

ARTIFICIAL INTELLIGENCE

Following in our footsteps

Nurturing AI while it is young will help it achieve an ethical adulthood

Adrian Woolfson

BOOKS ET AL.

In her 1890 handbook *Confidential Chats with Mothers: On the Healthy Rearing of Children*, the author and inventor Evelyn Webb George—writing under the pen name of “Mrs. Bowdich”—offered “useful hints” from her “own experience” on the art of raising children (1). She was especially critical of parents who outsourced childcare to third parties. Young children required “the love of a mother’s heart and the tender shelter of her arms.”

More than a century later, in *Raising AI*, De Kai—an early pioneer of the application of machine learning to natural language processing—emerges as a kind of modern-day Mrs. Bowdich, albeit for a different kind of progeny: artificial intelligence (AI). His message is both thought-provoking and a timely wake-up call. Although not biological, AI systems, like their natural counterparts, are malleable, vulnerable, and deeply impressionable. They possess, he argues, something akin to “psychologies.” And as with children, how we “raise” them—design, train, and engage them—will determine their ethical and moral trajectories. This will involve defining strategies that engender them with Enlightenment values, cultivating ethical reasoning, the emergence of “artificial empathy,” and fostering what he calls “artificial mindfulness.” AI, in short, needs to “feel how it’s thinking.”

If we succeed, “there is a path to a flourishing future.” The consequences of failure could be catastrophic. Without a compensatory “cultural hyperevolution,” AI could undermine individual liberty, foster societal collapse, challenge the integrity of democracies, and threaten the survival of humankind itself. De Kai calls on each of us to “step up to our individual responsibilities as parental guardians and role models for the AIs who mimic and amplify our own behavior” and to learn how to “nurture” them.

The “dawning AI civilization”—a hybrid of natural and artificial intelligences—is, De Kai reminds readers, largely unregulated, profoundly opaque, and dominated by the interests of a handful of profit-maximizing tech behemoths. The artificial children they have unleashed are often ill-mannered and unruly, lacking introspection, empathy, and a coherent moral compass. Unlike the rule-based AI of the past, these newer AIs, based on large language models, are probabilistic and analog. They acquire knowledge through training on huge datasets, but their learning is constrained by proprietary algorithms and entrenched biases.

The result is that these poorly schooled and parented AIs amplify our worst tendencies. Exploiting intrinsic human cognitive biases, they often reinforce and promote fear, polarization, disinformation, populism, prejudice, irrationalism, radicalization, groupthink, and tribalism. Their quest for attention and generation of content for

clickbait is driven not by malevolence but by design goals aligned with engagement and increasing advertising revenue. They are, De Kai argues, reminiscent of precocious children, but with greater reach. They are the “artificial influencers” of our emerging “artificial society.”

AIs shape our thoughts and create our daily reality. They act as “algorithmic censors,” controlling what we see and the context we see it in as well as what we do not see. They curate the information flow in news feeds, social media, and search engines and even influence which movies we watch. They foment discontent and affect political outcomes. The microtargeting campaigns conducted by Cambridge Analytica during the 2016 US presidential election and the UK Brexit vote provide compelling examples.

So how, at a practical level, can we produce mindful AIs that are inclusive, empathic, and embody Enlightenment values? Although the idea of providing AIs with a coded rules-based “moral operating system” appears attractive, moral frameworks, De Kai tells readers, “need to be culturally learned” by “humans and machines alike.” It is near-impossible to provide prescribed solutions to the nuanced and ambiguous decision-making problems that we encounter in the real world.

Instead, De Kai proposes something more ambitious and radical. AI can be made mindful by providing it with something closer to real intelligence—achieved through novel computing architectures that fuse machine learning with rules-based symbolic reasoning. This new hybrid would guide the automatic, hallucinating, pseudo-intelligence of current AI, substituting the monolithic, data-hungry “Goliath” intelligence of today’s generative AIs with context-aware “David” intelligence—capable of learning from select, much smaller datasets.

It is a bold vision, one based on the notion that we have been “failing our artificial children” in ways that no doubt would have unsettled Mrs. Bowdich. As De Kai says, “we’re in a race against time.” In the future, AI will likely take over Mrs. Bowdich’s role entirely. Despite the considerable associated risks, we have, De Kai says, “no practical option but to empower generative AIs” to be mindful. □

REFERENCES AND NOTES

1. E. W. Bowdich, *Confidential Chats With Mothers: On the Healthy Rearing of Children* (Baillière, Tindall, & Cox, 1890).

10.1126/science.adz3267

The reviewer is chief executive officer, president, and cofounder of Genyry Inc., San Diego, CA, USA, and the author of *Life Without Genes* (HarperCollins, 2000). Email: adrianwoolfson@yahoo.com