Delft University of Technology

Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT

Vieira, Romulo; Muchaluat-Saade, Debora; Cesar, Pablo

DOI
10.1007/978-3-031-53302-0_17

Publication date
2024

Document Version
Final published version

Published in
MultiMedia Modeling - 30th International Conference, MMM 2024, Proceedings

Citation (APA)
Vieira, R., Muchaluat-Saade, D., & Cesar, P. (2024). Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT. In S. Rudinac, M. Worring, C. Liem, A. Hanjalic, B. P. Jónsson, Y. Yamakata, & B. Liu (Eds.), *MultiMedia Modeling - 30th International Conference, MMM 2024, Proceedings* (pp. 234-248). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 14557 LNCS). Springer. https://doi.org/10.1007/978-3-031-53302-0_17

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Check for updates

Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT

Romulo Vieira[1(✉)], Debora Muchaluat-Saade[1], and Pablo Cesar[2,3]

[1] MídiaCom Lab, Fluminense Federal University, Niterói, RJ, Brazil
{romulo_vieira,debora}@midiacom.uff.br

[2] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
garcia@cwi.nl

[3] TU Delft, Amsterdam, The Netherlands

Abstract. The Internet of Multisensory, Multimedia and Musical Things (Io3MT) is a new concept that arises from the confluence of several areas of computer science, arts, and humanities, with the objective of grouping in a single place devices and data that explore the five human senses, besides multimedia aspects and music content. In the context of this brave new idea paper, we advance the proposition of a theoretical alignment between the emerging domain in question and the field of Artificial Intelligence (AI). The main goal of this endeavor is to tentatively delineate the inceptive trends and conceivable consequences stemming from the fusion of these domains within the sphere of artistic presentations. Our comprehensive analysis spans a spectrum of dimensions, encompassing the automated generation of multimedia content, the real-time extraction of sensory effects, and post-performance analytical strategies. In this manner, artists are equipped with quantitative metrics that can be employed to enhance future artistic performances. We assert that this cooperative amalgamation has the potential to serve as a conduit for optimizing the creative capabilities of stakeholders.

Keywords: Internet of Multisensory · Multimedia and Musical Things · Artificial Intelligence · Networked Performances

1 Introduction

The Internet of Things (IoT) [5] denotes a collection of embedded systems imbued with sensing and/or actuation capabilities, which collaborate with neighboring devices to execute specific tasks. The recent instantiation of these concepts has engendered noteworthy progress in the development and enhancement of applications spanning diverse domains such as home automation, smart cities, industry 4.0, e-health, security, and geospatial analysis, among others. Nevertheless, artistic and musical applications have hitherto received relatively scant attention in comparison to the aforementioned domains [4,28,29]. The existing instances within this realm predominantly reflect the technological and artistic

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
S. Rudinac et al. (Eds.): MMM 2024, LNCS 14557, pp. 234–248, 2024.
https://doi.org/10.1007/978-3-031-53302-0_17

preferences of their creators, thereby impeding the seamless integration of diverse multimedia content and the utilization of the inherent advantages offered by their counterparts.

To alleviate these limitations and move computing as much as possible to combine these domains, we have articulated an area called the Internet of Multisensory, Multimedia and Musical Things (Io3MT) [32]. This field is envisioned to serve as a unified platform where these three distinct content types can be seamlessly integrated, rendering them interchangeable and devoid of hierarchical constraints. The theoretical underpinnings of this area draw upon a confluence of various interdisciplinary subjects, including but not limited to Networked Music Performance (NMP) [27], Multiple Sensorial Media (Mulsemedia) [16], Interactive Art [13], among others.

In light of these considerations, Io3MT holds the promise of exerting a substantial influence and offering valuable support in the development of a diverse array of entertainment applications, such as multisensory cinema, digital television, games, and interactive museums and galleries. It is noteworthy that the current research endeavors within this domain are primarily concentrated on networked artistic performances.

Artistic performances can be understood through various lenses. For instance, Ervin Goffman [18] categorizes them within the context of theatrical paradigms, viewing every action performed by an individual during an interaction as a form of performance. Alternatively, Peter Berger and Thomas Luckmann [7] propose an interactive construction perspective, suggesting that these actions are not bound by predefined social, cultural, or natural prescriptions. From a practical and technical standpoint, a performance is considered the culmination of extensive theoretical and practical preparation. In this context, elements such as repertoire, arrangement, and gestures come together to materialize as auditory, aesthetic expressions, dances, movements, orality, vocality, and narratives, whether they are presented in recorded or live formats. These performances enable the audience to experience visual and interactive forms of expression.

Traditionally, artistic performance has always incorporated intertextual elements, featuring an inherent heterogeneity that often remains implicit in discourse. In the context of contemporary society's pervasive technological influence, new dimensions of the performative act have emerged, including the utilization of Artificial Intelligence (AI) as a tool, medium, mediator, or partner [9]. In response to this evolving landscape, this paper endeavors to provide an exploratory examination of the techniques, potentialities, and open challenges associated with the incorporation of AI within performances guided by the principles of Io3MT. This discussion can be used as a start point to build Io3MT environments using real-time AI techniques to enhance multisensory, multimedia and musical performances.

The remainder of the paper is divided as follows. Section 2 examines prior literature addressing the usage of AI in the generation of components essential to artistic performances. This encompasses areas such as AI-driven musical composition for accompaniment, the generation of lyrics, and choreographic movements aimed at enhancing the overall presentation. In Sect. 3, we expound upon the fundamental architectural and functional tenets underpinning Io3MT. Following this, Sect. 4 engages in a discourse regarding the ramifications arising from the intersection of these two domains of knowledge. Finally, Sect. 5 offers a succinct summation of the key findings and conclusions derived from this research endeavor.

2 Related Work

Artificial intelligence tools, particularly those of a generative nature, have witnessed extensive utilization for expediting and enhancing the production of diverse forms of creative content, encompassing domains such as music, animations, storytelling, and visual arts. Within this context, the ensuing section is dedicated to examining scholarly contributions that leverage these AI techniques to generate elements that substantially contribute to the realization and enrichment of artistic performances.

A current area of prominence within the domain of music composition is Generative Music, which is also referred to as Algorithmic Composition or Musical Metacreation [10]. Generative Music entails the creation of musical pieces facilitated by computational algorithms or autonomous systems guided by predetermined rules, parameters, and algorithms. It affords the capacity to generate compositions that dynamically evolve over time, adapting to different variables including environmental factors, audience interaction, or stochastic occurrences. Consequently, this approach has the potential to yield music compositions characterized by uniqueness in each performance iteration.

An example in this context is the Mind Band [26] initiative, a cross-media platform harnessing artificial intelligence to craft musical compositions by interpreting emojis, images, and vocalizations. This system operates on the foundation of the valence-arousal model, which translates emotional cues identified within user-contributed symbols or vocal inputs into musical compositions. The generation of these musical pieces is facilitated by a Variational Autoencoder-Generative Adversarial Network (VAE-GAN) architecture [11].

Their model consists of three main parts: input, music generation, and output. The first part receives multimedia information for transformation into music. The second part, at the core of the project, generates the music content using a combination of Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) techniques, where the VAE decoder also serves as the GAN generator. Those techniques are trained together with shared parameters, and element-wise losses apply to both the encoder and decoder, while the discriminator loss also involves the encoder. After that iterative process, the third phase produces various musical compositions as outputs.

An additional instance of a tool devised for algorithmic composition is exemplified by XiaoIce Band [34], designed to address concerns related to the assurance of harmonic coherence within multichannel musical compositions. In pursuit of this objective, the authors have introduced a model capable of producing melodies and orchestrations for various musical instruments while upholding the integrity of chord progressions and rhythmic structures. In this endeavor, they have formulated a Cross Melody and Rhythm Generation Model (CMRGM) and a Multi-Instrument Co-Arrangement Model (MICA) through the application of multi-task learning principles to both components.

Similarly, the MuseGAN project [12] aligns itself with the objective of generating multichannel music compositions. However, it diverges from the previous approach by replacing multi-task learning with a methodology grounded in GANs. Notably, MuseGAN introduces an objective evaluation framework that scrutinizes both the intra-track and inter-track characteristics of the generated musical compositions. A distinctive feature of this proposal is its capacity to extrapolate from a human-composed musical track provided as input. MuseGAN can autonomously generate four additional tracks to complement the original audio, thereby expanding its potential for active participation in a performance context.

Exploring further into works related to musical improvisation and accompaniment, it is worth mentioning the GEN JAM [11], which employs genetic algorithms to construct a computational model capable of improvising jazz-style compositions. Similarly, Papadopoulos and Wiggins' proposal [25] also utilizes genetic algorithms for the same purpose, albeit the difference of generating improvised melodies within a predefined chord progression. This approach includes the use of a fitness function to automatically assess the quality of improvisations, considering factors like melodic shape, note duration, and intervals between notes. It is worth noting that GEN JAM [11] relies on human evaluations to assess improvisation quality, unlike Papadopoulos and Wiggins' approach [25].

Two systems, DIRECTOR MUSICES [15] and Widmer's project [11], employ rules for tempo, dynamics, and articulation transformations in musical performances. DIRECTOR MUSICES follows a set of prescribed rules, including differentiation, grouping, and set rules, while Widmer's project achieves similar transformations through machine learning. In a similar way, the research conducted by Bresin [8] utilizes artificial neural networks, which go beyond conventional feed-forward networks by incorporating feedback connections for improved modeling of temporal and amplitude variations in musical output.

Artistic performances involve various aesthetic elements, including choreographed dance movements. AI techniques are utilized to create such choreography, exemplified by the Full-Attention Cross-modal Transformer (FACT) network [20]. FACT generates three-dimensional (3D) dance movements based on auditory input and initial motion encoding. It combines these inputs within a cross-modal network to capture the relationship between sound and motion. Given a musical piece and a brief initial motion sequence of about 2 s, FACT can generate lifelike dance sequences comprising joint rotations and translations. The model is trained to predict N upcoming motion sequences and, during testing, autonomously generates continuous motion, effectively learning the intricate connection between music and motion. This capability allows for diverse dance sequences tailored to different musical compositions.

AI-Lyricist [22] is a system capable of generating meaningful lyrics for melodies, receiving a set of words and a MIDI file as input. That task involves several challenges, including automatically identifying the melody and extracting a syllable template from multichannel music, generating creative lyrics that match the input music style and syllable alignment, and satisfying vocabulary constraints. To address those challenges, the system is divided into four modules: i) Music structure analyzer: to derive the music structure and syllable model from a given MIDI file; ii) Letter generator: based on Sequence Generative Adversarial Networks (SeqGAN) optimized by multi-adversarial training through policy gradients with twin discriminators for text quality and syllable alignment; iii) Deeply coupled lyric embedding model: to project music and lyrics into a joint space, enabling a fair comparison of melody and lyric constraints; and iv) Polisher module: to satisfy vocabulary restrictions, applying a mask to the generator and replacing the words to be learned. The model was trained on a dataset of over 7,000 song-lyric pairs, enhanced with manually annotated labels in terms of theme, sentiment, and genre.

In light of the aforementioned context, the primary goal of this study is to provide a theoretical and exploratory exposition of the key Artificial Intelligence (AI) methodologies applicable to artistic processes and events grounded in the principles of Io3MT. This investigation contemplates the specific facets that may be encompassed, including the automatic generation of musical and visual content, machine-facilitated musical accompaniment, the automated recommendation and extraction of sensory effects, the composition of lyrical content and choreographic movements. Moreover, this work also seeks to highlight the inherent challenges in these scenarios, and it delineates the forthcoming areas of research that we intend to pursue.

3 The Internet of Multisensory, Multimedia and Musical Things (Io3MT) Design

The Internet of Multisensory, Multimedia, and Musical Things (Io3MT) [32] represents an emergent domain of research characterized by the inclusion of heterogeneous objects capable of establishing connections to the Internet, thereby participating in the continual exchange of multisensory, multimedia, and musical data. The fundamental aim underlying this interconnectivity is to promote seamless interaction and the accrual of valuable information, consequently fostering the realization and automation of an array of services. These services span from those oriented towards artistic endeavors and entertainment applications to educational and therapeutic contexts. From a functional perspective, Io3MT seeks to streamline the integration and real-time communication among these interconnected devices. Furthermore, it endeavors to enable a specific class of data to either act upon or exert influence over another. For instance, this could entail scenarios where a sequence of musical chords induces alterations in the color palette of a visual display, or the exhibition of an image triggers the emission of scents reminiscent of the corresponding depicted environment.

Within this context, it becomes evident that Io3MT adheres to a horizontal framework with distinctive characteristics. These attributes include low latency, signifying the system's ability to swiftly report events within a predefined timeframe. Additionally, low jitter is crucial, indicating the acceptable variability in data packet arrival times. Reliability plays a pivotal role, underscoring the system's effectiveness in delivering data with minimal packet loss. Furthermore, Io3MT environments necessitate an adequate bandwidth to accommodate the diverse multimedia content they handle. The adoption of multi-path networking is vital, allowing the system to utilize multiple communication routes to mitigate overloading of a single path and reduce the risk of network failure. Alongside these aspects, fault tolerance, lightweight implementation, and efficient service coordination are essential components. These environments must possess the capability to handle and synchronize various signal types, spanning audio, video, images, sensory stimuli, and more. Moreover, they should facilitate the active participation of individuals with differing levels of musical expertise, thereby enriching and streamlining creative processes within such contexts.

As is customary in well-established IoT models and within various tools in the field of computer science, the adoption of a layered architecture is prevalent. This approach offers a significant advantage in ensuring that each layer's services are executed in accordance with the prerequisites established by the preceding layers. Typically, the lower the hierarchical layer, the more straightforward and efficient the services it provides, with increasingly intricate functionalities emerging as one ascends through the layers [14]. In line with this convention, the proposed architecture for Io3MT is structured into five distinct levels. The lowest of these layers is designated as the Things Layer. Functioning as the virtualization layer, it bears the responsibility of converting analog data originating from physical devices into digital formats. This transformation initiates the generation of substantial volumes of data that subsequently traverse throughout the Io3MT environment.

Subsequently, the Link Layer emerges with its primary objective being the facilitation of device connectivity to the overarching network. Devices integrated within this layer exhibit sensorial capabilities, processing data derived from the physical realm before transmitting it to centralized servers for further analysis and dissemination.

The third tier, referred to as the Network Layer, shoulders the responsibility for the foundational infrastructure upon which the Io3MT ecosystem is constructed. Within this layer, critical Quality of Service (QoS) parameters pertinent to the measurement of data transmission performance are taken into consideration. These parameters encompass attributes such as latency, bandwidth, packet loss, jitter, among others.

Next in line, the Middleware Layer assumes a pivotal role within the architectural framework. This level shoulders the critical responsibility of processing incoming data, making informed decisions, and providing essential services. Its multifaceted functions encompass tasks such as data translation to ensure compatibility across different formats, resource discovery to identify available assets, synchronization for temporal alignment, identity management for secure user and device identification, concurrency and transition control for managing simultaneous operations and state transitions, persistence for reliable data storage and retrieval, and stream processing to enable real-time analysis and manipulation of data streams to meet dynamic requirements and applications.

The fifth and ultimate stratum within the architectural framework is the Application Layer, charged with the responsibility of furnishing the services as solicited by customers or users. This layer can be further subcategorized into two distinct levels. The first of these is the service unit, tasked with offering functionalities pertaining to information management, data analysis, and decision-making tools. The second sub-division is the applications unit, which scrutinizes the services requested by end-users, subsequently processing data originating from the physical world into cyber manifestations to facilitate the accomplishment of specific tasks. Figure 1 presents a synthesis of this discussion, with the main technologies of each of these layers, illustrated for educational purposes.

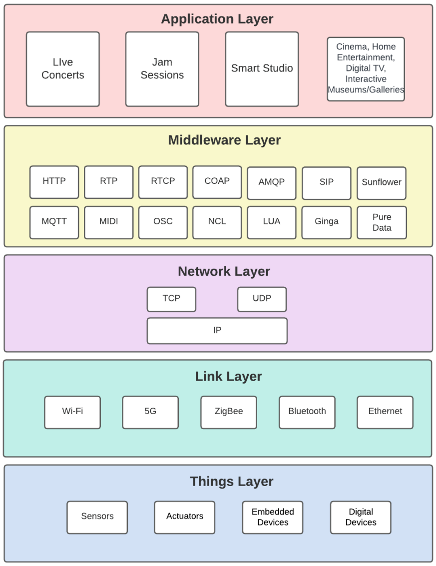

Fig. 1. Layered architecture of Io3MT.

Regarding artistic creation, the environment should be designed to deliver an immersive and captivating experience for the audience. It should also facilitate seamless integration and user participation, with resources accessible to those lacking artistic or computational expertise. To fulfill these objectives, several prerequisites must be met. Firstly, the presentation should be made more accessible to a wide range of users. Additionally, it should capture information from the audience and amalgamate it into a coherent artistic narrative, ensuring artistic security. Furthermore, the environment should encourage active participation, captivating the audience and fostering an environment that promotes initiation. Moreover, it should establish a clear and transparent relationship between participants' gestures and the outcomes achieved, providing a comprehensible and engaging experience. This approach allows the system to remain open and editable, enabling the combination of various elements and offering users opportunities to acquire and refine skills over time. Any imperfections within the system should not be attributed to functional errors but rather to potentially unsatisfactory results from a musical and aesthetic standpoint.

The constituent entities of these immersive environments are commonly referred to as multisensory, multimedia, and musical things (3MT). These objects are distinguished by their capability to engage in at least one of the specified actions, which may involve either generating or responding to content within these designated categories. They fulfill roles as providers of services and information, encompassing a diverse range of components, including hardware, software, sensors, actuators, cloud-based services, as well as electroacoustic and electronic musical instruments, along with smart devices. The structures of these devices may manifest either in physical or digital forms.

Moreover, these entities include embedded electronics, wireless communication, sensing, and actuation capabilities. They collaborate with other devices to perform tasks within project constraints. They require unique identification, scalability, persistence, reliability, and a loosely coupled architecture. Due to their artistic nature, these devices must prioritize aesthetics, expressiveness, and ergonomics. They should be user-friendly, flexible, efficient, and productive. Accessibility is vital for users with various abilities, and communicability is essential for conveying their purpose and design rationale while responding to user input [31].

The stakeholder can be categorized into six distinct groups. The initial category comprises musicians, performers, and artists who bear the responsibility of generating a diverse array of services. These services encompass a wide spectrum of creative expressions, spanning from musical compositions, live programming, dance performances, digital painting, graphic projections, 3D printing endeavors, and collaborative projects with robotic systems, among others. This category is inclusive of participants with varying levels of proficiency, encompassing beginners, amateurs, and experts, all actively engaged in the creation of multimedia art.

Subsequently, there are designers, entrusted with the task of conceiving interactive components encompassing electronic devices, software, and novel interfaces tailored for artistical expression. These professionals contribute to the design and development of innovative tools and interfaces that enhance the artistic experience. Following this group, there are the audience members, representing the end-users who consume the artistic content generated during presentations. These individuals actively engage with the system, either locally or remotely, facilitating the transmission and reception of multisensory and multimedia information. In doing so, they demonstrate the capacity to influence and modify the presented artistic outcomes through their interactions with the system.

Another significant category consists of members of the entertainment industry, comprising record labels, publishers, distributors, and manufacturers. These stakeholders are primarily concerned with the commercial aspects of applications within this domain, encompassing activities such as publishing, marketing, digital distribution, copyright management, product creation, and promotion, among others.

Lastly, the stakeholders encompass teachers and students who leverage the principles and theories derived from this field for educational purposes. They incorporate various disciplines related to science, technology, engineering, art, and mathematics (STEAM) into their educational curricula, utilizing insights from this domain to enhance learning and foster interdisciplinary engagement among students.

4 Discussion and Open Challenges

The preceding discussion highlights key AI techniques that enhance artistic performances. When considering the Io3MT layers, the Things Layer allows for real-time data analysis, aiding in equipment maintenance and gathering insights about the environment and audience reactions. This leads to adaptable environments and the potential for new artistic applications using computer vision and speech recognition, increasing device versatility [21]. The techniques that can be used in these tasks are as diverse as possible. The methodologies applicable to these endeavors encompass a wide spectrum, ranging from machine learning-based classification techniques that facilitate data categorization and label generation, to Recurrent Neural Networks (RNNs) adept at prognosticating failures. Additionally, signal processing algorithms are employed for noise filtration, feature extraction, and frequency analysis. Furthermore, Convolutional Neural Networks (CNNs) are leveraged for object recognition, motion detection, and the subsequent analysis and adaptation of content [17].

In the Link Layer, critical for establishing device connections and pre-processing physical data, integrating machine learning signal processing is crucial. These techniques assist in filtering, normalizing, pattern recognition, and anomaly detection, improving system efficiency. Deep learning methods like CNNs and RNNs are prominent for multimedia and time series data, but Isolation Forest, SVM for anomaly detection, and Reinforcement Learning for autonomous decisions are also valuable options [3,24].

In the Network Layer, responsible for data transmission and QoS management, the integration of AI offers several advantages. These include intelligent routing through Artificial Neural Networks (ANNs) and genetic algorithms, latency prediction via recurrent neural networks, bandwidth control with reinforcement learning and deep learning, jitter and packet loss detection using decision trees and clustering methods, optimization facilitated by genetic algorithms, and enhanced security, with the application of techniques like SVM, reinforcement learning, and genetic algorithms [19,33].

The Middleware Layer may be implemented by harnessing the capabilities of Natural Language Processing (NLP) for the translation or conversion of data characterized by distinct linguistic or semantic formats. In addition, machine learning techniques can be applied for the purpose of service recommendation and discovery, drawing upon historical data pertaining to individual participants within the environment. Furthermore, the synchronization of temporal data originating from various sources can be accomplished through the utilization of RNNs [2].

Ultimately, the Application Layer, supporting a wide range of services, can effectively utilize various AI resources, including the valence-arousal model, VAEs, GANs, and cross-generation models for algorithmic composition. Genetic algorithms aid in real-time tracking and improvisation, while machine learning and ANNs enhance performance. Cross-modal transformers are valuable for generating dance movements, and combining SeqGAN with multi-adversarial training enables automatic lyrics generation that matches musical tempo and more. A concise summary of this discussion is presented in Table 1.

Table 1. Synthesis of AI techniques that can be used in the Io3MT.

Layer	Tasks	AI Techniques
Things Layer	Specifies physical devices	Machine Learning, RNNs, CNNs
Link Layer	Connect devices to the network and pre-process data	RNNS, CNNs, SVM, Isolation Forest, Reinforcement Learning
Network Layer	Data transmission	ANNs, RNNs, SVM, Genetic Algorithms, Reinforcement Learning, Decision Trees, Clustering Methods
Middleware Layer	System control and management	NLP, RNNs, Machine Learning
Application Layer	Support services	VAEs, GANs, ANNs, SeqGAN, Cross-generation Models, Cross-model Transformers, Genetic Algorithms, Machine Learning

Henceforth, the workflow of Io3MT performances are poised to experience substantial enhancements through the integration of these procedures. Within the domain of audio processing, the real-time application of effects and corrections serves to elevate the auditory experience and the narrative dimension of artistic endeavors. This involves the capacity to dynamically adjust musical elements such as notes, rhythm, dynamics, tonal characteristics, and visual aesthetics, with the aim of achieving a more harmonious resonance with the emotional responses of the audience.

Encompassing the entirety of artistic creation, the benefits engendered by the incorporation of artificial intelligence are manifold. These encompass the real-time generation of content, whether it pertains to music, visual effects, or dance movements. Additionally, AI facilitates the production of diverse variations of content within predefined thematic or stylistic parameters, as well as the capacity to deviate from established norms. Moreover, artificial intelligence can be effectively harnessed to recommend or extract sensory effects [1]. Such advancements significantly expand the potential applications within this artistic context. Consequently, these advancements yield highly adaptable and interactive experiences that respond to audience actions, thereby sustaining engagement and sustained interest throughout the artistic endeavor.

The utilization of these techniques within the creative process is conducive to the attainment of precise aesthetic outcomes and the elicitation of intended emotional responses. Simultaneously, it aids in the judicious selection of color palettes, shapes, and musical compositions in a holistic manner. This iterative approach serves to perpetually align the ultimate artistic output with the initially envisioned aesthetic objectives.

Regarding infrastructure considerations, artificial intelligence exhibits the potential to enhance QoS by endowing devices with cognitive capabilities and contextual awareness. This capability facilitates the optimization of network traffic, amelioration of packet routing, mitigation of packet losses, efficient management of power consumption, and augmentation of the adaptability of network devices. Likewise, the integration, synchronization, and translation of diverse forms of information within an Io3MT environment play a pivotal role in bolstering QoE. In this context, the ecosystem is amenable to tailoring according to audience preferences, thereby rendering it more resonant and memorable performance.

Also, it contributes to the realm of business intelligence by enabling the examination of demographic data, both prior to and following a performance, with the objective of deriving critical insights. These insights play a fundamental role in repertoire selection, the formulation of efficacious marketing strategies, and the generation of comprehensive performance reports.

Despite numerous advantages, there are significant challenges in the realm of technology and AI. From a technological perspective, issues like network latency in real-time content creation and the diverse nature of data pose obstacles. Socially, disparities in technology access, standards set by the dominant class, and environmental concerns are pressing issues [30]. In AI domain, security, biases, and misinformation are overarching concerns, including equitable access, privacy, and addressing AI-generated misinformation.

Conversely, AI in artistic contexts faces unique challenges, such as the complexity of selecting attributes during system training, hindering adaptability to diverse artistic narratives and increasing development time and costs. Challenges also arise in hybridization and rendering in Io3MT, where various signal types and multisensory integration complicate content creation. Structural challenges involve limitations in cohesive organization and mapping issues in handling diverse content representations. Reproducing automatically generated content is difficult due to technical mastery and human irreproducibility, with potential solutions including increased training competence and defining difficulty levels.

In light of these considerations, AI holds significant disruptive potential in the realm of art, particularly when integrated with Io3MT. Unlike traditional methods, it enables autonomous learning through pattern recognition, shifting the artist's role to that of a curator and AI system refiner. This change prompts a reevaluation of originality and authenticity in art, allowing for more fluid and less constrained artistic performance while challenging existing paradigms [6,23].

In summary, AI techniques and systems are progressively integrating into artistic endeavors. While challenges such as copyright issues, work misappropriation, and plagiarism persist, this domain, particularly when intertwined with Io3MT, holds the promise of advancing digital literacy, fostering critical thinking, and democratizing the means of art production and reproduction.

5 Conclusion

In the modern era of widespread technology, especially in entertainment, the Internet of Multisensory, Multimedia, and Musical Things (Io3MT) has emerged at the intersection of various technological domains. This burgeoning field is characterized by a quest for new experiences and knowledge accumulation. This paper aims to draw parallels between Io3MT and Artificial Intelligence (AI), exploring techniques and potential impacts.

Incorporating AI entails exploring optimal tools for visual displays, experimenting with agency in digital performances, and utilizing programming languages for combining multimedia and multisensory elements. Despite presenting challenges and risks, the potential gains in artistic exploration and innovation are significant.

In summary, this paper contributes to the understanding of Io3MT and AI, providing a comprehensive introduction for artists and scientists. It encourages further investigation and outlines a future focus on creating practical, collaborative environments that integrate these areas, emphasizing equity, accessibility, and validation through fluency metrics and security.

Acknowledgment. The authors would like to thank the support of the following funding agencies CAPES (Grant Number 88887.668167/2022-00), CAPES Print, CNPq, and FAPERJ.

References

1. de Abreu, R.S., Mattos, D., Santos, J.D., Ghinea, G., Muchaluat-Saade, D.C.: Toward content-driven intelligent authoring of mulsemedia applications. IEEE Multimed. **28**(1), 7–16 (2021). https://doi.org/10.1109/MMUL.2020.3011383
2. Agarwal, P., Alam, M.: Investigating IoT middleware platforms for smart application development. In: Ahmed, S., Abbas, S.M., Zia, H. (eds.) Smart Cities—Opportunities and Challenges. LNCE, vol. 58, pp. 231–244. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2545-2_21
3. Ahanger, T.A., Aljumah, A., Atiquzzaman, M.: State-of-the-art survey of artificial intelligent techniques for IoT security. Comput. Netw. **206**, 108771 (2022)
4. Alvi, S.A., Afzal, B., Shah, G.A., Atzori, L., Mahmood, W.: Internet of multimedia things: vision and challenges. Ad Hoc Netw. **33**, 87–111 (2015). https://doi.org/10.1016/j.adhoc.2015.04.006, https://www.sciencedirect.com/science/article/pii/S1570870515000876
5. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**, 2787–2805 (2010). https://doi.org/10.1016/j.comnet.2010.05.010
6. Benjamin, W.: The Work of Art in the Age of Mechanical Reproduction. Penguin UK (2008)
7. Berger, P.L., Luckmann, T.: The Social Construction of Reality: A Treatise in the Sociology of Knowledge. Anchor (1967)
8. Bresin, R.: Artificial neural networks based models for automatic performance of musical scores. J. New Music Res. **27**(3), 239–270 (1998)
9. Brooks, A.L., Brooks, E.: Interactivity, Game Creation, Design, Learning, and Innovation: 5th International Conference, ArtsIT 2016, and First International Conference, DLI 2016, Esbjerg, Denmark, 2–3 May 2016, Proceedings, vol. 196. Springer (2017)
10. Brown, A.: Generative music in live performance. In: Generate and Test: Proceedings of the Australasian Computer Music Conference 2005, pp. 23–26. Australasian Computer Music Association (2005)
11. De Mantaras, R.L., Arcos, J.L.: Ai and music: from composition to expressive performance. AI Mag. **23**(3), 43–43 (2002)
12. Dong, H.W., Hsiao, W.Y., Yang, L.C., Yang, Y.H.: MuseGAN: multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
13. Edmonds, E., Turner, G., Candy, L.: Approaches to interactive art systems. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 113–117 (2004)
14. Floris, A., Atzori, L.: Managing the quality of experience in the multimedia internet of things: a layered-based approach. Sensors **16**, 2057 (2016). https://doi.org/10.3390/s16122057
15. Friberg, A., Colombo, V., Frydén, L., Sundberg, J.: Generating musical performances with director musices. Comput. Music. J. **24**(3), 23–29 (2000)
16. Ghinea, G., Timmerer, C., Lin, W., Gulliver, S.R.: Mulsemedia: state of the art, perspectives, and challenges. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **11**(1s), 1–23 (2014)
17. Ghosh, A., Chakraborty, D., Law, A.: Artificial intelligence in internet of things. CAAI Trans. Intell. Technol. **3**(4), 208–218 (2018)

18. Goffman, E.: The arts of impression management. Organ. Identity Reader, 11–12 (2004)
19. Hodo, E., et al.: Threat analysis of IoT networks using artificial neural network intrusion detection system. In: 2016 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6. IEEE (2016)
20. Li, R., Yang, S., Ross, D.A., Kanazawa, A.: AI choreographer: music conditioned 3D dance generation with AIST++. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 13401–13412 (2021)
21. Lopez-Rincon, O., Starostenko, O., Ayala-San Martín, G.: Algorithmic music composition based on artificial intelligence: a survey. In: 2018 International Conference on Electronics, Communications and Computers (CONIELECOMP), pp. 187–193. IEEE (2018)
22. Ma, X., Wang, Y., Kan, M.Y., Lee, W.S.: Ai-lyricist: generating music and vocabulary constrained lyrics. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 1002–1011 (2021)
23. Mazzone, M., Elgammal, A.: Art, Creativity, and the Potential of Artificial Intelligence. In: Arts, vol. 8, p. 26. MDPI (2019)
24. Osuwa, A.A., Ekhoragbon, E.B., Fat, L.T.: Application of artificial intelligence in internet of things. In: 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 169–173. IEEE (2017)
25. Papadopoulos, G., Wiggins, G.: AI methods for algorithmic composition: a survey, a critical view and future prospects. In: AISB Symposium on Musical Creativity, vol. 124, pp. 110–117. Edinburgh, UK (1999)
26. Qiu, Z., et al.: Mind band: a crossmedia AI music composing platform. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2231–2233 (2019)
27. Rottondi, C., Chafe, C., Allocchio, C., Sarti, A.: An overview on networked music performance technologies. IEEE Access **4**, 8823–8843 (2016). https://doi.org/10.1109/ACCESS.2016.2628440
28. Turchet, L., Fischione, C., Essl, G., Keller, D., Barthet, M.: Internet of musical things: vision and challenges. IEEE Access **6**, 61994–62017 (2018). https://doi.org/10.1109/ACCESS.2018.2872625
29. Turchet, L., et al.: The internet of sounds: convergent trends, insights, and future directions. IEEE Internet Things J. **10**(13), 11264–11292 (2023). https://doi.org/10.1109/JIOT.2023.3253602
30. Vieira, R., Barthet, M., Schiavoni, F.L.: Everyday use of the internet of musical things: intersections with ubiquitous music. In: Proceedings of the Workshop on Ubiquitous Music 2020, pp. 60–71. Zenodo, Porto Seguro, BA, Brasil (2020). https://doi.org/10.5281/zenodo.4247759
31. Vieira, R., Gonçalves, L., Schiavoni, F.: The things of the internet of musical things: defining the difficulties to standardize the behavior of these devices. In: 2020 X Brazilian Symposium on Computing Systems Engineering (SBESC), pp. 1–7 (2020). https://doi.org/10.1109/SBESC51047.2020.9277862
32. Vieira, R., Muchuluat-Saade, D., César, P.: Towards an internet of multisensory, multimedia and musical things (Io3MT) environment. In: Proceedings of the 4th International Symposium on the Internet of Sounds, pp. 231–238. IS2 2023, IEEE, Pisa, Italy (2023). https://doi.org/10.1145/XXX, in Press

33. Wang, Y., et al.: Network management and orchestration using artificial intelligence: overview of ETSI ENI. IEEE Commun. Stand. Magaz. **2**(4), 58–65 (2018)
34. Zhu, H., et al.: Xiaoice band: a melody and arrangement generation framework for pop music. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2837–2846 (2018)