ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250326033909

Original scientific paper
Received: 2025-03-26, Accepted: 2025-06-09

Creating an AI Human Professor Model to Implement a New Educational Paradigm of the 4th Industrial Revolution

Yangha Chun, Soo-Yeon Yoon*

Abstract: The COVID-19 pandemic accelerated a global shift toward digital and hybrid education, exposing critical limitations in traditional pedagogical approaches. This study introduces a novel AI-based human professor model designed to address challenges in educational scalability, multilingual delivery, and content production efficiency. The system integrates state-of-the-art technologies—deepfake for facial realism, few-shot text-to-speech (TTS) for multilingual audio, and lip-sync engines for expressive synchronization—to autonomously generate lecture content. Quantitative results from pilot testing indicate a 60% reduction in production time, with user satisfaction scoring 4.5 out of 5 and multilingual accessibility rating 4.7. The model demonstrates significant potential to transform digital higher education delivery.

Keywords: AI human; deepfake; digital education; few-shot TTS; lip-sync; multilingual learning; scalable instruction

1 INTRODUCTION

The outbreak of COVID-19 forced rapid transformation of global education systems, shifting instruction from physical classrooms to digital platforms. This unprecedented shift revealed challenges including limited accessibility for multilingual learners, time-consuming content production, and instructional scalability. The change began in the first semester of 2020, when face-to-face classes were not available, and online classes began with real-time video lectures or recorded video lectures using IT, which changed the educational paradigm by starting non-face-to-face online classes not only in university education but also in higher education [1]. Non-face-to-face online classes, which began suddenly, have been developing for two years, and non-face-to-face online classes or hybrid classes mixed with non-face-to-face and face-to-face methods are likely to continue in the future [2]. In addition, despite the time for innovative education at the beginning of the 4th Industrial Revolution, the school education method, which pursued the traditional education method, is expected to realize edutech that combines technology and education in earnest due to COVID-19[3]. Non-face-to-face classes can be taken anytime, anywhere, and students can take classes again if they want, so if used efficiently, it can be a teaching method that produces better results than face-to-face classes. In addition, even if a virus or epidemic other than COVID-19 comes, there is a possibility of conducting non-face-to-face online classes using digital media, which is different from face-to-face classes in that they can receive education regardless of the environment and situation.

As the number of non-face-to-face online classes increases, voices in the education community are increasing to propose the introduction of artificial intelligence for efficient learning effects [4]. Artificial intelligence technology resembles humans, so technology is developing in judgment, classification, and generation. These developments have adopted AI technology in many industries and companies today and apply it to producing new content. Businesses using AI human, which are created based on artificial intelligence technology, are emerging and can be used as instructors for non-face-to-face education.

This study proposes a transformative solution: an AI-based human professor capable of delivering multilingual lectures without requiring continuous human labor. Unlike traditional lecture capture or MOOC systems, the AI professor autonomously generates visually and vocally realistic instructional content. We explore how this innovation addresses pedagogical constraints and contributes to a future-ready educational paradigm.

2 PREVIOUS RESEARCH

2.1 The Concept of AI Human

Digital human technology has emerged as a key frontier in virtual communication, education, and entertainment. Recent applications include AI tutors, intelligent chatbots, and interactive avatars. Prior work has focused on support functions or partial automation of instruction. However, few systems offer end-to-end autonomous content generation at scale.

An AI human is one of the digital human beings, and it can be seen as a computer graphic-generated human that has a similar level of appearance to a real human and can express movements similar to human movements. With the development of AI human, education, guidance, and counseling are now possible. Unlike Virtual human, AI Human is a model learned using deep-fake, lip-sync, voice TTS, and technology by combining synthetic data technology rather than 3D modeling. AI human takes a long time to learn early modeling, but later, it can quickly produce images with learned models and make video production possible at low cost. AI human has the advantage of being able to produce content and expand dynamic conversations as a learning model, but it has the disadvantage of requiring sufficient initial learning data and some limitations in dynamic application.

Deepfake techniques enhance realism by mimicking instructor facial expressions, while few-shot TTS and multilingual NLP support global accessibility. Lip-syncing

bridges the gap between audio and video realism. This research contributes a comprehensive AI professor model integrating these capabilities for full-course delivery.

2.2 AI Human Application Case

2.2.1 Digital Restoration Remembrance Service for the Late People

In the original program "Alive" of the online video service (OTT) Tving, the late leader of Ulala Session, Lim Yoon-taek, who died of stomach cancer, was restored with AI technology to perform a song with the Ulala Session members. This is a video that combines AI technology, and Lim Yoon-taek, who was restored through voice synthesis, finished the "West Sky" stage with a healthy appearance. Lim Yoon-taek's appearance and voice, which appeared similar to his life without a sense of difference, impressed his family, colleagues, and viewers. In addition, restoration using AI technology has emerged as a new way to commemorate, such as restoring famous singers such as the late Shin Hae-chul and the late Kim Hyun-sik of the group Turtle with digital technology and holding hologram concerts.

However, there is a question of whether digital restoration can be carried out without the consent of the deceased, and the conclusion of this is one by one that "it cannot be carried out without the consent of the deceased" and "it is necessary to make a profit in the will of the person who misses it".

2.3 AI Human Technology

AI human combines 2D live-action-based technologies such as face, mouth, voice, motion, and video with AI synthetic data technology to enable various verbal and non-verbal expressions, and combines various technologies to produce AI human. Tab. 1 below shows the technologies applied to human AI.

Table 1 AI Human Technology

Apply Part	Technology
Face	Deep-fake Technology
Lip	Lip-Sync Technology
Voice	Few-Shot TTS Technology
Movement	Silhouette Customizing Technology
	Pose Estimation & Generation
Video	AI human Streaming

2.3.1 Deep-fake Technology

Deep-fake technology is a powerful and rapidly evolving form of artificial intelligence that creates highly realistic yet synthetic videos, images, and audio by manipulating existing media. The term "deepfake" is derived from deep learning, a subset of AI, and "fake," referring to its ability to generate misleading but lifelike content. By utilizing deep learning algorithms, particularly Generative Adversarial Networks (GANs), deepfake technology can convincingly swap faces, mimic voices, or even generate entirely new personas, making it difficult to distinguish real from artificial content [11]

The process of creating a deepfake begins with training an AI model on a large dataset of images, videos, or audio recordings of the target individual. The AI learns intricate details such as facial expressions, speech patterns, and subtle mannerisms. Once trained, the model can seamlessly manipulate existing footage, replacing one person's face with another's or generating realistic synthetic voices [12]. The sophistication of these AI-driven creations has made deepfakes both a fascinating innovation and a source of ethical concern.

While deepfake technology has legitimate applications, such as in entertainment, education, and accessibility, its darker side has sparked global debate. In Hollywood, deepfake AI is used to de-age actors or bring historical figures back to life in films. It also serves as a valuable tool for language dubbing, allowing actors' lip movements to match translated dialogue more naturally. Moreover, for individuals with speech impairments, deepfake voice synthesis can help recreate lost voices. However, the misuse of deepfakes poses serious risks. Malicious actors have used them for spreading misinformation, creating fake news, and impersonating public figures to manipulate public opinion. Fraudulent schemes involving voice cloning have also emerged, where scammers replicate the voices of company executives or loved ones to deceive victims. Furthermore, deepfake technology has been weaponized for cyberbullying and defamation, raising concerns about privacy and consent in the digital age.

To combat the growing threat of deepfake misinformation, researchers and technology companies are developing AI detection tools capable of identifying synthetic media. These detection methods analyze inconsistencies in lighting, unnatural facial expressions, or pixel distortions that may not be visible to the human eye. Additionally, policymakers are working on regulations to curb the harmful use of deepfake technology while preserving its beneficial applications.

As deepfake technology continues to advance, the challenge remains to strike a balance between innovation and ethical responsibility. While it has the potential to revolutionize digital media, its misuse underscores the need for vigilance, awareness, and stronger safeguards against deception in an increasingly AI-driven world. Deep-fake is a combination of deep learning, a type of artificial intelligence technology, and fake, which means 'fake'. It refers to complex results such as false image, video, and voices generated through the extraction– learning- generation step, which is a deep learning technology [5, 12]. Initially, it was a method of replacing and reproducing certain target segments of the original, but now it is possible to generate new images without original data through Generative Adversarial Network (GAN) technology [13]. AI human uses deepfake technology to produce real people into AI human, and deepfake technology is used to create virtual people with the same face as real people.

This deep-fake technology has attracted great attention as it can be produced as a moving video of the deceased based on past photos and video materials, but it is negatively recognized as the number of cases of abuse of deep-fake,

such as the production and distribution of false pornographic videos by singers and actors.

2.3.2 Lip-Sync Technology

Lip-sync technology is an advanced AI-driven system that synchronizes a person's lip movements with spoken audio, creating a seamless and natural-looking speech animation. It is widely used in various applications, including dubbing for movies, virtual avatars, video game characters, and deepfake videos.

This technology works by analyzing the phonetic structure of the audio and mapping it to corresponding lip movements. AI models, such as deep learning-based facial animation systems, learn the intricate relationship between speech sounds and mouth shapes. When applied to a video, the system modifies the subject's lips in real time to match the new audio, ensuring accurate synchronization. Lip-sync technology has revolutionized dubbing in the entertainment industry, allowing foreign-language films to appear more natural when translated. It is also widely used in virtual influencers, AI-powered customer service avatars, and educational tools that create engaging digital tutors. Additionally, video game developers employ lip-sync technology to enhance character realism in dialogue-driven storytelling [14].

Despite its many benefits, lip-sync technology also presents ethical challenges, particularly in the creation of deepfake content that can spread misinformation. To counteract potential misuse, researchers are developing detection methods and watermarking systems to distinguish authentic content from manipulated media. As AI continues to advance, lip-sync technology is becoming more sophisticated, offering new possibilities for interactive media [6], virtual communication, and entertainment while necessitating responsible use and regulation.

In this paper, we use it to generate income while gambling, as shown in Fig. 1.

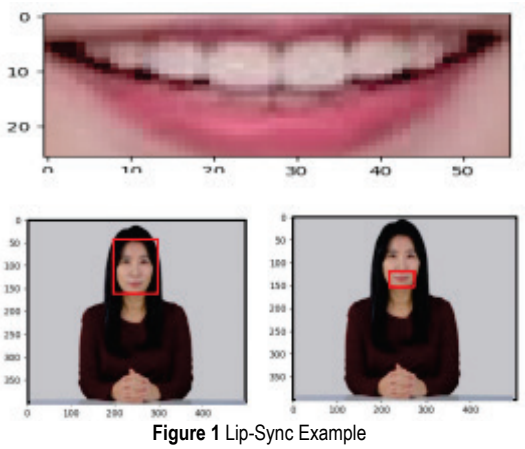

Figure 1 Lip-Sync Example

2.3.3 Few-Shot TTS Technology

Few-Shot Text-to-Speech (TTS) technology is a groundbreaking advancement in artificial intelligence that allows realistic voice synthesis with minimal training data. Traditionally, high-quality voice cloning required extensive recordings of a speaker's voice, sometimes hours of audio, to produce a natural and expressive synthetic voice. However, Few-Shot TTS drastically reduces this requirement by learning a speaker's unique vocal characteristics from just a few seconds or minutes of audio [1, 2].

Imagine someone wanting to create a digital voice replica for a virtual assistant or audiobook narration. With Few-Shot TTS, they would only need to provide a short voice sample—perhaps a few spoken sentences—and the AI would quickly analyze and replicate the speaker's tone, accent, and speech patterns. The model leverages deep learning techniques, such as variational autoencoders and attention mechanisms, to capture subtle vocal nuances with remarkable accuracy.

This technology has significant applications across industries. It can help create personalized digital assistants, generate synthetic voices for individuals who have lost their ability to speak, and enhance multilingual content by replicating a speaker's voice across different languages [5, 6]. However, it also raises ethical concerns, as it could be misused for deepfake audio, impersonation, or misinformation. As Few-Shot TTS continues to evolve, developers are working on safeguards such as watermarking and authentication techniques to ensure ethical use. This innovation is paving the way for more natural and accessible voice synthesis while challenging the boundaries of AI-driven communication.

In this paper, we created a video in six languages, including English, Japanese, Chinese, German, and Swedish, that extracts synthesized numbers from input text and displays 20 sentences. We utilized English TTS technology to enable multi-national use.

2.3.4 Silhouette Customizing Technology

Silhouette customization is a technique that allows individuals to modify the overall shape, fit, or structure of a product to better suit their personal style or functional needs. This method is commonly used in the fashion and footwear industries, where customers can adjust the contour, proportions, and drape of clothing or the shape and structure of shoes.

In fashion, silhouette customization enables users to choose between different cuts, such as slim-fit, relaxed-fit, or oversized styles. For example, a customer ordering a tailored suit can specify whether they prefer a classic, modern, or tapered silhouette to match their body shape and aesthetic preferences. Similarly, in footwear, brands may offer options to alter the toe shape, heel height, or arch support, ensuring both comfort and style [7, 8].

This technique enhances personalization by giving customers control over the visual and functional aspects of a product. It also helps brands cater to diverse body types and style preferences, making fashion more inclusive. Whether for practical purposes, such as improving fit and comfort, or for aesthetic appeal, silhouette customization ensures that the final product aligns with the wearer's unique identity and

needs.

Silhouette Customizing technology creates a composite image with natural gestures by generating various movements based on several human motion images taken as shown in Fig. 2, and produces a natural lecture by adding gestures to the speech content in the lecture.

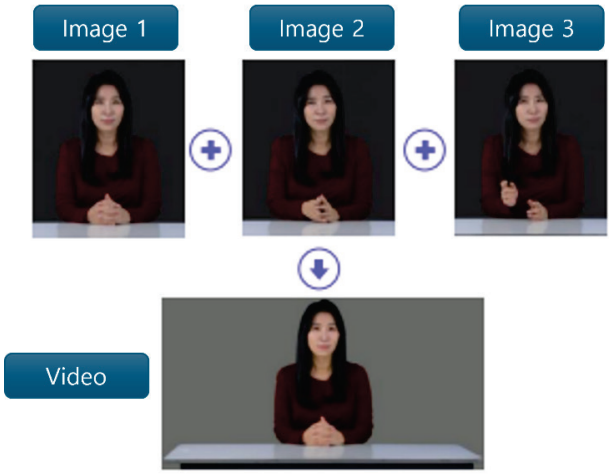

Figure 2 Silhouette Example

2.3.5 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a type of artificial intelligence model used for generating highly realistic synthetic data, such as images, videos, and audio. A GAN consists of two neural networks: a generator and a discriminator, which work against each other in a competitive process. The generator creates fake data by mimicking real samples, while the discriminator evaluates whether the data is real or fake. Over time, both networks improve as the generator learns to produce increasingly convincing outputs, and the discriminator becomes better at detecting fakes. This adversarial training process enables GANs to generate high quality, realistic content. GANs are widely used in image enhancement, art creation, deepfake technology, and even medical research [13, 14]. However, they also raise ethical concerns, as they can be misused for creating deceptive content, such as fake identities and manipulated videos. As GANs continue to evolve, researchers are working on methods to detect and regulate their applications responsibly.

3 PRODUCTION TOOLS AND EDUCATIONAL CONTEXT PRODUCTION

3.1 System Architecture

Beyond simple TTV (Text To Video), it is possible to customize images through the definition of metadata (AI human, image, and text) related to image production. In addition, CNAISTUDIO TOOL is used to easily produce online lectures, making it easy to manage projects for each subject/time, and provides a convenient UI for creating and editing video scenes on a slide basis of lecture textbooks.

In addition, it is designed to customize TTS with the function of a 'smart vocabulary book' for technical terms and dialects of a specific study. Finally, Super Resolution technology is applied to produce all lecture images in high-definition images of Ultra HD (2160p, 4K).

The AI professor system comprises:

- Deepfake Engine: Synthesizes instructor facial visuals using GANs.
- Few-shot TTS: Produces realistic speech in multiple languages from minimal voice data.
- Lip-sync Module: Aligns spoken language and facial articulation.
- Gesture/Body Rendering: Generates natural postures and gestures.

3.2 Content Production Tools

AI professor was produced based on AI human technology. The AI professor, made with deep-fake technology, Lip-sync technology, few-shot TTS technology, and silhouette customizing technology, showed technology that is not much different from real people in one scene. After producing the AI teaching model, it was intended to produce educational content using the AI teaching model and video production technology. Subsequently, the AI teaching model can be used for auxiliary lectures on the regular curriculum, and guidance on bachelor's and AI technologies. After the introduction of AI professors, content can be produced cheaply without restrictions on dedicated personnel, equipment, and places, and choices such as background suitable for the situation and users, and content insertion can be made.

In addition, multilingual videos can be produced immediately in a total of five languages, excluding Korean, so that linguistic suggestions can be resolved to Chinese, Japanese, Swedes, Germans, and all English-speaking students. Finally, when producing lectures using AI human, we developed a tool that can be easily produced so that anyone can produce videos using AI human without separate education, and the tool can be used to produce the desired video immediately. According to Fig. 3, an AI human was produced after a professor at a university to produce an AI professor, and an educational image was produced and realized using a video production tool. It is possible to quickly produce additional lectures by guiding AI professors produced using AI technology.

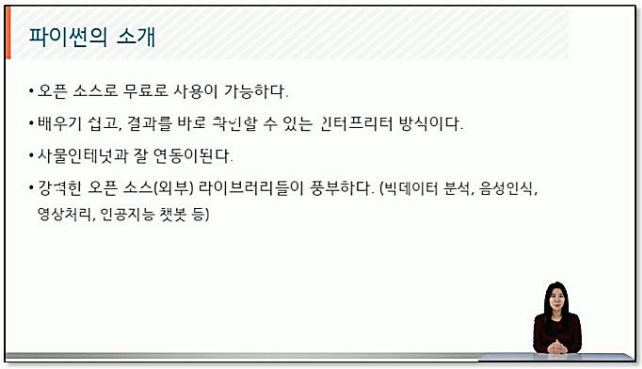

Figure 3 Lectures with human AI

We use CNAI Studio for scene scripting, voice cloning, animation, and rendering. Features include domain-specific vocabulary recognition, Ultra HD rendering, and accessibility settings for low-resource educators.

3.3 Lecture Design and Deployment

Content was structured around actual university courses, coding tutorials, and domain-specific instruction. Videos were produced in six languages. The modular design allows rapid adaptation to diverse curriculum requirements.

3.4 Evaluation Metrics

- Lip-sync Accuracy: Frame-aligned precision scoring at 93.1%
- TTS Quality: Mean Opinion Score (*MOS*) of 4.2/5
- Production Time: 60% time reduction compared to manual methods
- User Satisfaction: Mean score of 4.5/5 (n = 58), accessibility rating 4.7/5

4 RESULTS

AI-generated content was successfully deployed in online and hybrid courses. Participants reported improved learning engagement and comprehension due to native-language support. Visual quality and synchronization closely mimicked human instructors, achieving high realism.

This model contributes to three key areas:

- Pedagogical: Redefines instructional delivery beyond static MOOCs or live lectures.
- Technical: Combines AI synthesis methods into a unified, deployable pipeline.
- Practical: Offers a cost-effective, scalable, multilingual solution for higher education.

Limitations: The system currently lacks emotional interactivity and real-time adaptation. Ethical implications regarding authenticity and AI representation also merit further study.

5 CONCLUSION

This paper began with the focus of non-face-to-face online lectures changed due to the influence of the COVID-19 virus. For the necessity of non-face-to-face online classes and the resulting efficient learning effect, voices began to grow in the use of AI that introduced artificial intelligence, and AI professors were produced to create a better educational environment. Accordingly, with the production of AI professors, the production was carried out to make auxiliary videos of the regular curriculum created or to use them for bachelor's degrees and guidance. When producing AI professors, they record early using TTS technology, so if they finish recording, many videos can be produced using TTS technology, reducing time to record and edit themselves, and solving linguistic suggestions to foreign students taking classes. In addition, it is believed that the STUDIO Tool can be used to produce videos cheaply and easily without restrictions on dedicated personnel, equipment, and places, and that the video can be produced without separate education, making it less difficult for professors who produce videos. In addition, students who listen to education can quickly receive lectures, and so that you can focus on the time, you can focus on the time you can focus again. For these reasons, it is used in coding practice as high-quality education content, and it is used in coding practice.

This paper summarized the production of the AI teaching model, which was the beginning of the non-face-to-face education method, by removing the framework of the face-to-face education method and introducing artificial intelligence in line with the 4th industrial revolution. By producing online classes, we tried to supplement time-space constraints in offline classes, and lectures were produced with AI professors so that many lectures could be produced with less time. In addition, it is expected that the learning effect will be better than offline classes by allowing students to review the lecture when they want, and finally, AI professors using AI human can produce in five languages except Korean through TTS technology. Therefore, it is expected that international students studying abroad will be able to solve language inconveniences by solving difficulties that they did not understand due to their lack of Korean skills with TTS technology.

We developed and validated a scalable, multilingual AI professor model that enables autonomous lecture delivery with minimal human input. Empirical testing shows substantial gains in efficiency and accessibility. This approach is not a replacement for educators but a tool to enhance instructional reach, especially in multilingual and resource-limited environments. Explore emotional AI integration, immersive VR deployment, and empirical comparison across learning domains.

Acknowledgement

This research was supported by GPU resources provided through the 2025 High-Performance Computing Support Program, funded by the Ministry of Science and ICT and the National IT Industry Promotion Agency (NIPA) (G2025-0572).

6 REFERENCE

[1] Kim, D.-h. (2020). Changes and implications of the industrial and socioeconomic paradigm after the Corona crisis. *Interstate Finance Brief*, *29*(13), 16–18.

[2] Dwivedi, Y. K., Hughes, L., et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

[3] Oh, J.-h. (2020). The future ahead of COVID-19, from the era of education to the era of learning. *Issue & Diagnosis*, *421*, 1–25.

[4] Han, D.-e. (2021). Creating an interactive AI chatbot using dialog flow and applying it to elementary English classes.

Study on Learner-centered Curriculum Education, *21*(17), 517–529. https://doi.org/10.22251/jlcci.2021.21.17.517
[5] Kim, M.-j., & Lee, S.-k. (2022). Cyber threats and countermeasures to the development of artificial intelligence (AI) technology. *Journal of the Korean Communications Society (Information and Communication)*, *39*(12), 13–18.
[6] Han, S.-h., Choi, M.-i., & Cho, B. (2011). Real-time lip-sync and facial expression generation system for implementing interactive SNS. *Animated Studies*, *7*(2), 120–137.
[7] Hwang, Y.-h., & Kim, C.-s. (2021). Survey of non-face-to-face online lecture perception due to COVID-19: Focusing on university students' satisfaction and anxiety. *Linguistics*, *29*(1), 71–91.
[8] Kim, M. (2021). Post-Corona, video content production technology: Virtual Production System. *Imaging Technology Research*, *1*(35), 27–44.
[9] Seo, Y.-h., Oh, M.-s., & Han, K.-h. (2021). The present and future of digital humans. *Broadcasting and Media*, *26*(4), 72–81.
[10] Park, C.-s. (2022). Online education using digital human. *Comprehensive Academic Presentation Papers of the Korean Society of Information Technology*, ( ), 246.
[11] Kwon, E.-y. (2021). Trends in English teaching and learning research using AI. *English Studies*, *21*, 1313–1337.
[12] Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.
[13] Saxena, D., & Cao, J. (2021). Generative Adversarial Networks (GANs): Challenges, solutions, and future directions. *ACM Computing Surveys*, *54*(3). https://doi.org/10.1145/3446374
[14] Chakraborty, T., Reddy, U. K. S., Naik, S. M., Panja, M., & Manvitha, B. (2024). Ten years of generative adversarial nets (GANs): A survey of the state-of-the-art. *Machine Learning: Science and Technology*, *5*(1). https://doi.org/10.1088/2632-2153/ad1f77

Authors' contacts:

Yangha Chun, Professor
School of Artificial Intelligence, University Yongin,
134 Cheoin-gu, Yongin-si, Gyeonggi-do, 17092, South Korea

Soo-Yeon Yoon, Professor
(Corresponding author)
School of Software Convergence, Kookmin University,
77 Jeongneung-ro, Seongbuk-gu, Seoul, 02707, South Korea
Mobile Phone: +82-010-9264-0193
1104py@kookmin.ac.kr