



Final Statement of the Workshop: Power and Limits of Artificial Intelligence



On November 30th and December 1st 2016, the Pontifical Academy of Sciences hosted an international symposium on « Power and limits of artificial intelligence ».

In the past decade, computer science, robotics, and artificial intelligence (AI) have made remarkable progress. Those technologies hold great promise to address some of our most intractable social, economic and environmental problems, but they are also part of a long-term trend towards automatization, with consequences that may ultimately challenge the place of humans in society. This committee therefore reviewed the current trends of AI research, its potential utility and dangers, and formulated a number of recommendations.

Current trends. Major research is underway in areas that define us as humans, such as language, symbol processing, one-shot learning, self-evaluation, confidence judgment, program induction, conceiving goals, and integrating existing modules into an overarching, multi-purpose intelligent architecture. While progress is impressive, no evidence suggests the imminent emergence of a run-away intelligence with a will of its own. Artificial intelligence remains far from human and lacks an overarching mathematical framework.

Benefits. Used as a toolkit, AI has the potential to advance every area of science and society. It may help us overcome our cognitive limitations and solve complex problems such as energy management and ecology, where vast amounts of data present a challenge to human understanding. In combination with robotics and brain-computer interfaces, it may bring unique advances in medicine and care. By elucidating how we learn, it may bring dramatic changes in education. It may also help scientists shed light on the nature of intelligence, the organization of the universe, and our place in it.

Dangers. Unless channeled for public benefit, AI will soon raise important concerns for the economy and the stability of society. We are living in a drastic transition period where millions of jobs are being lost to computerized devices, with a resulting increase in income disparity and knowledge gaps. With AI in the hands of companies, the revenues of intelligence may no longer be redistributed equitably. With AI in the military, we may witness a new and costly arm race. While intelligent assistants may benefit adults and children alike, they also carry risks because their impact on the developing brain is unknown, and because people may lose motivation in areas where AI is superior.

Recommendations. The effort to develop intelligent machines must remain continuously directed to the greater good, reducing the poverty gap and addressing general needs for health, education, happiness and sustainability. All governments should be alerted that a major industrial revolution is underway and must take new measures to manage it. Scientists and engineers, as the designers of AI devices, bear a primary responsibility in actively trying to ensure that their technologies are safe and used for good purposes. We welcome the initiatives of some companies to create in-house ethical and safety boards, and to join non-profit organizations that aim to establish best practices and standards for the beneficial deployment of AI. We also

call for external civil boards to perform recurrent and transparent evaluation of all technologies including the military. The value functions that AI is asked to optimize require special attention, as they may have unexpected biases or inhuman consequences. Just like crash tests for transportation, the passing of ethical and safety tests, evaluating for instance social impact or racial prejudice, could become a prerequisite to the release of AI software.

SIGNATORIES

Prof. Werner Arber

President of the Pontifical Academy of Sciences; Biozentrum, Department of Microbiology University of Basel, Basel (Switzerland)

Prof. Antonio M. Battro

Academia Nacional de Educación (Argentina)

Prof. Olaf Blanke

Laboratory of Cognitive Neuroscience, Brain-Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL) Lausanne (Switzerland)

Prof. Patricia Churchland

University of California, San Diego, CA (USA)

Prof. Stanislas Dehaene

Collège de France, Paris; and Inserm-CEA, Cognitive Neuroimaging Unit, CEA/SAC/DSV/DRM/NeuroSpin, Gif sur Yvette (France)

Prof. John Donoghue

Henry Merritt Wriston Professor, Brown University, Department of Neuroscience, Providence, Rhode Island (USA)

Prof. Demis Hassabis,

Google DeepMind, University College London and University of Cambridge (UK)

Prof. Stephen W. Hawking

University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Cambridge (UK)

Prof. Yann LeCun

Director of AI Research, Facebook; Silver Professor of Computer Science, Neural Science, and Electrical and Computer Engineering, New York University, NY (USA)

Prof. Pierre Léna

Académie des Sciences, and Fondation La Main à la Pâte (France)

Prof. Laurie Ann Paul

University of North Carolina at Chapel Hill, Department of Philosophy, Chapel Hill, North Carolina (USA)

Prof. Alexandre Pouget

Université de Genève, Department of Basic Neurosciences, (Switzerland)

H.E. Msgr. Marcelo Sánchez Sorondo

Chancellor, The Pontifical Academy of Sciences (Vatican City)

Prof. Elizabeth Spelke

Harvard University, Department of Psychology, Cambridge, MA (USA)

Prof. Laura Schulz

Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, Cambridge, MA (USA)

Prof. Mariano Sigman

Universidad Torcuato Di Tella, Laboratorio de Neurociencia Integrativa (Argentina)

Prof. Wolf J. Singer

Max-Planck-Institute for Brain Research, Frankfurt am Main (Germany)

Prof. Elizabeth Spelke

Harvard University, Department of Psychology, Cambridge, MA (USA)

Prof. Josh Tenenbaum

Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, Cambridge, MA (USA)

Prof. Manuela Veloso

Carnegie Mellon University, Head, Machine Learning Department, School of Computer Science, Pittsburgh, PA (USA)

Prof. Cédric Villani

Institut Henri Poincaré (UPMC/CNRS), Paris (France)