Beyond the Turing test: Elucidating the similarities and differences between machine and biological intelligences.

A Templeton World Charity Foundation Challenge

June 2017

Goal:

The Turing test[1] has become a popular concept for exploring whether machine intelligence can rival that of humans. Seventy years since Turing's paper, a myriad of discoveries of animal (including human) intelligence call for a more nuanced understanding of this concept. Our challenge is to develop a unified explanatory description of the similarities and differences between animal and machine intelligence.

Opportunity:

Advances in computer technology have allowed machines to excel at tasks that were previously thought to be achievable only by humans. As human and machine intelligences seem to be becoming more and more alike, questions arise about whether they could become identical. Can machine intelligence match humans in cognitive capacity, and more critically, can machines be endowed with consciousness and emotion? If so, will they bear moral rights and duties? The answers to these questions will determine how we treat machines, and how we prepare for the future.

Thanks to the rapid efflorescence of AI and recent successes in the cognitive sciences and the philosophy of mind, there is a new opportunity to realize novel forms of machine intelligence and situate them in an overall framework spanning the wide variety of intelligences.

Popular efforts have focused on exploring how humans and computers are *alike*, which is only half of the picture. We propose a balanced exploration, including how humans (and our biological kin) are *different* from machines. This could influence whether we treat machines differently, to what extent we choose to integrate them into our lives, and what expectations we have of them. Such an exploration could bring balance to the quickly-evolving vision of human-AI interaction, and shape public policy as machine intelligence becomes an increasingly prevalent element of human life. It aligns with Sir John Templeton's intent to explore human nature, discover fundamental structures, and to invest early in areas with potential for big impact. We are open to projects that are intellectually high-risk but remain firmly committed to excellent, well designed, and carefully implemented science. We are also particularly eager to support catalytic work with the potential to support sustainable, long-term, and novel lines of inquiry.

---

[1] Turing, A.M. (1950). Computing machinery and intelligence. Mind, 59, 433-460

Roadblocks:

Much of the empirical research looking to map the entire space of biological and machine intelligence is hampered by the tendency to focus on shared capacity to perform particular functions. Thus, the Turing test has been seen as the *sine qua non* of machine intelligence by asking a machine to function in a limited human social engagement as if it were human. This perspective tends to channel thinking about multifarious machine intelligences into particular and limited functional categories.

Similarly, the centuries-old supposition that human consciousness is of a different, immaterial kind from the physically-produced actions and capacities of the body obstruct any attempt to understand general intelligence as a result of physical processes. While some may still argue that the highest functions of the mind derive from some immaterial essence, taking such a position *a priori* is incompatible with empirical inquiry into the relationship between biological and machine intelligences.

As a result of the general interest in the Turing test and the narrative trope of androids and human-like robots, much of the field has focused only on the similarities between human and machine intelligences. This same tendency has led many to speculate that duplication or recreation of human mental capacities will necessarily replicate human experience. Such an assumption bears close scrutiny and requires detailed verification. It is equally interesting to map out those ways in which our intelligences differ, perhaps fundamentally and irreconcilably.

Some philosophical and psychological schools of thought are strongly averse to the very idea that machines can acquire the feelings and affect that characterizes our basic notions of what makes a thinking being. Others simply assume the that such capacities can and will arise. Our challenge will be to bring many such perspectives into dialogue to generate a comprehensive and useful overview of the possibilities and roadmap for research.

Effective research requires the development of robust and falsifiable hypotheses. But some of the most fascinating issues in non-human intelligence are also the hardest to operationalize into research agendas.

There exists little overall consensus on the essential parameters within which to map the overall space of intelligences. There also remains substantial controversy in the philosophical community, for example, over the use of related concepts such as cognition, consciousness and selfhood. In order to make meaningful progress, it will be necessary to build bridges between academic communities and establish good semantic and logical order across fields to make an overarching framework.

Challenge Statement:

Draw from tools in philosophy and science to create an understanding of machine intelligence that is empirically grounded, conceptually sound, applicable to all kinds of machine intelligence, and also relatable to human intelligence. *(The following is meant as a descriptive, not exhaustive overview of the tasks at hand.)*

1. Create a **conceptual framework** that can relate machine intelligence to human intelligence in terms that extend beyond outputs and behaviour.
    a. Develop a comprehensive and rigorous theoretical structure to relate human and machine intelligence in exact terms, referring to specific key characteristics of their respective underlying hardware.
    b. Create an account of how human and machine intelligence are respectively connected to or dissociated from consciousness, emotions, and ethical issues.
2. Conduct **empirical and theoretical investigations** to establish where animal (including human) and machine intelligence have similarities, and to identify fundamental differences relating to intelligent capacities.
3. Develop **guideposts** for engaging with diverse machine intelligences.
    a. Explore what kinds of roles are suitable for what kinds of intelligences.
    b. Explore how machine intelligence relates to rights and duties, for example, whether it would ever be unethical towards a machine to damage it or turn it off.
4. Create **new explanatory narratives of machine intelligence** to make high-impact and positive contributions to the development of new machines, public policy, and public discourse.
    a. Create scholarly media through translational and review publications.
    b. Create popular narratives through creative media outlets.

Potential breakthroughs needed:

- Refining the language currently used to describe machine intelligence - and achieving consensus between scientists and philosophers.
- A robust conceptual framework for interpreting empirical data from observations of machines and biological organisms in a way that the two can relate to one another.
- Successful engagement with the broader scientific audience without veering into debate about reductionism and the mind-body problem.

What we do not fund:

- Projects that focus predominantly on whether humans and machines is perform the same function.
- Projects assuming that human intelligence and consciousness are disembodied phenomena, or composed of an immaterial substance.
- Projects for which TWCF staff can foresee immediate commercial value, or that fall within the remit of other major funding agencies (however we welcome co-funding opportunities for projects with specific elements focused on morality, ethics, human empowerment and other areas consonant with donor intent that would not be otherwise supported).
- Projects predominantly focused on developing new computing technology.