

1. First some context - who you are, and what Verses is.

Hello, I am Sean Lee, the founder and CEO of Verses.

I have had many titles in the music industry for a few decades. I have been a rock guitarist, jazz pianist, composer, professor, and policy advisor for the Korean Government. I have also founded a game studio, mobile TV company, and music streaming service as a serial entrepreneur.

My role in the industry is applying new technology to music to create innovative music products and revenue streams. For example, I was the first person to launch commercial mobile TV in the world, generating over \$100 million in revenue with my old company, Direct Media, which I co-founded as a joint venture with Samsung Electronics, Korea Broadcasting Service, and Korean Telecom. Also, I launched a music streaming service in the early days that is still one of the largest streaming services in Korea.

My journey is continuing in the era of virtual media like the metaverse, VR, and AR. Currently, Verses is presenting a new music product for what music should be in the virtual world. I believe that the music experience should no longer be just listening but it should be visualized, interactive, and communicative. For example, music fans can now interact with their favorite artists in the metaverse like playing music together, dancing, and making music videos to share with friends. Our company believes Music and fan engagement will become more diverse and powerful in virtual media and as much as experiences become more powerful, the value of music products will significantly increase.

Our efforts are receiving much recognition from the industry. We continued to receive the Best Innovation Award and two Innovation Awards at CES from 2022 to 2024, Billboard magazine selected us as the most innovative technology in 2023, and Universal Music Group selected us as one of the four best companies at UMG Bootcamp last year.

---- KT KIM

Hello, My name is Kyungtae Kim, one of the co-founders of Verses.

I work on developing audio, music signal processing, and AI tech in general for the company. I look for meaningful musical interactions to apply to our service, and I also find ways to incorporate new AI or signal processing technologies into our service.

Personally, I always wanted to make music and be successful with it. I felt that I needed something of my own to win a Grammy for example.

So I thought that it would be cool if I could create some instruments or new technology.

Like Yamaha's DX7 or Roland's drum machines created genres, I also wanted to create my own new musical territory.

What I focused on was not just music, but finding new musical value in other synesthetic situations including visuals, movements etc.

So, while doing my master's in audio machine learning, I did some interactive exhibitions and performances, and I met Sean Lee there.

At that time, I wanted to create performances that could immerse more by controlling music and visuals in real-time, so that I could contribute to increasing the value of music performances.

He asked me why not just increase the value of music directly, and I was like, 'Wait a second that makes sense!', So that is how I joined the company.

2. AI-generated music has been one of the years' hot topics. It's clear that one use for generative music is in responsive situations in games and metaverses. Can you explain how Verses works to create music for users in interactive game environments?

Sean Lee

Currently, AI-generated music has become a big topic with huge potential.

We are actively addressing this, so we are developing our original AI music engine with the Google startup program and using this engine in our service. We think AI is a big matter for the future of the music industry. So, we continue to do research and development.

Creating an interactive music environment is really a big deal because this is a new category of music and no one has enough experience to make this happen. So we put all of the cutting-edge technologies such as Generative AI engines, game engines, audio technology, and our exclusive music and video synchronization technology to build a solid tech system. Not just technology, we are trying to find new expressions and new aesthetics for how to create great interactive music moments.

So, how about making something impossible to be possible? Playing music, dancing, and creating music videos with your favorite stars like BTS. I think this is a dream for many music fans, but it's impossible in the real world.

However, we made this possible in virtual worlds. So, we made 'Aespa World,' we launched with the famous K-Pop idol star Aespa. Now music fans can select their favorite members, choose dance routines, pick backgrounds, and create music videos they can proudly share with friends. We made music fans' dreams come true.

Just like this case, I think music will become a more multi-modal, immersive, and comprehensive experience in the future. Historically, music media has continued to expand in terms of the medium itself and its expressive power.

The first musical media was a Music score that only contained visual symbols. And records, the next medium, contained only sound information. Music video contains both music and video information.

Now, virtual media is changing music. Music is expanding to new media form where music fans and artists communicate directly and music fans control that. These new elements and expressive power will bring a lot of fresh fun and excitement to music lovers. Also, this media expansion will create a new revenue stream. We are moving from an era where recorded music made money to an era where the relationship between musicians and fans makes money like what SW and game industry did it.

KT KIM

I think the biggest advantage of using virtual space, is that we can actually implement musicians' imagination into a real experience.

aespa World, we created, would be a good example of this.

Aespa is a K-pop group with four real members and a virtual member, called Naevis.

According to Aespa's concept story, Naevis is a character that assists the real-life Aespa members in communicating with the virtual world.

So we implemented Naevis as an entity that spreads music in the virtual world. So that Fans can experience new music when they acquire music items that she throws.

From this, fans can empathize more deeply with Aespa's music and become immersed in it.

This seems quite simple and straightforward but, based on our experience, we've realized that creating an immersive music world in virtual spaces is quite a challenging task.

Because it brings up a lot of questions.

Like, how should we express music visually. What would be the difference between a melody and beat in a visual format, etc.

So, we did a lot of research, from Disney's 'Fantasia' to experimental films of the 60s and 70s, all of music video games like rhythm games.

We are addressing these questions, and hypotheses one by one, exploring possibility, but there is still a long way to go.

From an engineer's perspective, I feel proud of the way Verses uses gen AI Technologies for creating interactive music experiences. Because we use gen AI technology in a non-disruptive way.

What I mean is, that whenever there's talk about generative technology, people tend to focus on how it might replace humans and destroy existing industries or ways of doing things right?

However, our technology requires the original source from the artist. It requires the artist's music and voice, and the concept of their musical world. And our job is to make it interactive way. and gen AI technology simply extends this world.

And another thing that I'd like to point out is that generative AI technology is becoming multimodal. From a technical standpoint, audio, image, or video, different types of data require different approaches according to the characteristics of each medium. For example, audio data has a concept of time, whereas image data does not.

But recently, many attempts have been made to transform between different media domains. For instance, creating music from an image or creating a text storyline from a piece of music.

We used to focus on monitoring music generation and audio signal processing technologies mainly. but recently, we've been working to integrate various generative technologies into our service, such as text generation, video creation, or creating 3D objects that fit our service. So it's not just music for us actually. We're exploring other areas too.

3. How is this way of experiencing music in games/virtual experiences technically different to how we experience music in games now?

The biggest difference is what is more important and what takes the spotlight. In video games, the focus is on gameplay mainly, listening to music is more of a background. As a result, when game players take actions, the visual responses are more important, sound

effects and music are the following things. Sometimes, there is no music at all, and many game players don't pay attention to music.

However, when a user tries interactive music in the virtual world, music is the most important thing. For example, when you enter K-Pop star Aespa World, everything within the environment dances to the rhythm of the music. Music becomes the leading role, visual objects play supporting roles. Another example is that when you get a music item, the music changes, and the windows of the buildings, and streets, as well as the sky, start sparkling in sync with the new beat. Every movement is controlled by music, and this music is controlled by music fans. This is a very powerful musical experience, and it will give users the satisfaction of having the power to control music. It's like a power that only a virtuoso music master can have.

--- KTKIM

We are cautious to ensure that music does not simply serve as background music. That is our number 1 point to think about. However, making users interested in the changes in music is also quite a challenging task, because most people are not used to focusing only on music.

Even basic audio effects that musicians or [anyone who has ever tried making music would know], such as reverb or delay, can make it difficult for people to distinguish the difference between sounds.

For instance, we had a user test, about our interactive music demo, and it was very simple concept. If a user touches or drags the screen, a simple reverb effect is applied to music as the user controls the screen. And people fail to perceive the difference in sound and simply think the volume has decreased. This happens even for music fans!

So, I believe creating synergy between musical changes and visual changes or interaction is an important point.

So we do a lot of experiments about them, like aligning musical changes with visual changes or make a situation that users have to focus on musical changes only, like that.

It's hard. Harder than we thought.

-- 2min

4. There have been some human-created soundtracks to video games or virtual experiences that have become enormously popular (for instance, even the Wii Shop menu music!) but if music is generated on an individual basis, everyone's experience

is different. So, will we miss out on a communal experience outside of the game if we don't all hear the exact same song?

Hearing the exact same song is a great experience. When we listen to the same music, we share similar emotions and felt connected.

We certainly like to listen familiar music, but at the same time, we also have a desire for new sounds and music. For instance, Electronic music fans love to hear new remixes. Because It's also fun to see familiar music become something different. I think listening to music has two sides of a coin like the comfort of the familiar and the excitement of the new.

I assume interactive music that is generated on an individual basis is not a contrast to traditional music listening but it is an extension of it. It is to provide users the option to continue enjoying familiar music without getting tired of it.

In addition, the younger generation really loves active media experiences. They like to have a power of control. They don't want just watch music videos but engage by dancing, singing, and creating TikTok videos. The trend is shifting to personalized and user-driven experiences, and music should follow this trend.

--- KTKIM

I believe that the answers to such questions should be obtained by introducing the concept of interactive music to the world.

Because the world is still divided into listeners and music creators, or the music experts, those who know what their doing with the music.

If people were told they could change and express music however they wanted,

many would respond with, "But I don't know how to play an instrument?" or "I don't know much about music," even though, in my view, they have firm musical tastes. It's like there are two parallel universes.

If we can create a new middle world between that two seemingly different worlds, and if we can provide fun and musically meaningful experiences intuitive way, **I think users will then decide whether they need concepts like 'original music' or not.**

5. Who owns the pieces of music generated by the Verses system?

We think our system is a tool or instrument. So, artists and music fans who use our system are the owners of the pieces of music generated by our system.

Fundamentally, we believe that rights and earnings should be shared with everyone who is a part of the creative process. That's why we believe that artists who provided the original

copyrights and music fans who remix the music both have rights to the music. Although a clear revenue-sharing model is not fully developed yet, we are in the process of drafting agreements that assign ownership of copyrights based on the level of participation.

--- KTKIM

In that context, we are also focusing on technologies like blockchain because we hope that it can let us manage copyrights for user-generated outputs properly.

6. How does the NFT system work?

--- KTKIM

Fundamentally, the Interactive music system pairs well with NFT. Because it can be one of the outputs generated by users.

The fact that new music is created every time the user plays, it means that a new output is produced each time.

So we basically support NFT systems, and did some research on it, however, we think NFT systems are still in their early stages, so we're kinda waiting for them to mature a bit further. We also met a lot of NFT companies when we were at CES. Perhaps, we might utilize them around next year.

But before that, we should focus on the different output formats too.

Because user-created output can be a video, image, or album cover, just music, or a whole scene we don't know yet.

7. Looking forward to future metaverse spaces and gaming: how much of the music do you think that we experience in games/virtual spaces will be AI-generated for individual users?

AI-generated music for individual users will be popular. and it will be a good solution for someone who needs interactive experiences. Plus, it's cost-effective, so video game and social media companies will like it.

However, big game studios will keep their traditional production process. High music quality and collaboration with famous composers or music producers can be important assets for

promotion campaigns. So, I believe that each will adopt the most reasonable music production method based on their specific needs.

---- KTKIM

I think our service users wouldn't care whether the music is AI-generated or made by humans, unless they are personal fans of the musician who worked on the music.

In my opinion, the strength of gen AI lies in its scalability, and the role of humans is to determine the creative direction. So AI-generated music, stories, or visual sources are good for stably expanding an existing universe. However, it's kinda difficult for them to introduce new concepts.

Even if a generative engine proposes a new concept based on provided input, which is generally text, recognizing it as a new concept and changing direction is ultimately up to the person controlling the engine.

I think, this is because, in the end, a generative engine is essentially about 'Sampling'. The basic principle of the most commonly used generative engine starts from complete noise and gradually improves the results that reflect our intentions in the way we have 'trained'.

It means it requires a certain dataset that I curated.

If engine asked to, for example, create 'innovate music' out of nowhere, the engine would start by looking for data captioned as 'innovate' in provided music piece as data and try to create something.

The engine could never propose other concept like John Cage's 4'33", like 'How about not make at all' like that,

At least for now. For now, it's literally just deep understanding from a large dataset.

So I think current generative engines are programs that recreate the afterimages of our civilization that we have digitized so far.