Probing Respiratory Care With Generative Deep Learning

HUGO SCURTO, UMRS1158, ISIR, Inserm–CNRS–Sorbonne Université, EnsadLab, EnsAD–PSL, France
THOMAS SIMILOWSKI, UMRS1158, Inserm–Sorbonne Université, R3S, Pitié-Salpêtrière, AP-HP, France
SAMUEL BIANCHINI, Reflective Interaction Research Group, EnsadLab, EnsAD–PSL University, France
BAPTISTE CARAMIAUX, ISIR, CNRS–Sorbonne Université, France

This paper combines design, machine learning and social computing to explore generative deep learning as both tool and probe for respiratory care. We first present *GANspire*, a deep learning tool that generates fine-grained breathing waveforms, which we crafted in collaboration with one respiratory physician, attending to joint materialities of human breathing data and deep generative models. We then relate a probe, produced with breathing waveforms generated with GANspire, and led with a group of ten respiratory care experts, responding to its material attributes. Qualitative annotations showed that respiratory care experts interpreted both realistic and ambiguous attributes of breathing waveforms generated with GANspire, according to subjective aspects of physiology, activity and emotion. Semi-structured interviews also revealed experts' broader perceptions, expectations and ethical concerns on AI technology, based on their clinical practice of respiratory care, and reflexive analysis of GANspire. These findings suggest design implications for technological aids in respiratory care, and show how ambiguity of deep generative models can be leveraged as a resource for qualitative inquiry, enabling socio-material research with generative deep learning. Our paper contributes to the CSCW community by broadening how generative deep learning may be approached not only as a tool to design human-computer interactions, but also as a probe to provoke open conversations with communities of practice about their current and speculative uses of AI technology.

CCS Concepts: • **Human-centered computing → Participatory design**; **Empirical studies in HCI**.

Additional Key Words and Phrases: Generative deep learning, Design research, Respiratory care, AI.

ACM Reference Format:
Hugo Scurto, Thomas Similowski, Samuel Bianchini, and Baptiste Caramiaux. 2023. Probing Respiratory Care With Generative Deep Learning. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 308 (October 2023), 34 pages. https://doi.org/10.1145/3610099

1 INTRODUCTION

Breathing as a phenomenon is increasingly being used in holistic approaches to healthcare, such as mindfulness-based cognitive therapy [49]. Within this space, a significant portion of work has focused on practicing and researching empathic attributes of breathing for human-human interaction. For example, respiratory care experts have explored breathing as a communication modality to slow breathing and appease minds of patients with chronic respiratory diseases [26]. Conversely, stressful social situations can impact an individual's breathing patterns [12].

Such clinical findings have led to an increased interest for breathing in HCI and interaction design. In these works, one of the main focus lies in endowing machines with interactive breathing behaviours, typically to entrain a person's breathing toward slower rates of relaxation [35], or

Authors' addresses: Hugo Scurto, hugo.scurto@ensad.fr, UMRS1158, ISIR, Inserm–CNRS–Sorbonne Université, EnsadLab, EnsAD–PSL, 91 Boulevard de l'Hôpital, Paris, France, 75013; Thomas Similowski, thomas.similowski@sorbonne-universite.fr, UMRS1158, Inserm–Sorbonne Université, R3S, Pitié-Salpêtrière, AP-HP, 91 Boulevard de l'Hôpital, Paris, France, 75013; Samuel Bianchini, samuel.bianchini@ensad.fr, Reflective Interaction Research Group, EnsadLab, EnsAD–PSL University, 31 Rue d'Ulm, Paris, France, 75005; Baptiste Caramiaux, baptiste.caramiaux@sorbonne-universite.fr, ISIR, CNRS–Sorbonne Université, 4 Place Jussieu, Paris, France, 75005.

to support breathing self-awareness [40]. However, these HCI works do not always engage in a collaboration with respiratory care experts, leaving practical insights on breathing, as well as clinical applications of technology, not included in the design process. One example lies in the use of elementary models of sinusoidal waveforms to design breathing behaviours in machines [21], while respiratory care researchers have long studied how fine-grained attributes of breathing waveforms convey rich physiological, activity and emotional information [47].

In recent years, advances in generative deep learning have greatly improved machines' ability to generate fine-grained, realistic data, encompassing text, audio or image [19, 32]. Generative deep learning relies on the training of complex computational models, such as generative adversarial networks (GANs) [25], over large sets of human-produced data. Applications of generative deep learning to HCI for well-being have started to burgeon, such as through the design of chatbots able to appease users that cope with emotional situations [51], or the generation of faces, bodies and voices for AI-powered virtual characters that support health assistance [39].

As an interdisciplinary group of designers, pulmonologists and researchers, we are interested in exploring the potential of generative deep learning to reproduce fine-grained attributes of breathing waveforms in machines. Specifically, we are interested in grounding such computational explorations within existing clinical practices of respiratory care. Recent research in respiratory care suggested that realistically reproducing breathing waveforms in a soft robotic object could provide patients with chronic respiratory diseases with a technological aid to reduce their anxiety [28]. Yet, understanding socio-material implications of applying deep learning to clinical practice requires careful collaboration and probing of healthcare experts early in the design process [9].

In this paper, we combine design, machine learning and social computing to explore generative deep learning as both tool and probe for respiratory care. Specifically, we present a deep learning tool for breathing waveform generation, which we named *GANspire* [46], and which we crafted in collaboration with one respiratory physician. We then report on a probe with a group of ten respiratory care experts, where each expert annotated a batch of breathing waveforms generated with GANspire to qualitatively describe their material attributes. The probe included semi-structured interviews that let experts reflect on GANspire and share their existing clinical practices of respiratory care with patients and AI technology. Our contributions are threefold:

- We crafted a deep generative model of breathing waveforms in collaboration with a respiratory physician. This crafting leveraged respiratory physician's qualitative annotation to discover aspects of physiology, activity and emotion in breathing waveforms generated with GANspire, letting us incorporate respiratory care insights into machine learning design.
- We found that material expressions of generative deep learning enabled to probe respiratory care experts' perceptions and expectations of technology. Beyond realism, ambiguous breathing materials generated with GANspire pushed experts to state their diverse positionalities toward AI, and speculate on deep learning applications to respiratory care.
- We inquired into socio-technical workings of technology within our group of respiratory care experts. Experts' interviews revealed their skepticism toward automation of care, along with their pragmatic acceptation of it, based on their scientific knowledge, and the material needs of patients with chronic respiratory diseases and their carers.

Altogether, our collaborative and probing approach to generative deep learning is valuable to CSCW researchers and practitioners. On the one hand, it enables to transcend engineering approaches to deep learning as a tool for HCI, by attending and responding to material aspects of data practices and deep generative models [11]. On the other hand, it also enables to reveal socio-technical implications of these algorithms, by acknowledging positionalities of machine learning researchers, and contextualising model evaluation within situated communities of practices [43].

2 BACKGROUND AND RELATED WORK

In this section, we review background on clinical practices of respiratory care, with a focus on mindfulness. We then review HCI work seeking to endow machines with breathing behaviours, and motivate our focus on generative deep learning. We finally review designerly explorations of machine learning, whose focus on materiality enables new forms of collaborative research.

2.1 Clinical Practices of Respiratory Care

Breathing is a vital function allowing gas exchanges between organisms and their environment [15]. Unlike other vegetative functions such as heart rate, breathing has the unique characteristic of being both automatically controlled (to maintain life and ensure homeostatis through servo-controlled adaptations to fluctuations in bodily metabolism) and voluntarily controlled. The automatic respiratory central pattern generators located in the brain stem can indeed temporarily be superseded by neural commands of cortical origin, allowing humans to perform voluntary respiratory maneuvers, and to use their respiratory system for non-respiratory purposes, the most important of which being speech production. Cortically-controlled modifications of breathing pattern can exert various influences on brain activity and emotional state. For example, voluntary deep and slow breathing can induce physiological, as well as emotional and cognitive calm in both healthy and unhealthy people [26]. Reciprocally, stressful situations can induce changes in breathing patterns and other physiological indicators [12]. Because breathing is the only vegetative function of which the activity is visible and audible, the intimate relationship between emotional state and breathing pattern makes it a very important agent of inter-human non-verbal communication.

In patients suffering from chronic respiratory diseases, breathing, which is normally not the object of any conscious perception, becomes both conscious and distressing. This defines "dyspnoea" or "breathlessness", a multidimensional experience that resemble pain by many aspects. In such patients, the automatic neural drive to breathe is increased to overcome abnormalities of the respiratory system and in response to emotional and cognitive distress. This makes voluntary breathing control more difficult, and a source of worsening of respiratory suffering. Respiratory diseases therefore not only impair gas exchange and generate dyspnoea and the ensuing handicap, they also have an impact on verbal and non-verbal communication. One major characteristic of respiratory diseases lies in their generally degenerative nature. This means that many patients will continue to experience respiratory suffering in spite of optimal treatments of the underlying condition [34]. Importantly, respiratory suffering can contaminate, to some extent, other groups of people in patients' environment, such as family members, but also caregivers themselves [16]. People indeed react to the observation of respiratory suffering in others by experiencing malaise and dyspnoea, in a way similar to pain. All these elements imply that the management of respiratory diseases cannot be limited to interventions targeting respiratory function. Interventions targeting brain physiology and psychological impact indeed become mandatory when dyspnoea becomes an intrinsic existential experience. Likewise, holistic approaches to respiratory care are increasingly considered, with data pointing at the interest of cognitive behavioural therapy, hypnosis and auto-hypnosis, mindfulness-based therapy and, most importantly, directed empathic concern [49].

Yet, the benefits of breathing techniques mentioned above may not be accessible to patients with respiratory diseases, because they may not be able to access the level of voluntary breathing control necessary to achieve the slowing down of breathing that these techniques involve. Soft robotic haptic objects could be of help in this case. Recently, Haynes *et al.* have developed an haptic aid to attenuate anxiety, under the form of a huggable cushion that can be pneumatically animated with various patterns [28]. More than one hundred volunteers participated in an anxiety-inducing experiment (mathematics test), either without any particular preparation, or with pre-conditioning using the cushion displaying breathing-like movements. This preconditioning was shown to significantly reduce pre-test anxiety. Yet the researchers concluded that capturing more fine-grained attributes of breathing waveforms could improve the efficacy of the device. As such, soft robotic objects with realistic breathing behaviour would made it possible for patients with high levels of respiratory drive to slow their breathing down without having to voluntary interfere with breathing control, both through their anxiety-relieving effects and through the induction of breathing mimicry.

2.2 Endowing Machines With Breathing Behaviours

In the last decade, HCI researchers and practitioners have sought to endow machines with interactive breathing behaviours, typically to promote mindfulness in human-machine interaction. Vidyarthi *et al.* designed *Sonic Cradle*, a tangible installation that maps a person's breathing to music generation to encourage meditation [50]. Prpa *et al.* created an immersive virtual environment that supports breath awareness by producing ambiguous responses to a person's breathing [40]. Frey *et al.* created a wearable pendant that measures a person's breathing and display it in real-time to foster interpersonal social connection [21]. Interestingly, artificially-breathing machines have started being studied by respiratory care research through in-lab studies [7]. Adler *et al.* observed how a virtual reality avatar that breathes synchronously with a human modulates bodily self-consciousness by changing the perceived location of breathing [1]. Czub *et al.* observed how a first-person virtual reality breathing avatar enables to entrain a person's breathing by introducing visuo-interoceptive conflicts between the avatar's breathing and the person's breathing [13]. Allard *et al.* observed how a second-person virtual reality breathing avatar may induce illusory self-identification when breathing synchronously with a person, while also attenuating the affective impact of dyspnoea when breathing asynchronously [2].

In these HCI works, only elementary features of breathing waveforms, such as breathing rate and amplitude, have been analysed or generated to design human-machine interactions. However, researchers in respiratory care have long highlighted the meaningfulness of fine-grained attributes of breathing waveforms [14]. Indeed, human breathing waveforms can express the absence or presence of respiratory disorders in patients [24]. They can also reveal a person's emotional state, encompassing anxiety, fear, happiness or sadness [33]. In certain experimental conditions, they can even reflect a person's breathing style—a feature known as *personnalité ventilatoire* [47]. Recently, generative deep learning was explored as a computational technique to generate fine-grained biosignals for healthcare research, removing confidentiality issues [29]. Technically speaking, an algorithm automatically tunes parameters of a large computational model to fit a large dataset of training examples—*e.g.,* images, sounds or biosignals. The resulting computational model can generate new data that resembles its training data [19]. One of the most successful architecture for generative deep learning is called the Generative Adversarial Network (GAN) [25]. In GANs, a deep neural network generates data examples from a latent space—*i.e.,* a high-dimensional, abstract parameter space—to increase the error rate of a discriminative model, which seeks to distinguish examples produced by the generator from the training dataset. After training, dimensionality of latent spaces can be reduced using generic algorithms such as principal component analysis [10, 27].

To our knowledge, generative deep learning remains unexplored to reproduce fine-grained attributes of breathing waveforms for interaction design. Rather than technically unfeasible, such research remains hard to lead due to methodological differences between communities of engineers, designers and health carers. Recently, the healthcare community highlighted that quantitative scores used by engineers to measure deep learning performance were not sufficient for clinical contexts [38]. Similarly, Yang *et al.* showed how HCI practitioners may defer technical understanding of machine learning to engineers, thus creating computational models that are driven by advances in engineering rather than design innovation, with less value for communities of users [52]. Thus, alternative research methods may be required to explore generative deep learning for breathing waveform generation in respiratory care.

2.3 Designerly Explorations of Machine Learning

In recent years, design researchers have explored alternative practices with machine learning, fostering collaborative research with other communities, including healthcare. Brand *et al.* used speculative design to inquire potential combinations of human data with machine learning that support introspective experiences of mindfulness [8]. Cai *et al.* used a probe approach to understand the needs of health experts toward a deep learning-based diagnostic agent [9]. The probe consisted in cancer grade predictions generated with the diagnostic agent over prostate tissue images. Such a materials-focused inquiry enabled expert pathologists to start a critical discussion about deep learning, eliciting implications and recommendations related to their own clinical practice, thus extending quantitative metrics of engineering. Through four analyses of design with machine learning, Benjamin *et al.* positioned uncertainty as one material expression of machine learning processes [6]. While engineering sees machine learning uncertainty as a metric to be curbed or explained to frame outputs more unambiguously, design postulates that "uncertainty offers an opportunity to design artefacts and scenarios that attribute or more fully exploit the characteristics of machine learning". As such, embracing material expressions of machine learning may create space for diverse communities to interpret its outcomes and tackle its socio-technical workings.

In design research, material expressions of technology can be thought as the set of structural and behavioural attributes that emerge at the boundaries of its pre-defined functions, from the materials and processes used to build it [42]. Researchers have started describing the materials and processes over which deep learning relies. Henderson *et al.* highlighted the costly human and material means required to automatically tune the millions of parameters of a deep computational model [30]. Scheuerman *et al.* have revealed how engineering processes of collecting, curating, annotating and packaging training datasets of deep learning "value efficiency at the expense of care; universality at the expense of contextuality; impartiality at the expense of positionality; and model work at the expense of data work" [43]. In parallel, practitioners have started exploring and analysing material expressions of generative deep learning. In semi-structured interviews with computational artists, Caramiaux *et al.* highlighted crafting, understood as the iterative process of dataset curation and model tweaking, as a method to discover glitches of deep learning deemed imperfect from an engineering's perspective [11]. Through analysis of artistic images created with GANs, Hertzmann positioned ambiguity as one material attribute emerging from craft and uncertainties [31]. While GANs' pre-defined function is to realistically reproduce classes of images, they also enable to morph different classes in a continuous manner, thus producing novel images whose aesthetics are characterised by their interfering between pre-existing classes. Going from art to design practice, Scurto *et al.* described socio-technical guidelines to enable collaborative crafting of machine learning with situated communities of practice [45]. These guidelines include relying on small datasets to enable practitioners grasp material attributes of data and their structural role in machine learning processes. They also include relying on somaesthetics to enable qualitative forms of model evaluation that embrace behavioural attributes.

We propose to adopt crafting and probes as research methods to lead designerly explorations of generative deep learning for respiratory care. On the one hand, we believe that crafting could help us collaborate with respiratory care experts in the discovery of material attributes of deep generative models, in addition to realistically reproducing fine-grained attributes of human breathing. On the other hand, we believe that probes may create space for respiratory care experts to share their broader perceptions and expectations of generative deep learning, in addition to illuminating socio-technical implications of its speculative applications in clinical contexts.

3 GANSPIRE: COLLABORATIVE CRAFTING

In this section, we present the collaborative crafting of *GANspire* [46], a deep learning tool that generates breathing waveforms. We use the term "craft" to emphasise the fact that our work with generative deep learning was iterative and driven by the interaction design, machine learning and respiratory care skills of the different actors involved along design and engineering. The first, second and fourth authors collaborated in such a designerly exploration of generative deep learning for respiratory care. The first author is an artist, designer and researcher working with machine learning as creative material. The second author is a respiratory physician and researcher working on relations between the breathing system and the nervous system. The fourth author is an HCI researcher working on interactions between machine learning, humans, and society.

The next sections describe our collaborative crafting process in a chronological order, starting from contextualising breathing data collection, to discovering deep model materialities, and facilitating breathing waveform generation. We attempt to report on the journey of our collaborative process, paying attention to the respiratory physician's responses to deep generative learning, and the singular design choices that resulted. Importantly, our development process took place in a respiratory care research unit, which gave us access to specific human and material resources, including research engineers and biosignal datasets. Figure 1 shows GANspire's workflow, which results from our situated collaboration. In GANspire, a sampling interface enables to generate breathing waveforms by navigating a three-dimensional latent space. As we will see, this latent space was computed based on subjective interpretations of a deep generative model trained over a custom dataset of human breathing waveforms.

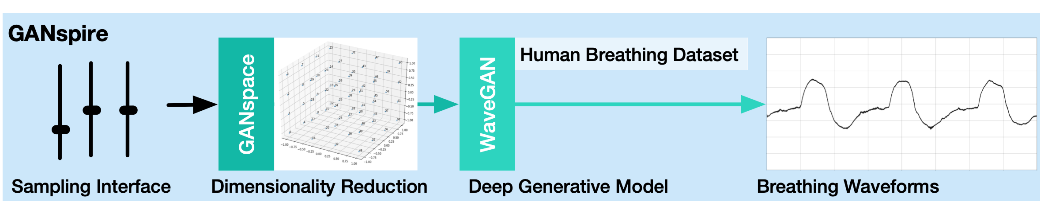


Fig. 1. Overview of GANspire workflow.

3.1 Contextualising Breathing Data Collection

Our collaborative crafting started with the intention of contextualising breathing data collection for generative deep learning. This was motivated by the design approach of first and fourth authors, who seek to develop an understanding of material attributes of breathing, rather than simply leveraging large biosignal datasets built for engineering [29]. By material attributes, we designate fine-grained variations of breathing waveforms, and their related expressiveness. The respiratory physician referred us to a research engineer from our medical research unit, who has access to several datasets of human biosignals recorded in previous physiological research. Among heart rate signals and electroencephalograms, they showed us different measures of human breathing, such as pleural pressure, esophageal pressure, gastric pressure, transdiaphragmatic pressure, and ventilatory airflow. These biosignals were measured using different measurement devices, at different locations of the human body, on different persons that accepted to take part in physiological research work. The first, second and fourth authors eventually opted for ventilatory airflow signals, as airflow is a central parameter in mechanical ventilators used to mitigate dyspnoea [36], and is also a central parameter for creative toolkits of pneumatic prototypes [48]. Additional materials were provided by the respiratory physician and the research engineer, such as scientific presentations aimed at students in respiratory care, as well as professional training tools aimed at respiratory care experts. Altogether, they helped us learn more about breathing as a phenomenon.

The first author received all recordings of ventilatory airflows held by the research engineer of our health laboratory at that time. Figure 2 shows slices of them, which we will refer to as breathing waveforms from now on. It consisted in six hours of human breathing waveforms, recorded over twenty-three healthy persons engaging in three types of activities: exercise, rest, or sleep. They were collected using a pneumotachograph, a device that measures the instantaneous airflow at the airway opening (see Appendix A). Typical breathing waveforms consist of a one-dimensional temporal vector sampled at a 1, 000 Hz frequency, with positive value when one breathes air in, and negative values when one breathes out (see Appendix A). In a virtual conferencing meeting, the respiratory physician described the material qualities of these human breathing waveforms, based on their expertise in neuroscience. We report notes taken from this meeting: *"While breathing is a cyclical phenomenon, it is aperiodic insofar as consecutive breathing cycles are not identical, but exhibit a high degree of breath-by-breath variability. This variability concerns the volume generated by each breath (tidal volume), the duration of each breath (breathing period and therefore breathing frequency), the duration of inspiration and expiration. [...] It derives from the nonlinear and mathematically-complex nature of the activity of central breathing pattern generators."* This argues in favour of modelling fine-grained attributes of breathing waveforms.

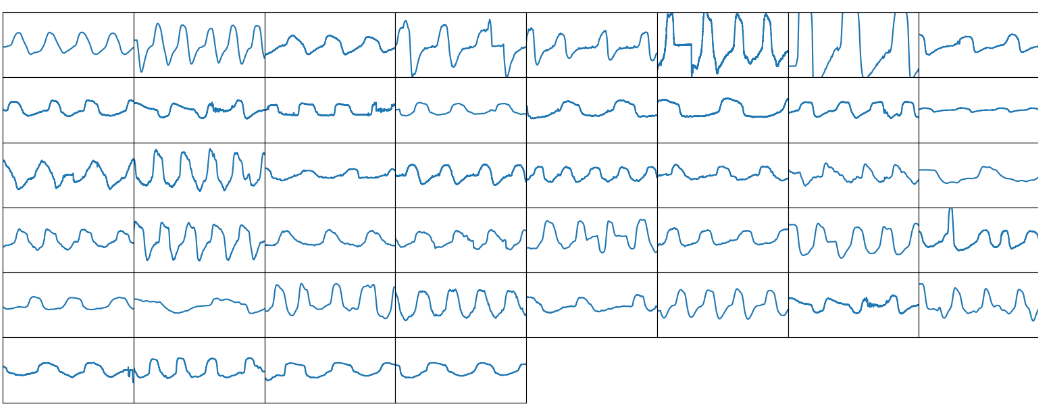


Fig. 2. Slices of our human breathing dataset. Each slice corresponds to 16 seconds (x-axis) of ventilatory airflow recording (y-axis), on one person doing a specific type of activity (exercise, rest or sleep).

Apart from anonymous codes used to discriminate different persons, this human breathing dataset did not have labels enabling to characterise data that was inside. One possibility would have been to ask our respiratory physician to label the breathing dataset, typically using annotations of physiology, activity or emotion. Yet, the first and fourth authors decided to leverage human annotation and interpretation at the end of our crafting process, that is, once we trained our deep generative model to produce breathing waveforms. In other words, we did not pre-define breathing classes in our dataset to attend to how the deep generative model would morph materialities of human breathing waveforms with those of generative deep learning. This is contrary to engineering-oriented approaches to generative deep learning, where evaluation often relies on quantitative metrics seeking to maximise similarities between generated data and labeled training data. Here, in line with design research presented in Section 2.3, our intention was to adopt a crafting approach, where material expressions of deep generative models are not considered as negative attributes to be minimised, but rather as opportunities for interaction design. In addition, we highlight that our dataset is representative of a small group of persons that underwent research with the research unit of our respiratory physician. As such, it values contextuality (*i.e.,* it is representative of the context in which our design process took place) over universality (*i.e.,* seeking to represent all types of breathing produced by humans) [43].

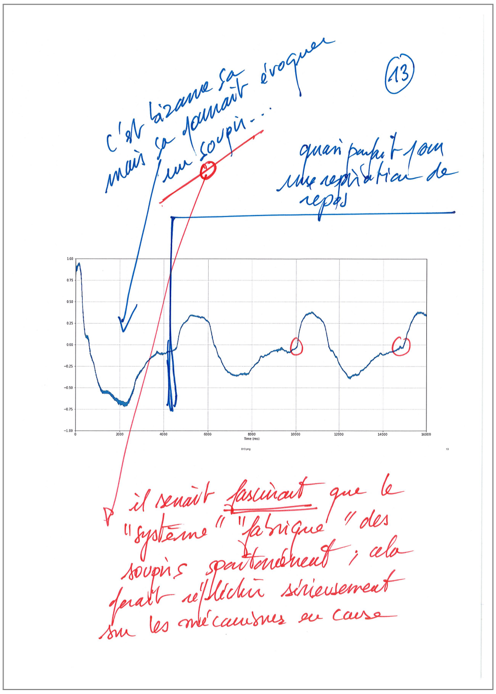
3.2 Discovering Deep Model Materialities

Our next step consisted in identifying a deep generative model that could learn from our human breathing dataset to generate raw breathing waveforms. The first and fourth authors opted for an unsupervised, deep generative model of waveforms, called *WaveGAN*, originally applied to audio signals [17]. WaveGAN is a generic model for fine-grained, one-dimensional temporal waveforms. Its architecture, based on five layers of neural networks, adapts that of DCGAN, originally developed for image generation [41], to be more receptive to the periodicity that is typical of audio waveforms. We thus expected that WaveGAN could be successfully trained on our human breathing dataset, since breathing waveforms display oscillatory variations at different time scales in a way similar to audio waveforms, despite not being strictly periodic. While more elaborated deep generative models for waveforms were engineered at time of work, our design intention was to give most importance to data work over model work [43], that is, to attend to fine-grained attributes of breathing waveforms, along with material expressions of generative deep learning, over accurate modelling from an engineering perspective.

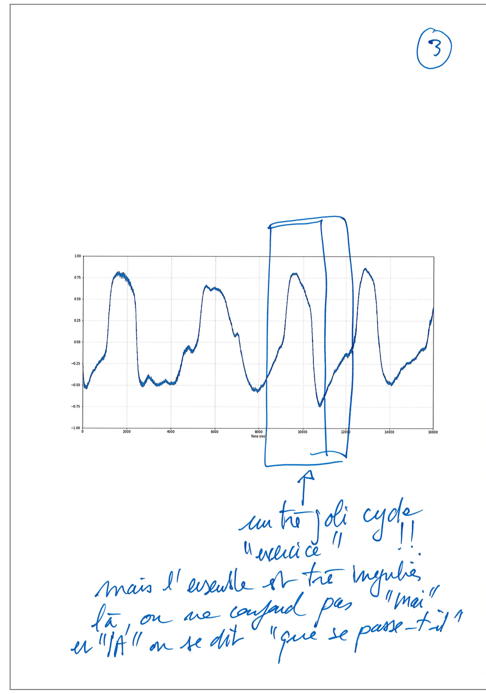
The first and fourth authors performed training experiments with WaveGAN on our human breathing dataset. Technically speaking, training experiments explored different latent space dimensions (10, 100) and parameter dimensionality multipliers (32, 64), as well as various losses (WGAN, WGAN-GP, DCGAN, and least squares) to optimise model's parameters over the training dataset. We also preprocessed our human breathing dataset by slicing breathing recordings as 16-second waveforms, without overlapping, so that training examples were adapted to WaveGAN's architecture. We relied on a GPU to perform training, which took approximately 8 hours for each training experiment. The generated breathing waveforms consist of 16-second slices of instantaneous airflow at 1,000 Hz (see Appendix B), similar to the shape of training examples.

Assessing deep generative models is notoriously difficult, and often rely on quantitative metrics used for engineering [4]. In contrast, we found that subjective interpretations enabled us to leverage the respiratory physician's expertise in breathing waveforms in discovering joint materialities of human breathing waveforms and generative deep learning. The first and fourth author started by curating training experiments based on qualitative observation and comparison of generated breathing waveforms with human breathing data. Then, the respiratory physician was presented with two batches of 100 generated breathing waveforms, randomly sampled from two training experiments curated by first and fourth author. Specifically, they observed overall structure and fine-grained variations of each generated breathing waveforms, and qualitatively assess their specific materialities, that is, their tightening or widening of materialities of human breathing waveforms. Interestingly, the respiratory physician reacted well to such a qualitative approach, as they spontaneously proposed to annotate generated breathing waveforms to support interdisciplinary communication on these aesthetic and material attributes. From a design perspective, such a qualitative evaluation can be seen as a form of somaesthetic appreciation, where breathing waveforms are appreciated through bodily sensory perceptions rather than quantitative criteria, creating space for discovery of unexpected material expressions and behaviours of machine learning [45].

Figure 3 shows example annotations produced by the respiratory physician for generated breathing waveforms resulting from two different training experiments. At first glance, we observed that the respiratory physician annotated breathing waveforms at a fine-grained level. This was expected, since they elicited these material attributes during dataset collection, as shown in Section 3.1. Based



(a) Top: *"that is strange. but it could evoke a sigh..."*; *"almost perfect for a calm breathing"*. Bottom: *""it would be fascinating if the 'system' 'produced' sighs spontaneously; this would make one think seriously about the mechanisms involved"*.



(b) Bottom: *"a very nice 'exercise' cycle!! but the whole is very irregular. here, one doesn't mistake 'myself' for 'AI', one rather asks 'what's going on"'*.

Fig. 3. Respiratory physician's qualitative annotations for two training experiments. On left, the breathing waveform morphs fine-grained attributes of breathing, thus producing ambiguity and enthusiasm for the expert. On right, the breathing waveform displays too much irregularities to be interpreted as realistic.

on this, the expert sought to find whether these fine-grained attributes realistically reproduce that of human breathing waveforms. Interestingly, we observed that ambiguous breathing waveforms systematically intrigued the respiratory physician in their annotations. Here, ambiguity means that fine-grained attributes of breathing waveforms remain realistic, but are hard to interpret, or combined in an unusual way. Figure 3a shows an example ambiguous breathing waveform generated through our first training experiment, presenting a continuous morphing between a calm breathing and a sigh, along with a *"strange"* glitch. We argue that this ambiguity is a typical material expression of GANs, as described in Section 2.3. Surprisingly, we observed that this material expression produced enthusiastic reactions from the expert, who sought interpretations from both respiratory care and technological perspectives. On the contrary, as shown in Figure 3b, our second training experiment and its material expressions produced too many irregularities in breathing waveforms to be perceived as realistic, or even ambiguous, by the respiratory physician. In view of feedback of the respiratory physician, we decided to stop our training experiments here, and opted for the first training experiment—a model trained on a DCGAN loss combined with a 64 parameter multiplier a 10-dimensional latent space.

3.3 Facilitating Breathing Waveform Generation

Inspired by our previous discoveries, we decided to facilitate breathing generation with GANspire to support further collaboration with respiratory physicians exploring material expressions of our deep generative model. The first and fourth authors first searched to reduce dimensionality of our deep generative model by implementing *GANSpace* [27]. GANSpace enables to reduce dimensionality of a latent space by applying principal component analysis (PCA) at one given network layer of a GAN, then using the first few components as latent parameters for generation. While GANSpace was shown to extract interpretable controls for image generation, similar PCA-based techniques are also used in the audio domain to identify the most informative parts of the latent space, without necessarily providing interpretability [10]. Our approach is in line with the latter work, that is, to propose to reduce dimensionality with no assumption on interpretability. Technically speaking, we implemented and applied the GANSpace algorithm at each of the five layers of our trained WaveGAN. Specifically, we randomly sampled 1,000 points from the latent space of our trained WaveGAN, from which we produced as many feature tensors at each of the five layers of our WaveGAN. We then computed one PCA on each of these batches of feature tensors, that is, one PCA for each of the five layers of our WaveGAN, and compared their explained variance ratios. We found that performing PCA at the first intermediate network layer enabled to capture most of the variability in breathing generation, with 75% of the generated breathing waveforms captured by the first three components of the PCA. In other words, GANSpace enabled to reduce dimensionality of our trained WaveGAN from a 10- to a three-dimensional latent space, while preserving 75% of its learned representation.

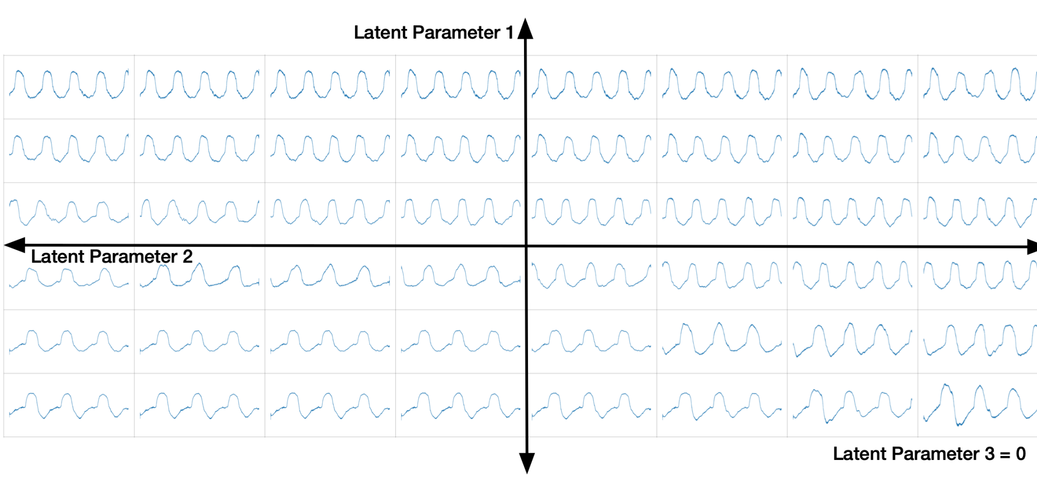


Fig. 4. Example 16-second breathing waveforms (x-axis: time; y-axis: airflow) sampled from GANspire's three-dimensional latent space (X- and Y-axis: latent parameters 1 and 2). Continuously changing one latent parameter triggers continuous morphing between generated breathing waveforms.

Figure 4 shows example breathing waveforms generated from this dimensionality reduction. For this specific figure, we regularly sampled two of the latent parameters computed with GANSpace, while leaving the other one at some fixed value. The respiratory physician qualitatively observed that continuously changing one latent parameter triggers a continuous morphing between generated breathing waveforms. In addition, they also remarked that the PCA-based latent space concentrated the most realistic breathing waveforms at its center, while producing the most ambiguous attributes at larger parameter values. From an engineering perspective, dimensionality reduction has led to a loss of 25% of the model's learned representation, thus to a loss of completeness toward our human breathing dataset. Yet, we argue that it enabled to focus on the materiality of generative deep learning, embodied by ambiguous breathing waveforms and revealed by continuous latent morphings, which is in line with our material approach.

We eventually designed an interface for our trained model to enable real-time exploration of generated breathing waveforms. We used *Marcelle*, a web-based toolkit for interactive machine learning [20], to design and implement a two-page user interface for our deep generative model. We used the WebSocket protocol to bridge our web interface and our trained model, respectively coded in Javascript and Python. We proposed to call *GANspire* our resulting deep learning tool, by conjunction with the noun GAN and the verb "respire". GANspire's main page enables to explore and generate breathing waveforms in real-time (see Appendix C). On the left panel, we put three sliders which we respectively connected to the three latent parameters computed with GANSpace. On the main panel, we plotted the generated breathing waveform corresponding to the current latent parameter combination. A toggle component and a slider on the left panel enables to activate and change cut-off frequency of a low-pass filter to remove noisy artifacts typical from WaveGAN [17]. GANspire's second page enables to easily sample a batch of breathing waveforms for further research. We will detail it in the next section, as we will use it to design our probe.

4 PROBE: BREATHING MATERIALS

In this section, we present the method we used to confront a group of practitioners from our respiratory care research unit with breathing waveforms generated with GANspire [46], described in Section 3. Together with the first, second and fourth authors, our first goal was to assess the scope of breathing waveforms generated with the model, encompassing joint materialities of human breathing waveforms and generative deep learning. Our second goal was to use this material to probe what such generated breathing waveforms evoke to respiratory care experts, based on their qualitative annotations and subjective interpretations. Finally, we used this material to probe respiratory care experts' expectations toward the socio-technical roles that such a deep generative model could play in speculative applications to clinical practice. In the following, we will call breathing materials the set of breathing waveforms generated with GANspire and collaboratively curated by our interdisciplinary research team. The next sections successively describe the collaborative curation of breathing materials led with our respiratory physician, the group of respiratory care experts we recruited, as well as the procedure, analysis, and limitations of our probe.

4.1 Collaborative Curation of Breathing Materials

Our goal was to create a batch of breathing materials produced with GANspire that would demonstrate its materialities, spanning a spectrum from realistic to ambiguous breathing waveforms. Simultaneously, we wanted this batch to be of reasonable size, since it was ultimately aimed at being found and analysed by respiratory care experts. We thus relied on GANspire's control interface to interactively explore model parameters and collaboratively design our batch of breathing materials.

4.1.1 Generated Breathing Materials. As explained in Section 3.3, GANspire's latent space is a 3-dimensional, continuous numerical space. Generating breathing material thus relies on numerically sampling this three-dimensional space. Figure 5 shows a graphical representation of GANspire's latent space. Specifically, each axis correspond to one latent parameter for the generative model, resulting from the algorithmic analysis described in Section 3.3. Thus, each discrete point in this latent space leads to the generation of one breathing waveform. Furthermore, each continuous

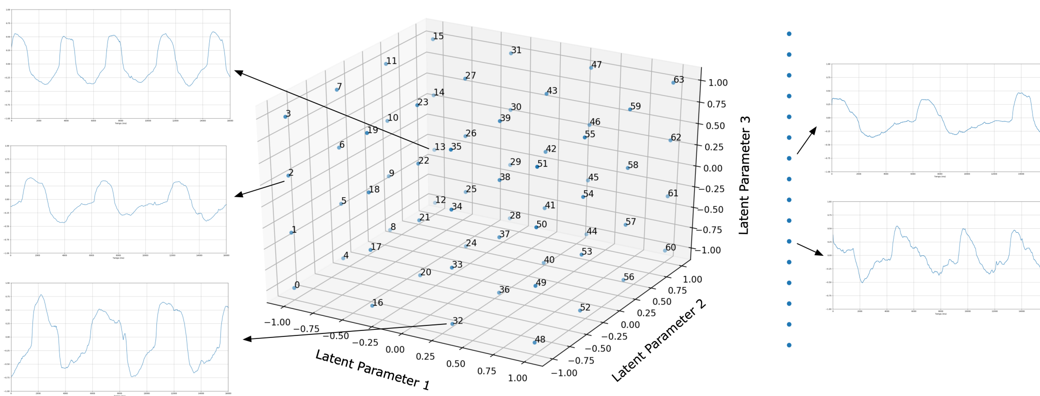


Fig. 5. Breathing material. We sampled 64 breathing waveforms from GANspire's three-dimensional latent space (center), to which we added 16 human breathing waveforms curated from our training dataset (right).

trajectory between two discrete points in the latent space leads to the generation of an infinite number of breathing waveforms, which corresponds to the morphing from origin to destination of the trajectory.

The first, second and fourth authors gathered to explore and sample GANspire's latent space (see Figure 5, center). We first decided to regularly sample the latent space to produce an analysis grid that would allow us to observe the distribution of materialities within the space. We chose a sampling step of 4, resulting in a batch of $4^3 = 64$ generated breathing waveforms, which we deemed sufficient to probe a wide range of breathing behaviours, while remaining sufficiently small to be evaluated by respiratory care experts.

We then sought to balance ambiguity in our batch of breathing materials, that is, balancing realistic breathing waveforms with some others that are hard to interpret, or that combine fine-grained attributes in an unusual way. We set the filter's cut-off frequency at 20 Hz, based on physiological criteria of breathing, so that generated waveforms balance noisy artifacts produced with WaveGAN with realistic attributes of human breathing, based on the respiratory physician's somaesthetic appreciation of the filtered, generated waveforms. Then, we set a numerical range for the three latent space parameters to balance realistic and strange fine-grained attributes. Indeed, we qualitatively observed that the PCA described in Section 3.3 concentrates the most realistic breathing waveforms at the center of its latent space, while producing the most strange attributes at larger numerical values. We thus showed our respiratory physician with a global display of all 64 sampled breathing waveforms, and varied the numerical range continuously until they found a just noticeable difference of balance, based on their expertise and subjective evaluation.

4.1.2 Human Breathing Material. We also decided to add human breathing materials to our batch of generated breathing material. This idea came from qualitative annotations produced by the respiratory physician, reported in Section 3.2, which frequently highlighted mistaking generated waveforms for human breathing biosignals. We thus wanted to probe whether other respiratory care experts would respond similarly to human and generated breathing waveforms.

We proposed to reuse our human breathing dataset, described in Section 3.1. We first sliced airflow recordings into 16-second waveforms so that they have similar shape than generated breathing waveforms (see Figure 5, right). We then set to 16 the number of human breathing waveforms that we aimed at curating, so that the total number of breathing material to be evaluated by experts would reach 64 + 16 = 80. In a video conferencing meeting, we eventually collaborated with the respiratory physician to curate human breathing waveforms. Specifically, we started by randomly selecting one 16-second slice for each person and type of activity represented in our human breathing dataset to present the respiratory physician with a wide range of breathing behaviours. They then leveraged their expertise in physiology to curate a batch of 16 human breathing materials that balance standard and strange human breathings. This way, batches of human and generated breathing waveforms would have subjectively similar ambiguous attributes.

4.2 Group of Respiratory Care Experts

We asked the respiratory physician to identify experts who might be able to respond to our probe. They provided us with a list of fourteen experts, including respiratory physicians, intensivists, palliative care physicians, nurses, research engineers and psychologists, either working in their research unit, or hospital department. We invited them to take part in a study on a "ventilatory activity-generating algorithm", through an email introducing both researchers and institutions leading the study, as well as a summary of the study's procedure, as described in Section 4.3. Out of the fourteen reached, thirteen experts responded to our invitation; two of them politely declined it due to excessive workload during our study period—to cite a few, intern supervision, clinical research, and global pandemic handling.

The ten remaining experts covered different areas of expertise and levels of experience in respiratory care. Specifically, six were respiratory physicians, two were intensivists, one was a respiratory physiologist, and one was a research engineer. Within them, four were senior physicians, while the six others had various years of practice in clinical contexts. Additionally, eight of them underwent health training in French universities, while the two others respectively studied medicine in Italy and Spain. Thus, the group of respiratory care experts were not constrained to one working methodology or one practical approach to interpret breathing materials. We believe this is a relevant aspect of our group of experts, since our wish is to probe generative deep learning in wider communities of respiratory care. Overall, the ten experts took between one and two-and-a-half month to complete the probe, that is, to engage in the annotation study as well as in semi-structured interview.

4.3 Procedure

The first, second and fourth authors collectively reflected on a procedure to lead the probe. In first meetings, the respiratory physician steered us toward quantitative measures used in cognitive science to evaluate breathing activity along different dimensions of emotion. Yet, our collaborative crafting process, and especially, the spontaneous annotations reported in Section 3.2, pushed us to explore more qualitative approaches to breathing waveforms, able to illuminate their ambiguity. We thus proposed a two-phase procedure for our probe, with the first phase consisting of an annotation study, combining qualitative annotations with one quantitative indicator, and the second phase consisting of semi-structured interviews with experts. The respiratory physician reacted well to this mixed procedure.

4.3.1 Annotation Study. In the first phase, experts were asked to annotate each of the 80 breathing waveforms described in Section 4.1 based on their specialised knowledge. Rather than systematically quantifying ratios between inspiratory and expiratory flows (as is often done by pulmonologists), they were encouraged to find words or phrases that best describe their qualitative impressions about the breathing waveform. For example, they were suggested to try to describe the type of breathing activity in presence, its physiological qualities or anomalies, the emotional or expressive states they may reflect, as well as any kind of ambiguous artefact that arouses their surprise or curiosity. They could also draw, circle, arrow, or freely annotate the breathing waveforms if they felt the need to. We underline that this is not a usual task for respiratory care experts, as breathing waveform observation in clinical contexts remains a tacit practice, mostly done at the patient's bed.

In addition to annotating, experts were asked to quantitatively assess the degree of anthropomorphism of each breathing waveform. This quantitative scale was inspired by two engineering-oriented metrics: the Godspeed questionnaire, a well-established tool for evaluating human perception of social robots [5], as well as human-centered evaluation methods for generative deep learning based on discriminating between real and generated data [18]. Here, experts used a 5-point Likert scale to rate whether the breathing waveform has a rather realistic appearance—*e.g.,* it is indistinguishable from a human breathing waveform—or a rather artificial appearance—*e.g.,* it presents many strange artifacts that seems to reveal material expressions of generative deep learning. Ambiguity would thus correspond to a medium score, balancing realistic with artificial attributes.

Experts were sent the 80 breathing waveforms as a series of empty forms (see Appendix D), constituting the probe, along with another document summarising the above-described instructions. We opted for such a remote procedure to provide them with more flexibility to perform the study, as they were known to have a high workload during the study period. Experts were encouraged to print the probe to annotate breathing materials by hand, but remained free to use a PDF reader application to annotate them if more convenient. We recommended that they spend approximately one hour to perform both qualitative annotations and quantitative assessments. Experts were not informed of the research purposes of our probe, nor of potential applications of GANspire, so as not to bias their annotation process, as well as to incite them to take part in the second phase of our study. Importantly, experts were not told that human breathing waveforms were included in the series of waveforms to annotate: rather, all waveforms were presented as being produced by an AI algorithm. Lastly, breathing waveforms were randomised for each expert to avoid ordering effects.

4.3.2 Semi-Structured Interviews. In the second phase, experts were proposed a semi-structured interview with the first author to provide feedback on the annotation task and discuss socio-technical workings of generative deep learning. First, they were asked about their own practice with breathing waveforms: typically, whether they use to perform such qualitative analysis of breathing waveforms in their everyday work, and whether breathing waveforms alone prevented them from sensing specific elements of breathing in patients. Second, they were asked about their knowledge of generative deep learning: typically, whether they experienced generative deep learning in their own practice, and whether they differentiate it from other types of technology they use in their breathing care practice. Lastly, they were presented with a speculative health intervention, consisting of a soft robotic object, whose inflatings and deflatings would be controlled by GANspire, potentially appeasing persons who experience it: this way, they could reflect on ethical aspects of deploying such a technology in clinical contexts, while potentially sharing their recommendations for designing such an intervention for respiratory care. Interviews took place over videoconferencing and were aimed at lasting 30 minutes approximately. When possible, they were conducted shortly after the annotation study, so that experts could report on their experiences of the probe.

4.4 Analysis

We used a thematic analysis approach for the analysis of qualitative annotations. We first reported handwritten annotations of experts as full words in a spreadsheet, then focused on retrieving themes related to physiology, emotion, activity and ambiguity of breathing waveforms. We eventually translated the extracted themes into English, and transformed them as adjectives or gerunds to reveal potential breathing behaviours. The majority of the coding was performed by the first author, while the fourth author cross-checked extracted themes and grammatical transformations and then discuss and resolve disagreements with the first author.

We used descriptive statistics for the analysis of quantitative assessments. We first reported handwritten assessments of experts in a spreadsheet, then computed means and standard deviations of perceived anthropomorphism for each of the eighty breathing waveforms. We also computed the mean and standard deviation of perceived anthropomorphism across all sixteen human breathing waveforms, and across all sixty-four generated breathing waveforms, separately.

Lastly, we used a thematic analysis approach for the analysis of semi-structured interviews. We first transcribed oral communications of both interviewer and interviewee in a spreadsheet, then focused on retrieving themes related to experts' healthcare practice with breathing waveforms, perceptions and experiences of generative deep learning, and ethical considerations on GANspire and its speculative intervention for respiratory care. Both first and fourth authors performed thematic coding on their side and eventually met to discuss and resolve disagreements.

4.5 Limitations

As detailed above, we designed our probe to provoke inspirational responses about AI technology in our group of respiratory care experts. Conversely, we did not design our probe to quantitatively assess anthropomorphism of breathing materials generated with GANspire. One way to lead such a perceptual study would be to opt for a controlled setup, using a batch of breathing materials, half of which would be produced with GANspire, and the other half by humans, then randomly assigning "human/AI" labels to breathing materials when presenting them to respiratory care experts. Here, the anthropomorphism assessments that we will report in Section 5.1.1 should not be taken as a quantitative measure, but rather as a qualitative probe of experts' perception of human- and GANspire-generated breathing materials.

Our study provides an in-depth investigation of how respiratory care experts interpret breathing materials generated with GANspire, but it does not measure experts' attitude toward AI directly. Furthermore, because experts received the misleading direction that all materials were produced by an AI algorithm, their quantitative assessments of anthropomorphism may have been mediated by their attitude toward it. Our quantitative assessments should thus be interpreted as representing a reflection of respiratory care experts' expectations toward AI. As described above, we computed the mean of anthropomorphic assessments to take into account such expectations into our probe, with typically extreme negative or positive assessments respectively testifying of experts' skepticism or enthusiasm toward AI. As changes in the Likert-scale ratings across participants are not guaranteed to be equal, we acknowledge that the descriptive statistics reported in Section 5.1.1 should not be taken as a quantitative analysis, but rather as a qualitative insight on experts' expectations on AI, which will be broadly uncovered through our semi-structured interviews in Section 5.2.

To investigate experts' skepticism or enthusiasm toward technology, our probe included questions about both their current practices of technology in respiratory care, along with speculative uses of technology in their clinical context. While this study design provided a window into their clinical practice, one limitation of this design is that we did not observe socio-technical workings of technology in situ. As a result, aspects of their collaborations with patients, relatives, nurses and colleagues of the care chain, administrators and authorities of the hospital, as well as of their uses of ventilators and other machines of the clinic, may be under-reported or absent from the data we collected. As argued above, rather than complete or objective information on respiratory care practices, our probe sought to uncover experts' beliefs, affects and concerns toward AI, by direct material engagement with breathing waveforms generated with deep learning. As shown below, our use of generative deep learning as a probe created space for critical discussion on technology, perhaps more balanced than if we introduced ourselves as machine learning engineers trying to solve issues in respiratory care as a whole, but with limitations regarding deeper uncoverings of socio-technical workings—we will discuss this in Section 6.1.

5 CLINICIANS' ANNOTATIONS AND RESPONSES TO BREATHING MATERIALS

In this section, we summarise the findings harvested with our probe of breathing materials, produced with GANspire, and led among our group of ten respiratory care experts. These findings contribute to the CSCW community on three points. First, our annotation study enables to understand how respiratory care experts communicate on breathing waveforms, leveraging subjective aspects of physiology, activity and emotion to interpret fine-grained aspects of human breathing, realistically or ambiguously reproduced by generative deep learning. Second, our semi-structured interviews provide empirical findings on current clinical practices of breathing waveforms by respiratory care experts, illuminating their expectations toward GANspire, and their ethical concerns on speculative clinical applications. Third, our overall findings validate our probe approach to generative deep learning, leveraging material expressions of deep generative models to establish an open conversation between us machine learning researchers and respiratory care experts, in an attempt to reduce cultural distances when discussing current uses of AI technology, and collaboratively speculating on novel ones.

5.1 Annotation Study

In this section, we report findings from the annotation study we led with our group of ten respiratory care experts, responding to the 80 breathing materials of the probe in Section 4.1.

5.1.1 Experts' Assessments of Generated Breathing Waveforms. We first analysed the extent to which breathing waveforms generated with GANspire were perceived as realistic and artificial by respiratory care experts. On average, experts assessed breathing waveforms generated with GANspire as more realistic than artificial ($\mu = 0.385, \sigma = 0.505$). This score suggests that breathing waveforms generated with GANspire are ambiguous on average. This validates our collaborative crafting process, described in Section 3.2, which sought to balance realistic attributes of breathing waveforms with material expressions of generative deep learning. In comparison, experts assessed human breathing waveforms with the same order of magnitude ($\mu = -0.449, \sigma = 0.513$). Looking at experts individually, only one of them (P8) assessed the generated breathing waveforms as more artificial, on average, than the human ones (see Appendix E). Still, high standard deviation values suggest that both batches (GANspire-generated and Human) provided with diverse levels of anthropomorphism. GANspire can therefore generate a wide range of breathing waveforms, from realistic to artificial.

Then, we assessed how experts' anthropomorphism assessments (from realistic to artificial) related to GANspire's latent space. Figure 6 reports on anthropomorphism assessments for each of the 64 breathing waveforms generated with GANspire averaged over the ten experts. Specifically, we reused the three-dimensional abstract representation of GANspire's latent space, as described in Section 4.1, and coloured each of the 64 generated breathing waveforms based on their mean anthropomorphism assessment. The visualisation shows that the most artificial breathing waveforms (color closer to violet) seem to be at the edges of the latent space, while the most realistic waveforms (color closer to yellow) seem to be at the center of the latent space. This finding confirms the qualitative observations of our respiratory physician, described in Section 3.3, when we collaboratively crafted the dimensionality reduction technique for GANspire. They also validate our collaborative curation process for our probe, described in Section 4.1, where we set the range of latent parameters to balance realistic with strange attributes of breathing materials. Mean

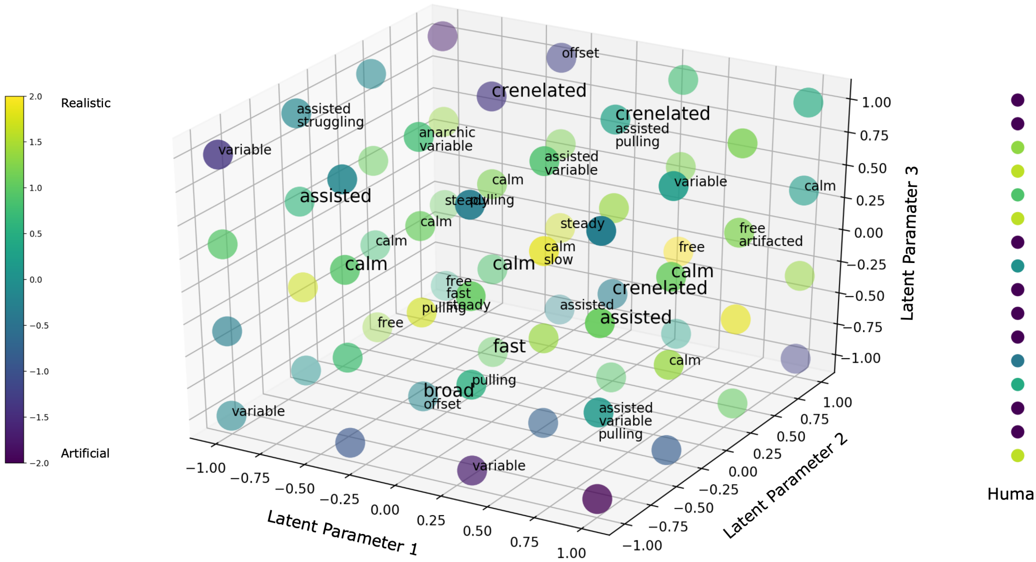


Fig. 6. Center: Mean anthropomorphism assessments for the 64 breathing waveforms generated with GANspire, plus consensus tags extracted from experts' annotations, presented within GANspire's three-dimensional latent space. Bigger and smaller tags respectively correspond to themes by four or three out of the ten experts. Right: Mean anthropomorphism assessments for the 16 human breathing waveforms.

anthropomorphic assessments for the 16 human breathing waveforms are also reported on Figure 6, right.

5.1.2 Experts' Interpretations of Material Expressions. We now look at annotations made by our group of respiratory care experts. Details on annotation strategies and their relationships with extracted tags can be found in Appendixes F & G. As described in Section 3.3, the first, second and fourth authors collaboratively crafted GANspire to research the joint materialities of breathing waveforms and generative deep learning, rather than to design interpretable controls for breathing generation. Thus, our goal here is to understand whether respiratory care experts managed to find a consensus in their interpretations of generated breathing waveforms, be they realistic or artificial.

Counting tag occurrences revealed some consensus in interpretations across experts. Figure 6 shows the tags for which three or four out of the ten experts agreed upon. These include breathing aspects such as type of activity (assisted, free, broad, slow, fast), physiopathology (pulling, struggling, anarchic), emotional state (calm, steady), as well as indeterminate artifacts (crenelated, offset, variable, artifacted). Importantly, these consensus tags were extracted regardless of the degree of realism or artificiality of generated breathing waveforms. This shows that both fine-grained attributes of human breathing waveforms and material expressions of generative deep learning can produce a wide range of interpretations for respiratory care experts. In addition, the scope of these consensus tags is broader than that covered by the three types of activities (exercise, rest or sleep) represented in our training dataset, as described in Section 3.1. This suggests that leveraging qualitative annotations of respiratory care experts after training of a deep generative model enables to discover specific attributes of generated breathing waveforms.

Looking back at the visualisation, we observed that tags such as artifacted, crenelated or struggling, seem to be found at the edges of the latent space. These may probably stand for strange artifacts resulting from material expressions of generative deep learning, since breathing waveforms assessed as artificial were also located at these edges, as anthropomorphism assessments reported in Section 5.1 showed. Yet, other tags located at the center of the latent space, such as calm, fast, free or assisted, does not seem to be ordered in a meaningful way at first glance. On the one hand, this suggests that interpreting breathing waveforms is a highly subjective task, as it is often led through tacit practice rather than verbal annotations, as explained in Section 4.3, but also, as our respiratory care experts all have different clinical experiences and trainings, as described in Section 4.2. On the other hand, this could suggest that leveraging material expressions of generative deep learning can lead to a loss of interpretability, in the sense that ambiguous breathing waveforms cannot lead to one shared interpretation across all respiratory care experts. Rather than opposite, we see these two suggestions as complementary, as it reveals that deep generative models cannot be evaluated universally, but rather contextually.

Overall, our findings shows that GANspire enables to reproduce fine-grained attributes of human breathing waveforms, while also showcasing material expressions of generative deep learning. Interpretations of physiology, activity and emotion from respiratory care experts suggest that these joint materialities could open up novel perspectives for breath-based interaction design.

5.2 Semi-Structured Interviews

In this section, we report findings from the semi-structured interviews we led with our group of ten respiratory care experts, based on the thematic analysis approach described in Section 4.4. Specifically, we detail experts' clinical practices with breathing waveforms, as well as their broader perceptions, expectations, and ethical concerns about GANspire.

5.2.1 Current Clinical Practices of Breathing Waveforms. The first theme relates to current clinical practices of respiratory care experts with breathing waveforms. All experts provided broad and rich descriptions of the context in which their practice with breathing waveforms situate. A first contextual factor for human breathing waveforms is mechanical ventilators. Mechanical ventilators are used in patients who suffer from respiratory failure, namely, who are physically unable to breathe adequately with respect to the metabolic needs of their body. Some of our group's experts testified that they never observe spontaneous breathing in humans: *"I see breathing waveforms all the time, but it's only people who are connected to a machine"* (P2). One expert detailed how our annotation study pushed them to question their own practice:

> *"In fact, I read a lot of waveforms, but of patients who are ventilated [...]. I don't see many waveforms of spontaneous breathing [...]. In fact, by reading the [generated] waveforms, I realised that I was very far from the... Well... Finally, the normality. I almost can't even see it. Well I have some doubts, but I had difficulties in identifying it, necessarily. Maybe that's why [the annotation study] took me so long eventually, because I was asking myself: 'Well, are these normal or not normal waveforms?'"* — P9

In addition, experts reported that analysing a breathing waveform is systematically put in relation with observing the patient who produced it. A patient's breathing often relates to a series of physical tests, the corresponding thoracoabdominal dynamics, their facial expressions while breathing, but also age, height, weight or gender: *"We're using all of this to analyse how patients are breathing, yes, absolutely."*, said one expert. They added that they may not even need to look at breathing waveforms to actually care for a patient's breathing:

> *"I actually come from the last century, and it was a time when you watched people breathe, and you listened to them breathe. So I tend to continue to use my eyes, my ears, and sometimes even touch the abdomen with my hand, for example to feel it inflate and deflate."* — P3

Beyond the body, respiratory care experts may put breathing waveforms in relation to the patient's social context. In fact, clinical questionnaires, examinations and effort tests underwent in functional explorations of breathing may not be enough to identify causes for certain breathing disorders. Depending on the type of practice and care institution, experts may inquire into the patient's social context and psychological environment to interpret their breathing. These clinical observations may in turn inform research work on the analysis of breathing waveform. *"There are a lot of factors"*, said one of the experts, with humility. They continued:

> *"It's not just the patient who's breathing, they have an environment, a family [...], so it's true that it's not something very fixed like that, but it goes beyond, really, the simple illness. We don't have a simple illness in front of us: we have a human being, with their illness, but in a social context, the psychological environment, and we pay a lot of attention to that."* — P11

Hence, presenting our group of experts with breathing waveforms generated with an AI algorithm produced a shift in their clinical practice of respiratory care, since it did not provide any bodily or social context about breathing, but a new, technological one.

5.2.2 Perceptions and Positionalities Toward AI Technology. The second theme relates to the perceptions and positionalities of respiratory care experts toward technology. Our analysis documents that experts witnessed a growing research interest for "artificial intelligence" across various healthcare domains, spanning radiology, ophtalmology, dermatology and pulmonology: *"it's all over the place"*, as put by P1. The most frequently-cited application was automatic disease diagnosis (*e.g.*, detecting an anomaly in a chest x-ray), with an exception being critical decision-making (*e.g.*, deciding on a volume expansion for a patient).

For experts with no research experience with deep learning, these impressive applications yielded personal interpretations for the notion of AI. Far from technical, such perceptions remained at a symbolic level, typically evoking aspects of automation toward data analysis and interpretation, to a somehow greater degree than that provided by machines that already operate in clinical contexts. One expert testified of their *"vague"* perception of AI by describing a mechanical ventilator, widely used in clinical practice, which automatically adapts ventilation to patients, but that is not perceived as AI yet:

> *"There are programs that calculate, that do a bit of the work that we do. They exist. We tell them how much volume we want, and they see in relation to the patient's breathing, they change all the pressures, inspiratory, expiratory, so they can vary everything. [...] It's not necessarily AI, but it's really the machine that decides more or less. We give them a minimum and a maximum pressure where to go, and the machine decides."* — P10

For experts that were senior physicians, these impressive results provoked skeptical reactions about the technology's potential to switch from research to clinical contexts. As one expert explained, machines that currently provide automation in health-related tasks do not adapt well to the improvisational nature of clinical practice: *"We're still very, very, very much based on intuition. The gut feeling, you know"* (P1). The above-quoted expert responded with a critical yet pragmatic posture about this potential:

> *"I still don't see the relevance of this, because for me, I still find it difficult to say that it will replace the clinical impression of a caring, benevolent doctor who has some experience in managing problems and who wants to improve the patient and manage their problems. So there you have it. Moderate level of conviction, it's in its infancy, it's simmering, but I need to be shown that a machine works well. That's it."* — P1

Interestingly, one reason for such a skepticism stemmed from previous experiences collaborating with deep learning engineers. One of the senior physicians reported communication issues with computer scientists: *"For the moment we're having trouble with AI, because basically you have to work with AI researchers, and... It's complicated"*, they uttered with a sigh (P2). Specifically, these communication issues involved an important loss of time to converge to an appropriate design for the technology. Furthermore, the resulting design had very limited abilities compared to the promises sketched by deep learning and its engineers. As one expert explained, this skepticism emerged beyond mere symbolic perceptions of AI, but through the technical understanding of the basics of deep learning (*e.g.*, data-driven pattern recognition):

> *"I learned a lot from mathematicians. I learned a lot about what AI was, and especially what it was not. I understood the limits. It's funny because it's quite paradoxical. By demystifying a little what AI is, by realising that it's all about algorithms, and by knowing these limits a little better, I understand much better the use I'm going to make of it in the future. It is much clearer for me, much clearer."* — P8

Hence, interpretations of breathing waveforms generated with GANspire were probably influenced by the diverse perceptions our respiratory care experts have of technology. In addition, we observe that our probe enabled to establish an open conversation between us and our group of respiratory care experts, since they felt comfortable sharing difficulties in overcoming cultural distances when collaborating with AI researchers.

5.2.3 Expectations Toward Materialities of GANspire. The third theme relates to expectations on GANspire. During the annotation study, some experts felt that many of the generated breathing waveforms they had to assess presented quite similar shapes. Interestingly, this perception led them to judge our deep generative model of breathing waveforms as *"incomplete"*. As one expert put it, *"In the series, some waveforms were missing, that are typically seen in patients with breathing diseases"* (P6). This quote suggests that respiratory care experts had a presupposition that GANspire should provide a universal representation of all breathing waveforms possibly produced by humans.

In a complementary way, one expert, who assessed generated breathing waveforms as more artificial than human ones (P8, as reported in Section 5.1.1), was disturbed by strange attributes of some waveforms generated with GANspire: *"what bothered me a lot was that some waveforms don't look at all like the physiological waveforms of a human"*. Despite this, they pointed at the materiality of GANspire as being conveyor of some form of machine-specific *personnalité ventilatoire*. The quote below testifies of this tension between the engineering-oriented expectation that GANspire should provide with perfect reproductions of human breathing waveforms, and the perception that its materiality could provide with design opportunities for breathing waveforms:

> *"There must be a problem in the creation of the algorithm, because it is a waveform which is, for us for example, a very abnormal waveform, which could be from machines to leaks... It looks more like a turbine... What you have reproduced there looks more like a machine, like a turbine, where there is not too much resistance in front of it, than a human."* — P8

Still, some experts seemed to understand the practical relevance of finding a compromise between representativity and materiality. One expert expressed their uncertainty toward the research method to adopt with generative deep learning to discover and design empathic attributes of human-machine interaction. They quoted surprising results from related research articles in the field of respiratory neuroscience [2], which we previously presented in Section 2.2:

> *"Given data we have on avatars that are synchronous or not with the patient's breathing, we'll have to determine how to do it so that we're not counter-productive on this object, to*

> *see if we... Because it was counter-intuitive for me, the fact that absence of [breathing] synchronisation would reduce dyspnoea... So yes, there remains work to do."* — P7

Hence, our group of respiratory care experts predominantly expected that GANspire provided a universal representation of human breathing waveforms, yet they also expected that ambiguous breathing waveforms could be of interest for respiratory care focusing on mindfulness. Our probe thus enabled to provoke inspirational reactions in our group of respiratory care experts, discovering and debating new design opportunities offered by materialities of generative deep learning.

5.2.4 Ethical Concerns on Speculative Clinical Applications. The fourth and last theme relates to ethical concerns about clinical applications of GANspire, based on the speculative intervention for respiratory care that we described in Section 4.3.2.

First, the training process of GANspire, building on human breathing recordings harvested in previous research experiments, raised no problem for our group of respiratory care experts. Experts were aware of data practices in healthcare research, and mentioned that people inevitably signed an informed consent before recording their breathing, stipulating of its future use for diverse research purposes. Some of them were also aware of technical apparatuses for breathing waveform analysis, and certified that no computer system enables to retrieve a person's identity based on a sole breathing waveform. Thus, anonymisation was guaranteed for both human breathing waveforms used for training of, and breathing waveforms generated with, GANspire. One expert added that GANspire was not intended to make money, and therefore its training process did not raise ethical issues related to the exploitation of human data.

From a clinical perspective, respiratory care experts suggested that the speculative health intervention based on GANspire (described in Section 4.3.2) was scientifically sound, and could be needed to alleviate the current lack of material means to take care of patients with chronic respiratory disorders: *"In my opinion, there is no answer that is given to them today"*, one of them said with confidence (P4). Specifically, most experts witnessed in their clinical practice the empathic distress that patients with dyspnoea convey in healthy persons, resulting in both caregivers and family members running away from these patients. Beyond circumventing such running away issue, the speculative intervention could be used to raise awareness on what dyspnoea actually is among families and communities of care: *"So the idea is to be able to program different types of breathing, right? Calm and soothing breaths for the dyspneic patient, but also jerky, polypneic breaths, to realise what dyspnea is?"*, as one expert suggested. They continued by explaining the shortcomings of current material strategies used to appease patients with dyspnoea, suggesting an opportunity for breath-based interaction design:

> *"We offer them small solutions, whether it's medication, or small ventilators so that they can feel the fresh air on their face, which lowers their dyspnoea threshold... We have Lasilix theophylline aerosols [for asthma, NDT], which in theory are supposed to work, but in practice I'm not at all convinced of. [...] And so in fact there is a place to take."* — P4

From a socio-economic perspective, two of these experts said that the automation of respiratory care potentially supported by the speculative health intervention could help, not replace, respiratory care experts: talking about patients with dyspnoea, P6 commented that *"There are 24 hours in a day, there are seven days in a week, so there is, I think, plenty of room for it to be done and to be a plus, and not an 'instead of'."* In contrast, one senior physician expressed a more cynical perspective about the fact that caregivers would not be replaced, as according to them, fewer and fewer people want to give care: *"Well, yes, there is a big crisis of vocations for the hospital"*, they coldly noted (P2). Another expert supported this argument by noting that automated care is already employed as an imperfect but necessary solution in less privileged socio-economic contexts than France:

> *"[In Colombia,] they follow patients' progress once a year, if not less. In the case of diseases that deteriorate very quickly, I think that [an automatically-adapting ventilator] is the best, so that we can be sure that at least, even if it's badly done, at least it adapts more or less during the patient's breathing deterioration. [...] First of all it's a lack of personnel trained in this type of ventilation [...]. And on the other hand, in fact it's a lack of means, so the patients can't afford a specialised consultation."* — P10

Hence, beyond scientific relevance, our probe illuminated complex socio-technical aspects that should be taken into account in future interaction design led with GANspire for clinical practice.

6 DISCUSSION

In this paper, our main contribution lies in providing insight into how generative deep learning may be used not only as a tool to design human-computer interactions, but also as a probe to provoke responses from communities of practice about their current and speculative uses of technology. Here, we deepen our reflections on how the collaborative crafting of GANspire, along with responses from the probe, led to such a contribution. Specifically, we discuss (1) ambiguity of deep generative models as a resource for inquiry, (2) socio-material research with generative deep learning, and (3) design implications for respiratory care.

6.1 Ambiguity of Deep Generative Models as Resource for Inquiry

In our work, we adopted a material approach to generative deep learning, taking inspiration from design research with machine learning, to discover alternatives modes of inquiry than engineering-oriented approaches [6, 11]. In the next sections, we detail how material expressions of deep generative models, and the ambiguities they produce, can be seen as meaning making, encouraging criticality and interdisciplinary imagination on technology in communities of practices, and balancing empirical investigation with inspirational responses along research.

6.1.1 Material Expressions of Deep Generative Models as Meaning Making. As described in Section 3.2, the collaborative crafting approach let the respiratory physician focus on fine-grained attributes of breathing waveforms and produce free-form annotations over them. This process enabled us to discover material attributes produced by generative deep learning over breathing waveforms, typically morphing between different modes of breathing. While usually leveraged in the computational arts [31], we believe that such material expressions of generative deep learning may be approached from a design perspective. Specifically, material expressions of generative deep learning may support all three types of ambiguity distinguished by Gaver *et al.* [23]. In our present work, ambiguity of information was produced through imprecise representations of breathing waveforms yielded by deep learning uncertainties and their inconsistencies compared to human breathing waveforms. Ambiguity of context was produced by the incongruous aspect of witnessing an AI algorithm generate breathing activities instead of automatically analysing those of humans. Lastly, ambiguity of relationship was produced by showcasing imperfections of deep generative models without explaining why, and introducing speculative clinical applications to question responsibility of engineers and health practitioners in their design process. We encourage further research to deepen conceptualisation of generative deep learning ambiguity as a resource for design, thus extending previous work on machine learning uncertainty [6], as well as other experts to embrace these ambiguous material attributes of generative deep learning to produce meaningful technology.

6.1.2 Encouraging Criticality and Interdisciplinary Imagination. As reported in both our collaborative crafting and probe of GANspire, our work showed how ambiguity, resulting from material expressions of generative deep learning, provided space for criticality on technology, both for the respiratory physician and our group of respiratory care experts. In our semi-structured interviews, respiratory care experts systematically sought to connect generated breathing waveforms to social and cultural contexts of their clinical practice. This enabled them to share their positionalities toward technology, along with some critical perspectives on the socio-economic factors that could drive clinical needs for generative deep learning. Presenting experts with ambiguous breathing materials, that is, showcasing uncanny glitches and morphings between human breathings, provided space for interdisciplinary imagination toward technology, helping them speculate on potential applications in clinical contexts based on material engagement with entanglements between humans and technology. Results from our semi-structured interviews go in this sense, as respiratory care experts were able to project themselves into the speculative health intervention that we presented to them, and elicit socio-economic factors that would drive such an application. Future work may further investigate how probes produced with generative deep learning may relate to cultural probes [22], as their materials can be familiar to given communities of people, yet ambiguous based on their information, context and relationship.

6.1.3 Balancing Empirical Investigation With Inspirational Responses. As discussed above, ambiguous probes created with generative deep learning can be fruitful to inquire on a specific community of practice, and their current and speculative uses of technology. Ambiguity also enables to reconfigure the positionality from which one researcher may inquire on a specific community of practice. In our semi-structured interviews reported in Section 5.2, one of the respiratory care experts criticised our work with GANspire, saying that we were not able to provide a universal representation for human breathing waveforms. Another expert told us that it was usually hard to collaborate and to communicate with AI researchers. As such, we were not perceived as machine learning engineers trying to solve experts' problems in respiratory care with cutting-edge technology along our study. Rather, we were perceived as design researchers, who wanted to debate and discuss current and speculative technology for respiratory care, and who shaped their probes according to specific hypotheses, which were seemingly not understood by our group of scientific experts. We believe that such a new, ambiguous relationship between researchers and experts enabled us to harvest inspirational responses from our respiratory care experts. While such affective responses were shown to be useful for design [22], they come with their weaknesses related to empirical investigations usually led in CSCW, typically seeking in-depths studies of practices, communication and collaboration through the use of official questionnaires or longitudinal field studies. Rather than replace them, we would like to suggest that both approach could complement each other, providing rich and broad qualitative insight into the social, collaborative, functional, aesthetic, cultural and political aspects of technology use, as our probe tended to show.

6.2 Toward Socio-Material Research With Generative Deep Learning

In our work, we sought to include diverse levels of collaboration with experts and materials in our design process with generative deep learning. Here, we discuss how our work may trigger new forms of socio-material research with generative deep learning, which could complement its mainstream use as a tool for human-computer interaction. In the next sections, we detail how attending to contextual materialities of datasets, interpreting generation with and for communities, and documenting socio-technical workings of deep learning may enable such research.

6.2.1 Attending to Contextual Materialities of Datasets. As described in Sections 3.1, the collaborative crafting process engaged for the design of GANspire started from the observation and contextualisation of human breathing waveforms. We first engaged in experiencing various signals, ranging from electroencephalograms to airflow waveforms eventually, which all originated from previous research led in our research unit with a small number of human persons. We also exchanged with a respiratory physician and a research engineer to build a scientific understanding of the neural origins of fine-grained variabilities typical from breathing data. We argue that attending and responding to these contextual materialities of our dataset was a crucial step in our work with generative deep learning. Specifically, we suggest that paying attention to the physical, biological, bodily, social and cultural phenomena that produced the materiality of datasets to be modelled with generative deep learning should be an inevitable step for these technologies to be sustainable and take into account fundamental human values. This is in opposition with most engineering-oriented approaches to dataset building, which often reduce the importance of data to focus on model training and performance [43], while also scraping datasets without necessarily having permission to do so [37]. Results harvested with the probe are in line with a material approach to data since the group of respiratory care experts found no ethical problem in our repurposing of an existing breathing dataset, nor in the generation of novel breathing waveforms based on it. Our findings thus extend previous design research with machine learning [45] to the specific case of generative deep learning.

6.2.2 Interpreting Generation With and For Communities. As described in Sections 3.2 and 4.2, our material approach sought to include experts from the community of respiratory care in the interpretations of materials generated with our deep generative model. Specifically, we leveraged the second author's expertise in respiratory medicine to interpret generation along different training experiences. Also, the wide range of interpretations for breathing waveforms produced by respiratory care experts seem to confirm that material attributes of generation cannot be evaluated universally, but rather contextually, using qualitative approaches. Importantly, our probe approach not only enabled to interpret generation with respiratory care experts, but also *for* their community of practice, that is, by discussing and illuminating socio-technical aspects of speculative clinical applications of our deep generative model. Our results thus extends the interest of probes to evaluate deep learning for health applications [9] to the case of deep generative models. Going further, our findings suggest that deep generative models cannot be evaluated based on the sole quantitative metrics used for engineering [4], nor on human judgments metrics used for HCI and well-being [51]. Rather, we believe that the human norms, values and representations that contributed to their material structures and behaviours should always be made explicit, or probed, when designing deep generative models. This is all the more relevant for the context of respiratory care, where complex relationships between patients, families, carers and technology are to be taken into account when designing technology for care [49]. Future work may probe GANspire with other groups of people, including patients with chronic respiratory diseases, as well as persons in patients' environment, to further its socio-material design and interpretation.

6.2.3 Documenting Socio-Technical Workings of Deep Learning. All along our research process, we sought to make explicit the subjective choices and values that we leveraged, both within our interdisciplinary group of designers, pulmonologists and researchers, as well as toward our group of respiratory care experts. Specifically, we reported both enthusiastic and skeptical responses on ambiguous aspects of GANspire from our group of experts, freeing ourselves from the necessity to verify research hypotheses on usability of deep learning, as is often the case in HCI and engineering [51]. We argue that acknowledging such diverse responses enabled us to go beyond mainstream approaches of deep learning as a tool for HCI and uncover socio-technical workings of deep learning. As shown in our semi-structured interviews in Section 5.2, experts expressed their skepticism toward automation of care, along with their pragmatic acceptation of it, based on their scientific knowledge, and the material needs of patients with chronic respiratory diseases. Some of them also elicited socio-economic factors that would drive current technological transformations of respiratory care, such as crisis of vocation for the hospital, or lack of training in clinicians. Certainly, the openness and holism of our group of respiratory care experts actively contributed to this socio-technical uncovering. For example, experts accepted to dedicate time to our probe, despite busy schedules, probably thanks to their curiosity for interdisciplinary research. Also and crucially, most of them got caught up in somaesthetic appreciation of breathing waveforms generated with GANspire, despite their common practice of using computational models to describe breathing phenomena [47], probably through their clinical practice of holistic care and mindfulness-based cognitive therapy [49]. Future socio-material research with generative deep learning may further include cultural and geographical specificities of healthcare practices to better document ethical and political implications raised by clinical deployments of deep learning, potentially revealing corporate infrastructures that underlie these technological transformations, along with their global disrupting of healthcare practice and training.

6.3 Design Implications for Respiratory Care

Beyond socio-material research, our work shed light on design implications to embody breathing waveforms in a soft robotic object to reduce anxiety in patients with chronic respiratory diseases. In the next sections, we deepen the former design implication, and discuss a second design implication related to using GANspire to help teaching clinicians diagnose respiratory diseases.

6.3.1 Reducing Anxiety With a Soft Breathing Object. Our findings showed that GANspire is able to generate a diverse range of breathing waveforms. Qualitative annotations from our group of respiratory care experts suggest that fine-grained variations of these breathing waveforms were interpreted as conveying diverse physiologies, activities and emotions, regardless of them being perceived as realistic or artificial. In future work, we will seek to embody breathing waveforms generated with GANspire in a soft robotic object, thus including fabrication of a pneumatic, inflatable object that enables to transcribe fine-grained variations of generated breathing waveforms. Indeed, previous research suggested that realistic breathing behaviours could be key to improve efficacy of soft robotic objects aimed at attenuating anxiety [28]. As shown in our collaborative crafting in Section 3.3, one advantage of using GANspire over simply replaying human breathing waveforms is that it supports continuous morphings between different breathing classes. Such continuous morphings between breathing classes have been shown crucial to produce sensori-motor effects for respiratory care, such as breathing entrainment [13] or affective impact [2]. Additionally, feedback harvested in our semi-structured interviews and reported in Section 5.2 suggested that there is room to take for technological aids for reducing breathing anxiety in patients with chronic breathing diseases, but that these aids could not—and should not—replace the complex care relationships actively maintained by physicians, nurses and family members within and outside the hospital. As such, our future work will aim at designing specific arrangements for this embodiment that take into account the social and material contexts of respiratory care, for example mixing quantitative measures of breathing entrainment with qualitative aspects of collaboration between carers and machines.

6.3.2 Teaching Clinicians Diagnose Respiratory Diseases. Another possible application of the realistic waveform generation capability of GANspire pertains to the training of respiratory healthcare professionals. Such an application may doubly benefit the CSCW community, through iterative development of GANspire in collaboration with respiratory care experts, and empirical investigation of communication between respiratory care experts and students mediated by generative deep learning. One possible development in this direction would be to enable clinicians to visualise breathing patterns characteristic of certain diseases without having to set individual parameters but rather by using a declarative interface. This would require the collection of actual breathing patterns in well-characterized patients, from which GANspire could be trained in a supervised manner, enabling "family wise" generation of breathing waveforms. While its current latent space remains ambiguous, as experts' annotations have shown in Section 5.1, such future development for GANspire would probably improve universal interpretability of latent space, thus enabling clinical training. Another putative pedagogic application of GANspire pertains to situations where clinicians have to rely on breathing waveforms for diagnostic reasons, or to decide therapeutic actions—some of which have been revealed in our semi-structured interviews in Section 5.2. These situations include home mechanical ventilation, where the analysis of breathing waveforms is the main tool used by clinicians to assess the quality of ventilatory assistance, particularly through the detection of leaks at the level of the patient-ventilator interface. They also include the clinical monitoring of patients receiving mechanical ventilation in the intensive care unit. Another example can be found in sleep medicine, where waveforms are central to the diagnostic of respiratory-related sleep disturbances, and to the evaluation of the effectiveness of treatment. In the above three examples, waveforms can be obtained either from real patient recordings or from simulation tools [3], e.g., high-fidelity mannequins that can be mechanically ventilated [44], and then used for teaching purposes. Future CSCW work may thus seek to iterate GANspire development by building breathing datasets labeled for specific care situations by respiratory care experts, and investigate collaborations between clinicians and students in educative applications mediated with GANspire.

7 CONCLUSION

In this paper, we have described the interpretations, perceptions and expectations of respiratory care experts towards a deep learning tool for breathing waveform generation. Our work suggests that a material approach to generative deep learning could be useful to support interdisciplinary collaboration between pulmonologists, designers and engineers, in the research of relevant attributes for interaction design, along with social and ethical aspects of breath-based technology.

As deep learning applications to healthcare and mindfulness are likely to increase in the next years, we believe that holistic design approaches that value care as a virtue should be considered for building such computational techniques. Our findings contribute to the growing literature that outlines designerly ways of exploring machine learning, negotiating values injected by engineering practices, and illuminating the entanglements between humans and deep learning technology.

ACKNOWLEDGMENTS

We thank all respiratory care experts and research engineer from our medical research unit for their precious time and feedback devoted to our study. This work was funded by a postdoctoral grant from Institut Universitaire d'Ingénierie en Santé (Sorbonne Université). Additional revenues come from two grants from Fondation du Souffle and Fondation de France for the art-science-health project *RÉESPIRATION* led by Samuel Bianchini and Thomas Similowski.

REFERENCES

[1] Dan Adler, Bruno Herbelin, Thomas Similowski, and Olaf Blanke. 2014. Reprint of "Breathing and sense of self: Visuo-respiratory conflicts alter body self-consciousness". *Respiratory physiology & neurobiology* 204 (2014), 131–137.

[2] Etienne Allard, Elisa Canzoneri, Dan Adler, Capucine Morélot-Panzini, Javier Bello-Ruiz, Bruno Herbelin, Olaf Blanke, and Thomas Similowski. 2017. Interferences between breathing, experimental dyspnoea and bodily self-consciousness. *Scientific reports* 7, 1 (2017), 1–11.

[3] Kristin T Avis, David J Lozano, Marjorie L White, Amber Q Youngblood, Lynn Zinkan, Julia M Niebauer, and Nancy M Tofil. 2012. High-fidelity simulation training for sleep technologists in a pediatric sleep disorders center. *Journal of Clinical Sleep Medicine* 8, 1 (2012), 97–101.

[4] Shane Barratt and Rishi Sharma. 2018. A note on the inception score. *arXiv preprint arXiv:1801.01973* 1, 1 (2018), 1–9.

[5] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.

[6] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–14.

[7] Sophie Betka, Elisa Canzoneri, Dan Adler, Bruno Herbelin, Javier Bello-Ruiz, Oliver Alan Kannape, Thomas Similowski, and Olaf Blanke. 2020. Mechanisms of the breathing contribution to bodily self-consciousness in healthy humans: Lessons from machine-assisted breathing? *Psychophysiology* 57, 8 (2020), e13564.

[8] Nico Brand, William Odom, and Samuel Barnett. 2021. A Design Inquiry into Introspective AI: Surfacing Opportunities, Issues, and Paradoxes. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event, USA, 1603–1618.

[9] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. " Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24.

[10] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011* 1, 1 (2021), 1–15.

[11] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. " Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[12] Evangelia Charmandari, Constantine Tsigos, and George Chrousos. 2005. Endocrinology of the stress response. *Annu. Rev. Physiol.* 67 (2005), 259–284.

[13] Marcin Czub and Marta Kowal. 2019. Respiration entrainment in virtual reality by using a breathing avatar. *Cyberpsychology, Behavior, and Social Networking* 22, 7 (2019), 494–499.

[14] Pierre Dejours, P Monzein, J Baynaud, and Bechtell. Y. 1961. Etude de la diversité des régimes ventilatoires chez l'homme. *Journal de Physiologie* 53, 2 (1961), 320.

[15] Christopher A Del Negro, Gregory D Funk, and Jack L Feldman. 2018. Breathing matters. *Nature Reviews Neuroscience* 19, 6 (2018), 351–367.

[16] Rebecca T Disler, Anna Green, Tim Luckett, Phillip J Newton, Sally Inglis, David C Currow, and Patricia M Davidson. 2014. Experience of advanced chronic obstructive pulmonary disease: metasynthesis of qualitative research. *Journal of pain and symptom management* 48, 6 (2014), 1182–1199.

[17] Chris Donahue, Julian McAuley, and Miller Puckette. 2018. Adversarial Audio Synthesis. In *International Conference on Learning Representations*. ICLR, New Orleans, USA, 1–16.

[18] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. CAN: Creative adversarial networks generating "Art" by learning about styles and deviating from style norms. In *8th International Conference on Computational Creativity, ICCC 2017*. ACC, Atlanta, GA, USA, 1–22.

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

[20] Jules Françoise, Baptiste Caramiaux, and Téo Sanchez. 2021. Marcelle: Composing Interactive Machine Learning Workflows and Interfaces. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event, USA, 39–53.

[21] Jérémy Frey, May Grabli, Ronit Slyper, and Jessica R Cauchard. 2018. Breeze: Sharing biofeedback through wearable technologies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Canada.

[22] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.

[23] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, Ft. Lauderdale, Florida, USA, 233–240.

[24] Jesus Gonzalez-Bermejo, Jean-Paul Janssens, Claudio Rabec, Christophe Perrin, Frédéric Lofaso, Bruno Langevin, Annalisa Carlucci, and Manel Lujan. 2019. Framework for patient-ventilator asynchrony during long-term non-invasive ventilation. *Thorax* 74, 7 (2019), 715–717.

[25] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014), 1–9.

[26] E Grossman, A Grossman, MH Schein, R Zimlichman, and B Gavish. 2001. Breathing-control lowers blood pressure. *Journal of human hypertension* 15, 4 (2001), 263–269.

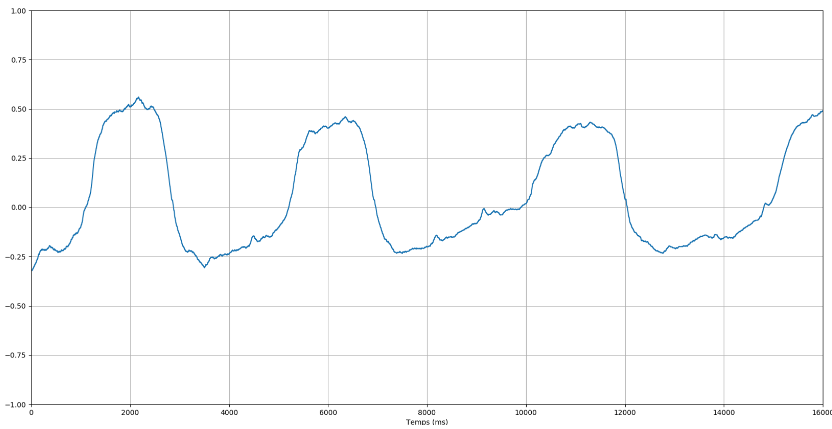
[27] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. GANSpace: Discovering Interpretable GAN Controls. *Advances in Neural Information Processing Systems* 33 (2020), 1–10.

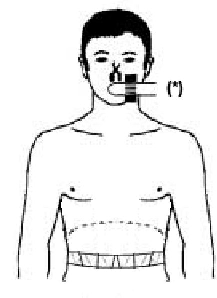
[28] Alice C Haynes, Annie Lywood, Emily M Crowe, Jessica L Fielding, Jonathan M Rossiter, and Christopher Kent. 2022. A calming hug: Design and validation of a tactile aid to ease anxiety. *Plos one* 17, 3 (2022), e0259838.

[29] Debapriya Hazra and Yung-Cheol Byun. 2020. SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation. *Biology* 9, 12 (2020), 441.

[30] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research* 21, 248 (2020), 1–43.
[31] Aaron Hertzmann. 2020. Visual indeterminacy in GAN art. *Leonardo* 53, 4 (2020), 424–428.
[32] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.
[33] Ikuo Homma and Yuri Masaoka. 2008. Breathing rhythms and emotions. *Experimental physiology* 93, 9 (2008), 1011–1021.
[34] Miriam J Johnson, Janelle Yorke, John Hansen-Flaschen, Robert Lansing, Magnus Ekström, Thomas Similowski, and David C Currow. 2017. Towards an expert consensus to delineate a clinical syndrome of chronic breathlessness. *European Respiratory Journal* 49, 5 (2017), 1–8.
[35] Grace Leslie, Asma Ghandeharioun, Diane Zhou, and Rosalind W Picard. 2019. Engineering music to slow breathing and invite relaxed physiology. In *2019 8th international conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, IEEE, Cambridge, UK, 1–7.
[36] Sangeeta Mehta and Nicholas S Hill. 2001. Noninvasive ventilation. *American journal of respiratory and critical care medicine* 163, 2 (2001), 540–577.
[37] Fabio Morreale. 2021. Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation. *Transactions of the International Society for Music Information Retrieval* 4, 1 (2021), 105–113.
[38] Myura Nagendran, Yang Chen, Christopher A Lovejoy, Anthony C Gordon, Matthieu Komorowski, Hugh Harvey, Eric J Topol, John PA Ioannidis, Gary S Collins, and Mahiben Maruthappu. 2020. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *Bmj* 368 (2020), 1–12.
[39] Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsanon, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3, 12 (2021), 1013–1022.
[40] Mirjana Prpa, Kıvanç Tatar, Jules Françoise, Bernhard Riecke, Thecla Schiphorst, and Philippe Pasquier. 2018. Attending to breath: exploring how the cues in a virtual environment guide the attention to breath and shape the quality of experience to support mindfulness. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, Hong Kong, 71–84.
[41] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* 1, 1 (2015), 1–16.
[42] Johan Redström. 2005. On technology as material in design. *Design Philosophy Papers* 3, 2 (2005), 39–54.
[43] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.
[44] Clara J Schroedl, Alexandra Frogameni, Jeffrey H Barsuk, Elaine R Cohen, Lakshmi Sivarajan, and Diane B Wayne. 2021. Impact of simulation-based mastery learning on resident skill managing mechanical ventilators. *ATS scholar* 2, 1 (2021), 34–48.
[45] Hugo Scurto, Baptiste Caramiaux, and Frédéric Bevilacqua. 2021. Prototyping Machine Learning Through Diffractive Art Practice. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event, USA, 2013–2025.
[46] Hugo Scurto, Baptiste Caramiaux, Thomas Similowski, and Samuel Bianchini. 2021. GANspire: Generating breathing waveforms for art-health applications. In *5th NeurIPS Workshop on Machine Learning for Creativity and Design*.
[47] SA Shea and A Guz. 1992. Personnalite ventilatoire—an overview. *Respiration physiology* 87, 3 (1992), 275–291.
[48] Ali Shtarbanov. 2021. FlowIO Development Platform–the Pneumatic "Raspberry Pi" for Soft Robotics. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–6.
[49] Thomas Similowski. 2018. Treat the lungs, fool the brain and appease the mind: towards holistic care of patients who suffer from chronic respiratory diseases.
[50] Jay Vidyarthi, Bernhard E Riecke, and Diane Gromala. 2012. Sonic Cradle: designing for an immersive experience of meditation by connecting respiration to music. In *Proceedings of the designing interactive systems conference*. ACM, Newcastle, UK, 408–417.
[51] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, USA, 3506–3510.
[52] Qian Yang, Nikola Banovic, and John Zimmerman. 2018. Mapping machine learning advances from hci research to reveal starting places for design innovation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal, QC, Canada, 1–11.

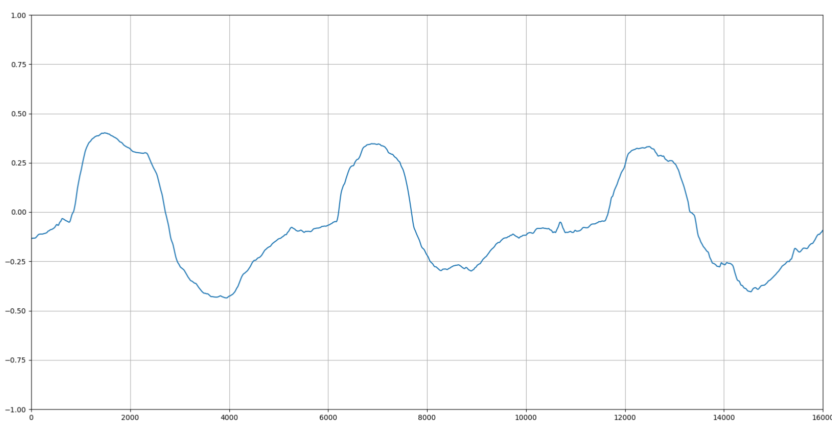
APPENDIX A





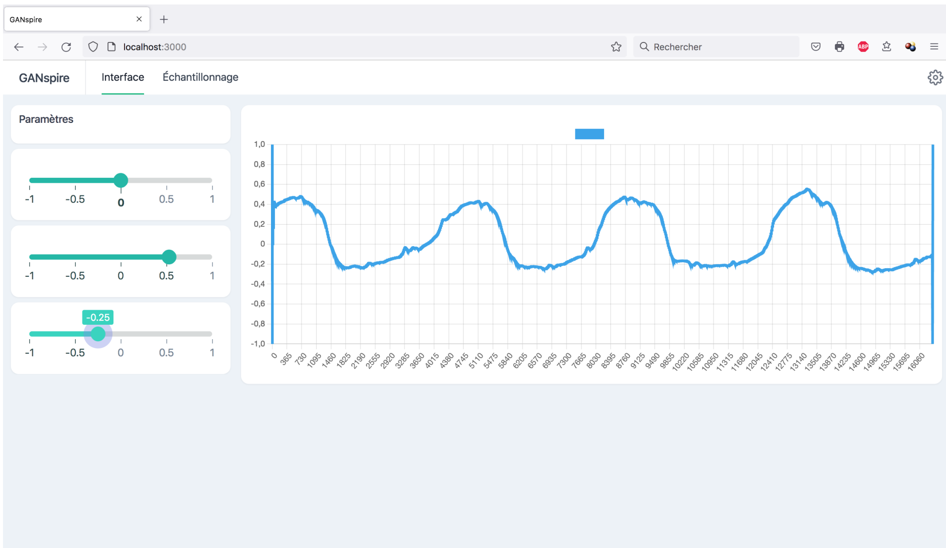
Left: Training example from our human breathing dataset, consisting of a 16-second slice of breathing airflow recording. Right: Pneumotachograph used to record breathing airflows in healthy humans during physiological studies previously led in our health lab.

APPENDIX B

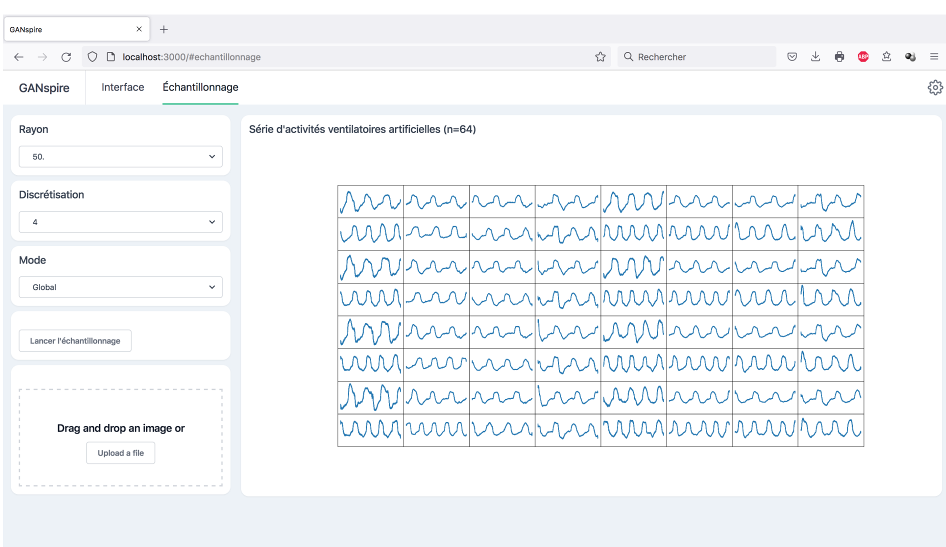


Sample breathing waveform generated with GANspire, consisting of a 16-second artificial breathing airflow.

APPENDIX C

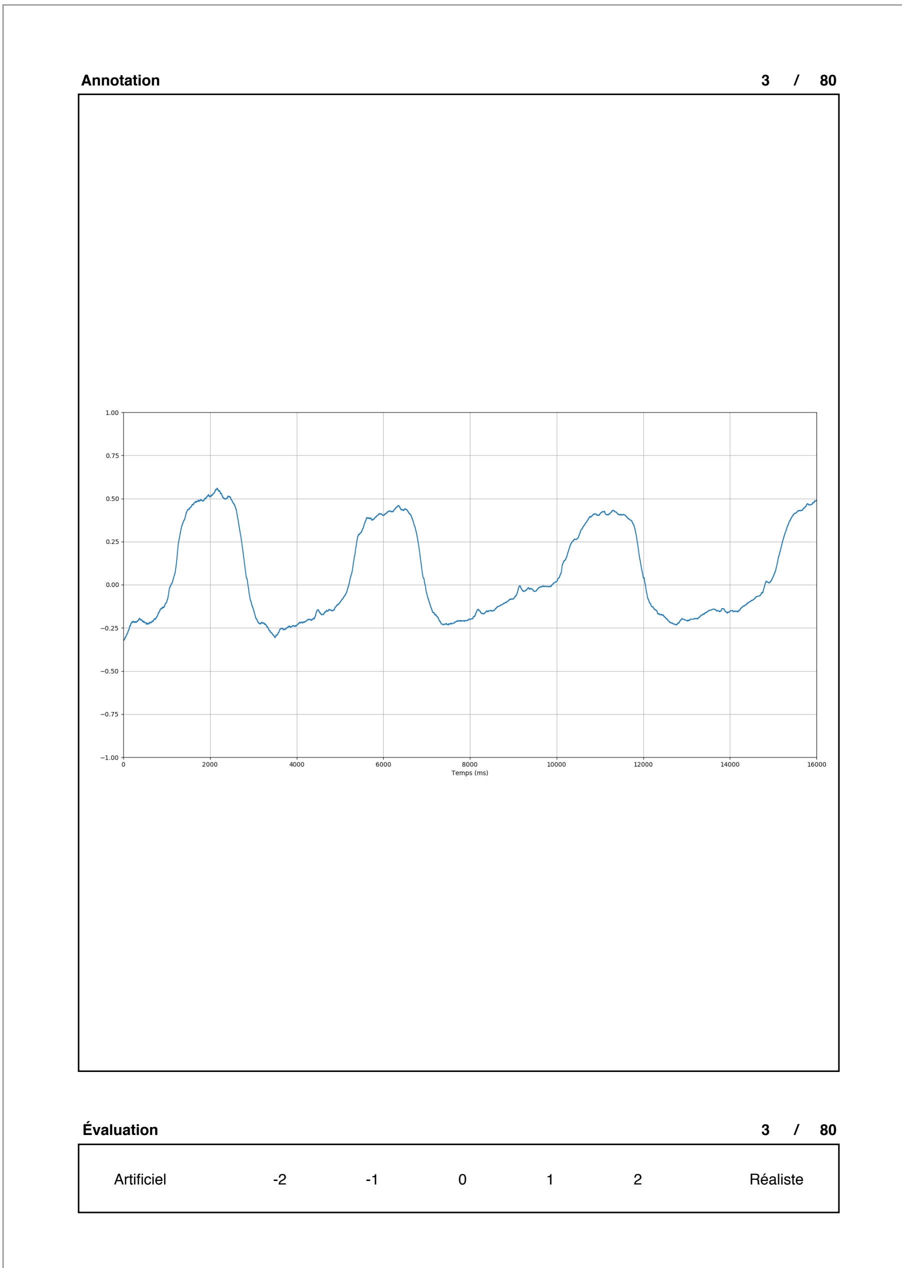


GANspire control interface, first page. The three sliders (left) enables to interactively set latent parameters computed with GANSpace's PCA, and as such, to explore diverse breathing waveforms (right) generated with our trained WaveGAN in real-time.



GANspire control interface, second page. The three menus (left) respectively enable to set sampling step for the latent space, parameter range for the PCA, and plot mode for the resulting batch of generated breathing waveforms (right).

APPENDIX D



Example empty form provided to experts for our annotation study. The main panel let experts freely annotate the breathing waveform. The lower panel let experts assess the breathing waveform's anthropomorphism on a 5-point Likert scale.

APPENDIX E

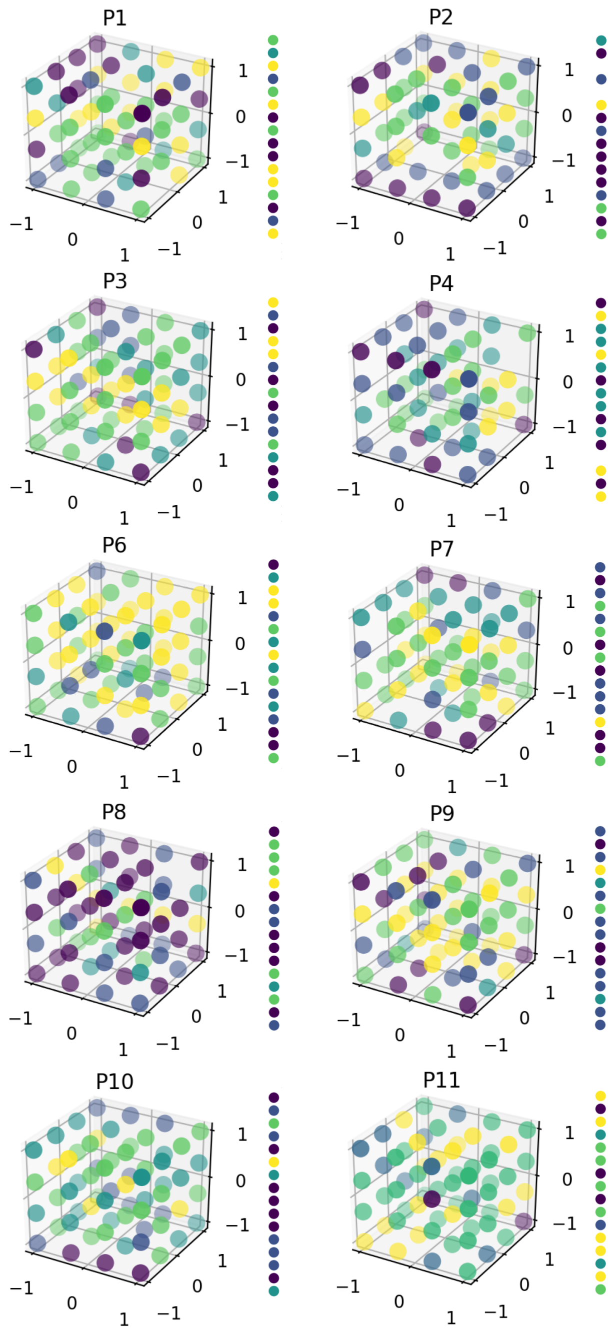


Fig. 8. Anthropomorphism assessments for each of our 10 experts, presented within GANspire's three-dimensional latent space for the 64 generated waveforms, and vertically for the 16 human breathing waveforms.

APPENDIX F

Here, we look at the time taken by experts to annotate breathing materials, as well as a measurement of the number of words they used for annotation. Experts spent between one and three hours annotating the breathing waveforms and assessing their anthropomorphism (μ = 122 minutes, σ = 42). This was more than the one hour that we expected and recommended to them, as described in Section 4.3.1. Table 1 shows mean number of words and standard deviations in the annotations. First, no correlation was found between number of words and time spent annotating breathing waveforms ($r = 0.4, p = 0.2$). Second, the relatively low number of words suggests that experts produced shorter annotations than the respiratory physician during collaborative crafting, which produced rich annotations, as reported in Section 3.2. In short, several idiosyncratic strategies were adopted by our group of respiratory care experts to annotate breathing waveforms, rather than a standard and shared one.

	P1	P2	P3	P4	P6
Time (minutes)	120	120	90	90	120
# of annot. ($\mu \pm \sigma$)	19 ± 14	6 ± 5	7 ± 5	5 ± 3	4 ± 5
# of tags ($\mu \pm \sigma$)	3.36 ± 1.37	1.21 ± 1.08	1.30 ± 1.05	2.14 ± 1.59	0.525 ± 0.821
Strategy (main)	qual.	qual.	quant.	quant.	qual.
	P7	P8	P9	P10	P11
Time (minutes)	105	150	210	150	60
# of annot. ($\mu \pm \sigma$)	7 ± 4	26 ± 14	8 ± 5	9 ± 4	0 ± 0
# of tags ($\mu \pm \sigma$)	1.88 ± 1.08	2.33 ± 1.24	2.71 ± 1.38	2.79 ± 0.904	0.038 ± 0.333
Strategy (main)	quant.	qual.	qual.	quant.	qual.

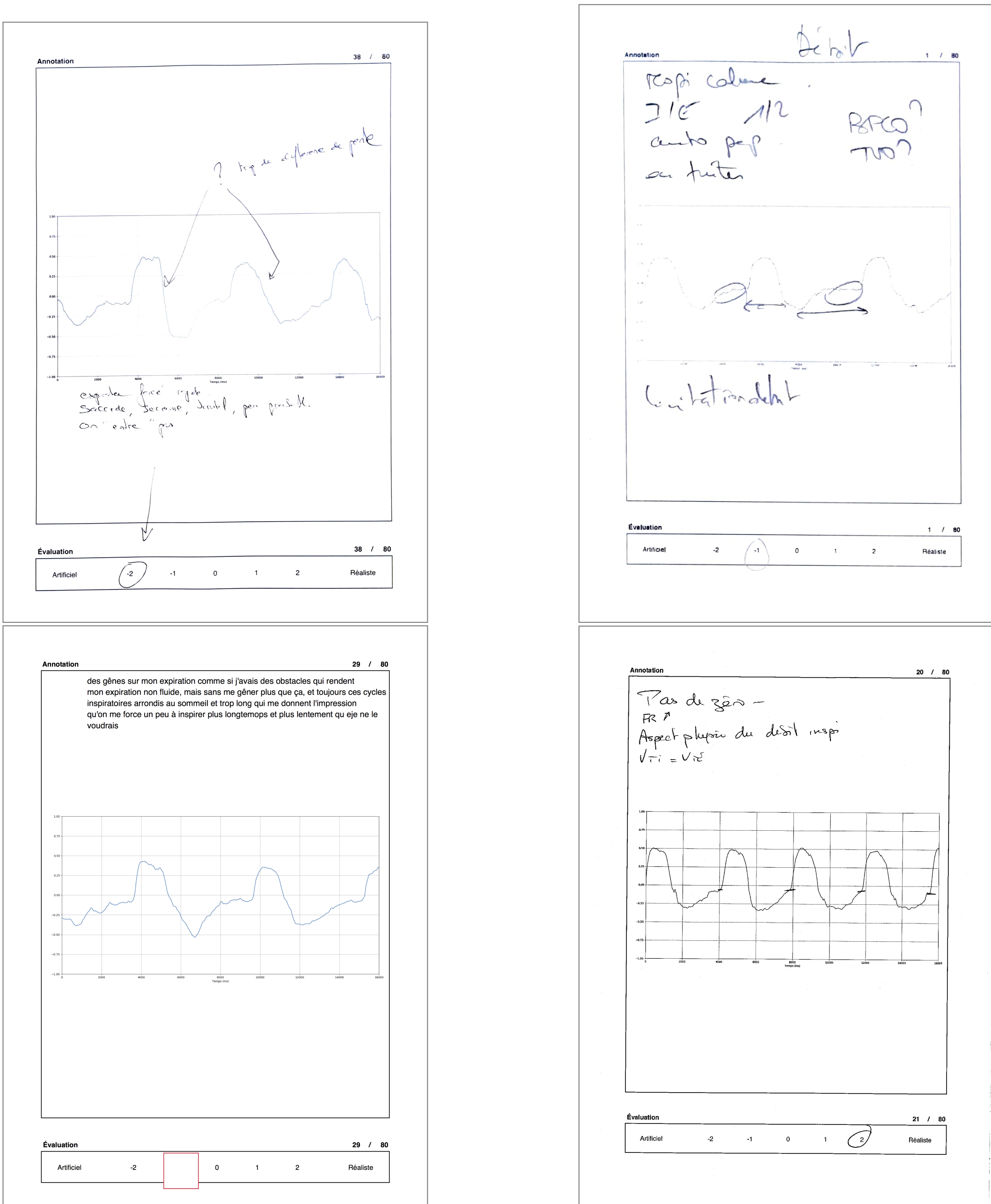
Table 1. Descriptive statistics of annotation study for each of our ten respiratory care experts.

Reading through all experts' annotations, we observed a continuum of strategies, which could be placed anywhere between the following two endpoints:

- The first is a **qualitative strategy**, where experts interpreted emotional or personal aspects of breathing by adopting a somaesthetic approach, typically imagining the temporal evolution of breathing within a patient's body, or even leading to breathing reenactments. Example annotations are shown in Appendix F and included: *"feeling of discomfort on my exhalation, as if I had obstacles that made my exhalation not fluid, but without bothering me more than that, and always these inspiratory cycles rounded to sleep and too long that give me the impression that I am being forced to breathe in a little longer and slower than I would like to"* (P8); *"forced exhalation, rigid. jerk, jolt, abrupt, unlikely. one doesn't 'go into it'."* (P1).
- The second is a **quantitative strategy**, where experts interpreted potential respiratory disorders by measuring physiological quantities, such as airflow frequencies and amplitudes, or inspiratory and expiratory volume ratios. Example annotations are shown in Appendix F and included: *"calm breath. $I/E = 1/2$. Auto-pep or leaks. BFCO? TVO? Flow restriction $\longleftrightarrow$"* (P4); *"No zero. FR $\nearrow$. Physio. aspect of inspi. $V_{T_i} = V_{T_e}$"* (P7).

These approaches seemed consistent for each expert (see Table 1). They are also in line with strategies observed in our collaborative crafting, as reported in Section 3.2, in which the respiratory physician alternated between a quantitative and a qualitative strategy to annotation.

APPENDIX G



Differences in Annotation Strategies. Left: Qualitative approaches from P1 (top) and P8 (bottom). Right: Quantitative approaches from P4 (top) and P7 (bottom).

Received January 2022; revised July 2022; revised April 2023; accepted July 2023