Trust Manifesto

arm AI

November 6th, 2019

OUR GUIDING OBJECTIVE

AI Must be Ethical by Design

Arm believes that the technology sector has a responsibility to ensure that the development of AI[1] is guided by certain principles in order to generate public trust in the underlying technology and its applications. Further, Arm believes that unless the technology sector gains this trust, then it will fail to deliver the full potential that we are confident AI technology can bring to individuals, businesses and societies across the globe.

Arm would like to see the technology sector come together to create an ethical framework to ensure that AI is developed in a fair and responsible way. Without such a framework, there is a risk that regulation will become onerous and fragmented, and will not allow AI to succeed.

Creating the right ethical framework requires collaboration and dialogue. We have drawn up some principles – and some challenges – which we believe could help drive the debate. We welcome others to join us in doing so.

We believe that ethics should be incorporated in the key design principles for AI products, services and components. However, at present there is no defining set of ethics to follow and so we call for the formation of an industry-wide working group to define and standardize a set of ethics that can be adopted by anyone deploying AI technologies.

Since ethics are so critical to AI, it is essential that anyone working in the field has a solid foundation in the issues. We call on all universities and colleges that teach AI to include mandatory courses on issues relevant to ethics in AI at undergraduate and graduate level. Further, we believe that all businesses developing AI technologies must ensure that their staff complete mandatory professional training in the field of AI ethics.

[1] For the purposes of this document, we define AI as the simulation of human intelligence processes by machines, especially computer systems. At a basic level, these new systems bring about the ability to correlate historical data to provide informed decisions, using techniques that describe solutions which may be beyond human ability.

Ethical Principles of Trust in AI Systems

There are many issues that must be addressed in the development of an ethical framework for AI that enhances trust. As a starting point, Arm proposes the following principles.

1/ WE BELIEVE ALL AI SYSTEMS SHOULD EMPLOY STATE-OF-THE-ART SECURITY

Given the risk of cyberattacks on critical AI systems causing major disruption, all AI deployments should take advantage of task-specific advanced hardware and software security.

Only by ensuring an end-to-end chain of security can we truly trust AI technology and the data, actions and insights that the technology will create.

2/ EVERY EFFORT SHOULD BE MADE TO ELIMINATE DISCRIMINATORY BIAS IN DESIGNING AND DEVELOPING AI DECISION SYSTEMS

The technology sector is global and the AI industry, in turn, is global and serves a community of diverse customs, values and perceptions about what is ethical. There is wide recognition that the careful use of training data is crucial in ensuring that discrimination and bias do not enter AI systems to the extent that the implementation of such data may be illegal or unfair. Concerns about this are already playing a role in the debates about using AI in areas as diverse as criminal justice and facial recognition technology. Standards must be developed to both assess the quality of training data and to enable traceability so that systems can be linked to the data sets on which they were trained.

3/ WE BELIEVE AI SHOULD BE CAPABLE OF EXPLAINING ITSELF AS MUCH AS POSSIBLE: WE URGE FURTHER EFFORT TO DEVELOP TECHNOLOGICAL APPROACHES TO HELP AI SYSTEMS RECORD AND EXPLAIN THEIR RESULTS

Where appropriate, the way an AI system works should be capable of being transparent, and the decisions that result from it should be explainable to a human interrogator including non-specialist users of AI.

4/ USERS OF AI SYSTEMS HAVE A RIGHT TO KNOW WHO IS RESPONSIBLE FOR THE CONSEQUENCES OF AI DECISION MAKING

In future, AI systems will make life-threatening decisions such as in autonomous vehicles or robotic surgery. The industry needs to work with standards and regulatory bodies to develop commonly acceptable frameworks for mapping liability where it is a concern.

5/ HUMAN SAFETY MUST BE THE PRIMARY CONSIDERATION IN THE DESIGN OF ANY AI SYSTEM

In general, if AI technologies are not able to demonstrate they can operate as well, if not better than humans, in instances where human harm is possible, they should not be deployed.

6/ WE WILL SUPPORT EFFORTS TO RETRAIN PEOPLE FROM ALL BACKGROUNDS TO DEVELOP THE SKILLS NEEDED FOR AN AI WORLD

Opinions differ as to the likely impact of AI on jobs, but most agree that the nature of work is going to change, and the tech sector cannot be complacent. All stakeholders should look afresh at the skills training provided in schools, colleges and elsewhere, and help our training providers to respond fast to the demand for new skills. This needs to include training people in appropriate skills across the whole of society.