

Machine learning, meaning making: On reading computer science texts

Louise Amoore , Alexander Campolo , Benjamin Jacobsen and Ludovico Rella

Big Data & Society
January–June: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517231166887
journals.sagepub.com/home/bds

Abstract

Computer science tends to foreclose the reading of its texts by social science and humanities scholars – via code and scale, mathematics, black box opacities, secret or proprietary models. Yet, when computer science papers are read in order to better understand what machine learning means for societies, a form of reading is brought to bear that is not primarily about excavating the hidden meaning of a text or exposing underlying truths about science. Not strictly reading to make sense or to discern definitive meaning of computer science texts, reading is an engagement with the sense-making and meaning-making that takes place. We propose a strategy for reading computer science that is attentive to the act of reading itself, that stays close to the difficulty involved in all forms of reading, and that works with the text as already properly belonging to the ethico-politics that this difficulty engenders. Addressing a series of three “reading problems” – genre, readability, and meaning – we discuss machine learning textbooks and papers as sites where today’s algorithmic models are actively giving accounts of their paradigmatic worldview. Much more than matters of technical definition or proof of concept, texts are sites where concepts are forged and contested. In our times, when the political application of AI and machine learning is so commonly geared to settle or predict difficult societal problems in advance, a reading strategy must open the gaps and difficulties of that which cannot be settled or resolved.

Keywords

machine learning, neural networks, ethics, reading, politics, computer science

Reading, if it happens at all, happens only in the encounter with difficulty and without guarantees (Keenan, 1994: 103).

In a series of interviews recorded in 2018, Turing laureates Geoffrey Hinton and Yann LeCun reflect on the breakthrough moments that shaped contemporary machine learning and artificial intelligence. Discussing the development of neural network algorithms, they describe how their work on backpropagation and convolutional neural networks had initially been rejected by the computer science community. “They rejected papers by Yann, even though they worked better on particular problems,” Hinton remarks, because “the referees thought it was the wrong way to do things” (Ford, 2018: 76). “The general consensus among statisticians and people in AI was that we were wishful thinkers,” Hinton continues, “who thought that just from the inputs and outputs you should be able to learn all these weights” (2018: 76). LeCun similarly comments on the difficulty of publishing on neural networks in the 1980s when it was “anathema within the community.

You couldn’t publish a paper that even mentioned the phrase neural networks because it would immediately be rejected by your peers” (2018: 122). In the accounts of Hinton and LeCun, the published computer science paper does not primarily reveal the state-of-the-art in machine learning, but rather it actively frames what can and cannot be accepted as knowledge in a machine learning community.

The computer science text – and whether and how it is read by others – is a lively and contested site through which machine learning shapes the world. When work is rejected by a major computer science conference, the failure of a paper to make it out into the world is said to obscure what is “really going on” in the field. As LeCun recounts, “Geoffrey Hinton and Terry Sejnowski published a very famous paper in 1983 [...] which described an early

Department of Geography, Durham University, Durham, UK

Corresponding author:

Louise Amoore, Department of Geography, Durham University, South Road, Durham DH1 3LE, UK.
Email: Louise.amoore@durham.ac.uk



deep learning or neural network model” but the authors “had to use code words to avoid mentioning that it was a neural network” and “even the title of their paper was cryptic” (Ford, 2018: 122). What is read in computer science, and indeed what remains unread because of its disavowal or rejection, matters to the account this science gives of itself and its machine learning discoveries. Alternatively, key papers mark the historical trajectory of the field: a single computer science paper – the 2012 “ImageNet Classification with Deep Convolutional Neural Networks,” with more than one hundred thousand citations – is widely understood to represent the disruptive moment when machine learning and deep neural networks broke through (Krizhevsky et al., 2012). As Hinton describes their AlexNet algorithm – when the results were so remarkable that they could no longer be ignored by the community that had long rejected papers on neural networks – “in the end, *science won out*” (Ford, 2018: 76).

These landmark texts of computer science may appear to offer material that can be read in order to *make sense* of what is taking place in otherwise opaque worlds of machine learning. Yet the readership of machine learning is heavily circumscribed, with texts explicitly stating “who should read” them and addressing “students of deep learning and AI” and “software engineers” who wish to “begin using deep learning in their product or platform” (Goodfellow et al., 2016: 8). Although machine learning certainly encompasses a huge range of practices and material technologies, such comments suggest that it also understands itself in a textual way, as a field of literature whose velocity is one of its signatures. As social science and humanities scholars we are interested in how machine learning knowledge transforms how societies understand themselves – and though we are not usually addressed as readers – these texts also form a significant part of the material that we read, research, and think about.¹ However, when we read computer science papers in order to better understand what machine learning means for societies, we are bringing to bear a form of reading that is not primarily about excavating the hidden meaning of a text, the models or techniques it describes, nor even exposing underlying truths about science. Not strictly reading *to make sense* or discern the definitive *meaning* of computer science texts, we find ourselves reading in order to engage with the *sense-making* and *meaning making* that significantly exceeds the text as such, connecting with ethical and political questions. Though many of the signature concepts of machine learning – features, gradients, functions, weights, representations, and so on – are introduced into the world in the types of papers discussed by Hinton and LeCun, in fact reading computer science involves engaging with a multiplicity of texts, from published papers and arXiv pre-prints, to podcasts, lectures, interviews, textbooks, and training manuals. As Peter

Galison writes on twentieth century experimental physicists, “some of the arguments [...] are contained in their published papers”, though “much remains unsaid in the published paper” and “significant features of the experimental conditions and procedures are omitted from the final articles” (1987: 4).

Our starting point in this essay is that researching machine learning involves reading the texts of computer science and, in the act of reading them, necessarily and inescapably engaging with the world-making that is taking place. Beyond the matter of *what* one reads, there is the question of *how* reading takes place and what kind of reading could offer ways to engage machine learning as it makes meaning in the world. What is it that we do when we read scientific papers on machine learning? What kinds of texts are we reading and what are the critical strategies that can be brought to bear? We are seeking a way of reading that avoids investing what Eve Kosofsky Sedgwick calls “faith in exposure,” wherein reading is imagined to uncover the real structures of power that are thought to underly the text itself. Of course, a degree of faith in the exposure of underlying power structures has been crucial to the critique of algorithmic worlds of surveillance capitalism, black box society, and automated systems (Zuboff, 2019; Pasquale, 2015). Notwithstanding the importance of existing critical readings, we propose a strategy for reading computer science that is attentive to the act of *reading* itself, that stays close to the difficulty involved in all forms of reading, and that works with the text as already properly belonging to the ethico-politics that this difficulty engenders.

We begin by mapping out what critical forms of reading might serve as exemplars for engaging with machine learning texts in computer science. We then discuss three specific cuts through the problem of reading computer science: *genre*, *readability*, and *meaning*, elucidating our reading strategy via specific passages drawn from texts. Throughout, we are inspired by philosophies of reading that advocate for engagement with particular kinds of passages from a text. Such passages are selected from a work not strictly because they are crucial to uncovering a definitive originary meaning, but because they contain moments when the author opens up an unresolved problem and signals the multiplicity and instability of meaning.² This kind of reading works against the foreclosures that address a narrow readership of science and industry practitioners, allowing a text to “land in unexpected places,”³ creating space for other readers not explicitly addressed, and for a wider society penetrated by the logics of machine learning. We draw on examples from passages from computer science texts, focusing on machine learning *textbooks* in the section on genre; on machine learning *papers* in computer science in the section on readability; and on *definitional concepts* in the section on meaning.

Science and strategies of reading

The study of science and technology is replete with strategies for reading the texts that scientists produce. Indeed, our emphasis on reading strategies for textbooks and scientific papers may appear retrograde to some readers from science and technology studies (STS) or the history of science. Thomas Kuhn, for instance, appeals to historians to treat scientific textbooks as critically as they would other sources, and to look beyond what he considered to be a deceptive image of linear progress, “disguising” the changes “in the aftermath of each scientific revolution” (1996: 137). In STS, published textbooks and articles are objects to be properly contextualised via ethnographic study of the actions of scientists, the living controversies in which they engage, before facts and theories are petrified in text, concealing their contested origins (Latour, 1987: 15).

Early works in these traditions tended to emphasise the strangeness and otherness of scientists and the deceptiveness of their texts, dramatizing the ethnographer’s *inability to make sense* of the papers that they encounter. In Latour and Woolgar’s (1979) *Laboratory Life*, for example, scientists are a “tribe” and the observer’s attempt to read, “to peruse some of their articles in order to ferret out possible reasons for their value” is thwarted, “it was all Chinese to him” and “complete gibberish” (1979: 75). One of the purposes of elaborating an alternative strategy of reading machine learning texts is to challenge such foundational formulations of reading as marking the mutual incomprehensibility of science, social science, and the humanities – and its founding upon particular gendered and racialised ideas about what and who can read and understand. In common with feminist accounts of science of Donna Haraway (1988) and Sandra Harding (1986), and N. Katherine Hayles’s longstanding work on textuality, reading, and computation (2005, 2021), we propose to read computer science as an engagement with the situated nature of all scientific ways of knowing.⁴ This is not a form of reading that treats scientists and those who encounter them as incompatible cultures, but instead regards the gaps and breaches between *any* author and reader as inescapable and an invitation to further ethico-political engagement.

STS scholars have developed many ways of addressing the notion of incomprehensibility between author and reader of scientific texts. For instance, in his book on the knowledge practices of machine learning, Adrian Mackenzie locates his study of machine learning algorithms within multiple textual worlds: “amid the books, documents, websites, software manuals and documentation, and a rather vast accumulation of scientific publications” (2017: xi). For Mackenzie, what is written in computer science thus forms part of an ethnographic situation where one gets closer to the knowledge practice as part of a broader effort to “reconfigure oneself as a machine

learner” (2017: xii). Getting close to the texts and practices of computer science, as Mackenzie’s and others’ influential ethnographic and socio-technical methods for studying algorithmic knowledge practices have done, is a necessary step for making sense of what is happening in machine learning and AI (Seaver, 2017; Dourish, 2016; Marres and Gerlitz, 2016). However, just as we do not wish to treat machine learning texts as incomprehensible, we also do not seek to *become* “machine learners” in quite the sense Mackenzie proposes. Rather, we attend to the gaps opened by reading texts for which we are not the intended audience, gaps that open the space for critical forms of reading. Although the texts of computer science significantly exceed the documented residue of a scientific practice to be excavated from the underlying text, the act of reading itself is worthy of some attention and development.

Machine learning texts are worthy of careful reading because they are a crucial archive of our contemporary condition (Thylstrup et al., 2021). Notwithstanding the hyperbole of the machine learning literature, its sense of its own cutting-edge historical importance, and often explicit engineering desire to make the world in its own image through prediction and classification, it also nurtures the kernels of major epistemic transformations in our world. Understandably, the manifest social and political harms involved in AI have created an acute critical injunction *to read in order to expose* the deeper structures at work – the mathematics, the code, the capital, the state power – within these texts. No one would wish to doubt the importance of such critical work. The reading strategy that we propose, however, draws inspiration from Sedgwick and others who challenge us to reflect on our habitual critical impulses to expose and unveil, asking instead different questions of our sources: “what does knowledge do” and “how does one move among its causes and effects?” (Sedgwick, 2003: 124). Sedgwick’s alternative mode of reading – perhaps misleadingly named “reparative reading”, for it does not entail reconciliation – “infuses critical projects” in a manner that relinquishes mastery of meaning in favour of the “experience of surprise” (146). The text is not a fixed object that conceals the truth but is a “phenomenon to be engaged”, formed through the act of reading with all its moments of surprise and future possibility.

To read computer science texts as phenomena in which we are engaged and implicated – even surprised sometimes – is also to address directly the foreclosures that would otherwise close down some of the critical entry points into machine learning technologies. Computer science texts impose a limit on logics of exposure by performing the idea that they are unable to be read by nontechnical specialists by virtue of their code, scale, mathematics, black box opacities, secret or proprietary models, or vast amounts of computing power. To engage these texts is to push on the moments of foreclosure, not to definitively

fix a meaning or diagnose a single underlying logic but rather to understand more precisely where and how these texts can surprise us. Machine learning techniques implicate us all, not least via our data and their penetration of our social world. The texts and concepts that make machine learning models intelligible draw on rich scientific and cultural histories. It is therefore critical to open them to a diversity of readers and viewpoints and crucially, at the point that their logics are still in experimental formation. It is by reading across this plurality of viewpoints that critical politics, ethics, and values can begin to be forged. Such a reading strategy implies that computer science texts are not only considered to be “about politics” or concerned “with society” when they address domains that are recognizably political and social – such as, for example, deep neural networks deployed in policing or in social welfare. In addition to these fraught moments, reading computer science texts potentially expands the scope for ethico-political engagement because reading itself confronts the difficulty of the open and contested meaning of the text, and the ways that concepts will move within the world and perhaps transform it.

There are philosophies of reading that can guide such ethical and political projects. Their insights do not treat the act of reading as merely or primarily cognitive or as “a matter of understanding what is said”, but as itself an ethical practice where “to live is to read, or rather to commit again and again to the failure to read which is the human lot” (Miller, 1987: 1, 59). This does not entail addressing computer science texts as literary or fictional but rather understanding their forms of difficulty, the demands they make on us as readers, and the accounts they ask us to give in reading. If we are to open ethical and political possibilities that are not reducible to fixing code or enhancing datasets, then the papers we are reading must surely also be sources not only of the historical record but of political questions of how we wish to live today. An ethics of reading takes as its object of study the act of reading as a fundamental engagement with ourselves and with the world. Reading computer science texts is not only a matter of discovering the meaning in an apparently technical paper but is significantly an engagement with the impossibility of reading as such. Impossibility here does not imply that the text is too impenetrable to be read, but more precisely that the text demands something of the reader that can only be responded to in the act of reading. It suggests that when one finds oneself thinking “this machine learning text is impossible” we ought to simultaneously consider that this is the case because reading is also “giving an account, telling a story, narrating” (Miller, 1987: 15). This is what it means to engage with a text as phenomena, to be attentive to the account that is given, the story that is narrated. As Thomas Keenan suggests in his ethico-political philosophy of reading, “by reading I mean our exposure to the singularity of a text, something that cannot

be organized in advance, whose complexities cannot be settled or decided” (1994: 1).

In the context of AI and machine learning, where the political application of technologies is so commonly to settle or predict intractably difficult problems in advance, we suggest that an ethics of reading serves to reopen the political difficulties of what cannot be settled and resolved. “Reading, if it happens at all”, writes philosopher Thomas Keenan, “happens only in the encounter with difficulty and without guarantees” (1994: 103). In Keenan’s philosophy of reading, there can be no predetermined pathway to a single meaning because reading will always “defy calculation in advance” and “refuse prediction” (1994: 102). Particularly resonant amid the algorithmic drive to predict and to calculate in advance, a critical reading strategy ought to engage with computer science texts as more lively, unpredictable and incalculable than their logics might presuppose. Put simply, in place of the injunction that reading computer science is too difficult/too technical/too different from literary and other forms of reading, we locate reading as precisely the injunction for why it is necessary and unavoidable to engage computer science through the encounter with difficulty. The act of reading implicates us because we cannot simply fall back on rules and codes to unlock the meaning of what is read; nor can one of our colleagues in computer science ever merely explain it to us in a way that entirely resolves these difficulties. The text “is read”, writes Derrida, but this act of reading is “not the site of a hermeneutic deciphering, the decoding of meaning and truth” (1988: 21). To read a computer science text ought never to be an exercise in deciphering a single output of meaning.

The different philosophies of reading proposed by Sedgwick, Miller, and Keenan offer entry points to an alternative strategy of reading that is potentially disruptive of a logic of resolving outputs of meaning. Indeed, there is a discernible method in their selection of passages of text, for example where Miller describes identifying “passages where the author reads himself” (sic) or moments when “an author turns back” on themselves (1987: 15). Such passages do not so much reveal or expose a hidden meaning as they give an account of the text’s instability and incompleteness. The selection of passages is more closely aligned to “a genealogy” of machine learning knowledge (where what is read contains “complete reversals”, and “jolts” and “surprises”) than it is to “a quest for origins” (where what is read “restores and unbroken continuity”) (Foucault, 1994b: 354–5). The passages from machine learning texts we discuss in the following sections have been selected not as a representative sample of canonical texts, but for their reversals and authorial breaches – or, as entry points into how apparently settled knowledge could have been otherwise. Despite their reputation for definitive landmark statements, computer science papers contain many passages where the author turns back and reflects

on the difficulty of a concept that may otherwise appear to be settled. To read passages in this way is also to shift the emphasis beyond a search for origins or the definition of concepts – what is a cluster?; what is a feature?; what is a loss function? – towards their multiplicity of meanings.

In the following sections, we address three distinct cuts through the problem of reading machine learning texts: genre, readability, and meaning. Though we propose that these three different cuts are co-present in many forms of computer science text, we treat them separately here in order to explicate the reading strategy for each case. In the first section, on genre, we focus on passages from François Chollet’s (2021) textbook on deep learning in order to map out a reading strategy that is attentive to the textual conventions of machine learning and, specifically, to how the genre deploys changes in register across image, text, and code. Second, on readability, we discuss passages from two computer science papers – Hinton’s (2014) paper on features and Bengio’s (2012) paper on deep learning of representations – where influential machine learning concepts are the fragile effects of a profoundly experimental process. Finally, on meaning, we select passages drawn from definitional work on a specific machine learning concept – “dropout” – to foreground how the work of conceptualization makes meaning and organises machine learning’s picture of the world. Together, the three cuts afford a critical reading of machine learning that disrupts linear narratives of advances in AI and opens onto the fraught difficulty of the world-making that is taking place.

Reading problem 1: “but what is a computer science text?” – on genre

It appears to be a curious juxtaposition: a strategy of reading drawn from philosophy and literary theory with a popular textbook introducing the field of deep learning with Python language. At first blush, the machine learning textbook may appear to be readily recognisable as self-explanatory or alternatively as a misleading or simplified picture of computer science. We propose that, on the contrary, the machine learning textbook significantly organises and builds a machine learning world view. More than this, machine learning textbooks embody an identifiable *genre* – they deploy a set of conventions that make them recognisable and intelligible as texts, and this genre matters to how they are read.⁵ In what they call “genre flailing”, Lauren Berlant understands genre broadly as the attempt “to read with things”, “to control the object enough to say something about it” whilst simultaneously “to change it enough that it comes to organize surprising kinds of exemplary association” (2018, 156; 161). Understood as a genre in Berlant’s terms, machine learning textbooks are objects with a distinctive type of composition that we wish to say

something about, while retaining our own differences from their imagined readers. Looked at slightly askew, they are lively literature with a capacity to surprise us as they build new exemplary associations.

The machine learning textbook has a distinctive genre that assembles a collage of formally written text, abstract concepts, mathematical formulae, graphs, images, different coding languages, diagrams, and so on. At one level, the difficulty of reading appears clear and obvious – how to hold the text together as a whole across the manifest differences in its composition. Very often, the textbooks demand not only to be read but technically performed, with coding exercises supplementing theoretical discussion. The reader is invited into the formation of the text so that they actively build novel technical senses to otherwise abstract ideas like “depth”, “representation”, or “layer”. This creative eclecticism and multiplicity is a defining feature of the genre of machine learning textbooks. In contrast to the notion that these textbooks present a misleading picture of science, our strategy for reading involves looking for the characteristic features of the genre and how they build a form of world view. Of key significance to the genre are the shifts in register and gaps that emerge across the heterogeneous elements of images, concepts, and code.

Consider, for example, François Chollet’s (2021) widely read textbook *Deep Learning with Python*. We select here a passage from Chapter 5 of this text titled “Fundamentals of Machine Learning.” In the spirit of our ethics of reading, this selection is not revelatory of a hidden core meaning or underlying ideology, even if the title itself seems to promise this. Rather, our reading interrupts the instructive and synthetic style of the textbook and invites the reader to forge the associative connections across the gaps. The textbook opens onto a substantial topic – “the nature of generalization in machine learning” – with a disarmingly simple coding exercise, using the MNIST dataset, which is composed of standardized, labelled images of handwritten digits, also pictured in the text. Unlike many elementary coding exercises, which commonly give readers a small sense of mastery over some operation, this one opens up a puzzle: if one shuffles the dataset’s labels (so that, for instance, the image of a written five is associated with the label “three”), it is still possible to fit a model on it: “deep learning models...can be trained to fit anything, as long as they have enough representational power” (Chollet, 2021, 127). Here the coding exercise undercuts the expectation that labels capture some real properties of the images reproduced on the page. Instead, the narrative of the story is the ubiquitous power of the deep learning model.

This disjunctive relationship between image (of the MNIST figure), explanatory text, and coding exercise gives rise to Chollet’s next question for the reader, “how come deep learning models generalize at all?” How is it possible that they learn some relevant way to identify

new instances, not contained in the training dataset. His answer begins with yet another change in register, with an invitation to “take a look at what’s really going on here” (2021: 128). This shift moves from the deep learning models that readers will be expecting to learn about, to “the structure of information in the real world” – a meta-physical territory. Referring once more to the image of handwritten MNIST digits, Chollet argues that the subspace of possible inputs is structured in a “continuous” way. To illustrate this claim, he relies on the reader’s ability to imagine transformations of the images reproduced on the adjoining page: “if you take a sample and modify it a little, it will still be recognizable as the same handwritten digit” (128). His broader point is that machine learning works by interpolation – “the source of generalization” – in structured, lower dimensional manifolds of possible input spaces (129). In order to understand how these models really work, the reader is invited to speculate about the informational structure of things themselves.

The heterogenous collage and shifting registers of machine learning textbooks make reading to decipher a single intended meaning complicated, to say the least. There is an understandable critical impulse to try to stabilise the meaning of each term in order to reveal a deeper incompatibility between them: “this thing is not really like that other thing”. To do this, however, risks missing the ways that these analogies constitute powerful claims with traction on the world (Lury and Day, 2019). So, to suspend the impulse to stabilize, limit, or neutralise meaning is an important methodological strategy in our reading of computer science texts. In place of “what does it really mean?”, our reading looks at the fragmentary, eclectic moments of these texts to open onto other questions: “what else could it mean?”; “how does it produce certain effects?”; “how might the address land differently for other readers or in other contexts?” In other words, regardless of the ultimate metaphysical structure of information in the universe, it is significant that machine learning textbooks make it possible for their readers to see concrete problems (in our case, images of digits) in terms of abstract lower-dimensional manifolds.

In Chollet’s account of the nature of generalisation in machine learning, then, there is also at work a significant ordering and building a machine learning world view. The conventions of the genre, as we have described, invite the reader to experience how and why a model “can be trained to fit anything”, significantly laying the groundwork for machine learning’s promise to offer generalisable solutions to singular problems in the world. Understood in this way, one important characteristic of the genre is to deploy the structure and analogies of the fable. The fable is a story that offers a general narrative that nevertheless invites the reader to locate a moral message that is singular to them and their lives.⁶ To read the fables in computer science textbooks is not to say that

the examples are “fictions”, deceptive, or illusory. On the contrary, fables are powerful ways to make meaning in the world. Computer science textbooks fabulate in the sense that they use recognisable general characters – MNIST figures, images from the ImageNet dataset, illustrations of backpropagation – to invite the reader to instantiate the example in their own models for their own specific problems. To be attentive to the fables of machine learning texts is also to consider how government and industry authorities absorb the promise of the *general* story of machine learning’s power to “fit anything”, and connect their *particular* political problem, from welfare systems to pandemic models. As Keenan describes the politics of fable, it is “an exemplary ethico-political mode” aimed at “producing and securing” a moral message (1994: 47).

Like all reading strategies, ours involves risks. We may appear to take machine learning textbooks at their word, or even let the architects of AI and deep learning systems “off the hook” for their very real harms. Far from abandoning critical capacity and political traction, we suggest that a certain receptiveness to the genre of a machine learning textbook actually expands the scope for critical reading and intervention. Such a reading is attuned to how the text is engaged, and to the gathering of fragments or written text, equations, images and code, so that the practice of reading allows for speculation on how the text could be read differently. Refusing the definitional distillation of what is really meant by generalization, prediction, representation, weight, or parameter, we are proposing to read “between” these terms, both in their connections with each other and in their worldly operationalization in terms of data, code, model, and so forth. Together, they ask readers to think about and even problematize the world in these terms and in so doing they open onto a range of effects not fully captured by their definition. An understanding of the genre of machine learning textbooks alerts us to the ways that algorithmic systems make demands that society will envision and enact how their work (from concepts to models to technologies) will be used in the world. We are concerned to be better equipped to recognise the ethico-politics and power of these machine learning texts so that there might also be points of opening and indeterminacy where one might imagine different pathways or formulations of problems that are not yet foreclosed.

Reading problem 2: “but do you understand it all?” – on readability

In our research group’s discussions on the value of reading computer science there has been a dilemma that we have sought to interrogate. On the one hand, we experience being inspired and challenged by reading machine learning papers, not least because we are curious about their concepts and their effects on the world. In this sense, reading

computer science involves encounters with some familiar ideas in unfamiliar textual settings, and we engage it in ways that productively interrupt our own habits of reading and critique. On the other hand, though, among the most common questions posed to us by computer scientists are those concerning what can be read and understood: “but do you understand it?”, “are you able to read it?”, “do you know code/math?”, “isn’t the algorithm blackboxed, vast, and unavailable to be read?”.⁷ This is of course a problem with any specialized or technical literature. How, then, to reconcile the engagement with computer science papers as generative difficulty with the question of whether computer science is readable and intelligible in particular ways? Put differently, do computer science papers pose special, even characteristic, problems for readability that one would not encounter with other forms of text?

In contrast to the notion that in order to be readable a text must be available to us, its meaning fully transparent, we propose a different approach to the reading of computer science papers. The published texts of computer science are among the many sites where machine learning algorithms and models are giving accounts of themselves.⁸ Beginning from the position that computer science papers can be read (are available to be engaged via reading) and are unreadable (not reducible to the semantic transference of intended meaning), what interests us is how often computer science papers seem to acknowledge this difficulty, sometimes in quite explicit ways.

Hinton’s (2014) paper ‘Where do features come from?’ addresses how Boltzmann machines learn layers of features. Though the paper is certainly not the most significant of Hinton’s papers in terms of citations or influence, it is fruitful for understanding how to read the genealogy of an idea in computer science. Hinton describes a series of reversals in the paper in the sense that he reflects on something that has subsequently appeared as a settled question in machine learning: how distributed representations are learned. The paper reflects on the genealogy of the Boltzmann machine and offers an explicit “speculation on the future of neural network models” (2014: 1095). It is an example of a fork in the road of machine learning knowledge, where there is uncertainty on the route of the past pathway and a certain doubtfulness regarding how to build better models in the future. As readers, we are engaging the phenomenon, speculating on the branching pathways in the formulation of a machine learning practice, just as the author also engages speculatively with potential directions their science could take. There is a certain disruption here in the idea that the science holds the knowledge necessary to unlock meaning about which social science or philosophy can “merely speculate”.

Hinton’s paper contains an important technical account of the backpropagation of error derivatives, and even in the technical detail he describes “a curious twist of fate” where the search for a “way of training large Boltzmann

machines” accidentally “ended up with a method for making backpropagation work much better in deep feed forward neural networks” (2014: 1092). One can read the reversals and surprises of machine learning papers, then, even amid the most technical of descriptions. Moreover, Hinton reflects on the multiple possible pathways that could have been taken, and he gives an account that is inescapably normative in its pursuit of a “better” approach to generalising features. Some of the most powerful assumptions of contemporary machine learning – that generalisation is desirable, that algorithms should learn from unlabelled data, that hidden layers should distil what matters – are available to be read in the proliferation of normative assumptions in computer science papers. For example, Hinton seeks a way of “overcoming the need for labelled data” and becoming “better at recovering the true causes in the layer above” (1091), describing “a very good way to initialize the weights” (1096). These passages in the text function in a similar way to Keenan’s and Miller’s selected passages where the author reflects on their use of terms, the potential rejected alternatives, and the decisions they have taken. After all, deciding how to initialize weights is also deciding what should be afforded greater weight, what matters, which parameters count. In short, reading such passages for their explicit accounts of making better models is also a matter of reading for world-making. Reading these passages momentarily affords an appreciation of the rejected pathways that were not taken, how those that were not “better” fell away from view. It is precisely these contingent decisions and residual alternative pathways that become entirely absent once an experimental machine learning system has become an “automated decision system” in use in the world.

Even where Hinton’s text interrupts the written narrative account with mathematical equations, or via the ubiquitous flow diagrams of inputs, hidden layers, and outputs, these nonetheless contribute to giving an account of the work that is necessary for computer science to make something legible even on its own terms. In a sense, the equation and the flow diagram do end the discussion (at least cease the narrative account) about the proper meaning of concepts. But mathematical representations such as those in Hinton’s paper are also fundamental to how machine learning is shaping the wider world. Indeed, one of the most interesting aspects of deep learning is precisely the way it connects a series of mathematical *operations* (matrices, gradients, and so on) to ethical and political *problems* in the world. Reading the equations in Hinton’s paper does not provide full and unequivocal access to a definitive meaning of “features” and their role in machine learning practice. Quite the contrary, the reading opens onto the uncertain pathways and contingent decisions that led to a specific, situated, and partial account of feature discovery in machine learning. It is just such contingent forks in the path that offer entry points to a critical reading of

machine learning texts. We propose that such a reading does not curtail nor limit the scope for holding to account the onward harms of a machine learning model. On the contrary, it makes it possible to read a text not as a fossilized residue but in its fullest potentiality and process of becoming, and thus to multiply the sites where machine learning's effects are at work. To speculate with and on a text is thus also to read the potential that it harbours and the possible political futures it imagines.⁹

In passages drawn from a second machine learning paper – Bengio's (2012) 'Deep learning of representations for unsupervised and transfer learning' – we locate a similarly reflexive account of the experimental pathways of machine learning. In common with Hinton's paper, Bengio's text is significant to our reading strategy because it mobilizes a set of ideas and concepts that some social scientists and humanities scholars would say they had spent their careers reading and thinking about, such as depth, generalization, representation, and composition. There are passages in Bengio's text where the language and the questions appear curiously familiar to the reader, or somewhat like those of philosophy or social theory. For example, where he grapples with the problem of representation, he poses the questions to himself and to the reader: "what can a good representation buy us?", "what is a good representation?", and "what training principles might be used to discover good representations?". In this pursuit of the "good" – and in common with very many other computer science papers – Bengio makes an explicitly normative commitment to the "good model" and how the good model represents the world. For instance, Bengio aligns a good deep learning model with notions of generalizability to new problems. A deep learning model, in Bengio's account, is able to "take advantage of out-of-distribution training examples" in order to make predictions on examples that are not from the same distribution as the training distribution (2012: 30). For Bengio, deep learning is "well suited to transfer learning because it focuses on learning representations and in particular 'abstract' representations" (2012: 30).

What is the significance of reading Bengio's paper for its normative exposition of the good model? In the context of longstanding social science knowledge of the historical alignment of statistical ideas of standard distribution (normal curves) and the good society, machine learning ideas about out-of-distribution learning reconfigure the relationship between the model and social norm. Indeed, it is by virtue of deep learning's promise to "make subsequent learning tasks easier" that it so voraciously enters new social and political domains, its "transfer learning" signalling a transferability to multiple policy problems (Goodfellow et al., 2016: 527). In short, reading Bengio's paper in its own terms – as an exposition of the potential form of the good model, written at a critical moment in the history of deep learning – is not so much concerned

with readability as such, but rather with how it gives an account of making a set of ideas comprehensible, and how this comprehensibility has effects on the world. It does matter how something like a feature or a representation is rendered comprehensible as an idea in computer science. After all, a feature is always also a thing of interest or concern, something discoverable in a scene or dataset that comes to attention, and a representation is a means of making something present in the world. These are powerful ideas in computer science, just as they are also fundamental to the ethico-politics of who or what can be represented or of interest. To read computer science is also to intervene to reinstate some of the profound contingencies that actually dwelled within the making of a deep learning model – how it became understood as a "good model" before it became something active within society.

Reading problem 3: "but it doesn't mean the same thing": On meaning

When discussing key concepts in machine learning with computer scientists, a strikingly common refrain is that a particular concept – function, bias, cluster, error, rule – simply "does not mean the same thing" as it might be interpreted to mean in social science or the humanities, or in everyday language. This problem of difference in meaning tends to foreclose reading and engagement on the basis that the terms, vocabularies, and language are fundamentally incompatible. Thus, for example, it may be said that what a social scientist may mean by "bias" (e.g., prejudice, imbalance, unfairness) is not equivalent to what a computer scientist may mean when they use the concept "bias" (e.g., prediction errors, the distance between actual and predicted values, or a term added to the weight matrices in the hidden layer). A whole array of otherwise familiar concepts such as rule, class, generalization, or value become estranged and unfamiliar because they are thought to hold unique and specific technical meanings in the world of machine learning. Put simply, what we have come to call the "it does not mean the same thing" problem asserts an impediment to reading on the basis that the author and the reader may not share the common vocabulary necessary to decode the meaning of the text.

The "it does not mean the same thing" problem is of concern to us because it lends to machine learning concepts a clarity and coherence of meaning that is not matched by the computer science texts we engage. From the perspective of an ethics of reading, no concept could ever mean the same thing since it never fully means the same thing to itself, never fully coincides with its own meaning. Reading machine learning's key concepts seems to us to involve not strictly a deciphering of settled concepts, but rather engagement with an active process of invention, experimentation, and conceptualization. It does not strike

us that reading is difficult because we lack the correct conceptual keys to unlock the accurate meaning, but because reading always involves the difficulty of the shifting and slipping transformations of a concept in its exposition.

The question we pose is whether it is possible to locate ways of reading machine learning concepts that insist upon the slippages in language and vocabularies as forces that are at work within a text. Instead of matching a technical term to a definition, we want to understand how machine learning concepts render the world intelligible in certain ways, according to certain perspectives. That is to say, the differences and gaps between concepts are not so much a lack of understanding to be overcome (an opacity to be rendered transparent) as they are openings onto the plural processes of conceptualization already present in the text. “A concept is never given”, writes Adi Ophir, but is “performed or played in the act of conceptualization” (2012: 1). For Ophir, it is in the multiple acts of conceptualization – the discursive framing, the connection to other concepts, the claims to usefulness – that a concept is invented and discovered, so a concept involves “presenting a question, sharing it with others, looking for an answer” (2012: 7). It is precisely this active work of conceptualization – proposing questions to others, sharing the associated experiments, reaching towards potential answers – that animates the computer science literature on machine learning. If one reads the signature concepts of machine learning, attentive to the questions posed and shared, the plurality of meaning is not an obstacle to understanding but precisely the object of study. Such a reading involves sustaining uncertainty as to what a concept might mean and resisting the temptation to foreclose the differences that exist. Indeed, it is in the plural acts of conceptualization that computer science texts *become political texts* in the sense that they decide what is at stake in the parameters of a problem. “The request to explicate a term”, writes Ophir, “to explain ‘what is x?’ is what is responsible for its appearing as a concept” (2012: 6).

In the demand for explication of a term, there is something rendered at stake politically, and there can be no pre-programmed set of instructions on which to rely for clarification. Concepts appear in the world – in political discourse, in science, in policy – because the question of what something is cannot be answered with a set of codes or instructions. To read computer science concepts is thus also to engage in this discussion, to make and receive a demand for exposition without transparent explanation. Computer science texts are replete with “what is x?” type questions. Though one may read these primarily as definitional problems, our reading strategy loosens the grip of a search for definitions in order to open space for machine learning concepts that significantly organise the field’s picture of the world. Let us explore what happens when we read some of computer science’s defining concepts as processes of invention, explication, and discovery.

In the passages we select here, the concept has become a signature concept for machine learning, but it has also necessarily required discussion and invention in relation to “what is x?” type provocations.

Consider, for example, the concept “dropout” and how it is proposed in machine learning as a “simple way to prevent neural networks from overfitting” (Srivastava et al., 2014). The concept has become crucial to contemporary machine learning and is a significant element of the genealogy of current large language models (LLMs). Understood strictly as a machine learning term, a regularization technique, dropout refers to training a neural network by randomly dropping out or ignoring certain units or neurons (and their connections). To read dropout as a concept, however, one would need to consider how it is placed at the centre of discussion and explication, and what account is given. As a technique for “reducing overfitting” of neural networks to their training data, the computer scientists explain that “training a network with dropout” results in “significantly lower generalization error on a wide variety of classification problems compared to training with other regularization methods” (2014: 1931). In common with most computer science texts, the invention of the concept happens via a set of claims about the algorithm’s performance, calibrated against benchmark data sets and across key application domains such as object classification and speech recognition. For example, the performance of the algorithm on the MNIST database with “dropout refining” is reported as a 0.79% classification error, compared to 1.6% error for a standard backpropagation algorithm. The conceptualization of dropout takes place in relation to the readily recognisable datasets of ImageNet and MNIST and already established problem domains such as image and object recognition. Working back from the benchmarks of performance, dropout is explained through descriptions of experimentation, approximation, and intuitive adjustment of probability that are said to account for the algorithm’s performance.

The conceptualisation of dropout does much more than define a term; it also significantly organises machine learning’s picture of the world. It is presented as an experimental process that more efficiently “approximates” the same effect as “combining many different neural network architectures” (Srivastava et al., 2014: 1936). The overfitting to training data observed in standard neural nets is described as the building up of “brittle co-adaptations” that “work for the training data but do not generalize to unseen data” (1931). The process of random dropout “breaks up” the brittleness of the neural net by introducing uncertainty and “adding noise” into the “presence of any particular hidden unit” (1932). These are accounts of a concept that we find substantially shape the world view of the machine learning algorithm. In effect, the dropout concept foregrounds unknowing by not considering some neurons, selected at random. One could begin to read the accounts

to decipher what computer scientists *mean* by dropout, but in the act of reading one follows the *meaning making* that takes place through the dropout concept. A desirable world from the perspective of machine learning, for example, is one in which algorithms are able to generalise to unseen data, avoid overfitting to what they know or, put differently, to incorporate and profit from the introduction of noise, absence, and uncertainty.

The conceptualization of dropout not only draws in other existing machine learning concepts such as backpropagation and generalization, but it also borrows from the concepts and worldviews constituted in other fields. For example, the computer scientists explain how their work on dropout was in part inspired by the biological sciences, suggesting that “a motivation for dropout comes from a theory of the role of sex in evolution” (2014: 1932). The biological concepts of selection advantage, randomness, and evolutionary fitness are drawn into relation with dropout: “each hidden unit must learn to work with a randomly chosen sample of other units. This should make each hidden unit more robust and drive it towards creating useful features on its own” (p 1932). Analogies with concepts from other scientific fields – once again, a feature of the genre – is a common way for machine learning texts to actively incorporate a conceptual multiplicity that masquerades as the settled unity of a defined term. The concepts we read in machine learning are thus not uniquely alien to readers outside computer science, but actually alien and different from themselves (Parisi, 2019). So, of course, the computer science concept does not mean the same thing as ostensibly the same concept used in social science or biology or philosophy, but it also does not strictly coincide with itself. It is resonant with multiple other meanings, some of them purposively enrolled (e.g., “fitness” from biology or “spiking” from neuroscience) and others overflowing, slipping and sliding into new potential uses.

What are some of the ethico-political stakes of revisiting a concept like dropout, of suspending the impulse to close it down to a definition, and reading it for its actually existing multiplicity, contingency, and uncertainty? Other work has signalled the political power of the semantic slippages in computer science concepts, where ideas like “rule” and “function” begin to overflow a machine learning worldview and enter a wider episteme and political project (Amoore, 2023). The reading strategies we are advocating align with this work because they refuse the deciphering of a settled concept, to simply dismiss all uses of analogical reasoning, and instead take as their starting point the pluripotentiality of all concepts. To read a computer science concept is to be alert to its malleability and mobility and how it might shape new political paradigms. For example, the concept “generalization” in machine learning alters conventional statistical notions of generalizing observations from samples of a dataset or a population. In a path-breaking paper on “rethinking generalization”, the computer

scientists set out to interpret their “experimental findings” by “comparison with traditional models” (Zhang et al., 2017: 1). Their experiments with fitting deep neural networks to random labels lead them to claim that “the effective capacity of several successful neural network architectures is large enough to shatter the training data” (2017: 9). A concept of generalization is advanced here that stretches what is “general” to potentially encompass everything or, as the authors put it “convolutional neural networks can fit random noise” (2017: 2). To be clear, it is not that the reinvention of a concept in machine learning – generalization – is then set free to reshape societies, economies, or politics in its own image. Certainly, the concept of generalization in these texts is not at all the same thing as claiming that an algorithmic solution can be applied *in general* to any new problem or domain, shattering the specificity of the training data, so to speak. But the renewed concept of generalization does achieve a degree of slippage into the claim that a model can break free of all context, becoming generally useful across multiple domains of life.

The normative propositions about how a model should work – shaped via concepts such as dropout and generalization – have an epistemic reach that far exceeds the use of the specific algorithm itself. How very swiftly the computer science introduction of randomness suggests itself as a way of addressing difficult political and social problems in general. Reading concepts as openings of potential rather than closed definitions multiplies the possible sites for ethico-political intervention, keeping space for critique even long after a specific algorithmic architecture has been discredited or superseded by something new. Our mode of reading – opening onto difficulty and plurality – does not stand in the way of direct critique of actually occurring deployments of algorithmic systems in society – Palantir at the border or in the UK’s NHS, for example – but redoubles and expands accountability into the invention of concepts and worlds. Understood in this way, the collaboration of corporate tech such as DeepMind and Google Brain’s collaboration on the very concept of generalization (Zhang et al., 2017), for example, can be held accountable beyond the deployed algorithmic system and into the realm of knowledge and sense making.

Conclusions: against disambiguation

In comments following his 2018 Turing lecture, computer scientist Geoffrey Hinton reflects on the most significant challenges facing machine learning research. It is the capacity of a deep learning model “to disambiguate” meaning, Hinton suggests, that is one of the key tests for contemporary computer science. Discussing the future of natural language models and machine vision, he describes how algorithms will be judged on their ability to disambiguate and infer meaning to ever more finite specific examples. It is in the very etymology of “disambiguation” – to rid

something of ambiguity, to establish a clear meaning – that machine learning locates a virtue in a single incontrovertible reading.

In this essay, we have proposed a strategy for reading computer science texts that are intended to work against the grain of the logic of disambiguation and offer new entry points for critique. The logic of a single and disambiguated meaning substantially underpins the allure of machine learning models. When algorithmic systems for object recognition or speech generation go out into the world, the claim to a single unambiguous output that can be “read off”, even parsed, is part of what makes machine learning so seductive to those addressing complex social worlds. Our reading strategy does not call for clearer or less ambiguous meanings in ML texts. On the contrary, we are concerned with seeking out the multiplicity and instability that remains lodged in every signature concept of today’s computer science. The published texts of computer science are among the many sites where machine learning algorithms and models are actively giving accounts of themselves. Where machine learning builds a logic of single definitive outputs of meaning, our reading strategy is alert to the many approximations, reversals, narratives, norms, and assumptions that are always present in the accounts of building a model.

In part, then, the critical motivation for reading computer science texts differently is a practical concern with how to engage the technologies that are so fundamentally reconfiguring our world. Much of the critical engagement with machine learning systems’ impact on the world has done important work on the injustices and inequalities of sites and domains where technologies are deployed. There is a potential danger, though, that the critical voices from the social sciences and humanities are circumscribed as cleaning up the ethics issues after the event, responding to the harms of AI as they penetrate policy and society. Notwithstanding the significant harms of the disambiguated output as it is deployed in the world, we propose extending attention to the worldviews and normative commitments forged in the very textbooks and papers of the machine learning field of computer science.

The reading strategy we propose in this essay – inspired by the different philosophies of reading offered by Sedgwick, Miller, and Keenan – is intended as a direct response to a reading problem we have encountered throughout our research on machine learning and society. That is to say, we have encountered in different forms the claim that machine learning texts are not accessible to critical reading – they are blackboxed, opaque, proprietary, coded – or, that if they are readable then their concepts do not correspond to the meanings of their twined concepts in philosophy, social or political theory. It is precisely this opacity of meaning and absence of conceptual correspondence or shared vocabulary that forms our starting point for a critical reading of machine learning texts. There are no absolute

and disambiguated meanings on which we might rely to orient ourselves in a world of algorithms. To read machine learning texts is necessarily to engage the full weight of having no absolute points of reference. In the passages we have discussed, the absence of fixed points of reference has afforded a shift of focus, from questions such as “what does x mean?” or “how does x work?”, to questions of “how does the meaning of x emerge?” and “what work is done to make x (dropout, generalization, backpropagation and so on) work?”

We have explained the dimensions of our reading strategy through the use of passages selected from machine learning textbooks, published papers, and definitional concepts. At each step, our concern has been to outline a strategy that offers entry points into reading a diverse range of computer science literature – from the most famous papers on influential new algorithms to the more specialized discussions of definitional concepts. Much more than sites for excavating hidden meaning, we have proposed that the texts of computer science offer moments of reversal that can serve to disrupt linear narratives of advances and progress in AI. One way to think about the ethico-political significance of such reversals is that they are moments when the genealogy of machine learning could have been otherwise, or where the exposition of a concept (even one that later becomes pivotal to machine learning) addresses its multiplicity and the other things it could mean, other contexts it could address. Through our focus on the problems of genre, readability, and meaning we do not intend to delimit the sole parameters of a strategy, and we are hopeful that readers will find others. Where computer science papers seek “better approaches to generalizing features”, or machine learning textbooks construct fables for deep learning, they actively give accounts of the world that is being wrought and the meanings being made. To begin to read in this way is to identify similar moments where a machine learning worldview is being invented and discovered.

Acknowledgements

Our thanks to the editors and four anonymous reviewers for their generous critical engagement with the paper. The paper has benefited greatly from their comments and provocations. We are grateful to Volha Piotukh and Paul Langley for their feedback on early drafts of the paper. 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’.

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) received no financial support for the research, authorship, and/or publication of this article

ORCID iDs

Louise Amoore  <https://orcid.org/0000-0001-6728-8553>
 Alexander Campolo  <https://orcid.org/0000-0003-3159-4131>
 Benjamin Jacobsen  <https://orcid.org/0000-0002-6656-8892>
 Ludovico Rella  <https://orcid.org/0000-0001-5468-9526>

Notes

1. The use of “we” captures our group of four scholars with research backgrounds across the social sciences and humanities, specifically in human geography, sociology, and the philosophy and history of science and technology. Though our interdisciplinarity is in no sense intended to capture a general orientation to reading computer science, it does animate our collective account of what it means to read and to develop a strategy for reading computer science texts.
2. Our selection of texts reflects different cuts through the problem of meaning making in machine learning: genre, readability, and definitional concepts. Though the texts selected may appear to be pivotal interventions that definitively shaped machine learning paradigms, they are chosen in order to examine how even apparently path-defining texts contain instabilities that are worthy of critical attention. The selection of the “passage”, then, serves as an opening onto the traces of rejected alternative pathways, or how things could have been otherwise.
3. As Foucault depicts the significance of reading for a “potential reader” without definitive address, “the effects of the book might land in unexpected places and form shapes that I had never thought of” (1994a: 174).
4. Donna Haraway reflects on the limits of a strong deconstructive approach to “the truth claims of science”, warning of the limits of a feminism that says “they’re just texts anyway, so let the boys have them back” (1988, 578). She advocates for a feminist “successor science” that offers a “richer, better account of a world, in order to live in it well” (581).
5. We are grateful to Volha Piotukh for helpful discussion of the function of genre in translation studies.
6. As Deleuze writes, “there is no literature without fabulation” because literature must discover “beneath apparent persons, the power of an impersonal”. This impersonal being is, for Deleuze, “not a generality but a singularity: a man, a woman, a beast” (1998, 3).
7. An observation from our research team’s undertaking of machine learning training courses is that one imagines getting closer to being able to read the texts of computer science papers. In practice, what is most striking is that the use of fables and examples on such courses (e.g. building a simple model using a Tensorflow library) is an important means of sense-making. Once again the question of how to make sense of machine learning texts is subverted by the powerful forces of sense making that are at work.
8. The distinction between accountability and giving a partial and incomplete account is drawn from Judith Butler, for whom the partial account is an ethical demand, “the question of ethics emerges at the limits of our schemes of intelligibility, where

one is at the limits of what one knows and still under the demand to offer and receive recognition” (2003, 18). The concept of algorithms giving an account of themselves is developed in Amoore (2020).

9. Reflecting on whether the humanities are “outgunned and out-classed by capital”, Alexander Galloway proposes to “continue to pursue the very questions that technoscience has always bungled”, the questions of history, society, aesthetics and culture that are simultaneously crucial to a digital world and yet inadequately grasped through its methods (2021, 258).

References

- Amoore L (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.
- Amoore L (2023) Machine learning political orders. *Review of International Studies* 49(1): 20–36.
- Bengio Y (2012) Deep learning of representations for unsupervised and transfer learning. *JMLR: Workshop and Conference Proceedings* 27: 17–37.
- Berlant L (2018) Genre Flailing. *Capacious: Journal for Emerging Affect Inquiry* 1(2): 156–162.
- Chollet F (2021) *Deep Learning with Python*, 2nd edition Shelter Island, NY: Manning Publications.
- Deleuze G (1998) *Essays Critical and Clinical*. London: Verso.
- Derrida J (1988) *Limited Inc*. Evanston, IL: Northwestern University Press.
- Dourish P (2016) Algorithms and Their Others: Algorithmic Culture in Context. *Big Data & Society* 3(2). DOI: 10.1177/2053951716665128.
- Ford M (2018) *Architects of AI: The Truth About AI From the People Building It*. Birmingham: Packt Publishing.
- Foucault M (1994a) The Masked Philosopher. In: Rabinow P and Rose N (eds) *The Essential Foucault*. New York: The New Press, 174–179.
- Foucault M (1994b) Nietzsche, Genealogy, History. In: Rabinow P and Rose N (eds) *The Essential Foucault*. New York: The New Press, 351–369.
- Galison P (1987) *How Experiments End*. Chicago: University of Chicago Press.
- Galloway AR (2021) *Uncomputable: Play and Politics in the Long Digital Age*. London: Verso.
- Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning*. Cambridge, MA: MIT Press.
- Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.
- Harding S (1986) *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.
- Hayles NK (2005) *My Mother Was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press.
- Hayles NK (2021) *Postprint: Books and Becoming Computational*. New York: Columbia University Press.
- Hinton G (2014) Where do features come from? *Cognitive Science* 38(6): 1078–1101.
- Keenan T (1994) *Fables of Responsibility: Aberrations and Predicaments in Ethics and Politics*. Stanford, CA: Stanford University Press.

- Krizhevsky A, Sutskever I and Hinton G (2012) ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 2: 1097–1105.
- Kuhn T (1996) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.
- Latour B and Woolgar S (1979) *Laboratory Life: The Social Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.
- Lury C and Day S (2019) Algorithmic personalization as a mode of individuation. *Theory, Culture and Society* 36(2): 17–37.
- Mackenzie A (2017) *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press.
- Marres N and Gerlitz C (2016) Interface methods: Renegotiating relations between digital sociological research, STS and society. *The Sociological Review* 64(1): 21–46.
- Miller JH (1987) *The Ethics of Reading: Kant, De Man, Eliot, Trollope, James, and Benjamin*. New York: Columbia University Press.
- Ophir A (2012) Concept', in *Political Concepts: A Critical Lexicon*. Available at: <https://www.politicalconcepts.org/concept-adi-ophir/> (accessed 10 December 2022).
- Parisi L (2019) The Alien Subject of AI. *Subjectivity* 12: 27–48.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Seaver N (2017) Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. *Big Data and Society* 3(2).
- Sedgwick EK (2003) *Touching Feeling: Affect, Pedagogy, Performativity*. Durham, NC: Duke University Press.
- Srivastava N, Hinton G, Krizhevsky A, et al. (2014) Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15(56): 1929–1958.
- Thylstrup NB, Agostinho D, Ring A, D'Ignazio C and Veel K (eds) (2021) *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: MIT Press.
- Zhang C, Bengio S, Hardt M, et al. (2017) Understanding Deep Learning Requires Rethinking Generalization', *ArXiv:1611.03530v2[Cs]*.
- Zuboff S (2019) *The Age of Surveillance Capitalism*. London: Profile Books.