TEAMING.AI 2ND PRESS RELEASE

February 2022

Auditable ethics as a solution for trustworthy AI

An important challenge for any new technology is ensuring that it satisfies legal and ethical requirements, in terms of safety, privacy protection, intellectual property rights, confidentiality, or fundamental rights protection in general. An arguably even greater challenge is ensuring that a technology not only satisfies such requirements, but can be tested and proven to do so. This is where the relatively new concept of 'auditable ethics' comes in.

In the context of TEAMING.AI, auditable ethics refers to the ability of verifying at any time that a technology has not only been designed with a certain value system in mind, but also that it is currently operating within the boundaries of that framework. In a sense, the vision builds upon the existing data protection principles of privacy-by-design and accountability, as they are enshrined in European data protection law. The privacy-by-design principle implies that a technology must be designed in such a way that it supports the protection of personal data (and thus also that it protects privacy), and the accountability principle implies that it must be possible for the responsible parties to prove how compliance with data protection law has been ensured.

Auditable ethics in TEAMING.AI

Auditable ethics applies the same logic, but increases the stakes a little bit, since it includes but is not limited to data protection and privacy protection. In TEAMING.AI, the ambition is therefore to firstly establish a generic ethical framework for artificial intelligence applications, and secondly to find a way to integrate this framework into AI applications in a way that allows compliance to be verified.

The creation of an ethical framework – a value system – is already not trivial. Within TEAMING.AI, this has been done by defining high-level TEAMING.AI policies, consisting of a series of requirements that are relevant to AI industrial manufacturing use cases. These requirements of course include data protection compliance, and existing product liability and product safety laws, but also the requirements that have been proposed in the European Commission's draft AI Act (e.g. in relation to transparency, intervenability, data quality, and so forth). The objective is not to define a one-size-fits-all value system that could cover all AI use cases, but rather to identify relevant questions that could be asked for any AI application. How is supervision ensured? Have data protection impacts been considered? Is the system certified in any way?

Secondly, it is the ambition for this ethical framework to be implemented in TEAMING.AI's architecture, in a way that allows compliance to be continuously and automatically verifiable. In practical terms, this would

ideally mean that a user of the TEAMING.AI solutions – such as e.g. a company using a TEAMING.AI driven AI application, or independent verifiers such as auditors or even labour unions – would be able at any time to assess precisely which controls have been applied to the application, and precisely in which way the requirements have been satisfied. Each compliance question would be linked to a specific response, and (if available) to a specific evidence of that response's factual accuracy.

The general approach for modelling compliance in a verifiable way consists of rendering the legal and ethical policies as part of the Teaming.AI knowledge graph. In this manner, an AI application can indicate which requirements it claims to satisfy and how. Where evidences are structured and standardised – which will be rarely the case initially – validation can also be standardised. The audit is however not normally fully automatic, since (in the current state of play) human intervention for the interpretation of evidences is usually required.

The principal goal of the ethics framework in Teaming.AI is to integrate these human interventions seamlessly into the teaming workflow, so that we can assure and verify that every change in the system complies with the requirements, thus achieving auditable ethics. In that way, TEAMING.AI can contribute to an efficient, safe, ethically sound and legally compliant workspace.

For additional information please contact

Project Coordinator: Software Competence Center Hagenberg GMBH (SCCH)

Bernhard Moser bernhard.moser@scch.at

Mario Pichler Mario.Pichler@scch.at

Communication and Dissemination Manager: Core Innovation

Ilia Kantartzi ikantartzi@core-innovation.com

Connect with us