

## GenexAI Call - 16th May 2025

Transcribed by [TurboScribe.ai](#). [Go Unlimited](#) to remove this message.

So, I think you are also used to manage multiple time zones. Yeah, we do. We work with USA, we work with UK.

Hi Shahana. We do have like our colleagues from New Zealand, totally different. So, it's a different time.

I was telling Farid that I once had a call with Japan and the call was 4am my time, 4am. I woke up at 3.30, at 3.45 I remember my name. At 3.55 the call was cancelled because of a tsunami in Japan.

The problem is that I hadn't drank like one litre of coffee. I could not go back to bed. So, I had a miserable day.

But, you know, those are the things that happen. That's part of our lives normally. Okay, is Jessica going to join today? Yes, I think she should be joining very much soon.

And I'm happy that my Japanese story is also captured by the AI note taker. So, we have a lot of ground to cover today. So, really looking forward for this conversation.

Thank you all for your time. Let me just check if Siska is joining. Fabio might not join, but Siska, let me just check her.

You are the organiser, right, Faisal? So, you should see Siska. Let me just confirm with her. She is in court.

She's joining now. Okay. Hi, Siska.

Hi, Ernesto. Hi, everyone. How are you? I'm good, thank you.

And you? Excellent. Faisal, from your end, are we waiting for anybody else? We have you, Shahana, Mesup, Hamsa. Anyone else that we are waiting for? No, we are good to go.

Farid might not be joining today. Tayyaba has some issues. So, she is not joining today.

Okay. Very good. So, thank you very much for your time.

We are really looking forward to start working with you. We spoke with Farid earlier today, of course, to get a little bit more about the background of GeneXus AI. And what we had agreed is that we're going to have the opportunity to work with you at least during a one-month sprint to see how far we can go from there.

The main goal, definitely, at least from our side, is to advance as much as possible on the user

interface. We know this is a project. When I say project, I mean the platform of GeneXus AI that normally would take at least six months, plus another six months, total one year, to be in a fine-tuned version.

So, we don't expect any miracles in this one month, but we expect to learn from you a lot, to interact with you a lot, and to have some interface. And I'm going to show you some examples so that we can share to our clients, which will also help us get more funding and then continue working with you. So, that's basically how we see this next one month.

And it would be great to schedule some weekly calls. This time, I know it's 5 p.m. for you. We can move a little bit earlier if you prefer.

But it would be nice to have some weekly catch-ups, even if it's 30 minutes, on Fridays, whatever you prefer, to ensure that you're moving forward. But making a long story short, let me stop here and ask you, Faisal, and everyone else, in terms of our overall goal and expectations, if we are on the same page. Yeah, I do have a chat with Fareed, and he let me know about what your plans are in terms of development for the product, face-by-face.

And for the very first month, you want just to have a UI for the one system you had, which you are interacting with the JSONs or something, if I'm not wrong. Yes. Then you want some interactive things to work with it.

And then on the first phase, we should have to build a vector database first with the data. Correct. Which I will be asking questions about how you are going to provide us and what are the different channels we should look for these data.

What are the types of data, abstract, reviews, and other things, how these data are linked together. What type of queries you guys will be asking from this data, expecting to work. The answers will be these sort of things we will need from you guys to build a first phase, which Fareed mentioned, to have a vector database and interact, to have a chat with that vector database.

As you guys mentioned, if I'm not wrong. And in the second phase, then we will proceed toward the SaaS application. Like, first of all, we will get a UI interface.

We will decide the user management stories. What will be the different roles? What will be the different types of users? What users can do what on the basis of their subscriptions, plans, etc. And what features will be in this SaaS application other than the shared system? Like how the data will move from one user to different users.

So these features are defined. If I have understood it correctly. 100% correct, Faisal.

That's exactly the point. And I'm going to go a little bit more into this, but it's exactly as you've described. I'm going to share my screen a few moments because we have some mockups of what we visualize.

By all means, you are the experts. So forgive us for our lack of expertise that you have. We just wanted to put some visual things because I think it's better than explaining.

So feel free to interrupt any time. So I'm going to share my screen. And the way that I wanted to organize the call was to spend this initial 15, 20 minutes or so describing the process that we visualize.

And we're going to focus on what the basic process, because there are many other things we can build later on. But let's get the, as we say in Brazil, the rice and bean first. Then later we put on the meat.

And I love rice and bean, by the way. So we basically put on that and then we move forward the next. Then after that, let's discuss about the types of inputs that we have.

I will anticipate everything is structured data and the outputs that we're going to have as well. And then we can maybe leave the last part of the call, the final five, ten minutes to discuss next steps, cadence of updates, things like this, if that's okay for you. Yes.

Fantastic. I will share my screen. Let me know if you can see my screen now.

Yeah, we can. Okay. So I will share all this information with you.

I will send the email to you, Faisal, and everyone else that is copied in this meeting invitation. So the starting point will be our website. This is our website.

So you can get a look and feel about the fonts, the colors, et cetera. So it's mainly blue, white, and green. And the idea would be to the starting point of the platform would be here, in which it's coming soon, of course.

So that would be the starting point. Once the user logs into this, then, of course, log into the platform. As you said, Faisal, and by the way, everyone, feel free to interrupt me anytime.

I don't see your cameras. I can hear to listen to you. So if you have any questions, feel free to interrupt.

Very good. Let me just put the clock here so I can keep track of time. Okay.

So we log in the platform. Forget about this select product. This goes later.

But we're going to do the literature review. Literature review means we're going to evaluate either abstracts and or PDFs, which are from clinical articles. Both examples are structured data.

So let's say we're going to look into databases. And we say, yes, we're going to look into databases. So the way we go is, and I will share some screens for you in a few seconds, is basically to create a user interface that they can enter the search terms, extract information

from the database, and ask questions to this information to generate a report.

Now, the databases, the most famous database is PubMed, as you can see here. We already have the API to PubMed. It's a free API.

And this is exactly what we wanted to replicate. So once again, let's say I'm going to search for studies talking about the cardiac and implant and I don't know, mitral valve, for example. And those are the fields that we can take.

And, of course, we can select the publication date for a certain period. So, for example, since the beginning of 2024. And we should be able to replicate all of that into the GenXCI user interface.

We already have the API. So the same results that we got here, 925 results, we already have the codes to receive that. This is an abstract.

So you have the journal that it was published, the title, of course, the authors, and the summary of the abstract. And here's the text that the user will ask questions to, to ensure that this goes into the future in criteria. So, for example, I'm doing a research and I wanted to evaluate mitral valve implant in cardiac.

Very good. But then I wanted to look into only into individuals with a high risk. So, for example, I can say I can ask, is this talking about high risk? And then, of course, in this case is yes.

So this would be the screening process that we have here that we enter the search terms. We run the search on the databases. We extract the abstracts and identify duplicates.

This is something we already have the code for that. And then if there are more search terms, the user can go back and do more research. It will go into the same knowledge base.

Give me one moment. I will show to you visually how we envision that. And then the user will ask questions for eligibility.

And then, of course, is this talking about X, is talking about Y. And then at the end, we have the end of screening process. Those are the abstracts that are that you searched, plus and they are talking about the topic that you discussed. Before I show you the screens that we visualize for this process, let me stop here.

Any questions, any comments so far? Yeah, I have a question. Like you said, we already have like you have some frameworks API or you have just the code to hit that API request or something. Or you have you have some back end written on it to get these data already from the APIs of PubMed.

Yes, I have the full back end. I have the full code to connect to the PubMed, extract the abstracts, as you can see here in JSON format and save those abstracts in my computer. So I have all of that.

OK, so you saved it on your computer, not into any database. Exactly. OK.

And that's one of the things we need to change. OK, you need to change this to save this to your DB also. Exactly.

OK. Very good. Let's proceed to the next step.

I'm going to basically show to you how we envision this. Just to confirm, can you see my screen? Yes, I can. OK, I know it's the resolution is not so good because this is an example that we took a screenshot from another vendor.

But it will give you the idea. So once again, let's imagine you are on the GeneXus website. You click on here.

And this would be the initial screen. So you log in with your details. We can discuss this later.

And then you get into the platform. Yeah. In the platform, you can enter different knowledge bases.

Think of this as folders. So, for example, Cisco and I, we are working in the same company, but we may be doing a research on cardiac technologies. And we also are working on another completely different independent research on diabetes, for example.

Those would be two separate knowledge bases. One has nothing to do with the other. So I will select my knowledge base.

In this case, I will enter a name saying cardiac implant, whatever. And I will start populating this knowledge base with the abstracts that I will extract from different databases. But at this moment, let's start with PubMed.

OK. This is basically a screen that we envision. So you see it's very similar to the screen that we have with PubMed.

OK. And those are the databases that we wanted to select. The same search terms are accepted widely across different databases.

But at this moment, let's focus on PubMed. But ideally, in the future, that will be Cochrane, Lelux, Cielo, different databases here. But basically, we enter the search terms.

And then we enter that. And, of course, that is the time to extract that because the API that we use extracts a database in batches. So it takes a few seconds.

It's not instantaneously like we saw in PubMed, but it's a few seconds. And then we can see this screen. And then once the search is completed, of course, it informs the search is completed.

However, something that I will emphasize to you, for any single research that is done, there are multiple search terms. And these search terms may lead to the same results. Our API already

detects the duplicates.

So we already separate what is a duplicate, what is not a duplicate. So let's say we can do like three search terms. Each will generate 100 abstracts.

But at the end, we're going to have 150 unique abstracts. But this is something we already have. So basically, that's the extraction of the search term.

Excuse me, the abstracts. The extraction of the abstracts. They are stored, ideally, in a vector database.

Now, if they are stored in my computer, there's no vector database here. But they are stored in the vector database. And it's there.

So now let's start working with these abstracts. This is the screen that we envision for the questions. So there are some standard questions we should always consider.

Like, excuse me, who are the authors? What is the title of the study? Et cetera. But then we also have types of questions which may be inclusion or exclusion. So there are basically three types of questions.

Questions that you are just asking to the abstract. Who is the author? Where this was published? And you just wanted to retrieve the information. So you ask, and information comes.

That's it. But then you have what is called eligibility questions. And the eligibility could be an inclusion or exclusion.

Eligibility means this information will be used to filter the study. So let me give a better example here. So, for example, let's say question number seven.

We say we have an eligibility criteria. So we check, yes, it's eligibility criteria. If it is eligibility criteria, it's inclusion or exclusion.

It is inclusion. If it is inclusion, the system already understands in order for this abstract to be eligible, it must mention et cetera. So, for example, in my previous example, it must mention high risk or must talk about high risk patients.

And that's where we are using LLMs to read the abstract, understand the question, and understand if that is mentioned in the abstract or not. The eligibility criteria also can be exclusion. So this abstract must not mention whatever or must not include.

So, for example, I want only abstracts that are talking about children. So they should not mention adults, adult population, for example. And then looking into this, and then we have question number nine.

We say, no, it's not eligible. And then we can say it must mention something. But then this is

just for informational purposes.

So at the end, what we want is a system that we understand based on this is eligibility criteria, inclusion, exclusion. In this case, for example, question seven and eight above would be used to filter the studies. And then we would have some result like this, for example.

Like, oh, out of those 150 studies, 70 abstracts match your criteria. And those are the abstracts that we will incorporate. Once again, let me stop here and see if you have any questions or comments.

OK, till now, what you have done is like in terms of searching through the LLMs, already you have what you showed us with the JSON one. And one thing when you want to, like you said, we have already implemented the duplicate logic that it should not show the duplicate reviews and abstracts in the results. So will you be also tracking the delta? Like when you search some theories and it's already asked what you want to like take it from the database or you want to like compare the data from it and then show the data of it.

Like if you understand my question. Yes. No, no.

Yeah, I understand perfectly your question. And this goes back to something that we can discuss later, which is the Prisma diagram. So let me show an example here.

So, for example, let's say. It's a little bit too small. Yes.

Let me make it bigger. Is it better now? Yeah. Yeah.

OK, so this is at this moment. Our codes, they only detect the duplicates. But basically, it's very simple.

We just needed to count the number of the results and then calculate the delta. So, for example, here you see a research that was done on different databases, MEDLINE and BASE, etc. And they are using the same search terms.

So in MEDLINE, they acquire 215, etc., etc. And then they include another four. Probably they include them manually on that.

So look into here. Records after duplicates removed. So 305.

We have all these already in the code. So we just need something to take this plus that minus that. And then we say how many duplicates were removed.

And this means, OK, all these duplicates, all of those will be screened. And remember the questions that I had mentioned to you. This is what is called eligibility.

So, OK, so we excluded 144. So we have 161 that match those eligibility criteria. So what I've discussed is this part here until here.

And this is something that we needed to generate. It doesn't need to be now. This is called a Prisma diagram.

And this is an official visualization for medical research, which is what we're doing now. But I just wanted to show to you visually what is the logic behind that. And leading to this, I have another question.

If you ask the same theory, like the same thing, and now you want it to be into a database, like currently you don't have a database, you are just fetching it on your local. And how you want it to be like fetch it from your local database or your own database, or it should always hit these platforms to search this theory to get the result? That's a very sophisticated question, Faisal. And I will ask your help to answer this.

So when we are now talking about abstracts, let me go back to an abstract here. Just to give an example. OK, so now we're talking about abstracts.

Let's say we extract the abstracts. Now we're going to ask questions. When we extract the abstract, not only we extract this, but we also download the metadata from the abstracts.

The metadata from the abstract is gold, has a lot of information. So when we do our codes, and I will show to you an example of that. The first line of a question is for the metadata, for in case of abstracts.

So, for example, can you see the gold here? Yes. So we first ask the metadata. If we don't find the data in the metadata, in other we're always going to find, but if we don't find then, and of course the metadata, the tag is also released, then we use OpenAI, in this case GPT-4.0, to read the abstract.

What I'm envisioning is, but this is only for abstract, metadata first, then use the vector database results, but I've never tried that, but it seems like it could be worthwhile, and then if it doesn't respond, then we could use an NLM to read the abstract and return the answer to us. You mean you have this information in metadata, already have, because once you had, you did the same thing, and if it's the metadata didn't find in your database, then it will go to OpenAI to read that abstract. Otherwise, it will get the same abstract from your database.

It will not get the NLM to get again the abstract. Okay. Exactly.

This is also happening in your API currently. Yes. Now, basic questions such as author, title, journal publication, this will always be in the metadata, always, but when we go to a specific question, so for example, what is the ideal treatment reported to prevent preterm birth? This was a research we did for preterm birth.

This is something we don't even get into metadata. The reason that I'm saying this is because if we go back to the questions, when we have that screen that you enter the questions, this is starting on question number six, because we always going to ask five questions first. Author,



study title, journal, document, object, identifier, which is like an ID code for abstracts, which in this case is here.

Yeah, document ID, right. And publication date. So we're always going to ask that.

So these are the ones that whoever uses GeneXus AI, these will always appear, and this is going to be metadata. These ones, on the other hand, we can ask for metadata, but most likely they will be available for interpreting the text. Okay.

And sorry if I ask a silly question. These abstracts never get updated? No. You can have an abstract that is published in the following year.

That's a great question, by the way, but no. So, for example, let's say this Peter van der Buijl did this analysis, mitoannular disjunction, et cetera, et cetera. He might do the same analysis next year with different results because it has more updated data, but this will not change this year.

This will always remain the same. And we get the publication date in the metadata? Yes, yes. Then that will make it distinct from the one it does last year or the next year.

So the publication date will be the unique one. Yeah, but the DOI will change because it might also publish with different authors, et cetera. But even if you publish the same study with the same authors on the same topic, as long as it's a different publication, it will have a different DOI.

Okay. Very good. So that's basically what we are aiming to do.

I believe this screen here is the one that is the most specific one. So basically five questions that we always ask, who are the authors, et cetera, et cetera. Then starting on question number six, there are two options.

Either a drop-down list because it's very common for people to ask about language, the keywords, number of patients, what is the control intervention. Anyway, that is a drop-down list for standard questions, plus the possibility for the user to enter their own questions, customized questions. So that's basically the screen that we envision.

**This file is longer than 30 minutes.**

**[Go Unlimited](#) at [TurboScribe.ai](#) to transcribe files up to 10 hours long.**