P6

Interviewer: Could you describe your practice process by combining an example of code auditing?

P6: If I were to give an example, it would probably be something like taking a publicly available dataset of audit reports, which is perhaps used more frequently by everyone. This dataset contains some bug reports publicly announced by major audit firms regarding many projects.Before the audit, it was correct, but there is a drawback that we may not be able to see it from here. It has done classification, but we may not be able to see a complete project from it. For example, some submitted projects may not have published the complete project, but it has published problematic code snippets. Take this, this is a report I audited before. For example, if we didn't have ChatGPT or these large models before, when we got something, the first step in the audit process was to evaluate this thing. Whether it was a company or individual audit, now that we have these contracts, I will first give a pricing based on the number of lines of your code, your business model, and the complexity of, for example, mathematics implemented in the business model. Then the process of giving this pricing depends on the auditor's experience. Maybe based on your business model, we charge a certain amount of money. If there is a problem now, these companies rely heavily on or like these experienced auditors. But now, after the emergence of large models, what's the situation? When we interpret the business model of these projects or read the semantics of these codes, we no longer rely so much on the auditor's experience. We just need to Ctrl+C and Ctrl+V copy it into the corresponding ChatGPT, for example, throw it in here, and we can, for example.

Interviewer: It's like you're using it to understand the code you want to audit, right?

P6: Right, from manually assessing the complexity of a project, to for example, now I may need a lot of manual operations, but in the future, for example, we can use some APIs of ChatGPT to write a prompt, and then we can very quickly conduct an assessment for a specific project. Of course, many companies are already doing this now. Then, as long as we have enough datasets, we can provide, for example, how many days we can save and how much money we can charge for this project. Previously, it might have taken two or three days to assess, but now we may only need one or two minutes to get the job done.

Interviewer: Right, the efficiency has improved a lot.

P6: Right, this is during the audit period, and it is relatively fair. It won't be the case that, for example, in a company with 5 auditors, the time given by each auditor may differ by two or three days or three or four days. No, because we can help you, for example.

Interviewer: So it means that some auditors work quickly while others work slowly, right?

P6: There are still issues with the given time gap.

Interviewer: But if we talk about fairness, is it more fair for the company or for each auditor?

P6: It is relatively fair for clients because in many cases, audit fees depend on the difficulty of your project, and what does the project difficulty determine? The length of time we conduct the audit determines the amount of money received. In the early stage, during the audit, for example, when a project party engages an audit firm for an audit, the first step is to conduct an audit assessment, and ChatGPT provides us with some assistance.

Interviewer: Now that the time has been reduced, will the amount of money received also decrease?

P6: No, it's because for an auditing firm, it's not necessary to deliberately extend the cycle of every project. What's more important for an auditing firm is its external reputation in auditing. Among the 10 projects you may have audited, only one project may have a problem, or there may be 0 problems, i.e., 0 bugs. Yes, these are the fundamental competitiveness and market competitiveness of an auditing firm.

Interviewer: So it's equivalent to, for example, a process that previously took 10 days to complete. For instance, if our project charged 10,000, now it can be completed in 5 days but still earn 10,000.

P6: That's not what I meant. It's that our assessment cycle has been shortened. Previously, when my client engaged an auditor, okay, I needed to provide you with an assessment cycle for your project within three or two days. Then, during the assessment process, the auditor still had to continuously communicate with the project party to understand what your project specifically did, what the model was, and review the code, etc. Right, but if there is...

Interviewer: Right, but in terms of the result, the time you spent has still decreased.

P6: It saves a lot of trouble for auditing firms, and also saves a great deal of trouble for both auditing firms and auditors. Moreover, for example, the audit time assessment for projects provided by such large models is relatively fairer. That's one thing. Then, the first thing a project party does when approaching an auditing firm is to evaluate whether we should enter into a cooperation, and this evaluation process is basically carried out by professional auditors.

Then, the second thing is that after officially starting the audit, what is the difference between the process without ChatGPT and the process with ChatGPT? After entering the audit function, many current supply chains may implement their own EVMs. After implementing their own EVMs, what difference will there be? Different EVMs have different languages, but these different languages implement the same business model. So, at this time, there will be a difference.

Right, another characteristic is that perhaps this supply chain has been popular or emerged for only one year or half a year. There's a question: do auditors have enough energy to quickly and proficiently master this language? Previously, for auditors, if, for example, a certain supply chain emerged, and there were several mainstream languages used in this supply chain, such as GS, Move, and Rust, these three languages might be involved. Before ChatGPT, for auditors, if they wanted to audit a project, they first had to be relatively familiar with this language. They might not use it for large-scale tool development, but they definitely had to be more familiar with it if they were to conduct an audit. They needed to know where problems might occur at the language level, and only then could they audit for vulnerabilities at the business level while ensuring that there were no issues at the language level.

Interviewer: Let me interrupt. So now, with the advent of ChatGPT, it's equivalent to spending less time on the language level.

P6: There are fewer because for many language learning scenarios, for example, when I passed JavaScript, it may be more of a scripting language, and its one-week learning cost is relatively low. However, for some languages, if you want to learn them to a level where you can truly conduct audits, you need two to three months to study. Right, but actually most vulnerabilities are logical vulnerabilities, which are independent of the language. There's nothing wrong with the language itself, and the code written in it is also fine, but when different logics are assembled, vulnerabilities will appear.

Yes, so actually when we want to solve this kind of problem, the first difficulty is the language issue. For example, some languages are relatively more difficult. Take Rust, a language that gives many people headaches, especially in the supply chain. Many people, including auditors in their work, actually use this language. Most of the projects that private auditors in auditing firms receive are on Ethereum, but the contracts on Ethereum are basically not in the form of Rust. So there are many supply chains that are relatively niche compared to Ethereum, which use Rust to write their smart contracts. However, an auditing firm may receive only two or three such orders out of 10 or 20 orders.

Interviewer: So when the auditor audits your project, sometimes they actually haven't studied this language either, right?

P6: Right. Another question is what? Actually, our company rarely uses this language; it's just that occasionally a project pops up that requires using this language. So, if I were to learn this language, would it take up too much energy? Right. So, the requirement from the auditor is that I won't spend too much time learning this language, but I need a tool to help me quickly understand the logic in smart contracts when I don't know this language or am not very proficient in its grammar. Because I just want to see if there are any such vulnerabilities I've seen in its logic. Right, this is the benefit that ChatGPT brings. For example, we may have corresponding ones here. Let me look and see if we can search by language. Sometimes we want to find this vulnerability, a bug at the non-language level, that is, at the logical level, and we also want to quickly understand this code. For example, the same piece of code can be written in Study, Python, or Rast. Maybe this code is written in three different languages, and they are language-independent, but they will have the same type of vulnerability. However, if I don't understand Rast, it may take four or five lines to write in Python, but it may take five or six lines, seven or eight lines, or even ten lines in Rast. It will be very difficult and challenging for you to analyze.

Yes, another point I'd like to make is that so many languages bring challenges to auditors, and as I just mentioned, many supply chains only stay popular for half a year or a year. During this period, they may develop many intelligent science projects based on the supply chain, but after half a year or a year, the popularity of the supply chain drops sharply, and basically no developers are working on it anymore.

Interviewer: So it's equivalent to reducing your input, cutting down on unnecessary input, right?

P6: Right, it has reduced a great deal. In fact, it has reduced a great deal of unnecessary investment in this area, because for human moderators, learning a lot of language related to the supply chain is actually a very headache-inducing thing.

Interviewer: I see. So generally, when a company assigns a project to your company, does your company then assign this project to, say, three or four auditors to conduct human moderation simultaneously, or does it break the project down and have each person moderate a portion?

P6: From the companies I've seen and heard of, they should all have three or four people conducting human moderation on a single project together, because the vulnerability may not only exist in a certain part of this project; it may be caused by the logical relationship between these two modules.

Interviewer: So everyone has to look at the same code, right? It's like everyone points out some bugs, no, some vulnerabilities, and then sits down together to discuss which vulnerabilities should be eliminated.

P6: That's right, and this is the vulnerability under review. When dealing with different languages, ChatGPT has provided us with a great deal of assistance, allowing us to avoid the need to master, or proficiently master, different languages, thus saving us the cost of learning these languages. Right.

The second point is that, first, after evaluating the project, second, look at the language required for this project, and third, when we actually review and enter this project, as an auditor, you may need to read the code. For example, there are two methods. Before ChatGPT, what did we do first when we got a project? We first read its documentation to see what this thing is for. Right, after you master the documentation, then compare whether there are any differences between the code implementation and the documentation, and then compare whether there are any differences between the implementation of the comments and the documentation. There will be many comments on the code. Are there any differences between the implementation of these comments and the actual implementation? Right, then you will look, but if ChatGPT is available, it is actually very simple. We have a lot of things. Only after you read this can you know what this code is roughly doing, and then you have a general basic framework. For example, we check whether there is, of course, some companies will draw this diagram, such as drawing an architecture diagram similar to this, while some companies won't.

When you read code and comments, you may first need to have such a mental map in your mind, and then delve into it step by step. This is the process of auditing in the companies I've been involved with or in my personal auditing. This was the practice before ChatGPT, and then you go in and read it thoroughly. For example, when we look inside, say from the very beginning entry point, okay, what kind of issues might this contract have? This is the experience of reviewers. In your mind, okay, what kind of responsibility does it assume in this project, and what kind of vulnerabilities might this type of file have given the logic it implements? You go to look at the corresponding contracts and time limits with these questions in mind. For example, if we randomly click into a pre-sale, say in this pre-sale contract, okay, could there be overselling due to allowlist restrictions? Okay, we go through it line by line. That's roughly the idea. ChatGPT can help us understand the semantics of this stuff. For example, auditors audit in this way, and after having ChatGPT, another thing it can do for us is that, of course, many auditing companies may have their own large models, some trained by themselves, and of course, some may only do prompt engineering. For example, they would give it such prompt words in the front, and I'll show you what its output might look like.

He can, for example, help us conduct detailed analysis, even identifying what you have imported, which contracts you have imported, what potential issues might occur with the imported contracts, including the logical relationships within the contracts, such as what variables I have defined, in which functions I have manipulated these variables, and so on. And it can help us analyze what a function does. Currently, we might be inputting prompts ourselves to analyze the contract, but if the company is working on this, it will have its own set of prompts, and we won't need to input prompts ourselves. It will analyze the contract based on different business models, such as the business model we just mentioned, where the contract is for pre-sale or airdrop, checking if there is an allowlist, and then analyze the contract using the company's internal prompts, and output corresponding content to help us understand the contract. This way, we can read the contract faster.

Interviewer: Right, if you think from a more high-level perspective, you first understand it, and after understanding, then you go on to look for the loopholes. So when ChatGPT helps you find different loopholes, what role does it play?

P6: Helps me understand the logic of the contract more quickly, because of what? For example, when I'm looking at this contract, it's because I've previously reviewed relevant contracts or seen similar business logic, and I know what kind of vulnerabilities it might have. So, with these ideas in mind, when I look at the contract again, I'll be much clearer. However, the role of auditing is that, in many cases, it's these vulnerabilities or this business model that no one has written before, and logical vulnerabilities that have never occurred online before will also occur in this business model that no one has written before. But at this time, because no one has written this business model, the auditor has never seen it either. This is when ChatGPT is very useful. When we throw this code in, it will tell us what this thing is doing. When we read it again, it's like ChatGPT can help us draw a picture like this, except it's in text form, and it will be very convenient for us to read.

Interviewer: Understood. After finding this vulnerability, how do you use ChatGPT to help you, for example, write reports or provide repair suggestions? How do you use it at this stage?

P6: Actually, regarding ChatGPT, I'd like to mention one more point. For example, when we provide it with some good prompt words along with the contract we need to review, it will point out potential issues, such as token price calculation and emergency withdrawal.

Interviewer: It's equivalent to giving you a list, a list of its possible vulnerabilities.

P6: Right, so we should look into the corresponding parts of the contract to see if such issues will occur. Right, this provides us with a significant amount of meaningful guidance, if your prompt is written well enough.

Interviewer: Do you have any prompts that you've written well and use frequently? I'd like to know your thought process when coming up with prompts.

P6: Similar to this, the first one tells it what it is for. The second one tells it what I'm feeding it. The third one, for example, when we want it to analyze vulnerabilities, you should narrow down the scope of vulnerabilities as much as possible. For example, with short contracts, based on our manual experience, there may be some issues like those in the reports we manually generated. For example, there may be a lack of effective purchase verification when you make a purchase. For example, the allowlist, which is supposed to allow each person to buy only once, but here it allows unlimited purchases, which means the allowlist is ineffective. For example, whether there will be DDOS and so on. We will tell it in advance that the following vulnerabilities may exist, and then ask it to help me confirm whether they do.

If not, please provide your reasons; if so, also provide your reasons. Right, the first thing is to tell it what it is, the second is to tell it what we want to do, and the third is to tell it what vulnerabilities might occur in the stage code if it's a vulnerability. Then, for example, give it some additional hints, and don't miss any details, etc. Based on its output, we will then go to the actual contract to confirm whether there is such a problem. This is the benefit that ChatGPT brings to us. Compared to our previous approach of directly looking at the code, our process has been streamlined a lot. Secondly, we used to spend a lot of time looking at the logic in useless comments and functions. After having ChatGPT, we will have a better sense of direction and save a lot of time.

Interviewer: So, in the process of having a conversation with ChatGPT, how would you improve your prompt to achieve better performance? For example, if the initial prompt is not good and the results it gives are also not good, how would you adjust it to achieve the results you want?

P6: In fact, it has relatively little to do with auditing, and may be more about how the large model is designed behind the scenes.

Interviewer: I'd like to know what your practice is as a user, and the environment you were in when you were a ChatGPT customer.

P6: For example, the information provided here might be quite extensive, and there could be many vulnerabilities. Then, there are things that might not be provided, such as the simplified experience feature, which is only given in a very cursory manner. For example, it doesn't indicate in which function or which lines the issue might occur. We personally think that after reviewing these potential vulnerabilities, there are still issues with the allowlist, but we also feel that the responses provided are rather cursory or scarce.

Ok, if you continue to talk to it, you can tell it, for example, in which function allowlist control might occur and which state variable causes it. However, this requires more expert knowledge on your part. You need to be a relatively experienced auditor before using ChatGPT for auditing, and ChatGPT helps you increase efficiency. But you must guide it through such prompts, rather than being a developer who has never audited or someone who has never been exposed to this, and then trying to have a more refined conversation with it, which might be more difficult. Of course, this is just my personal opinion.

Interviewer: Right, indeed. Then the next question is what challenges do you think you've encountered in the process of using ChatGPT to help you find vulnerabilities?

P6: The challenge is that sometimes the logical analysis it provides, for example, during the process of analyzing logic, does not achieve the effect I want. Sometimes it is too coarse, and sometimes it is too detailed. By "coarse", for example, for a very simple piece of code where we assume we already know there is no problem, but after we throw some specific or what we consider good prompt words at it, it analyzes a lot of false positives. For example, it gives so many suggestions, right? Then after further communication with it, it identifies in which functions these problems are likely to occur. But when we actually look, we find there is absolutely no problem, so it gives a large number of false positives.

Interviewer: What's your attitude towards these false alarms? Do you feel that it has wasted your time?

P6: Yeah, of course, what I'm demonstrating now might be too reliant on ChatGPT, but if it's a personal review, I might only have it do the first step. When it comes to a detailed review, I'll do it myself. Right. I won't completely guide it to do everything and then have it give suggestions while I verify whether some of its suggestions are correct. I'll only have it do part of the work. Right, because this is a time compromise. In fact, you'll also spend a lot of time guiding it.

Interviewer: Spend a lot of time thinking about the prompt.

P6: This is actually okay. The most important thing is that after you write this prompt and it gives you suggestions, verifying those suggestions will take you a relatively long time.

Interviewer: We plan to introduce a tool to help you use ChatGPT for code auditing. What do you think are the most desired features? What are some features you most want?

P6: Actually, the ability to read code like this is no longer lacking for me. For example, when I get a contract, the first step I take here is not lacking. For instance, take the online editing tool for Solidity as an example; it should support ChatGPT. That is, when you input a piece of code, it will read this code and then tell you what this code you've written is about. The official has already been working on this; just look for it, and it should be available. Here, you can see that ChatGPT already has this feature, and I remember Etherscan should also have it. For example, when we get a contract and don't know what it does, but we don't want to read it line by line.

Interviewer: So the way they use ChatGPT with these tools is just by adjusting its API, right?

P6: I'm not sure if they used specific prompts to do something behind the scenes, but these two are mainly the two websites and power builders that Ethereum developers must be relatively proficient in. They have introduced ChatGPT very well. For example, it can provide various information analyses of our transactions, such as where the points it gives are located. Actually, it doesn't provide all the details here. In fact, it should be like later, for example, for some public key transactions, it will tell you that this is a public key transaction, and it will provide information to that extent. Right, like when it reads code, it should also have the function of reading code. I forgot. And like the website we just looked at for bugs, it should also have ChatGPT. I remember it seems to have introduced ChatGPT too, for these vulnerability websites.

If it's something I hope to do, for example, when I get a project, I would like to know what its architecture or module design is like. If you can output this, is it this diagram? Then, if it's ChatGPT, you should make it as rich as possible. For example, when you input an architecture diagram, you can do it based on its parts or UML, which is very simple. But what I really want to know is what this module does. And which functions are called between modules, so that developers or auditors can quickly understand what this protocol does, which external protocols it interacts with, and what its internal interactions are like. Right.

Interviewer: Understood. When you are confirming the results of vulnerability submissions, how do you go about it?

P6: Pure manual labor, I don't plan to introduce any of that anyway while I'm here.

Interviewer: Is there anyone who can use ChatGPT for vulnerability confirmation?

P6: No, we can only say that individuals may do so when reviewing vulnerabilities, but the company will not use them at the corporate level.

Interviewer: Personally, what methods would you use to confirm underreporting if you were to use ChatGPT?

P6: It's actually just the process we just talked about. What does this code do? Right? If so, I'd basically only do the first step, because there are actually too many false positives in the subsequent steps, and the effort you spend is not worth it compared to just looking at it yourself.

Interviewer: That is to say, what you actually get from ChatGPT is information, while the thinking is still done by you.

P6: Right, give me an idea or at most tell me at what level it is, so that you can give me a direction to think about.

Interviewer: At most, give a direction for thinking, but have you ever tried to let it think for you?

P6: For example, when doing this step, this step may help us think, but actually in this situation, I feel that its output is not good. Maybe it's because my prompt is not well-written, that is, its output is not good. For a relatively experienced auditor, actually all these outputs are not good, or rather, it's a waste of my time to check and verify them.

Interviewer: In fact, sometimes ChatGPT is exactly what experienced auditors need more. What ChatGPT outputs are some pieces of information that are difficult for humans to discover or obtain in a short period of time.

P6: Right, if we're talking about the process of confirmation, we mostly, for example, manually classify it as whether it can be exploited or what the cost of exploitation is. In fact, we're mostly determining a vulnerability level.

Interviewer: Apart from that, what other strategies are there? For example, if I give you a vulnerability or a piece of code, through what means can you quickly determine that there is something wrong with what I said?

P6: If we can only rely on the current situation, we can only rely on experience, and here experience means you must have sufficient familiarity with this vulnerability.

Interviewer: Can you give an example?

P6: For example, a certain piece of code you posted, say a certain piece of code has a reentrancy issue. But if, for example, there is a possibility that the manual assurance you provided is considered to have a reentrancy issue, but when I am reviewing and confirming whether this vulnerability is a reentrancy vulnerability, based on my experience, for example, ok, whether an external call has occurred, if an external call has occurred, which states have changed, whether these states are important, and whether they involve loss in capital? For example, I will distribute and verify step by step according to my experience, verifying step by step in my thought process.

Interviewer: That is to say, for example, in the specific case of a duplicate vulnerability, our experience suggests that if a duplicate vulnerability is to occur, it must violate certain Python patterns like "check if". So I think there is a duplicate vulnerability here, but in fact, this code does not show that it violates such a pattern, so you may think that my reported vulnerability is a false positive.

P6: Right.

Interviewer: So, in fact, what you have here should also be a so-called expert experience and knowledge. If I were to say it like I just did, the repetitive parts can actually be summarized into one paragraph. For example, it must violate certain things. What other such things can you come up with that can be expressed in words?

P6: You can provide some keywords, such as in what area, or something like that.

Interviewer: For example, the over-issuance in short selling.

P6: Similarly, in fact, when auditing this type of situation, I can only say that it is based on experience. For example, if there is short over-issuance, okay, I'll take a look at your specific logic. What is the logic behind the short position? What kind of short issuance logic did you use? Did you use Merkel trees or some specific flag bits, such as Vector's flag bits, to verify, or did you use some mapping to verify whether this region has received anything, etc.? I can only make a judgment based on my specific accumulated experience. If my experience is lacking, for example, if I have never seen this type of short issuance logic, it is personally very difficult for me to make a judgment. If such a situation occurs, our internal practice is to first put this vulnerability on hold, have everyone discuss it, and confirm whether it exists. If no one has ever seen it, we will hand it over to the project party for handling, and our audit team will not be involved.

Interviewer: Ok, I understand. One last question. During the past audit process, have you summarized any information in this regard?

P6: Yes, we have a very complete set of data regarding different business types, potential vulnerabilities that may occur within the same business type, their ratings, and their impacts. This set of data is very comprehensive.

Interviewer: What I'm talking about is the manual verification methods for different types of vulnerabilities.

P6: Yes, we have a verification standard.

Interviewer: The verification criteria are okay. So, actually, compared to the vulnerability discovery process where we might have the concept of a checklist, there is also something, a checklist, in the vulnerability verification process.

P6: But it cannot be done in the short term and highly depends on the auditor's experience. You need to create a check list that is either universally applicable across the entire company or relatively applicable to a specific type of business. It may require several [iterations]. For example, we initially proposed the first version, but when we actually used the check list, we found it was too detailed. It was so detailed that for some vulnerabilities, we could originally make judgments based on manual experience, but if you require us to conduct inspections and make judgments through the check list every time, it becomes too cumbersome and actually increases the complexity of the work. Therefore, the level of detail needs to be controlled by relatively experienced auditors, as not all vulnerabilities are suitable for a check list. Right. So this data needs to be continuously iterated and requires a lot of time and effort to develop. The final conclusion is that I didn't have it at the time and didn't make a backup.

Actually, there's another point, which is why we need to summarize the checklist. It was a consideration within the company at the time because different auditors may give different suggestions. For example, when we're checking a vulnerability, I might think it's high-risk while you think it's medium-risk. In the end, whose suggestion should we follow? There will be such conflicts. So we need a relatively standardized checklist. If there's a relative dispute, we can use the standardized checklist to resolve this issue.

Interviewer: Got it.

P6: Yes, when we were doing this thing, everyone in the entire company was involved.

Interviewer: Is it mandatory, or?

P6: For mandatory tasks, the whole company is involved, and there is 2nd-round Moderation after completion, which is a very troublesome thing.

Interviewer: May I ask how many people are there in your company approximately? Is it a very large company, so it is more standardized, or what?

P6: No, it's about six or seven, if it's an auditor.

Interviewer: So what determines whether a company has these things?

P6: Business. Could you share your screen again? For example, this type of company must have such a thing, with many auditing platforms. That is, compared to some auditing firms, it will organize competitions for the projects received from its clients. Right, for example, auditing competition platforms like Code4rena and Sherlock. It doesn't mean that they hand over the projects to their own auditors for review. Instead, they will organize these projects in the form of competitions. OK, everyone can participate, and there are no barriers. You can participate, submit bugs, and based on the quality of the bugs you submit, it will determine how much money you can earn during this auditing period. So they have an important business. After everyone submits bugs, they will make judgments and evaluations. This type of company, including this one, is all the same.

Interviewer: So they have a checklist to verify the submitted reports. That's okay, it's pretty much all of it. Right, there should be nothing else to add. Let's stop here today, and I'll end the meeting.