Visions, needs and requirements for Future Research Environments: An Exploration with Biologist and Science Fiction Author Peter Watts

Katharina Flicker (TU Wien), Florina Piroi (TU Wien), Andreas Rauber (TU Wien), Peter Watts

We live in remarkable times: the world is changing at an increasing pace, our societies face challenges that extend across national and geographical borders, and we are flooded with (dis)information. The scientific process has already changed extraordinarily in the past half century with research environments evolving from isolated and loosely connected islands to dense networks of researcher and institutional cooperation.

Still the world is changing and we need to ensure that science remains a global effort. Building a global network and infrastructures to support that aim, however, takes time. We need to start such building processes now and – most importantly – we need to develop and explore visions for research, science and society that give us ways into desirable futures. Thus, we launched an exploration series to elaborate visions on how research will be conducted in the future and to explore different perspectives on research.

“The question is not whether an AI-centered system would be perfect. It's whether it would be better than what we have now”

TU Wien: Research is undergoing a transformation. What challenges are essential to be addressed and solved in order to allow research and society to prosper?

PW: I think the big issue to solve is not so much the limitations of technology as the motives and cognitive capacity of people using it and of people being used by it. I think we have a chronic tension between two equally pernicious forces. One being that systems of power, which control most of the technology, are always going to act to hold on to that power. Thus, I think the problem is not the use of AI, but the motivation behind the people who control the AIs, and I don't know if we can trust that. Improving technology leads to ever-increasing control and surveillance of citizens; the imbalance will always favor disenfranchisement of the citizenry.

The other issue is—well, human nature. We were shaped by natural selection, and natural selection has no foresight; it only selects for what works in the moment. As a result, we find it very difficult to internalize future consequences. On a gut level, today's inconvenience is always worse than tomorrow's catastrophe.

If you want research and society to prosper, the most effective approach might be to rewire Human Nature. It's not out of the realm of possibility—everything from brain lesions to parasites already does that to some extent, and we have a pretty good idea of which neurotransmitters we need to tweak to control

our insatiable greed (one word: nociceptin). But saving humanity by changing it into something else isn't exactly a winning campaign platform.

TU Wien: What mechanisms or information would you need to trust such systems, or decisions based on AI?

PW: I would argue that a system trained on biased datasets is bound to be deployed as long as people get rewarded for them. If you could have some kind of an [information-based] system that does not incentivize getting a particular product out the door ASAP, that alone would go a long way towards preventing the kind of biased training sets we have, if you don't emphasize quick results. Admittedly I don't know how you do that, because in so many ways we need results last week. We need something *fast*, but the downside of that is when you need something fast, you get sloppy results like racist algorithms.

> "There's going to come a point where, while you can't say the AI is "saving lives", it's at least killing fewer innocent civilians than humans would"

There's a difference between *error* and *bias*. I would rather have a system with greater variation, one that produces unbiased error on both sides of the prediction line. That's probably naïve, but am I being naïve to hope that the problem might come out in the wash when you have a big enough, sufficiently unbiased data set and ensure that people have less control over how that data is collected?

Probably.

> "What you want is a constant state of constructive evolutionary chaos within competing systems that keep each other in check."

TU Wien: Let's have a closer look at big data scenarios then: More data is becoming available. An increasing number of actors derive information from the data, which leads to a plethora of different views on the same subject. Trying to weed out the most observed views and truths poses an interesting challenge. It is actually quite hard on a data level because what constitutes truth is always dependent on the context. The more we move towards anything that's not as binary, the harder it will be even for the most sophisticated and benevolent AI to determine truth.

PW: The question is not whether an AI-centered system would be perfect. It's whether it would be better than what we have now. A particularly relevant example is the use of AI in the battlefield. I am cognizant of the dangers of letting an AI-driven military drone off the leash and allowing it to choose and attack its own targets. Nevertheless, there's going to come a point where, while you can't say the AI is "saving lives", it's at least killing fewer innocent civilians than humans would. At that point, even for an AI that makes mistakes, you could almost consider it a war crime if you *don't* let robots loose in the battlefield.

TU Wien: AI also has a higher transparency potential than humans. There are mechanisms

to check it. It is much more difficult to understand how humans really make decisions, or what considerations and influences led to this or that decision. That is, observing actions is easy, but understanding the reasons for them at a truly fundamental brain or perceptual level is not.

PW: I don't know if I buy that. As I understand it, one of the problems that keeps cropping up with neural nets is that their logic *isn't* transparent; we train them on inputs and they generate outputs, but to a large extent the way those patterns form is relatively opaque. They get the right answer, but we're not entirely sure how they got there. On the other hand, while it may be well-nigh impossible to predict the individual behavior of a given person, it's disillusioningly easy to predict the behavior of *groups* of people. For example, it's been argued that even our ability to reason, to use rhetoric and logic etc., did not evolve to seek "Truth". That all evolved as a means of social control, to inspire other people to do what you want them to; the whole "search for truth" thing just tagged along as a side effect. We have studies in which information presented to a group by an outsider is rejected, while the exact same information, presented by a tribe leader, is accepted. So the question is not what is being said, or what information is being conveyed. The question is who is saying it.

There are evolutionary reasons for why we think this way. I still find it depressing.

TU Wien: Within the digital realm, we are currently experiencing a concentration of power in the hands of a very small number of companies. What kind of counter mechanisms could help us to push towards access that is more egalitarian to tools, information resources, benefits from whatever the world is offering us?

> "You probably want to invest more in the ability to adapt than you want to invest in something that's adapted to the current moment."

PW: Acts of sabotage! Malware! Which of course already exists—what I'm saying is, let's make that a feature instead of a bug. The big problem now is that governments or corporations prevail not because they are the most efficient, but because they got there first. By virtue of their inertia and their massive influence, they are essentially like great Redwood Trees which have rotted from the inside out, but are still capable of blocking the light to keep all the other saplings from growing. So let's build a system in which that kind of inertia is maladaptive. Let's build a system rife with malware, a vibrant ecology of digital life competing and trying to hack each other, distribute each other's trade secrets and wipe out each other's bottom lines. As long as you don't break the infrastructure of the Internet, what you are doing is resetting the conditions of fitness for nimbleness and rapid growth, rather than for sheer 800-lb-gorilla inertia. Companies would have to come up with a business plan that allowed them to start again from scratch at any moment. Lacking a stranglehold on the market, they'd have to—here's a thought—actually serve their customers to survive. This would grant the upstarts and kick-starters a fighting chance; at any given point a behemoth could crash, opening up sunlight in the canopy and allowing

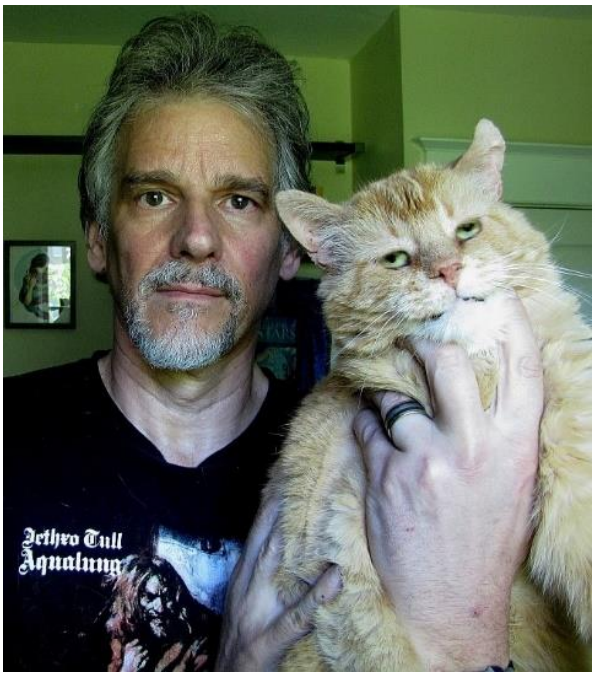
new entities to compete on an even footing again.

This kind of approach would of course be catastrophic in a monoculture—imagine a bug that took out every copy of Windows 10—but that's more an indictment of monocultures than of the approach itself. The whole point is to get rid of the monocultures, to replace them with a multispecies complex in which local speciations and extinctions are ongoing and relatively benign. What you want is a constant state of constructive evolutionary chaos within competing systems that keep each other in check.

TU Wien: So when creating new systems, we need to build in mechanisms that ensure the diversity and prevent monopolization of whichever niche in that ecosystem exists.

PW: We don't know what's going to work in 20 years. Therefore, you probably want to invest more in the ability to adapt than in something that's adapted to the current moment. It's also important to remember that we're trying to subvert monopolies and legacy power asymmetries here; you don't want your vibrant ecosystem targeting scientific data, for example. You'll need certain safeguards.

No, since you ask. I have no idea how those would work.



Peter Watts studied Zoology and Resource Ecology at the University of British Columbia in Vancouver, and is most reknown for his book Blindsight*. Watts won several awards for his work, including the* Shirley Jackson Award *and the* Hugo Award*.*