

Durham Research Online

Deposited in DRO:

08 October 2019

Version of attached file:

Accepted Version

Peer-review status of attached file:

Not peer-reviewed

Citation for published item:

Halina, Marta and Martin, Joseph D. (2019) 'Five ways AI is not like the Manhattan project (and one way it is).', 3 Quarks Daily .

Further information on publisher's website:

<https://www.3quarksdaily.com/3quarksdaily/2019/08/five-ways-ai-is-not-like-the-manhattan-project-and-one-way-it-is.html>

Publisher's copyright statement:**Additional information:**

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Five Ways AI Is Not Like the Manhattan Project (and One Way It Is)

Joseph D. Martin and Marta Halina

Calls for a Manhattan Project–style crash effort to develop artificial intelligence (AI) technology are thick on the ground these days. Oren Etzioni, the CEO of the Allen Institute for Artificial Intelligence, recently issued such a call on [The Hill](#). The analogy is commonly used to [describe](#) DeepMind’s initiative to build artificial general intelligence (AGI). It is similarly used to describe military initiatives to build AGI. At a [conference last year](#), DARPA announced a \$2 billion investment in AI over the next five years. Ron Brachman, former director of DARPA’s [cognitive systems initiative](#), said at this conference that a Manhattan Project is likely needed to “create an AI system that has the competence of a three-year old.”

In one sense, the goals of such analogies are clear. AI, the comparison implies, has the potential to be as transformative for our society as nuclear weapons were in the mid-twentieth century. Whoever masters it first will enjoy a massive head start on the next wave of technological development, economic competition, and, yes, the [arms race](#) of the twenty-first century. It’s a project that comes with ethical implications that demand focused and well-resourced attention. These consequences are so important that we should not bat an eye at ploughing limitless resources into its development.

But if this analogy is to sustain such a bold claim, it bears closer scrutiny. First, analogies of this sort are not innocuous. Invocations of historical examples, especially examples so iconic as the Manhattan Project or the Apollo program, aim to borrow the authority—and implications of success—that such historical episodes command. It is prudent to examine analogies to see if that authority is merited, or if it has been unjustly swiped.

Second, analogies might highlight unexpected features of a problem by highlighting similarities between something we don’t understand well and something we do, but they also constrain our thinking. They compel us to discuss problems in certain terms and think about their consequences in certain ways. The historical analogies we allow to dominate our discussion of AI therefore have crucial consequences for how we think about its future. This is particularly important to note in the case of AI, where broad foresight and flexible strategies are required to ensure that we harness new technologies in ways that are beneficial for all.

Is the current development of AI like the Manhattan Project? In many critical ways, it is not.

This is not to suggest that AI doesn’t have the potential to visit widespread changes on the world—it might very well do just that. But the analogy, on closer examination, doesn’t hold up. The following five ways in which it breaks down tells us a great deal about how we might think about AI in more useful ways.

1. The hard scientific work was done before the Manhattan Project started.

One of the biggest myths about the Manhattan Project is that it was a triumph of nuclear physics. It was not. By the time it was instituted in 1942, it could rely on about a half-century of atomic and nuclear physics.

The neutron was discovered in 1930, joining the proton and electron as the particles that explained the structure of all known elements. The general properties of the nucleus and the essential principles governing nucleon interactions were worked out through the 1930s—including the principles of nuclear fission. Very little nuclear physics had to be done during the Manhattan Project in order to accomplish its goals.

This is not the circumstance that currently prevails in AI research. We do not understand the principles that ground intelligence—biological or artificial. Although we have advanced considerably in our understanding of the cognitive and neural mechanisms underlying sophisticated behaviour in human and nonhuman animals, we still lack even a basic understanding of how animals succeed in flexibly learning and reasoning about the physical and social world. One can of course reframe the goal of AI research from one of achieving artificial general intelligence to one of developing systems that excel at particular tasks, such as playing chess, driving a car, or [generating cake recipes](#). In the case of artificial narrow intelligences (ANIs) such as these, however, there is no overarching project that lends itself to being compared to the Manhattan Project, but rather a multitude of distinct projects and aims, each of which must be evaluated on a case-by-case basis.

The Manhattan Project rewarded the resources invested in it because it proceeded from well-understood principles. A crash funding program for an area where the principles are unknown or in flux would be a much riskier investment.

2. The Manhattan Project addressed a highly constrained technical problem.

The question of how you build a nuclear bomb is complicated, but clear. The Manhattan Project's goal was therefore well defined. If you know that once you pack enough of a certain type of atom into a small enough space, it will explode, then it's just a matter of how you do the packing.

The Manhattan Project, to be sure, required a massive engineering effort. The \$2 billion dedicated to the cause was directed primarily at technical problems: How do you separate enough of Uranium-235, the fissionable isotope, from the much more abundant and much more stable Uranium-238? How do you manufacture enough Plutonium to sustain an explosive nuclear reaction? How do you get that Plutonium to detonate properly given that its properties prevent it from working like a simple Uranium bomb in which it's possible to unite two sub-critical masses into one supercritical mass?

These were the difficult problems of nuclear weapons design, and as a result, the answers to them were classified. What was not classified was the basic research carried out by nuclear physicists before and after the war: this was published and accessible, contributing in large part to the myth that the Manhattan Project was a triumph of physical research.

But most problems are not like this. General AI is not like this. We do not have a set of clear, well-defined problems that can form the focus of a massive technical effort. We don't even agree about how we would know if we had genuine AGI. Debates in animal cognition research show how difficult it is to come to a consensus concerning what counts as generally intelligent and what constitutes an appropriate test for identifying sophisticated cognition and behaviour. In contrast, after the Trinity test, everyone at the Manhattan Project knew that they'd succeeded. One could again shift the focus from AGI to ANI—the criteria for success in the latter case is typically unambiguous (often it's to outperform the best human in the task). However, one must then be clear that the comparison being made does not concern artificial general intelligence or AI generally, but rather a particular program designed or trained to succeed on a particular task, like facial recognition or missile guidance.

3. Massive brain drain in the 1930s led to a high concentration of world experts in one place, all working toward one (military-enforced) goal.

The 1930s saw a massive migration of intellectuals out of Central Europe, many of whom had been instrumental in developing the new physics underwriting the very possibility of nuclear weapons. A great many of these individuals came to the United States, which, conveniently, was just maturing as a scientific nation.

The European physicists and mathematicians coming to the United States and gaining posts in American universities therefore integrated with a young, energetic generation of American physicists. This was a historically unusual concentration of expertise, and the entry of the United States into World War II meant that almost all of the people with the most relevant talents were able to be conscripted into the effort to build the bomb.

The disanalogy with AI in this case is clear. Certainly, many talented people work on AI, but by no means are they all bending to the same oars. The fact that AI is being developed within a corporatized, profit-driven environment means that it is responding to fundamentally different incentives, which imply a fundamentally different distribution of expertise.

Many of those advocating for AI development would be aghast at the suggestion that the world's (or at least a country's) experts be plucked from their jobs, sequestered in a remote government laboratory and placed under a draconian secrecy regime. But such were the measures necessary for the Manhattan Project to achieve success, even in its comparatively narrow technical goals.

AI is a twenty-first century technology being developed primarily by twenty-first century businesses. Any large-scale, Manhattan-style investment would have to come from twenty-first century states struggling to balance it against many other competing technical and social priorities, with neither the clout nor expectation of compliance necessary to attain a similar concentration of expertise. Of course, things could move in this direction, and may already be doing so with, for example, China's [civil-military fusion](#), but we imagine many working in AI outside of the military context do not mean to encourage this when drawing on an analogy with the Manhattan Project.

4. The level of investment in the Manhattan Project was unprecedented.

The Manhattan Project cost \$2 billion. In today's money, that's about \$30 billion or £22.5 billion—all concentrated on the same well-coordinated project. This remains the sort of funding that you can only get for a technical or scientific project if you have the single-minded devotion of a wealthy state behind you. And then probably only during wartime.

When the US Congress canceled the Superconducting Super Collider in 1993, its cost had crept up to just \$12 billion, or about \$1.4 billion in World War II-era dollars. That scale of funding for major projects has proved extremely difficult to commit in peacetime, especially to non-military projects. We might see that scale of investment in AI worldwide over the course of several decades, but that does not amount to the concentrated funding of the Manhattan Project.

Dozens of governments have released or announced their intention to release AI strategies in the last two years, including the new [American AI Initiative](#) signed February this year. Many of these initiatives have money attached to them, ranging from [millions to nearly two billion USD](#). Although most of these strategies have industrialization and scientific research as their top priorities, they concern a wide range of technologies and applications. In no case are we near Manhattan-like numbers on any given project.

5. The Manhattan Project proceeded with the fierce urgency motivated by the threat of Nazism.

Hitler's bomb project, it turned out, was a dud. Many brilliant physicists, including Werner Heisenberg and Carl Friedrich von Weizsäcker, remained in Germany and worked on it, but it lacked all the other factors that we've described here and was nowhere near creating a working weapon by the end of the war.

But, of course, no one among the allies knew that at the time. Germany had been the heartland of modern physics before World War II, and the assumption remained that they had therefore begun with a head start. A singleness of purpose and the sense of a real and immediate existential threat lit a fire under the researchers and technicians working on the bomb.

Whatever our concerns about AI, they don't quite reach that scale. And because, as a technical project, it lacks the specificity of the bomb, it's difficult to see how it could.

One way that it is...

For these reasons at least, we should be suspicious of comparisons with the Manhattan Project. It was the product of a peculiar and delicate set of historical circumstances. Starting from the state of affairs in 1942, it is remarkable that, from a standstill, the Manhattan Project was able to create two types of working weapon in three years.

For that to happen, the science had to be in just the right place, the problem had to be clear and focused, the right concentration of expertise had to materialize, nearly inexhaustible funds had to be devoted to the project, and those working on it had to be scared to death of what would happen if they failed.

None of those properly apply to the current state of affairs in artificial intelligence.

But one strong point of analogy remains between AI and the Manhattan Project. We tend to think of the Manhattan Project as a success. But that's only a straightforward assessment if we limit our criteria of success to gadget production. "Success" in AI means a great deal more than creating something that works. It also means managing it successfully.

Here it's not so clear that the Manhattan Project was a success. The bombs it created were used against civilian populations—over the objections of many of the scientists involved. The arms race it set off reshaped global politics for decades, and we are still grappling with those effects. These were challenges Manhattan scientists anticipated, at least in part, but the project itself was ineffective at formulating a response.

However good technical specialists are at bringing technology into being, their visions for how, when, and why that technology should be used often hold little sway with the people charged with making those decisions.

Nuclear scientists notably failed to control how nuclear weapons were used. [The Franck Report](#), written by a group of concerned nuclear physicists at the University of Chicago, urged strongly that the bomb should not be deployed in a military capacity. Their motivation involved anticipating a great deal of what would come to pass during the Cold War: an arms race, increased international tensions, the difficulty of international control, erosion of American moral authority.

But the technology was nevertheless used in ways that suited the systems of authority in place at the time. That's something that we can expect to carry over to any products of AI research. Both the Manhattan Project and AI are inextricably bound up in global politics. Countries like the United States, Russia, China, and South Korea already take AI to be critical to future military

power. As Russian president [Vladimir Putin](#) put it, the country that takes the lead in AI will be "the ruler of the world." Within this context, the analogy to the Manhattan Project is clear: AI, like the development of the atomic bomb, is a tool that many want to develop in the service of securing global military power. Although this is true, the analogy still falls apart for the reasons stated above, and many who make this comparison do so without the military implications in mind. Thus, we would encourage alternative framings.

What is AI like, if not the Manhattan Project?

Both the proselytes and naysayers of artificial intelligence undoubtedly have some things right. But any claims for new technologies need to be evaluated both on their own merits and in light of an understanding of history that goes beyond glib sloganeering.

Happily, an abundance of such understanding is available in the history of science and technology. AI is widely distributed, connected to many other technologies, and managed by intricate and overlapping regimes of funding and influence. In that sense, it has more in common with the electrical grid than it does with nuclear weapons.

The electrical grid arose in a piecemeal way. Its implementation was accompanied by considerable disagreement about the form it should take and who should control it. It was integrated gradually with existing systems. Its effects were pervasive, but diffuse. Arising first in the private sector, it had to be subjected to regulatory structures before it could work at large scales.

Examples like this much better capture the effects of most technologies we care about. But we tend to avoid them, for the obvious reason that they lack a certain flair. Nuclear weapons are evocative and eye-catching; the electrical grid is pedestrian and mundane.

But perhaps we need a little mundanity in AI. The comparison with nuclear weapons has the unfortunate side effect that it tends to motivate competing utopian and dystopian visions, crowding out discussion about the ways AI might make small but influential changes in existing systems. The problems that threaten humanity (poverty, disease, ecological destruction, war) will not be solved by engineering a magic bullet, no matter how many talented people we bring to the table. Although new AI technologies are extremely powerful and this power can be harnessed for good, this must be done through the painstaking coordination of diverse groups and infrastructures and in conversation with those who are ultimately meant to benefit.