

From rules to examples: Machine learning's type of authority

Alexander Campolo^{1,*}  and Katia Schwerzmann^{2,*} 

Big Data & Society
 July–December: 1–13
 © The Author(s) 2023
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517231188725
journals.sagepub.com/home/bds



Abstract

This paper analyzes the effects of a perceived transition from a rule-based computer programming paradigm to an example-based paradigm associated with machine learning. While both paradigms coexist in practice, we critically discuss the distinctive epistemological and ethical implications of machine learning's "exemplary" type of authority. To capture its logic, we compare it to computer programming rules that date to the middle of the 20th century, showing how rules and examples have regulated human conduct in significantly different ways. In contrast to the highly constructed, explicit, and prescriptive form of authority imposed by programming rules, machine learning models are trained using data that has been made into examples. These examples elicit norms in an implicit, emergent manner to make prediction and classification possible. We analyze three ways that examples are produced in machine learning: labeling, feature engineering, and scaling. We use the phrase "artificial naturalism" to characterize the tensions of this type of authority, in which examples sit ambiguously between data and norm.

Keywords

Machine learning, artificial intelligence, rules, examples, data, authority, naturalism

Machine learning researchers often narrate the history of artificial intelligence in terms of a transition from programming by rules to training by examples. In his book *Deep Learning with Python*, Chollet (2021: 3) explains: "...for a fairly long time, most experts believed that human-level artificial intelligence could be achieved by having programmers handcraft a sufficiently large set of explicit rules for manipulating knowledge stored in explicit databases." However, this approach ran into limitations when it came to "solving more complex, fuzzy problems, such as image classification, speech recognition, or natural language translation" (Chollet, 2021: 3). These problems needed to be addressed in a new way: "A machine-learning system is *trained* rather than explicitly programmed. It's presented with many examples relevant to a task, and it finds statistical structure in these examples..." (Chollet, 2021: 4).

Chollet's description gives the impression of a symmetrical "Copernican" reversal between the two paradigms. In the classical programming paradigm, rules are applied to data to produce outputs. In machine learning, the order is reversed, with examples used to generate representations applicable to new data—to generalize. This paper takes the machine learning community's trope opposing programming rules and examples as a starting point to draw out its deeper sociological and ethical ramifications in

terms of authority and the reasons people accept it—legitimacy. We argue that not only has the relative position of rules and examples changed, but these relations themselves have been substantially transformed. The phrase "type of authority" is meant to evoke the sense developed by Weber (1978: 212; 1980: 124) (*Typ der Herrschaft*): a specific way of producing obedience. In other words, authority designates the "conduct of conducts" (Foucault, 2008: 186).

Although some may object that words like "rules," "examples," "training," and "data" point to neutral technical realities, our position is that these are inescapably ethical concepts whose semantic shifts register and make possible different ways of exercising authority. In this sense, machine learning's specific way of classifying our world and regulating our conduct according to its

¹Department of Geography, Durham University, Durham, United Kingdom

²Institut für Medienwissenschaft, Ruhr-Universität Bochum, Bochum, Germany

*Equally contributed.

Corresponding author:

Alexander Campolo, Department of Geography, Durham University, Lower Mountjoy, South Road, Durham DH1 3LE, United Kingdom.
 Email: Alexander.Campolo@durham.ac.uk

predictions takes on ethical and political significance. This type of authority contributes to what Amoore (2022: 2–3) has termed “a machine learning political order,” which “does not merely change the political technologies for governing state and society, but is *itself* a reordering of that politics, of what the political can be.” What we call “exemplification” in machine learning produces a new, naturalistic kind of normativity that can transform the way society conceptualizes, exercises, and legitimizes authority.

Our approach to understanding these changes is comparative, drawing out conceptual distinctions between an authority based on programming rules and what we call an exemplary type of authority in machine learning. In the first section, we analyze the role of rules within a larger rational type of authority and then focus on the moment in the middle of the 20th century when digital computing intensified this calculative form of rationality. Beyond mere historicization, this genealogy situates machine learning within a constellation of concepts—rules, rationalization, and calculation—making machine learning’s own emerging type of authority intelligible in the broader sweep of social and political theory. Against the idea that machine learning intensifies an inherited modernist, bureaucratic, rational type of authority (Farrell and Fourcade, 2023), we identify distinctive features of machine learning’s exemplary type of authority. Unlike the highly explicit and rigid formalisms of computational rules, machine learning’s norms emerge implicitly when data is aggregated and processed at great scale by models. To be sure, in practice there is considerable mixing between rule- and example-based paradigms, and our argument is not that there is a total historical discontinuity between rules and examples. Beyond the machine learning community’s self-narration, we draw out and theorize formal features of an emerging, exemplary type of authority as a Weberian “ideal type” (*idealtypisch*, Weber, 1978: 21; 1980: 4), an analytical construct, which, although not encountered in pure form, can nonetheless serve as a conceptual tool for critically grasping both continuities and new dynamics.

The second part of the paper turns to the practices and knowledge involved in producing this exemplary type of authority. We analyze ways by which the aggregation and processing of data create examples that serve as a training resource in machine learning systems. These examples interact with model architectures to produce the representations that make classification and prediction possible. By now, it is widely accepted in studies of science that it is misleading to think of data according to its etymological sense as somehow given or naturally available, waiting to be processed. Instead, we say that data is constructed (Gitelman, 2013; Latour, 1999). In the area of machine learning, there are compelling accounts of the human labor required to produce seemingly autonomous or unsupervised learning systems (Bechmann and Bowker, 2019). Other studies have

shown how existing social norms and hierarchies seep into input data through engineering choices and proxies, mixing fact and value (Chun, 2021; O’Neil, 2016). Closer to our own interests, Grosman and Reigeluth (2019: 5) have shown that algorithms have their own “technical normativities,” and Jatón (2017, 2021) emphasizes circularities between the construction of algorithms and the referential ground truth datasets. Here, we situate these perspectives within a comparative account of machine learning’s exemplary type of authority. In addition to the constructedness of data and circularities of reference, we analyze how machine learning data is made exemplary to *elicit* norms, which, unlike in the rule-based programming paradigm, are not explicitly programmed or *prescribed*. Instead, the connection of inputs to desired outputs circumscribes a space in which norms can emerge. These norms are expressed not as prescriptive commands but rather in the form of model parameters that are learned during training (Grosman and Reigeluth, 2019: 6).

We analyze three concrete moments—labeling, feature engineering, and scaling—in which norms are elicited in machine learning. By situating these moments in a rough (and very recent) conceptual–historical sequence, we show how the machine learning community has conceptualized a movement from a more interventionist, “hand-crafted” creation of examples through practices like labeling and feature engineering to one in which exemplary representations emerge from the structure of the data itself through scaling. In practice, however, these tendencies mix intractably. We are less interested in characterizing this movement in terms of a distinction between apparently interventionist supervised forms of learning and unsupervised ones, but rather by the range of practices that make examples to engender norms. In processing these examples, machine learning models produce representations, which express regularities found in the data and seem to naturally emerge from it. These representations then become normative in a more traditional sense when they influence our behavior. The “ought” of examples turns into the “is” of the representations, and, in a further twist, these representations become normative (“ought”) again when the algorithmic outputs are used to make and legitimate predictions and classifications that regulate our conduct.¹ We designate this process as an “artificial naturalism.”²

This apparently contradictory phrase is meant to evoke associations with artificial intelligence and to express the ambivalent position of machine learning’s normativity—oscillating between constructedness and givenness. Unlike an earlier 19th-century naturalism associated with writers like Emile Zola (1893),³ machine learning’s artificial naturalism entails three aspects: first, it presupposes a world determined by deep statistical structures, the source of norms, that are accessible not to human perception but only through the representations produced by models. Second, these representations should increasingly be

learned from the data itself rather than from human-specified interventions. In the abstract, this valuation of non-intervention is not new, as historians of objectivity have shown (Daston and Galison, 2010). And in practice, machine learning is characterized by a messy tension between a desire to let data speak for itself (or through models) and the engineering practices that permit it to do so, albeit ones that look very different from the explicit specification of rules. The form of truth that emerges from training by examples is characterized not by a perfect correspondence to or reflection of training data (this would be mere “memorization”) but rather by an inductive ability to generalize, to discover regularities that function as norms that make it possible to accurately classify new data in new contexts. Third, machine learning introduces a new sense of scale into the logic of examples by aggregating them at massive scales. Instead of an example as the singular, concrete expression of an essence or type, in machine learning, examples form a multiplicity in which patterns and differences form a more accurate composite representation. The horizon of machine learning is to make the map of its models and data converge asymptotically with the territory of the phenomenon through scaling. Machine learning’s naturalism has political implications since its norms appear not as human-made commands expressed explicitly as rules but as emerging from reality through the mediation of a technological assemblage of models and examples. In order to understand the significance of these changes, we turn first to the authority of rules to develop a point of comparison.

A genealogy of computer programming’s rule-based type of authority

The widespread claim of a transition from a rule-based programming paradigm to an example-based one in machine learning begs a number of questions. What, precisely, do rules and examples refer to in these technical contexts? How did this distinction arise in the first place? Scholarship on rules has considered these questions. In a major study, Lorraine Daston argues that we live in the shadow of a historically specific algorithmic form of rules, which aligns with what we call the rule-based programming paradigm. According to Daston (2022: 7), the recent hegemony of such rules has caused us to lose sight of a much older and more flexible sense of rules, which held together other forms, such as “models” and “paradigms.” These older rules often worked hand in hand with examples as illustrations. Politically, this historicization works to recover these more diverse meanings in order to create new spaces for human discretion and judgment in the application of rules. Rather than attempting to recover the possibilities latent in the past by opposing inhuman rules to the human values of discretion and

judgment, we begin with an actually existing challenge to the hegemony of Daston’s algorithmic sense of rules expressed by the machine learning community. To account for the opening of this conceptual gap between rules and examples, we first need to understand the emergence of the specific type of rules that can be opposed to an equally particular type of examples.

Rules have long regulated human conduct. Most relevant for our purpose is the relationship between, on the one hand, a methodological (or even technical) sense implying functional instructions to produce knowledge and, on the other hand, the normative principles guiding the conduct of people and things. This relationship was elaborated throughout modern philosophy, where rules became identified with universal reason (Erickson et al., 2013: 37). With René Descartes, philosophy’s primary object became the method itself, meaning a series of steps guided by rules that the mind must follow in order to secure its way toward knowledge. These rules originate in the subject’s reason or what Descartes (2006: 5) calls “good sense,” which is the universally shared nature of the human mind. Immanuel Kant elevated rules to the level of the transcendental: Rules are *both* the conditions of possibility for knowledge and a guide for moral action. They originate and are grounded in the subject itself. In Kant’s striking formulation, the universality of rules becomes the “categorical” source of their legitimacy. Only when subjects obey a categorical imperative regardless of particular circumstances, interests, and the pleasure or displeasure resulting from it are they free, in other words, non-conditioned by contingency (Kant, 2015).

This Enlightenment celebration of rules as the source of freedom seems quite remote from the computational programming paradigm whose rules seem to reduce us to machines. By the beginning of the 20th century, Weber sensed that—*contra* Kant—the legitimacy of rules could not be so easily grounded in the structures of subjectivity itself. Instead, rules gain their legitimacy from external cultural “orders,” like the legal order, whose authority has to be enforced and legitimized in the eyes of the governed. This led Weber to place rules at the heart of a rational type of authority, distinct from authority based on tradition or charisma. Weber identified five more specific characteristics of the rational type of authority, which fed into the computational understanding and legitimation of rules.

First, rational systems relocate the source of authority from persons to an *impersonal* body of rules. Somewhat circularly, the governed owe their obedience not to a ruler but to a set of rules that guarantee their own authority, a “rule of law” (Weber, 1978: 217). Second, these bodies of rules have a *systematic* character. Instead of gradually accumulating from empirical precedent or freely mixing rules and examples, rational systems are established intentionally through a process of abstraction. Particulars are removed so that a small number of self-consistent, axiomatic rules

can cover a wide variety of situations. As opposed to the implicit, habitual rules associated with the traditional type of authority—or, as we will see, the exemplary type of authority—rational rules must also be written down, *encoded*, and made publicly available (Weber, 1978: 657). This encoding must be rendered in *explicit* terms, as formal and unambiguous as possible. Finally, these rules make life *calculable*.⁴

Rules that are impersonal, systematic, explicitly encoded, and calculative produce a distinctive form of authority. Whereas for Kant and Descartes rules are authoritative due to the intrinsic, universal claims of reason itself, the rational type of authority is legitimated extrinsically by its calculative results: it permits the “highest degree of efficiency ... superior to any other form in its precision, its stability, in the stringency of its discipline, and in its reliability” (Weber, 1978: 223). Metaphorical connections between the authority of rules and calculating machines crystallized at this time: such rules “operate like a technically rational machine...” (Weber, 1978: 811). This instrumental, machinic sense of rules deepened with the rise of digital computing in the 20th century, giving rise to the sense of rules that is now opposed to that of examples.

Foundational work in computer science in the middle of the 20th century linked the technical characteristics of digital programming with questions of rules and authority. For instance, although Alan Turing’s classic paper “Computing machinery and intelligence” is best known for its tantalizing question, “can machines think?,” and its thought experiment, “the imitation game,” a new computational understanding of rules conditions his formulation of these problems. Turing opens with a then-common comparison between digital and human computers (very often women in practice) who execute rules for large-scale calculations controlled by a predetermined division of labor (Light, 1999). He imagines these human computers as *totally* subject to the authority of rules—a moment where technical forms of control and the authority to guide human conduct coincide.⁵ “The human computer,” writes Turing (1950: 436), “is supposed to be following fixed rules; he has *no authority to deviate from them in any detail*. We may suppose that these rules are supplied in a book, which is altered whenever he is put on to a new job” [emphasis ours]. In Turing’s understanding of computation, the activity of rule-making is entirely distinct from the activity of rule-executing. The art of creating these rules is named “programming” and once again compared to the control of human behavior: “If one wants to make a machine mimic the behavior of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table” (Turing, 1950: 437). Rather than aspiring to the freedom of the Enlightenment subject, rule-making becomes purely instrumental, oriented to controlling the specific task at hand.

Rules for calculation have long consisted of dividing operations into elementary steps (Daston, 2022: 82). As the historians Erickson et al. (2013: 3) observe, with digital computing, rule-making became intensely analytic: “complex tasks and episodes were analyzed into simple sequential steps ... analysis took precedence over synthesis.” By extending this sense of analysis, programming rules expanded the scope of what can be governed by rules: not just numerical calculations but any “behavior” that might be encoded in a formal language that is totally unambiguous to a computer. As opposed to Weber’s rational rules, condensed from a systematic process of logical abstraction, these programming rules multiply, even fracture into different applications for each objective. While a single program might constitute a coherent whole, programs put together do not necessarily form any sort of rational, systematic corpus of rules. Coherence becomes purely formal—guaranteed by the programming language and its syntactic rules.

These characteristics of computational rules point to a horizon not of application or enforcement, which requires adapting an abstract rule to a concrete situation, but rather of total control over a movement between narrowly specified inputs and outputs. The theoretical discreteness of digital computers, the discontinuous movement between “states sufficiently distinct for the possibility of confusion between them to be ignored,” assures that an input signal leads invariably to a determinate output state, evoking the total determinism of Pierre-Simon Laplace (Turing, 1950: 439–440). Computers are capable of an “enormously large” number of discrete states, and this quantitative explosion in the number of rules leads to a qualitative one: a new form of universality in which “digital computers ... can mimic any discrete state machine” (Turing, 1950: 441). Universality no longer designates the ability of a small set of principles to cover all possible empirical situations but rather to have a space of discrete states so large that an enormous number of rules can be specified, a space large enough to emulate any possible behavior with utter predictability, leaving no space, as Daston laments, for interpretation or judgment. The potential for a huge number of rules means that any ambiguity becomes potentially fatal: an error at any step throws off the entire calculation. Therefore, special programming languages are required to express computational rules in a totally explicit and unambiguous way. Natural languages harbor far too much ambiguity.

This totalizing authority of computational rules might seem to constitute a logical endpoint, a sort of apotheosis of rules. Instead, the machine learning community began to identify their limitations. The extremely high requirements for explicit, unambiguous specification became onerous, making both symbolic forms of artificial intelligence and narrower “expert systems” brittle when confronted with more open-ended problems (Mitchell,

2021).⁶ To address such problems, it would no longer be possible to rely on “human operators to formally specify all the knowledge that the computer needs,” as another machine learning textbook puts it. Instead, “the solution is to allow computers to learn from experience” (Goodfellow et al., 2016: 1). Against this backdrop of computational rules, we can now illuminate what these types of statements entail; how, precisely, they differ from machine learning’s examples; and what type of authority they make possible.

A genealogy of machine learning’s exemplary type of authority

The criterion of explicitness at the heart of programming rules provides an entryway into this comparison. As rules become more numerous and specified, less seems to escape them. Programming rules tell us and the machine ever more precisely how to behave. This prescriptiveness is justified in familiar terms, echoing the instrumental legitimation of Weber’s rational type of authority: complex calculative processes can be executed with great speed and precision and at low costs. In contrast, recall Chollet’s (2021: 3) characterization: “A machine-learning system is *trained rather than explicitly programmed*” [emphasis ours]. How might comparatively implicit, “silent” examples guide our conduct through “training”—rather than explicit programming?

The idea of examples, like that of rules, encompasses a huge range of ethical and epistemological positions. Examples can point to lives or ways of living that we can emulate, not through explicitly articulated principles but rather by imitating exceptional individuals. Kant (1997: 4:408–4:409) criticized examples on exactly these grounds: an ethics based on exemplarity—for instance, emulating the person of Jesus—is insufficient because it relies on implicit moral principles that allow us to recognize someone as exemplary in the first place. In Kant’s view, examples function only as embodiments of rules or principles that have been left implicit but remain fundamental. However, this weakness can also be a strength. We argue that an important source of examples’ authority is located precisely in their ability to implicitly and naturalistically “show” rather than explicitly “tell” in the abstract. Rather than demanding obedience to the rule—a construct, an abstraction—examples implicate us as living subjects as we attempt to follow them.

Examples are not exclusively premodern or religious; they sprout up wherever mediation is required between individuals and kinds. In modern science, examples have long served to manage the tension between specimens and species. While these ideas date to ancient Greek philosophy, Daston and Galison (2010: 58) argue that during the Enlightenment, the observational talent of identifying

the universal from particulars constituted a holistic form of genius in which “the eyes of both body and mind converged to discover a reality otherwise hidden to each alone.” These exemplary images “aspired to generality—a generality that transcended the species or even the genus to reflect a never seen but nonetheless real plant archetype” (Daston and Galison, 2010: 60). This ability to point to unseen structures or levels of reality also is characteristic of examples.⁷

The term “example” has a technical sense in machine learning. Like the older statistical notion of “observation” (Upton and Cook, 2014), examples in machine learning refer to “a collection of features that have been quantitatively measured from some object or event that we want the machine learning system to process” (Goodfellow et al., 2016: 96). It is important to note that these examples are not equivalent to the data. Instead, the term “example” designates the complex assemblage by which data is aggregated, formatted, and related to an objective so that norms can emerge to enable predictive or classificatory activity. Machine learning decisively changes the numerical logic of exemplarity, transforming the example from the singular instantiation of a type to a multiplicity. And whereas programming rules prescribe explicitly in advance, in machine learning’s artificial naturalism, norms *emerge* recursively through exposure of models to examples at scale.

In the three sections below, we trace a very recent historical movement that begins with data labeling as a form of interventionist example-making and moves toward feature engineering, where engineering knowledge is used to identify useful representations of data. These practices point toward an idealized horizon of full self-supervised automation and zero-shot learning (the use of models on tasks for which they were not trained), making possible the application of these norms across domains (Brown et al., 2020). In practice, we stress that things are considerably more complicated, with human-specified norms continuing to intervene in fine-tuning or reinforcement learning with human feedback.

Labeling

In contrast to the rule-based programming logic, where rules are prescribed explicitly and abstractly in advance, the example-based logic of machine learning begins at the end: with a set of desired concrete outputs. The choice and form of desired output obviously play a critical role as a source of norms. A common way of making data exemplary is to label it (Jaton, 2021: 9). We stress that even this well-understood process involves subtleties that point not only to a single normative moment of *ascription* of labels but also to a host of other practices that permit the emergence of norms through aggregation, making certain features intelligible while discarding others.

The influential ImageNet dataset stands as a paradigmatic instance of example-making by labeling in deep learning (Azar et al., 2021; Denton et al., 2021; Gershgorin, 2017). A set of studies has shown that the production of datasets is not only labor and resource intensive, as it relies on the outsourcing and exploitation of precarious labor on platforms like Amazon Mechanical Turk; dataset production is also *reproduction* of a preexisting onto-epistemology (Crawford and Paglen, 2021: 1113) or ground truth (Jaton, 2017; 2021): here, the hierarchical classificatory schema derived from the lexical database WordNet. In addition to the idea that referential relationships between images and labels are contingently constructed, we wish to emphasize that other practices of standardization and aggregation are required to elicit the norms that can produce classificatory decisions.

An even earlier moment of normalization evokes an ancient etymological sense. The philosopher Canguilhem (1991: 125) observes that the original meaning of the word “norm” derives from the Latin word for a carpenter’s tool, a T-square, which allows for the creation of right angles. Much work in computer vision dataset creation involved orientation: centering and cropping images so that invariant features and patterns can be identified by algorithms. An important predecessor of ImageNet, the CalTech 101 dataset, cites orientation as a motivating problem: “...can we find a natural alignment for images of octopuses, of cappuccino machines, of bonsai trees?” (Fei-Fei et al., 2004: 178). This is not a mere engineering problem; normalization in this sense of standardized spatial arrangement of objects is required to make classes and structures emerge.

The concrete attribution of class labels to images by human workers is at the heart of exemplification in the labeling paradigm. This is usually understood not as an explicit prescription of norms but as a more neutral description built on implicit forms of consensus. Considerable ingenuity is required to match a given image to a *single* label at scales required for the production of exemplary norms in ImageNet. On Amazon Mechanical Turk, the worker is provided with a label (“concept”), a definition of the concept, and a set of images from which to choose the ones corresponding to the label.⁸ Label attributions are validated by groups of workers, giving a pragmatic, even democratic flavor to the normative process of example-making. The more difficult a concept is to illustrate, the greater the number of labelers required to reach an agreement (Deng et al., 2009: 6; Kovashka et al., 2016: 212). The designers of ImageNet are aware that thresholds of consensus are to some degree arbitrary and become more elusive as categories become more and more fine-grained in so-called edge cases; they even created “an algorithm to dynamically determine the number of agreements needed for different categories of images” (Deng et al., 2009: 5). These crowdsourcing

methods extend the ascriptive moment of ground-truthing to a wider but often hidden set of annotation workers who draw on implicit knowledge and reach classification agreements by majority decision. ImageNet nonetheless circumscribes the spaces in which norms can emerge by predictively prepopulating a small set of images (some random and others corresponding to the label) from which the labeler chooses the best match. Human labelers function as cogs in an algorithmic system (“humans in the loop”), which constantly evaluates both their efficiency and their accuracy (Kovashka et al., 2016: 213). Their discretion is reduced to a minimum, save perhaps the ability to give time-consuming and unpaid feedback in the case no good answer can be given to a prompt.

Machine learning datasets introduce a new logic of scale into example-making. Traditionally, an example was a concrete *singularity* that embodies a type by expressing its most distinctive features. Data labeling processes involve a conceptually more minimal (but still labor-intensive) process of matching single instances, which are only partially representative of the general type, with a generic label. Epistemologically, this relationship is not reflective but inductive in the sense that by collecting many diverse instances (scale), a model can be trained to produce a more robust representation of the type, a representation that captures its invariant features. This process is normative in the sense that desired outputs and “problematizations”—task definitions—guide the selection of what attributes can be named and made intelligible (Jaton, 2021: 2). Whereas a rule-based procedure would preselect these attributes and attempt to explicitly specify the rules allowing for moving from inputs to outputs, in machine learning, this intermediate stage is left implicit in training.

We have emphasized some of the ways in which labeling produces examples that can be processed by a model. We now turn to some of the effects of the exemplary type of authority, particularly the distinctive form of subjectivity that it produces. How do we behave when we think of the observable world as a particular instance of some more fundamental data-generating distribution, which can be accessed through the representations produced by machine learning algorithms?⁹ The ancestry and ethnicity analysis services offered by corporations like 23andMe illustrate some of these aspects of machine learning’s authority. Patterns discovered in the DNA of clients are made to fit cultural assumptions about ethnic and racial identification, which are then applied to subjects (Cachoian-Schanz and Schwerzmann, 2023). The connection between ethnoracial labels and carefully drawn clusters grants the output an epistemic authority based on structures claimed to be present in the data and thus in the DNA but inaccessible to human perception.

To categorize a consumer’s ancestry via DNA from a saliva sample, 23andMe relies on a group of initial testing subjects who are asked to self-identify based on what

they know about their ancestors and their geographic origins (23andMe, 2020a). 23andMe then draws the lines around what it considers as forming a distinctive population by identifying and labeling genetic patterns discovered in the dataset. The appearance of these structures depends on assumptions about the relation of human DNA to a geographic origin and the subsequent connections between ethnicities and national borders. Issues like population migration that may compromise the purity of the data, in other terms, the homogeneity of the DNA supposed to exist inside a population, are explicitly excluded:

When a 23andMe research participant tells us that they have four grandparents all born in the same country—and the population of that country didn't experience massive migration in the last few hundred years, as happened throughout the Americas and in Australia, for example—that person becomes a candidate for inclusion in the reference data ... And we remove outliers, people whose genetic ancestry doesn't seem to match up with their survey answers. To ensure a clean dataset, we filter aggressively—nearly ten percent of reference dataset candidates don't make the cut [emphasis ours] (23andMe, 2020a).

The explanatory documents provided by 23andMe characteristically combine sociohistorical assumptions about what constitutes countries of migration with technical considerations regarding issues of overfitting: “Most country-level populations overlap to some degree, though. In those cases, we *experimented* with different groupings of country-level populations to find combinations that we could distinguish with high confidence” [emphasis ours] (23andMe, 2020a). Here, the process of clustering and, critically, labeling these clusters in ways that are meaningful is described as “experimental” and is done at the discretion of the company. The implicit and relatively silent character of examples justifies, according to proponents, these interpretive acts. However, this exemplification process is covered over in the staging of the output (in terms of likelihood of belonging to one or several ethnicities). This likelihood is instead framed as naturally emerging from the data, itself a faithful reflection of our DNA.

In Figure 1, the shape and color of each data point symbolize one of the labels listed on the right, referring to the “experimental” grouping or clustering of data by 23andMe. Without this interpretation of output structures—that is, without ethnic categories chosen without a

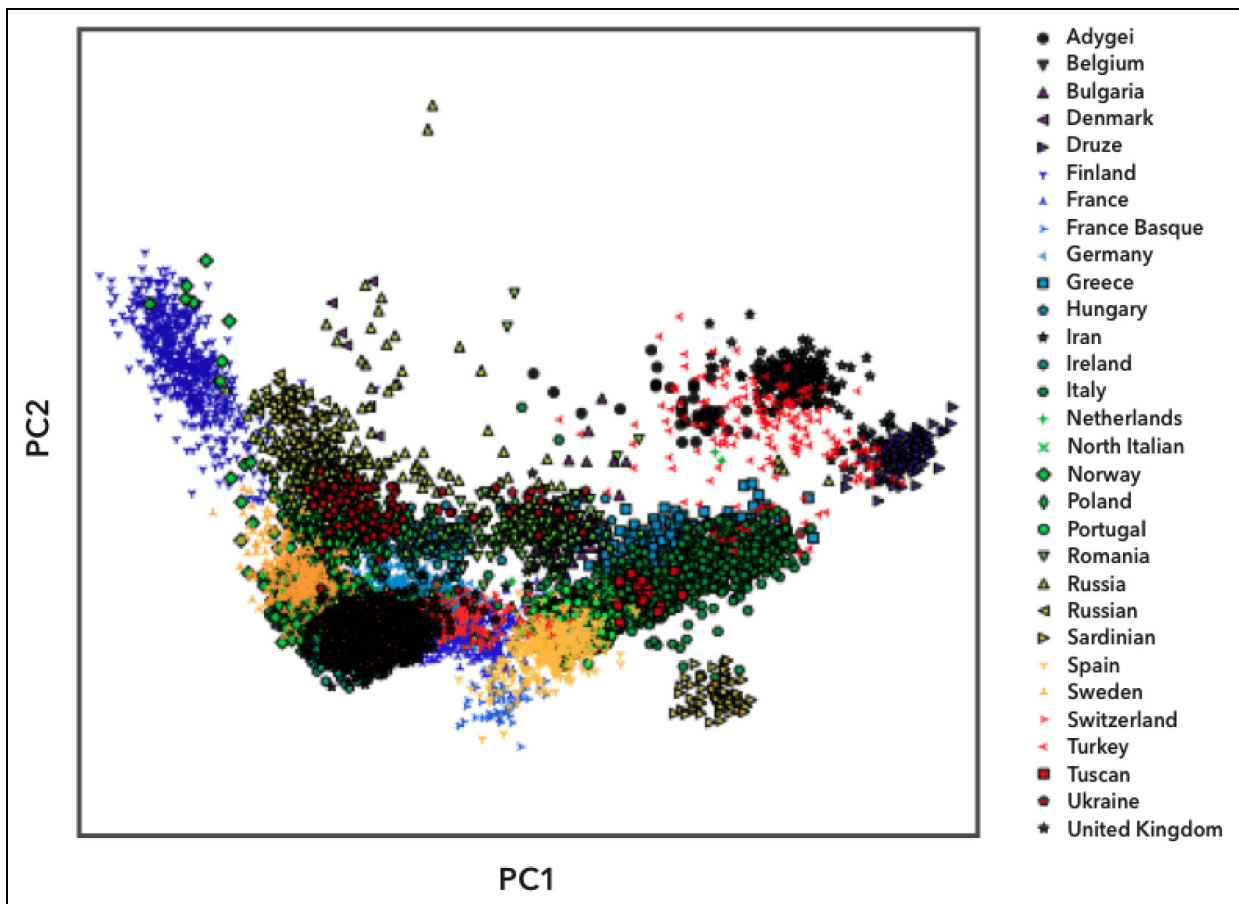


Figure 1. 23andMe graph visualizing the “principal components plot of 23andMe reference European populations” (23andMe, 2020a).

clarification of the supposed connection between DNA, ethnicity, and nation-states—one would only see a tight cluster of points spatially distributed on the graph, with the exception of a few isolated clusters. Meaning only emerges through the “experimental” labeling of the output clusters, which itself relies, among others, on the personal account of participants who assure that their ancestry can be tracked to one and one place only. While the implicit character of machine learning norms tends to produce ambivalent and complex mathematical distributions or estimates, human interpretation makes these results actionable, for instance, permitting lines to be drawn around clusters and attributing meaning to them. These interventions are legitimated epistemologically by the belief that the processing of genetic data reveals apparently natural patterns in the DNA that could not have been otherwise perceived. At the same time, the staging of the results tends to downplay or even hide the production of implicit norms in both labeling of inputs and interpretation of outputs.

Unlike the supervised form of learning using the ImageNet dataset, it may seem that there is a less intentional, ascriptive normalization of data in unsupervised methods like clustering, which are said to only reflect what “is”: the underlying structure in the data. However, we wish to relativize these differences, to focus instead on the continuum of practices required to produce examples, elicit norms, and exercise authority through them. As the 23andMe example shows, the structures or patterns discovered in the data have to be “produced–interpreted” to form meaningful clusters in terms of the given task. By “produced–interpreted,” we mean that the discovery of a structure and its interpretation are inextricably entangled with what the researcher is looking for and the standards and expectations she sets, even when these remain implicit (unlike, for instance, in the explicit formulation of programming rules or ground-truthing). This forms an important source for the authority of examples in machine learning. Note, for instance, the naturalization of the learning process associated with its unsupervised character: “[Clustering] involves the *automatic* identification of *natural data groups* (the clusters)” [emphasis ours] (Marzell, 2021). While they may appear accurate in relation to a given task or benchmark, the patterns discovered through unsupervised algorithms do not convey their meaning explicitly. They do not make their principles of composition intelligible (Miller, 2019). Nor do they command or prohibit in any prescriptive sense as rules do.

This implicit type of authority produces a form of subjectivity that differs from the alienation said to accompany the rational type of authority exercised by rules. Those subjected to machine learning’s classifications are forced to search for and construct rationales for unintelligible results. There are a distinctive set of subjective effects and affects tied to the authority of exemplification. The angry, disappointed, or skeptical comments to a blog in

which 23andMe explained an update to their model dramatize the conflict between family narratives, rule-based, bureaucratic modes of identity production, and machine learning’s identity attribution. After the update, the clients are critical of the new classifications, which they see as having become less accurate. Consumers seem to expect to receive a scientific confirmation of what they know about themselves. When this fails to materialize, it is not the legitimacy of the apparatus of DNA ethnic analysis and its myriad implicit assumptions about race and ethnicity that are attacked as inconsistent or problematic. The user “RPK” demands: “Give me back my Irish grandparents! They disappeared with this update. I have documentation that they emigrated from County Kidare in 1848. Both their parent sets were also Irish. The previous report nailed it to within 20 miles! Now I have no Irish blood. Something is wrong with your update. Please fix it” (23andMe, 2020b). Here, the algorithm is granted the power to symbolically make the user’s grandparents disappear by entirely rewriting what the user knows about themselves, which was established through bureaucratic modes of identification. In a similar way, “Mrs. C.” describes her self-conception as shattered by the updated results: “I’m almost half the woman I used to be thanks to the errors in your new and ‘improved’ algorithm. Your new algorithm incorrectly removed all of my French and German heritage that was accurately listed prior to the algorithm update ... I’ve lost all faith in the accuracy of your data and I’m make sure [sic] to let others know of your errors as well. Please make this right and fix these errors” (23andMe, 2020b). While errors are attributed to the algorithm and the accuracy of the data, the connection of race and ethnicity with DNA ridden with problematic socio-political assumptions remains entirely unchallenged. Again, our claim is that machine learning’s logic of exemplarity grants this connection the legitimacy of a natural fact emerging from the processing of data through a model.

The 23andMe example showcases the relationship between two types of authority: the older rational type legitimized by law and bureaucracy and the emerging exemplary type legitimized through the assemblage of data and models. While the bureaucratic rules associated with the rational type of authority—here the rules of identity attribution—come with their own violence, opacity, and arbitrariness, they apply to every individual in the same way (at least in principle) so that individuals can orient themselves in relation to the rules and potentially contest them. Bureaucratic and computational rules point back to a foundational moment of prescriptive specification or encoding, which can be contested. By contrast, being ruled by examples means never accessing the implicit, experimental norms elicited from examples and constantly updated through the optimization of the model. With machine learning’s exemplary type of authority, norms are *emergent* and constantly subject to

change due to both alterations in the models and modifications in the very behaviors modeled.

Feature engineering

The success of deep learning methods on large, labeled datasets like ImageNet inspired the research community to look for a more radical understanding of exemplarity in which the structure of the data itself, rather than the ascription of external labels (always an expensive and time-consuming process), could produce the norms required to classify and represent. Up to this critical period around 2010, it was widely acknowledged that the most important factor in the success of a machine learning project was in preprocessing of data or feature engineering. Feature engineering (the latter word connoting intentionality and design) transforms data so as to produce examples that can effectively train the model (Bengio et al., 2013: 1798). Originally, this was based on the implicit knowledge accumulated by the programmers while working with models. Tactile metaphors predominate in descriptions of these practices—instead of intentionally “handcrafting” programming rules, engineers “turn knobs” back and forth, searching for optima. In an influential 2012 article, Pedro Domingos explains how examples may be constructed from data in this way: “Often, the raw data is not in a form that is amenable to learning, but *you can construct features* from it that are. This is typically where most of the effort in a machine learning project goes. It is often also one of the most interesting parts where *intuition, creativity, and ‘black art’* are as important as the technical stuff” [emphasis ours] (Domingos, 2012: 84). Feature engineering takes the form not of labeled outputs but of more general assumptions or hypotheses about one’s data and models (an implicit form of knowledge production, contrasting with the universalist, prescriptive, and explicit logic of rule-making). This reverses the directionality of the input–output dynamic of algorithmic rules. Instead of specifying rules to govern the movement from input to output as tightly as possible through an exhaustive and explicit specification of steps, engineers seek transformations of the data in aggregate that will elicit desired outputs. Ideally, these transformations capture essential information about the task and make mathematical optimization more tractable. Authority shifts from tight control over the behavior of a calculative process to a more speculative, provisional form of authority based on hypotheses about the features that can most effectively discriminate between inputs.

However, as Bengio et al. (2013: 1798) explain, these ad hoc engineering practices, while effective, share some of the same limitations as labeling: “Such feature engineering is important,” they write, “but labor intensive and highlights the weakness of current learning algorithms: Their inability to extract and organize the discriminative information from

the data.” The implication, as in self-supervised learning, is that models *themselves*, not human engineers, should be able to extract learning norms from data. This leads to a second tension within the artificial naturalism of machine learning’s type of authority. Instead of an explicit series of rules or task-dependent engineering, machine learning should rely on “generic priors,” which “capture the posterior distribution of the explanatory factors of the observed input” (Bengio et al., 2013: 1798). This view rests on an idea that the more minimal these mediations are, the more the data can speak for itself. However, this naturalism is complicated by the fact that extreme care is required to select such priors capable of creating a context in which these exemplary norms can emerge.

Scaling

Perhaps, what most distinguishes machine learning’s type of authority from prior senses of exemplarity is its awesome sense of scale, expressed in many different ways: the number of model parameters, the amount of compute required to train a model, and the size of datasets. From the standpoint of contemporary large language models (LLMs), which tend to be trained on unlabeled data, labeled datasets like ImageNet constitute an intermediate logic of exemplarity. Such collections of labeled images, to be sure, are not individuals in the prototypical sense of the singularity that most perfectly embodies the essence of some type or class—a one-to-many relationship. Nonetheless, a smaller set of (labeled) individuals elicits the norms required to characterize a potentially larger set by the identification of common, distinctive features. However, when representations emerge less and less from ascriptive interventions like labeling and more from immanent relationships within the data itself, this numerical imbalance between the sample and the population (to somewhat anachronistically deploy a statistical concept) becomes murky. Instead of isolating the special case from prosaic everyday reality (and indeed, this exceptional, perfect status was a former source of exemplary normativity and authority), in machine learning, examples promise to *converge* with what they map, an important aspect of machine learning’s naturalism.

LLMs illustrate this asymptotic naturalism. These systems (like GPT-3) tend to be self-supervised, deep learning, decoder-only models with huge numbers of parameters trained on very large datasets of natural language. Their “objective [norm-giving task] is ... to predict the next token [a unit of text] given the preceding tokens in the example” (Chowdhery et al., 2022: 3). This apparently narrow remit seems remarkably adaptable to numerous domains in what is termed few-shot learning; LLMs can be adapted for other tasks simply by using them as natural language interfaces, with comparatively little

domain-specific training. A group of researchers based at Stanford University describes their qualities of “emergence” and “homogenization” in a way that resonates with our understanding of examples. “Emergence,” they explain, “means that the behavior of a system is *implicitly induced rather than explicitly constructed*; it is both the source of scientific excitement and anxiety about unanticipated consequences. Homogenization indicates the consolidation of methodologies for building machine learning systems *across a wide range of applications*” [emphasis ours] (Bommasani et al., 2022: 3). In their implicitness and generality, these models crystallize machine learning’s exemplary type of authority. From a comparative perspective, it is striking that these models tend to be controlled using natural language in the form of prompts rather than the formal and unambiguous programming languages used to express computational rules. Prompts elicit rather than explicitly command. Indeed, early work shows that the same prompt can sometimes produce different results, suggesting a more open, probabilistic relationship between input and output.

Scale shifts this logic of exemplary authority from a singular transcendent source of norms to one that appears to be immanent with the data itself—yet retains a norm-giving power. Where “ought” was traditionally marked out from “is”—as the example reflected the essential features of the type and differed from a mere instance (*token*)—data, compute, model, and scale asymptotically *merge ought* and *is*. The appearance of slogans like “scale is all you need” even implies that scale is some sort of sufficient condition. However, it would be misleading to conclude that large-scale datasets are free of human interventions required to produce smaller labeled datasets. Indeed, talk of scale in the abstract can obscure the various normative operations that characterize dataset construction methods and assumptions, although these details are becoming more difficult to publicly scrutinize.¹⁰ One large, unlabeled dataset, Google’s Generalist Language Model (GLaM), does include some high-level discussion on its training set. Invoking size appears to justify claims that it is representative of natural language in some vague broader sense: “To train our model, we build a high-quality dataset of 1.6 trillion tokens [a linguistic unit] that are representative of a wide range of natural language use cases” (Du et al., 2021: 2). However, the authors later provide a much more concrete description of the way that they engineered a normative mixture of linguistic sources; one key innovation was the use of “a text quality classifier” to engineer a mixture of predominantly “high-quality” linguistic sources: “Wikipedia, books and a few selected websites” (Du et al., 2021: 2). The authors subsequently observe that smaller, higher-quality filtered training sets of 143 billion tokens produce better results than an unfiltered set of 7 trillion tokens, especially on language generation

tasks (Du et al., 2021: 7–8). Thus, scale, while clearly playing an important role in the performance of the model, is not sufficient; the selection of dataset sources also constitutes a better-defined norm-giving space.

In machine learning’s exemplary type of authority, scale points to a different horizon of universality than that of the rational authority of computational rules. The latter relied on the quasi-infinite space for the specification of explicit rules in computer memories theorized by Turing. Machine learning’s universality instead makes datasets as large as possible so as to cover the widest range of possible situations and applications, as can be seen in the interest in few- and zero-shot learning (“homogenization” in the words of the Stanford researchers). These datasets are treated unproblematically as notional samples of some prior distribution that can be grasped through generative modeling. Norms are elicited from these datasets through generic priors that are flexible and implicit in their minimalism, encompassing many possible sources of variation in order to produce exemplary representations. This points to a change in the *form* of predictability at the heart of machine learning’s type of authority. It no longer means total control over an analytically specified calculative process but rather the probabilistic association of desired input–output relationships on the basis of some implicit-level structure imputed from datasets engineered according to emergent norms. These structures or patterns in high-dimensional spaces are beyond thresholds of human recognition, which become more and more intelligible to models at great scale and with high dimensionality, constituting a horizon of convergence between normative examples and the phenomena that they supposedly represent.

Conclusion: The new authority of examples in machine learning

Our genealogies have focused on points of distinction between a rule-based programming paradigm and a more recent machine learning paradigm based on examples. In the middle of the 20th century, figures like Turing produced a radical account of computational rules that are (a) totally explicit and expressed in a formal programming language and (b) created for specific tasks through an exhaustive analysis of a behavior while (c) “universal” in that the potential space for their elaboration is large enough to cover any task. Finally, (d) rules work in discrete states, which makes the movement from input to output entirely predictable. This understanding of rules was made possible by a long and heterogeneous set of developments shaped by a series of transformations that put rules at the center of our accounts of knowledge (Cartesian method) and ethical life (Kantian imperatives). Critically, Weber’s sociological account of the rational type of authority in the modern bureaucratic order shows how rules became organized in the explicit,

abstract, and systematic bodies that would later characterize computational rules.

Computer scientists often present machine learning in terms of a discontinuity with the rule-based programming paradigm, which, after collapsing under its own brittle weight, is said to have been superseded by a paradigm of training based on examples. The reality is more complex, as both paradigms coexist. Our analysis nonetheless isolates key features of the emerging example-based paradigm in terms of authority. While Weber's rational type of authority operates by rules, machine learning's authority produces obedience through norms elicited from examples. The very recent history of machine learning further illuminates this exemplary logic, beginning with the relatively intentional introduction of norms by human labelers for computer vision datasets alongside "feature engineering," which designates the use of implicit and intuitive engineering knowledge to identify relevant aspects of the data. Although different from the unequivocal explicitness of algorithmic rules, these relatively direct interventions played a major role in making data exemplary, allowing it to elicit the representations required to associate inputs and outputs. However, the machine learning community perceived these interventions ambivalently, lamenting that models were not yet capable of learning something informative from the structure of the data *itself* rather than simply "memorizing" human-assigned input-label associations. "Scale" has emerged as a more naturalistic alternative to human intervention, blurring "ought" and "is."

It is important to reiterate that "example" in machine learning is not congruent with "data." Instead, it designates the complex assemblage by which data is aggregated, formatted, and processed so that norms can emerge. Machine learning's exemplary authority contrasts with the constructed, transcendent authority of rules: whereas rules are imposed from above on the diversity of things they govern, norms emerge from tasks and examples and are immanent to them. What can it mean to be governed according to machine learning's type of authority and its implicit normativity? To conclude, we offer a few hypotheses. While both algorithmic rules and machine learning norms guide our conduct by making it more predictable, the nature of this predictability, the connection between input and output, has changed. It may be tempting to think that machine learning's exemplary type of authority is more open to variability and change. Instead of thinking in quantitative terms, we prefer to emphasize qualitative differences between these regimes in terms of the sources and legitimacy of their authority as well as possible modes of resistance. While examples are instrumental—produced in light of a task or objective—they elicit norms that appear to express natural regularities or types.

The proponents of machine learning claim that machine learning-based algorithms can predict or identify behaviors more accurately—especially more accurately than human

counterparts—thanks to the large amount of information they can process. This instrumental, task-oriented logic is underpinned by naturalist assumptions that combine epistemological and ethical aspects: the predictive efficacy of machine learning is made possible and justified by claims to identify some "underlying explanatory factors for the observed input"—even if these are unintelligible to the governed or even those who built the models (Bengio et al., 2013: 1798). Whereas rules tend to acknowledge their "constructedness," machine learning norms are construed as being of the same kind as prior probability distributions, natural regularities.

The 23andMe case provides some insights on what it means to be ruled by two conflicting types of authority: the exemplary and the rational. Instead of producing alienation and friction between rules as general principles and empirical diversity, examples *implicate us*, inducing anger and anxiety. Because the principles of prediction or classification are implicit, we can never know what parts of our behaviors, characteristics, or identities might have caused us to be associated with a certain output category or label. The explicit intransigence of rules dissolves into a series of ever-modifiable but unintelligible correlations. How might we resist being ruled by examples when each subject is interpellated by machine learning's type of authority in a slightly different way, as no individual produces exactly the same data and, thus, is classified following the same norm? We cannot offer any definitive prescriptions, only concepts and comparisons that need to be tested in further empirical research and in local contexts. However, it seems like being ruled through machine learning's type of authority implies a specific disassembling of a liberal, rule-based form of political subjectivity, as data extraction latches on always partial parts of ourselves (any kind of contingent behavior that can be extracted) to then reassemble those parts following ever-changing criteria. The foundational moment that characterizes the political disappears into myriad microdecisions (Sprenger, 2020), obeying implicit norms. This transformation could lead to the dissolution of the common ground necessary for collective agency and struggle.

There have been some visionary attempts at theorizing such subjectivities—Deleuze's (1992) idea of the "dividual" comes to mind—but thus far they have been underspecified and tentative. We hope that further concrete engagement with machine learning's exemplary type of authority at both technical and theoretical levels will help us further specify the way we are affected by it and how we can resist it. Perhaps, the "architectures" of machine learning models will take their place alongside the agora, the square, the theater, and the street as sites of political contestation.

Acknowledgments

The authors wish to thank the editors and two anonymous peer reviewers whose comments greatly improved this article. We

also wish to thank Louise Amore, Benjamin Jacobsen, Ludovico Rella, Eamonn Bell, and Dmitry Muravev for their comments on earlier drafts.



Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research has received funding from the European Research Council (ERC) under Horizon 2020, Advanced Investigator Grant ERC-2019-ADG-883107-ALGOSOC “Algorithmic Societies: Ethical Life in the Machine Learning Age” and “INTERACT! New Forms of Social Interaction with Intelligent Systems,” Ruhr-Universität Bochum.

ORCID iDs

Alexander Campolo  <https://orcid.org/0000-0003-3159-4131>
Katia Schwerzmann  <https://orcid.org/0000-0002-7938-2608>

Notes

1. The terms “is” and “ought” are often associated with the philosopher David Hume, who controversially denied that moral conclusions (“ought”) could be derived from factual premises (“is”). We reprise these terms to signal the ambiguous positions of facts and norms in machine learning (Hume, 2007: 302). See also Grosman and Reigeluth (2019: 2).
2. We note that there are many other important moments in machine learning pipelines, which these norms are refined that we cannot cover here, notably in testing and benchmarking models.
3. We wish to thank Anastasia Klimchynskaya who introduced the authors to Zola’s text.
4. See Daston (2022: 82–105) for a history of calculative rules and Jones (2016) for a longer history of calculating machines extending to the early modern period.
5. Where Daston emphasizes the “hybrid intelligence” of algorithmic rules—a co-constitution of human and machine calculation from which humans were never definitely excluded—we emphasize differences (Daston, 2022: 126).
6. The actual history of expert systems shows that many of its practitioners themselves made this critique. J. Ross Quinlan and Donald Michie initiated an example-led paradigm that led toward more contemporary forms of machine learning (Jones, 2023: 204).
7. In this sense, our perspective on examples may share commonalities with the “indexical” characterization of artificial intelligence of Weatherby and Justice (2022).
8. For an idea of the interface, see Fei Fei (2010: 24–25).
9. Phan and Wark (2021: 4) question this naturalistic paradigm by emphasizing that machine learning, while making perceptible things that have so far escaped human regimes of perception and in particular vision, in fact, produces “novel, non-visual ground” for old logics like, for instance, race.

10. Rather notoriously, OpenAI’s recent GPT-4 model contains no discussion of its training dataset, citing concerns about competition and safety (OpenAI, 2023: 2).

References

- 23andMe (2020a) Ancestry composition guide. Available at: www.23andme.com/ancestry-composition-guide-pre-v5/ (accessed February 14, 2022).
- 23andMe (2020b) The 23andMe ancestry algorithm gets an upgrade. Available at: blog.23andme.com/ancestry-reports/algorithm-gets-an-upgrade/ (accessed February 14, 2022).
- Amore L (2022) Machine learning political orders. *Review of International Studies* 49(1): 1–17.
- Azar M, Cox G and Impett L (2021) Introduction: Ways of machine seeing. *AI & Society* 36(4): 1093–1104.
- Bechmann A and Bowker GC (2019) Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society* 6(1): 205395171881956.
- Bengio Y, Courville A and Vincent P (2013) Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8): 1798–1828.
- Bommasani R, et al. (2022) On the opportunities and risks of foundation models. Report, Center for Research on Foundation Models, Stanford University, July.
- Brown T, et al. (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33: 1877–1901.
- Cachoian-Schanz D and Schwerzmann K (2023) “One unique you”: Affective attachments and DNA-testing as ethnotechnological apparatus. *Social Text* 41(1): 71–97.
- Canguilhem G (1991) *The Normal and the Pathological*. New York: Zone Books.
- Chollet F (2021) *Deep Learning with Python*. Shelter Island, NY: Manning Publications Co.
- Chowdhery A, et al. (2022) PaLM: Scaling language modeling with pathways. Available at: arxiv.org/pdf/2204.02311.pdf.
- Chun WHK (2021) *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.
- Crawford K and Paglen T (2021) Excavating AI: The politics of images in machine learning training sets. *AI & Society* 36: 1105–1116.
- Daston L (2022) *Rules: A Short History of What We Live By*. Princeton: Princeton University Press.
- Daston L and Galison P (2010) *Objectivity*. New York: Zone Books.
- Deleuze G (1992) Postscript on the societies of control. *October* 59: 3–7.
- Deng J, et al. (2009) ImageNet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. New York: IEEE, 248–255.
- Denton E, et al. (2021) On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society* 8(2): 1–14.
- Descartes R (2006) *A Discourse on the Method of Correctly Conducting One’s Reason and Seeking Truth in the Sciences*. Oxford: Oxford University Press.

- Domingos P (2012) A few useful things to know about machine learning. *Communications of the ACM* 55(10): 78–87.
- Du N, et al. (2021) GLaM: Efficient scaling of language models with mixture-of-experts. Available at: arxiv.org/pdf/2112.06905.pdf.
- Erickson P, et al. (2013) *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. Chicago: University of Chicago Press.
- Farrell H and Fourcade M (2023) The moral economy of high-tech modernism. *Daedalus* 152(1): 225–235.
- Fei-Fei L (2010) ImageNet: Crowdsourcing, benchmarking & other cool things. Talk at the Robotic Institute, Carnegie Mellon University, 29 March. Available at: https://www.image-net.org/static_files/papers/ImageNet_2010.pdf.
- Fei-Fei L, Fergus R and Perona P (2004) Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In: *2004 IEEE conference on computer vision and pattern recognition workshop*. New York: IEEE, 178–186.
- Foucault M (2008) *The Birth of Biopolitics: Lectures at the Collège de France, 1978–1979*. New York: Picador.
- Gershgorn D (2017) The data that transformed AI research—and possibly the world. Quartz, 26 July. Available at: <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world> (accessed 19 March 2023).
- Gitelman L (2013) *“Raw Data” Is an Oxymoron*. Cambridge, MA: MIT Press.
- Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning*. Cambridge, MA: MIT Press.
- Grosman J and Reigeluth T (2019) Perspectives on algorithmic normativities: Engineers, objects, activities. *Big Data & Society* 6(2): 1–12.
- Hume D (2007) *A Treatise of Human Nature: A Critical Edition*. Oxford: Clarendon Press.
- Jaton F (2017) We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science* 47(6): 811–840.
- Jaton F (2021) Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society* 8(1): 1–10.
- Jones M (2016) *Reckoning with Matter: Calculating Machines, Innovation, and Thinking about Thinking from Pascal to Babbage*. Chicago: University of Chicago Press.
- Jones M (2023) Decision trees, random forests, and the genealogy of the black box. In: Ames MG and Mazzotti M (eds) *Algorithmic Modernity: Mechanizing Thought and Action 1500-2000*. Oxford: Oxford University Press, 190–215.
- Kant I (1997) *Groundwork for the Metaphysics of Morals*. Cambridge: Cambridge University Press.
- Kant I (2015) *Critique of Practical Reason*. Cambridge: Cambridge University Press.
- Kovashka A, et al. (2016) Crowdsourcing in computer vision. *Computer Graphics and Vision* 10(2): 103–175.
- Latour B (1999) Circulating reference: Sampling the soil in the Amazon forest. In: *Pandora’s Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press, 24–79.
- Light J (1999) When computers were women. *Technology and Culture* 40(3): 455–483.
- Marzell T (2021) Clustering with machine learning: A comprehensive guide. Available at rocketloop.de/en/blog/clustering-machine-learning-comprehensive-guide/ (accessed 6 April 2023).
- Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267: 1–38.
- Mitchell M (2021) Why AI is harder than we think. In: *Proceedings of the genetic and evolutionary computation conference*. New York: Association for Computing Machinery, 1–12.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- OpenAI (2023) GPT-4 technical report. Available at: arxiv.org/pdf/2303.08774.pdf.
- Phan T and Wark S (2021) Racial formations as data formations. *Big Data & Society* 8(2): 1–5.
- Sprenger F (2020) Microdecisions and autonomy in self-driving cars: Virtual probabilities. *AI & Society* 37: 619–634.
- Turing A (1950) Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy* 59(236): 433–460.
- Upton G and Cook I (2014) *A Dictionary of Statistics*. Oxford: Oxford University Press.
- Weatherby L and Justie B (2022) Indexical AI. *Critical Inquiry* 48(2): 381–415.
- Weber M (1978) *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.
- Weber M (1980) *Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie*. Tübingen: Mohr Siebeck.
- Zola É (1893) *The Experimental Novel and Other Essays*. New York: Cassell Publishing Co.