir ein System so trainieren können, dass es bei einer hundertprozentigen Steigerung des Transaktionsvolumens nur eine etwa zehnprozentige Erhöhung der Personalkapazitäten benötigt. #38:00#

Interviewer Also ein weiteres Ziel in Bezug auf die Skalierbarkeit, ein perfekt skaliertes System sollte in der Lage sein, das Volumen zu erhöhen, ohne dass die Personalkapazitäten im gleichen Masse ansteigen müssen. #38:19#

Interviewee Genau. #38:36#