

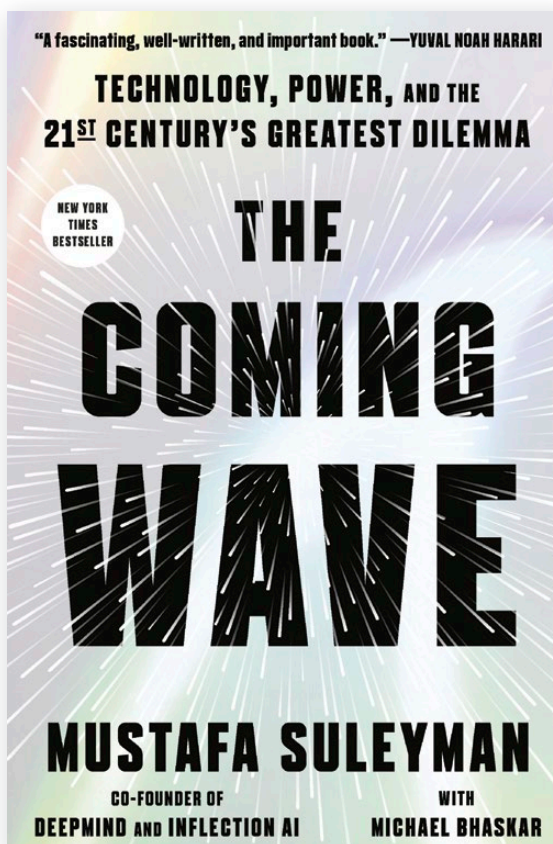
BOOKS

AI's Wave

CARL MITCHAM AND
LUKAS FUCHS

Although informative and bold—not to mention much endorsed and promoted—Mustafa Suleyman's new book, *The Coming Wave*, is ultimately unsatisfying. Suleyman, cofounder of the Google-acquired artificial intelligence company DeepMind and now CEO of Microsoft AI, wrote the book with assistance from technology journalist Michael Bhaskar. They attempt four interlocking tasks: to call out the existential threat of uncontained artificial intelligence, admonish readers not to ignore the dangers, situate the warning within a historical context of ever-increasing waves of techno-societal transformation, and make concrete policy proposals for achieving containment. The policy proposals are the most provocative and problematic aspect of the book.

The arc of Suleyman's argument is given by the titles of the initial, penultimate, and ultimate chapters: "Containment Is Not Possible," "Containment Must Be Possible," and "Ten Steps Toward Containment." Three-fourths of the book is dedicated to compelling arguments supporting the "not possible" thesis, which is nevertheless salted with "must be possible" counterpoints. With impassioned seriousness, Suleyman's rhetoric becomes an urgent plea to confront a unique threat. "If this book feels contradictory in its attitude toward technology, part positive and part foreboding, that's because such a contradictory view



The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma
by Mustafa Suleyman, with Michael Bhaskar. New York, NY: Crown, 2023, 352 pp.

is the most honest assessment of where we are." Suleyman might be likened to the concerned creators of *Bulletin of the Atomic Scientists* in 1945, nuclear engineers who feared the new weapons they'd created.

At least three arguments differentiate Suleyman's alarm from other jeremiads about AI. One is the way he places AI in the longer history of technological change by employing a popular science-technology-society boilerplate about how waves of agricultural, mechanical, chemical, and electrical innovations have challenged people to either catch and

ride them or be dragged under and swept away. Against this waveform of evolving techno-reality, he posits that any societal effort to restrict or somehow contain new technologies will be fighting the tide. Despite its emotional resonance, professional historians of technology would criticize the wave metaphor as simplistic.

Suleyman's core analytical contribution is to conceive artificial intelligence as an omni-use, hyperevolving technology that is transforming a broad spectrum of other technologies similarly to how electricity transformed manufacturing,

communication, urban life, and more. “Technologies of the coming wave are highly powerful, precisely because they are fundamentally general.” Deploying AI in chemical engineering and synthetic biology ups the ante on creating new materials and organisms, posing potential environmental and societal disruptions of an unprecedented speed and catastrophic magnitude.

Without denying the possible benefits to AI and synthetic biology, Suleyman simply argues that too much attention is given to benefits at the expense of risks and threats. He attributes this tendency to what he calls “pessimism aversion”: motivated reasoning makes humans too optimistic. In his telling, worry about the coming wave is warranted because of AI’s “on-demand utility that permeates and powers almost every aspect of daily life.” AI is being adopted and tested in a wide variety of contexts, propelling development, decreasing costs, and spreading use. The technology is hyperevolving (through fast, iterative learning processes), developing with increasing autonomy (with AI systems, according to Suleyman, “conducting their own R&D cycles”), and can have asymmetric impact (by design or by hacking). As Suleyman predicts, “Containing something like this is always going to be much harder than containing a constrained, single-task technology, stuck in a tiny niche with few dependencies.”

Low-cost, widespread adoption constitutes an especially critical threat. Global competitors, rogue nonstate actors, or millenarian fanatics now possess the tools—or will soon—to disrupt global infrastructure, challenge established power structures, and threaten public health. In the past, such challenges would demand a massive build-up of military weapons, industrial capacity, or social organization; with artificial intelligence and synthetic biology, disruption of the global order might come from a local lab or laptop computer.

Suleyman’s third distinctive contribution is an argument for containment as the necessary precondition for managing the oncoming AI wave. He is a resolute critic of fellow Silicon Valley technophiles under the spell of libertarian antistate sentiments. With an innovator’s can-do spirit, he outlines 10 concrete steps that could open the door to containment or prudent management. Suleyman rejects freeze-and-flight responses in favor of fight—or at least inventorying all possible tools that he can imagine to fight with.

His first recommendation is to “encourage, incentivize, and directly fund much more work” on safety engineering. “It’s time for an Apollo program on AI safety and biosafety.” Second, safety measures must be audited; such measures “will struggle to be effective if you can’t verify that they are working as intended.” Third, he wants to slow down AI development, perhaps with national export controls. “The wave can be slowed, at least for some period of time and in some areas,” he writes, and “buying time in an era of hyperevolution is invaluable.”

Fourth and fifth, Suleyman argues that critics need to become makers (“credible critics must become practitioners”), and corporations must integrate high purpose into the pursuit of profit. Critics too often “fall into the pessimism-aversion trap that is hardwired into techno/political/business elites.” Unwilling to recognize their own impotence, they have too much faith in “writing theoretical oversight frameworks or op-eds calling for regulation.” Suleyman presents himself as a model here. He recalls the emphasis he placed on factoring in ethics and safety alongside profit in founding DeepMind.

Proposals six and seven address the state. Democratic governments, he writes, must “get way more involved, back to building real technology, setting standards, and nurturing in-house capability.” States can better steer AI

toward the public interest if they are involved in creating it. Additionally, states should pursue moderating international agreements. “We need our generation’s equivalent of the nuclear treaty to shape a common worldwide approach ... setting limits and building frameworks for management and mitigation that, like the wave, cross borders.”

Proposals eight and nine shift to individuals. Specific policies must be generally supported by national and international technoscientific cultures—as Suleyman writes, they’ll need “real, gut-level buy-in from everyone involved in frontier technologies.” And the public must also be on board. Throughout this section, Suleyman discusses what “we” need to do. This “we” refers variously to the author and coauthor, AI researchers and entrepreneurs, scientists and engineers generally, the global West, or all humanity. “When people talk about technology—myself included—they often make an argument like the following. Because *we* build technology, *we* can fix the problems it creates. This is true in the broadest sense. But the problem is there is no functional ‘we’ here. Insofar as “the invocation of the grand ‘we’ is at present meaningless, it prompts an obvious follow-up: let’s build one.” Recommendation nine is to create social or “we” movements for containment.

Finally, the tenth step is “coherence, ensuring that each element works in harmony with the others, that containment is a virtuous circle of mutual reinforcing measures and not a gap-filled cacophony of competing programs.”

Despite Suleyman’s awareness of danger, sincere effort at a response, and appreciation of current fragilities in liberal democracy, there is something deeply naive and unrealistic about many of his proposals. Take the idea of an Apollo program for AI safety. Suleyman ignores the difference, as economist Richard R. Nelson once

framed it, between “the moon and the ghetto”—the difference between putting a man on the moon and raising people out of poverty. Apollo may have been a daunting problem, but an international AI safety program, even if funding were available, would be a wicked problem of the highest order.

Still, not all of his proposals are so crazy. Recent actions by both the European Parliament and the US Congress to regulate AI can be read as efforts to operationalize proposals six and seven. But can anyone genuinely imagine the European Union or United States as models for a global commonwealth? Is EU or US leadership sufficient to institute global rules? Is there a conceivable nation or multilateral body capable of detecting and preventing uncontained AI developments from posing global dangers? Is it possible that Suleyman is practicing the pessimism-aversion he otherwise warns against?

In an epilogue, Suleyman makes a final appeal: he presents a vision for technology as a beneficial, progressive force that the elusive “we” must “never lose sight of.” “Too many visions of the future start with what technology can or might do and work from there.” Instead, society should first imagine how technology can “amplify the best of us, open new pathways for creativity and cooperation.... It should make us happier and healthier, the ultimate complement to human endeavor and life well lived—but always on our terms, democratically decided, publicly debated, with benefits widely distributed.” Alas, this sounds like the kind of bland cliché that ChatGPT would write.

Carl Mitcham is professor emeritus of Humanities Arts & Social Sciences at Colorado School of Mines and author of *Steps toward a Philosophy of Engineering* (2020). *Lukas Fuchs* is a philosopher of political economy and a research fellow in the Philosophy and Ethics Group at Eindhoven University of Technology in the Netherlands.