ECIS European Committee
for Interoperable Systems

ECIS ivzw-aisbl
Louizalaan 65 avenue Louise Box 2
B-1050 Brussels, Belgium

T/F +32 (0)2 706 24 15
info@ecis.eu
www.ecis.eu

ECIS roundtable event – 29 November 2016

Summary of Presentation on Artificial Intelligence

Author – session moderator
(more detailed report by the subject matter expert, Adrian Weller, to follow)

Artificial Intelligence ("**AI**") is not a new phenomenon; it has been around for at least 50 years as possibly the grandest of all challenges for computer science. Recent developments have led to AI systems providing remarkable levels of progress and value in different areas – from robotics in manufacturing and supply chain, to social networks and e-commerce, and systems that underpin society such as health diagnostics. As with any technology there is an initial period of hype, with excessive expectations and then a period of reality and measurable results – we are at the beginning of such a period right now.[1] Recent progress has been enabled by three factors:

- Advances in algorithmic theory;
- Greater computational power; and
- Availability of data

Our session focused on three questions – (i) how do we deal with liability; (ii) how can privacy be respected; and (iii) how can we achieve a greater degree of transparency and accountability in algorithms and avoid discrimination. Algorithms arguably need to attain a "higher" standard than humans, they need to be more transparent than "human decision making" and while they may follow the same logic as humans they should be held to account.

First, with regard to liability, it is important among other aspects, to consider the complex supply chain in AI services. The self driving car is a good example. In the case of an accident, who is at fault – the software developer, the manufacturer or the driver? AI agents are rather like children, they are strongly influenced by what they learn from data – outputs can be substantially changed depending on the data set used. Examples are the Tay chatbot that emitted hate speech after being exposed to Twitter. It is therefore important to consider what standards of behavior are appropriate and what actual normative standards should be in place. In addition, algorithms should be inspected at a technical level, so that the reasons for malfunctions can be established. From a legal point of view, a key consideration is whether algorithms have a legal personhood.

Second, with regard to privacy, it can be said that we live in an age where "tracking" and various kinds of surveillance have become commonplace. Health apps are widely used and our phones collect and transmit large amounts of data about us. Giving up personal privacy, while maybe not such an issue for today's millenials, can clearly have significant consequences down the road (think Facebook). The ability to manage one's data, and third party data ownership are important factors in this regard. Who actually owns your data? For instance, if

[1] http://www.nextgov.com/emerging-tech/2016/12/microsofts-new-plan-flood-your-entire-life-artificial-intelligence/133908/?oref=nextgov_today_nl

you allow a smart meter to be installed in your household, who owns the data collected about your energy usage and patterns and who is given access to it? Has the data been retained by the primary business group or is it passed on to third parties? If data is "tweaked", how does this affect ownership since it has been transformed? It should be noted that tweaked data can be used to train algorithms.

What about the "right to be forgotten" as set out in data protection laws including the General Data Protection Regulation ("GDPR")? And, if your data has been used to train an algorithm, what happens when your data is erased? Does this erasure influence the output?

Another growing area of interest is the concept of "Differential Privacy" and the potential to mask certain data fields. This will be covered in a follow-up paper.

Third, with regard to transparency, a key concern is to ascertain "how well" an algorithm is working, set against its original objectives. Can a system be trusted? Why was a specific recommendation made in one decision instance? Is the algorithm "fair"? What about instances of bias in the system, which may be influenced by certain attributes such as gender or ethnic origin? A good example would be an application for a bank loan online, which may refuse credit to minority groups and cause follow-on adverse impacts on an individual's ability to get credit. Another example is the COMPAS recidivism risk prediction system used in the US to help decide whether parole can be granted.[2] This has been claimed to be biased in certain ways in terms of race. How do you remove bias from such systems to avoid discrimination? Is the right approach to delete certain data fields, or to add other data fields to make the system fairer? Arguably it can be said that humans are by nature biased so is it possible for AI systems to function better?

In the follow up discussions we also discussed the European regulatory context. A key consideration here is that AI is changing very rapidly and policy actions tend to lag behind. As such, there is a risk that regulatory action by the EU as in other technology areas, may fall behind the curve of technological development and deployment. It is therefore difficult to imagine how such regulatory intervention in the EU would usefully guide the development and deployment of AI systems in a meaningful and timely way.

On the other hand, key issues underlying AI, such as data protection, safety and security, liability, and accountability are already to a greater extent than is immediately obvious, covered by current and soon to be implemented legal frameworks. A good example of this is the GDPR, that will take effect as law across the EU in 2018 and will restrict automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which "significantly affect" users. The law will effectively create a so-called "right to explanation," whereby a user can ask for an explanation of an algorithmic decision that was made about them (see Goodman, Flax et al).[3]

Policy action should make sure that society can take full advantage of the capabilities of AI systems while minimizing the possible undesired consequences on people. Safety is very important, as well as fairness, inclusiveness, and equality. These and other properties should be assured of in AI systems, or at least we should be able to assess the limits of an intelligent machine, so that we do not "overtrust" it.

It is therefore of utmost importance for societal well-being that policies and Regulations help society use AI in the best way. Ethical issues, including safety constraints, are essential in this respect, since an AI system that behaves according to our ethical principles and moral values would allow humans to interact with it in a safe and meaningful way.

[2] https://www.nytimes.com/2014/02/17/opinion/new-yorks-broken-parole-system.html?_r=0

[3] https://pdfs.semanticscholar.org/25d7/d975a12fd657b4743acd262cbdfe2dc2e6e9.pdf

While it is clear that a complete lack of Regulation would open the way to unsafe developments, this may not be the case in the EU due to the strong legal framework that will frame the development and deployment of AI.