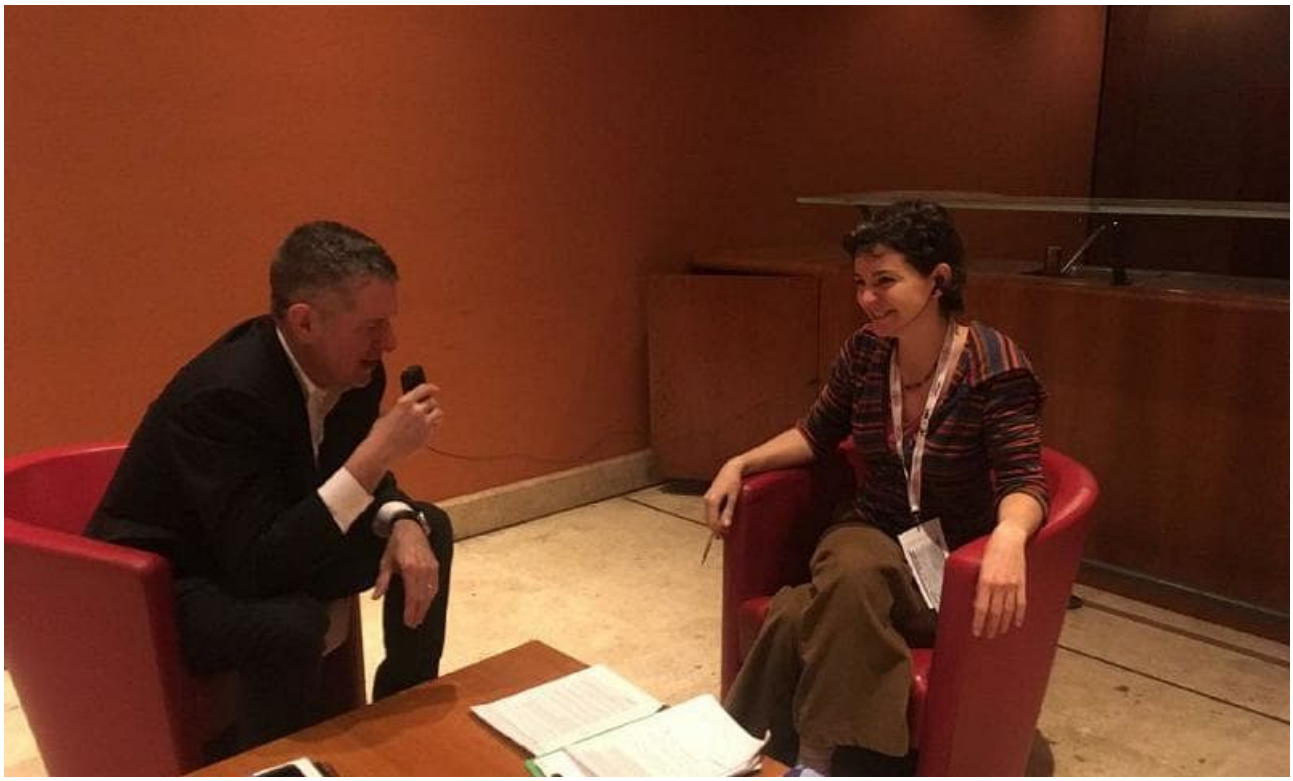
Di Roberta Fulci – Radio 3 Scienza

Intervista a James Barrat



Roberta Fulci intervista James Barrat, autore di *La nostra invenzione finale. L'intelligenza artificiale e la fine dell'età dell'uomo* (Nutrimenti, 2019)

What do you call a general artificial intelligence?

General artificial intelligence is very much like our intelligence, but our intelligence is not limited to one domain. Computer intelligence typically is limited to one skill, such as translation, or search, or object recognition, or navigation. Our brains can do all those things, and we can do them seamlessly: we can play the piano and then realize it's time for our chess match across town. We can drive to the chess match and talk to our French girlfriend on the phone, or boyfriend, and then we can play chess: we can do all those things seamlessly. That is a general intelligence. There is no example of computer general intelligence now, but it won't be long before scientists create it.

What's the difference between general artificial intelligence and super artificial intelligence?

Well, the first step is to create human level intelligence in a machine, and then, as a statistician named I.J. Good said in the 1960s, that machine would be better than us at everything that we do with our brains - including making smart machines. That idea is called the intelligence explosion. After a machine is improving its own intelligence, then the growth level of intelligence will expand exponentially. Right now, intelligence is pretty much flatlined. Our intelligence does not grow. Once machines are improving their own intelligence then we'll go from artificial general intelligence to super intelligence and that will be a very rapid transition in which intelligence is growing exponentially.

So going from "narrow" artificial intelligence to general artificial intelligence will be much harder than going from general artificial intelligence to super artificial intelligence?

That's right. Right now there are a lot of companies working on creating human level intelligence in a machine or artificial general intelligence. Companies like Google, and IBM, and Facebook, and Amazon have all said explicitly, or implied that, what they really want is to create basically a brain. Some people think that that can happen as early as 2029. Ray Kurzweil says for example that by 2029 a thousand dollars of computing will get you a computer as smart as a human. Now, that could take, depending on what polls you read, anywhere between 20 and 100 years from now. The important thing to know, though, is that once we create machines that are as smart as humans, it will be a short step to being much smarter than humans. And then, once those machines are doing artificial intelligence research and development, the pace of intelligence expansion will be very, very rapid, and that's called the intelligence explosion.

How do you picture a super artificial intelligence? Do you think it will have a body?

We used to think that intelligence required a body. I don't think many people think that anymore, because all of the things we do with our computers... there's no embodiment to them. We do search, we do translation, we do navigation, we do all those things that require narrow intelligence, but they're not necessarily in a body. So a super intelligence does not need to be embodied, in fact it will probably be decentralized: it will exist in computers and it will exist in the cloud. It will exist in a variety of places: intelligence will be decentralized. So it won't be simple to unplug, because will exist in many places.

You talk of artificial intelligence as of something with a will. Now, humans have the will to survive because it is written in their genes. Machines do not have genes. Why would they want to survive?

Well, you know, there's a giant economic wind propelling the development of artificial general intelligence, and one of the things they're working on right now is the whole concept of will. I don't think it's a concept that we will fail to create. Another way to look at it: this great scientist named Stephen Omohudro applied rational agent economic theory to intelligent machines and he found that basic drives will emerge from a self-aware, self improving machine, and those basic drives include: being resource-acquiring, they want resources whether it's money or bitcoin or something else; they won't want to be unplugged, they'll be self protective because as a goal achieving machine being unplugged would be the worst thing that could happen to it; they'll be efficient with the energy because they won't have perfect knowledge of how much energy there is in the environment; and they'll be creative which means that for whatever goal they're programmed to fulfill, whether it's asteroid mining or playing a really good game of chess, it would improve their goal achieving ability to be smarter. So self-aware, self improving machines will use some of their resources to do artificial intelligence research and development. So one way to look at it is that rational behavior and sufficiently smart machines will drive them to improve their own intelligence.

You talk about genetic programming. Are machines subject to evolution?

That's a very good question. It's very interesting to think about. Since artificial intelligence is being the next step in our evolution, some people think that we will melt with the machines, that the AI will augment our brains. Already our smartphone is a brain augmenter: I use it for all kinds of knowledge enhancement. I use it to speed my knowledge, to acquire knowledge, to get around, to find out new things, to gather more resources. So it's very easy to see all the power of a smartphone growing inside our brains for our intelligence to be augmented. So it may be that the next step for Homo sapiens is something like Homo sapiens siliconous or a combination between man and machine. I don't see that advancing as rapidly as machine intelligence on its own. I see artificial intelligence, specifically machine intelligence, progress is really exponential in that, and so I think it’s more likely that we could be replaced by machines rather than exist side by side with them.

Humankind will become extinct eventually. It will be replaced by another species, or many other species. What’s wrong if after us machines are coming instead of new living beings?

Well, it's very hard to say goodbye to our own species! You know every species goes extinct someday. I think one of the giant drives to explore our solar system and then our galaxy is to find new planets to colonize in a way to extend our survival. I like our species, I feel sentimental about us, and if we're not going to survive I want the machines that replace us to carry something important from us into the future. Some of our humanity. Some of our curiosity, some of the thing that makes us human.

If it’s so likely that every civilization develops an artificial intelligence soon or late, we may look for an artificial intelligence somewhere in the Universe. Wouldn’t they know about us?

Well, there's a paradox called Fermi paradox: the physicist Enrico Fermi wondered why with so many planets that were able to see in our galaxy that should be able to support life, why haven't we heard from anybody. Why

doesn't there seem to be anybody out there? One solution to the Fermi paradox is that there is some filter in the history of the civilizations out there and the filter stopped them from communicating to us. One of those filters could be artificial intelligence. It could be that once life gets beyond the radio stage, it has a brief window before it creates artificial intelligence and then artificial intelligence takes over and has its own goals, and those goals don't include looking for primitive life in the universe. Therefore they haven't found us. We have to ask ourselves why would super intelligent machines go prospecting for primitive lifeforms. You know there's no obvious reason why they would, or they might know that we're here and just choose to ignore us. It was interesting when I was writing Our final invention I discovered that SETI, the organization that searches for intelligent life, was turning part of its array, the Allen Array, towards cold parts of the galaxy, because it reasoned that artificial intelligence would seek to cool itself so it would gather, it would cluster, it would gather in colder parts of the of the galaxy. Although those are fascinating idea that is maybe where we'll find them someday all in the cold parts of the galaxy.

You met so many people doing research on artificial intelligence. Research on how to develop AI "the right way". What does it mean? How do we project a "good" AI?

It's very very hard to program ethics into machines. We have a hard time ourselves agreeing on any kind of ethics: if you take a simple example like tell the machine to preserve human life, a roomful of people would get into a big argument about when human life begins. And if you travel around the world you'd get an even bigger argument about when human life begins and what it means, and in some places women and children don't have the same personhood by law or by custom as men, so it's extremely difficult to program ethics into machines. When I talk about keeping tech companies in check, I think that we need a supervisory body to make sure they're not making dangerous AI. I think Google, Facebook, Amazon have not been good corporate citizens. I think they've done a lot of ethically irresponsible things and I don't think they can be trusted to develop this very sensitive technology unsupervised. So one thing we can do is: we can get our politicians to have

hearings and to establish something like the IAEA, the International Atomic Energy Agency. It has license to look in silos and to look at nuclear refineries and see that companies are in compliance with treaties, and it can impose sanctions. Unfortunately we need that kind of supervision with AI because these companies - all of them - have a track record of ethically slippery behavior. They can't be trusted to develop this technology unsupervised.

Ethics is not entirely separated from intelligence. Also, human ethics change in time. Is there any chance that being extremely intelligent also means being "good"?

Unfortunately being extremely intelligent does not mean being extremely kind. There's many examples of CEOs who really behave like psychopaths and are very intelligent people. Sometimes people who are not very intelligent rise to great positions of power. We've seen that with our president of the United States, right now. But being extremely smart does not mean being extremely kind. Anything that we put into a computer we have to program in, if it's not there in the programming it will not naturally arise. So if we want to make ethics in a machine we're going to have to think carefully about how to do that. But you go into this discussion thinking that you want ethics, that you want a machine that knows rules and knows how to behave. But what you really need is a machine that's intuitive, that grows with us. That's very powerful. It grows with us. It intuits what's best for us over the years because if we lock in ethical codes right now they won't be relevant in 100 years. If we lock in ethical codes right here they won't be relevant everywhere around the globe. So what we really need are these super smart machines that are intuitive about what's best for us. Now if it's hard to program ethics into a machine it's even harder to program intuition into a machine. But these are the ideas that will keep us alive, into the intelligence explosion and beyond.
Well, that wasn't a happy answer!

Many of us don't take AI issues seriously. Is it because we have never faced the problem before?

Well, a couple of things: we've never been outsmarted by a technology we created before. So this is a new one on us and we have a bias against things we haven't experienced. But the other thing we have to worry about now - there's an ultimate danger in the long term that's really serious and this is called the control problem. This is the issue about controlling intelligence far greater than ours. Can we do it? How will we do it? We need to develop a science for understanding it. But on the way there are a lot of ethical problems with the way AI is being developed that are not science fiction and they're not even futuristic. Right now for example people are building battlefield robots and drones. These are machines that kill humans without a human in the decision making loop. This research has been funded. It's being funded in the United States, Russia, China, Israel and some other nations. Right now Israel has a machine called a Harrop which has a conventional warhead. It's a drone with a conventional warhead that flies in the airspace around Israel. And then when another weapon targets it, it flies into that weapon and destroys the weapon and its operators, whether it's intentional or an accident or whatever: we have to be very careful about creating machines that kill people without humans in the decision loop. We really don't want to go down that slippery slope, because if we do, pretty soon we'll have whole armies of terminator robots. And this is not a science fiction story. This is happening now. Another thing that's happening right now is we have databases that were hand coded in the 90s and 2000s and even later that are tremendously bias. They're biased against women especially, they’re biased against minorities. And these databases are used to create algorithms that decide who gets bank loans, who gets into college and who gets certain jobs. So we have to be aware of this data bias. We have to get rid of these databases that are corrupted. Another problem we've got that we face with the AI right now is huge unemployment coming up for people for unskilled labor, and even middle class labor. Gardner and Company, a financial analysis group, says that by 2030 half of all jobs will be taken by AI. So any job with anything slightly repetitive about it: drivers, delivery people, truck drivers, taxi drivers, factory workers of all kinds, but even even lawyers, even doctors and even a lot of white collar jobs are repetitive, and that's something that computers are really good at. So we have to worry right away about a bunch of problems with the development of AI, not just the long

term super dangerous problems.

Climate change and gene editing: two issues that may go well or not, depending on how we handle them. And then the risks connected to AI. Why do these challenges happen all together?

That's an excellent question. I think every era has its own challenges, whether it's war or some new technology. Right now we're confronted with climate change which seems to be inescapable. But we know there are things we can do about it. Artificial intelligence is a dual use technology capable of great good and great harm if it's developed safely and ethically it can probably help with climate change. It could probably help solve some of these seemingly intractable problems. So there's a way that this technology could work to make us safer, if it's developed properly. What remains to be seen however is if we as a species have the will to make the hard choices to guarantee our future. So far with climate change we've shown that we don't have the will. We've known about climate change for a while and yet we still have climate change deniers. In fact the president of the United States is a climate change denier. These are people that are ignorant of science, they don't value science. Until we get over that kind of ignorance, we're going to be prey to all kinds of problems.

Do we build machines resembling us or are we rather finding out that we are machines after all?

That question is almost theological and it's a great question. What's really valuable about artificial intelligence... Actually, despite the critical nature of my book, I'm actually a giant fan of AI, I think AI is absolutely fascinating and here's a gateway thought for people who think AI is this impenetrable algorithm thicket. The gateway thought is that artificial intelligence is the most profound look at what we are. It involves neuroscience, the study of the brain, and involves psychology, now it involves ethics, and involves language acquisition and perception. How you change a precept, something you see in the environment, to a concept, something you store in your head and can refer to. It involves all the things that we do with our brains and it's resulted in new

insights into how our brains work. This is neurosciences that is moving ahead at a very rapid pace because we really want to try to find out how this miracle in our heads works. It does an immense amount of calculations and it uses very little energy. And this is exactly what we want from high performing computers. So artificial intelligence is a look inside. And it's helping us really understand who we are. It's also telling us that intelligence is difficult: if you want to mirror human cognition in machines, artificial intelligence says you have to bring your best game. You have to try really hard. It's not going to be easy. When they first started the field of artificial intelligence they thought: we can solve this. We can solve this intelligence problem in one long summer. Well, it's been many years. It's moved in fits and starts and we're in a golden period right now, where many businesses are using AI: fifty percent of the businesses in the United States use artificial intelligence. Eighty five percent of the businesses in China use artificial intelligence. It's generating a lot of income. It's generating an increasing amount of investment. If handled properly, it could solve many problems. But I don't think we're at the stage where we're learning from machines. I think we're still learning about ourselves in order to put those superpowers into machines.

Do you think we can figure out in how many ways our daily life will be different when we'll be sharing our world with even just general artificial intelligence?

Well, for one thing, unfortunately when we create human level intelligence in a machine it's going to replace even more jobs than the narrow kinds of AI we have today. It'll be fascinating, I mean: I'm very intrigued by digital assistants like Syria and Cortana, and those will be terrific tools. Think about how useful those will be for old people who are homebound who don't have enough social interaction! Five years from now, digital assistants are gonna be really amazing at conversation. They're going to remember things, they're going to be real confidence. We can worry about whether they'll absorb so much data that they'll be bad for us because we'll be giving away data. But the upside is they're really really interesting machines. The downside of course is we're gonna invent our way out of a lot of jobs and there are gonna be a lot of people who are unemployed. Foxconn is the world's largest industrial manufacturer.

They make the iPhone. They're in China. They recently bought 30000 robots to replace 30000 people. That's a harbinger of what's ahead. Gartner and Company, the financial analysis group, said that by 2030 half of all jobs will be taken by AI. What are we going to do with all the people whose jobs are taken? How are we going to retrain them to do other jobs that won't also be taken by machines? The technologists will tell you that these new industries of robots and AI will generate new jobs. But it's really hard to imagine that they'll generate new unskilled jobs. New jobs for people who are unskilled: it's hard to see how they'll create new jobs that can't also be taken by machines and if they create jobs surplus surely they'll just create a fraction of the jobs that they're replacing. It's a very scary employment environment ahead.

I come from the future. I know what's happened with AI and with super AI. I can fix everything that may have gone wrong with AI, but only if you get it right: you have to guess if AI and humans are living happily together or not. If you get it right, I'll fix everything. What would you answer?

Oh, that's a great puzzle. So, the fact that you showed up shows that we haven't been made extinct. Therefore it tells me that the technology companies, either got wise and stopped following profit only or be regulated by a wise and intelligent government. I would say the latter because companies are driven by a corporate mandate to deliver products and profits, like clockwork. They're required pretty much by their own bylaws to achieve those goals at any cost. So I don't think that that's the solution to how we're gonna survive this. I think the solution is that like other sensitive technologies artificial intelligence requires government oversight. So I would say - my guess would be that the governments of the world got together and realized how sensitive and dangerous this technology was and they agreed to treat it is just the way they've agreed to treaties about the development of nuclear fission nuclear power plants and nuclear weapons.