Towards a Human-Centric Design Framework for AI Assisted Music Production

Augoustinos Tsiros, Alessandro Palladini
Machine Learning Innovation Centre
Music Tribe
5 Brindley Road
Manchester, United Kingdom
{augoustinos.tsiros, allesandro.palladini}@musictribe.com

ABSTRACT
In this paper, we contribute to the discussion on how to best design human-centric MIR tools for live audio mixing by bridging the gap between research on complex systems, the psychology of automation and the design of tools that support creativity in music production. We present the design of the Channel-AI, an embedded AI system which performs instrument recognition and generates parameter settings suggestions for gain levels, gating, compression and equalization which are specific to the input signal and the instrument type. We discuss what we believe to be the key design principles and perspectives on the making of intelligent tools for creativity and for experts in the loop. We demonstrate how these principles have been applied to inform the design of the interaction between expert live audio mixing engineers with the Channel-AI (i.e. a corpus of AI features embedded in the Midas HD Console. We report the findings from a preliminary evaluation we conducted with three professional mixing engineers and reflect on mixing engineers' comments about the Channel-AI on social media.

Author Keywords
Audio Mixing, Human-AI Interaction, MIR Systems, Design Frameworks

CCS Concepts
• **Applied computing** → **Sound and music computing**; • **Information systems** → *Music retrieval*; **Human-centered computing** → **Interactive systems and tools**;

1. INTRODUCTION
Intelligent automated systems are no longer a utopian vision of the future, but rather our new reality. In a world where algorithmic intelligence is rapidly penetrating all areas of our life, humans are often left to wonder about their role in this emerging ecosystem. This is particularly relevant when AI technologies are deployed in products designed to automate tasks that were previously carried out primarily by human experts and require the utilisation of creativity, intuition and subjective judgement such as in disciplines related to creative design. Creativity is often thought to be a distinct feature of human intelligence, e.g. painting, writing, music and architecture. Because of the inherent subjectivity involved in these creative activities, AI tools that aim to automate these processes are often received with a degree of skepticism by the target users, i.e. designers and artists. According to the literature, the skepticism stems primarily from the following factors: i) practitioners' fear of being replaced by these new tools [1], [2], ii) a sense of doubt that a machine can perform as well as a trained human on subjective tasks, [3]–[5], and iii) distrust of AI recommendations resulting from the lack of understanding of the underlying reasoning that underpins the models' outputs, [6]–[8].

This creates a considerable challenge for designers and developers of AI technologies for creative applications. Since, they have to overcome three major obstacles, i.e. develop AI tools that will be able to match the performance of expert creative practitioners, design tools that seamlessly blend in existing workflows of the practitioners and identify interaction models that will cultivate trust. All of these factors are critical in the success of the system in the marketplace as well as the acceptance by the end user. In this paper, we discuss the conceptual framework which we have appropriated in order to deal with the aforementioned issues. We highlight these issues drawing on a series of practical examples from our experience in developing AI features for assisting live audio mixing engineers, and through reflection on mixing engineers' feedback about the Channel-AI from interviews and comments posted on social media. Furthermore, we present the design of the Channel-AI (a corpus of embedded AI features of the HD audio mixing console). Finally, through engagement with relevant literature we discuss the design principles which we apply to validate our design choices and understand where we stand in terms of level of automation, trust calibration and feedback provision.

2. BACKGROUND & PREVIOUS WORK
Advances in machine learning techniques have led to the development of creative systems that offer higher levels of automation for music creation, production and consumption. These changes are redefining the relationship between audio practitioners and creative systems. While prior to the machine learning era, creative systems were viewed as passive tools the intervention of machine learning, content analysis and processing capabilities are increasingly shifting our perception of these tools from being passive to being active. These changes have led human perceive creative systems more like collaborators capable of analyzing, interpreting the underlying structure of sound and music and provide suggestions as well as implement creative solutions.

In the context of music creation, a number of new tools have been developed that offer the aforementioned capabilities. For instance, Folk-rnn is a corpus of statistical models for algorithmic composition that utilises machine learning approaches to generate variations of input note sequences [9]. There is a growing interest in the application of machine learning for music creation and processing [10], [11] and the industry has begun to engage in research and development activities related to this area, e.g. Magenta at Google and CTRL (Creator Technology Research Lab) at Spotify. Similarly, in the context of music production many tools are available in the marketplace

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

that utilise machine learning for computer assisted audio processing: iZotope's Neutron a system for audio mixing that utilises machine learning for instrument recognition. Sonible's Smart EQ performs audio analysis and generates custom equalisation curves to achieve tonal balance. LANDR, CloudBounce are both cloud based mastering services that aim to automate decisions that traditionally were made by a mastering engineer by utilising machine learning to control signal processing parameters of various DSP processes.

Online services for music mastering have been extensively criticized [12]–[14]. In our opinion these negative reactions derive primarily from the failure to address the aforementioned design and performance challenges. For instance, LANDR has been criticised of trying to replace human labor rather than to augment it [14]. LANDR would have received less criticism if its design supported existing mastering engineering workflows and assisted users to achieve well mastered track while the process through which the mastering was achieved was more transparent. A more transparent approach in the case of LANDR would be to expose the parameters of the digital processing chain used for mastering the music tracks to the users and allow users to override the recommended parameters until they are satisfied with the results. The criticism directed towards LANDR from both mastering engineers and the research community confirms that a more human centric approach to the design of ML tools is needed. We need to design considerate creative systems that leverage both human creativity and listening skills rather than precludes them.

The shift in the paradigm from content agnostic to content aware creative systems which offer higher automation has raised several important questions. First, how to best approach the design and evaluation of such creative systems from an engineering perspective. Second, how to best design the interaction between the user and the intelligent audio or music software in order to foster creativity, trust and maintain the human in the loop by encouraging co-exploration and co-creation [15]. Third, how to approach the evaluation of the output of the system as this relate to computational creativity. The research presented in this paper engages primarily with the first two questions. We propose two frameworks that are considerate of the human in the loop paradigm and help MIR system developers and designers guide the development process.

3. PROBLEM DEFINITION

We have identified four main application domains in which users are likely to utilise the Channel-AI: music concerts, broadcasting, theatre and church. In the context of live sound there is zero tolerance for errors. This is due to fact that an error, could potentially damage the reputation of the mixing engineers involved in a live event, but also due to the health and safety issue that could arise from violating loudness specifications and compliance standards. Moreover, the mixing consoles in which the AI tools are embedded, are often used in very high-profile venues were the stakes are high for the engineers, the organisers and the artists. Hence, users are very quality conscious and conservative in incorporating new tools in their workflow. Moreover, the users are domain experts and have invested a lot of time to master the tools they use; hence they can be dismissive of a tool that could perform the task they have invested so many years to master. Additionally, they have well established workflows and are reluctant to changing their current practices. Below follow comments by mixing engineers we collated from Facebook and Twitter that highlight this problem:

*"Just saw the AI thing. What a waste of R&D. You are allowing the console to do my job based on what? This sh**t does not belong in a pro-desk. Can't we use our costly R&D on important things we have to deal with every day that may make the console even more useful. I personally don't know anyone that spends $35K and up on a console and uses EQ suggestions from the console."*

A comment by another mixing engineer: *"All this AI thing is going to do is produce a crop of lousy engineers who have no idea how to achieve a desired result, or indeed , know if the AI results actually sounds good, preferring to simply BELIEVE that it does since the AI said so. Intentionally or not, this will further cut the ears out of the equation."*

"*Has anyone involved in the AI design actually mixed a few thousand shows?"*

As the comments above suggest, some people have a negative predisposition against automation before they even try the technology and the assistance these tools can provide. These comments could be attributed to the reluctance to accept the concept that computers can be creative and produce results that are comparable to that of a professional audio practitioner. However, we also received many positive comments, which actually suggest that other mixing engineers are keen to embrace automation:

"*Pro touring guy here...pretty excited for the AI. It can intelligently gate a vocal and reduce bleed. When I tried this on the Midas HD console and it worked I was so happy."*

Another engineer said: "*This is the stuff of dreams for engineers walking up to a board with a 3-minute soundcheck."*

"This is the new channel presets library that most other consoles offer - except this is dynamic and learning and adaptive."

Furthermore, there is a strong element of subjectivity in the practice and preferences of each individual mixing engineer (e.g. some might set trim levels very loud while others prefer to allow a lot of headroom, these decisions might also be genre depended). Hence if the AI outputs diverge from their current practices, automated systems such as the Channel-AI could easily be dismissed. These creates considerable interaction design challenges for the adoption of intelligent automated systems, since they should be carefully designed so that they can co-exist harmoniously with the existing user workflows, without getting in the way. It is critical to provide the right level of automation and feedback, so that experienced users maintain control over the process and the audio output. In the following two sections, we will present the Channel-AI and discuss the design framework that we have been using to make sure we are not violating these requirements.

4. SYSTEM ARCHITECTURE

The Channel-AI is an evolving corpus of embedded AI features that live alongside the existing software and hardware of the Midas HD console. The Channel-AI aims to support mixing engineers' workflow by offering a number of unique MIR features including: i) instrument recognition, ii) provision of adaptive settings for audio levels, equalisation, gating, and dynamics processing. Figure 1, provides an overview of the system architecture. The Channel-AI can be invoked by the user only when needed. Users decide when they want to capture an audio buffer. Feature extraction is performed on the audio buffer. The data extracted is fed into a classifier which returns the detected instrument. The system provides feedback regarding the quality of the sampled buffer, i.e. source vs spill and also indicate the level of confidence regarding the detected instrument class. The user is required to select whether to accept the classifiers output or not. Up to four buffers can be captured, invoked and deleted by the user per audio channel. A series of adaptive 'presets' (i.e. signal aware suggestions of parameter values) for trim, phase, eq, compressor and gate are returned taking into consideration the instrument type and analysis data of the captured audio signal. The mixing engineer can audition each preset and select which ones they wish to apply. Preset values can be overridden by the user. An Auto-Setup functionality is also provided that allows to apply all recommended audio presets on a channel at once. For a video demonstrating the different functions of the system please see [1].

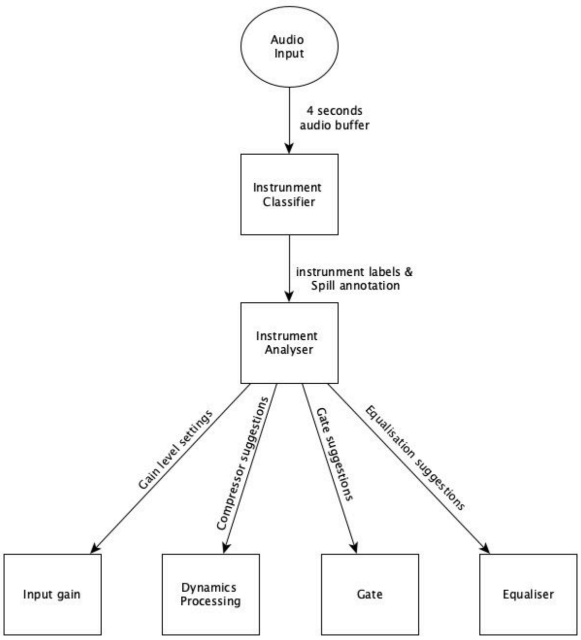

Figure 1. Block diagram shows an overview of the system architecture.

5. DESIGN FRAMEWORK

Given that our tools are to be used mainly by highly skilled mixing engineers, we consider of paramount importance that the human-machine cooperation is balanced in terms of the level of user control, type and level of automation. After reviewing a large body of literature related to human AI interaction and automation, a set of principles were collated and applied to guide the design of the Channel-AI and ensure that the models of interaction adopted mitigate the following risks which the authors believe could impede the adoption and utilisation of the system:

- Automation making undesired, suboptimal, and non-rectifiable decisions.
- Removing engineers' authority and control to do their jobs in the best way they see fit.
- Forcing users to radically change existing workflows.

According to the literature, human-machine cooperation must be carefully designed to allow mutual responsibility, authority and autonomy [16]. A growing body of literature suggests that trust is an important component to enable two agents to co-operate effectively [17]–[21] this is true for both human-human and human-AI collaboration. Moreover, trust calibration, i.e. the degree to which a person trusts another agent is also an extremely important variable that should be taken into account when designing human-AI interaction. A mismatch between the system capabilities and the users' level of trust can lead to under-trusting or over-trusting the system. According to [17], inappropriate reliance on a system can lead to problems. For instance, if the trust exceeds the capabilities of the automated system it can lead to misuse, i.e. delegate tasks that the system is not able to perform. While when trust is lower that the capabilities of the system it can lead to disuse, i.e. underutilisation of the features of the system. According to the literature other important factors for developing trust between the user and the system include feedback provision, indication of how confident the system is about the validity of its outputs [22], users' understanding of how an AI reaches a particular conclusion and why it has reached it. The provision of appropriate feedback is important to facilitate user understanding, to provide justification and enable user control [23].

To understand where the Channel-AI stands in terms of human-machine cooperation and help our team identify risks and plan future system development, we performed a system analysis using the stage model suggested by [24]. The model consists of four functions that can be performed by either a human or an intelligent automation system, these are: i) Information Acquisition, ii) Information Analysis, iii) Decision and Action Selection, and iv) Action Implementation. Most complex MIR systems consist of many layers and each layer can exhibit different levels of automation, ranging from: No Automation to Complete Automation [25]. We combined models suggested by [24], [25] to analyse what level and type of automation each of the different components of the Channel-AI, as shown in Table 1. Below follows a set of examples that demonstrates how the different levels of automation apply to the interaction between an audio mixing engineering and the automated features offered by the Channel-AI.

0. **No Automation:** the mixing engineer is completely in control of the console and the AI system does not interfere in decision making. For instance, deciding when to take an audio buffer for analysis or applying the adaptive presets can only be triggered by the user.
1. **Assistance:** the mixing engineer is in control of the creative system, but the system can control some parameters such as reduce gain levels if the audio repeatedly exceeds 0 dB.
2. **Partial Automation:** the mixing engineer needs to take over control only when corrections are needed. The AI classifies an instrument under one of the categories. The mixing engineer needs to monitor if the instrument has been classified correctly and correct the classifier label if it is incorrect.
3. **Conditional Automation:** the AI is completely in control given that certain conditions are met. When the instrument has been correctly classified, there is no need for the mixing engineer to intervene. The AI prompts the mixing engineer to intervene only when required; i.e. either when the AI cannot recognize the instrument, or the classifiers confidence measures are low indicating potentially erroneous classification.
4. **High Automation:** the AI is in complete control and the mixing engineer is not needed, but the AI can abort certain function under specific conditions; e.g. the AI has been set up to apply a particular corrective preset such as apply compression make up gain equal to the gain reduction resulting by the threshold and ratio of

[1]https://drive.google.com/open?id=17bBWLfrgmsDHPqaDY_rCOj23ifVbzhDI

the compressors. The AI should decide whether the make-up gain should be aborted or set to lower level if it is going to cause the signal to distort.

5. **Total Automation:** AI makes all decisions and no input by the mixing engineer is needed. For instance, the AI based on the analysis of the input signal that has been captured decides the setting of the input signal, the user can't intervene in this decision, except for overriding the value retrospectively.

Additionally, to aid the interaction design process of the AI-System, we identified a set of design principles (derived from, see [26]) and applied them to further optimise the human-AI interaction and evaluate our current design. A total of 18 design principles are proposed by [26], these are grouped into 4 categories: Initial; During Interaction; When Wrong; Over Time. We only utilized the first 11 principles since the remaining seven principles apply only to AI systems that implement interactive machine learning techniques such as reinforcement learning. We utilised these principles to ensure appropriate trust calibration (principles 1, 2, 5), feedback provision (principles 3, 4), maximisation of user control and minimisation of disruption to current workflows (principles 6-11). Using these principles at both the conceptualisation and evaluation phases has proven very rewarding since it led to a major redesign of the interface and the workflow, a more detailed explanation of the principles used and how they map to the user interface, see Figure 2

Table 1. System analysis showing different the levels of automation of the Channel-AI features across the four functions.

Automation Levels and Interactions	**Info Acquisition**	**Info Analysis**	**Decision Selection**	**Action Implementation**
	0 - Channel Selection	3 - Instrument Detection	0 - Profile Selection	0 - Name Channel
	0 - Profile Creation	3 - Audio Analysis	0 - Profile Retention	0 - Invoke Auto-setup
		5 - Feature Extraction	3 - Channel Name	0 - Audition Settings
		5 - Noise Detection	3 - Setting Generation	0 - Apply Settings
		5 - Quality Assessment	3 - Setting Selection	0 - Override Settings
AI Capabilities		Bayesian Inference Instrument Detection Unsupervised Audio Analysis	Settings Generation Channels Comparison	

0 = No Automation, 3 = Conditional Automation, 5 = Total Automation

Purpose

1) Explain the purpose of the AI.
2) Make clear what the system can do and how well it can do it.
3) Show the performance of the system choosing appropriate feedback strategies.
4) Show when the system is not confident.
5) Design for appropriate trust, not for higher trust.

Interaction

6) Minimize the impact on the existing workflow
7) Support efficient invocation.
8) Support efficient correction.
9) Support efficient dismissal.
10) Make the level of automation adaptable.
11) Design clear transitions between the different levels of automation.

Figure 2. Design guidelines applied to the Channel AI.

6. INTERVIEWS WITH PROFESSIONAL MIXING ENGINEERS

We report on interviews held with three professional audio mixing engineers to analyse how the Channel-AI influences their workflow and discuss the feedback we received on this version of the software tool.

6.1 Interview setup

6.1.1 Participants

We held individual interviews with three high profile live mixing engineers. One participant had experience with an earlier version of the Channel-AI that was being tested, while the other two had no previous experience. We aimed to gather feedback on the effectiveness of the adaptive preset suggestion for Gain levels, Compression, Gating and Equalisation and obtain insights on the influence of the MIR tool on the participants' workflow. We also hoped to collect information on ways to improve the current implementation of the AI-System.

6.1.2 Study Design

Each mixing engineer was given a demonstration of all of the features of the Channel-AI. During the demonstration we ensured that the participants understood the functionality of the features and answered any questions they may had in relation to functionality of the system. The engineers had been instructed prior to the interview to bring with them a multitrack recording they had recently mixed and felt comfortable with. We believe that asking the participants to test the Channel-AI using multitrack recordings of a song that they are familiar with, and consequently have a reference point of the setting configuration they had already applied, would be better from an ecological validity point of view, rather than asking them to mix a multitrack recording they never heard before. After the demonstration of the system, the engineer was instructed to mix the multitrack recordings utilising the functions of the Channel-AI. Interviewees were told that they could diverge if they wanted from the system suggestions and that they could take as long as they needed until they achieved a satisfactory mix. After the mixing session, a semi-structured interview was conducted to elicit information regarding the engineers' experience of the interaction with the system and receive feedback on the workflow and the parameter settings suggestions offered by the adaptive presets. Finally, they were asked to make suggestions for improvement of current functionality and propose other features that would be useful.

6.1.3 Findings

Regarding the integration of the Channel-AI on existing user workflows, participants seemed to value the capabilities offered by the system but two out of three felt that the workflow imposed by the current system could become much faster if instead of sampling and analysing each channel separately a bulk analysis and set-up function was available. Regarding the Auto-setup function, which applies Gain, EQ, Gate and Compression settings with a single touch of a button, one engineer suggested that it would be useful to have a configuration page that allows the engineer to decide which settings and DSP processes to apply and which ones not to apply, e.g. configure the system so that it applies settings only for audio equalisation and gating. The engineer made this suggestion on the basis that he does not use compression to tame the dynamics of the sound, but he only uses dynamics processing as a limiter, i.e. to make sure the signal does not clip under any circumstances. Currently the Channel-AI provides several presets for each channel, e.g. on average the system returns five equalisation settings suggestions per channel. One of the participants thought that this could potentially increase set up time, because out of curiosity he will need to listen to each adaptive preset. However, a more optimistic view regarding the presets which another participant expressed, is that more setting configurations could be evaluated per instrument that would have been possible through manual adjustment.

Regarding suggestions of gain levels, we found that there are two divergent views on how mixing engineers approach this step. The system at the time we performed this study would suggest setting the gain of the input signal close but below 0dB. A common approach is to set all faders on the console to 0dB and then adjust the trim levels until the instrument is loud enough in the mix (i.e. performing gain staging from the trim, pre-fader). The second approach involves adjusting the gain level to get the loudest possible signal while avoiding clipping and then proceed with the mix by adjusting the levels using the faders. One of the engineers who follows the first approach (i.e. setting all faders at around 0dB) thought that the gain was set to a high level for the way he works. The other two though that the gain levels were good. One of the participants correctly suggested that the levels should be adjusted based on loudness and not on based on the RMS or decibel scales of the input signal. In more recent releases of the software, gain levels are adjusted based on the equal loudness measures. We also found that there were differences in the approach to adjusting the levels between front of house engineers and monitor engineers, with monitor engineers be more inclined to use the first approach in order to provide consistency to the level of the musicians on stage.

Gate settings were considered to be very effective at dealing with spill, i.e. the phenomenon whereby sound is picked up by a microphone from a source other than the main instrument. The only suggestion for improvement we received from the engineers was that the release parameter of the gate was consistently set too short which sounded unnatural as it trimmed natural resonances of the signal. Engineers also thought that compression settings and equalisation setting were appropriate, except from one of the interviewees who does not apply make-up gain since he does not use dynamics processing to reduce the dynamic range of the signal. Regarding the EQ suggestions, two of the engineers said that the recommended settings sounded interesting although in some instances they would not have thought of setting the EQ frequencies and gains this way.

Although we received very few negative comments regarding the performance of the system and the adaptive presets, there were a few interesting outcomes. First, we found that despite the few negative comments which we received, the engineers who participated in the study seemed to be sceptical when asked if they would use the system. Our understanding is that scepticism stems primarily from conservativism. For instance, one of the participants said he would not use the adaptive presets as he stated: *"I have my ways of doing things, but the kids will love it"*. Second, we realised that it can be difficult to conduct rigorous and structured testing session with expert mixing engineers as they seem to prefer to guide the exploration. Also, it is very important to explain the purpose and context of use of the features which are being tested thoroughly in order to avoid misunderstandings that can waste both the participants' and the researchers' time. For instance, when we asked our betta tester to test some of the new features we had added to the console in-situ (in the field), they ended up using the technology in ways that was not designed for and the results of the testing session was uninformative. All of the engineers seemed to agree that speed is key to the work they do, and they would not be interested to use a technology that would slow them down. They would prefer higher level of automation that they can trust than having to review many options. Additionally, we noted that many of the new feature requests which we received by the participants were related to usability and user experience.

7. CONCLUSSIONS

The best approach to the design and evaluation of human-centric AI for expert users is an open question for debate. While research communities such as music informatics and computer music have made huge progress in the development of new tools and methods for incorporating information retrieval and machine learning techniques, the human factors have by and large received less attention in these subject areas. We reported on the feedback we received from users online and also reported on the findings from interviews we held with three professional mixing engineers. Our findings highlighted the issues that could potentially arise in designing automated systems that support professional audio practitioners. We observed that although domain experts valued the assistance of the system, they were skeptical when asked about their willingness to adopt intelligent automated systems in their workflows. The results from the interviews suggest that usability and set-up time are very important when designing tools for live mixing engineers. This led us to the conclusion that finding the right balance between automation and user control is of paramount importance for the adoption of automated music production systems by domain experts. Hence, there is need to devise conceptual and design frameworks that are considerate of the human in the loop paradigm and help intelligent music production system developers and designers guide the development process.

In this paper we presented a method for analysing AI assisted music production systems, which in our experience is useful in both the design and evaluation phases. The proposed system analysis method aims to aid developers analyse complex system which consist of multiple MIR and ML components and incorporate different levels of automation in the different components of the system. This method is extremely useful since it allows to consider the system architecture in tandem with the human factors and optimise the human-AI cooperation prior to the development of the system. Moreover, this could be used for evaluation of existing systems.

Finally, we applied a set of design guidelines, which we demonstrated are well suited for designing human-centric intelligent music production systems. We showed how these principles where applied to guide the design of the Channel-AI and ensure that the human-machine cooperation is balanced, does not remove engineers' authority and level of control as well as enables the co-existence of the automation alongside the current user workflows. Future work includes further refinement of the two frameworks through evaluation and feedback by the users in controlled usability studies which we will be conducting in the near future.

8. REFERENCES

[1] A. Campolo, M. Sanfilippo, M. Whittaker, and K. Crawford, "AI Now: 2017 Report," New York, 2017.

[2] B. L. T. Sturm, M. Iglesias, O. Ben-Tal, M. Miron, and E. Gómez, "Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis," *Arts*, vol. 8, no. 3, p. 115, 2019.

[3] A. Jordanous, "Evaluating Evaluation: Assessing Progress and Practices in Computational Creativity Research," in *Computational Synthesis and Creative Systems.The Philosophy and Engineering of Autonomously Creative Systems*, T. Veale and F. A. Cardoso, Eds. Springer, 2019, pp. 211–232.

[4] G. Ritchie, "The Evaluation of Creative Systems," in *Computational Synthesis and Creative Systems.The Philosophy and Engineering of Autonomously Creative Systems*, T. Veale and F. A. Cardoso, Eds. Springer, 2019, pp. 159–194.

[5] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behav. Brain Sci.*, vol. 40, no. 2012, pp. 1–58, 2017.

[6] V. Dominguez, P. Messina, I. Donoso-Guzmán, and D. Parra, "The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images," *Proc. 24th Int. Conf. Intell. User Interfaces - IUI '19*, pp. 408–416, 2019.

[7] M. Ribera and A. Lapedriza, "Can we do better explanations? A proposal of user-centered explainable AI," *CEUR Workshop Proc.*, vol. 2327, 2019.

[8] Q. Roy, F. Zhang, and D. Vogel, "Automation accuracy is good, but high controllability may be better," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–8, 2019.

[9] B. L. Sturm, J. F. Santos, O. Ben-Tal, and I. Korshunova, "Music transcription modelling and composition using deep learning," in *1st Conference on Computer Simulation of Musical Creativity*, 2016.

[10] J. Nika, K. Déguernel, A. Chemla-Romeu-Santos, E. Vincent, and G. Assayag, "DYCI2 agents: Merging the 'free', 'reactive', and 'scenario-based' music generation paradigms," in *43rd International Computer Music Conference*, 2017, pp. 227–232.

[11] J. P. Briot and F. Pachet, "Deep learning for music generation: challenges and directions," *arXiv Prepr. arXiv1712.04371.*, pp. 1–17, 2018.

[12] J. Sterne and E. Razlogova, "Machine Learning in Context, or Learning from LANDR: Artificial Intelligence and the Platformization of Music Mastering," *Soc. Media + Soc.*, vol. 5, no. 2, p. 205630511984752, 2019.

[13] S. Inglis, "LANDR, CloudBounce & the Future of Mastering.," in *Sound on Sound*, 20016.

[14] T. Birtchnell, "Listening without ears: Artificial intelligence in audio mastering," *Big Data Soc.*, vol. 5, no. 2, p. 205395171880855, 2018.

[15] H. Scurto *et al.*, "Appropriating Music Computing Practices Through Human-AI Collaboration," *Journées d'Informatique Music.*, 2018.

[16] C. A. Miller and R. Parasuraman, "Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control," *Hum. Factors*, vol. 49, no. 1, pp. 57–75, 2007.

[17] J. D. Lee, K. A. See, and I. City, "Trust in Automation: Designing for Appropriate Reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[18] M. Körber, "Theoretical considerations and development of a questionnaire to measure trust in automation," in *20th Triennial Congress of the IEA. Springer*, 2018, pp. 13–30.

[19] M. Yin, J. W. Vaughan, and H. Wallach, "Understanding the effect of accuracy on trust in machine learning models," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2019.

[20] C. Castelfranchi and R. Falcone, "Trust and control: A dialectic link," *Appl. Artif. Intell.*, vol. 14, no. 8, pp. 799–823, 2000.

[21] K. Yu, S. Berkovsky, D. Conway, R. Taib, J. Chen, and F. Zhou, "Do I Trust a Machine? Differences in User Trust Based onSystemPerformance," in *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, Human-Comp., J. Zhou and F. Chen, Eds. Springer, 2018, pp. 245–264.

[22] S. A. Cambo and G. D., "User-Centred Evaluation for Machine Learning," in *Human and Machine Learning*, Human–Comp., J. Zhou and F. Chen, Eds. Springer, 2018, pp. 315–340.

[23] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[24] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst. Man, Cybern. Part A Syst. Humans.*, vol. 30, no. 3, pp. 286–297, 2000.

[25] SAE, "Automated Driving – Levels of Driving Automation are Defined in New SAE International Standard J3016," 2014.

[26] S. Amershi *et al.*, "Guidelines for Human-AI Interaction," *CHI Conf. Hum. Factors Comput. Syst.*, 2019.