Deeply Listening Through/Out the Deepscape

Hugo Scurto[1], Axel Chemla–Romeu-Santos[2]
[1] EUR ArTeC / Université Paris 8 / EnsadLab
[2] STMS Lab Ircam–CNRS–Sorbonne Université
Paris, France
hugo.scurto@ensad.fr, chemla@ircam.fr

Abstract

This paper presents early artistic, conceptual and technical work toward practicing and theorising through/out the *deepscape*. I first introduce the concept of deepscape, which may designate the global flows of media intensively computed by deep learning throughout the Internet, entangled with the material, human and cultural resources they capitalise on throughout corporate infrastructures of artificial intelligence (AI). I then propose to explore deep listening of soundscapes generated by deep learning as a practice to raise awareness of the planetary scale of the deepscape. I relate the diffractive prototyping of a deep generative model of soundscape, based on the multichannel hacking of the Realtime Audio Variational autoEncoder (RAVE), trained on worldwide soundscape data that I transversally recorded over 28 places in late April 2022, using the Locustream online sound map. I argue that listening to the planetary soundscape that continually flows from this deep generative model may reveal the ethico-onto-epistem-ology of deep learning, by recalling the landscapes that are being exploited by infrastructures of AI, while situating data collection practices and training costs of deep learning. The paper ends by discussing art and science work that might be engaged to reveal and reconfigure the deepscape in depth.

Keywords

Deep Learning, Deep Listening, Diffractive Art Practice.

Introduction

For over half a century, a vast majority of scientists have been developing and describing Artificial Intelligence (AI) through anthropomorphic conceptions and representations. From symbolic AI to modern Artificial General Intelligence [27], simulating human-like intelligence through quantification and computation have been standard goals and methods for most researchers in the field. These anthropomorphic conceptions stem from millenary Western representations of AI as humanoid robots, or artificial objects endowed with human intelligence [37], which are still actively shaping popular representations, along with corporate narratives, on AI [11].

The technological revolution brought by deep learning in the last decade paradoxically destabilised anthropomorphic conceptions of AI [12], while simultaneously reinforcing anthropocentrism. On the one hand, difficulties in interpreting internal representations of deep neural networks, as well as images they can generate [32], pushed scientists and artists to approach AI as a form of non-human intelligence [34, 6]. On the other hand, systems powered by deep learning, from facial recognisers to text and image generators [43, 14], entailed worldwide AI applications, both public and private, across almost all sectors of human society [42].

As an artist-researcher, I am interested in developing a non-anthropocentric representation of AI that raises awareness of the planetary costs of deep learning. I am inspired by Kate Crawford and Vladan Joler's *Anatomy of an AI system* (2018) [22], whose large-scale map reveals the globalised, extractive infrastructures that lie beneath modern AI applications, and that are barely tackled by anthropocentric discourses. I am also interested in engaging in diffractive art practice with deep learning [48] to perform and reconfigure this non-anthropocentric representation. I am inspired by Memo Akten's *Deep Meditations* (2018) [6], whose deep learning-generated continual flows of images and sounds let us embrace the interconnectedness of AI with all socio-material bodies of time and space, including us, and the planet.

In this paper, I introduce the concept of *deepscape*, along with deep listening experiences of planetary soundscape live-generated by one deep neural network. The deepscape designates the global flows of deep learning-generated media intensively computed by deep learning throughout the Internet, entangled with the material, cultural and human resources they capitalise on throughout corporate infrastructures of AI. Deep listening experiences [39] of deep learning-generated continual flow of soundscape aim at raising awareness of the deepscape, by leveraging a planetary representation for a deep neural network, and recalling the landscapes that are being exploited by infrastructures of AI. By including non-human voices and planetary scales within research and practice with deep learning, this research-creation thus seeks to resist the anthropocentric culture of modern AI research, illuminating its media aesthetics and socio-ecological impacts.

The next sections successively describe an early conceptual mapping of the deepscape, the diffractive prototyping of a deep generative model of planetary soundscape, and deep listening experiences led through/out the deepscape. The paper ends by discussing art and science work that might be engaged to reveal and reconfigure the deepscape in depth.

* The first author uses the pronoun "I" to relate their research-creation and their technical collaboration with the second author.

Mapping the Deepscape

In this section, I sketch an early map of the deepscape, by describing the global media flows, corporate computational infrastructures, and planetary socio-material resources that compose it. I argue that the concept of deepscape better represents the media aesthetics and planetary scales of deep learning than mainstream anthropomorphic conceptions and representations of AI. Furthermore, I suggest that the deepscape may inspire new practices with deep learning, which may complement those with/in AI, as we will see.

Global Media Flows

Since DeepDream almost a decade ago [38], the Internet has been flooded by media generated by deep learning. These media encompass image, sound and text, and more recently, video, 3D movement or biosignals [47]. Sounds and images generated by deep learning also permeate our physical environments, with voices of AI assistants resonating in smart homes and cities, and mobile-based face filters re/acting over our bodies across the globe [33]. The concept of deep fake has made popular such media across fiction and reality, raising awareness of deep learning's threats to steal one's identity [13]. As AI-based virtual worlds are rapidly evolving, we can expect these media flows to grow in the near future [1].

Taking inspiration from Arjun Appadurai's notion of *scape* [8], I propose to introduce the notion of deepscape to point to these global flows of media intensively computed by deep learning and permeating digital and physical worlds. Rather than fake, I argue that these media are real, based on the impact they have on cultures and societies, and the planetary computation they build on to exist, as we will see. Importantly, such media are materially configured to exist under the form of generative flows, due to computational architectures of deep learning, which can either support streamable media generation [16], or continuous interpolation between images or sounds [9]. Artists have well explored these material attributes of deep learning in stream-based installations [10], including dadaBots and their infinite online music livestreams [19], or Anna Ridler and her moving images [44].

While humanoid robots surrounded by blue zeros and ones keep on being summoned to promote modern AI applications [11], I suggest that the singular aesthetics of media flows generated by deep learning are starting to replace anthropomorphic representations of AI at a global scale in our collective imagination. Indeed, ambiguous attributes of images generated by GANs [32] (Generative Adversarial Networks, a specific type of deep neural network), along with the almost infinite interpolative flows from which they emerge, have permeated our representations of machines across art and science, along with our conception of how they may exhibit creativity [29]. In this paper, I limit my analysis to witnessing such global flow aesthetics of media generated by deep learning, and how they are transforming our representations of AI.

Corporate Computational Infrastructures

Different types of computational infrastructures actually support the deepscape and its global media flows. The first is the Internet [20], since media generated by deep learning actively flows through social networks [1], helped by mobile applications. A second type is deep generative models [26], which are compressed representations of datasets, computed by deep learning, enabling stream-like generation of media that more or less resemble the dataset. A third type is supercomputers and data centers combined with cloud computing [52], since deep learning requires intensive computation and large amounts of data to produce deep generative models.

Over the last decade, such computational infrastructures have been in majority invested, developed and owned by global corporations, including the Big Tech. In addition to owning Instagram and Facebook, Meta recently introduced their AI Research Supercluster, consisting of 16,000 graphical processing units (GPUs) especially designed to develop deep learning applications, parallel to their metaverse [4]. OpenAI—recently followed by Meta, Google and Microsoft—developed several popular deep generative models, including Dall-E [43], GPT-3 [14] and Jukebox [24], enabling to generate image, text and music respectively, based on text prompts. Google developed Colaboratory [18], which leveraged the corporation's GPU-based cloud computing services to accelerate deep learning applications, in turn enabling the emergence of a global community of practice dedicated to deep generative models [40]. Importantly, Google also develops and maintains TensorFlow, one of the most widely used library to develop deep generative models and modern AI applications [2].

I suggest that corporate computational infrastructures are going to exponentially increase the scale of the deepscape and its global media flows in the next coming years. In addition to the Big Tech, hundreds of worldwide corporations have been investing in deep generative models recently, constituting what have been called a "generative AI landscape" or "creative new world" by investors and analysts [3]. It is highly probable that these global corporate investments are to modify distribution of work in the coming years, especially in the creative and cultural industries [7]. Many artists have made such political aspects of AI central to their practice with deep learning [17]. Holly Herndon and Mathew Dryhurst proposed artist-led platforms addressing legal issues related to ownership of training datasets [50], and of deep generative models themselves [31]. In this paper, I limit my analysis to pointing out to corporate computational infrastructures that support the deepscape, and how they may generate governmental issues at a planetary scale.

Planetary Socio-Material Resources

Last but not least, a wide array of planetary resources enables to fuel the deepscape, exploited by corporate computational infrastructures. A first type is material resources [22, 41]. Rare elements such as lithium are required to build the supercomputers enabling deep learning computations, along with personal computers and smartphones required to interact with deep generative models [30]. Extracting such material elements devastate landscapes across the globe, while contributing to huge levels of atmospheric and water pollution, which in turn generate political conflicts between local populations, opaque intermediaries and global corporations [23]. In addition, electricity consumption of deep generative models breaks records and should continue increasing every year, de-

spite engineering efforts to reduce computational costs [52]. This leads to huge carbon emissions that threaten the lives of both humans and non-humans across the globe.

A second type is human resources. Workers are required to extract material resources needed to build computational infrastructures of deep learning, often enduring awful economic and health conditions [23]. Large teams of engineers actively maintain such computational infrastructures [2], from deep learning libraries to mobile AI applications, across various hardware and software, as computational technology keeps on rapidly evolving. Computer scientists are required to design and implement deep generative models. While they sometimes build and collect the datasets over which they train their models, they essentially delegate data work to disadvantaged people across the globe [28], thus reinstantiating work inequalities and data colonialism [21] through deep learning. Far from objective, deep generative models thus perpetuate values and world views of a few [45], reinforcing cultural bias of the global computational infrastructure.

Indeed, a third type of resource that directly relates to this data work is cultural resources. Beyond mere numbers, datasets carry historical and cultural value, as images, sounds or text they represent are always and already entangled within communities that produced it, and labels and classifications they encapsulate reflect certain world views [45]. When a person or a corporation trains a deep generative model over a given dataset, it exploits its cultural value to generate media that somehow reproduces this value, while distorting it, or even biasing it in unethical ways [51]. Artists such as Stephanie Dinkins have sought to illuminate responsible ways of dealing with datasets and deep generative models, especially toward communities that are underrepresented in deep learning, such as the Black community [25]. Cultural accountability may eventually become difficult to tackle, as corporate AI infrastructures have started using datasets produced by deep generative models to train deep generative models [43]. In this paper, I limit my analysis to pointing out to socio-material resources that fuel the deepscape, and how they damage socio-ecological landscapes of the planet.

Diffractive Deep Learning Prototyping

In this section, I describe the diffractive prototyping of a deep generative model of planetary soundscape, based on a transversal soundscape dataset that I collected over the globe in one week-end, and the multichannel hacking of the RAVE variational auto-encoder [15], made in collaboration with the second author. I argue that such a diffractive practice with deep learning reveals and reconfigures the deepscape, by performing the planetary scale of deep learning, while illuminating the material costs that its computations imply.

Transversal Soundscape Dataset

I identified three requirements for my data collection practice to engage with the concept of deepscape. First, data should be collected across worldwide places to transcribe the planetary scale of deep learning. Second, data should be in the form of multichannel soundscapes to convey a sense of spatial landscape through sound. Third, data should combine anthropophony, biophony and geophony [35] to have the model embody a sonic environment rather than one individual form of intelligence, such as a musician or a bird.

I used Locustream to collect this soundscape dataset [49]. Locustream is an online sound map developed by the Locus Sonus research group that enables to listen to live stereophonic soundscapes across the globe. Practitioners can setup their own microphone in the place of their choice, and livestream the sound it captures online, using either a PureData patch running on a RaspberryPi, or a mobile application for smartphone. The map has been used over sixteen years by artists and scientists to promote novel forms of listening to, and performing with, the planet and its diverse environment.

I decided to lead online field recording through the Locustream online sound map. Specifically, I opted to restrain my field recording to one local portion of time—namely, one week-end in late April 2022—, while widening it to all portions of the globe accessible on the map over that period—namely, twenty-eight places across seventeen countries and five continents. I called it transversal, following scientific dataset nomenclatures used in statistics. On the one hand, such a transversal dataset let me situate ecological properties of soundscapes related to planetary seasonal shifts, be they related to human or non-human activities. On the other hand, it enabled me to materialise the planetary scale of deep learning, since all places, cultures and species of the transversal dataset would be eventually normalised and approximated as one planetary soundscape by deep learning.

The resulting dataset consists in nearly 16 hours of stereophonic soundscape (see Figure 1), that I have deeply listened to while recording, with a mean duration of 31 minutes per location (see Figure 2). Such *small data* approach come within the scope of diffractive art practice [48].

Deep Generative Model

Along with the second author, we opted to work with the Realtime Audio Variational autoEncoder (RAVE) [15]. RAVE is a state-of-the-art deep generative model that learns to generate raw audio waveforms using a two-stage training procedure. In the first stage, a variational auto-encoder is trained to learn a high-level latent space of the training dataset. In the second stage, only the decoder is trained with an adversarial generation objective to enforce continuity of the latent space. One advantage of RAVE over other deep generative models are that audio signals are modeled and generated at a 48 kHz sampling frequency, *i.e.,* standard audio quality.

The second author hacked RAVE's implementation[1] by adding a multichannel audio loader, which enables to load audio data with c channels directly for use in RAVE, including the multichannel factor c into encoder, decoder, and adversarial parameters. From a methodological perspective, this add-on could be described as a form of *learnable algorithm*, which comes within the frame of diffractive art practice [48]. The source code for our modified implementation will soon be publicly available. Our deep generative model can be used with the nn$\sim$ object for neural audio synthesis[2].

[1] https://github.com/domkirke/RAVE/tree/multichannel

[2] https://github.com/acids-ircam/nn_tilde

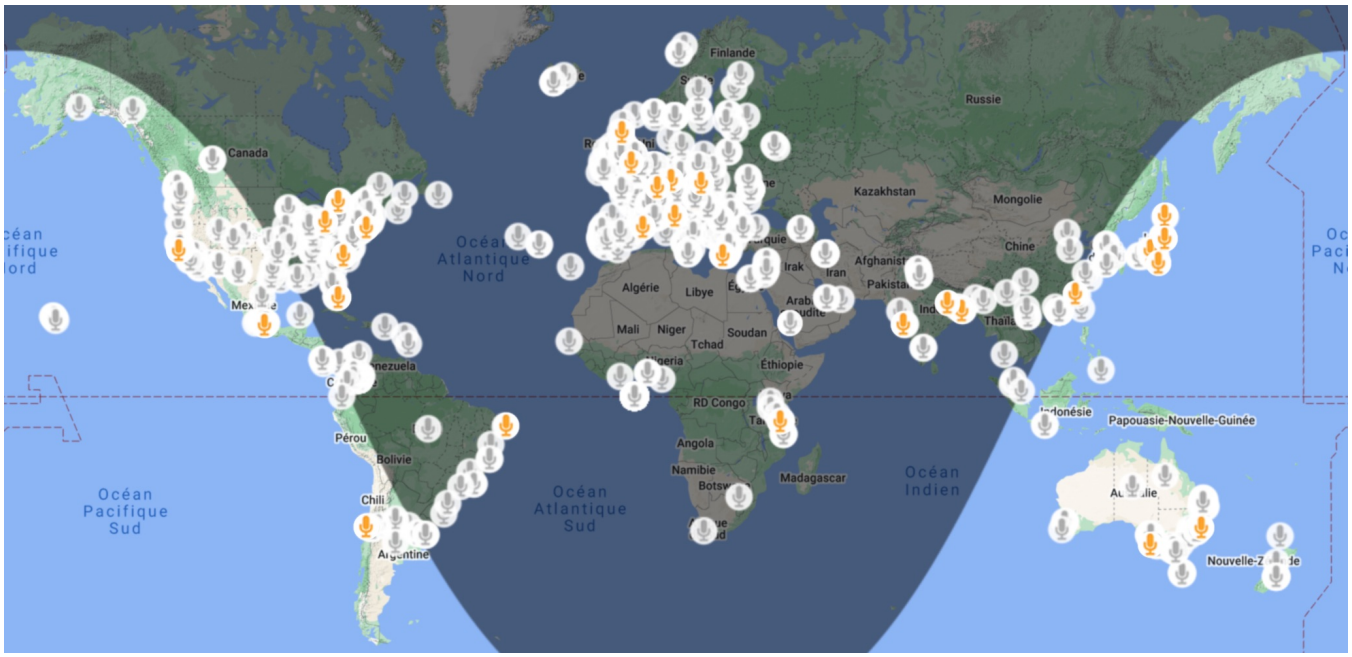

Figure 1: Transversal soundscape dataset. In orange: The 28 places where I led online field recording over one week-end in late April 2022. In grey: all microphones previously opened on Locustream since 2006 [49].

Training and Streaming Costs

We trained our deep generative model on our transversal soundscape dataset for three million steps, specifically two and one million steps for first and second stages. Trainings were conducted using a private infrastructure, which has a carbon efficiency of 0.432 kgCO_2eq/kWh. A cumulative of 360 hours of computation was performed on GPU hardware of type Titan V (TDP of 250W). Total emissions are estimated to be 38.88 kgCO_2eq of which 0 percents were directly offset. These estimations were conducted using the MachineLearning Impact calculator presented in [36]. Streaming costs are equivalent to working with a personal computer, as RAVE support streamable audio generation on standard CPUs [16].

We used the same model hyperparameters and architecture as provided in the original RAVE implementation [15]. Specifically, we preprocessed our transversal soundscape dataset by cropping it into one-second audio slices to apply dequantization, random crop and all-pass filters with random coefficients as data augmentation strategy. Training evaluation was led by combining quantitative and qualitative dimensions of sound generation, with the first training stage stopped based on joint quantitative observation of loss minimisation and qualitative listening of generated audio, and the second stage stopped based on qualitative listening criteria.

Deep Listening Experiences

In this section, I relate first deep listening experiences created through/out our deep generative model of planetary soundscape. I group them within an ongoing artwork called *Deepscape: Transversal*. Deep listening is a meditative aesthetic proposed by composer Pauline Oliveros "to inspire both trained and untrained performers to practice the art of listening and responding to environmental conditions in solo and ensemble situations" [39]. I suggest that deep listening may be helpful to raise awareness of the deepscape and its interconnectedness with all socio-material bodies of the planet.

The following paragraphs describe online radiophonic broadcast, collective acousmatic listening, and open listening and discussion situations, inviting diverse audiences to experience planetary soundscape flows computed by our deep generative model, and describe their spatial, temporal and material attributes. These *situational wholes* enabled to illuminate *somaesthetic behaviours* of our deep generative model, which jointly come within the scope of diffractive art practice [48].

Online Radiophonic Broadcast

The first experience consists in the online radiophonic broadcasting of stereophonic audio live-streamed by our deep generative model. It is mounted on the Locustream online sound map[3] since October 17th, 2022, while being temporarily stopped at times. On the server side, I used the nn~ Max/MSP object for neural audio synthesis, and the Locus Sonus Locustream Pure Data patch [49] to mount the live stream online. On the deep learning side, I experimented with RAVE prior for unconditional audio generation [16]. On the sound map, the stream is located at the material place where the deepscape is computed, that is, from my home as of now. To inspire deep listening, the broadcast is accompanied with a note of intent that include two questions: *Who terraforms this deepscape? Whose scapes are getting threatened as AI sucks our attention away from planetary issues?*

[3] http://locus.creacast.com:9001/deepscape_transversal.ogg

Collective Acousmatic Listening

The second experience consisted in a collective acousmatic listening session led at Cirque Électrique, Paris, France, on November 5th, 2022, for the opening of an AI-based experimental music concert. It consisted in 15 minutes of planetary soundscape[4], recorded in one shot from our deep generative model, based on real-time reconstruction of stereophonic audio coming from a microphone placed at my home's window. The latter was picking up a flow of environmental soundscape, which included keyboard sounds, human sounds and discussions, bird vocalisations, car engines, wind howlings and children screaming in their playground.

The session started by an on-stage presentation of the piece, as well as listening guidance standing as deep listening instructions. Specifically, I told the audience that they were going to be put in the dark to listen to 15 minutes of soundscape generated by one deep learning model; that they may approach this experience as sound art rather than music, and that they may let themselves immerse into imaginary worlds while listening. As we set the volume relatively high to produce strong bass, I warned the audience that the experience was going to be quite loud, and therefore, that they should not hesitate to leave if it became difficult.

The audience—approximately one hundred people—was standing up while put in the dark for acousmatic listening. During listening, several people were witnessed closing their eyes and collectively relaxing their necks and bodies while standing up. At the end of the session, some audience members reported how the listening experience felt shorter than 15 minutes, which somehow testifies of their immersion in deep listening. They also felt a sense of space while listening throughout the stereophonic stream. They reported encountering non-human voices throughout the experience, such as humid rocks, strong wind, screaming dogs and singing birds.

Open Listening and Discussion

The third type of experience consisted in open listening and discussion sessions led with art-research practitioners, teachers and students. The first was a workshop at Kallio-Kuninkala, Järvenpää, Finland, on August 24th, 2022. The second was a public event at École d'Art d'Aix-en-Provence, France, on October 9th, 2022. The third was a seminar at École des Beaux-Arts de Paris, France, on November 7th, 2022. Sound materials included both unconditional audio live-streamed by our deep generative model, and the 15-minute planetary soundscape previously recorded.

The sessions started with conceptual elements about the deepscape, as well as technical information on how we prototyped our deep generative model. Then, I communicated listening guidance standing as deep listening instructions. Specifically, I told practitioners and students to listen attentively to the stereophonic soundscape flowing from our deep generative model, and ask themselves the following three questions: *Where am I? What am I listening to? To whom does it cost?* In addition, I asked them to imagine five words that would best describe their listening experience. Due to the openness of these sessions, most participants did not strictly follow these instructions, but rather used them as listening guidance to foster their imagination.

The conceptualisation of the deepscape found echo in most art-research practitioners and teachers. Participants that involved in the discussion generated various interpretations of the deepscape through/out listening. One artist had the feeling to be *"on another planet inhabited by Martians"*. One composer described *"an impossible landscape of windy ocean with small insects and land grass"*. Rather than a soundscape, one teacher felt they were under a *"strange living creature"*, or *"monster"*, attempting to regurgitate things with its throat. Students felt they witnessed *"noises of a diver"*, moving *"inside and outside water"* while *"adjusting their oxygen level"*. One of them added: *"It's a good profile of what a planet is sounding like. It's like a blue planet."*

Discussion

In this section, I discuss further art and science work to be done to reveal and reconfigure the deepscape in more depth. Specifically, I discuss the planetary media computation driven by AI and revealed by the deepscape, the inseparability of ethics, ontology and epistemology when engaging with deep learning, and current work led toward bodily intra-acting through/out the deepscape.

Deepscape as Planetary Media Computation

Discourses on AI almost always suggest anthropomorphic conceptualisations and representations among expert and novice audiences. These conceptualisations often dismiss the media aesthetics of deep learning, as its computed flows of images, sounds and texts intensively permeate worldwide physical and digital environments. By evoking a sense of spatiality with the suffix "-scape", and of deep learning with the prefix "deep-", the concept of deepscape may reveal the planetary media computation driven by AI, recalling the intensive socio-material extractions led by corporate infrastructures to support its media flows across the planet [41]. I am looking forward to critiques from other researchers to reconfigure this concept in more depth and entangle a plurality of perspectives and disciplines within a potential theoretical framework.

From a practical perspective, the concept of deepscape also emphasises the landscape aesthetics that accidentally emerged from deep learning. Indeed, first art practices with deep learning consisted in navigating so-called latent spaces computed by deep learning over scientific datasets built for image generation [17]. Rather than the realism originally sought by scientists, the most intriguing findings essentially lied in ambiguous images, that deep learning interpolated or extrapolated from the dataset [32]. Artists then curated and showcased such findings as art, or sought to recreate this navigation process in art installations, typically presenting the audience with deep learning-generated flows of sounds and images [10]. I suggest that this process is reminiscent of field recording, where sounds and images are experienced in a continuous flow, while navigating unknown places, sometimes producing strange encounters that exceed human interpretation skills, and careful data collection that might reveal the interconnectedness of the recorder with the planet [35].

[4] https://soundcloud.com/hugoscurto/deepscape-transversal-cirque-electrique-20221105

Recently, artists have been exploring spatial concepts for AI, describing it as "imaginary landscapes" [53] conveying a "worldview" through their training data [5], or even crafting deep generative models to create virtual places [44]. My *Deepscape: Transversal* artwork might contribute to these works, by situating data collection practices and planetary costs of deep learning within art practice, while seeking to resist the anthropomorphic culture of modern AI research.

Ethico-onto-epistem-ology of Deep Learning

Our diffractive prototyping of deep learning illuminates the inseparability of ethics, ontology and epistemology when engaging with deep learning. As we discussed above, collecting a dataset entails an ethics of ownership, representativeness and positionality. In our case, all soundscapes recorded online were open sourced by the Locustream online sound map [49]. Specifically, each soundscape comes with a complex history of negotiating with local populations and authorities, ensuring that an open microphone would hinder human and non-human communities as little as possible. Furthermore, the transversal soundscape dataset pragmatically represents socio-economic abilities of worldwide populations and countries to buy microphones and provide access to the Internet. Lastly, my positionality was described through the relying on a transversal data collection method, of which I attempted to be transparent on the advantages, motivations and limitations to it, as well as through my choice to deeply listen to one stream at a time while recording, instead of automatically recording all streams simultaneously using machine listening.

I suggest that dataset building inevitably produces an ontology of deep learning-generated media. Reactions harvested along the three deep listening experiences testify of the planetary soundscape flowing from our deep generative model, approximating and normalising worldwide soundscapes in the dataset. Making training and streaming costs transparent through our technical collaboration, as suggested by diffractive art practice [48], also illuminates the material entanglement of such planetary soundscape with planetary resources exploited by infrastructures of AI to produce deep generative models. Beyond artwork analysis [10, 17, 29], I suggest that further research might be done on corporate deep generative models, such as Dall-E, GPT-3 or Jukebox, to reveal the ontologies they produce, since they are often trained unethically over large datasets automatically scraped from the Internet.

Lastly, deep listening experiences of my *Deepscape: Transversal* artwork may also promote a alternative epistemology for deep generative models. Indeed, scientists often evaluate deep generative models through fast and quantitative methods, typically automatically-generating short samples with deep learning, and measuring reconstruction quality using Likert scales [15]. I suggest that approaching media flows computed by deep learning through slower and careful crafting may reconfigure how we conceptualise deep generative models, and perhaps, how we develop them. While such crafting practices are emerging naturally among communities of practices dedicated to deep generative models [40], I suggest that researchers also have a responsibility to promote such slower ways of producing knowledge about deep generative models, in a rapidly-evolving computational society.

Bodily Intra-Acting Through/Out the Deepscape

While deep listening may help grasp the planetary scale of the deepscape, it might remain opaque for certain audiences that are not accustomed to music or sound art practices. Fortunately, there remains several ways of practicing and bodily intra-acting [48] through/out the deepscape. Along with other artists [53, 44], I am currently exploring landscape image generation with deep learning. While images remain two-dimensional, they might recall the landscapes of the planet that are getting damaged as the deepscape rapidly evolves.

I am also currently leading sound and image field recording in one local portion of space, while widening it to all portions of time accessible over one year. Such a longitudinal data collection practice should complement the transversal method presented in this paper. Instead of a planetary soundscape, it may enable to learn a model of the planet itself, as sound and light recorded over the year from one identical place may reflect the planet's motion in time and space, with corresponding climatic variations. Such a deep longitudinal model should further my diffractive art practice with deep learning, revealing further ecological entanglements of AI with all socio-material bodies of the universe.

Lastly, I am currently exploring soundwalk-inspired modes of listening to, and practicing with, deep learning. I am collaborating with other artists-researchers to create virtual soundwalks in latent spaces computed by deep learning, as an art-based method to illuminate spatial attributes of the deepscape [46]. We are also exploring computational methods to map bodily movements with soundscape modulations by deep learning to stimulate embodied ways of navigating, and producing knowledge, through/out the deepscape. I am looking forward to further suggestions from practitioners and researchers in other fields.

Conclusion

This paper presented early artistic, conceptual and technical work toward practicing and theorising through/out the deepscape—the planetary media computation intensively led by corporate infrastructures of AI over material, human and cultural resources. I proposed deep listening experiences of deep learning-generated continual flow of soundscape to raise awareness of the deepscape, by recalling the landscapes that are getting exploited by infrastructures of AI. By including non-human voices and planetary scales within research and practice with deep learning, this research-creation thus sought to resist the anthropocentric culture of modern AI research. I am looking forward to further art and science work that might be engaged to reveal and reconfigure the deepscape in depth.

Acknowledgments

Authors would like to thank audience members who took part in deep listening experiences, as well as humans and non-humans who contributed to the transversal soundscape dataset, and research-creation collaborators Anne Dubos, Antoine Caillon, Bertha Bermudez, Emanuele Quinz, Jan Schacher, Lucile Vareilles, Philippe Esling, Peter Sinclair, Samuel Bianchini, Vincent Rioux, *et al.*

References

[1] What if 99% of the Metaverse is generated by AI? https://cifs.dk/news/what-if-99-of-the-metaverse-is-made-by-ai. Accessed: 2022-11-29.

[2] Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. 2016. Tensorflow: a system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283.

[3] Generative AI: A Creative New World. https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/. Accessed: 2022-11-29.

[4] Introducing the AI Research SuperCluster, Meta's cutting edge AI supercomputer for AI research. http://ai.facebook.com/blog/ai-rsc/. Acc: 2022-11-29.

[5] Akten, M.; Fiebrink, R.; and Grierson, M. 2019. Learning to see: you are what you see. In *ACM SIGGRAPH 2019 Art Gallery*, 1–6.

[6] Akten, M.; Fiebrink, R.; and Grierson, M. 2020. Deep meditations: Controlled navigation of latent space. *arXiv preprint arXiv:2003.00910*.

[7] Amato, G.; Behrmann, M.; Bimbot, F.; Caramiaux, B.; Falchi, F.; Garcia, A.; Geurts, J.; Gibert, J.; Gravier, G.; Holken, H.; et al. 2019. Ai in the media and creative industries. *arXiv preprint arXiv:1905.04175*.

[8] Appadurai, A. 1996. *Modernity at large: Cultural dimensions of globalization*, volume 1. U of Minnesota Press.

[9] Arvanitidis, G.; Hansen, L. K.; and Hauberg, S. 2017. Latent space oddity: on the curvature of deep generative models. *arXiv preprint arXiv:1710.11379*.

[10] Audry, S. 2021. *Art in the Age of Machine Learning*. MIT Press.

[11] Better Images of AI. https://betterimagesofai.org/. Accessed: 2022-11-29.

[12] Bory, P. 2019. Deep new: The shifting narratives of artificial intelligence from deep blue to alphago. *Convergence* 25(4):627–642.

[13] Boylan, J. F. 2018. Will deep-fake technology destroy democracy. *The New York Times*.

[14] Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33:1877–1901.

[15] Caillon, A., and Esling, P. 2021. Rave: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011*.

[16] Caillon, A., and Esling, P. 2022. Streamable neural audio synthesis with non-causal convolutions. *arXiv preprint arXiv:2204.07064*.

[17] Caramiaux, B., and Alaoui, S. F. 2022. " explorers of unknown planets": Practices and politics of artificial intelligence in visual arts. *Proceedings of the ACM on Human-Computer Interaction*.

[18] Carneiro, T.; Da Nóbrega, R. V. M.; Nepomuceno, T.; Bian, G.-B.; De Albuquerque, V. H. C.; and Reboucas Filho, P. P. 2018. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access* 6:61677–61685.

[19] Carr, C., and Zukowski, Z. 2018. Generating albums with samplernn to imitate metal, rock, and punk bands. *arXiv preprint arXiv:1811.06633*.

[20] Chatonsky, G. M. 2016. Esthétique des flux (après le numérique).

[21] Couldry, N., and Mejias, U. A. 2019. Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20(4):336–349.

[22] Crawford, K., and Joler, V. 2018. Anatomy of an ai system. *Retrieved September* 18:2018.

[23] Crawford, K. 2021. *Atlas of AI*. Yale University Press.

[24] Dhariwal, P.; Jun, H.; Payne, C.; Kim, J. W.; Radford, A.; and Sutskever, I. 2020. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.

[25] Dinkins, S. 2020. Community, art and the vernacular in technological ecosystems. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 221–221.

[26] Foster, D. 2019. *Generative deep learning: teaching machines to paint, write, compose, and play*. USA: O'Reilly Media.

[27] Goertzel, B. 2007. Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to ray kurzweil's the singularity is near, and mcdermott's critique of kurzweil. *Artificial Intelligence* 171(18):1161–1173.

[28] Gray, M. L., and Suri, S. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[29] Grba, D. 2022. Deep else: A critical framework for ai art. *Digital* 2(1):1–32.

[30] Heelan, J.; Gratz, E.; Zheng, Z.; Wang, Q.; Chen, M.; Apelian, D.; and Wang, Y. 2016. Current and prospective li-ion battery recycling and recovery processes. *Jom* 68(10):2632–2638.

[31] Herndon, H. 2022. Holly+.

[32] Hertzmann, A. 2020. Visual indeterminacy in gan art. *Leonardo* 53(4):424–428.

[33] Javornik, A.; Marder, B.; Barhorst, J. B.; McLean, G.; Rogers, Y.; Marshall, P.; and Warlop, L. 2022. 'what lies behind the filter?'uncovering the motivations for using augmented reality (ar) face filters on social media and their effect on well-being. *Computers in Human Behavior* 128:107126.

[34] Korteling, J. H.; van de Boer-Visschedijk, G.; Blankendaal, R. A.; Boonekamp, R.; and Eikelboom, A. 2021. Human-versus artificial intelligence. *Frontiers in artificial intelligence* 4:622364.

[35] Krause, B. 2008. Anatomy of the soundscape: evolving perspectives. *Journal of the Audio Engineering Society* 56(1/2):73–80.

[36] Lacoste, A.; Luccioni, A.; Schmidt, V.; and Dandres, T. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

[37] McCorduck, P.; Minsky, M.; Selfridge, O. G.; and Simon, H. A. 1977. History of artificial intelligence. In *IJCAI*, 951–954.

[38] Mordvintsev, A.; Olah, C.; and Tyka, M. 2015. Inceptionism: Going deeper into neural networks.

[39] Oliveros, P. 2005. *Deep listening: A composer's sound practice*. IUniverse.

[40] Oppenlaender, J. 2022. The creativity of text-to-image generation. In *25th International Academic Mindtrek conference*, 192–202.

[41] Parikka, J. 2015. *A geology of media*, volume 46. U of Minnesota Press.

[42] Parloff, R. 2016. From 2016: Why deep learning is suddenly changing your life. fortune.

[43] Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; and Sutskever, I. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, 8821–8831. PMLR.

[44] Ridler, A., and Sinders, C. 2021. Mechanized cacophonies.

[45] Scheuerman, M. K.; Hanna, A.; and Denton, E. 2021. Do datasets have politics? disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2):1–37.

[46] Scurto, H., and Postel, L. 2023. Soundwalking deep latent spaces. In *Proceedings of the 23th International Conference on New Interfaces for Musical Expression (NIME'23)*.

[47] Scurto, H.; Caramiaux, B.; Similowski, T.; and Bianchini, S. 2021. Ganspire: Generating breathing waveforms for art-health applications. In *5th NeurIPS Workshop on Machine Learning for Creativity and Design*.

[48] Scurto, H.; Caramiaux, B.; and Bevilacqua, F. 2021. Prototyping machine learning through diffractive art practice. In *Designing Interactive Systems Conference 2021*, 2013–2025.

[49] Sinclair, P. 2018. Locus stream open microphone project. In *Proceedings of the 2018 International Computer Music Conference (ICMC 2018)*.

[50] Spawning. https://spawning.ai/. Accessed: 2022-11-29.

[51] Srinivasan, R., and Uchino, K. 2021. Biases in generative art: A causal look from the lens of art history. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 41–51.

[52] Strubell, E.; Ganesh, A.; and McCallum, A. 2019. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*.

[53] Tokui, N. 2018. Imaginary soundwalk.

Authors Biographies

Hugo Scurto is an artist, designer and researcher, born and based in Marseille. Their research employs art, design and science to craft, prototype and diffract machine learning in an ecology of music. Their practice consists in creating, listening and performing with learning machines that reveal and reshape our musical entanglements with our environments. Hugo is currently postdoc at EUR ArTeC (Paris 8 / EnsadLab), and co-founding member of w.lfg.ng, a post-AI music collective. Before this, they completed a PhD at IRCAM, graduated in Physics from École Normale Supérieure Paris-Saclay, and were visiting research student at Department of Computing, Goldsmiths University of London.

Axel Chemla—Romeu-Santos is a composer, performer and researcher, born and based in Paris. Their scientific research focuses on generative models, bayesian learning, and signal processing to design unsupervised sound synthesis approaches based on perceptual inference, symbolic extraction, and raw signal generation. Their artistic work focuses on the creative aspects of such algorithms in composition and performance configurations. Axel is currently postdoc in the ACIDS group of IRCAM, CNRS, Sorbonne Université, and a co-founding member of w.lfg.ng. They completed a PhD in Deep Learning and Sound Synthesis between IRCAM and the Laboratorio d'Informatica Musicale, and graduated from CRR93 cursus of Computer Music.

continent	country	city	place	day	time (UTC+2)	duration
europe	greece	chania	preservation park	28/04/2022	16:23	00:31:09
	england	londres	walworth	29/04/2022	15:22	00:40:24
	czechia	brno	luzanky	29/04/2022	17:12	00:30:00
	spain	emporda	aiguamolls	29/04/2022	17:42	00:30:07
	germany	furtwangen	black forest	30/04/2022	15:12	00:30:01
	scotland	dumfries	loch patrick	30/04/2022	16:19	00:30:01
	germany	berlin	floating	01/05/2022	02:49	00:30:05
	france	marseille	le frioul	02/04/2022	08:07	00:30:24
asia	india	calcutta	survey park	28/04/2022	16:56	00:30:00
	japan	nagano	otanomo	29/04/2022	16:05	00:30:00
	india	bankura	farm	29/04/2022	18:52	00:11:47
	japan	yamanashi	yamanakako	29/04/2022	19:05	00:38:58
	india	mumbai	lab	30/04/2022	15:43	00:32:09
	japan	hokkaido	maeyama	30/04/2022	20:48	00:39:56
	china	hong kong	nau chi wan	30/04/2022	23:38	00:26:00
	japan	otsuchi	otohama	01/05/2022	01:41	00:30:07
	japan	nagano	otanomo	02/05/2022	09:24	00:30:00
america	mexico	cerro pelon	biosphere reserve	28/04/2022	17:29	00:30:00
	usa	jasper ridge	birdcast	29/04/2022	16:36	00:30:05
	brasil	olinda	thelmo cristovam	29/04/2022	18:21	00:30:06
	chili	las salinas	losriosdezunino	30/04/2022	12:47	00:30:30
	canada	south river	warblers roost	30/04/2022	13:19	00:30:11
	usa	miami	driftwood	30/04/2022	17:00	00:30:18
	usa	greensboro	north carolina	01/05/2022	00:44	00:30:59
	usa	new york	wave farm	01/05/2022	02:16	00:31:01
	canada	leamington	point pelee	02/05/2022	08:40	00:30:27
africa	tanzania	dar	es salaam	30/04/2022	12:19	00:30:00
	tanzania	dar	es salaam	01/05/2022	00:02	00:31:44
oceania	australia	coolum beach	cassiawildlife	30/04/2022	21:47	00:30:00
	australia	cleland	clelandpark	30/04/2022	22:30	01:06:06

Figure 2: Transversal soundscape dataset spatiotemporalities.