10 Deep Remixes

A music artist sings with the voice of her collaborator, their faces intermingled in a hazy algorithmic collage. A video self-portrait in which closeups of the artist's face become galaxies and nebulae as seen through a neural network's perspective. Real-life politicians perform strange AI-generated songs as a deceased singer speaks about contemporary politics.

These cutting-edge works constitute early examples of the future of generative art. Specifically, they belong to a new range of remix practices enabled by deep learning technologies and by the availability of large data sets. These remixes happen at different levels. First, as discussed in the previous chapter, to a very large extent machine learning art rests on the creation of a data set, which can be done by generating one's own material, crowdsourcing the work, or appropriating existing materials. The careful construction of a training set by blending data from different sources to train a desired model constitutes a first form of machine learning remix. Second, once a model has been trained, it can be reused as is to generate different outputs. This feature of deep learning has already opened new possibilities for artists by giving them access to models which would otherwise be quite expensive to train, for example through the release of Google's BigGAN (Brock, 2019) as well as the GPT-n language models (Radford et al., 2019; Brown et al., 2020) and the music sample-generator Jukebox (Dhariwal, 2020) developed at OpenAI.

Third, deep learning neural architectures are to some extent modifiable. Not only can a pretrained model be reused in different contexts but under certain conditions, it can be upgraded by, for example, retraining the neural network in whole or in part using a different data set, or by adding new layers of neurons on top of the existing ones.

Finally, novel machine learning techniques such as style transfer (Gatys et al., 2015), pix2pix image-to-image translation (Isola et al., 2018; Wang et al. 2018) and Flow Machines (Pachet, Roy, & Carré, 2021) allow for a whole new range of algorithmic remixes in which a transformation from one domain to another is automated via training over a data set.

Machine learning thus opens the door to a new era of remix culture in which not only content but content-generation *processes* can be easily reproduced, copied, modified, and remixed. This change is supported by the capacity of machine learning (in particular, deep learning) to rematerialize large collections of content (data) into dynamic structures (models) that can be activated as generative processes. Early examples are found in recent advances in image and voice synthesis and also in the emergence of new techniques such as style transfer (Gatys et al., 2015), program synthesis (Kant, 2018), and transfer learning (Thrun and Pratt, 1998). These techniques automate the production and transfer of complex, often nonintelligible pattern production mechanisms such as a writer's style or a music genre.

Remix Culture

In the late 1960s, Jamaican sound engineer Osbourne Ruddock was working as a disc cutter. As he was removing vocal tracks from recordings for DJs and MCs, he discovered that the mixing desk also allowed him to make other modifications to the musical tracks, thus creating different effects. Ruddock, who would later be known as King Tubby, thus started to use the mixer to create his own version of existing records; these versions could then be talked over by DJs. Through the 1970s, Tubby would become a leading figure in Jamaican dub music, as he further developed his remixing techniques by augmenting and mutating songs using echoes, phrasings, and reverbs.

The mixing desk and sound systems allowed dub artists such as Tubby to invent a complete new genre of music built on pre-existing records rather than on sounds recorded from the real world. Later with the emergence of the computer, remixing became the dominant approach to music making and would spread to other media (for example, to photography with the introduction of digital editing tools such as Photoshop in the 1990s), favoring the emergence of what American scholar Lawrence Lessig has dubbed *remix culture* (Lessig, 2009).

Sampling and remixes have defined the evolution of artistic practices and cultures throughout the twentieth century. In his book *Remix Theory*, media theorist Eduardo Navas explains these transformations in relationship to technological developments (Navas, 2012), tying the emergence of the remix to mechanical reproduction. Starting in the nineteenth century, new technologies allowed the recording of samples of the real world, first through photography and then through sound recording.

Navas explains that in the 1920s, photo collages and photomontages constituted early examples of recycling existing materials and recombining them. The Dadaists used these strategies to undermine and subvert painting, photography, and poetry. Hannah Höch's *Cut with the Kitchen Knife Dada Through the Beer-Belly of the Weimar Republic* (1919–1920) exposes the failure of Germany's Weimar Republic to create a democratic and egalitarian regime by appropriating images and text from the mass media.[1] In 1920 Dadaist poet Tristan Tzara published "To Make a Dadaist Poem," a procedure that involved cutting up a newspaper article using a pair of scissors and randomly reorganizing the contents. Later, in the 1960s and 1970s, similar recombination techniques and other algorithmic procedures were used within avant-garde literary movements such as OULIPO and the Beat Generation.

Beginning in the 1980s, Navas argues, new media pushed the tendency further by increasingly *privileging* sampling of existing material over real-world records. In the 1990s music remix became a genre of its own, pushed in the US by an industry that found in these principles an efficient program for music production, commercialization, and consumption.[2] The increased availability of computers and the development of new software since the 1980s has facilitated processes of remixing in other fields, as well. For example, Photoshop and other image-manipulation tools have radically transformed photography, and copy-pasting has become one of the basic principles of writing thanks to word processors and the World Wide Web.

As remixing practices were thus increasingly applied to media outside of the musical realm in the late 1990s, they quickly became the norm. Lawyer Lawrence Lessig sees the emergence of this remix culture as a positive development. He has argued that remixing (the practice of creating derivative works by transforming existing materials) is a beneficial and natural method to enhance human creativity that has been a common practice throughout human history. While copyright laws established in the late twentieth century have restricted such practices, in the twenty-first century remix culture is nonetheless thriving thanks to technological advances that facilitate copy-pasting, modifying, and mixing media. This is true not only for images, sounds, and video but also for source code.

Open-Source Cultures

In the early 1980s in the midst of the growing spread of personal computers, commercially owned software had become the norm. Richard Stallman, then a programmer at MIT's Artificial Intelligence Laboratory, started to create software tools for the Unix environment for which he made the source code publicly available. In 1983 he launched the GNU[3] project to create a Unix-like environment that would run using only free, open-source software. In 1989, he wrote the GNU General Public License (GPL), which protected the rights of users and developers to use, copy, and modify software—in other words, to freely remix code.[4]

In new media art, interconnections with free software and open-source communities can be traced to the early 1990s, for example, through the release by composer and musician Miller Puckette of Pure Data, a visual programming language for real-time interaction and sound. In the 2000s open-source software such as Arduino/Wiring, Processing, and Scratch, created for both educational and professional purposes, have pervaded the field.

This softwares directly contributes to the remix culture described by Lessig within the niche field of new media art. They are surrounded by rich, multidisciplinary communities who share code on public forums and on source code repositories. As most new media practitioners know, producing technology-intensive art works such as audiovisual performances and interactive installations often involves patching together a number of pre-existing audio, video, hardware, and software components. Thus, artists fix, improve, and create new open-source code such as libraries and code snippets that they then redistribute to the community.[5]

Machine learning is tightly knit with open-source culture. The core software tools currently used for deep learning such as Tensorflow and PyTorch are free and open source. Furthermore, deep learning research communities have historically been defined by a culture of sharing and open access; new results are quickly disseminated on the internet, often with source code, on open-access platforms and journals such as *arXiv* and *Journal of Machine Learning Research*.

It is one thing to create a collage by combining different pictures; it is another to copy-paste code from different sources to create a new piece of software. But what happens when we can remix not just media content such as image or sound but processes that *generate* such content—for example, by applying the voice signature of a singer to one's own voice, or redrawing a portrait in the style of a famous painter? And what new art forms and practices are allowed when we can share, modify, retrain, augment, and otherwise alter machine learning models?

Machine Learning Remixes

While it has become trivial to sample content such as images, sounds, and even programming code, there are two kinds of content that still cannot be easily copy-pasted. First, until recently it was still difficult and expensive to modify certain components of time-based media, such as replacing or modifying the face of an actor in a movie or creating a synthetic voice indistinguishable from the original. Second and perhaps more importantly, it is difficult to copy and paste the generative *process* (or at least a model of the process) allowing the creation of new media content such as a new song, photograph, or painting.

Indeed, copying and remixing the *style* of an author or a certain *genre* is a difficult task that until recently could hardly be automated by technological means. But what if it were possible to automatically get a snapshot not so much of a specific media work but of the process that generated it? What if one could then preserve that snapshot in a digital format? Furthermore, what if we could cut, paste, modify, and rearrange such algorithmic snapshots in the same way that Dadaists such as Hannah Höch were using pictures to make collages and photomontages or that King Tubby was reusing existing music samples to create his own tunes?

Previous attempts have been made to imitate styles and genres using computation. For example, in computer graphics the field of non-photorealistic rendering (NPR) focuses on ways to render two-dimensional (2-D) and three-dimensional (3-D) objects in expressive styles; however the algorithms developed usually lack flexibility and scalability because a different algorithm needs to be created for each style (e.g., image filters for watercolor, impressionism, etc.). These algorithms often take years of research and development to design.

What if we could design generative programs as easily as we can create a photomontage? What if we could as easily remix these programs to create new ones?

Deep learning technologies foster the emergence of a new genre of algorithmic remixes that involves models and data rather than source code. Machine learning algorithms are able to turn a media-generation process into a model that can be saved on disk, exchanged, modified, and so forth. Moreover, some of these algorithmic objects are flexible enough that users can modify and mix them with one another to produce new models.

Exploring Pretrained Models

The dazzling development of deep learning and its artistic applications are inseparable from the appearance of huge pretrained neural network models made available to users. These developments are promising considering the associated expertise and computing costs needed to train these systems at the moment. For example, the training costs of the GPT-3 language model is estimated at $4.6 million, which would make its training from scratch inaccessible to the vast majority of artists.

In particular, GPT-3, which is 116 times larger than its predecessor GPT-2[6] is able to infer new linguistic patterns or tasks from textual instructions given by the user. This capacity allows the user to interact with the model in a fluid fashion by giving instructions in natural language. Hence, it constitutes an alternative to traditional programming, where one can actually interact with the system to train it to do what one wants. Writer Gwern Branwen, who has worked extensively with GPT-3, explains that this form of prompt programming feels akin to teaching your pet some new tricks. "You can ask it, and it will do the trick perfectly sometimes, which makes it all the more frustrating when it rolls over to lick its butt instead—you know the problem is not that it can't but that it won't." (Branwen, 2020)

The use of pretrained models such as BigGAN and GPT-3 supports new forms of remix based on the exploration of the quasi-infinite generative spaces offered by these models. However, training these models requires so much resources that only big businesses are currently able to afford it. These models are thus biased and out of the immediate control of their users. This is a major difference with the free software movement that has supported the emergence of software developed by users, for users. The technoscientific open cultures fuelling machine learning research behind BigGAN and GPT-3 remain under the aegis of big multinationals, keeping these tools out of the direct control of its users.

Alternative Faces

Deep learning remixes are not only supported by big IT: community-driven technologies are also appropriated for creative purposes, sometimes in unpredictable ways. In December 2017, it was reported that users in the Reddit community r/deepfakes were using a software tool to swap faces of actors in porn movies with those of other people such as celebrities, thus creating realistic-looking sex remixes. The technology was later banned on several websites such as Reddit, Twitter, and Pornhub (Cole, 2017).

While the motivation of the r/deepfakes community was to generate new pornographic flix, the technology behind that phenomenon is a user-friendly software called FakeApp that can be applied to any kind of video content. One important trend is the insertion of actor Nicolas Cage in Hollywood movies such as *Forrest Gump*, *Raiders of the Lost Ark*, and *Superman*. Another popular style involves remixes of politicians, such as a remix featuring the president of Argentina, Mauricio Macri, superimposed on Adolf Hitler's character in a scene from the film *Downfall*, and several experiments involving *deepfaking* US president Donald Trump.[7]

Other recent developments allow even more fine-grained modification of audio and video content. Suwajanakorn, Seitz, & Kemelmacher-Shlizerman (2017) present a method to generate new videos of people speaking in a frighteningly realistic way. These computer scientists at the University of Washington trained a neural network on many hours of footage from US ex-president Barack Obama. In a new form of digital puppetry they were then able to generate convincing video sequences of Obama using only audio records of his voice (Suwajanakorn, Seitz, & Kemelmacher-Shlizerman, 2017).

Artist Mario Klingemann used a similar technology in his 2017 video piece *Alternative Face v1*. In this work the ghostly figure of deceased French *yé yé* singer Françoise Hardy appears as an animated collage of original footage taken from different sources. The

Figure 10.1
Still from *Euro(re)vision* two-screen video, 2019 by Libby Heaney. Courtesy of Libby Heaney.

soundtrack is taken from a well-known interview with counselor to the president Kellyanne Conway on CNN on January 22, 2017, in which she used the expression *alternative facts* to describe press secretary Sean Spicer's incorrect statements concerning the number of attendees at president Donald Trump's inauguration. Hardy seems to be looking at the camera and following the movements of Conway's face, eyes, and lips in the original footage, producing a chimerical creature constantly morphing between reality and fiction. Klingemann's goal with producing this work was to mark the beginning of a new era of alternative facts in which people cannot trust their own eyes anymore (McMullan, 2018).

More recently, British artist Libby Heaney used a similar technique to create *Euro(re)vision* (2019) (see figure 10.1). In this dual-screen video piece, two well-known female EU leaders (chancellor of Germany Angela Merkel and UK prime minister Theresa May) who were important figures during the Brexit negotiations perform in a contest-like setting that recalls legendary singing competition Eurovision. Throughout the performances, the glitchy figures of the two leaders speak in alien languages that sound like their native tongues—perhaps as they would sound to a nonspeaker of those languages. The parodic speeches eventually converge to a strange poetic mix between English and German, which Merkel and May perform simultaneously, symbolizing for Heaney the possibility of unification in the face of adversity.

The artist used different machine learning techniques to create this algorithmic collage directly inspired from Dada poetry and more specifically Hugo Ball's sound poems, which Ball describes as "verses without words" (Ball and Pinoncelli, 2011). The lyrics were composed from three different character-level RNNs trained on different political text corpora: one on the German Federal Parliament debates, one from the UK House of Commons, and one on a mix of those two.

To create the work, Heaney herself performed in front of the camera, reading and singing the jabberwockian generative texts as she impersonated the two women and then used deepfake to replace her own face with the politicians' faces. Throughout these contemporary performances of technological puppetry, the two animated caricatures appear dressed up in sequined gowns, gesticulating and speaking in tongues. The imperfections of the face replacement technology, which reveals itself at times in image glitches, have been left for the viewer to see.

Heaney's satiric work is a critique of political rhetoric and of its interconnections with artificial intelligence in a post-truth era—in which anyone can literally put words into the mouths of politicians. Her work embraces an aesthetics of technological glitch, which Heaney calls "good noise," which depends on using machine learning systems "in ways for which they were unintended," in other words, to "shake them, almost to [the] breaking point, to see what comes out, to see ourselves anew" (Heaney, 2019).

Both Heaney's and Klingemann's works reveal the ethical, legal, and social repercussions of the new forms of media alterations made possible by deep learning tech. In the 1990s, image editing software such as Photoshop profoundly transformed advertising, culture, and the public's relationship to the news. By expanding media manipulation into the realm of video and audio, deep learning technologies in the age of fake news will contribute to the transformations previously initiated by media editing technologies.

Remixing the Generative

There is a profound consequence underlying machine learning software's ability to *photoshop* audio and video content: the ability to automatically create and remix generative *processes* of media production. This is the natural result of the fundamental properties of contemporary machine learning systems. As we have described, machine learning algorithms are able to automatically transform generative processes into data structures (models) that can be stored, copied, and modified. These processes can then be used to generate new, unforeseen media content such as images and sound and also new agent behaviors for games and robotic art.

Consider for example the emergence of digital technologies that allow the creation of synthetic voices able to convincingly imitate someone's voice on the basis of a few samples. These digital voices are in fact complex mathematical functions embodied in a neural network's weights that take text and transform it into words spoken using someone's voice. In other words, it is a record or sample not of the voice itself but of the process that can generate the voice. Once trained, this process is turned into a piece of data that can be copied, pasted, and modified.

The ability provided by deep learning systems to automate the creation of a content-generation program is unprecedented in human history at least to such a scale and with such flexibility. Another example is *style transfer*, a technology that allows the application of a specific style to an image. The first paper on the subject appeared in 2015 on *arXiv* (Gatys et al., 2015) explaining how to use a convolutional neural network in which the representations of content and style are separated and can thus be manipulated independently to produce new images.

Similarly, research project Flow Machines aimed to transform musical styles into computational objects that could easily be remixed by composers and musicians. In September 2016 the group released two songs whose scores were generated by a style-imitating algorithm—one was trained on scores from the Beatles and the other on scores by a mix of

Figure 10.2
Holly Herndon and Jlin, *Godmother*, 2018. Still from the music video. Courtesy of Beggars Group Media Limited.

American songwriters. Not long after, Google research group Magenta released a software tool called NSynth that can generate new digital instruments by mixing models of musical instruments (for example, a hybrid between an electric guitar and a trumpet).

A recent work by electronic musicians Holly Herndon and Jlin, in collaboration with Mat Dryhurst and Jules LaPlace, provides a strong example of this new form of algorithmic remixes. The performed song, called *Godmother* (Herndon, 2018a), was generated using a complex set of algorithmic processes embodied in a "machine intelligence" named Spawn to which the artists attribute a true authorship (see figure 10.2). The system employs a kind of custom-built of augmented vocoder that uses style transfer, allowing Holly Herndon to sing using Jlin's voice. The piece uses vocal fragments to produce an uncanny montage of machininc beatboxing. According to the creators, this feature was entirely "dreamed up" by the algorithm, presumably "learning from the stops and starts in Holly's speech" (Kirn, 2018).

In an official statement about the work, Herndon referred to Spawn as her infant, describing how the piece was created by the "nascent machine intelligence" listening to her "godmother" (Jlin) and reinterpreting her art in the voice of her "mother" (Herndon). "In nurturing collaboration with the enhanced capacities of Spawn," she said, "I am able to create music with my voice that far surpass the physical limitations of my body." She adds:

> Going through this process has brought about interesting questions about the future of music. The advent of sampling raised many concerns about the ethical use of material created by others, but the era of machine legible culture accelerates and abstracts that conversation. Simply through witnessing music, Spawn is already pretty good at learning to recreate signature composition styles or vocal characters, and will only get better, sufficient that anyone collaborating with her might be able to mimic the work of, or communicate through the voice of, another (Herndon, 2018b).

The scalability of deep learning techniques posits that these early examples will open the door to many new creative options, profoundly impacting art practice and outcomes. New techniques of video style transfer will allow the hybridization of visual styles sampled from existing movies and other visual content. Romance novels could be automatically rewritten in the style of Margaret Atwood, and *The Handmaid's Tale* turned into a coming of age novel. New variations will be automatically generated from a single movie genre or mixes of many (e.g., dramatic comedy and zombies). Machine learning will allow the creation of new character personalities for fiction, movies, and video games, including speech, movements, appearance, and other characteristics, by mixing existing characters.

It will also be possible to make transfers across different media forms. For example, artists will be able to automatically generate a soundtrack based on a video scene captured from their smartphone, or correspondingly to generate image filters that match a soundtrack. It will also be possible to transpose natural patterns found in nature, such as the movements of a flame, into text, sound effects, or the behavior of a robot.

With this revolution in remix culture will come a number of challenges and issues related to authorship and copyright. Herndon warns about the ethical implications of these new forms of remixing, which could result in a "permission-less mimicry" fostered by a "data-driven new musical ecosystem surgically tailored to give people more of what they like" with decreasing regard for the artistic identity of ideas. Referring to experimental composer George Lewis, she argues for a "more beautiful, symbiotic, path of machine/human collaboration" offering an opportunity "to reconsider who we are, and dream up new ways of creating and organizing accordingly" (Herndon, 2018b).

An AI Opera

Euro(re)vision and *Godmother* are distinguished by the integration of several different techniques. The piece *Legend of Wrong Mountain* (2018) takes this principle to the extreme, in an attempt to create a Chinese *kunqu* opera entirely generated by learning algorithms (Huang et al., 2019). Created by an interdisciplinary team of artists, computer engineers and designers, the project is described by the authors as "a machine's attempt at *Gesamtkunstwerk*,"[8] the project is a complex technological remix integrating a plethora of learning algorithms trained on four different sets of data.

Every component of the piece is thus computer generated. To create the music score, a RNN was trained on a hundred images of traditional Kunqu sheet music. For the script, the authors relied on a custom hierarchical system of Markov chains trained on a database of sixty traditional Kunqu scripts. This approach, they claim, allows them to preserve a logical structure with chapters, dialogues, and lyrics. The script was used for the generation of a video performance and also rendered as a computer-generated book of the script in traditional woodcut style, which was also generated by a neural network. Finally, the music and script are accompanied by generated visuals created from texture analysis and performer video clip poses as well as scene images.

Whereas the result is a bit shaky, the project is nonetheless relevant as an experiment in machine learning art. It pushes the automation of a creative process to the extreme by the interplay of different techniques and data sets, all made possible by the distribution of machine learning tools under open-source licenses and by the availability of data sets. *Legend of Wrong Mountain* sets the stage for future machine learning based total works of Wagnerian magnitude in which data and algorithms are intermingled in spectacular machinic collages.

Conclusion

Art always happens in context. From mere inspiration to copy and counterfeit, creators have always used each other's works to support their own. In the twentieth century the automated reproducibility of media content led to the emergence of a remix culture that plainly embraced the production of derivative works.

Until the emergence of deep learning, remixes always somehow followed the principle of the collage. Whether it was printed images, poetry lines, sound tracks, or code snippets that were remixed to generate new content, the basics remained the same: copy; paste; repeat.

New machine learning technologies such as deepfakes just push this principle further, expanding the possibilities offered to artists. However, machine learning also opens the door to entirely new forms of remixes mediated by models—autonomously adapted structures processes crystallized into data structures. This happens, for example, through the indirect influence of machine learning models via the remix of training sets. Furthermore, the current fascination with machine learning has pushed a range of productions in which multiple media such as text, sound, and image are automated using learning algorithms to create new works.

Perhaps more importantly, machine learning introduces the possibility of not just remixing content but also of generating media with processes that use technologies such as style transfer. The work of musicians Holly Herndon and Jlin is emblematic of this new kind of algorithmic remix, in which a performer can now sing with the voice imprint of another singer. We are only at the beginning of this technological revolution and as we move forward, the capabilities of machine learning systems are likely to drastically transform the conditions of artistic production and the landscape of remix cultures.