A Generative Framework for Designing Interactions to Overcome the Gaps between Humans and Imperfect AIs Instead of Improving the Accuracy of the AIs

Hiromu Yakura
hiromu.yakura@aist.go.jp
University of Tsukuba / National Institute of Advanced Industrial Science and Technology (AIST)
Japan

ABSTRACT

My research focuses on improving human-machine collaboration in the context of machine learning, particularly by recognizing the limitations and potential for errors in machine learning techniques and designing effective interactions for filling the gaps between humans and them. To this end, I have explored the application of machine learning in a variety of domains, such as malware analysis, music recommendation, conversation analysis, photo editing, and video-based learning. I also worked on clarifying the limitations of the current technologies by using adversarial approaches and qualitative methods. My thesis is planned to synthesize what I learned from these projects into design principles for constructing interactions that take full advantage of imperfect machine learning models. I particularly put emphasis on deriving principles that do not depend on the fine-tuning of the models, thereby providing a generative framework allowing researchers and practitioners to design a range of effective intelligent interactions without incurring significant computational and data collection costs.

ACM Reference Format:
Hiromu Yakura. 2023. A Generative Framework for Designing Interactions to Overcome the Gaps between Humans and Imperfect AIs Instead of Improving the Accuracy of the AIs. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544549.3577036

1 INTRODUCTION

My primary research interests are at the intersection of human-computer interaction and machine learning. In particular, I am focusing on how we can overcome the gaps people have against machine learning techniques that require a well-defined application scheme and can produce wrong results. I acknowledge that a number of researchers have contributed to improving the accuracy of machine learning, but I believe that it is also important to investigate how humans can benefit from imperfect machine learning systems. In addition, to promote sustainable collaboration between humans and machine learning systems, we need to carefully design how the systems obtain feedback from humans and intervene with them. Considering that an increasing number of machine learning systems are employed in many people's daily lives, we need not only to improve machine learning techniques themselves but also to cultivate effective ways of leveraging them.

My previous projects are in line with this context; for example, the one on malware analysis support [29, 30] focused on the gap that, while computer security experts wanted to automate costly analysis, the introduction of machine learning had not been so favored because it might put errors in the analysis results that are used for forensic purpose. Then, instead of automated end-to-end analysis, I proposed a new approach that offers hints for experts who are beginning to analyze by highlighting characteristic behaviors using the attention mechanism. Another project proposing a music recommender dedicated to being used while working [26, 27] stemmed from the gap behind the scientific result that people should listen to songs they feel mediocre for keeping concentrated. That is, it is difficult for humans to find their mediocre songs by themselves, while it is much easier for a recommender to find such songs rather than to find their best songs.

In my thesis, I am planning to discuss the principle of interaction designs that fill such gaps based on a series of research that I have conducted and am working on. My approach is based on both recognizing the limitation of current technology and providing users with control to overcome the limitations, as presented in the following section.

2 REPRESENTATIVE PAST PROJECTS

2.1 Computational support for executive coaching

In executive coaching, coaches are required to analyze the nonverbal behavior of a coachee during a conversation, which is mentally demanding. Therefore, I started this independent project with my undergraduate friend, assuming that computationally analyzing such nonverbal behavior can help coaches. However, in the early attempts, we found that conventional methods based on supervised machine learning cannot consider the contextual semantics of the behavior and often contradict the intuition of experienced coaches.

Hence, we developed a computational support system that captures sudden changes in coachees' nonverbal behavior in an unsupervised manner using the Gaussian mixture model (GMM) [32], as in Figure 1. This is principled in the separation of observation and interpretation, which is inspired by the fact that expert coaches do actually not have difficulty in interpreting the implications of nonverbal cues they observed. Rather, they are afraid of missing out on some cues while mentally demanded to manage sessions by, for

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3577036

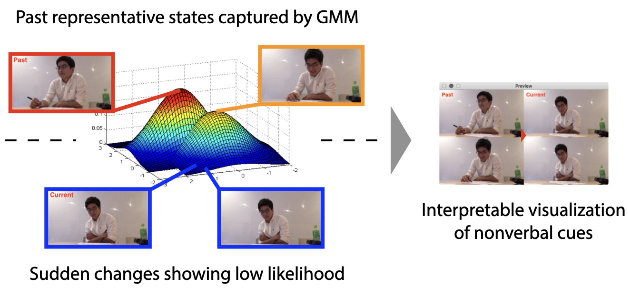


Figure 1: GMM allows us to capture sudden changes in the nonverbal behavior of a coachee, which help coaches infer the coachee's internal status [3].

example, finding an appropriate question that can provide a chance for coachees to have an effective reflection. Thus, our design is dedicated to supporting only the observation part and delegates the interpretation part to coaches via visualizing the observed cues in an interpretable manner, instead of directly inferring the meaning of the cues based on rules or simplified classifications. We confirmed the effectiveness of this approach with expert coaches, and this work was accepted to ACM CHI '19 and '20 [3, 4].

2.2 Practical adversarial examples against daily-use devices and cars

To overcome the gap between humans and machine learning, we also need to understand the limitations of machine learning techniques. To this aim, I have worked on research about adversarial examples to reveal the risk of machine learning systems being intentionally misguided. In IJCAI '19, we showed that, by broadcasting a subtle noise, we can make deep-learning-based speech recognition devices in the physical world transcribe unuttered speech from normal music clips, which can result in abuse of voice assistants [28].

In AAAI '20 [23], we showed that placing a moth-like patch on a STOP sign can make autonomous driving cars misrecognize it as Speed 80 (Figure 2). This work revealed the existence of the gap from another perspective; while human drivers would not feel the patch suspicious by assuming it to be a moth stopping on the sign, it can fool the cars as an adversarial example. Given that adversarial examples are common to deep learning models, this work emphasizes the importance of designing interactions in which humans can cope with such a malfunction of machine learning by providing users with transparent control.

2.3 Explorable content editing for human-centered design processes

The importance of such transparent control is also emphasized in my other project, which focused on the gap between deep-learning-based content editing techniques and the creative process of humans. Here, our design process is often exploratory, that is, an open-ended journey starting with an under-specified goal [19]. Therefore, "one-shot" editing provided by end-to-end techniques is not optimal for serving such a serendipitous process. For example, many people

Figure 2: Moth-like patches on a 'STOP' sign can make autonomous driving cars misrecognize it as 'Speed Limit 80' [23].

are still enjoying editing photos using Instagram, exploring possible results, whereas state-of-the-art photo enhancement transfer methods can produce high-fidelity results.

Given that, I came up with the idea of letting users know how to obtain a stylized result in a tool they are familiar with, instead of providing only the stylized result (Figure 3). In photo editing, users can know which transformations should they apply to what extent in order to obtain a stylized result, and then, they can enhance contrast or disable a specific filter. This is enabled by combining black-box optimization with a perceptual metric retrieved from a pretrained style transfer model. Importantly, this approach allows us to leverage various pretrained models accumulated via machine learning research. We confirmed in experiments its applicability to stylizing pictures and making up facial photos, which was accepted to IJCAI '21 [25].

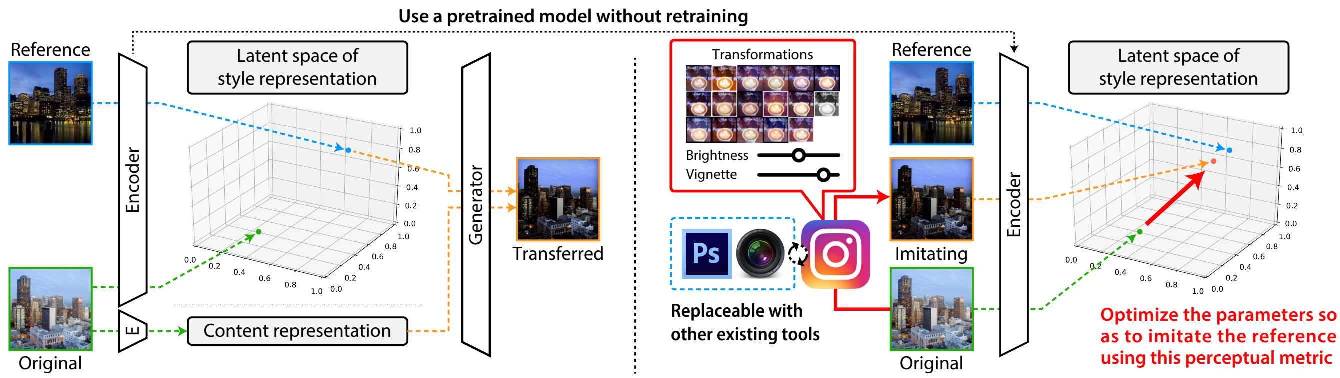
2.4 Designing false-positive resistant interventions for behavior change

I have also worked on designing machine-learning-based interventions for behavior change under the assumption that errors in machine learning systems are inevitable. Specifically, in our ACM CHI '21 paper [5], we explored the design of an intervention to guide users' attention to video communication (e.g., video-based learning) when they are detected as not focusing on by a machine learning system. Here, explicitly alerting users would lead them to ignore it, especially when the detection is false positive, resulting in the deterioration of their trust in the system. Thus, we designed an intervention that perturbs the voice in a video to direct their attention without consuming their conscious awareness based on the nature of human speech communication, as humans often create such paralinguistic changes so as to draw listeners' attention.

We experimentally showed that our approach is effective not only in terms of helping users refocus their attention but also maintaining their trust. This is because it was not likely to frustrate users even though the intervention is activated by false-positive detection of machine learning. This paper shows that the idea of Mindless Computing [1] becomes more powerful when combined with imperfect machine learning systems and is awarded Best Paper Honourable Mention at ACM CHI '21.

3 TOWARDS COMPLETING MY THESIS

As described above, I have consistently worked on how to fill the gap between machine learning techniques and humans in various application fields, which further includes virtual reality concerts



(a) Typical workflow of an end-to-end style transfer with GANs (b) Parametric transcription of the transfer with a set of provided transformations

Figure 3: Our approach presents a way to stylize so that a user can explore the variations of a stylized result in a familiar tool.

[24], vocabulary learning [8], and audio transcription [7]. Currently, I am working on investigating the effect of supporting intellectual tasks (e.g., writing texts or preparing presentations) using large language models [9], which includes not only making progress but also maintaining users' task engagement (i.e., avoiding procrastination). For the latter purpose, models do not always have to be accurate but can produce some interesting or funny outputs, which makes their relationship with the users robust and long-lasting. Therefore, we are evaluating the effectiveness of using the output of such models as an intervention to guide users to resume their tasks.

Simultaneously, I have completed some concomitant projects that leverage both empirical methods and state-of-the-art techniques. For example, my CHI' 21 [22] and CSCW '22 [6] papers used qualitative approaches, mainly semi-structured interviews, to figure out people's responses to newly introduced intelligent systems. On the other hand, we proposed a new representation learning method for singing voices [31], which was necessitated to expand the application of our content editing method [25] to the audio domain. This method exploited self-supervised contrastive learning [13] to enable measurement of similarity in voice timbre or singing expressions in an analogous manner to human perception.

However, toward completing my thesis, I have to construct a strong narrative that connects all of my research activities. Through the above projects, I found that my key research question can be reworded as follows.

> *What are the design principles required to construct interactions that fully benefit from imperfect machine learning systems without spending large computational and data collection costs to fine-tune them?*

As presented in Section 2, I have explored ways to exploit machine learning techniques without employing regular supervised-learning-based approaches. In this section, I would like to explain the current plan for turning what I learned from these projects into a single set of design principles.

3.1 Lens of Parasuraman's framework of automation

Currently, I presume that Parasuraman et al.'s framework of automation [14] can be a basis to connect these projects. They presented a pioneering framework for automation design, which can be applied to four classes of functions: information acquisition, information analysis, decision & action selection, and action implementation. Then, they defined a 10-point scale to evaluate the automation level for each function, from level 1 ("the computer offers no assistance: human must take all decisions and actions") to level 10 ("the computer decides everything, acts autonomously, ignoring the human"). This framework allows us to assess the suitability of automation design by determining whether the automation level of each function is appropriately situated within the trade-off between human performance, automation reliability, and cost of consequences.

Parasuraman et al. take an example of the ground proximity warning system (GPWS) that alerts pilots when their aircraft is flying into the ground. They classified it as level 4 ("suggests one alternative") in decision & action selection because it does not execute any countermeasures but makes suggestions for pilots with an audio message that says "pull up!" However, when such an alert is triggered, the performance of pilots is often degraded but its cost of consequences can be too high, i.e., resulting in a crash. Therefore, GPWS is now being replaced with automatic ground collision avoidance, which exhibits level 7 ("executes automatically, then necessarily informs the human") automation by taking control automatically if the pilot does not.

This framework can be applied to the design of my past projects, for example, the computational support for executive coaching (Section 2.1). Our developed system retains level 10 for both information acquisition and information analysis while exhibiting level 1 for both decision & action selection and action implementation. Specifically, it acquires information about a coachee's nonverbal behavior and analyzes its outlierness, working independently from the coach. However, it touches neither decision & action selection nor action implementation; it leaves the coach to interpret the observed cues and take the necessary action (e.g. asking the coach another question to determine the reason behind it).

This is rationalized by considering the trade-off between human performance, automation reliability, and cost of consequences. Here, coaches' performance in observing coachees' nonverbal behavior is not consistent due to the mental demand to carry out the sessions. In contrast, the reliability of GMM-based unsupervised learning in detecting nonverbal cues is quite high because it does not introduce any heuristics or biases caused by the training data. Furthermore, its interpretable visualization (see Figure 1) allows the coach to ignore the cues when they are not implicational, which minimizes the negative cost of consequences. A similar discussion can be applied to the interpretation part, i.e., high human performance with low automation reliability. In this part, the cost of consequence is also higher than the observation part because misinterpretation of the nonverbal cues can mislead the session.

The situation of guiding users' attention (Section 2.4) provides another example; in this case, we cannot expect the human performance of noticing oneself being distracted. Nevertheless, we cannot guarantee automation reliability because the information that machine learning systems can access via a webcam would be sometimes insufficient to achieve high-accuracy detection of distraction. Furthermore, the cost of consequences of false positives is quite expensive considering that being intervened while concentrating in fact would be particularly frustrating. We can say that we have resolved this trilemma by introducing a new intervention that is not frustrating even when false positives occur, which in turn reduces the cost of consequences.

However, this framework does not fully cover all possible designs, given the expansion of the capabilities and application areas of machine learning techniques. Specifically, applying machine learning techniques for creativity support (Section 2.3) does not involve typical decision making and action implementation processes. Rather, the most important factor in this situation is how machine learning systems can help the rapid and iterative exploration within users' design processes. Thus, we should consider the automation level not as the balance of the ownership between humans and systems but as the level of their fluent collaboration. This point suggests the demand for reframing Parasuraman et al.'s framework so as to conform with a wider range of machine learning applications.

3.2 Current plan

I am now working to survey and critically analyze recent research on guidelines for human-centered applications of machine learning techniques. In particular, I acknowledge that a large body of research has contributed to the proposal of guidelines for various related concepts, such as human-AI collaboration, human-AI teaming, human-in-the-loop intelligent interactions, etc. [2, 11, 18, 20, 21]. For example, Amershi et al. [2] presented 18 items to consider when designing AI-infused systems, which include "make clear what the system can do," "mitigate social biases," and "provide global controls." As implied by these items, most research has focused on general recommendations to avoid common pitfalls that occur upon the use of machine learning techniques.

However, I would like my thesis to present a *generative* framework that can guide researchers and practitioners to explore new ideas to exploit machine learning techniques, rather than to avoid pitfalls when implementing the ideas. This is one of the reasons why I settled Parasuraman's framework [14] as a basis of the discussion. Their framework allows us to explore new automation designs by considering "what if we change the automation level?" or "how can we overcome the trilemma of human performance, automation reliability, and cost of consequence?" I believe that this point is important to differentiate my discussion from the line of recent research on guidelines for the human-centered application of machine learning techniques.

In this sense, another line of research that can provide inspiration would be creativity support. For several decades, many ideas about how we can exploit computers for supporting human creativity have flourished [16, 17], which have consequently invited researchers to construct taxonomies of supporting tools [10, 12]. These taxonomies often serve as a generative framework, as they highlight under-explored types and application areas of supporting tools. Still, discussions in this line have not been situated in the context of the latest advancement of machine learning techniques, except for the one by Rezwana and Maher [15]. Therefore, I would like to complement the literature by applying an analogous approach to discuss how we can exploit recent machine learning techniques based on what I learned from my past projects (Section 2).

4 CONCLUSION

My consistent research theme is how we can overcome the gaps between humans and imperfect machine learning techniques. Although improving their accuracy would be a solution, I decided not to make such an effort considering the computational and data collection cost of customizing them. Instead, I chose to explore interactions that can bridge the gap so that we can benefit from the collective intelligence of researchers and practitioners of both machine learning and HCI. I completed a series of research that presented such interactions in various application areas, such as executive coaching, content editing, and behavior change. Now, I am working toward turning the findings from my past projects into a generative framework that can foster the collective intelligence to induce future interaction designs that leverage unseen machine learning techniques to be proposed. This requires an overarching understanding of the related topics, which includes a long history of human-centered computing and creativity support, to situate my thesis. Therefore, I believe that interacting with other doctoral students and senior researchers in this community certainly helps me to construct a bold narrative.

Acknowledgments

This work was supported in part by JST ACT-X (JPMJAX200R) and JSPS KAKENHI (JP21J20353). I also express my appreciation for the support from Google Ph.D. Fellowship and Microsoft Research Ph.D. Fellowship.

REFERENCES

[1] Alexander Travis Adams, Jean Marcel dos Reis Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless Computing: Designing Technologies to Subtly Influence Behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, 719–730. https://doi.org/10.1145/2750858.2805843

[2] Saleema Amershi, Daniel S. Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi T. Iqbal, Paul N. Bennett, Kori Inkpen,

Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3. https://doi.org/10.1145/3290605.3300233

[3] Riku Arakawa and Hiromu Yakura. 2019. REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 572. https://doi.org/10.1145/3290605.3300802

[4] Riku Arakawa and Hiromu Yakura. 2020. INWARD: A Computer-Supported Tool for Video-Reflection Improves Efficiency and Effectiveness in Executive Coaching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1–13. https://doi.org/10.1145/3313831.3376703

[5] Riku Arakawa and Hiromu Yakura. 2021. Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 99:1–99:15. https://doi.org/10.1145/3411764.3445339

[6] Riku Arakawa and Hiromu Yakura. 2021. Reaction or Speculation: Building Computational Support for Users in Catching-Up Series Based on an Emerging Media Consumption Phenomenon. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–28. https://doi.org/10.1145/3449225

[7] Riku Arakawa, Hiromu Yakura, and Masataka Goto. 2022. BeParrot: Efficient Interface for Transcribing Unclear Speech via Respeaking. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. ACM, New York, NY, 832–840. https://doi.org/10.1145/3490099.3511164

[8] Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. 2022. VocabEncounter: NMT-powered Vocabulary Learning by Presenting Computer-Generated Usages of Foreign Words into Users' Daily Lives. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 6:1–6:21. https://doi.org/10.1145/3491102.3501839

[9] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Proceedings of the 2020 Annual Conference on Neural Information Processing Systems*. Curran Associates, Red Hook, NY, 1877–1901.

[10] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Proceedings of the 2021 ACM Conference on Designing Interactive Systems*. ACM, New York, NY, 1817–1833. https://doi.org/10.1145/3461778.3462050

[11] Christopher Flathmann, Beau G. Schelble, Rui Zhang, and Nathan J. McNeese. 2021. Modeling and Guiding the Creation of Ethical Human-AI Teams. In *Proceedings of the 4th AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, 469–479. https://doi.org/10.1145/3461702.3462573

[12] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions. In *Proceedings of the 2018 ACM Conference on Designing Interactive Systems*. ACM, New York, NY, 1235–1257. https://doi.org/10.1145/3196709.3196732

[13] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. CVF/IEEE, New York, NY, 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975

[14] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2000. A Model for Types and Levels of Human Interaction With Automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30, 3 (2000), 286–297. https://doi.org/10.1109/3468.844354

[15] Jeba Rezwana and Mary Lou Maher. 2022. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *arXiv* 2204.07666 (2022), 27 pages. https://doi.org/10.48550/arXiv.2204.07666

[16] Ben Shneiderman. 2000. Creating Creativity: User Interfaces for Supporting Innovation. *ACM Transactions on Computer-Human Interaction* 7, 1 (2000), 114–138. https://doi.org/10.1145/344949.345077

[17] Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Communications of ACM* 50, 12 (2007), 20–32. https://doi.org/10.1145/1323688.1323689

[18] Ben Shneiderman. 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems* 10, 4 (2020), 26:1–26:31. https://doi.org/10.1145/3419764

[19] Jerry O. Talton, Daniel Gibson, Lingfeng Yang, Pat Hanrahan, and Vladlen Koltun. 2009. Exploratory Modeling With Collaborative Design Spaces. *ACM Transactions on Graphics* 28, 5 (2009), 167:1–167:10. https://doi.org/10.1145/1618452.1618513

[20] Almar van der Stappen and Mathias Funk. 2021. Towards Guidelines for Designing Human-in-the-Loop Machine Training Interfaces. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. ACM, New York, NY, 514–519. https://doi.org/10.1145/3397481.3450668

[21] Austin P. Wright, Zijie J. Wang, Haekyu Park, Grace Guo, Fabian Sperrle, Mennatallah El-Assady, Alex Endert, Daniel A. Keim, and Duen Horng Chau. 2020. A Comparative Analysis of Industry Human-AI Interaction Guidelines. *arXiv* 2010.11761 (2020), 8 pages. https://doi.org/10.48550/arXiv.2010.11761

[22] Hiromu Yakura. 2021. No More Handshaking: How have COVID-19 pushed the expansion of computer-mediated communication in Japanese idol culture?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 645:1–645:10. https://doi.org/10.1145/3411764.3445252

[23] Hiromu Yakura, Youhei Akimoto, and Jun Sakuma. 2020. Generate (Non-Software) Bugs to Fool Classifiers. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. AAAI Press, Palo Alto, CA, 1070–1078.

[24] Hiromu Yakura and Masataka Goto. 2020. Enhancing Participation Experience in VR Live Concerts by Improving Motions of Virtual Audience Avatars. In *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Washington, DC, 555–565. https://doi.org/10.1109/ISMAR50242.2020.00083

[25] Hiromu Yakura, Yuki Koyama, and Masataka Goto. 2021. Tool- and Domain-Agnostic Parameterization of Style Transfer Effects Leveraging Pretrained Perceptual Metrics. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. ijcai.org, Somerset, NJ, 1208–1216. https://doi.org/10.24963/ijcai.2021/167

[26] Hiromu Yakura, Tomoyasu Nakano, and Masataka Goto. 2018. FocusMusicRecommender: A System for Recommending Music to Listen to While Working. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. ACM, New York, NY, 7–17. https://doi.org/10.1145/3172944.3172981

[27] Hiromu Yakura, Tomoyasu Nakano, and Masataka Goto. 2022. An Automated System Recommending Background Music to Listen to While Working. *User Modeling and User-Adapted Interaction* 32, 3 (2022), 355–388. https://doi.org/10.1007/s11257-022-09325-y

[28] Hiromu Yakura and Jun Sakuma. 2019. Robust Audio Adversarial Example for a Physical Attack. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. ijcai.org, Somerset, NJ, 5334–5341. https://doi.org/10.24963/ijcai.2019/741

[29] Hiromu Yakura, Shinnosuke Shinozaki, Reon Nishimura, Yoshihiro Oyama, and Jun Sakuma. 2018. Malware Analysis of Imaged Binary Samples by Convolutional Neural Network with Attention Mechanism. In *Proceedings of the 8th ACM Conference on Data and Application Security and Privacy*. ACM, New York, NY, 127–134. https://doi.org/10.1145/3176258.3176335

[30] Hiromu Yakura, Shinnosuke Shinozaki, Reon Nishimura, Yoshihiro Oyama, and Jun Sakuma. 2019. Neural Malware Analysis with Attention Mechanism. *Computers & Security* 87, 101592 (2019), 1–15. https://doi.org/10.1016/j.cose.2019.101592

[31] Hiromu Yakura, Kento Watanabe, and Masataka Goto. 2022. Self-Supervised Contrastive Learning for Singing Voices. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), 1614–1623. https://doi.org/10.1109/TASLP.2022.3169627

[32] Kenji Yamanishi, Jun'ichi Takeuchi, Graham J. Williams, and Peter Milne. 2004. On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms. *Data Mining and Knowledge Discovery* 8, 3 (2004), 275–300. https://doi.org/10.1023/B:DAMI.0000023676.72185.7c