

Célestino Guemes Seoane, Solutions R&D, Atos



Embedding ethical values into intelligent systems

“Technology is neither good nor bad; nor is it neutral.”

Melvin Kranzberg, technology historian (1917-1995)

Some of the most evident technology-driven ethical issues result from the changes brought about by Artificial Intelligence (AI). These are highlighted in the news almost daily and include racial discrimination in facial recognition systems, gender bias in business applications, the widely-perceived “exploitation” of privacy by digital platforms, and inequality caused by automation.

As active players in this society-shaping evolution, technologists are starting to adopt a more involved, proactive stance that seeks to minimize negative impacts of technology and foster positive societal progress. While we need to avoid imposing our own values on others, as a minimum we should be aware of

the ethical impacts of our own technological decisions. And as with many other high-impact concepts such as security, privacy or accessibility, our involvement cannot be an afterthought. If you try to retrofit this kind of thinking you will never get an optimal solution.

The conceptual framework that Ethics by Design represents can provide a solid foundation to address the way we build socially responsible intelligent systems. Our approach aims to form a set of standards as an ethical blueprint that our partners and customers will also be able to make use of. Importantly, Atos’ Ethics by Design has the potential for incremental improvements to address future challenges that new advances in technology are set to bring in the future.

The conceptual framework that Ethics by Design represents can provide a solid foundation to address the way we build socially responsible intelligent systems.





Case study

AI, health and ethics

The application of AI in the health sector provides a useful illustration of the ethical considerations that could typically be built into application design.

Imagine a team of data scientists who are developing a deep learning system for predicting cardiac problems, based on multiple data sources such as health records and electrocardiograms. They develop a model that works well in some of the scenarios they tested, but which does not work at all well when applied to women. This means that further development on the model is unlikely to progress as the risks are too great.

Atos' Ethics by Design comprises tools to detect bias in data to avoid these kinds of anomalies from the outset. For example, it may be that there is significantly more data available about men of a certain age with heart conditions, and that this has been the principal source of data for the model. At the same time, tools need to be developed to explain why the AI algorithms are detecting certain patterns as this will help developers and doctors to gain a clearer understanding of the results.

By having a series of checklists and guidelines in place as part of a process methodology, the development team will be guided to anticipate problems early on in the design and the development cycle of AI systems. These methods should validate the outcomes of the model and pilot the data, and there will be use-case analysis from an ethical perspective before the first data is analyzed.

The best solutions are developed by talented teams of data scientists coming from diverse backgrounds. While training in ethical issues in AI is critical, incorporating culture as an element in design should cover latent biases. In the example of the cardiac solution, it could be that an all-male data scientist team might have preconceived ideas about heart disease based on male symptoms. Further, there are many examples of digital solutions that do not incorporate accessibility factors for people with disabilities.

Increasingly policy makers are considering the role of regulation in relation to AI. The data scientist team developing an AI health application will need to be aware of regulations, both for AI in general and also specific to health.