Although there exists significant opacity surrounding the collection, use, and storage of big data, questions of exactly how long digital traces persist online generate perhaps the widest array of answers, ranging from “digital information lasts forever—or five years, whichever comes first,”¹ to “it depends. . . .” to user data “will exist as close to forever as matters.”² Scholarly examinations of these data life cycles have taken interest in how big data challenges the notion of life as directly indebted to the corporeal body in what has been termed digital afterlives and haunted data.³ Magnetically inscribed on servers, data metrics are “haunted” insofar as data “bear traces of human, material, technical, symbolic, and imaginary histories.”⁴ Big data is collected, sorted, classified, and hierarchized by algorithmic analysis for the generation of predictive correlations that function toward the systematic digital monitoring of people or groups to regulate or govern behavior.⁵ Put another way, data traces—at once living and nonliving, human and technical—are essential elements in machinic compositions that channel affective intensities, enable agency, and produce subjects.

Data life cycles are further complicated by renewed interest in learning-based artificial intelligence (AI) vis-à-vis machine learning (ML) and a subset of ML called deep learning (DL). Machine learning algorithms use large datasets to learn to perform tasks through pattern recognition and probability. While machine learning AI requires human supervision and prestructured datasets, more recent forms of deep learning AI take the
machine learning paradigm further in that layers of algorithms compose artificial neural networks modeled after the neurons of the human brain that are capable of unsupervised learning from unstructured datasets. Structured, human-aided machine learning remains more common than deep learning; yet both ML and DL have become increasingly embedded in everyday sociotechnical interactions. For example, learning-based AI is used in social media and streaming media platforms, automated driving, medical research, industrial automation, and home assistant devices. Additionally, so-called creativity and invention machines function not to associate patterns, but instead to generate new concepts, designs, and strategies resulting “from a noise-driven brainstorming session between at least two neural assemblies.” Regardless of specific task, machine and deep learning demonstrate how data collected from individuals is made part of machinic apparatuses in what can be considered “becoming AI.”

If, in fact, our data is ourselves, capable of existing as a productive component of artificial intelligence even after the death of the corporeal human body, then questions of data life cycles are inexorably linked to the process of subjectivation. Despite differences in theories of how the subject is formed and disciplined, subjectivation works toward the engineering of “docile subjects as functional components of the sociotechnical megamachines of war, bureaucracy, and/or capital.” While the subject has been often conceptualized as a bounded human organism, Gilles Deleuze and Félix Guattari’s three syntheses of subjectivation offer an explanation of how the subject is formed through the connection, inscription, and arrangement of partial objects that allow for the momentary emergence of the subject. Deleuze and Guattari situate the subject as formed by machinic capture and channeling of desire within a social field of bodies, organizations, and institutions; however, Stephen Wiley and Jessica Elam amend Deleuze and Guattari’s emphasis on the social to the sociotechnical in accounting for the critical role of technical machines and media to the formation of the subject. Accordingly, Wiley and Elam position synthetic subjectivation “as a way to conceptualize subject formation as grounded in compositions of heterogeneous elements, and not in humans or hominid organisms.” Stated differently, the subject is not merely a delineated human organism, but a subject of sociotechnical compositions that engineer embodied potentials toward productive ends.

Taking inspiration from Wiley and Elam, I examine machine learning data life cycles as a form of synthetic subjectivation, in which the subject of the three
ontogenetic syntheses of desiring-production is a machinic composition of big data and neural networks. AI neural networks, then, demonstrate an emergent subject not directly reliant on a body of flesh and blood, but one that can instead operate as an entanglement of big data, machines, servers, datapoints, software, infrastructures, and algorithms. Such an examination requires attention to the multitude of continually expanding forms of learning AI; however, I focus on invention, generation, and Creativity Machines, including DABUS (device for the autonomous bootstrapping of unified sentence) as an experimental “artificial inventor.” These new forms of AI provide a glimpse of the possibility of subjectivity where the human is a mere figure in the sand—where the hominid component is absent entirely or present only through haunted data that transcends corporeal life and death. Taken together, AI challenges not only the binary distinction between life and death, but also the notion of the subject as necessitating a delimited hominid component.

**Artificial Intelligence**

In popular and science fiction imagination, AI invokes images ranging from subservient humanoids to out-of-control machines set upon the destruction of the human race. The contemporary reality of AI is far more banal, embedded, and infrastructural; contrary to science-fiction, AI often exists not as a unified machine but instead as a composition of lines of code, magnetic and silicone inscriptions, flows of digital packets, warehouses of servers, entanglements of cables, and datasets. Not all forms of artificial intelligence are the same, however, as an important distinction exists between symbolic AI (also known as “Good Old-Fashioned AI or GOFAI) and “connectionist” machine learning. At the heart of this division is a difference in how “intelligence” is conceptualized. GOFAI attempts to reproduce human intelligence through procedural knowledge—intelligence is achieved by manipulating human-readable symbols according to underlying rules. By contrast, machine learning is a form of automation that utilizes algorithms and datasets to identify structural relations among datasets, detect and uncover patterns, as well as classify, transcribe, and make decisions. It should be noted that both GOFAI and machine learning AI continue to be used in different ways, and the two are, at times, not mutually exclusive. Nevertheless, advances in the past decade—specifically deep learning artificial neural networks—have led to increased dominance of the machine learning paradigm.
Machine learning requires prestructured data and human intervention, but recent developments in deep learning exist as a subset of ML capable of self-training through unstructured datasets. Deep learning neural networks are compositions of algorithmic nodes designed to resemble human brain neurons, arranged as the input layer, hidden layer(s), and output layer. These neural networks are trained in a “bottom-up” statistical approach by large, aggregated user-generated datasets for the purposes of object recognition, processing, identification, and automation. Hence, deep learning is reliant on sociotechnical shifts—specifically, the paradigm of big data collection—that captures human participation in an ever-increasing variety of contexts. As such, deep learning “is less about replacing human cognitive labor by an intelligent machine but about embedding and harvesting human cognition in computing networks through new forms of labor and machinized power relations.” Or, stated differently, deep learning AI is at once human and nonhuman, composed of neural networks reliant on data collection, storage, and processing.
Creativity Machines and invention AI also utilize neural networks, but to different ends. Aforementioned forms of learning AI mostly complete tasks of pattern recognition and probabilistic output. By contrast, Creativity Machines operate through a relationship between two neural networks—the imagitron and perceptron. The imagitron or imagination engine is a generative neural network that is fed input datasets, which are then “perturbed by any form of random, semi-random, or systematic disturbances so as to drive the generation of potential ideas.”\textsuperscript{17} The perceptron or discriminative neural network operates as a critic by calculating the merit of potential ideas stimulated by the imagitron and in turn injects increasing levels of synaptic noise back into the imagitron.\textsuperscript{18} As such, Creativity Machines utilize random or systematic disturbances in the production of patterns representing potential ideas, plans, or strategies.\textsuperscript{19} Put another way, Creativity Machines operate through disruption—noise in the system “detrains” a network, leading to abstracted and original variations to existing products and processes. Creativity Machines are capable of feats that include generating alternative Christmas carols, designing personal hygiene products, generating potential words, allocating resources, and interpreting content. Still, there remains pushback that neural networks underperform when compared to humans in tasks of creativity, including the identification and interpretation of symbols, modeling social action, and predicting the future.\textsuperscript{20} Viewed instead from the perspective of synthetic subjectivation,
however, these dichotomies between human versus AI become irrelevant, as both are intertwined in a process of ongoing becoming.

Generative and invention machines represent just one form of AI, whereas other machine and deep learning AI are now imbricated social media, streaming media, scientific discovery, medical research, industrial automation, self-driving cars, mapping crime, facial recognition, personal home assistants, customer experience, language identification and translation, aerospace and defense, and news aggregation. There are, of course, ethical issues that accompany these uses, including the fact that there are many aspects of artificial intelligence output that remain beyond human comprehension and explainability, even while this output is increasingly used to automate processes and decision-making. In addition to these moral quandaries, AI necessitates a reconsideration of the ontology of life and death, as well as the question of what the synthetic subject is.

**Digital Life Cycles and the Synthetic Subject**

Life and death are often positioned as dichotomous; nevertheless, complicating this distinction is a conceptualization of life and death as coextensive, as human and nonhuman bodies exist in a constant state of becoming. As explained by Rosi Braidotti, life itself is a relentlessly generative force that requires interrogation of human and nonhuman entanglement. The imbrication of users, technologies, data collection practices, infrastructures, platforms, and digital flows contributes to an erosion of the neatly demarcated human being, giving way to a more productive, nonbinary distinction “between same and other, between different categories of living beings, and ultimately between life and death.”

By extension, the origins and endpoints of corporeal life and death cannot account for the shifting forms of existence enabled by compositions of users and their data traces that are deeply entangled with media long after the human body’s corporeal demise.

The relationship between life, death, and media predates the era of computational big data. Beginning with the inception of the telegraph, media allowed for new forms of simultaneity, liveness, and copresence that has been framed as ghostly and haunted. In this way, “Media enable us to establish, maintain and develop relations with the dead without being present in the same space-time continuum.” The “reanimation” of the dead
is demonstrated further in social media, as profiles of deceased and memorialized users are projected to overtake living users by the year 2100.26

Social media is driven by participation and algorithmic interaction—sharing and socializing foster popularity and visibility, which in turn begets more popularity and visibility.27 For Grant Bollmer, social media produces the human as a posthuman subject, as “humans and technology become interchangeable through the privileging of connectivity and flow above all else.”28 Humanness, then, is not defined by a bounded body but instead through connection, as “nodal citizens” are expected to engage in and internalize proper conduct that includes connecting and maintaining flows while simultaneously managing the definitions and limits of their own life.29 Bollmer notes that in this way, “social media do not inherently rely on human subjects, but rather derive economic value from anything that can maintain connections and circulations—which, consequently, reduces the subject of social media to data itself.”30 Social media platforms enable novel relationships between humans and technical systems, operating as what N. Katherine Hayles terms a cognitive assemblage, which “emphasizes cognition as the common element among parts and as the functionality by which parts connect.”31 Cognitive assemblages perform roles related to cognition—incorporating knowledge, adapting, evolving, and transforming complex sociotechnical systems. Hayles specifically notes the amplification of technical autonomy in examples of automated drones, facial recognition, and self-driving vehicles as a redistribution of both cognition and agency with significant consequences for what it means to be human in developed societies.32 Social media platforms are an oft-studied site of connectivity and ensuing data collection, with significant attention paid to how data is monetized through targeted advertisement. Machine and deep learning AI, though, represent new strategies of making productive data life cycles not only by social media platforms, but also by businesses, commercial interests, administration, government agencies, scientific research, healthcare, stock exchanges, and other institutions. Unitig these uses of AI is the fact that data can exist independently of the human body, perform agency, and possess social value through circulation and flows. Accordingly, binary notions of life and death and conceptualizations of the delimited human subject are challenged by sociotechnical compositions in which an individual’s data can connect and be made productive even without direct reliance on a living homininid component.
The Three Syntheses and AI

ML and DL processes are embedded in a growing number of quotidian productive interactions—drawing from datasets to suggest, sort, remove, identify, create, aggregate, and invent. For instance, Netflix’s machine learning algorithms use thousands of screenshots to suggest personalized thumbnails to attract users to specific content.⁴³ Deep learning AI is used to identify and moderate social media content, make predictions about how to allocate resources, create “deepfake” videos of fictitious events, develop new drugs and cures, and drive cars. Uniting the myriad forms of AI is the need for datasets, which requires the ontogenetic act of making and breaking connections between partial objects—the first synthesis of synthetic subjectivity. As previously noted, Wiley and Elam eschew the quotidian signifying distinction between human and technology to consider the composite body of natural-sociotechnical arrangements. AI does not recognize the user as a whole human, but instead functions through engagement with data points, creating hybrid human-machine networks.⁴⁴ Yet generating this data first requires connections that exist prior to organization and molar hierarchization—eyes, hands, remote controls, modems, routers, and servers create contingent connections of composite bodies that facilitate the creation of datasets needed for operations of machine and deep learning. Through the use of mobile and fixed technologies that connect to a variety of platforms, individuals click, rate, like, watch, search, ask, and share. They move, drive, and check-into locations in conjunction with GPS signals. They are sites of acknowledged and unacknowledged biometric data extraction. And through often unrecognized techniques, these data flows are made productive components of artificial intelligence.

As previously noted, creativity and invention AI neural networks are predicated on the connection between a generative and discriminative neural network. Critical to this process is a dataset. For example, German AI artist Mario Klingemann utilized a generative adversarial network (GAN) AI as a 2018 art installation called Memories of Passerby I. The exhibit, described as “an autonomous machine that uses a system of neural networks to generate a never-ending, never-repeating stream of artistic portraits of nonexisting people” was trained via an enormous collection of seventeenth, eighteenth, and nineteenth-century portraits.⁴⁵ Similar exhibits, along with the sale of AI-generated art, are not without controversy, as some have questioned the originality of both the art itself as well as the code used in
some AI artwork generation neural networks. These ethical considerations notwithstanding, these examples highlight the importance of the connective synthesis to AI. Neural networks require datasets—in this particular case, collections of centuries-old artworks. To wit, complaints that these AI art generation exhibitions lack humanity ignores the fact that centuries’ worth of digitized art—these works themselves sociotechnical compositions of human and nonhuman becoming—are essential to these neural networks. Whether AI generates original artwork, invents products, or recognizes faces, the connection with datasets is exigent to AI functionality.

Deep learning is predicated on the relationship between datasets and the layers of a neural network. While specific network architectures vary, critical to deep learning is an iterative review of data so that the artificial neural network can learn to identify trends and features from datasets. This process is exemplary of the second synthesis of recording and enregisterment, in which machinic bodies are inscribed and arranged. Neural network nodes are connected to one another and communicate in that the output signal of one node becomes the input signal of another. In the input layer, data can often take the form of human-readable text or recognizable images, but neural networks must reinscribe and arrange data in machine-readable form—deep learning operates by breaking down the features of text or images and passing these input signals through layers of nodes. These inputs are transmitted between the hidden layers where nodes apply weights to connections, represented by real numbers. As the network trains, weights adjust to bring the output closer to the correct value by deciding how much influence the input of neural networks has on the output. Here, algorithms, long defined as a set of instructions, are better understood as performing entities, or “actualities that select, evaluate, transform, and produce data.”

Applied specifically to AI, machine and deep learning inscribe and enregister the data as identifiable to neural networks, which allows for recognition, sorting, aggregation, suggestion, and identification. Overall, this process demonstrates Wiley and Elam’s contention that “digital media technologies can both open up and close off the disjunctive process, either expanding what a composite body can do or limiting its potential to predefined pathways.” Put another way, data is connected to and inscribed by neural networks in the process of becoming a composite body of the synthetic subject.

As a specific example, the AI inventor DABUS departs from the neural network architecture of other Creativity Machines by using a swarm of thousands of disconnected neural nets containing interrelated memories that
constantly combine and detach into longer chains of complex concepts.\textsuperscript{42} A camera and computer are employed to detect and isolate promising neural chains that are fed into a “thalamobot” that then injects both noise and reinforcement back into the system.\textsuperscript{43} Another key departure for DABUS is that the AI adapts through “episodic learning rather than careening through databases and racking up lots of training epochs. In other words, a single exposure to a data pattern is often sufficient to make a lasting impression on this new neural network paradigm that links conceptual spaces rather than neurons.”\textsuperscript{44} To this end, creativity AI works through synaptic perturbation—creation is the result of connection and disruption. This generative process of DABUS is far from disorganized, as the AI undergoes what Thaler calls a “subjective feel” that forms “as chains that incorporate a succession of associated memories, so-called affective responses, or associative gestalts.”\textsuperscript{45} Explained by Wiley and Elam, “In the first synthesis, connections are selected and made; in the second synthesis of recording and disjunction, these connections are inscribed into a grid of possibilities that are registered in diverse ways across the arrangement.”\textsuperscript{46} DABUS’ creative process is one of connection and inscription, in which the goal is the actualization of generative possibility. The thalamobot both opens and closes the disjunctive process by selecting, disrupting, and mutating data previously entered into the creative apparatus.

The third synthesis—the conjunctive synthesis in which the subject is recognized—leaves the most questions for learning-based AI and data life cycles. Wiley and Elam specifically invoke the challenges of AI when asking “How do we understand the moment of recognition or consummation (‘So that’s what it was!’) when the conjunctive synthesis involves no human components at all?”\textsuperscript{47} Due to the emergent compositions of deep learning, the challenge of recognizing the synthetic subject is likely to only become more difficult. For example, the US Patent Office, UK Intellectual Property Office, and European Patent Office all ruled that despite designing two products, DABUS could not lawfully be listed as the inventor on patent applications.\textsuperscript{48} Legal recognition is, admittedly, only one form of inscribing and recognizing a subject through signifying semiotics, but this situation speaks to the complexity of the subject as sociotechnical apparatus as opposed to a neatly bounded human body. Equally important is how AI is framed. For instance, in an anthropomorphic description of DABUS, inventor Stephen Thaler has likened the AI’s creative process to characteristics of mental illness, hallucinations, attention deficit, mania, reduced cognitive flow, and depression.\textsuperscript{49} The figures of the schizophrenic and the neurotic are important
to Deleuze and Guattari—the deterritorialized and decoded flows of desire of
the schizophrenic are contrasted with the Oedipalised repression and feeling
of lack of the neurotic. While the schizophrenic offers potential as
revolutionary agent, able to transgress social codes, the neurotic is made to
decline its own repression. The synthetic subject of creation and invention AI
may seem removed from a human component that primarily exists via data
echoes, but the process of continual generation certainly demonstrates a
neurotic subject of desiring production (and algorithmic repression) within
the megamachines of global capital flows.

Must a subject have a human component capable of recognition? For Wiley
and Elam, the answer may lie in attention to the complexity of “organized
inorganic matter,” along with the tendency of that technical assemblage to
select technologies—and humans—that further ‘its own’ development.”51 For
instance, autonomous self-driving vehicles powered by deep learning AI
represent a rapidly emergent future for automobility. Self-driving deep
learning AI is trained on data collected not only by human drivers, but also
from self-driving AI, as well as so-called synthetic datasets (algorithmically
generated datasets created to mimic data captured by measured events).
These synthetic datasets come in various forms, but one of the most common
ways to generate synthetic data is to use a generative adversarial network—
the same neural network model used in many of the aforementioned forms of
AI creation and invention.52 Perhaps, then, it should come as no surprise that
the head of machine learning and perception at the self-driving technology
development company Waymo emphasized that the goal of development is
not merely autonomous vehicles, but “AI machine learning that’s creating
other AI models that actually solve the problem you’re trying to solve.”53 Of
course, it should be noted that true separation of human components from AI
is likely further away than tech companies might publicly promote. An
examination of the micro-work within the AI industry found steady
structural demands for humans to prepare AI data, create synthetic datasets,
verify output, and even impersonate AI tasks.54 As such, “technological
progress has not eliminated the need for micro-tasking, but transformed it,
integrating humans and computers more tightly.”55 These possibilities reveal
the developing complexity of the synthetic subject—shifting compositions at
times reliant on and removed from the components of living hominid bodies
and even human-generated data.
The Future of Synthetic Subjects

The era of big data is one in which connectivity begets possibilities for the collection, storage, and processing of data points. Demonstrated by machine and deep learning AI, datasets generated from a multitude of everyday practices are made a productive part of sociotechnical apparatuses that then act back on these interactions by suggesting, sorting, aggregating, identifying, creating, and inventing. Data flows—haunted by human, material, and social histories—are recorded, unregistered, and made readable by neural networks that learn to more effectively perform tasks, in turn expanding and delimiting the potential of the composite body. To this end, AI is a docile and economically productive component of sociotechnical megamachines—even if this subject is unbounded from the traditional notions of the bounded, living human body. Wiley and Elam’s synthetic subjectivation allows productive analysis of AI as subject, formed through the investment of desiring-production in sociotechnical machinic compositions of big data and neural networks. Taken together, emergent compositions of artificial intelligence and big data necessitate reconsideration of some of the critical questions of data life cycles including “how long will data exist online?” and “what is this data used for?” In accounting for the emergence of machine and deep learning AI, perhaps a far more relevant question is instead “what is data becoming?”

Notes


Ibid.


Ibid., 1870.


Ibid., 449.

Ibid., 451.


Ibid., 209.


Ibid., 7; 119.

Ibid., 133.


Ibid., 34–35.


Mühlhoff, “Human-Aided Artificial Intelligence.”
37 Deleuze and Guattari, Anti-Oedipus, 75-78.
38 Arkoudas and Bringsjord, “Philosophical Foundations,” 52.
39 Ibid.
44 Thaler, “What is DABUS?”
45 Ibid.
47 Ibid., 222.
50 Deleuze and Guattari, Anti-Oedipus.
55 Ibid., 6.

Dr. Justin Grandinetti is an assistant professor in the Department of Communication at the University of North Carolina Charlotte. His research interests include mobile and streaming media, media archaeology, and spatial materialism. Justin’s work has appeared in Information,