Universiteit Leiden
The Netherlands

Performative transactions: worlding compositional ecosystems

Lukawski, A.

Citation

Lukawski, A. (2025, November 21). *Performative transactions: worlding compositional ecosystems*. Retrieved from https://hdl.handle.net/1887/4283663

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4283663

Note: To cite this publication please use the final published version (if applicable).

Chapter 1. Artists as Operators

Navigating the Complexity of Musical Space

Although what can be considered a musical work comprises countless material and immaterial parts—scores, manuscripts, performances, recordings, economic value, aesthetic judgments, historical contexts, symbolic aura, and more (de Assis & Łukawski 2024, 1)—even if we focus only on sound, the imaginary space of all possible music would still constitute an enormous space of possibilities. As observed by Martin Rohrmeier—researcher and director of the Digital and Cognitive Musicology Lab at the École Polytechnique Fédéral de Lausanne (EPFL)—even the results of well-defined rule systems such as counterpoint or voice-leading already "span an enormous search space, which is hard to traverse and affords for rare, original solutions and strategies to be found" (Rohrmeier 2022, 53). A quick thought-experiment shows the scale. Write twelve chords using only the span of two chromatic octaves (we could imagine it is a harmonic material for a miniature to be performed on a toy piano). Any chord may contain any subset of those notes, from none to all twenty-four. The number of all possible solutions is about five million times larger than the estimated number of protons in the observable universe (Tegmark 2003, 462) ($2^{24\times12} = 2^{288} \approx 4.97 \times 10^{86}$). Such a space is so large that not even the most powerful supercomputer could check all options one by one within any reasonable timeframe.

The problem evokes Jorge Luis Borges' *Library of Babel*, a thought experiment imagining a library containing every possible combination of letters across countless books (Borges, 1941/1962; see Bottou & Schölkopf, 2023 for a discussion on Borges and AI). In Borges' story, the library is so vast that any meaningful book is lost among endless volumes of random text. Similarly, the musical "library" generated by such simple parameters vastly exceeds what could ever be systematically explored, placing the search for musical solutions beyond the reach of exhaustive computation or human memory. Statistically speaking, one could imagine a monkey seated at a grand piano, randomly pressing keys for millions of years. Given enough time—and enough random strikes—it is not impossible that the monkey would eventually stumble upon a performance of a Chopin Mazurka (this is known as an infinite monkey theorem). Assuming we have no time to lose, and no army of monkeys at our disposal, we need a smarter way to search.

Despite the difficulty of the described problem, musicians do not rely on infinite randomness to discover meaningful structures. In fact, musicians effectively handle enormous musical spaces every day. Improvisers craft new progressions on stage in real time, and composers sketch ideas long before they test every alternative. It is possible, because the brain is built for good-enough answers when time and data are scarce. As Samuel Gershman—researcher in computational cognitive neuroscience—writes, "the brain is evolution's solution to the twin problems of limited data and limited computation" (Gershman 2021, 154). To visualise the problem of limited data, imagine that you stand in front of a tree and wonder if your ladder will reach the lowest branch. You have never measured the tree, but using your own height, the distance to the trunk, and the way the trunk tapers, you form a quick estimate. It is not exact, yet usually close enough to decide whether to fetch the ladder. The problem of limited computation can be pictured with another example. After a concert you must catch the last train. Rather than modelling every possible route, you recall the simple rule "avoid the main avenue at rush hour" and pick a side street you know. The path may not be the absolute shortest, but it costs almost no mental effort, and you reach the station on time. Artificial Intelligence research copies these strategies. Deep Learning pioneers admit that a large part of the inspiration for deep learning is the human brain (Bengio 2021, xviii). Thinking of music as a multidimensional space of parameters—where pitch, rhythm, dynamics, timbre, and other characteristics each define independent dimensions—is not uncommon. For example, composer Davor Vincze—who often uses electronics and AI tools in his practice—envisions musical works as collections of points in a multidimensional matrix, an idea he operationalises in his microllage compositional strategy (Vincze, 2023, 312–314). However, as Vincze notes, manually mapping even a simple musical piece into such a space would be "a tedious and next-to-impossible task" (Vincze, 2023, 312).

Artificial Intelligence research combines the brain-like strategies with the ability of computers to quickly process large amounts of structural data. When a neural network is trained on musical material, it learns so-called latent spaces—internal maps where similar events lie close together (Audry 2021, 104). Intuitively, a latent space is the model's own map of the music it has "heard": chords, rhythms, or timbres; the model judges similar cluster together in the same region. Technically, each layer of the network turns its input into a new set of numbers. After training, these numbers form a distributed representation: no single unit means "major triad", but a pattern across many units does. The model has compressed the raw, high-dimensional data

into a smoother, lower-dimensional surface where gradient search is possible. A process that imitates what human brains do in the process of learning enables computers to solve problems that were previously reserved exclusively for humans (Bolt, 2023, 97).

The key premise of this chapter is, then, that artificial intelligence systems are increasingly able to automate tasks long considered central to human musical creativity, such as navigating large possibility spaces, identifying patterns, and producing stylistically meaningful output. However, such compression brings a risk. Bias helps us find useful solutions quickly, but it can also hide unusual ones that might matter artistically. Creative work therefore juggles two needs: focus—to locate workable ideas fast—and openness—to leave room for surprise. To understand how artists engage with such systems, this chapter investigates the conceptual and practical foundations of AI-assisted composition. What follows is both a theoretical reflection and a practical report—much of the material presented here was developed and tested through a postgraduate course I designed as part of this doctoral trajectory and taught in 2023–2025 at the Conservatorium van Amsterdam and the Iceland University of the Arts. The course introduced students to creative uses of AI in composition, combining technical experimentation with artistic reflection, and it played a central role in shaping the methods and examples discussed in this chapter.

The chapter is structured in three parts. **Part I – On Automation** sets the conceptual groundwork by comparing symbolic and connectionist AI, then examines how automation reshapes creative labour and artistic agency. **Part II – On Methods** presents practical compositional strategies: first, workflows that generate symbolic musical data (MIDI) with machine-learning models and a discussion of a case-study composition *Ani(mate)*; second, interactive approaches that enlist large language models as conversational or performative agents. The **Conclusion – Artists as Operators** synthesises these strands, redefining the composer as an active operator who navigates, curates, and contests semi-autonomous systems.

Part 1—On Automation

Two AI Lineages: Connectionism and Computationalism

Artificial Intelligence (AI) is a notoriously elastic term, but its coinage is quite precise. Computer scientist John McCarthy introduced the term when he convened the 1956 Dartmouth Summer Research Project—a six-week workshop that aimed to discover "how to make machines use language, form abstractions, and improve themselves" (McCarthy, Minsky, Rochester, & Shannon, 2006/1955). From the start, then, AI referred not just to a technical agenda but to an ambitious cultural dream. The early researchers already envisaged "programs that pushed the boundaries" of machine capability (Mollick, 2024, Part 1, chap. 1). Six decades later, Tiziana Terranova—a theorist and activist whose work focuses on the effects of information technology on society—can describe AI as both the hidden operating system of daily life—controlling logistics, finance, and social-media feeds—and a magnet for enduring hopes and anxieties about consciousness, agency, and control (Terranova, 2023, viii). Beneath that lore lie two technological lineages whose rivalry still organises contemporary debate.

The first lineage often labelled symbolic AI (also computationalism or "good old-fashioned AI"), approaches intelligence as the orderly manipulation of explicit symbols under hand-written rules. In such systems every fact is encoded as a discrete token—for instance the rule IF chord_is_major THEN mood_is_happy—and reasoning proceeds by syntactic inference. Early milestones include Newell and Simon's *Logic Theorist* (1956), often considered the first program to prove mathematical theorems, and the expert systems of the 1980s, whose thousands of if-then clauses diagnosed diseases or advised mineral prospecting. In each case cognition was specified in advance by the programmer.

The second lineage, connectionism, seeks to model intelligence by training adaptive networks to autonomously identify statistical regularities in data rather than by specifying rules in advance. A foundational conceptual step was taken by Canadian psychologist Donald O. Hebb, who proposed that neurons strengthen their mutual connections when repeatedly activated together, thereby forming distributed "cell assemblies" that encode experience (Hebb, 1949). Building on that principle, psychologist Frank Rosenblatt introduced the Perceptron (Rosenblatt, 1958), a simple

learning algorithm that trains a model to classify handwritten symbols by incrementally adjusting the strengths of the connections during training. In parallel, Oliver Selfridge proposed the Pandemonium model (Selfridge, 1959), which described pattern recognition as a hierarchy of simple processes. At the lowest level, basic "detectors" identified simple features, such as lines or curves. Higher levels combined the outputs of these detectors to recognise more complex structures, such as letters or symbols. Selfridge called these competing units "demons", a term that draws metaphorically on the idea of autonomous agents—echoing earlier scientific metaphors such as *Maxwell's demon*, which imagined independent entities making selective decisions based on observed inputs.

These early prototypes established two characteristics that continue to define contemporary neural networks. First, learning is achieved by optimising weights rather than editing symbolic code; as the authors of the Future Art Ecosystems briefing note, the model "is not manually defined through a sequence of instructions" but is produced by algorithms that search large datasets for regularities (Serpentine Arts Technologies 2024, 25). Second, pattern recognition is hierarchical: successive layers formulate progressively abstract representations, a principle illustrated by mid-century neurophysiological research on feature detectors in the frog retina (Lettvin et al., 1959; as discussed in Tilford, 2023, 128).

Connectionism nevertheless entered a period of reduced funding after Marvin Minsky and Seymour Papert (1969) demonstrated a serious limitation of early neural networks. A single-layer perceptron could only separate input patterns using a straight line, which meant it could solve simple tasks like distinguishing light from dark, telling whether a point lies above or below a diagonal on a plane, or deciding whether a musical note is higher or lower than a given pitch, but not slightly more complex problems. It could not compute the so-called exclusive-or (XOR) function—a basic logical operation that returns true if either of two inputs is true, but not if both are. Because the XOR problem cannot be separated by a straight line, it exposed that single-layer perceptrons lacked the flexibility needed even for elementary forms of reasoning. This finding severely undermined optimism about neural networks at the time, and research attention shifted back to symbolic approaches.

Only decades later, when new training methods, faster hardware, and large-scale datasets became available, did multi-layer networks re-emerge as a viable path. The crucial breakthrough was the development of backpropagation algorithms, which

made it possible to efficiently adjust the weights of neurons across multiple layers (Rumelhart, Hinton, & Williams, 1986). By computing how much each connection contributed to the final error and propagating corrections backward through the network, backpropagation allowed deep architectures to learn complex, non-linear relationships—such as the XOR function—that single layer models could not handle.

In parallel, theoretical work strengthened the case for connectionism. The universal approximation theorems proved that sufficiently large neural networks could approximate any continuous function to an arbitrary degree of accuracy (Cybenko, 1989; Hornik, 1991), showing that the architectural limitations identified by Minsky and Papert could be overcome when deeper structures were employed. Combined with the exponential growth of data and computing power, these advances made training multi-layer neural networks practically feasible for the first time. The resulting field of deep learning retains the perceptron's core mechanisms—weighted connections, layered abstraction, and data-driven optimisation—while scaling them to contemporary computational resources.

Deep learning became practically viable only after several key advances converged. New training methods such as unsupervised pre-training made it possible to stabilise very deep networks (Hinton, Osindero, & Teh, 2006, as discussed in Audry, 2021, 98). At the same time, the rise of inexpensive graphical processing units (GPUs) and the availability of web-scale datasets provided the computational resources and data volume necessary to train large models effectively (Audry, 2021, 98). By the mid-2010s, deep learning was achieving state-of-the-art results in fields such as speech recognition, computer vision, and natural language processing (Goodfellow, Bengio, & Courville, 2016, 1). Crucially for this study, it also began to underpin a new generation of generative models capable of producing text, images, audio, and other creative media. The commercial enthusiasm that followed led to the development of today's foundation models: vast neural networks trained on massive datasets, whose billions of parameters allow them to generalise across multiple modalities (Serpentine Arts Technologies, 2024, 25).

Since such systems learn from experience rather than from explicit programming, designers relinquish direct control and instead curate training regimes—the data, objectives, and evaluation metrics that steer each model's optimisation. This shift carries political weight: the architectures now driving artistic exploration also underwrite what Martin Zeilinger (2021, 13) calls the "frightful surveillance tools" of

platform capitalism. Artists such as Keith Tilford therefore read neural networks ambivalently, both as instruments that expose our behavioural clichés—no more mysterious, perhaps, than a frog's reflex to leap toward shadow—and as engines for re-imagining creative agency (Tilford, 2023, 128).

In what follows, I focus on this connectionist lineage, the ability of which—to discover latent musical structure without exhaustive rule design—makes it central to questions about the role of artists as operators of tools automating tasks associated with music composition. Symbolic approaches will re-enter the discussion in Chapter 3, where the composer's role shifts from explorer of statistical spaces to explicit system-builder and legislator of musical rules.

Automation and Creative Work

Automation of creative tasks has fascinated and unsettled humans for centuries. Already Aristotle envisioned mechanical instruments relieving enslaved musicians from their endless toil, imagining machines capable of taking over tedious human labour (Clancy, 2022b, 169). Today, economic, writer and social theorist Jacques Attali speculates that artificial intelligence might radically expand creative output, foreseeing a future in which hundreds of symphonies are composed in the unmistakable styles of Beethoven or Tchaikovsky—indistinguishable from authentic historical works. For Attali, such scenarios illustrate how the boundary between artistic reality and virtual possibility will blur, dissolving the distinction between past works and future creations yet to be realised (Clancy, 2022c, 8).

This vision resonates with the capabilities demonstrated by recent AI models such as NotaGen, "a symbolic music generation model aiming to explore the potential of producing high-quality classical sheet music" (Wang et al., 2025). NotaGen allows users to specify a composer's name as the target style for the generated music and can generate scores either from scratch or conditioned on prompts written in ABC music notation. While the compositions produced by the model do not yet perfectly simulate the styles of the referenced composer, they often exhibit surprising nuance, evoking an uncanny familiarity that reveals why a listener might recognise the intended composer in the generated work. Further exemplifying the trend of fully automating the music generation is the *Stochastic Pirate Radio* project by Bob Sturm et al. (2024).

This project employs an integrated pipeline of multiple AI models, including SunoAI, Boomy, and Stable Audio for generating audio content, alongside various Large Language Models for textual content, and additional models for speech synthesis. This aggregated system, which simulates a complete live radio station evoking the radio stations in the early Grand Theft Auto game series, is continuously live-streamed on YouTube, situating itself within a broader tradition of generative AI live-streams. Notably, the authors reference earlier influential examples such as Dadabots' *Relentless Doppelganger*, a stream which has been generating an uninterrupted flow of metal music since September 4, 2019, and remains active in various iterations to this day.

The automation of creative labour poses a real threat to artists whose societal roles and livelihoods often already exist in precarious balance. Many artists express a profound unease about being rendered obsolete by generative AI tools. As reported in the Future Art Ecosystems briefing, there is a widespread fear among artists: 77% feel that AI threatens to replace their jobs and opportunities, echoing broader societal anxieties where 60% of workers across various sectors anticipate significant changes due to AI-driven automation (Serpentine Arts Technologies 2024, 82). Furthermore, the briefing suggests that while AI-driven automation might lead to deskilling in certain crafts, history demonstrates that artistic skills continuously evolve; some disappear, others emerge, and some even experience a revival as cultural needs and technological tools change (2024, 91-92). At the same time, artists who possess skills predating the rise of AI, and whose expertise aligns with the sensibilities these new technologies leverage, hold an essential advantage. Their specialised knowledge positions them uniquely to master and creatively expand the possibilities offered by AI tools, illustrating the continuing value of human expertise even in highly automated contexts (2024, 90). In practical terms, current approaches to AI integration in creative workflows span a spectrum. Some organisations adopt a transformative "Minotaur" approach (the head of a bull and the body of a human), entirely restructuring workflows around AI capabilities. Others take a "Centaur" approach (upper body of a human and the lower body of a horse), empowering individual users to choose precisely where and how to integrate AI into their practice (2024, 66-67). Crucially, the interaction between artists and AI technologies is more complex than simple substitution narratives suggest. Instead, it initiates a dynamic process of feedback and emergence, where artists not only adapt to technological tools but also reconfigure the very notion of what creative work involves. This interplay, according

to Future Art Ecosystems, transcends narrow framings of automation as mere replacement, opening new possibilities for creative expression and collaboration across various artistic domains and technological layers (2024, 81).

Ethan Mollick notes that the current wave of automation uniquely impacts highly compensated, highly educated, and creative roles—jobs traditionally viewed as secure from technological displacement. He raises pointed questions about why clients or audiences should continue paying human artists when AI systems produce similar results rapidly and inexpensively (Mollick, 2024, Part 1, chap. 2). However, despite these concerns, automation in creative fields may not necessarily lead to outright replacement but rather to transformation of existing roles. Historical precedents show automation typically shifts job responsibilities rather than eliminating them. For instance, accountants who once manually calculated figures now use sophisticated software; they remain accountants but perform fundamentally altered sets of tasks (2024, Part 2, chap. 6). This trend suggests a future where artists increasingly curate and manage AI-generated content, focusing on tasks that demand uniquely human qualities such as creativity, critical judgment, and interpretative skill. Mollick further emphasises that despite AI's proficiency, genuine expertise requires humans to retain foundational knowledge and skills (2024, Part 2, chap. 8). He argues against the notion that basic factual knowledge is obsolete in the age of AI; instead, he highlights a paradox where foundational skills become even more crucial because true expertise depends on deeply interconnected knowledge stored in long-term memory. Consequently, the future of creative and expert labour involves both mastering foundational knowledge and becoming proficient in integrating AI tools into professional practice. Indeed, working effectively with AI itself constitutes a new form of expertise, where some individuals may excel dramatically, becoming the "new kings and queens" of AI-assisted tasks, while others may benefit less significantly (2024, Part 2, chap. 8). Moreover, Mollick speculates that the rise of AI could catalyse renewed interest in the humanities. Expertise in fields like philosophy, literature, history, and art uniquely qualifies individuals to interact with AI creatively and effectively, enabling the development of nuanced prompts and innovative applications (2024, Part 2, chap. 5). This suggests a potentially counterintuitive future where humanistic knowledge becomes increasingly valuable precisely because it enhances human capacities to manage and collaborate with sophisticated AI systems. Thus, rather than simply diminishing the role of human creativity and expertise, AI may redefine these roles, placing greater emphasis on uniquely human insights and

interpretative capabilities.

Media theorists Anna Munster and Ned Rossiter (2023, 48) highlight a polarised debate around automation, with some forecasting dire consequences for human labour, while others envisage human-machine collaborations enhancing value and productivity. Crucially, they point out that certain automated processes—such as AI-generated images—can operate indefinitely without direct human oversight, suggesting a future where creative machines operate increasingly autonomously. According to the authors, the emergence of the automated image represents a profound shift in the relationship between humans and machines, where the external world increasingly appears to itself mediated solely through computational processes, creating a continuum of porosity between machine operativity and social structures (Munster & Rossiter, 2023, 51). This transition marks the post-cybernetic era characterised by a redistribution of labour between conscious human cognition and non-conscious machine operations, signaling a diminishing role for human agency. Munster and Rossiter emphasise that traditional humanist notions of performance provide little solace against the backdrop of growing machine autonomy, suggesting instead that critical gestures—such as simulation, fakery, and the use of machine-generated 'tonality'—are necessary to performatively investigate and expose the operational grammar specific to contemporary automated image production (2023, 52-53). They provocatively question the place of human brilliance and creativity when AI assumes artistic roles, raising fundamental concerns about the ongoing significance of human inventiveness in a machine-driven context (2023, 60).

Marek Poliks and Roberto Alonso Trillo (2023, 29)—researchers in music and philosophy of technology—further illustrate how autonomous creative production increasingly diminishes human roles within cultural production and consumption chains. They use the example of ChatGPT to demonstrate how, despite stable or even expanding total labour input for cultural creation, the proportion directly executed by humans is progressively shifted to computational systems. This phenomenon results in fully autonomous chains of cultural production and consumption circulating online content with minimal human oversight, exemplified by bot-generated YouTube channels (previously mentioned *Stochastic Pirate Radio* being an illustrative experimental case), algorithmically optimised Spotify tracks, and entirely artificially generated and maintained websites. They position these developments within the philosophical frameworks of "computational realism", which proposes that the

universe itself operates on discrete, computable principles, and "New Empiricism", a Silicon Valley epistemology considering that "data itself, simply by virtue of sufficient quantity, is a generator of meaning" (2023, 26-27). Critiquing these ideologies, they suggest that reducing experience, intelligence, and consciousness to mere externalised data flows and recursive computational processes significantly flattens human cognitive and experiential depth, relegating human creative roles primarily to passive interpretation—little more than "cloud-reading, storytelling about patterns in the mist". Adding another layer to these discussions, Mattin (2023, 429)—musician known for his stance against copyright and for publishing his works under the no-licence—argues that the genuine threat to human autonomy is less about a singular, dominant artificial intelligence, and more about the complex and unpredictable collective behaviours of human groups mediated by networked computational systems. Mattin suggests that rather than a unified, malevolent AI force, the real danger lies in the subtle yet pervasive computational structures and ecosystems, powered by ordinary algorithms, driving human society into unpredictable and potentially troubling directions.

What to Automate and What to Control

The discussion around automation in creative labour necessarily engages deeply with the concept of autonomy inherent to AI systems. In this context, Artemi-Maria Gioti (2021, 1–2)—composer and artistic researcher working in the fields of artificial intelligence, musical robotics and participatory sound art—emphasises two crucial attributes defined by Michael Young (2008) for what he terms 'live algorithms': empowerment—the ability of algorithms to make independent decisions influencing future actions—and opacity, referring to their inherently complex and nonlinear generative processes. This autonomy can lead to significant indeterminacy, necessitating a form of trust from human collaborators in AI-driven systems to consistently deliver outcomes within anticipated bounds. Sofian Audry (2021, 161) similarly observes that the autonomy exhibited by sophisticated AI systems frequently results in outcomes that surprise even their creators, eluding complete rational comprehension. Audry cautions against overly simplistic narratives claiming that machine learning straightforwardly eases artistic labour or entirely replaces human authorship. Instead, they argue that machine learning reshapes and

redistributes creative agency, cultivating novel relationships between humans and machines within artistic workflows (2021, 159). Indeed, once models are designed, trained, and set in motion, they begin operating with a substantial degree of autonomy, shaping artistic outputs independently of direct human intervention (Audry, 2021, 35). Artists who engage with AI, thus, navigate a delicate balance, managing the high levels of autonomy intrinsic to these systems while strategically directing them to align with their aesthetic intentions (43). This tension between autonomy and authorship, as further elaborated by Audry (292), characterises a fundamental spectrum in artistic practice. Some composers, such as John Cage, are comfortable granting substantial autonomy to external processes, allowing materials and agents to express their intrinsic characteristics freely. Others, like Brian Ferneyhough, construct tightly controlled systems in which every parameter is rigorously specified, ensuring that the outcome adheres closely to a predetermined compositional logic. In such practices, complexity is not a byproduct of autonomous processes but the result of deliberate design, where the composer maintains interpretive authority over every layer of the musical material. In either case, the integration of AI within artistic processes does not represent a mere technological substitution; rather, it necessitates deliberate artistic negotiation between human intentionality and machine autonomy, highlighting ongoing creative tensions and opportunities within human-machine collaborations.

Emanuele Arielli (2021)—philosopher and researcher interested in aesthetics, cognitive science and communication theory— provides a useful classification of how machine learning intersects with aesthetic processes by mapping two fundamental dimensions: object versus subject and description versus generation. From these dimensions, four distinct types of machine learning tasks in aesthetic domains can be identified. First, AI systems can study objects—analysing stylistic and formal patterns in existing artefacts, such as harmonic structures in music compositions. Second, they can generate objects, producing stylistic variations that closely mimic the analysed corpus, thereby extending and reinterpreting existing aesthetic traditions. Crucially, Arielli emphasises that a comprehensive account of aesthetic processes must also consider the dimension of human reception. Thus, the third type involves studying subjects—tracking, modelling, and predicting aesthetic preferences of audiences through behavioural data. Finally, the fourth type concerns generating subjects: AI systems can be trained to simulate human aesthetic judgments, autonomously evaluating novel artefacts without direct human input (Arielli, 2021, 12–14).

Although various techniques exist to automate aspects of creative workflows, it is crucial to distinguish between different kinds of automation. For instance, methods such as Self-Organising Maps and K-means clustering can automate the classification of large sets of generated objects, grouping musical fragments based on feature similarity. Meanwhile, social media profiling algorithms automate the classification of subjects, modelling and predicting aesthetic preferences based on user behaviour. From the perspective of Arielli's framework, both activities—classifying objects and modelling subjects—could theoretically be delegated to machines. However, for artists seeking to maintain an active, critical engagement within the creative process, automating these two aspects can be counterproductive. It is precisely through curating musical artefacts (classifying objects) and anticipating the aesthetic responses of listeners (modelling subjects) that composers retain agency, creative intuition, and aesthetic control over their works. The emphasis on maintaining an active human role aligns with recent theoretical work on human-AI co-creativity. Rather than treating AI models as autonomous generators of finished musical works, the recent studies foreground interactive, performative engagements with the outputs of machine learning systems. For instance, the *Human-AI Musicking* framework proposed by Valverde, Agres, Herremans, and Chew (2023), published in the proceedings of The International Conference on AI and Musical Creativity (AIMC), conceptualizes the creative process as a dynamic co-activity between human and machine, where the human maintains a central curatorial and interpretative role throughout the generative process. Within this model, creativity is understood not as a product delivered by the AI, but as an ongoing, reciprocal negotiation in which human aesthetic agency remains paramount.

In AI-assisted composition, two broad modes of generation can be distinguished. Some models aim to produce relatively complete musical outputs in a single pass, requiring the composer to engage primarily in selection and refinement of the initial input. Once the input parameters are set, the model does the rest and generates a full ready-made composition. An example is the NotaGen system (Wang et al., 2025), which generates stylistically conditioned full scores. Other models are better thought of as generators of modular material: smaller fragments, motifs, or textures that invite recombination and creative assembly. In this approach, the composer does not expect a finished work but rather gathers musical strata to construct larger forms.

These very broadly described generation strategies could be seen as, importantly, independent of any tool or platform and operate across different forms of musical data, whether symbolic (e.g., MIDI files, piano-roll representations) or audio-based. Nevertheless, symbolic data currently afford far more precise manipulation of musical structures than audio data, offering finer control over parameters like pitch, rhythm, dynamics, and articulation. While advances in audio generation are rapidly narrowing this gap, symbolic representations remain particularly advantageous for composers seeking detailed, granular level of control over musical material. This distinction, while important today, is likely to blur over time as machine learning models for audio synthesis continue to improve.

Part 2—On Methods

Teaching Creative Workflows with Artificial Intelligence

Much of the material discussed in this chapter stems from an experimental composition course I developed and taught during this doctoral research trajectory, titled *Posthuman Creativity Labs: Artificial Intelligence and Blockchain in Music*.

The course was first taught in 2023–2024 as a master's elective at the Conservatorium van Amsterdam, where it unfolded over the academic year through a series of monthly seminars and workshops. Each session combined short theoretical lectures with hands-on experimentation, and students were asked to bring back creative assignments using the presented technologies. Students were invited to reflect on how the tools they used changed their approach to composition, their sense of authorship, and the identity of the musical material they produced. The student cohort included composers, sound designers, and performers, many of whom had no programming experience. This constraint shaped the course's design: it favoured accessible, Graphical User Interface-based tools like Magenta Performance RNN, Riffusion, and ChatGPT, while also scaffolding more advanced practices for those who wished to go deeper.

In a second iteration of the course, also hosted at the Conservatorium van Amsterdam but now as part of the *Creative Performance Labs* program in January and February 2024, the structure was condensed into four intensive sessions. This time, the focus shifted more explicitly toward preparing a final performance: the course culminated in a presentation of students' works. Because of the compressed format, the second iteration built directly on the practical experiences and examples developed in the first version. It featured refined teaching strategies and compositional tasks that had proven effective, while also placing greater emphasis on workflow design and real-time interaction with AI systems. Again, the class included students from a broad musical background, most without programming knowledge, which required maintaining a high degree of technical accessibility without simplifying the conceptual ambitions of the course.

A third institutional iteration took place at the Iceland University of the Arts in Reykjavík, where the student cohort included individuals with prior experience in computer programming, including live-coding practices using programs like

SuperCollider. In response, this version of the course adopted a more advanced technical orientation, building upon the methodological foundations laid in the earlier versions. Students were invited not only to interact with existing AI tools, but also to prototype their own workflows, combining generative reasoning models (such as ChatGPT) with sound synthesis and control environments. In several cases, this led to the development of custom interfaces or API extensions that allowed for real-time interaction between reasoning models and live-coded music systems.

Finally, the course was presented and adapted in various guest lectures to different musical communities, including classical composers (at Geneva University of Art and Design), film music students (at ArtEZ University of the Arts), and artistic researchers (at Orpheus Institute). These presentations tested the portability of the core workflows and confirmed their utility across a wide range of artistic contexts. In each case, the course prompted its audience to rethink not only what composition entails, but what it means to collaborate with systems that are both generative and opaque. In all its versions, the course pursued a dual aim: first, to demystify AI and blockchain by giving students concrete tools for integrating these technologies into their artistic practice; and second, to foreground the conceptual and aesthetic shifts introduced by these technologies. Topics included the structure of latent spaces, the automation of creative labour, the reconfiguration of authorship, and the emergence of posthuman agency in music. Students were encouraged to consider not only what AI systems can generate, but how they generate, and under what assumptions. The guiding questions were both technical ("how do I get this to work?") as well as operational and reflective: "what kind of composer do I become by using this?", "what kind of work is this tool capable of producing?", and "what parts of the process am I delegating, shaping, or reclaiming?".

The course served as a live, iterative creative testing ground for both artistic inquiry and conceptual development of this doctoral trajectory, as well as a generator of engagement with the topics discussed in this thesis. Throught experimenting with various AI tools in the context of students' creative practices, it enabled a testing ground for the iterative development of compositional workflows with new technologies. This led to developing two distinct compositional workflows: one that I further describe as *Generate, Select, Edit, Continue*, and another based on prompt-responsiveness and iterative model control, titled *"Could you continue my MIDI, please?"*. Both emerged from experiments in the classroom, where students negotiated

the limits and possibilities of machine-generated musical material. These methods—presented in the following sections—do not claim to be comprehensive or universal, but they illustrate how composers can function as operators: users who activate, guide, and curate automated systems within constrained creative tasks.

In this sense, the course was not only pedagogical but also research-driven. It allowed for direct observation of how AI systems affect compositional decision-making, how artistic intentions are refracted through non-human agents, and how authorship becomes distributed across systems of generation, selection, and refinement. The role of the composer as operator—proposed at the end of this chapter—is grounded in this experience of collaborative learning, experimentation, and critique.

Part 2.1—Composing with AI Models Trained on Music

Generate, Select, Edit, Continue

Currently, most AI models capable of generating high-quality audio outputs are trained predominantly on datasets of popular music. Examples include models like Suno, Udio, and Stability AI's Stable Audio, all of which excel at producing tracks in popular genres. However, there are no widely available models capable of producing high-quality, stylistically sensitive generations of experimental or contemporary classical music. There are several reasons for this. First, by definition, experimental music resists standardisation, making it difficult to assemble sufficiently large and coherent datasets for training. Second, much of the contemporary music repertoire remains under copyright protection, limiting the available training material. Third, training state-of-the-art audio models remains prohibitively expensive, and commercial entities have little economic incentive to invest in models aimed at small, specialised audiences. Perhaps most fundamentally, experimental music often prioritises exploration of sound-production methods themselves rather than adherence to recognisable stylistic patterns. As a result, generative models trained purely on existing examples would struggle to capture the experimental ethos unless they were specifically designed to model creative processes rather than surface outputs—a challenge that touches on deeper questions about the role of experimentation and innovation in art music, to which we will return in the next chapter. Nonetheless, even models primarily trained on popular music corpora can

often be creatively "hacked" or subverted for experimental purposes. For instance, at the course I taught at the Conservatorium van Amsterdam, some students took the initiative to experiment with ElevenLabs' Voice Changer model, which is normally designed to take a human voice as input, and output the same spoken text using a cloned voice. Instead of using speech, they recorded the sounds of musical instruments and fed them into the model. The result was a striking hybrid: the instrumental sounds were transformed, carrying an unexpected blend of instrumental timbre and human vocal qualities, creating an uncanny and novel listening experience. By exploiting their biases, breaking their expectations, or creatively manipulating their inputs and outputs, composers can repurpose these models in ways that defy their original design intentions. For the purposes of this chapter, however, which focuses on compositional workflows based on using AI systems in ways roughly aligned with their intended operation, it is useful to examine concrete examples based on symbolic music generation.

The first example is Performance RNN model (Simon & Oore, 2017), developed within Google's Magenta project. Performance RNN was trained on a dataset of expressive, good quality performances (Yamaha e-Piano Competition dataset), specifically MIDI recordings rich in timing deviations, dynamic variations, and articulation nuances—elements crucial to human musical expressivity. The model is optimised not simply for generating correct notes, but for producing new performances that imitate the timing, dynamics, and phrasing characteristics found in the training data. In this sense, it represents an important step beyond earlier symbolic models that focused purely on pitch and rhythm without expressivity (Briot, Hadjeres, and Pachet, 2020, 190). Crucially for non-programming musicians, Performance RNN can be used without technical expertise through an online demo page provided by Magenta, allowing real-time interaction with the model. This makes it a convenient tool for demonstrating AI capabilities to non-coding artists, for instance in a conservatoire context. The generated MIDI output can be routed to a Digital Audio Workstation (DAW) or into a score notation software, enabling musicians to record, edit, and manipulate the AI-generated performance as part of their compositional process[1]. The model's output can be further conditioned by constraining it to selected scale patterns,

[1] *To enable this setup, users must create a virtual MIDI channel. On macOS, this is done through the Audio MIDI Setup utility: Audio MIDI Setup → Window → Show MIDI Studio → Double-click IAC Driver → Enable "Device is online". Windows users can achieve a similar setup using external tools such as the free software LoopBe1.*

either manually via the demo interface or dynamically through MIDI input.

When recording the output of Performance RNN, composers can interact with the model in real time, influencing its responses and capturing the results. This might be seen as situating the process within a tradition of using improvisation as a compositional method. As Artemi-Maria Gioti notes in her dissertation on distributed creativity (Gioti, 2021), composers like Giacinto Scelsi famously used improvisation as a strategy for composition, recording spontaneous sessions and later transcribing or adapting them into notated works (78). Similarly, the concept of "seeded improvisation" by composer and researcher Richard Barrett describes a hybrid approach in which composed materials are interspersed with spaces for improvisation, allowing emergent musical phenomena that could not arise from notation or free improvisation alone (Barrett, 2014, as emphasised in Gioti, 2021, 78). Within this lineage, using Performance RNN can be seen as a digital extension of these improvisatory compositional strategies. The AI model serves not as a generator of finished works, but as an improvisational partner producing rich musical material to be further edited, selected, and assembled into coherent artistic structures.

It is important to emphasise that such generated material should not be treated as performance-ready or self-sufficient. On the contrary, its greatest creative potential lies in treating it as raw material for further experimentation. A common strategy among artists working with generative systems is to produce large quantities of varied outputs before any compositional decisions are made. This abundance creates a fertile space for exploration, enabling composers to navigate, classify, and sculpt the generated material according to the needs of their larger artistic vision. By maintaining a clear separation between the generative phase and the curatorial phase, composers preserve their aesthetic agency and integrate AI models into a broader experimental workflow. However, once such musical material is prepared and ready to be incorporated into a larger musical work, there are additional strategies that can be employed to further develop it using AI tools, assuming the composer wishes to continue engaging with generative processes. Two particularly useful approaches are continuation and generative in-painting.

In the continuation approach, an AI model is tasked with generating a stylistically coherent extension of an existing musical fragment. The composer provides an initial musical prompt—whether written manually or generated earlier by another model—and the system predicts how the musical material might plausibly continue based on

the stylistic features embedded in the input. In the generative in-painting approach, the model fills in the musical gap between two given fragments, creating a smooth and stylistically unified transition. This technique is particularly valuable for connecting independently generated or composed sections into a cohesive musical whole.

Both strategies are made accessible through the Allegro MIDI Transformer model, developed by programmer Alexandr Sigalov as part of his broader Project Los Angeles, which explores novel architectures and applications of AI for music generation (Sigalov, 2020). Allegro MIDI Transformer builds on the Transformer architecture, a deep learning model originally designed for natural language processing, but now widely adapted for symbolic music tasks. The model allows composers to either extend an existing MIDI fragment into a longer sequence or interpolate between two different fragments, producing fluid transitions that can bridge stylistically distinct ideas. Importantly, Allegro MIDI Transformer can be used not only to continue musical material originally written by the composer but also to further develop material generated by other AI systems, such as Performance RNN. This opens the possibility of chaining multiple models together within a single workflow, leveraging the strengths of each at different stages of the creative process.

Bringing these practices together, a flexible compositional framework can be summarised through the iterative workflow of generate, select, edit, continue:

- Generate: Create a large pool of musical material using one or more AI models. This phase prioritises diversity and richness over immediate coherence.

- Select: Curate the generated material, classifying fragments according to their usability, stylistic suitability, or experimental potential within the emerging work.

- Edit: Actively modify, combine, and transform the selected materials to align them with the composer's artistic aims. Editing can involve anything from minor rhythmic corrections to radical re-contextualisations.

- Continue: Use continuation or in-painting models to extend, connect, or reimagine the edited materials, further integrating them into a larger compositional structure.

Through this iterative cycle, composers maintain their central creative role, while AI systems function as sources of inspiration and musical material. Each phase invites human aesthetic judgment, ensuring that the outputs remain embedded within a

broader artistic vision rather than dictated by the statistical patterns learned by the models.

Case Study: *Ani(mate)* (Movements 1–5)

The methods discussed above informed the composition of *Ani(mate)*, a commissioned work for piano and string ensemble. The piece was commissioned by the Chamber Music at Lundsgaard festival to open their concert series at Lundsgaard Gods in Kerteminde, Denmark on the 7th of August 2025, and was also subsequently performed again in Copenhagen on the 12th of August 2025. The commissioners invited me to compose a new work that would explore the creative use of artificial intelligence tools. It was premiered by Trio Con Brio Copenhagen together with the Marmen Quartet, pianist Enrico Pace, violists Hélène Clément and Michael Germer, violinist Andrej Bielow, and cellist Jonathan Swensen. The title, *Ani(mate),* carries multiple meanings. To "animate" is to bring to life, to give motion or spirit. The term *mate* suggests companionship or partnership yet simultaneously references *checkmate* evoking notions of conflict, strategy, or finality—as in the decisive end of a chess match.

The resulting composition consists of six short movements. The first five were composed using the connectionist AI-based methods described throughout this chapter, while the sixth movement was created using a distinct set of methods, which will be discussed later in Chapter 6. This section focuses on the first five movements as a concrete case study of the operator role: how the techniques of "generate, select, edit, and continue" can be used to develop musical material in dialogue with machine learning systems.

In these movements, I worked extensively with the Magenta Performance RNN model to generate a large corpus of MIDI material conditioned by a system of curated pitch-class distributions. By adjusting the input parameters, I could influence the statistical frequency of particular pitch classes in the generated sequences, creating long, unstructured passages of material. I recorded hours of such "seeded improvisations" from the model, then curated fragments that seemed musically promising according to my compositional judgment. This process required the same form of operational decision-making discussed earlier in this chapter: engaging with the model as an

opaque yet responsive environment, and navigating its latent space through iterative exploration.

Once selected, the fragments were imported into a Digital Audio Workstation software, where I edited the MIDI data to remove noise, refine durations, and shape local rhythmic profiles. I then transferred the material into a score-editing software to create a clear piano score. At the final stage of the process, I orchestrated the selected fragments for the string ensemble and piano. The piece exists as a score in two versions: for a single pianist (performed by Jens Elvekjaer in Copenhagen) and for two pianists playing four hands on one piano (performed by Jens Elvekjaer and Enrico Pace in Kerteminde).

While much of the musical material in these movements was derived from model-generated sequences, I was struck by how strongly the resulting music aligned with my existing compositional idiom. This was not an outcome I anticipated. Given that the model was trained on external performance data and conditioned only by statistical constraints on pitch-class distributions, I expected the outputs to diverge significantly from my established stylistic tendencies. Instead, after curating and normalising the material (besides the desired material, the model outputs various types of noise in the generated signal), the resulting music sounded surprisingly similar to what I might have composed without AI tools at all.

This observation highlighted how central the acts of selection and structural positioning are to the compositional process. The final character of the piece did not appear to result primarily from the local content of the fragments produced by the model, but from the decisions about how these fragments were ordered, how they shaped musical tension and release, and how they were evaluated against my own aesthetic preferences. These decision points—shaping temporal progression, controlling suspense, and filtering material according to taste—seem to act as strong attractors that pull diverse material toward a personal style. Even when the source material is heterogeneous, the composer's operational choices can impose continuity and identity. This finding offers an important perspective on the operator role: curatorial navigation of large generative spaces can, paradoxically, intensify rather than dissolve authorial voice.

To test this observation, I composed the fifth movement with a deliberately different intention. Here, I selected a longer fragment from the AI model without intervening

to reshape it according to my usual criteria, and allowed material I would typically exclude as cliché or overly sentimental to remain. To my ear, the resulting movement is markedly more traditional and emotionally overt than the surrounding ones. I decided to retain it in the final score precisely to underscore this artistic point: that the suspension of curatorial filtering can produce a distinctly different stylistic result, making the influence of those filtering decisions visible by their absence.

Another noteworthy aspect of the piece is the tension between its macro-level coherence and its micro-level idiomaticity. Although the six movements form a coherent overall structure, the instrumental writing often departs from idiomatic ensemble practice. This reflects the origin of the material: the model generated only piano performances (not scores), which I then normalised and orchestrated. Because the model was not conditioned on ensemble-specific affordances, the resulting parts occasionally diverge from conventional instrumental habits, even as they contribute to the macro-level continuity of the piece. This disjunction illustrates how integrating non-orchestrated, model-generated material into ensemble writing can produce music that is structurally cohesive yet locally idiosyncratic, and it became an important feature of the work's character. Nevertheless, it is important to remain aware of the limitations inherent in current generative models. One of the major challenges is that AI models are inevitably shaped by the properties of their training datasets.

Biases in Training Data and Artistic Control

To perform well, models must be trained on thousands of examples, but this necessity introduces stylistic and structural biases that can be difficult to overcome. The dataset itself becomes a latent boundary on what the model can generate, and it is often infeasible to manually curate such large datasets for stylistic purity or quality control.

One promising solution to this challenge lies in fine-tuning pre-trained models with a much smaller, curated dataset tailored to the artist's own stylistic preferences. Fine-tuning involves taking an already trained model and adjusting it slightly to better reflect a new, specialised corpus. An illustrative example of this approach can be found in the *Nobody's Songs* project by programmer Sebastian Macchia (Macchia, 2020). Macchia fine-tuned Magenta's Music Transformer model on the music of Erik Satie, adapting the model to generate material more closely aligned with Satie's style.

The generated material then served as the basis for further creative work, ultimately resulting in a full album. In this way, fine-tuning offers composers a way to personalise generative models, allowing for a deeper alignment between AI outputs and individual aesthetic goals. Rather than being constrained by the biases of large, generic datasets, artists can sculpt the creative potential of AI systems according to their own choices.

However, the dependency on large, general-purpose datasets can often reveal itself as a significant limitation to artistic experimentation. Models trained on extensive corpora excel at continuing or extending materials that resemble their training data but can struggle when confronted with radically divergent musical languages. For example, the Allegro MIDI Transformer is highly effective when tasked with continuing a classical music fragment, such as a short excerpt by Mozart. It can also produce relatively satisfying results when asked to extend a post-tonal musical fragment in a style reminiscent of Mussorgsky, Prokofiev, or Bartók—a combination I tested with quite promising outcomes. However, when presented with a strict twelve-tone series, the model's output quickly collapses into nonsensical continuations, resembling geometrical shapes rather than meaningful musical structures. In such cases, the model's internalised assumptions about harmonic progression, phrase construction, and rhythmic regularity—patterns absorbed from the training corpus—reassert themselves, making it difficult to maintain the structural integrity of more experimental idioms. This example highlights a crucial limitation: when working with models trained on broad stylistic data, artists must always account for the latent stylistic biases encoded within the model. No matter how flexible a model may seem, it cannot easily extrapolate beyond the structural habits ingrained in its training data without substantial retraining or fine-tuning.

At the same time, it is important to avoid over-interpreting this limitation as an indictment of generative methods altogether. As Philip Galanter (2003)—theorist, curator, and researcher interested in generative art, sound art, and complexity science—points out, "(…) the generative approach [itself] has no particular content bias, and generative artists are free to explore life, death, love, war, beauty, or any other theme" (Galanter, 2003, 17). In other words, while the data used to train a model may introduce stylistic limitations, the generative methodology remains fundamentally neutral. It is not the act of using generative processes that imposes aesthetic restrictions, but rather the choices surrounding what data, assumptions, and

frameworks those processes are based upon. This distinction is crucial for composers and artists using AI tools: the limitations of existing models should not be seen as intrinsic shortcomings of AI-assisted composition itself, but rather as challenges that can be addressed through thoughtful interaction with the system. One way to overcome these challenges is to provide generative models with richer, more stylistically targeted input data or contextual guidance. However, achieving such fine-grained controllability within current symbolic generation models remains a technical challenge.

In this context, it is worth acknowledging that some recent systems, such as Google's Lyria 2 and Meta's MusicGen, use text as input to guide music generation. In these models, users prompt the system by describing in natural language the style, mood, or instrumentation they wish to hear. However, from the user's perspective text-based prompting does not constitute a fundamentally different generative paradigm, but rather just a different interface for interacting with models that are still grounded in learned latent musical spaces. Their outputs remain shaped by the biases and constraints of their training data, just as with symbolic input models. While the interaction moves from providing musical fragments to crafting verbal descriptions, the underlying generation dynamics and limitations are comparable.

This issue—how to guide generative models toward more stylistically sensitive, context-aware outputs—points toward another compositional strategy. In what follows, we will explore a workflow based on text-based generative AI agent for symbolic music generation. These conversational systems offer a different mode of collaboration, enabling users to "teach" the model specific theories, stylistic principles, or creative constraints through natural language instructions. Rather than relying solely on prior training data, these agents can negotiate musical ideas dynamically, allowing composers to develop musical materials together with the model in an interactive, iterative process.

Part 2.2—Composing with Large Language Models

Large Language Models as AI Agents for Musical Tasks

Among the many forms of Artificial Intelligence, one class of models has recently come to dominate both research and popular imagination: Large Language Models

(LLMs). LLMs are a type of neural network trained to process, predict, and generate human language by learning from vast textual datasets. They represent a specific approach to AI that privileges statistical pattern recognition over explicit symbolic reasoning, continuing the connectionist lineage described earlier. Although statistical models for language have existed for decades, LLMs distinguish themselves by their unprecedented scale and versatility, making it possible for a single model to perform a wide range of tasks—from translation and summarisation to creative writing—without task-specific retraining. The foundational architecture for most LLMs is the Transformer, introduced by Vaswani et al. (2017), which replaced earlier sequential models with a parallelisable attention mechanism capable of capturing long-range dependencies in data. However, it was the scaling experiments conducted by Brown et al. (2020) in the development of the GPT family that revealed the surprising few-shot learning abilities of very large Transformer models, demonstrating that capabilities could emerge from scale alone. Bommasani et al. (2021) later offered a more conceptual framing, describing LLMs not merely as Transformer-based models trained on web-scale corpora, but as systems whose latent representations compress, generalise, and transform broad domains of human knowledge, enabling emergent capabilities that extend beyond their original training objectives. Over time, the Transformer architecture has become so central to AI research that it now practically defines the public understanding of AI itself, with the term "AI" increasingly evoking Transformer-based systems like ChatGPT (Poliks & Alonso Trillo, 2023, 24).

Human reasoning is traditionally anchored in deduction: the logical extraction of truth from known premises. Deduction operates according to formal rules that guarantee validity—if the premises are true and the reasoning steps are valid, the conclusion must also be true. In contrast, the reasoning exhibited by Large Language Models (LLMs) is based not on deduction but on inference: the statistical prediction of the most probable continuation given prior data. LLM predicting "Neil Armstrong" as the first person on the moon does so purely through statistical association, not through comprehension or deduction (McCormack, 2023). LLMs do not know facts, meaning, or truth; they predict the next most likely token based on the patterns learned in training. This distinction explains why LLMs can often produce outputs that appear deductively valid without possessing any actual understanding. As programmer and researcher Andrej Karpathy (2023) explains, while we have a detailed understanding of the architecture and training processes of LLMs—including the mathematical operations, optimisation procedures, and engineering principles

involved—the question of how these models generalise so broadly remains open. Despite being trained primarily on the task of next-token prediction, LLMs spontaneously develop complex capabilities such as reasoning, planning, and abstraction. Research efforts in mechanistic interpretability are actively working to reverse-engineer these internal processes, aiming to map how specific circuits and patterns within the networks contribute to emergent behaviour. While there have been important advances (e.g., in understanding induction heads and simple algorithmic circuits), our current understanding remains partial and fragmented.

Inference imitates deduction sufficiently well for many practical purposes because human reasoning itself, limited by cognitive and informational constraints, often substitutes heuristics and probabilistic estimations for strict logical tasks (such as in the example of estimating the height of a tree at the beginning of this chapter). The distinction between different types of cognitive processing, famously popularised by Daniel Kahneman in “Thinking, Fast and Slow”, offers a helpful analogy for understanding LLM behaviour. Kahneman distinguishes between System 1, which is fast, instinctive, and automatic, and System 2, which is slow, deliberate, and effortful (Kahneman, 2011). As Karpathy (2023) explains, human responses to simple problems like "2 plus 2" are instant because they are cached in System 1, whereas more complex tasks, such as multiplying "17 times 24”, require the slower, rational operations of System 2. Current LLMs, in their core, primarily operate through a process resembling System 1: they produce fast, pattern-based predictions without sustained deliberative thought. However, on top of these systems, various techniques such as chain-of-thought prompting and inference-time scaling are used to simulate the multi-step processes associated with System 2. As LLM research engineer Sebastian Raschka (2025) explains, reasoning in LLMs involves generating answers that require multiple intermediate steps rather than jumping directly to a conclusion. Regular LLMs can answer straightforward factual queries, but true reasoning tasks—such as solving mathematical problems, puzzles, or coding challenges—require models to simulate an extended, stepwise thought process. This can be encouraged more deeply through specialised training methods aiming to scaffold internal representations that can handle multi-step problem solving more effectively, moving LLMs closer to System 2-like behaviour when tackling complex tasks. In that sense, LLMs become versatile models for solving a broad range of general tasks.

An important further development is the concept of AI agents. Unlike standalone LLMs, agents are systems designed not only to reason internally but also to interact with external tools to achieve specific goals. As researchers Wiesinger, Marlow, and Vuskovic (2024) explain in a guide paper on technical application of AI agents released by Google, just as humans use external aids—such as calculators, books, or search engines—to supplement their reasoning, Generative AI models can be trained to use tools via Application Programming Interfaces (APIs) to access real-time information, retrieve data, or trigger actions. An AI agent is thus defined as an application that observes its environment and acts upon it through the tools it has available, aiming to fulfil a specific objective (Wiesinger et al., 2024, 5). The technical mechanism that enables this interaction is the API—a set of tools that facilitates the exchange of data and functionality between systems (Serpentine Arts Technologies, 2024, 24). By calling APIs, agents extend their abilities beyond internal text prediction, enabling them to fetch, update, manipulate, or generate information dynamically. Tools exposed through APIs typically operate through standard methods like GET, POST, PATCH, or DELETE, allowing agents to perform complex tasks such as querying databases, retrieving web data, or interfacing with external software (Wiesinger et al., 2024, 7). To facilitate the use of APIs, agents often employ structures known as Extensions. An Extension teaches the agent how to interact with a particular API by providing examples of correct usage, specifying the necessary arguments, and modelling the format of successful API calls (14). Functions operate similarly but are executed client-side rather than agent-side: a model can output a function and its arguments without making a live API call, leaving the execution to the client (18). Together, the concepts of Extensions, Functions, and APIs build the operational scaffold that enables agents to transform LLMs into active participants capable of modifying and acting upon digital environments.

The ability of AI agents to interact with external tools opens wide-ranging possibilities for creative applications in music. Rather than limiting generative AI models to producing text-based outputs, agents can be configured to perform practical musical tasks by interfacing with digital environments. For instance, an agent can control a Digital Audio Workstation (DAW) to automate recording sessions, adjust parameters, trigger samples, or manipulate MIDI tracks. A web-based platform called WavTool temporarily demonstrated such possibilities by integrating an LLM-based agent directly into a DAW interface. WavTool enabled users to manage the production workflow through natural language commands, allowing the AI to automate tasks

such as recording, editing, and arranging, as well as to trigger embedded AI models to generate, extend, or modify musical fragments within the same environment. Although at the time of writing WavTool has gone offline with a notice that it will return with expanded capabilities, it is likely that similar integrations of AI agents into DAW ecosystems will soon be replicated and further developed by major DAW manufacturers, progressively embedding AI-assisted workflows as standard features within music production software.

"Could you continue my MIDI, please?"

During one of the institutional editions of my course Posthuman Creativity Labs: Artificial Intelligence and Blockchain in Music, taught at the Iceland University of the Arts, the students expressed a strong interest in experimenting with LLMs for music making. At the same time, WavTool—the primary tool I had previously used to demonstrate this capability while teaching in Amsterdam—was no longer available. To enable the use of ChatGPT's reasoning and text-writing capabilities for working with MIDI files, I thus developed a simple API function[2]. This tool allowed students to upload MIDI files during a conversation and have them converted into a JSON format understandable by the LLM in the realtime of their conversation and from inside of the ChatGPT's user interface. The experiment opened a rich space for exploration of the tool's capabilities and limitations. What I have quickly discovered when working with the tool to prepare the class is that successfully implementing such a workflow in music experimentation requires a very careful strategy.

The agent must correctly interpret the context of its operations: it must understand who is issuing a command, what kind of output is expected, and how the tool it controls functions within that specific environment. One thing to remember is that plain LLM, without further fine-tuning to better align with a specific expertise, does not have any knowledge on how to successfully write and analyse music. However, it does know how to successfully structure the MIDI file format and how to "reason" with natural language. This gives us the opportunity to teach the model any music theory that we'd like it to know, before applying this in practice. The agent must

[2] The code for the API function is available on my GitHub (@sunsetsobserver) under the name MIDI-JSON-API.

embody a human-like understanding of the tasks it performs. For instance, to analyse a musical score, it must first grasp the theoretical foundations—such as harmonic analysis, form, or voice-leading—before attempting to classify or critique musical structures. Similarly, for generative composition tasks, the agent must internalise stylistic, structural, and aesthetic constraints, rather than applying surface-level transformations without deeper comprehension. Ensuring such context-sensitive, musically informed behaviour remains a key challenge when deploying agents in professional creative workflows. The agent, asked briefly to compose something interesting and to return a MIDI file, will almost certainly fail to do the job as it does not know who is asking, why is such question relevant, and what would be the context for an appropriate evaluation of such a task. However, two strategies tested during my course, provided more than extraordinary results in terms of musical outputs.

In one scenario, the student would first explain to the LLM their own musical background and artistic expectations when working with a musical material. Then, they would input a MIDI file of a musical fragment that they like, to show the model what is considered a "good" generation. To enable deeper understanding of that fragment, the model would be further queried to take its time and analyse the file according to its various parameters, such as to perform an analysis of the musical fragment's melodic, harmonic, and rhythmical content. After such an extended analysis, the model would be asked to try and conceptualise what a stylistically-coherent continuation of that musical fragment could contain—what would be the changes in the musical material and how could this be implemented. In that situation the model would come up with a few possible ideas and confront them with the student. The student would then ask the model to choose their preferred strategy and ask the model to generate the file. The model would write the extended composition and use the API tool to convert that text into a downloadable MIDI file, which the student can then open in their preferred DAW and listen to for evaluation. While the evaluation of the results of such a creative workflow will always be inevitably subjective, the high-level controllability with which the user can teach their personal assistant to become gradually more helpful, enables a way of achieving highly satisfactory musical results, comparable and actually surpassing the capabilities of pre-trained musical transformer models such as Allegro MIDI Transformer for experimental creative applications.

We quickly discovered that the presented workflow, while successful, is quite labour intensive. At least at first, to prepare the context for further creative conversations. The general rule is that the more creative context is provided to the model, the more useful it becomes for assisting the creative collaboration with the specific user. However, it is possible, at least partly, to strategise a quicker contextual jump enabling the right creative mode of the foundational model. For instance, a specific music theory can be evoked to immediately adjust the model by asking it to impersonate an expert in a field associated with a specific music theory. To give one example, a user can ask the model to act like an expert in transformational theory by David Lewin, a strictly mathematical method of music analysis. The model can then help the user analyse their musical fragment within the framework of that methodology and it can further use that knowledge to generate a coherent musical continuation. While there is of course a risk that the model does not know a specific theory, or only simulates to know a certain theory well, for an expert in that theory itself it will be easy to recognise the hallucinations of the model and to intervene accordingly. At the end of the day, interacting with an LLM is like interacting with a random person (Mollick, 2024, Introduction: Three Sleepless Nights). At the beginning we don't know what the specific beliefs, knowledge and skills that person has, but one thing that we can rely on is that given enough time, we can learn each other better and learn to work with each other on gradually more aligned common goals.

AI Agents Controlling Music Environments

In the second class of my course at the Iceland University of the Arts, some of the students expressed their interest in using the reasoning capabilities of LLMs from within SuperCollider—a platform for audio synthesis and algorithmic composition in which users write code to design, control, and manipulate sound in real time, enabling the creation of dynamic musical structures, custom instruments, and complex generative systems. In their creative practice, the students used SuperCollider primarily in live-coding performances, where they write and execute code in real time to produce live music, with this code often projected visibly to the audience as part of the performance itself. For this reason, switching back and forth between a browser-based ChatGPT interface and their live-coding environment was impractical. Furthermore, they expressed a desire for a more integrated interaction, where the

outputs of the reasoning model could directly influence their SuperCollider code without interrupting the performance flow. To address this need, I developed a simple SuperCollider extension that enables direct API communication with OpenAI's ChatGPT models from within the SuperCollider environment. This extension allowed students to send prompts and receive responses programmatically, making it possible to integrate large language model outputs dynamically into their live-coded performances without leaving the SuperCollider interface. For their final concert, students used the extension to let ChatGPT dynamically decide on the semantic meanings of their natural language inputs in the console visible on the screen to the audience, to dynamically control the sound parameters of their live-music performances.

Because LLMs and agents can be accessed through APIs, it becomes possible to integrate them into other software environments commonly used in experimental and professional music production. Similarly, in a MAX/MSP environment, an agent could be employed to generate control sequences, manipulate audio streams, or dynamically reconfigure patches based on high-level commands provided by the user. This integration bridges generative reasoning with real-time computational environments, enabling fluid, on-demand augmentation of creative processes. Looking further into the future, these developments point toward a generalised paradigm of operator-like interfaces, where reasoning-enabled agents could coordinate and control a wide variety of software tools across an operating system. Rather than scripting or programming each individual task manually, users could delegate complex, multi-step operations to agents that understand high-level goals, learn tool-specific protocols, and autonomously compose workflows across diverse applications. In such a vision, the computer becomes a dynamically reconfigurable creative partner, responding to abstract artistic intentions through concrete, tool-mediated actions executed with the support of large-scale reasoning models.

Because the environmental and economic costs of AI tools have become a growing concern, especially in educational contexts, this workshop was also designed to foreground such considerations. We discussed with the students how large-scale language models can entail significant energy consumption and corporate dependencies, and we deliberately investigated how to work within these constraints in a responsible way. To this end, we used the *gpt-4o-mini* model, selected for its efficiency and minimal resource footprint. Over more than a week of continuous

artistic exploration by the entire group, the total API usage amounted to less than one US dollar. This experience demonstrated that conceptually ambitious and musically innovative work with AI can be achieved at minimal computational and financial cost. It also modelled for students how to critically assess the resource implications of their own creative tools.

While these developments dramatically expand the creative possibilities for composers, it is important to maintain perspective on the current state of AI technology. As AI pioneer Yann LeCun emphasises, Large Language Models—despite their impressive achievements—are not themselves a path toward human-level intelligence (Plumb, 2024). They still lack fundamental aspects of reasoning, planning, and understanding necessary for true autonomy, and no amount of scale alone will bridge this gap. Nevertheless, the methodologies explored in this chapter—building collaborative workflows with text-based AI agents, developing contextual strategies for reasoning, and creatively reconfiguring generative models within musical environments—remain artistically valid regardless of whether future AI systems achieve more advanced forms of intelligence. These strategies will remain applicable to any form of text-based collaboration with future intelligent systems, and, perhaps more humorously, with actual human collaborators as well.

Artists as Operators

The notion of the artist as operator of tools emerges with clarity within contemporary practices of machine learning-based creation, where authorship, materiality, and process are fundamentally reconfigured. Iannis Xenakis described the role of a composer working "with the aid of electronic computers" as "a sort of pilot: he presses the buttons, introduces coordinates, and supervises the controls of a cosmic vessel sailing in the space of sound, across sonic constellations and galaxies that he could formerly glimpse only as a distant dream. Now he can explore them at his ease, seated in an armchair" (Xenakis, 2001 [1963], 144). As Tilford (2023, 142) suggests, the appearance of a "Subject" in contemporary generative systems does not originate from a stable source of meaning but instead indexes "the emergence of an operator" working with randomness as its material. In this view, the operator is neither the originator nor the result but the local status of an ongoing process—a temporary configuration within an indeterminate field. Composer Holly Herndon—known for her work with emerging technologies and especially AI and blockchain— similarly articulates an accumulative logic of operator-practice when describing her trajectory across projects: with each new body of work, she does not rebel against previous methods but gathers accumulated tools and skills to extend and transform her practice (Clancy, 2022d, 46–47). This continual aggregation and modulation of technological means embodies an operator-function in the strongest sense: not the repetition of known procedures, but the dynamic accumulation of operational capabilities in response to new experimental terrains. Audry (2021, 35) further emphasises that artists working with machine learning must embrace the unpredictable: while algorithms are precise in their computational structure, their generative behaviour—particularly in machine learning contexts—emerges as unstable and contingent due to stochastic processes, data dependency, and nonlinear interactions that resist full control. Rather than imposing rigid control over outcomes, artistic practice in this context becomes an engagement with volatility, an ongoing negotiation with systems that resist full determination.

In this framework, Paulo de Assis' method of archaeological, genealogical, and creative investigation of musical material provides a powerful operational model for how composers-as-operators can structure their engagement with AI systems. De Assis' methodology unfolds in three steps. First, the archaeological phase collects all available material traces of a musical work or practice, categorising them into six types

of strata: substrata (background documents and conditions preceding the work), parastrata (materials generated during the creation process), epistrata (derivatives, interpretations, and analyses over time), metastrata (future reinterpretations and extensions), interstrata (hybrid materials resulting from contamination between strata), allostrata (external materials not directly related but capable of productive interaction). Second, the genealogical phase maps the relations between these strata, tracing connections, influences, and singularities—points of high energy where creative tensions cluster and transformation is possible. Third, in the problematisation phase, selected materials are arranged into experimental arrangements that actively challenge the established orders of the work, generating new actualisations rather than restoring or reifying an original (de Assis, 2018, 109–111).

This methodology can be applied with the iterative generative workflows described earlier in the chapter. When working with AI models, composers first generate large pools of material (analogous to assembling strata), then map and classify these materials according to their stylistic, structural, or experimental potentials (analogous to the genealogical phase). Finally, they intervene through editing, continuation, and recombination, producing new experimental assemblages that problematise the original materials, not by reconstructing their sources but by opening them to divergence, mutation, and further transformation. In both cases—whether operating on historical musical materials or AI-generated outputs—the operator engages in a practice that refuses stabilisation and definitive closure. Instead, the composer-operator curates fields of virtuality, navigates unstable topologies of musical possibilities, and constructs provisional assemblages that expose rather than resolve the tensions latent within the material. Through this continuous process of divergence, genealogy, and reconfiguration, the composer reclaims agency not by controlling the creative field, but by sustaining its metastable openness.

Latent spaces—those internal representations where AI models organise relations between concepts—offer novel ways of seeing, exposing unexpected connections and contextual proximities between ideas, sounds, or structures (Serpentine Arts Technologies, 2024, 65). Yet navigating these spaces demands an operational craft: the ability to formulate effective prompts, to guide the model's attention, and to shape emergent outputs with precision and intentionality. As Future Art Ecosystems 4 observes, prompt engineering is an evolving craft-based skill and a form of conceptual labour, akin to querying an unstable search space (Serpentine Arts Technologies, 2024,

93–94). The artist-operator must therefore remain vigilant, embodying what Mollick (2024) calls the "human in the loop"—the critical presence necessary to detect hallucinations, errors, and misalignments within AI outputs (Part 1, Chap. 3). Without this active vigilance, there is a real danger of "falling asleep at the wheel", allowing AI systems to silently erode human expertise, skill development, and the meaningful labour of creative work itself (Mollick, 2024, Part 2, Chap. 6). In a landscape where AI offers universally accessible shortcuts, compressing complex processes into instantaneous outputs, the value of deliberation, critical thinking, and aesthetic judgment becomes even more urgent (Mollick, 2024, Part 2, Chap. 5). While some creative tasks may be delegated entirely to AI—what Mollick terms “Automated Tasks”—others, especially those tied to personal values, deep meaning, or ethical stakes, must remain within the domain of what he calls "Just Me Tasks”, activities reserved for irreducibly human engagement (Mollick, 2024, Part 2, Chap. 6). The future of artistic agency, then, does not lie in rejecting AI technologies but in mastering how to live, work, and create with them critically, retaining and reconfiguring human expertise precisely where it matters most.

Seen through this lens, the artist as operator is redefined. The role of the operator expands into a field of new responsibilities: to navigate latent spaces with sensitivity, to craft interactions through prompt-based conceptual labour, sustain critical vigilance against automation’s numbing effects, and to deliberately reserve domains of irreducible human engagement. As Mollick notes, the introduction of calculators once provoked fears like today's anxieties about AI; yet rather than abolishing the need for arithmetic or critical thinking, it re-situated these skills within new cognitive ecologies (2024, Part 2, Chap. 7). Similarly, AI will not eliminate the necessity of learning, writing, or thinking critically—it demands their intensification. The imperative becomes clear: to teach and cultivate not only artistic techniques but AI literacy, critical judgment, and operational acuity (Mollick, 2024, Part 2, Chap. 7). Moreover, at a planetary scale, the potential benefits of AI—such as massively expanding access to education and knowledge in underserved regions—constitute a profound ethical mandate (Mollick, 2024, Part 2, Chap. 7). In this view, to refuse the development and intelligent use of AI would be to abdicate the responsibility to address urgent global needs. The task of the artist-operator, then, is neither nostalgic resistance nor blind acceleration, but the careful invention of new practices: to operate within a world reshaped by artificial intelligence without relinquishing the depth, intensity, and responsibility of human creation.