PRE-PRINT: Pedwell, Carolyn (in press). "The Problem of Algorithmic Common Sense Learning' in *Learning Under Algorithmic Conditions: A Compilation of Critical Memoranda*. Eds. M. Curringa, E. de Freitas, E. Dixon-Román and T. Webb. Cambridge: MIT press.

The problem of algorithmic common sense learning

Carolyn Pedwell, Lancaster University, UK

Keywords: Abduction, affect, AI, common sense, Cyc, Open Mind Common Sense

The 'common sense problem' within AI is longstanding; notable early failures include computing systems "suggesting boiling a kidney to cure an infection, and attributing reddish spots on a Chevrolet to a case of the measles" (Cantwell Smith, 2019, p. 37). Difficulties with common sense ways of knowing persist in the machine learning age: "A toddler can easily identify a dog, construct simple sentences and figure out how to use an iPad. Ask any single AI to perform all three tasks, and the algorithm – if not explicitly trained on all three – fails" (Fan, 2019, p. 86).

What both Good Old Fashioned AI (GOFAI) and machine learning are interpreted as lacking here is "the ability to reason intuitively about everyday situations and events, which requires background knowledge about how the physical and social world works" (Choi, 2022, p.139). For Aristotle, common sense entailed how the five senses (sight, hearing, touch, taste, smell) in human and other animals combine to allow perceptual discrimination of objects. Contemporary invocations of common sense foreground ordinary wisdom, shared rules of thumb, and elemental 'laws' (Hall and O'Shea, 2015), yet vestiges of Aristotle's formulation persist in how common sense is today associated with both 'the sensible' (what is logical or reasonable) and 'the sensable' (what is obvious or perceptible to the senses). When it comes to solving problems in novel conditions as humans do, common sense learning remains AI's holy grail.

In this memorandum, I consider how, in different ways, both 'first wave' and 'second wave' AI[1] generate computationally ordered modalities of common sense which struggle to address the sensory-cognitive entanglements and sedimented power relations that shape everyday knowledge production – with implications for (more-than-human) learning, perception, and discovery in digital media ecologies.[2]

Codifying Common Sense

GOFAI approached common sense primarily as "the capacity to make inferences on a body of object-led information" (Davis, 1990, p. 3). It focused on establishing "primitive and social formal rules that captured everyday knowledge" (Dreyfus, 1992, p. ix) and could be encoded into rules-based algorithms to enable inference-making in logic-based systems (e.g. McCarthy, 1959).

[1] AI literatures refer to various 'waves' of AI, with some studies claiming three or four distinct waves. For clarity, I refer to two key waves in AI research corresponding broadly to 'logic-based' and 'machine learning' systems, while acknowledging the diversity of techniques within these schematic approaches, as well as the overlaps and interconnections between them.

[2] Key ideas in this memorandum are expanded upon in Pedwell (In press).

GOFAI's quest to "codify common sense" through translating tacit human knowledge into machine-readable knowledge is perhaps best encapsulated by the Cyc project; launched by computer scientist Doug Lenat at the US-based Microelectronics and Computer Technology Corporation in 1984, Cyc's initial methodology involved encoding in machine-readable terms 99 per cent of a one-volume American desk encyclopaedia, followed by all of the common sense 'facts' (e.g. that an object can't be in two places at once) that the encyclopaedia's creators "presumed the reader already knew" (Dreyfus, 1992, p. 76). This design, the team hoped, would enable the system to infer further rules directly from ordinary language.

Cyc's learning architecture foregrounded analogical reasoning. Operationalizing the computer scientist Marvin Minsky's (1984) claim that humans "assimilate new information by finding similar things we already know about and recording the exceptions to that analogy", the Cyc team sought to "swap the problem of 'telling the system about *x*' for the problem of 'finding an already known *x* that's similar to *x*'" (Lenat et al, 1985, p. 66). In other words, Cyc would navigate new situations by drawing analogical links to what it already 'knows' – which would require that it be programmed with both a substantial base of knowledge and an organized body of reasoning methods.

Through mobilizing higher-order logic, Cyc began to abstract, generalize, and learn from its experience. It could, for example, "infer 'Garcia is wet' from the statement 'Garcia is finishing a marathon run', by employing its algorithmic rules that running a marathon entails high exertion, that people sweat at high levels of exertion, and that when something sweats it is wet" (Copeland, 2016, online). Framed as a forerunner to IBM's Watson supercomputer, which in 2011 famously beat two reigning human champions on *Jeopardy* (Havasi, 2014), Cyc influenced the emergence of other AI common sense reasoning projects including, most notably, Open Mind Common Sense, set up by Minsky, Push Singh, and Catherine Havasi at MIT Media Lab in 1999 (Havasi et al, 2014).

As time passed, however, logic-based AI projects like Cyc failed to achieve robust flexibility and intuitive sense-making – deficits linked not only to difficulties in scaling up from isolated "micro-worlds" (Davis, 1990) but also to GOAFI's account of "intelligence as a passive receiving of context-free facts into a structure of already stored data" (Dreyfus, 1992, p. 34). As the philosopher of science Hubert Dreyfus argued from the 1970s, human learning entails broad embodied "know-how" shaped by contextual particularities, unfolding moods, and sensory-motor skills. A key problem with symbolic processing AI's model of learning is that "one cannot substitute an extractable web of beliefs for the whole cloth of our concrete everyday experiences" (Dreyfus, 1992, p. 54).

Common sense, as such, is never merely a cognitive operation – affect, sensation, and embodied habituation are vital to its dynamics, to how common sense shapes and is shaped by situated modes of attention, perception, and interaction (Keeling, 2007). When CyCorp (Cyc's corporate owner since 1995) claims that Cyc's "human-like" cognitive skills mean that it "understands (represents fully) real world contextual nuance … like culture, emotions, time, space, beliefs and bias" (CyCorp, n.d.), the significance of the words "cognitive" and "represents" is crucial. What Cyc offers is a logic-based account of cognitive elements of "emotion" amenable to symbolic representation. Whatever cannot be made legible in precise machine-readable instructions, codes, and categories lies outside the scope of the system.

To "codify common sense" may thus be precisely to elide its affective elements, as immanent visceral and sensorial processes are reframed as schematic cognitive ones. Yet as computer science's struggles with formalizing tacit human knowledge emphasize, intuition may be powerful to learning precisely to the extent that it *remains* implicit, incoherent, and unarticulated. In turn, common sense's resistance to explication is what makes "the practices it teaches significantly inaccessible" (Shotwell, 2011: 13) – and therefore not easily (if ever) replicated by artificial means.

Abduction, affect and discovery

In associating common sense with inference-making via programmed knowledge, GOFAI built computing systems capable of inductive and deductive reasoning, but not what the philosopher Charles Sanders Peirce (1958) called "abductive reasoning" – the generation and evaluation of explanatory hypotheses amid uncertainty. By contrast, in its capacity to go beyond what it has been taught to find "hidden patterns" in its data, machine learning is said to specialize in abduction. Similar, it might be said, to how Peirce describes abduction as hypothetical thinking with "maybes" (Fortes, 2022: 2), generative AI constitutes an unfolding "speculative experiment" in which "data inputs and the algorithm mutually modify to optimize the output" (Amoore, 2020, p. 48).

Yet even state-of-the-art 'abductive' machine learning systems struggle with common sense ways of knowing. As the computer scientist Gary Marcus notes, deep learning programmes excel in "closed-end classification problems" with millions or billions of training examples to learn from, but do less well when data is limited, the test set differs significantly from the training examples, or the "space of examples is broad and filled with novelty" (Marcus, 2018, p. 15). Problems that have less to do with categorization and more with everyday sense-making are less successfully addressed by statistical approximation and recursive learning from examples.

Key AI actors thus recommend a hybrid 'solution' to AI's persistent common sense problem that integrates machine learning with aspects of GOFAI. Marcus (2018), for instance, calls for a "neuro-symbolic approach" that marries the perceptual classification of deep learning with the inference and abstraction of symbolic processing. In this vein, computer scientist Yejin Choi advocates a synthesis of the abductive capacities generative machine learning (such as that associated with ChatGPT-3) and language-based (as opposed to logic- or rules-based) formalisms – which, in her view, are "sufficiently expressive and robust enough to encompass the vast number of commonsense facts and rules about how the world works" (Choi, 2022, pp. 139-40). In other words, language based formalisms would supplement the large-scale, speculative pattern recognition of generative AI with neuro-symbolic training sets that aim to simulate tacit human learning.

Yet critics argue that hybrid neuro-symbolic approaches *still* neglect what is fundamental to human common sense: its imbrication with and dependence on situated socio-affective contexts and relations (Suchman, 2024). While computer scientists champion language-based formalisms claiming that "it is language, not logical forms through which humans acquire knowledge about the world" (Choi, 2022: 140), this shifts the site of learning from 'logic' to 'language' without robust means of addressing the implicit and sensory elements of common sense that exceed language. What both first and second wave AI lack, from this perspective, is a genuine ability to *learn affectively* – a capacity, that is, for tacit, experimental, sensory-oriented learning in which particular implicit qualities may never become explicit.

The 'learning' in machine learning refers to algorithmic recursion in which the outcomes of previous actions are taken as inputs for future action, which allows a given program to recognize new items not part of its original training data. While AI research has, since the 1990s, increasingly explored how affect may figure in synthetic learning and sense-making, common computational techniques such as sentiment analysis provide "statistical proxies for affective intensities [which can] displace, reference, meaning and comprehension" (Andrejevic, 2013, p. 54). Unable to generate modalities of common sense that synthesize deep entanglements of cognitive, sensory-motor, and socio-cultural processes, machine learning systems compensate by substituting probabilistic correlations for relational processes of affecting and being affected (Pedwell, 2023).

Current AI architectures may not, from this perspective, be as 'abductive' as they are commonly framed – particularly if we understand sensory-cognitive processes as central to abduction's capacity for speculative hypothesis generation and the creation of 'newness'. It is worth emphasizing, in this respect, that, although Peirce's understanding of abduction developed across his career, it retains an emphasis on embodied processes of learning premised on affective intensities and relations. Abduction, as such, is a multi-modal model of discovery in which what is most important is not necessarily the data itself but rather subjects' sensorial reactions: abduction is initiated by "the feeling of puzzlement and ends at the satisfaction of knowing" (Fortes, 2022, p. 5). As an intuitive (yet also logical) mode of inhabiting plausible possibilities, Peircean abduction thus explicitly departs from probability-based modes of inference, including machine learning systems which generate shifting common sense 'truths' as the outcome of "aggregate feeling tone[s]" (Andrejevic, 2013, p. 46).

Truth, ideology, and analogical reasoning

Common sense has been widely understood in AI research to involve physical, biological, psychological, or economic 'truths' about "how that world works" (Lenat et al, 1985; Choi, 2022), that must either be programmed into logic-based systems or immanently generated by training machine learning programs with relevant data sets. Yet as critical thinkers from Antonio Gramsci (1971) onwards have argued, common sense is far from impartial or neatly extractable from hegemonic worldviews; the question of "how the world works" is thus political and ideological as much as ontological and physical.

Take, for instance, Cyc's investment in analogical reasoning as a route to common sense learning. In practical terms, once the Cyc team had encoded 400 articles from their chosen encyclopaedia into the system, a group of research assistants was enlisted to enter "the final 99 per cent of the knowledge base". Each research assistant would "take an article, locate the already-represented similar article(s), and perform a machine-assisted "copy & edit" procedure to produce a machine-understandable version of the new article". An existing article on "Pewter" might, for instance, be copied and edited to populate a new entry on "Britannia-Metal" (Lenat et al, 1985, p. 77). Drawing analogies between types of metal may not seem particularly contentious, yet the ideological interests at stake in machine-enabled analogical reasoning slide into relief when we consider other founding terms within Cyc's knowledge base – such as "Ronald Reagan", "female animal", "abortion", "AIDS", "homosexuality" and "terrorist".

While the early Cyc literature touches briefly on what kind of "light" practical training the project's research assistants might require for their "copy & edit" task, there is scant discussion of what role the social positionalities, affective orientations, and ideological worldviews of these "knowledge enterers" – or indeed of the project's creators, designers, and funders – might play in the system's ongoing production of common sense.

MIT's Open Mind Common Sense Project (OMCS) also employed analogical reasoning in its learning schema, though its use of machine learning techniques and internet-sourced data distinguished its approach from Cyc's. OMCS combined a semantic network, ConceptNet, "built from a corpus of knowledge collected and rated by volunteers on the Internet", with a reasoning engine, AnalogySpace, which "used factor analysis, to build a space representing the large-scale patterns in common sense knowledge" (Havasi et al, 2014, pp. 24-25). In this distributed system, common sense 'truth' emerges, in part, as the shifting effect of popular ratings. From 2007, internet volunteers rated common sense statements from OMCS using the set "generally true", "sometimes true", "not true", "doesn't make sense", and "not true but amusing" (Havasi et al, 2014, p. 32) – data which was fed back into the system to refine and expand its knowledge base. Through OMCS's mediating lens, then, the validity of common sense claims become a recursive function of algorithmically calibrated 'popularity'.

While Gramsci frames popular knowledge, values, and rules of thumb as crucial sites of social learning and political struggle, popularity within machine learning systems is the unstable product of popular ratings and statistically adjudicated matters of fit that remain largely hidden from public view. Generative AI, in turn, increasingly incorporates "counterfactual reasoning", which computer scientists frame as "closely related to abductive reasoning" and thus necessary to engage a world "filled with previously unseen situations" (Choi 2022, p. 144). Yet if Peircean abduction thinks with "the possible" to generate "momentary truths" for the purpose of discovery (Fortes, 2022, p. 2), the statistically adjudicated modes of common sense produced by machine learning systems may limit rather than expand experimental possibilities for social life. This is the case not only given how computational common sense may embed social normativity at the levels of logic, procedure, and data, but also because deep learning's probabalistic mode of working with hypotheticals and inverse inference from a vast field of counterfactuals can *proceduralize* abduction – so that it tips more heavily towards regulation and away from creativity (de Freitas, 2022).

To work effectively with "inconsistency, subjectivity, and generally noisy data", OMCS, for instance, sought to make "rough conclusions" based on analogies, similarities, and tendencies rather than on an "assumption of absolute truth" (Havasi et al, 2014: 27). The system employed singular value decomposition (SVD), a technique in linear algebra, which compresses data by sharing information between items deemed similar to each other. This enabled "a method of commonsense inference called *cumulative analogy*" (ibid). If a general risk of analogical reasoning is that it flattens complex relations and elides affective, cultural and socio-political particularities, cumulative analogy intensifies this risk because it seeks "to moderate elements of the cultural field that may present themselves as typical or outstanding, so that they can be led to make sense relative to other, more even-keeled examples" (Hallinan and Striphas, 2016, p. 122) – dynamics indexing the wider historic links among statistics, machine learning, and the racialized, gendered, and classed constitution of "normalized standards of behaviour" (Amaro, 2022, p. 24).

'Learning', from this perspective, may reproduce a recursive loop of dominant cultural associations – or make probabilistic speculations based upon iterative biases and prejudices projected into the future. Human learning, of course, might be said to work similarly insofar as it involves analogical resoning and other epistemic shortcuts, yet what is at stake as machine learning architectures become increasingly 'environmental' is how probabalistic analogizing is operationalized on grand populist scales.

More-than-human futures

While AI research tracks back and forth between first wave symbolic processing and second wave deep learning techniques – both of which bely an enduring attachment to cognitivism – alternative frameworks for conceptualizing human-machine intelligence and sense-making remain dormant (Suchman, 2024, p. 88). What would (or even could) meaningful (processual, multi-dimensional) affective learning and abductive experimentation –attuned to the complex and shifting interactions among common sense, computation, and social relations of power – look like in intelligent systems?

Post-war genealogies of AI offer glimpses of different routes to common sense with the potential to decentre cognitivism and foreground affective attunement – from approaches associated with '4E' (embodied, embedded, enactive, extended) cognition of the 1980s to MIT's affective robotics of the 1990s. What may be vital to the future of expansive common sense learning and discovery in algorithmic conditions, however, is to avoid holding cognition and affect in rigid opposition. While normative AI imaginaries position cognitive 'chess-playing intelligence' and affective 'childlike learning' as oppositional, for instance, cognitive reasoning, as the feminist science studies scholar Elizabeth A. Wilson (2010) argues, "emerges out of circuits of physical, psychological, and aesthetic expertise" and is "more imbricated in networks of affectivity than has usually been supposed" (p. 5). In Peircean mathematics, moreover, abduction deeply entangles cognition, reason and affectivity – in a world in which "mathematical habits [are] embodied, situated, cultural, planetary and more than human" (de Freitas, 2022: 2).

Given that generative machine learning architectures give rise to distributed and relational capacities premised on human-algorithm entanglements, the imperative is not, in my view, to negate the possibility of synthetic common sense by reiterating that machines 'lack feelings' or that they 'are not made of the right kind of materials'. Rather, what is at stake now is how varied AI systems are (re)constituting both "the sensible" and "the sensable" in ways which return us to Aristotle's ancient philosophy of common sense.

The composite forms of perception, knowledge, and learning enabled by algorithmic architectures are transforming what can be made perceptible to the senses, as well as the range of entities which participate in such cognitive-sensorial agencies. While 'the human' cannot be excised from conversations concerning computational common sense, we do, as the media scholar Lisa Blackman contends, "need radically revised notions of body-world-consciousness" which are "compatible with twenty-first century media" (2019, p. 178). If, moreover, human experience is being re-mediated (experientially and affectively) in line with the affordances of algorithmic architectures (Pedwell, 2019), it may be that we are learning to conform to computational categories, expectations, and common senses as much as (or more than) they are learning to capture and/or simulate us.

Within these emergent more-than-human media ecologies, what might it mean to inhabit common sense as a site of socio-affective and "political struggle" (Gramsci, 1971)? How might dwelling experimentally within these unfinished histories of human-machine relations enable a different philosophy and politics of AI with more expansive abductive possibilities for (un)common sense, collaborative sensing, and affective learning?

References

Amoore, L. (2020). *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press.

Amaro, R. (2022). *The Black Technical Object: On Machine Learning and the Aspiration of Black Being*. Sternberg Press.

Andrejevic, M. (2013). *Infoglut: How Too Much Information is Changing the Way We Think and Know*. Routledge.

Blackman, L. (2019). *Haunted Data*: *Affect, Transmedia, Weird Science*. Bloomsbury Academic.

Cantwell Smith, B. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgement*. MIT Press.

Choi, Y. (2022). The Curious Case of Commonsense Intelligence. *Dædalus*, *151*(2) Spring: 139–155.

Copeland, B. J. (2016). Cyc. In *Britannica*, Retrieved January 10, 2023, from https://www.britannica.com/topic/Cyc

CyCorp. (n.d.). Cyc Technology Overview. CyCorp website. Retrieved January 10, 2023, from https://Cyc.com/wp-content/uploads/2021/04/Cyc-Technology-Overview.pdf

Davis, E. (1990). *Representations of Commonsense Knowledge*. Morgan Kaufmann.

de Freitas, E. (2022). The Role of Abduction in Mathematics: Creativity, Contingency, and Constraint. In L. Magani (Ed.), *Handbook of Abductive Cognition*, 1–24. Springer Nature.

Dreyfus, H. (1992). *What Computers Can't Do*: *A Critique of Artificial Reason*. HarperCollins.

Fan, S. (2019). *Will AI Replace Us? A Primer for the 21st Century*. Thames and Hudson.

Fortes, G. (2022). Abduction. In V.P. Glăveanu (Ed.), *The Palgrave Encyclopaedia of the* Possible. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-98390-5_44-1

Gramsci, A. (1971). *Selections From the Prison Notebooks* (Q. Hoare and G. Nowell Smith, Eds.). Lawrence and Wishart.

Hall, S. and O'Shea, A. (2015). Common-sense Neoliberalism. In S. Hall, D. Massey, and M. Rustin (Eds.), *After Neoliberalism? The Kilburn Manifesto*, 52–68. Lawrence and Wishart.

Hallinan, B, and Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society*, *18*(1), 117–137.

Havasi, C. (2014, August 9). Who's Doing Common-Sense Reasoning and Why it Matters. *TechCrunch*. https://techcrunch.com/2014/08/09/guide-to-common-sense-reasoning-whos-doing-it-and-why-it-matters/?guccounter=1

Havasi, C., Pustejovsky, J., Speer, R. and Lieberman, H. (2014). Digital Intuition: Applying Common Sense Using Dimensionality Reduction. *IEEE Intelligent Systems*, July/August, 24-35.

Keeling, K. (2007). *The Witch's Flight. The Cinematic, the Black Femme, and the Image of Common Sense*. Duke University Press.

Lenat, D., Prakash, M., and Shepard, M. (1985). Cyc: Using Common Sense Knowledge to Overcome Brittleness sand Knowledge Acquisition Bottlenecks. *The AI Magazine*, *6*(4), 65–85.

Marcus, G. (2018). Deep Learning: A Critical Appraisal. *arXiv preprint*: 1801.00631.

McCarthy, J. (1959). Programs with Common Sense. *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, Her Majesty's Stationary Office, pp. 75–91.

Minsky, M. (1984). Afterword. In Vinge, V., *True Names.* Bluejay Books.

Pedwell, C. (2019). Digital Tendencies: Intuition, Algorithmic Thought and New Social Movements. *Culture, Theory and Critique 60*(2), 123-138.

Pedwell, C. (2023). Intuition as a "trained thing": Sensing, Thinking and Speculating in Computational Cultures. *Subjectivity* 30, 348–372.

Pedwell, C. (In press). The Intuitive and the Counter-intuitive: AI and the Affective Ideologies of Common Sense. *New Formations*.

Peirce, C.S. (1958). *Collected Papers*. Harvard University Press.

Shotwell, A. (2011). *Knowing Otherwise: Race, Gender, and Implicit Understanding*. Pennsylvania State University Press.

Suchman, L. (2024). The Neural Network at its Limits. In R. Dhaliwal, T. Lepage-Richer and L. Suchman (Eds.), *Neural Networks* (pp. 87–112). University of Minnesota Press.

Wilson, E. (2010). *Affect and Artificial Intelligence*. University of Washington Press.