Personalized Remote Piano Education via an AI-Empowered Teacher–Student Framework

Melody Han* and Geoffrey Ye Li†
*St. Johns School, Houston, TX, USA
†Department of Electrical and Electronic Engineering, Imperial College London, London, UK

***Abstract*—After COVID-19, remote education has emerged as a new pedagogical paradigm. However, traditional teacher-to-many-students instructional systems remain ill-equipped to deliver personalized guidance and granular feedback, thus online teaching quality needs to be enhanced urgently. To overcome this limitation, we propose an AI-powered personalized remote education framework that leverages artificial intelligence's generalization capabilities and adaptability to achieve large-scale personalized instruction. Within our proposed two-stage teacher-AI-student paradigm, the teacher first instructs a general AI agent, which is subsequently specialized to tutor individual students. We validate this framework through a personalized remote piano instruction case: A Transformer-based music transcription model converts performance audio into symbolic MIDI sequences, while the teacher-instructed AI agent refines these sequences into expressive performance versions. The resulting teacher-student performance comparison data provides students with personalized feedback on note onset timing and dynamics. To assess the effectiveness of the proposed framework, we collected piano proficiency examination data encompassing approximately 30 distinct pieces by different composers. Each piece's performance is treated as an independent student, thereby allowing us to analyze and evaluate the effectiveness of our proposed framework in providing personalized feedback to students. Experimental results across multiple piano pieces demonstrate that this framework effectively captures performance variations, delivers explainable feedback, and balances efficiency with personalized needs in remote learning environments.**

***Index Terms*—Remote education, Artificial Intelligence (AI), personalized instruction.**

I. Introduction

The rapid digital transformation of society has fundamentally reshaped how knowledge is disseminated and acquired. While traditional face-to-face education achieves teaching outcomes, it remains constrained by geographical limitations and infrastructure. The COVID-19 pandemic further exposed the vulnerability of traditional classrooms [1], [2]. Notably, remote education effectively safeguards learning continuity, equity, and accessibility, particularly benefiting learners in educationally underserved or geographically isolated regions [3], [4]. Consequently, remote education has become an indispensable component of the modern educational ecosystem.

Despite the promise of remote education, most current remote education systems still follow the conventional one-teacher-to-many-students paradigm (based on 18th century Prussian education system), merely transferring traditional classroom models into virtual platforms [5], [6]. This learning paradigm overlooks the inherent diversity of learners, where different students might possess different knowledge backgrounds, learning pace, and cognitive preferences, leading to disengagement and uneven learning outcomes [7]. Moreover, there lacks of interactive feedbacks for online teaching. In contrast, personalized remote education seeks to tailor the learning experience to each student's profile—enabling differentiated guidance, adaptive content delivery, and dynamic progress tracking [8]. To this end, we intend to propose a personalized remote education framework to advance the learning efficiency of students.

Artificial intelligence (AI) has exhibited exceptional capabilities in learning complex representations, adapting to dynamic environments, and generalizing across diverse tasks [9]–[11]. Such abilities make AI particularly suitable for modeling heterogeneous learner behaviors and tailoring instructional responses. By integrating these capabilities into remote education, AI can enable concurrent personalization, where each student's learning trajectory is treated as an individual yet interrelated task. This allows teaching content, pace, and feedback to evolve dynamically with learners' cognitive states and engagement patterns. Motivated by this potential, this paper aims to explore how AI can be systematically leveraged to construct a personalized remote education framework that achieves scalable, adaptive, and equitable personalization across diverse learners and learning scenarios.

We present our proposed AI-driven personalized remote education framework in Fig. 1, which illustrates a teacher-to-AI-to-student interaction process. In the first stage, the teacher instructs a general AI agent by providing curated instructional content and guidance. The AI agent then serves as an intelligent intermediary capable of generating diverse, context-aware instructions for individual students. In the second stage, each student engages in an independent study session while interacting with the AI agent. When students encounter any questions regarding the study materials that they are learning, the AI agent will provide feedback (answer) to students. With our proposed framework, a single teacher can simultaneously supervise and personalize the learning experiences of multiple students. In summary, the contributions of this paper are:

1. We propose an AI-driven personalized remote education framework. We treat each student's instruction as an individual AI task and utilize the generalizability of AI on diverse tasks to achieve personalized remote education.
2. We demonstrate the proposed framework via a personalized remote piano education case. In this case, we enable students learning different piano pieces, and the AI agent instructed by the teacher can provide personalized feedback to students for improvement.
3. We assess our proposed framework from the two perspectives of effectiveness and personalization. Our proposed framework can provide feedback in terms of the onset timing and velocity of each note for students. Moreover, the feedback is personalized according to the piano pieces selected by different students.

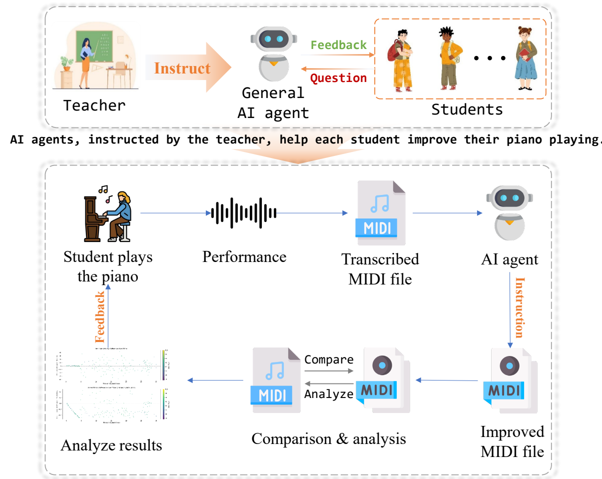

Fig. 1: Illustration of our proposed AI-driven personalized remote education framework. In our proposed framework, the instruction for students is not directly given by the teacher. Instead, the teacher will instruct AI agents at first. Subsequently, the AI agent will provide personalized instruction for each student as per their learning status based on the individual background and feedback.

We organize the remainder of this paper as follows: Section II presents the details of our proposed framework, key roles, and workflow. We illustrate the personalized remote piano education case in Section III, and assess the proposed framework in Section IV. Finally, we conclude the paper in Section V.

II. Personalized Remote Education Framework

In this section, we first introduce key roles in our proposed framework in Section II-A. Subsequently, we delineate the workflow of our proposed method in Section II-B.

A. Framework Key Roles

According to the framework depicted in Fig. 1, we consider one teacher and N students. In order to provide personalized instruction for N students, the teacher needs to instruct a general AI agent that can handle different learning trajectories of students. In the student instruction stage, the teacher will instantiate N AI agents to instruct each student, and therefore provide personalized education. To this end, in our proposed framework, three key roles exist, teacher, AI agent, and student.

Teacher: In our proposed framework, the teacher will not directly interact with the students but with the AI agent. In other words, the teacher will convey all of their teaching experience and knowledge to a general AI agent, i.e., training an AI agent.

AI Agent: The AI agent is responsible for interacting with students and solving the questions posed by students in our proposed framework. In particular, the teacher will create N AI agents and assign to each student an AI agent. With this setting, the AI agent can provide personalized instruction to the students.

Students: The students in our proposed framework can select any study material that they are interested in, instead of all students taking the same class together. Moreover, the students can interact and ask the assigned AI agent any questions during the study.

B. Framework Workflow

We present the workflow diagram of our proposed AI-driven personalized remote education framework in Fig. 1, which includes the teacher instructing a general AI agent and the student learn from the AI agent in two stages.

- At the first stage, the teacher will instruct a general AI agent (train a general AI model) that can handle multiple tasks. In our proposed framework, we consider the learning session between a student and an AI agent as an AI task. Therefore, the general AI agent instructed by the teacher needs to handle at least N different AI tasks.
- At the second stage, students select the topics that they are interested in studying. The student can ask the AI agent any questions encountered during the study, and the AI agent gives the solution to the student as feedback. Notably, at the second stage, we can deem the AI agent assigned to each student as a personalized teacher.

III. Piano Education Case Study

In this section, we demonstrate the feasibility of our proposed framework via a personalized remote piano education case. Specifically, we introduce the workflow of the personalized remote piano education in Section III-A. In Section III-B, we present the technical details that are utilized in our proposed framework.

A. Personalized Remote Piano Education

The personalized remote piano education framework includes two stages. In the first stage, the teacher will train a general AI agent that can transform the ordinary Musical Instrument Digital Interface (MIDI) [1] file into a professional one. In this way, the AI agent can help students improve their piano playing. Subsequently, in the second stage, the student will study piano playing via interaction with the personalized AI agent as follows.

1. Students will select any piano piece that they are interested in and obtain the audio-form performance.
2. We utilize a music transcription AI model to transcribe the audio-form performance into the MIDI file, which is the cornerstone for providing precise feedback to students.
3. The teacher instructed the AI agent will improve the transcribed student MIDI file into a professional one.
4. Upon the comparison and analysis between the transcribed student MIDI file and the improved MIDI file, we can derive precise feedback for students on onset timing and press velocity per note to augment the students' piano playing proficiency.

[1] MIDI is a standardized communication protocol that allows electronic musical instruments, computers, and other devices to exchange performance data such as pitch, velocity, and control signals.

In the next subsection, we will delineate the technical details regarding the music transcription AI model, AI agent training, and AI agent-aided personalized instruction.

B. Technical Details

In this subsection, we first introduce the music transcription AI model, which converts students' performances from audio recordings into MIDI representations. Next, we present the technical details of how the teacher trains a general AI agent to provide performance guidance to students. Finally, we describe the design of personalized instructional strategies generated by the teacher-instructed AI agent.

1) Music Transcription AI Model: To convert a student's piano performance from audio into a symbolic form for feedback, we employ a sequence-to-sequence Transformer-based music transcription model [12]. The model directly maps spectrogram frames $\mathbf{X} \in \mathbb{R}^{T\times F}$ to MIDI-like event sequences $\mathbf{Y} = (y_1, \cdots, y_L)$, where each token y_i represents a musical event such as Note, Velocity, Time, or end of sequence (EOS). Formally, the model learns the conditional probability

$$P(\mathbf{Y} \mid \mathbf{X}) = \prod_{t=1}^{L} P(y_t \mid y_{<t}, \mathbf{X}; \theta), \quad (1)$$

where θ denotes the Transformer parameters.

Each MIDI-like vocabulary token follows the structure proposed in [13]:

- **Note(pitch)**: 128 discrete pitch values,
- **Velocity(v)**: 128 levels representing key press intensity,
- **Time(t)**: absolute time quantized in 10 ms bins,
- **EOS**: end-of-sequence indicator.

During training, the objective is to minimize the cross-entropy loss between the predicted and ground-truth sequences:

$$\mathcal{L}_1 = -\sum_{t=1}^{L} \log P(y_t^* \mid y_{<t}, \mathbf{X}; \theta). \quad (2)$$

2) AI Agent Training: Based on the transcribed MIDI sequence $\mathbf{Y} = (y_1, \cdots, y_L)$ obtained from the student's performance, the AI agent is trained to generate an expressive and pedagogically enhanced performance for feedback and demonstration. Following the ScorePerformer framework [14], the agent adopts a Transformer-based encoder–decoder architecture with hierarchical variational style modeling. It learns multi-level latent embeddings $\{z_G, z_B, z_b, z_o\}$ that capture performance nuances at the global, bar, beat, and onset levels.

To ensure smooth and interpretable latent representations, the model minimizes the maximum mean discrepancy (MMD) [15] between the learned posterior distribution $q(z)$ and a prior Gaussian $p(z)$:

$$\begin{aligned}\mathcal{L}_{\text{MMD}} =& \mathbb{E}_{p(z),p(z')}[k(z,z')] + \mathbb{E}_{q(z),q(z')}[k(z,z')] \\ &- 2\mathbb{E}_{p(z),q(z')}[k(z,z')],\end{aligned} \quad (3)$$

where $k(z,z') = \exp(-\|z-z'\|^2/2\sigma^2)$ is the Gaussian kernel function.

Conditioned on $\mathbf{Y}$ and the learned hierarchical style embeddings, the performance decoder predicts expressive control tokens $\hat{\mathbf{Y}} = (\hat{y}_1, \cdots, \hat{y}_L)$ that refine timing, dynamics, and articulation for personalized learning feedback:

$$\mathcal{L}_{\text{perf}} = -\sum_{t=1}^{L} \log p_\theta(\hat{y}_t \mid \hat{y}_{<t}, \mathbf{Y}_{\le t}, z_{\le t}). \quad (4)$$

The total training loss combines reconstruction and regularization terms:

$$\mathcal{L}_2 = \mathcal{L}_{\text{perf}} + \mathcal{L}_{\text{MMD}}. \quad (5)$$

Through this training process, the AI agent learns to enhance the student's transcribed performance into a more expressive and stylistically accurate rendition, forming the basis for adaptive and personalized piano feedback within the remote education framework.

3) AI Agent-Aided Personalized Instruction: After obtaining the transcribed MIDI sequence $\mathbf{Y} = (y_1, \cdots, y_L)$ from the student's performance and the AI-refined version $\hat{\mathbf{Y}} = (\hat{y}_1, \cdots, \hat{y}_L)$ generated by the trained agent,

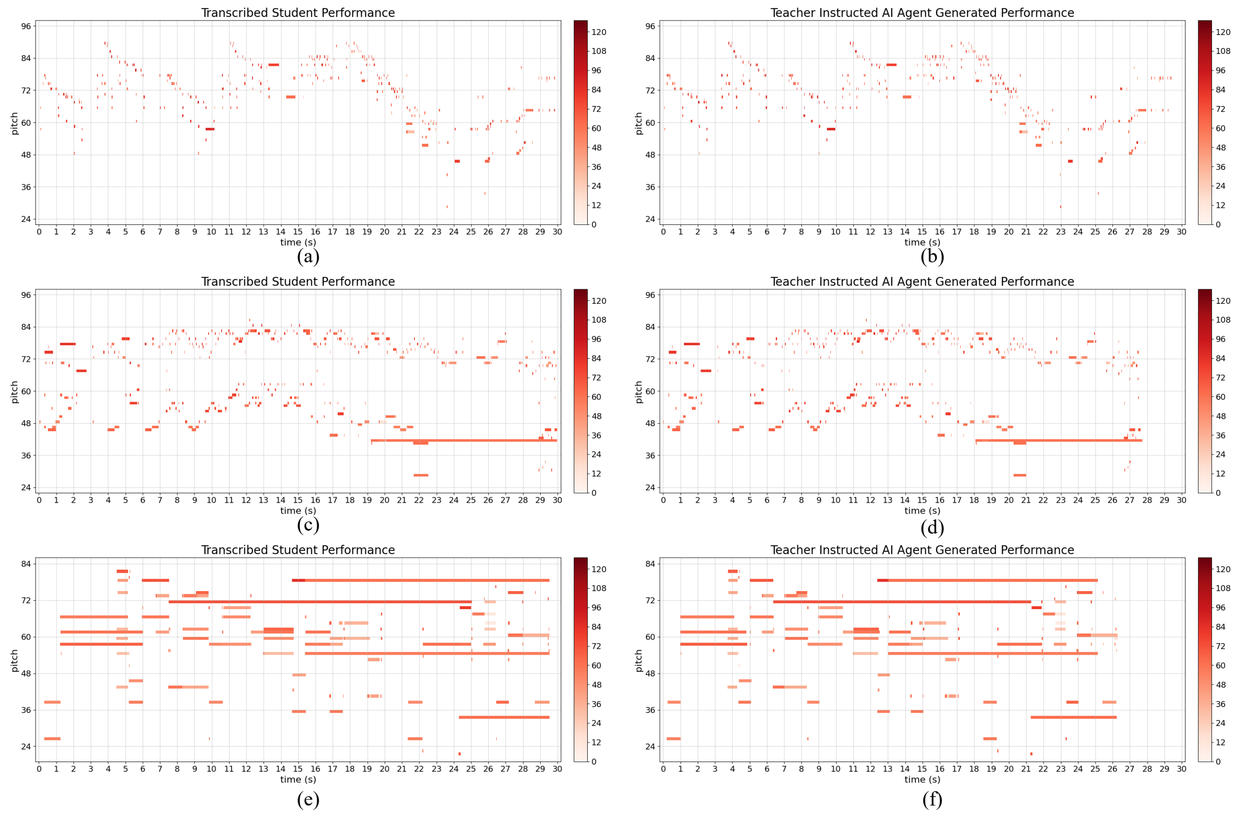

Fig. 2: MIDI visualization of three different students' performance and the associated personalized augment version via AI agents. Here, (a) and (b) are Curious Story, Op. 138 No. 9, which was composed by Heller. (c) and (d) are Fugue in B, BWV 866, which was composed by Bach. (e) and (f) are Gymnopédie No. 1, which was composed by Satie.

a personalized instruction module is implemented to analyze the performance differences between the two sequences. Specifically, each MIDI file is first converted into a set of note events $(t_{\text{on}}, t_{\text{off}}, p, v)$, where t_{on} and t_{off} denote the onset and offset times, p is the MIDI pitch, and v is the key velocity.

A greedy pairing procedure aligns the student and AI notes by pitch and onset time within a tolerance τ:

$$|t_{\text{on}}^{(\text{stu})} - t_{\text{on}}^{(\text{AI})}| \leq \tau, \tag{6}$$

thereby ensuring that only musically corresponding notes are compared. For each paired note, the module computes two expressive dimensions:

$$\Delta t_{\text{on}} = t_{\text{on}}^{(\text{AI})} - t_{\text{on}}^{(\text{stu})}, \tag{7}$$

$$\Delta v = v^{(\text{AI})} - v^{(\text{stu})}. \tag{8}$$

Here, Δt_{on} measures timing deviation and Δv represents velocity (dynamics) difference between the student and AI performances.

The calculated differences are visualized as personalized feedback, enabling learners to understand how the AI model refines their playing in terms of timing precision and expressive control. Consequently, the AI agent not only generates an enhanced rendition $\hat{\mathbf{Y}}$, but also serves as an interactive instructor that guides students toward expressive and technically accurate performance within the proposed remote education framework.

IV. Experimental Evaluation

We state the experimental configurations regarding the personalized remote piano education framework in Section IV-A. Subsequently, in Section IV-B, we assess our proposed frame-

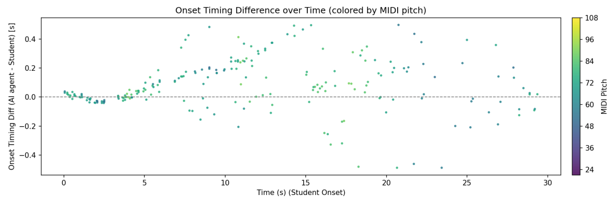

Fig. 3: Feedback on the onset timing via comparison between the AI agent's improved performance and the student played performance.

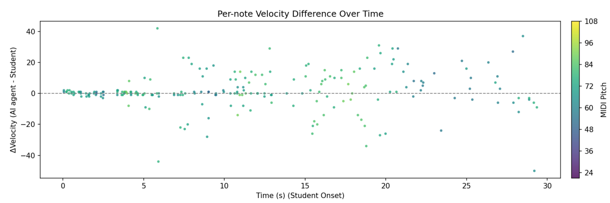

Fig. 4: Feedback on the note press velocity via comparison between the AI agent's improved performance and the student played performance.

work from the effectiveness and personalization perspectives.

A. Experimental Configurations

To simulate that different students possess distinct preferences and proficiency in piano playing study, we collect almost 30 student performances with different proficiency levels. Concretely, we collected data from three students' piano proficiency examinations. Each examination lasted for more than one hour and included performances of over ten piano pieces composed by different composers. Regarding the music transcription AI model, we load the model weights from [12]. Analogously, we load the model weights from [14] to represent the AI agent model that is already well-instructed by the teacher. Lastly, we consider the onset time tolerance $\tau = 0.5$ seconds in the note pairing procedure by default.

B. Effectiveness of Proposed Framework

To demonstrate the effectiveness of the teacher-instructed AI agent model, we randomly select three different student performances to visualize the MIDI comparison between the transcribed student performance and the AI agent-generated performance in Fig. 2. Generally, observing Fig. 2, the MIDI visualization between the student performance and AI agent improved performance is similar, with a slight difference in onset timing and note press velocity. We present the numerical comparison regarding the onset timing and note press velocity of Curious Story, Op. 138 No. 9 between the student and the AI agent in Figs. 3 and 4, respectively. Observing Fig. 3, we can conclude that the onset timing is precise at the first 5 seconds; However, after the first 5 seconds, the onset timing of the student performance is slightly slow. Analogously, observing Fig. 4, the note press velocity of the student is maintained well at the first 10 seconds. However, the note press velocity of student does not sufficiently stabilize after the first 10 seconds. Therefore, we can conclude that our proposed framework is indeed effective in giving feedback to students for piano playing improvement.

TABLE I: Statistical Feedback Results for three different students in Terms of Onset Timing

	Mean	Median	Std	MAE
Heller	0.0800	0.0429	0.1761	0.1386
Batch	0.0244	0.0252	0.2090	0.1575
Satie	0.0579	0.2021	0.2889	0.2641

Observing Tables I and II, we can conclude that the AI agent provides distinct and personalized feedback to students for piano playing improvement. Specifically, as per the results depicted in Table I, the onset timing of student 2 in playing the Bach composed Fugue in B, BWV 866 is excellent. However, the results depicted in Table II indicate that Student 2 needs to pay more attention to the note press velocity. Student 3 cannot control well in both onset timing and note press velocity when playing Satie's composed Gymnopédie No. 1. Therefore, our proposed framework can provide personalized feedback to students for piano playing.

V. Conclusion and Future Work

In this paper, we introduced an AI-driven personalized remote education framework that

TABLE II: Statistical Feedback Results for three different students in Terms of Note Press Velocity

	Mean	Median	Std	MAE
Heller	1.5251	0	12.1961	7.744
Batch	1.0876	0	11.2046	7.023
Satie	0.0967	1	14.0602	9.5806

enables scalable, adaptive, and student-specific instruction. By modeling each student–AI interaction as an independent learning task, the framework harnesses the cross-task generalization power of AI to provide customized educational experiences. Through a remote piano education case, we demonstrated the end-to-end pipeline—from music transcription and AI-based expressive performance rendering to note-level feedback analysis. Experimental results confirmed that the AI agent can generate refined MIDI outputs closely aligned with professional interpretations and produce personalized feedback on timing and dynamics for different students. These findings verify that the proposed framework not only supports individualized improvement, but also enhances overall engagement and learning efficiency.

In future work, we plan to expand the dataset to include a larger and more diverse group of students, as well as a broader range of musical pieces and styles, to enhance the generalizability and robustness of the findings. We will also implement comparative studies by benchmarking the AI-generated feedback against human expert evaluations and existing intelligent tutoring systems to better contextualize the system's performance. Furthermore, we aim to conduct a comprehensive user study or pilot experiment with real students to evaluate educational outcomes, including learning improvement, engagement, and usability in real-world settings. In addition, we plan to analyze potential limitations such as model bias, scalability, generalization to other instruments or subjects, and computational requirements. These efforts will provide a more complete assessment of the system's educational effectiveness, fairness, and practicality, and guide the development of a more adaptive and scalable AI-based tutoring framework.

References

[1] A. Bozkurt and R. C. Sharma, "Emergency remote teaching in a time of global crisis due to coronavirus pandemic," *Asian journal of distance education*, vol. 15, no. 1, pp. 1–6, Apr. 2020.

[2] C. Hodges, S. Moore, B. Lockee, T. Trust, and A. Bond, "The difference between emergency remote teaching and online learning," *Educause review*, vol. 27, no. 1, pp. 1–9, Mar. 2020.

[3] S. Dawadi, F. Goshtasbpour, and A. Kukulska-Hulme, "Equitable access to higher education learning and assessment: Perspectives from low-resource contexts," *Journal of Interactive Media in Education (JIME)*, vol. 2024, no. 1, pp. 1–15, May. 2024.

[4] F. Gottschalk and C. Weise, "Digital equity and inclusion in education: An overview of practice and policy in oecd countries," *OECD Education Working Papers*, vol. 14, no. 299, pp. 1–75, Aug. 2023.

[5] M. Joshi and R. Vaidya, "A review of paradigm shift from conventional to personalized e-learning," in *International Conference on Advances in Computing, Communications and Informatics*, Mysore, India, Aug. 2013.

[6] H. J. Kim, J. Choi, and S. Lee, "Teacher experience of integrating tablets in one-to-one environments: Implications for orchestrating learning," *Education Sciences*, vol. 9, no. 2, pp. 1–13, Apr. 2019.

[7] B. Means, M. Bakia, and R. Murphy, *Learning online: What research tells us about whether, when and how*. Routledge, 2014.

[8] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, vol. 10, no. 3, p. e1355, Jan. 2020.

[9] H. Wen, H. Xing, and O. Simeone, "Pre-training and personalized fine-tuning via over-the-air federated meta-learning: Convergence-generalization trade-offs," *arXiv preprint arXiv:2406.11569*, Jun. 2024.

[10] J. Qiu, J. Wu, H. Wei, P. Shi, M. Zhang, Y. Sun, L. Li, H. Liu, H. Liu, S. Hou *et al.*, "Visionfm: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence," *arXiv preprint arXiv:2310.04992*, Oct. 2023.

[11] B. Y. Lin, K. Tan, C. Miller, B. Tian, and X. Ren, "Unsupervised cross-task generalization via retrieval augmentation," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Dec. 2022.

[12] C. Hawthorne, I. Simon, R. Swavely, E. Manilow, and J. Engel, "Sequence-to-sequence piano transcription with transformers," in *Proc. 22nd Int. Society for Music Information Retrieval Conf. (ISMIR)*, Virtual, Nov. 2021.

[13] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, "This time with feeling: Learning expressive musical performance," *Neural Computing and Applications*, vol. 32, no. 4, pp. 955–967, Nov. 2018.

[14] I. Borovik and V. Viro, "Scoreperformer: Expressive piano performance rendering with fine-grained control," in *Proc. 24th Int. Society for Music Information Retrieval Conf. (ISMIR)*, Milan, Italy, Nov. 2023.

[15] X. Jia, M. Zhao, Y. Di, Q. Yang, and J. Lee, "Assessment of data suitability for machine prognosis using maximum mean discrepancy," *IEEE transactions on industrial electronics*, vol. 65, no. 7, pp. 5872–5881, Jul. 2018.