Towards a Hierarchy of Trust in Human-AI Music-Making

ADRIAN HAZZARD, University of Nottingham, United Kingdom

CRAIG VEAR, University of Nottingham, United Kingdom

SOLOMIYA MOROZ, University of Nottingham, United Kingdom

This paper investigates Human-AI collaboration between an AI and robotic arm performing with an ensemble of musicians, over an iterative process of improvised music-making. Music-making is an inherently social task, with relationships built on layers of implicit trust that are typically developed over prolonged social engagement and practice. We were interested in useful patterns for collaboration between humans and AI's, and in this setting specifically, whether the human musicians would accept and develop a 'trusting' relationship with our system, called *Jess+*. We introduced a 'hierarchy of trust' model as part of our methodology which placed an emphasis on trust-building of not only the AI and robot's contribution to the ensemble's music making, but also to the social aspects of the process across the wider project team. Our approach appeared successful in promoting creative and musically expressive relationships with the system, one they considered 'trustworthy'.

CCS Concepts: • **Human-centered computing → Interaction design theory, concepts and paradigms**.

Additional Key Words and Phrases: trustworthy autonomous systems, human-AI interaction, human-robot music, social models of trust

1 INTRODUCTION

Our increased engagement with autonomous systems raises questions about their trustworthiness, how it is made manifest in the systems, factored into their design [13], regulation, and use [1, 8, 11]. Matters of trust and trustworthiness are broad [5, 12, 17], malleable [7, 14], highly context specific [4], and in relation to technology, often contested [1, 10]. In creative settings, the artists intentions [9], processes, use of technologies may be open to interpretation, or deliberately playful or provocative, adding to the complexity of understanding what trustworthiness means in relation to collaborative human-AI [2]. Our work is interested in seeking useful patterns for collaboration in Human-AI music making.

This paper draws on a research project which studied an inclusive ensemble of three musicians (one disabled, two non-disabled) improvise with a prototype AI and robotic arm (called *Jess+*) over an iterative process of musical exploration, reflection and discussion. We were interested in the development of their creative practice with the system and the development of the notion of trustworthiness in this practice. The focus on trustworthiness here, represents one of several threads of our on-going investigation into creativity and inclusively in collaborative music making between disabled and non-disabled musicians and AI and robotics [16], and also the evolution of the music score through interactive and intelligent technologies [15].

Over the course of the workshops Craig Vear (as PI and workshop leader of *Jess+*) developed a hierarchy of trust approach to support the musicians' to form expressive and 'trusting' relationships with the system. This involved building trust across many aspects of the social, creative, technical and behavioural relationships engendered throughout the team and its project. What emerged was a framework of principles and approaches that governed the project, which later, upon reflection and evaluation by Craig Vear and Adrian Hazzard developed into a hierarchy of trust across, and interlinked with, different domains. Currently, our hierarchy is non-validated and exploratory, but we are interested in further careful scaffolding of the process of human-AI collaboration to better understand how relationships with, and trust in such technologies and systems is developed, understood and made manifest.
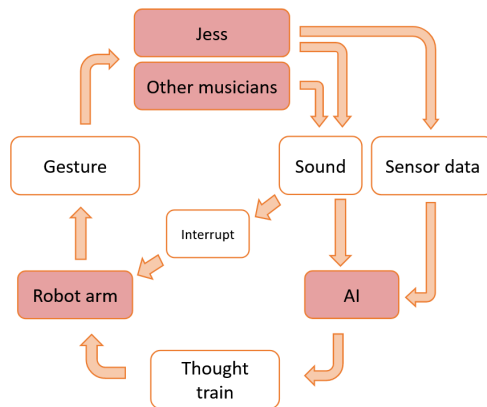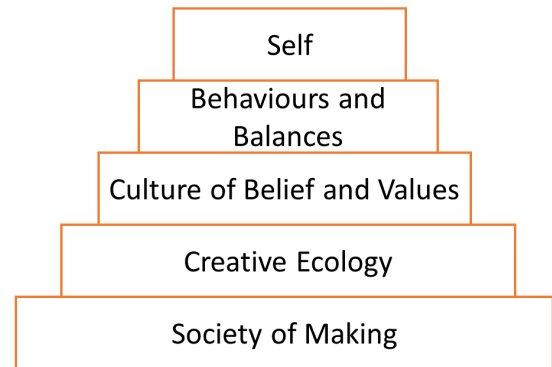
Fig. 1. Interaction Design of the Jess+

Fig. 2. The Jess+ Hierarchy of Trust

2 *JESS+* SYSTEM OVERVIEW

Jess+ is an intelligent robotic system for real-time shared musical creativity, employing an AI and robotic arm. *Jess+* does not contribute sonic material to the ensemble, rather the robot arm performs gestures akin to a conductor or dancer. These gestures can result in geometric markings on paper via a pen fixed to the robot arm (recalling experimental graphic score notations of the mid-twentieth century [6]), or they can be gestures in 3D space.

Figure 1 illustrates the interaction design, which is a closed-loop system where the human ensemble are sensed by the AI through sound (a microphone in the room) and on-body measurements (brainwave (electroencephalography) and skin conductance (electro-dermal activity) of the disabled musician). This data is used to generate a real time stream of predicted data from an AI (consisting of 7 trained models), which in turn is translated into a robot arm gesture. This is then interpreted by the human musicians as an abstract form of notation, who make a response via their instruments. Their response is then sensed by the AI (audio, EEG, EDA), and the loop continues. For further design details see [16].

3 METHOD AND REFLECTIONS

We employed a practice-based approach [3] over five iterative workshops of improvised music-making between three musicians (Jess, Clare and Deirdre) and the AI Robot Arm *Jess+*–supported by the researchers and developers. Each workshop was followed by further system development in response to the musicians' feedback and requests. Semi-structured interviews were conducted at the beginning and end of this process to capture further reflections. We used these activities to capture the musicians experience of working with–and perceptions of trust regard–*Jess+* (see [16]).

We augmented our methodology by employing what we refer to as the 'Hierarchy of Trust' model (Fig 2), to help us understand the perspectives and characteristic that might arise from engaging in creative collaboration with AI's. This hierarchy is a non-validated exploratory pathway to emphasize the social and ecological nature of creative work and supporting working relationships. Heavily influenced by Maslow's Hierarchy of Needs,[1] we structured ours so that each tier relies on the integrity of the tier below to maintain the bonds of trust. Ensuring that each tier was maintained independently drove the management of the workshop and creative processes. Although our work is specific to music

---

making, given its sociological focus on creative engagements we hope this hierarchy might be applicable for any creative human-AI project. In the following we outline each tier and introduce some discussion points from our project.

4 HIERARCHY OF TRUST: IN PRACTICE

Society of Making: the first tier focuses on growing and maintaining trust in the relationships between the group of makers, musicians, technologists and managers of our project. Schoenherr et al. [13] argue the design of AI's should be human-centred and a "multi-stakeholder response." We aimed to flatten any hierarchy of contribution and to develop a shared sense of ownership, where each member's views and actions were valued. For example the musicians were chosen specifically because of their prior experiences of working together, and the research and development team were also experienced musicians, so could communicate on equal terms. The musicians music-making centered on improvisation to remove the hierarchy of composer > performer, and matters of musical direction or expectation. This laid the groundwork for the development of trust within the core social relationships and emphasised the project's inclusive character. Our musicians prior experience highlighted that successful music making requires a shared social or musical connection, ideally both. For example, Deirdre noted that "a good musical experience is based on a good relationship and connection between the musicians. It rejuvenates itself in that you want to do it over and over and it gets better and better." In our workshops, the musicians were initially uncertain about working with *Jess+*–"We all had our guards up" (Jess)–and they assumed the system should lead, "Initially when I was playing with it, I wanted to do what it was telling me to do. But it's not that at all, you literately need to play with it" (Deirdre). These feelings soon diminished and the musicians went on to describe *Jess+* as an accepting and non-judgemental actor in their ensemble.

Creative Ecology: ecology is a study of the relationships between a community and their environment, and the balances between these relationships. We worked exclusively in a 'messy' workshop space, free from external pressures and the formality of traditional recital halls, so the team could 'own' this space. To foster a creative, open and honest environment–a space to "play" music not to "serious" music–we drew on Benford et al's [2] call for 'responsible irresponsibility'; provide a safe, bounded space (responsibility) within which irresponsible action is encouraged, so as to foster creativity. Over the course of the workshops the musicians were free to contribute to the music in any way they felt comfortable, the aesthetic was developed by them, the workshops did not follow tightly planned schedules and they explored different performance configurations and approaches, which also extended to physical reconfiguration of the space and the robotic arm. The latter included transitioning from the robot arm marking on paper by removing the pens, and adding other appendages to the head of the arm such as a glove or a cat toy. On a related note, Clare highlights the relationship between freedom to take risks and trust, "you need trust that everybody will respond to you, and you will respond to them. But it's also being allowed to make mistakes or to risk doing something [unexpected]."

Culture of Belief and Values: we wished to be open and honest to all regards the design and development of the AI and its behaviours, so they could contribute to its on-going development. There was never any attempt to 'pull the wool over the eyes' of the musicians, or employ 'Wizard of Oz' interactions, and questions about its design were tackled head-on (while acknowledging the challenges of differing levels of technical knowledge). In simple terms, by trusting the creative process and team scaffolding it, by proxy the musicians could start to trust this robot active within their music making. The AI also had a belief structure with which it understood the world (a world of music making). This belief structure controlled the language of its generative gestures, which was made transparent to all through discussions and was modified throughout the process (see [16]). Literature on trust disagrees as to whether technologies can be considered trustworthy. For example [1] posits that an AI cannot be trustworthy as it is unaware of the trust placed in them, thus trust in a technology is better understood as the trust placed in the makers of it [10]. Contrary to

this position, our musicians uncoupled their trust in the research and development team from their trust in the *Jess+* system. Specifically, as much as they trusted the team, if the *Jess+* system did not "listen" and respond to them with the care and consideration they attributed to it, they would not view the system as trustworthy in their ensemble practice.

Behaviours and Balances: trusting how the AI behaved was very important to the musicians within their music-making, in addition to the developers who sometimes struggled to explain how the black-box AI made its decisions. Equally important was how the team maintained and balanced the bonds of trust and understanding that grew through the process. An illustration of this was observed in how the musicians formed differing understandings of, and relationships with *Jess+*. Jess considered the system to be an extension of her, "an extra layer of [her] creativity," whereas for Clare and Deirdre it was an equal partner, a "creative accompanist." Our musicians emphasised the importance of forming shared connections with fellow performers and collaborators, "(you) need to share the enthusiasm. You can play with the best instrumentalist in the country and there will not be that connection, you need to be of the same mind" (Deirdre). In our project, the musicians' did form a shared connection with *Jess+*, but it could break down. Twice we discovered the microphone level through which the AI sensed the musicians' musical gestures was set too low, meaning the AI was not 'listening' to them. Clare illustrates the implications of this, "that's why you need to be in a safe ensemble. And you know, if you trust the people that you're playing with, then it's fine to be vulnerable. And that brings out some really good music making. But then if you find that one of the players, and in this case it's the robot and it's just switched off. It makes you feel awful, I put my trust in you and it's not being reciprocated."

Self: the final tier concerns the role of the individual in how they contribute to maintaining bonds of trust. Initially our musicians were uncertain about what to expect of the AI and robot arm, and were equally uncertain as to what they could bring to the project, "I don't know much about technology." We observed the development of trust in their own behaviour and integrity of contribution, which in turn deepened their sense of fulfillment and project ownership. This was illustrated in the final workshop–a sharing of the work done with wider project partners. The musicians–rather than researchers–volunteered to lead in explaining and advocating for the value of the system and wider project. Furthermore, our musicians also reported a new found freedom of expression in their improvisational practice that emerged directly from their experience, "[*Jess+*] gives you that inspired creativity, it helps to force the music out of you" (Jess), and, "definitely freed me up an awful lot more and made me a freer improviser than I was" (Deirdre).

5 FINAL THOUGHTS

Our interest is in seeking useful patterns for collaboration between humans and AI's, and we are at the early stages of investigating strategies to scaffold the development of human's understanding, relationship building and shared creative practice with collaborative-AI's, with the ultimate aim of comprehending how trust is made manifest in such settings. We propose an initial model for these strategies–the 'Hierarchy of Trust'–which emphasises the social and ecological nature of creative work and supporting working relationships.

REFERENCES

[1] Pepijn Al. 2023. (E)-Trust and Its Function: Why We Shouldn't Apply Trust and Trustworthiness to Human–AI Relations. *Journal of Applied Philosophy* 40, 1 (2023), 95–108.
[2] Steve Benford, Adrian Hazzard, Craig Vear, Helena Webb, Alan Chamberlain, Chris Greenhalgh, Richard Ramchurn, and Joe Marshall. 2023. Five Provocations for a More Creative TAS. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*. 1–10.
[3] Linda Candy, Ernest Edmonds, and Craig Vear. 2021. Practice-based research. In *The Routledge International Handbook of Practice-Based Research*. Routledge, 27–41.
[4] Alan Chamberlain, Steve Benford, Joel Fischer, Pepita Barnard, Chris Greenhalgh, Ju Row Farr, Nick Tandavanitj, and Matt Adams. 2022. Designing for Trust: Autonomous Animal-Centric Robotic & AI Systems. In *Proceedings of the Ninth International Conference on Animal-Computer Interaction*.

1–4.

[5] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.

[6] Jennie Gottschalk. 2016. *Experimental music since 1970.* Bloomsbury Publishing USA.

[7] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[8] Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. 2023. Trustworthy AI: From principles to practices. *Comput. Surveys* 55, 9 (2023), 1–46.

[9] C Thi Nguyen. 2021. Trust and sincerity in art. (2021).

[10] Philip J Nickel. 2012. Trust in technological systems. In *Norms in technology.* Springer, 223–237.

[11] Karoline Reinhardt. 2023. Trust and trustworthiness in AI ethics. *AI and Ethics* 3, 3 (2023), 735–744.

[12] Eike Schneiders, Alan Chamberlain, Joel E Fischer, Steve Benford, Simon Castle-Green, Victor Ngo, Ayse Kucukyilmaz, Pepita Barnard, Ju Row Farr, Matt Adams, et al. 2023. TAS for Cats: An Artist-led Exploration of Trustworthy Autonomous Systems for Companion Animals. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems.* 1–5.

[13] Jordan Richard Schoenherr, Roba Abbas, Katina Michael, Pablo Rivas, and Theresa Dirndorfer Anderson. 2023. Designing AI using a human-centered approach: Explainability and accuracy toward trustworthiness. *IEEE Transactions on Technology and Society* 4, 1 (2023), 9–23.

[14] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad Van Moorsel. 2020. The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the 2020 conference on fairness, accountability, and transparency.* 272–283.

[15] Craig Vear. 2019. *The Digital Score: Musicianship, Creativity and Innovation.* Routledge.

[16] Craig Vear, Adrian Hazzard, Solomiya Moroz, and Johann Benerradi. 2024. Jess+: AI and Robotics with Inclusive Music-Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3613904.3642548

[17] Stephen Wright. 2010. Trust and trustworthiness. *Philosophia* 38 (2010), 615–627.