Routledge
Taylor & Francis Group

Check for updates

Tuning sound for infrastructures: artificial intelligence, automation, and the cultural politics of audio mastering

Jonathan Sterne[a] and Elena Razlogova[b]

[a]Department of Art History and Communication Studies, McGill University, Montreal, Canada; [b]Department of History, Concordia University, Montreal, Canada

ABSTRACT
This paper traces the infrastructural politics of automated music mastering to reveal how contemporary iterations of artificial intelligence (AI) shape cultural production. The paper examines the emergence of LANDR, an online platform that offers automated music mastering, built on top of supervised machine learning branded as artificial intelligence. Increasingly, machine learning will become an integral part of signal processing for sounds and images, shaping the way media cultures sound, look, and feel. While LANDR is a product of the so-called 'big bang' in machine learning, it could not exist without specific conditions: specific kinds of commensurable data, as well as specific aesthetic and industrial conditions. Mastering, in turn, has become an indispensable but understudied part of music circulation as an infrastructural practice. Here we analyze the intersecting histories of machine learning and mastering, as well as LANDR's failure at automating other domains of audio engineering. By doing so, we critique the discourse of AI's inevitability and show the ways in which machine learning must frame or reframe cultural and aesthetic practices in order to automate them, in service of digital distribution, recognition, and recommendation infrastructures.

The ad copy for LANDR's artificial intelligence-based music mastering promises a lot: 'think self-driving cars and Shazam' (About LANDR n.d.). LANDR is a new media company based in Montreal but now with international offices – and $10.4 million CAD in venture capital investment (LANDR n.d.). Alongside some fairly common services like music distribution assistance, they offer online, automated music mastering, built on top of supervised machine learning branded as artificial intelligence (AI). Their audacious claim is that they can automate the aesthetic decisions of a highly specialized class of audio professionals, the way self-driving cars promise to automate the decisions of drivers. And just like self-driving cars, the process of automation is much

---

CONTACT Jonathan Sterne ✉ jonathan.sterne@mcgill.ca

more technically, culturally, and ethically complex than it might initially seem (Crawford and Joler 2018). At the end of their study of the Netflix Prize, Blake Hallinan and Ted Striphas (2016, 131) ask 'what happens when engineers – or their algorithms – become important arbiters of culture, much like art, film, and literary critics?' In this issue Jeremy Morris examines the influence of 'infrastructures of discovery' on cultural values in podcasting and Nick Seaver analyzes the 'infrastructural point of view' of engineers who develop music recommendation systems. In this paper, we focus on cultural *production*: we examine the less studied role of automated mastering in music circulation as an infrastructural practice. Increasingly, machine learning will play a role in the sound and look of the media around us (Manovich 2018). While contemporary discourse treats AI as an inevitable creative force beyond human control (Winner 1977), the political consequences (Edwards *et al.* 2009) of this transformation can only be seen if we understand the concrete ways in which machine learning reframes aesthetic practices and cultural values.

Mastering can best be understood as the audio equivalent of typesetting and the creation of page proofs for a publication. Every media text bears the mark of its anticipated modes of circulation and audio is no different. As Mandy Parnell, an engineer who has mastered recordings for artists like Björk, Feist, The XX, and Tim Hecker explains, 'we need to make it fit inside the world. How is it going to sound on the radio or on a playlist?' (Parnell 2017). A mastering engineer usually receives a stereo recording (or in the case of video game or film audio, a multichannel mix) and then prepares it for translation and circulation by manipulating aspects of its timbre, loudness, and stereo image (among other things). Mastering engineers mediate between art and formats by bringing an 'objective set of ears' to a recording (Parnell 2017, here objectivity should be understood as a professional construct, not a realized fact), a result of artistic and institutional distance from recording project. They are stand-ins (Mulvin, 2021) for future playback scenarios and audiences, translating a recording so that it would work across different across the variety of reproduction systems that would play recorded music (Auld 2004; Shelvock 2012: 9-10; Rumsey 2011; Katz 2015).

To support its claim as a stand-in for mastering engineers, LANDR uses the term 'artificial intelligence' to describe what they do. In fact, LANDR's technology is built around supervised machine learning. This semantic slippage is part of the new media industrial landscape at the time of writing. Within the field of computer science, machine learning is one particular kind of approach to artificial intelligence, and even the kind of machine learning currently in vogue represents only one approach within the field (Langley 2011; LeCun et al. 2015; Schmidhuber 2015). As the term AI has become more commercially useful in recent years, it is often used interchangeably with, or instead of, the specific kind of machine learning that companies and labs

are doing. For instance, since 2016, tech companies like Microsoft and Google, and laboratories like the MIT Media Lab have begun renaming their machine learning operations as 'artificial intelligence' operations (e.g. Darling 2017; Lundgren 2018)[1], and claiming AI research as one of their central priorities (Schwartz 2018).

Despite its ubiquity, the emerging roles of AI in cultural production are only now coming under critical scrutiny. To contribute to this project, we consider the intersecting histories of machine learning and mastering, and ask why LANDR's service is even possible at this historical conjuncture. Such a question might seem obvious, but we aim to trouble exactly that obviousness. As of yet, relatively little scholarship on AI has questioned the rhetoric of inevitability that surrounds it, despite similar work in other software and new media contexts (see, e.g. Wyatt 2004; Turner 2006; Zuboff 2019). However, the technical, infrastructural, industrial, and cultural conditions have to exist for AI to operate, and users have to be convinced to use it. While much social and advertising discourse trumpets AI as an inevitable force, all AI applications depend on cultural, technical, and industrial conditions of possibility that cannot be known in advance (Slack and Wise, 2014).

For the purposes of this article, we approach LANDR from the standpoint of what a recording sounds like, in isolation from other factors like how it works and how it fits into the social and sonic practices of musicians and audio producers. From this standpoint, LANDR can process individual recordings in such a way that listeners may hear them as 'mastered,' though not in every case, with some limitations that we discuss below, and not exactly in the same way that a human engineer would. For instance, mastering that is more attentive to questions of artistic intent might matter more for a piece of music meant to be heard as art than a radio ad that is not. Musicians with no financial backing, or who are churning out material, may find LANDR to be a cost-efficient solution. Other users and observers, including some audio professionals, do not. For instance, Larry Crane, editor of *Tape Op* (the world's largest circulation recording magazine), and a working engineer and producer, argues that LANDR devalues the creative part of audio work, whether by amateurs or professionals; the company intentionally simplifies a complex process in order to claim they can do it themselves (author interview). Like recommendation engines, LANDR makes judgments about the sonic character of the music in order to suggest where it might fit into a musical world. Unlike most recommendation engines, LANDR makes these judgments based on the sound of the music rather than from user data or metadata. LANDR's success is defined in part by limiting the problems it tries to solve while finding ways to market new uses for its products, offering a new chapter in the history of mastering, a set of audio techniques that have evolved alongside the history of audio formats. To do its work, it must intervene in the 'technological imaginations' (Balsamo 2011) of audio, mastering, and AI.

Recent work in infrastructure studies has turned scholarly attention to the question of circulation and its conditions of possibility. Mastering is about shaping sound in anticipation of its future modes of circulation, using signal processing to 'tune' music to the possible conditions of its future transmission, storage, and audition. As Lisa Parks and Nicole Starosielski argue, ' a focus on infrastructure brings into relief the unique materialities of media distribution – the resources, technologies, labor, and relation that are required to shape, energize, and sustain the distribution of audiovisual signal traffic on global, national, and local scales' (2015, 6; see also Jones 2000). Machine learning is, among other things, a fundamentally infrastructural question, relying as it does on massive stores of data, 'naturalized' artificial standards for media (Bratton 2016, 45), and tremendous resources of human labour and processing power that come with equally massive costs. As Kate Crawford and Vladen Joler write, 'a full accounting for these costs is almost impossible, but it is increasingly important that we grasp the scale and scope if we are to understand and govern the technical infrastructures that thread through our lives' (2018, II). LANDR is much more than simply mastering + machine learning; it is a whole set of techniques, practices, and mediations (Born 2005).

From the big Bang of AI to LANDR

For a person new to audio production, there are many more resources available to learn to mix music than to master it. There are countless books, tutorials, and educational programmes dedicated to mixing music. Mastering has many fewer educational resources available to amateurs. Yet LANDR's corporate story is an interesting inversion of this skill hierarchy: for them, it was easier to use machine learning to master recordings than to mix them. This difference elucidates much about the cultural politics and context of AI.

LANDR comes at a particular point in the history of machine learning and would not have been possible before. But it also comes at a particular moment in the history of mastering and audio production. From a machine learning perspective, LANDR would not have been practically possible much earlier. Simultaneously with LANDR, other companies – notably Cloud-Bounce and Izotope – also began to apply machine learning to mastering, making it a classic case of what Robert Merton calls 'simultaneous discovery' (1973: 371).[2] Yet this is only half the story. Machine learning cannot simply enter an industry because it exists – one of the great conceits of today's AI hype is that any intellectual or cultural task can be automated via machine learning. This is far from reality. Rather, machine learning – and its industrial, aesthetic, and operational dimensions – can only work for certain kinds of techniques and industries, and only if it defines its tasks in a certain way. Just as 'friend' takes on a certain meaning in social media contexts that is

not exactly the same as what 'friend' meant before social media (boyd 2006), 'mastering' means different things depending on whether we are talking about what LANDR does or what a mastering engineer does.

The boom in artificial intelligence technology – and AI investment – in recent years has resulted from a variety of factors. Popular accounts often refer to 'Big Bang' moments like Google's application of NVidia graphics processor units (GPUs) to neural networks in 2009 and the 2011 application of a convolutional neural network to the problem of image recognition (Alpaydin 2014: 15-18; LeCun et al 2015; Schidhuber 2015; Wang and Raj 2017, 39-44).[3] Yet a number of other crucial social factors also clearly play a role. The neural networks through which machine learning applications run require vast amounts of data. These data have to be generated by people or organizations, often for free or as a side-effect of other processes, and they have to be in a form available for processing (boyd and Crawford 2012; Levy 2015; Radin 2017). In other words, they are only available because of a more fundamental infrastructural system where a wide range of commensurable data can be collected from people and then transmitted, circulated, collected, and repurposed. Institutional imperatives also play a role, both at the university level and in the world of venture capital (Hoffman 2017). In other words, the 'Big Bang' of machine learning is as much social and industrial as it is technical. Without a massive institutional interlocking of data, infrastructure, and the everyday practices of users, there would be no AI as we currently know it today.

In the case of audio, the vast number of digital recordings in existence would seem to make it a perfect resource for training and testing machine learning applications. But sound – especially the sound of music – has been a difficult area for artificial intelligence. Although speech recognition and machine translation have improved in recent years, the most successful music recommendation engines are still built around hand-coded genre tags, and data built off of user behaviour rather than conclusions derived from sound (Eck 2014; Hallinan and Striphas 2016; Durham 2018; for a history of speech recognition's early years, see Li and Mills 2019). More recent experiments have moved into the use of neural networks, big data sets, and machine learning (Lewis 1999, 2018; Dubnov and Surges 2014; Collins 2016a, 2016b) for composition, and companies like Spotify have revealed an interest in AI-assisted and AI-based based composition (Titlow 2017). But as of yet, that has not happened on a large commercial scale.

LANDR's emergence follows from the broader context of increasing technical capacity, expertise, and interest combined with the generation and acquisition of large data sets for free or cheap. In 2009-10, members of the Music Department at Queen Mary University, London founded a project to see what audio engineering tasks could potentially be automated via machine learning (Barchiesi and Reiss 2010; Strauss 2014). LANDR co-

founder Stuart Mansbridge was a student who was part of the QMUL project. While the use of machine learning for automation was new, automation has a long history in knowledge work, characterized by the closure of 'spaces for play in the structure of skills' – a routinization of procedure and outcome, a formalization of tasks (Aneesh 2001: 363, 382). The difference with machine learning is that automation did not necessarily require a formalization of tasks ahead of time (though it did often require a *redefinition* of tasks, from Turing's imitation game on down). Rather, in theory, neural network-based machine learning develops its own procedures through an iterative process, in a manner that is often opaque even to the programmer. In actual practice, however, machine learning is often only one processing routine among many. The results are experientially similar to older programming models: tasks and skill structures get solidified into place.

LANDR's debts matter – to venture capital, academia, local music scenes, and to the history of audio mastering itself.[4] As Crawford argues, we need to understand algorithms as 'participants in wider institutional and capitalist logics' (2015, 87). Jenna Burrell (2016) has offered a slightly different valence of this argument, claiming that the opacity of machine learning cannot be separated from practices of corporate secrecy and legal obfuscation. The rhetoric of AI's inevitability obscures the intertwined technical and corporate origins of LANDR.

When we first encountered LANDR, they were a company called MixGenius that aimed to automate all the tasks of music mixing. A standard studio recording of music has a variety of (often) acoustically isolated tracks, sometimes with multiple performances. An individual track may have one or more instruments recorded on it, and may or may not have been recorded at the same time as the other tracks. But when played back simultaneously with the other tracks, it gives the impression of people or sounds playing together (for more on the theory of multitracking as a technocultural practice, see Stanyek and Piekut 2010; and Horning 2013). A mixing engineer edits these together, sets the relative volume of different instruments – for instance, boosting a guitar track for a solo, dropping it for a verse where there is singing – and applies a variety of signal processing techniques to touch up the audio, making each part of the mix fit together. The result of a mixing session is a stereo or multi-channel recording that is ready to be mastered. MixGenius' plan and business model was to automate this process. They did not succeed. In our interview with LANDR co-founder Justin Evans, he describes the problem thus:

> The mixing part was really challenging around nuances of genre; it was incredibly difficult to solve. With mastering, the variance is less, there is a lot less. It's still a problem, but it's a solvable problem with data, wherein mixing it seemed like the dataset would be so massive that it would be next to impossible to do. Because there's such a limited scope in what happens in mastering, it seemed

> like 'OK cool we can do this with data.' […] We started working on that and then launched the product very quickly. Because of course, when you're funded [by venture capital] you only have a certain amount of runway and you need to get proof points to get there.[5]

LANDR emerged from MixGenius by passing an academic software project through the filters of a tight venture capital timeline and the limits of machine learning. Put simply, they learned that music mixing is not the right *kind* of problem to address by machine learning. It does not produce coherent and commensurate data sets that can easily be compared with one another, and so it is not susceptible to LANDR's kind of computerized classification. The variability in numbers of tracks across projects alone is potentially infinite: there is no common standard or format within or across genres. The decisions of a mixing engineer vary not only by genre and sound, but also by a host of contextual factors not available in the music itself (Frith and Zagorski-Thomas, 2016; Horning 2013; Hepworth-Sawyer and Hodgson 2017). If MixGenius could not make sense of all the factors going into a project through the available tools of Music Information Retrieval at the time, it could not produce a satisfying mix. As Crane told us, 'how could [MixGenius] be intuitive and creative? There's no formula' (author interview). Not only are there more data points in mixing, they are also the wrong kind of data for machine learning.

Mastering was different. Evans' 'Ok, cool, we can do this with data' hides a myriad of conditions: it makes something contingent sound inevitable, translating a social condition into a technical problem. LANDR needed consistent data and more circumscribed sonic goals than producing a satisfying mix. Mastering provided both because the range of parameters was more limited. But to understand that, we have to consider LANDR not only in the context of machine learning, but also mastering, which has its own dynamic and vexed history. LANDR arrived on the scene at a moment when mastering was technically and commercially available for co-optation by machine learning.

A compressed history of mastering

In the early days of studio recording, performers would be recorded directly onto wax cylinders that would then be sold commercially (the following discussion is drawn from Meintjes 2003; Auld 2004; Katz 2015; Shelvock 2012: 9-10). Recording studios that recorded to disc had a lathe in the studio for cutting master discs that, through a process of positives and negatives (as in photography), would be used to stamp copies. But as tape became an important medium for recordings, and as the final formats of audio proliferated – 45 singles, LP records, compact cassettes, 8-track tapes, etc. – *mastering engineer* became a separate job description. At first, mastering engineers

were largely thought of as technicians or translators; their job was to prevent too much bass from ruining a master record, or worse, the record-cutting lathe itself. Their job was thought of as that of a traffic cop, or a translator.

The meaning of mastering has changed along with the conditions of music's circulation. But professional history also plays a role; as mastering engineers sought to aggrandize their role (as professions often do), they created a mythology around their work (Shelvock 2012, 11). If recording studios were cordoned off from the street and daily life through a visual and architectural rhetoric of space travel and advanced technology (Meintjes 2003, 72, 84), mastering engineers were more like wizards in distant castles, working magic on music and sending it back to its makers so that it could go out in the world. That few musicians attended mastering sessions added to their mystery. Until recently, mastering engineers were an interesting exception to the DIY 'revolution' in music creation and production. A small number of high-end mastering engineers more or less defined the sound of recorded music for a vast swath of genres and applications. The industry was highly specialized, very highly concentrated at the top, and poorly understood by non-specialists. In contrast, boundaries around other specialized music creation roles began to blur as DIY became a key marketing strategy for music technology and software companies. Musicians were told they could do everything themselves, allowing amateurs to realize fantasies of access to technology and control that only the most professional musicians could previously entertain. In many genres and musical scenes, formerly distinct roles like recording and mixing engineer, musician, songwriter, producer, and promoter are collapsed into one another (Bell 2014; Baym 2018; Crane, author interview). While musicians and engineers have experienced a certain amount of role collapse, mastering mostly has not.[6] Larry Crane told us 'This happens to me on a weekly basis. […] I'll be making a record with someone, producing or mixing in whatever facility I'm in, and they'll be asking me 'What does mastering do? I don't understand.' (Author interview).

Though hardware and software existed for DIY mastering, it remained largely a practice done by specialists, or not at all. Attempts to create hardware that automated part of what mastering engineers did through combinations of processes or 'wizards' – like the TC Electronics Finalizer – were still expensive and not widely taken up by amateurs (Massey 1996). Mastering presets in software were more successful (probably because the software was less expensive and widely pirated) but still did not fully bring mastering into the DIY world, at least not compared with audio engineering or promotion. The other problem with DIY mastering is that it can 'ruin an album […] It's easy for someone to fool themselves into thinking they are making a track better when it becomes louder than a previous iteration of the track, and when it becomes brighter, when there is more high end,' yet those

changes often create problems with the relative loudness of vocals, snare drums, or other important parts of a mix (Crane, author interview).[7] Poorly understood and difficult to do on your own, mastering was still available for LANDR's DIY sales pitch. Because it was already removed from the every-day experience of musicians, AI-assisted mastering wouldn't feel as different to users than AI-assisted mixing.

Mastering also provided one other solution to a problem specific to LANDR's application of supervised machine learning: their extraordinary hunger for coherent, commensurable data sets. Getting from a stereo mix to a master is a more finite problem than getting from a bunch of recorded tracks to a mix. Crucially, the data exist for tracking the difference between a finished stereo mix and a master: record company archives are full of final mixes before *and* after they were mastered, and the digital data all conform to well-established standards. Access to this 'before and after' mas-tering data is perfect for supervised machine learning, the kind employed by LANDR. Supervised machine learning involves a known input and a desired output; in this case, a finished mix was the input and a mastered recording was the output (Alpaydin 2014: 21-48). Essentially, supervised machine learn-ing algorithms work from a set of training data and then surmise how one gets from point A to B, which is how Evans intimated that LANDR works in our interview. It is crucial to note that we can't verify his account. LANDR's practices of corporate secrecy prevent us from verifying how it works and, even if we were allowed in, an analysis of the algorithm or its outputs might not tell us much. In our interview, Evans told us that LANDR acquired access to a set of unmastered mixes from the vault of a record company (Warner is a likely candidate since they are an investor in LANDR) and com-pared them to the finished masters, using one subset as training data and another subset as testing data. They could then begin building a method for mastering unmastered audio, which could be augmented by large, infor-mal datasets generated by its users. In the years since it went live, LANDR's most important resource for data has been its own users.

In addition to working with the data set provided by a record company, as well as user data, LANDR routinely hired audio engineers as consultants, including working mastering engineers. This was an easy thing for them to do: Montreal has a relatively large and successful music scene, and four large universities with audio technology programmes of one sort or another. On the day we visited, they had two large, working mastering studios sitting vacant. Evans explained that in addition to the automated mastering they provided to regular customers, LANDR was also working with corporate clients, like movie studios, where they would use a combi-nation of machine learning and a human mastering engineer to master the large number of recordings that would go into, for example, a Hollywood motion picture. The end goal may have been to automate the process, but

they were not yet there. LANDR may well use machine learning, but it is only one part of what they do.

Like other big-data based music services such as Shazam and Spotify, LANDR thus masks their own reliance on musicians and music practices outside of the corporate mainstream for growth and AI development. Both Shazam and Spotify perfected their algorithms by relying on datasets they procured informally. In a barter deal in 2002, Shazam digitized a large vinyl library in exchange for extracting musical fingerprints from the records to test its recognition algorithm (Razlogova 2018). While they were starting out, Spotify engineers reportedly downloaded millions of tracks from Pirate Bay on to their platform, to test out an early version of their algorithm – a fact the company has tried hard to suppress (Eriksson, M. *et al.* 2018). LANDR may have been granted access to a company's back catalog and its users' recordings, but it follows in the same path.

Mastering as a service for infrastructures of circulation

When LANDR retunes tracks for digital circulation, it becomes an integral part of music distribution and recommendation infrastructures (Sterne 2012; Bratton 2016, 45). To LANDR, people upload unmastered stereo audio tracks, see animated graphics as LANDR analyzes and masters their music, and then audition the recordings, choose different settings, pay for them, and download them in a variety of formats. In this way, LANDR looks more like any other cloud-based file service – like Dropbox or Box.com – and less like an audio product. Evans described their design goals like this: 'How do you create a new behavior that isn't threatening to people? We did a lot of thinking about interfaces that are not going to feel like 'oh my god, what am I doing here?'' (authors interview). Inasmuch as it is a music company, LANDR takes its cues from other web-based and app-based services for music, especially major music recommendation and recognition services such as Shazam and Spotify. LANDR's description of its mastering engine builds directly on the success of Shazam, the oldest and best-known music recognition app. This is evident in their 'think Shazam' analogy, and in the description of the mastering process as extracting and analyzing a 'musical fingerprint,' a term popularized by Shazam (Acoustic Fingerprinting n.d.).[8] Other cues are visible in their strategies for self-presentation, algorithm development, and fundamental underlying ideological assumptions.

LANDR relies on the same surprise factor as music recognition and recommendation. Originally Shazam, and to a lesser extent, early music recommendation services like Spotify and Pandora, benefited from the users' delight at seeing a machine perform a task, such as recognizing and recommending music, that they thought only human memory or a knowledgeable human DJ could accomplish. At launch, LANDR also had received a lot of

criticism, and still does, on audio community sites such as the unfortunately-named Gearslutz and the music press (Gearslutz 2014). As mastering engineer Mike Wells told a reporter in 2014, 'as an artist and an audio engineer, I can't imagine anyone with a serious investment in their music giving final control of their art over to an algorithm' (Heiselmann 2014). Of course, this criticism romanticizes audio production; certainly this criticism works for music as consecrated art, but of course advertising spots, subway announcements, interface sounds, ringtones, and countless other forms of mundane audio are also mastered on a daily basis. Some users argued that LANDR's algorithm worked little better than non-AI based software options already available to musicians, and could not stand up to a collaboration with an experienced and responsive human mastering engineer. But that talk only encouraged users to try LANDR to test it against presets and engineers. Similarly, for years after their launch, users and journalists continually tested Shazam, trying to confuse it with a particular genre or track, or waited for the Spotify to suggest something the listener particularly hated. LANDR was tested in a similar fashion. In 2014, one reviewer found that LANDR cut off low frequencies necessary for EDM tracks, using 'a tonal balance [that] might be perfect for folk or classical, but it [didn't] cut it for EDM [Electronic Dance Music], hip hop, or even pop.' (Hazard 2014). After years of targeted testing with EDM musicians and human EDM mastering engineers, LANDR fixed its low frequencies problem and, as of 2018, works better with EDM (J. Evans, Interview by authors, August 24, 2016; Anonymous mastering engineer, Interview by authors, September 6, 2018). LANDR used that initial skepticism to demonstrate the promise of AI-based mastering. The mere fact of making the comparison cedes the terms of the argument to the company. LANDR – and those who accepted its terms of debate – set up an equivalence between what their app does and what a human mastering engineer does.

LANDR, Shazam, and Spotify share another key ideological conceit: that all music will eventually be subsumed in data banks and analyzed by AI – a bizarre infrastructural fantasy for music. As LANDR co-founder Justin Evans affirmed, 'all music will be in the big data cloud eventually' (Author interview). Breathless mainstream press coverage of AI for music trumpets the inevitability of comprehensive coverage of big data and complete machine learning, touting a 'the universe of music' offered by Pandora and the 'soundtrack to your life' offered by Spotify (Thompson 2014). But just as no infrastructure can be total, no database can hold all of the world's recorded music.

This AI-data-inevitability story echoes decades of industry fantasies about a 'celestial jukebox' (Burkart and McCourt 2006), which erases these services' contingent and human origins and futures. But as many artists and scholars have pointed out, that is just not possible. Much of the world's historical, non-Western, and unlicensed music will remain outside of the commercial cloud (Gitelman 2010; Bennett 2012). Machine learning applications work by

analyzing commensurable data; audio data in the case of LANDR or Shazam, or metadata in the case of recommendation engines like Spotify. Shazam and Spotify are limited by digital licensing. Shazam cannot recognize a track that has not been digitally licensed because its database cannot include unlicensed tracks. While LANDR's limitations are different, they also stem from its financial goals. LANDR works well with EDM tracks because it collaborated with EDM musicians and engineers in tweaking its mastering engine. The mastering algorithm could potentially work with genres that do not have a sizable LANDR user base, but LANDR will not invest in training it to do so. They follow their users, 'getting good' at genres that show up on their service (Authors interview).

Crucial to understanding LANDR as an infrastructural service, and to the critical analysis of corporate applications of signal processing more generally, whether through machine learning or other protocols, is that we cannot actually know how it works. LANDR *could* conceivably work entirely by machine learning, comparing a set of inputs to a set of outputs and rendering them. If it worked this way, it would compare each uploaded track to the massive data set of recordings they have at their disposal. The system would then look for patterns and features to develop a basis for comparing it to other recordings in its database. From this, it would develop a profile (or 'fingerprint'). This much is like other companies such as Shazam or Spotify. If the whole process were completed through machine learning, it would then diverge from the applications used by those companies, and use the fingerprint to decide how the audio *should* sound, and what problems it might have. From there, it would make a series of automated decisions about the sound of the audio to get it from 'unmastered' to 'mastered' by passing the file iteratively through layers of a neural network. But what actually happens after the fingerprinting is unclear. While it is theoretically possible that LANDR operates entirely by machine learning, this is unlikely.

LANDR has no patent for such a process and we have found no evidence of success at such a task in the audio engineering literature. Another option is that Evans could have been lying to us, and LANDR could be a form of fake artificial intelligence, where people are doing the work behind the putative algorithm. This is also unlikely because the number of mastering engineers required to do their work would be cost prohibitive. A more likely scenario is that LANDR does do the feature extraction, but then selects from a grid of possible presets for its outputs. It would, as in facial recognition technologies, simply decide which preset is 'most like' the track submitted for mastering and then apply that preset. It could have just a few presets – under ten. Or it could have many dozens. Or it could have shades in-between. Magdalena Piotrowska, an engineer who has been auditing automatic mastering services, confirmed in an email (to Sterne, 21 September 2018) that, in her experience, given the same input file, LANDR consistently outputs the

same mix in her experience. Jonathan also tested LANDR with the same exact file on different days and found it appeared to make the same decisions. For our interview, Larry Crane also tested LANDR and said to him it sounded like a preset, or that 'a mastering engineer spent 5 minutes on the recording.' By applying a mix of compression and equalization, he was able to achieve the same results as LANDR in a few minutes.

None of this is meant to suggest that LANDR listens like a person. Although LANDR uses the language of music genres for its advertising, it does not actually work according to genre, instead operating within a set of categories generated by the machine learning process itself. This is because genre cannot be encoded such that an algorithm can discover it. As David Brackett writes, genres are 'associations of texts whose criteria of similarity may vary according to the uses to which the genre labels are put. 'Similar' elements include more than musical-style features, and groupings often hinge on elements of nation, class, race, gender, sexuality, and so on […] the 'effects' of musical genre cannot be traced to the 'cause' of musical style in a direct, one-to-one relationship' (2017: 4). Machine listening, then, is nothing like human listening: it does not operate according to the same categories of music genre and style distinctions, it cannot account for extra-musical factors that are not coded as metadata, and while LANDR's feature extraction gets as aspects of the music like loudness and the relationships among different frequencies, it does so differently than a person would (Piotrowska et. al 2017).

As figure 1 shows, LANDR hedges some of these differences. While their advertising speaks in terms of genre, here they turn to 'production styles,' since genre is not an operative category in their software because of its heavily non-sonic definitions. In the final step in the flow chart, they explicitly compare themselves to mastering engineers with the phrase 'adjusts all the knobs,' which implies their algorithm makes decisions in the ways that a mastering engineer might. But of course, its decision process is totally different: it extracts features from a recording, compares them to a dataset, and then makes a decision, whatever the actual process of decision-making and application is. A mastering engineer will work through a process of trial and error, based on repeated auditions of the track, and comparisons with other recordings by the same and other artists (Shevlock 2012; Katz 2015).[9] LANDR appears to take only a single pass at listening, or more accurately, auditing a track.

Mastering engineers draw from the history and traditions of audio engineering. LANDR draws from the tradition of web-based audio applications from Winamp on down (Morris, 2015). Mastering's opacity from its clients is a major part of its history and cultural significance. LANDR effectively leverages that history to promote its own commercial aims. While users may interact with the interface in myriad ways, the overall effect of
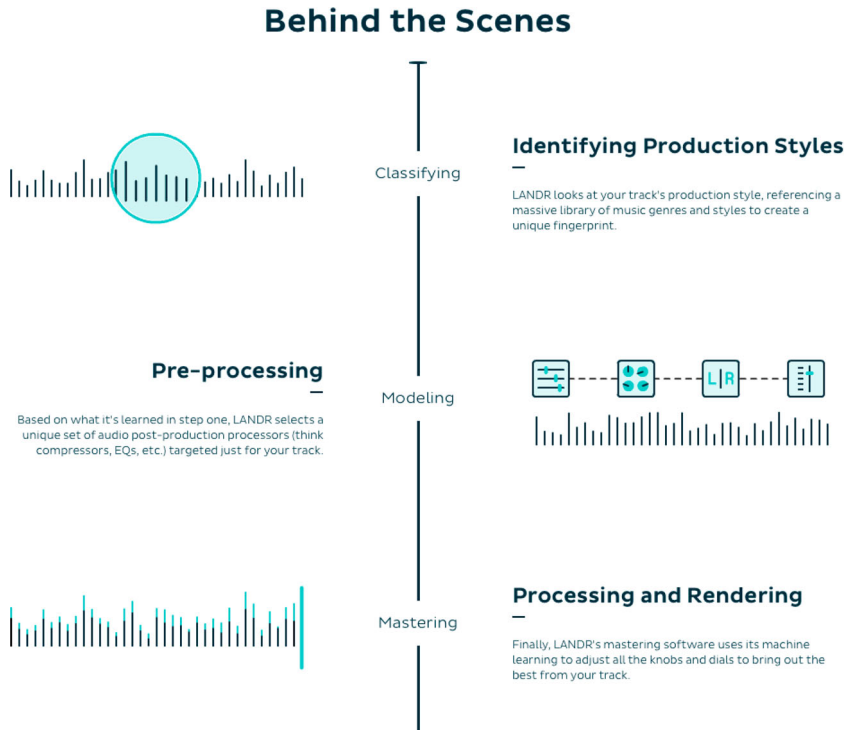
Behind the Scenes

Figure 1. LANDR explains its process to its users (source: Jonathan's screenshot of https://www.landr.com/en/online-audio-mastering 28 July 2018, reproduced under fair use provisions).

LANDR's approach is to transform the status of mastering from something whose inner workings are obscured for the user because of the structure of the audio industry to something that is obscured from the user because of the inner workings of its status as a web-based software service. This is not a trivial change. It is, to use Alex Galloway's term, 'technical transcoding … that nevertheless coexists with an exceedingly high level of ideological fetishism and misrecognition' (2012: 60). Of course, industrial and technical practices cannot exist outside of ideology. Infrastructures of music distribution and discovery, and automatic mastering that tunes tracks for circulation, operate as meshes of discourses, materials, and practices that aim to shape a corner of the audio-technical universe.

Conclusion

Whether or not LANDR has long-term success as a mastering company, it stands as an early example of how automated aesthetic judgments are used to shape the sound, look, and feel of media. But those judgments are

completely different from that of a mastering engineer. Mastering tunes sound for infrastructure, but the infrastructures through which music will circulate are not fixed sets of phenomena that can be operationalized. As Paul Edwards (2016, 25) has written, 'many modern infrastructures are not systems at all, but complex clusters of interacting elements captured by organic or ecological metaphors.' In this article, we have documented the possibilities and limits under which LANDR emerged as a passage point for music circulation as an infrastructural practice. Broadly speaking, it required four prior conditions:

(1) A data condition: it began with a large dataset of mastered and unmastered recordings, and then built out a massive database from its users' submissions. These data had to be abstracted from social practice and rendered commensurable with one another for the purposes of machine learning. This was only possible because of the conditions under which digital music circulates.
(2) The specific institutional conditions of audio mastering: its removal from the everyday experience of most musicians, its black boxing as a cultural and technical process, a high degree of specialization and concentration in the mastering industry that made it more amenable to co-optation by machine learning, a label's ownership of a large number of comparable recordings pre- and post-master, and a user base willing to upload their own work for comparison.
(3) An aesthetic condition: the standardization of the sonic profile of commercially released music, an effect of the mastering industry's subtle mediation between recording practice and incredibly diverse contexts of audio playback, afforded an opportunity to delegate aesthetic decisions to machine learning algorithms.
(4) A technical condition: the development, in other fields, of large-scale machine learning techniques that could make use of the available datasets, in the context of a state of musical, mastering, and audio-technical practice.

Each of these conditions was necessary, but alone none is sufficient. LANDR's failure at mixing shows that AI systems will need data sets that are internally coherent – each datum must be commensurable with the others in order for the system to work. The likelihood of LANDR's use of presets, or something like them, rather than true machine learning at the signal processing stage, also means that those of us interested in culture should be especially concerned about the shortcuts that media companies make in the service of rendering aesthetic practices as subject to potential automation. Scholars will need to analyze the conditions under which the products and elements of the production processes of culture – music,

literature, journalism, cinema, art, dance, video games, fashion, and on and on – can be rendered and manipulated as commensurable data.

But commensurable and usable data – alongside technical capacity – are never enough on their own for AI to take hold in a particular field. The culture industries' discourse of AI's inevitability and imminent omnipotence is misleading. To understand, assess, and intervene in the cultural politics of AI, scholars will need to consider broader questions of how work processes operate, the meanings of the work performed by the AI to its users, the ideologies operating in the interface, what kinds of data can be generated or acquired and the infrastructural conditions within which it must operate.

The politics of current and future digital cultural infrastructures, then, are inseparable from the role of automation in aesthetic practices. Formats and signal aesthetics designed for circulation dominated the media of the twentieth century, and we should expect the same to be true for the 21st. In the coming years, we expect to see many more applications that use machine learning, always in combinations with other technical processes and arrangements of people, practices, and technologies, to automate aspects of signal processing, conditioning the sound, look, feel, and texture of aesthetic practices across media and formats. When it takes textual form, like music, literature, TV, games, journalism, visual art, or some other form, culture can circulate through infrastructures, shaped and quantified according to the needs and protocols of technology and industry. But people engage with textualized or datafied forms of culture because they matter in shifting, contingent, and contested ways. Human categories like genre – or any other major classification – constantly change with use (Bowker and Starr 1999). As of now, this is something to which machine learning must react or adapt; it is not something that can be accounted for within the protocols of machine learning itself. In this way, machine learning is just the latest chapter in a long story of capitalism failing to fully account for culture. But how might cultural studies account for this dimension of machine learning? Faced with the monumentality of today's cultural infrastructures, built on pulsing rivers of data, we must attend to the politics of signal processing in any stories we tell about meaning and contestation in a datafied world. For the time being, machine learning will be an important part of the tales we tell.

Notes

1. From Jonathan's informal conversations with colleagues in these organizations, it appears no other major changes occurred apart from rebranding.
2. We have chosen LANDR over competitors for reasons of approach. Izotope uses its machine learning to design software plugins to be installed on users' computer, not in a dynamic 'real time' arrangement and therefore is not exactly comparable. Their algorithms are more 'hand crafted' with a machine learning

supplement. Cloudbounce is less heavily capitalized, but also less accessible to authors.

3. There are numerous internalist accounts of the recent history of machine learning by computer scientists eg., http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/. As of yet, we have not found a history of this period of the field that treats the science as itself a cultural and social phenomenon.
4. For more on LANDR's connection to the local scene, as well as how its algorithm works, see our companion essay (Sterne and Razlogova 2019).
5. Spoken language in interviews is lightly edited to read better as written language.
6. Crane did point out that there are services which provide both mixing and mastering, and some artists do their own mastering, but these are an exception. No mastering engineer we spoke with seemed threatened by LANDR; Crane reported the same impression.
7. This may be in part because of the contexts of DIY music production: a home studio with significant acoustic irregularities and consumer grade speakers would create more problems for a mastering engineer to solve than a recording done by professional engineers in a commercial studio.
8. Based on a technique called Music Information Retrieval, music recognition analyzes parts of a recording to compare it to an available database of recordings, and then makes a 'guess' as to a match.
9. We discuss this further in our companion essay (Sterne and Razlogova 2019).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes on contributors

Jonathan Sterne teaches at McGill University and is, most recently, author of Diminished Faculties: A Political Phenomenology of Impairment (2021). Visit his website at https://sterneworks.org

Elena Razlogova is an Associate Professor of History at Concordia University. She is the author of The Listener's Voice: Early Radio and the American Public (2011).

References

"About LANDR.". n.d. *LANDR*. Accessed July 28, 2018. https://www.landr.com/en/about.
"Acoustic Fingerprint". n.d. *Wikipedia*. Accessed July 5, 2018. https://en.wikipedia.org/w/index.php?title=Acoustic_fingerprint&oldid=848922505.

Alpaydin, E., 2014. *Introduction to machine learning*. Edited by Francis Bach. 3rd ed. Cambridge, MA: MIT Press.

Aneesh, A., 2001. Skill saturation: rationalization and post-industrial work. *Theory and society*, 30 (3), 363–396.

Auld, R. 2004. "Mastering-then and now." *Recording Magazine*, 2004. http://www.auldworks.com/articles/mastering.htm.

Balsamo, A., 2011. *Designing culture: the technological imagination at work*. Durham: Duke University Press.

Barchiesi, D. and Reiss, J., 2010. Reverse engineering of a mix. *Journal of the audio engineering society*, 58 (7/8), 563–576.

Baym, N.K., 2018. *Playing to the crowd: musicians, audiences, and the intimate work of connection*. New York: NYU Press.

Bell, A.P., 2014. Trial-by-fire: a case study of the musician–engineer hybrid role in the home studio. *Journal of music, technology & education*, 7 (3), 295–312.

Bennett, V., 2012. Collateral damage. *The Wire*. Issue: 237, http://thewire.co.uk/articles/8439.

Born, G., 2005. On musical mediation: ontology, technology and creativity. *Twentieth-Century music*, 2 (1), 7–36.

Bowker, G.C. and Star, S.L., 1999. *Sorting things out: classification and Its consequences*. Cambridge: Mass.: The MIT Press.

boyd, d. and Crawford, K., 2012. Critical questions for big data. *Information, communication & society*, 15 (5), 662–679.

Brackett, D., 2017. *Categorizing sound: genre and twentieth-century popular music*. Berkeley and Los Angeles: University of California Press.

Bratton, B.H., 2016. *The stack: on software and sovereignty*. Massachusetts: MIT Press.

Burkart, P. and McCourt, T., 2006. *Digital music wars: ownership and control of the celestial jukebox*. Lanham; Toronto: Rowman & Littlefield.

Burrell, J., 2016. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big data & society*, 3 (1), 1–12.

Collins, N., 2016a. A funny thing happened on the way to the formula: algorithmic composition for musical theater. *Computer music journal*, 40 (3), 41–57. doi:10.1162/COMJ_a_00373.

Collins, N., 2016b. Towards machine musicians Who have listened to more music than Us: audio database-led algorithmic criticism for automatic Composition and live concert systems." *Computers in entertainment*, 14 (December): 1–14. doi:10.1145/2967510.

Crawford, K., 2016. Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, technology & human values*, 41 (1), 77–92. doi:10.1177/0162243915589635.

Crawford, K. and Joler, V., 2018. "Anatomy of an AI system: the Amazon Echo as an anatomical map of labor, data and planetary resources." New York: Artificial Intelligence Now. https://anatomyof.ai.

Darling, K., 2017. "AI and public perception." Presented at the AI Now Expert's Workshop, MIT Media Lab, 25 July 2017. Available online at https://www.youtube.com/watch?v=P2nPHYhU_EE. Last accessed 28 July 2018.

Dubnov, S. and Surges, G., 2014. Delegating creativity: use of musical algorithms in machine listening and composition. In: Newton Lee, ed. *Digital Da Vinci: computers in music*. New York: Springer, 127–158.

Durham, B., 2018. "Regulating dissemination: a comparative digital ethnography of licensed and unlicensed spheres of musical circulation." PhD Dissertation, Oxford: Oxford University.

Eck, D., 2014. "Google, USA: music recommendation and discovery at scale — CIRMMT." CIRMMT Distinguished Lectures in the Science and Technology of Music, October 30. https://www.youtube.com/watch?v=GAzVZOZe2f8.

Edwards, P.N., 2016. We have been assimilated: some principles for thinking about algorithmic systems. In: Ulrike Schultze, Margunn Aanestad, Magnus Mähring, Carsten Østerlund, and Kai Riemer, ed. *Living with monsters? social implications of algorithmic phenomena, hybrid agency, and the performativity of technology*. Cham, Switzerland: Springer, 19–27.

Edwards, P.N., *et al.*, 2009. Introduction: an agenda for infrastructure studies. *Journal of the association for Information systems*, 10 (5), 364–374. doi:10.17705/1jais.00200.

Eriksson, M., *et al.*, 2018. *Spotify teardown: inside the black box of streaming music*. Cambridge, MA: MIT Press.

Frith, S. and Zagorski-Thomas, S. eds., 2016. *The art of record production: an introductory reader for a new academic field*. London: Taylor and Francis.

Galloway, A.R., 2012. *The interface effect*. Cambridge, UK; Malden, MA: Polity.

Gearslutz, 2014. "I guess we are all out of a job then?" (Thread). May 28. https://www.gearslutz.com/board/mastering-forum/931986-i-guess-we-all-out-job-then.html#post11778143.

Gitelman, L., 2010. Welcome to the bubble chamber: online in the humanities today. *Communication review*, 13 (1), 27–36.

Hallinan, B. and Striphas, T., 2016. Recommended for you: the netflix prize and the production of algorithmic culture. *New media & society*, 18 (1), 117–137.

Hazard, B., 2014. "Can LANDR replace your mastering engineer?" *Passive Promotion*. June 12. http://passivepromotion.com/can-landr-replace-your-mastering-engineer.

Heiselmann, J., 2014. "Is Instant Online Mastering Worth It?" *Sonicbids*. May 12. http://blog.sonicbids.com/is-instant-online-mastering-worth-it.

Hepworth-Sawyer, R., and Hodgson, J., eds, 2017. *Mixing Music*.

Hoffman, S.G., 2017. Managing ambiguities at the edge of knowledge: research strategy and artificial intelligence labs in an Era of Academic capitalism. *Science, technology, & human values*, 42 (4), 703–740. doi:10.1177/0162243916687038.

Horning, S.S., 2013. *Chasing sound: technology, culture, and the Art of studio recording from edison to the LP*. Baltimore, MD: Johns Hopkins University Press.

Jones, S., 2000. Music and the internet. *Popular music*, 19 (2), 217–230.

Katz, R.A., 2015. *Mastering audio: The Art and the science*. Burlington, MA: Focal Press.

"LANDR.". n.d. *Crunchbase*. Accessed July 28, 2018. https://www.crunchbase.com/organization/mixgenius.

Langley, Pat. 2011. The changing science of machine learning. *Machine learning*, 82 (3), 275–279.

LeCun, Yann, et al. 2015. Deep learning. *Nature*, 521 (7553), 436–44. doi:10.1038/nature14539.

Levy, K.E.C., 2015. The contexts of control: information, power, and truck-driving work. *Information society*, 31 (2), 160–174.

Lewis, G.E., 1999. Interacting with latter-Day musical automata. *Contemporary music review*, 18 (3), 99–112. doi:10.1080/07494469900640381.

Lewis, G.E., 2018. "Why Do We Want Our Computers to Improvise?" *The Oxford Handbook of Algorithmic Music*, February. https://doi.org/10.1093/oxfordhb/9780190226992.013.29.

Li, Xiaochang, and Mills, Mara. 2019. Vocal features: from voice identification to speech recognition by machine. *Technology and culture*, 60 (2), S129–60. doi:10.1353/tech.2019.0066.

Lundgren, I., 2018. "Google Goes All-in on Artificial Intelligence, Renames Division Google AI." *Techcrunch*, 20 July. Accessed July 28, 2018. http://social.techcrunch.com/2018/05/08/google-goes-all-in-on-artificial-intelligence-renames-research-division-google-ai/.

Massey, H. 1996. Off to see the wizard. *Musician Magazine* (December): 77–79.

Manovich, L., 2018. *AI aesthetics*. Moscow: Strelka Press.

Meintjes, L., 2003. *Sound of Africa! making music zulu in a South African studio*. Durham: Duke University Press.

Merton, R.K., 1973. *The sociology of science: theoretical and empirical investigations*. Edited by Norman W. Storer. Chicago: University of Chicago Press.

Morris, J., 2015. *Selling digital music, formatting culture*. Berkeley and Los Angeles: University of California Press.

Mulvin, D., 2021. *Proxies: The cultural work of standing In*. Cambridge: MIT Press.

Parnell, M. 2017. "Mandy Parnell in Conversation." Ableton Loop Fest 2017 (Berlin, Germany). Accessed July 28, 2018. https://www.youtube.com/watch?v=D-XWRXnnWL0.

Parks, Lisa, and Starosielski, Nicole. 2015. Signal traffic: critical studies of media infrastructures. Open WorldCat, http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=992537.

Piotrowska, M., *et al.* 2017. "A Study on Audio Signal Processed by 'Instant Mastering' Services." Paper presented at the Audio Engineering Society Convention. http://www.aes.org/e-lib/browse.cfm?elib=18597.

Radin, J., 2017. Digital natives: How medical and indigenous histories matter for Big data. *Osiris*, 32 (1), 43–64.

Razlogova, E., 2018. Shazam: The blind spots of Algorithmic music recognition and recommendation. In: J. W. Morris, and S. Murray, ed. *Appified: culture in the Age of apps*. Ann Arbor, MI: University of Michigan Press, 257–266.

Rumsey, F., 2011. Mastering: Art, perception, technologies. *Journal of the audio engineering society*, 59 (6), 436–440.

Schwartz, O., 2018. "'The Discourse Is Unhinged': How the Media Gets AI Alarmingly Wrong." *The Guardian*, July 25. Online edition. Accessed July 28, 2018. https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong.

Schmidhuber, Jürgen. 2015. Deep learning in neural networks: an overview. *Neural Networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003.

Shelvock, M., 2012. "Audio Mastering as Musical Practice." M.A. Thesis, University of Western Ontario.

Slack, J.D. and Wise, J.M., 2014. *Culture and technology: a primer*. 2nd ed. New York: Peter Lang.

Stanyek, J. and Piekut, B., 2010. Deadness: technologies of the intermundane. *TDR/the drama review*, 54 (1), 14–38.

Sterne, J., 2012. *MP3: the meaning of a format*. Durham, NC: Duke University Press.

Sterne, J., and Razlogova, E., 2019. Machine learning in context, or learning from landr: artificial intelligence and the platformization of music mastering. *Social media and society*, 5 (2), 1–18.

Strauss, K., 2014. "Making Money on the Music Mix: Meet Mixgenius." *Forbes*, June 16, 2014. http://www.forbes.com/sites/kastenstrauss/2014/06/16/making-money-on-the-music-mix-meet-mixgenius/.

Thompson, D., 2014. "The Shazam Effect." *Atlantic*, December. http://www.theatlantic.com/magazine/archive/2014/12/the-shazam-effect/382237/.

Titlow, J.P., 2017. "Why Did Spotify Hire This Expert in Music-Making AI?" *Fast Company*, July 13. https://www.fastcompany.com/40439000/why-did-spotify-hire-this-expert-in-music-making-ai.

Turner, F., 2006. *From counterculture to cyberculture: stewart brand, the whole earth network, and the rise of digital utopianism*. Chicago: University of Chicago Press.

Wang, H. and Raj, B., 2017. "On the Origin of Deep Learning." *ArXiv*, February 24. http://arxiv.org/abs/1702.07800.

Winner, L., 1977. *Autonomous technology: technics-out-of-control as a theme in political thought*. Cambridge, Mass: MIT Press.

Wyatt, S., 2004. Danger! metaphors at work in economics, geophysiology, and the internet. *Science, technology, & human values*, 29 (2), 242–261. doi:10.1177/0162243903261947.

Zuboff, S., 2019. *The age of surveillance capitalism: the fight for a human future at a new frontier of power*. New York: Public Affairs.