

Continuous Ethic Compliance (CEC)

1st Hirsekorn, Hendrik
Wirtschaftsinformatik:
IMBIT 18A
Duale Hochschule Baden-
Württemberg Mannheim
Mannheim, Germany
h.hirsekorn@gmx.de

2nd Schork, Simon
Wirtschaftsinformatik:
IMBIT 18A
Duale Hochschule Baden-
Württemberg Mannheim
Mannheim, Germany
simon.schork@gmail.com

With the increasingly development of more intelligent like systems called AI systems, we see an increasing need for a structure that controls and oversees these systems within every company that develops and utilizes so called AI. We present a concept within the framework of Continuous Engineering which can achieve this task and will help guide the process of developing and controlling AI solutions. Therefor it is necessary to line out the current state of the art according to guidelines and possible upcoming standards from various selected resources. After all the challenge of implementing the concept and generating benefits can not be underestimated which leads to a discussion about the implementation and scaling within a company. The conclusion gives a prospect on what to come and how it could be handled in an ethical way.

Keywords—artificial intelligence, ethic, compliance, continuous engineering, machine learning, deep learning

I. INTRODUCTION

Many people would describe the development of computer systems, that can perform tasks that usually need human intelligence, better known as artificial intelligence, as an ongoing revolution. We know by now that the development of better and more complex AI systems will outperform humans in many ways and with the many ways it will improve our society and economic sectors, by performing tasks more precise, in a faster way and at a lower cost. Looking to the future, this development also brings a lot of threats. For sure there are many jobs endangered and not as it may seem the jobs that need lesser education than others. To the job of a nurse the threat is not as high as for highly educated doctors, since an AI system could in future easily replace a doctor but it will take far more time to have efficient and reliable care robots, that might take over a nurse's work [1].

Already many movies or novels dealt with the topic that an AI might become so powerful that it starts to become a threat to the whole human life on earth. None of them shows a scenario in which everything goes well in a result of the AI's fulfilling not only human tasks but also acting according to human ethics [2]. The main difficulty in real life is to build AI's and intelligent machines in a way that it won't violate a given ethical code and to stay in that given range.

The importance of AI systems following ethical guidelines can be explained briefly in some cases that can occur in autonomous driving. Since approximately 94% of car crashes are due to critical driver related reasons [3], it might be possible to reduce accidents on the road by that same number through autonomous driving. However, the AI system controlling the car can still end up in situations

for it to be impossible to prevent an accident. For a specific case like the car driving around a corner towards an old man and a child crossing the street, while the speed of the car allows only to avoid one of them, the system knows already which one to avoid because it was programmed. A human could not act rationally in such a situation while AI system does exactly that. This means that the programmers implementing the AI into the car's software decide upon who will survive in a situation like that. In Germany such a decision would always be left to coincidence, but that might vary in other cultures having a different perception on whose life is worth more. Furthermore, the question is raised about who will come up for the damage when an autonomous driving vehicle causes material damage in order to avoid a deadly crash for a pedestrian [4].

In order to control that these ethical questions concerning the use of AI systems are discussed by the Developers and the standards are followed, although they might vary from culture, there needs to be a continuous ethical compliance making sure every requirement is met.

II. CURRENT SITUATION

Most companies already have a structure to deal with compliance topics e.g. quality assurance or compliance management, there are typically dedicated departments surveilling and ensuring this. The employees must sign an ethical guidelines document to which the company commits its actions and products. Currently regarding the framework of *Continuous Engineering* there is no dedicated structure which includes the supervising of artificial intelligence, machine learning and deep learning projects. This part outlines on which basis *Continuous Ethic Compliance* will operate.

Countries and institutions all over the world are making their move trying to provide their citizens and companies with guidelines and standards to follow when developing intelligent or semi-intelligent systems. The following list contains a few of those, published by a major institution or country. We will highlight those impacting our region the most (European Union and IEEE).

A. European Union: 'Ethics Guidelines for Trustworthy Artificial Intelligence'

On 8 April 2019 the EU presented their 'Ethics Guidelines for Trustworthy Artificial Intelligence' [5]. According to the document the EU wants to create a framework for the development of AI systems. They extract three components which an AI system should met: lawful, ethical and robust. To summarize an AI system should always *fit the legal rules* made on national or international level, for a scenario in which those are not up to date the

system should always *fit the ethical norms* of our society. To complement those two principles an AI system should also be *robust* in technical perspective and considering its context and operating environment [6].

To fulfill those components an AI system should be developed according to four ethical principles and meet seven key requirements. The *ethical principles* are:

- Respect for human autonomy
- Prevention of harm
- Fairness
- Explicability

For the *requirements* the expert group agreed on:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountability

These requirements must be achieved by different groups of stakeholders amongst others: developers, deployers and end-users [6].

The EU guidelines provide a first draft of an assessment list and recommendations for the implementation.

B. IEEE: 'Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems'

The IEEE as an international and accountable institution, with many standards published so far, especially for the IT sector. In 2019 an IEEE driven initiative published a document with the to advance and align public discussion to defined values, inspire the creation of standards and facilitate the creation of national and global policies according to their principles.

1) Goals

- Human Rights
- Well-being
- Accountability
- Transparency
- Awareness of misuse

2) Objectives

- Personal Data Rights and Individual Access Control
- Well-being Promoted by Economic Effects
- Legal Frameworks for Accountability
- Transparency and Individual Rights
- Policies for Education and Awareness

In their key values the IEEE and EU guidelines are overlapping, which is not unusual since the first version of IEEE has been published in 2017 and probably was one of the sources for the EU guidelines [7].

C. More policies for AI systems

In February 2019 President Donald J. Trump has signed an Executive Order launching the American AI Initiative

which will focus on investing in research and development and making federal resources accessible for them, setting AI Governance Standards and building an AI workforce while protecting the USA's national and economic security interests [8].

The South-Korean government has also made their move by publishing strategy guidelines not only including vision, strategy, policy aims and success factors but tasks and the draft of a framework to implement AI solutions [9].

III. THEORETICAL SOLUTION

To implement those approaches with Continuous Engineering it is necessary to create a dedicated structure within the framework. The Continuous Ethic Compliance should be performed by a dedicated team. According to the need and the number of projects including AI systems, this team could be part- or fulltime working together. The final goal is to prevent AI systems like mentioned in *I. Introduction* from violating not only laws but ethical guidelines and values like mentioned in *II. Current Situation*.

A. Structure of the team

It is necessary to create a cross-functional team since there is no profession which covers all the essential skills. Even though a cross-functional team comes with some benefits there are also some requirements for it to work. Since team members are all from different functions communication is not easy and misunderstandings are likely to happen. This can lead to dissatisfaction and disruption of the project. To avoid this the team leader must act as a mediator and create an environment of open communication among team members. In addition to a good communication among the team he as to facilitate a constant two-way communication with his team to generate motivation and trust in their work. A team leader with strength in communication and coordination will lead this team to success [10].

If the members fulfill the requirements it will enhance the effectiveness and efficiency of the team compared to non-cross-functional teams especially when working on new and innovative concepts [10]. The constant communication with people of different professions will enhance their ability of external communication which also leads to faster problem solving, therefore better project performance and is key to generating transparency and therefor trust among stakeholders and other employees.

B. Roles within the team

Beside the team leader the member must fit a few required roles. These are not fixed to the according profession if an employee meets the same level of knowledge by experience and education other than degree this could be also an option.

1) Project Manager

As mentioned in *A. Structure of the team* it is highly recommended to put someone with a lot of experience in project management and team leading to meet the requirements in communication and coordination.

2) *Compliance Officer*

The Compliance Officer must ensure, the practices used by the company to create value stay within the legal borders [11]. His knowledge and expertise should be used as the foundation. It is most likely not enough to judge AI systems by its own, since the laws restricting those are limited to not existing. Nevertheless, this will probably change soon which will make his role even more valuable.

3) *Machine Learning Engineer*

There are many definitions for an ML Engineer, most likely it requires strong mathematical skills, programming experience and knowledge of different algorithms used in Machine Learning and Deep Learning. It is necessary to understand how the systems work and how algorithms come to decisions when trying to categorize or judge an AI system.

4) *Data Scientist*

This seat should be taken by someone with deep understanding for the data AI system work with. Either someone from the field of Data Science or Statistics. At least he requires skills in mathematics, ML, AI and statistics [12]. One of the main duties will be reviewing training data to identify possible bias or discrimination within.

5) *Changing seat according to the different projects*

In order to fully understand the application of the system and generate trust within other departments. It is useful to have intermediaries between the CEC team and the different departments they are currently working with.

C. *Tasks of the team*

The first thing for every AI system will be a categorization how much regulation is necessary and which control mechanisms will be applied. First it needs to be clear whether and to what extend humans will be affected. Next is operationalization. It is required for an assessment of the system to set metrics and substantiate its results. After setting up performance indicators the model can be assessed. The model in case of an AI system is a statistical model which takes data as an input and generates the output, e.g. a prediction, through mathematical functions. The last assessment will be the algorithms of the model itself, e.g. the team could look at the written implementation of the model [13].

Probably the most argued when talking about more independent AI systems like self-driving cars, is the responsibility for the mistakes such a system could make. The CEC team has to work close with the development and the department which will apply the system to decide and record most likely a chain of responsibility in case of a mistake [13].

Another duty of the CEC team will be a “medium”-technical reporting for stakeholders and externals. One option to achieve this are Model Cards. They contain key information regarding the model and will facilitate a better understanding and therefor support creating Continuous Trust among the target groups. A Model Card includes:

Model Details (Developer, Date, Version, Type, License,...), Intended Use (Use cases envisioned during development), Factors (Different conditions under which the model was tested), Metrics (Performance measures, Decision thresholds,...), Evaluation Data (Datasets, Preprocessing,...), Training Data (same as Evaluation data), Quantitative Analyses (Results of evaluation according to different metrics), Ethical Considerations, Caveats and Recommendations (Additional concerns) [14].

IV. PRACTICAL IMPLEMENTATION

There are different approaches to a practical implementation of an ethical compliance team in a company that creates an AI system. The best and efficient way for making no mistakes and following all the guidelines would be an extra department that only does the compliance. However, that might be an efficient way to stick to the book, it would be rather expensive for a company to open up a whole new department and hire people to form a team that checks if all is in the targeted ethical specifications. In the future it might be necessary for extremely big and complex projects but for smaller ones still relevant enough to need a continuous ethical compliance team it is more cost efficient to pick the team from already hired people in the company and assigning them to the project for compliance.

The more cost-efficient team must be scaled by the project manager. He decides after consultation with other superiors, how many and which team members work full time, which part time and decides whether some might only need to attend jour fixes. The order of priority for full time work on the project is top down form the project manager to the changing seat but can vary in certain cases.

Looking in the future of autonomous driving cars that don't have a steering wheel nor a gas pedal or breaks a human can't interfere anymore. The model behind that idea is called Human-Out-Of-The-Loop. Today there are only cars allowed on the roads that are a mix between HOTL and Human-In-the-Loop, which describes the model that still needs interaction of a human being. The driver should always be ready to intervene in case the system fails [15].

Taking a step away from cars, it may be the case for other AI's that there should not be any use of a complete HOTL in order to be able to deescalate or improve things “by hand”. When some systems get out of control and humans can't intervene anymore, things might turn out bad. It is an individual decision that should be made well thought [16].

The individual assembled teams in a company will discuss such decisions and of course the controlling of sticking to the ethical guidelines.

V. CONCLUSION

Future AI trends will cause the disappearing of jobs while also creating new ones. They will increase people's life expectancy and many accidents or mishaps can be prevented, but unfortunately not all. As well as bringing these benefits, it also carries many risks and threats to the well-being of humans, their rights, by misuse, by

bringing questions of accountability and a lack of transparency. So, for more complex and advanced AI systems in the future, it is inevitable to assure through Continuous Ethical Compliance teams that the ethical principles and requirements are reached and kept.

Even though many institutions and countries are developing and publishing guidelines, there are no laws guiding and restricting the development of AI systems. This should be on the top of the agenda for the constitutional states which praise their legislation often enough. It is required for the concept to work, since every human being has its own understanding of ethical behavior and most likely acts according to that.

The CEC team needs to be built in the mentioned manner in order to work efficient and consist from members that have different viewpoints on the project. The different roles must be placed by competent people from project management for all others to work efficiently together by being appropriately coordinated. The Compliance Officer, Machine Learning Engineer and the Data Scientist need to be qualified for the project and have a deep understanding of their knowledge field and specialty. The individual person or even person group from the department that is also involved in the project should also have a slight understanding of what the others do in order to build a succeeding team.

All of them should understand the importance of their work by working through their tasks with conscience. The right decisions in the development process will prevent accidents. If it is only reacted when something bad happens through artificial intelligence, then any regulation comes too late [17].

- [1] Y. N. Harari, Reboot for the AI revolution., Nature 550, 324-327 (2017)
- [2] J. C. Havens, "Creating the human standard for ethical autonomous and intelligent systems", 2018, 28-31
- [3] S. Singh, Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey (Traffic Safety Facts Crash-Stats. Report No. DOT HS 812 115), National Highway Traffic Safety Administration, Washington D.C.
- [4] C. Keese, "Silicon Germany: How we can manage the digital transformation", Knaus, 2016
- [5] Author unknown, Report on "Ethics Guidelines for Trustworthy AI", European Commission, Brussels, 2019, <https://ec.europa.eu/digital-single-market/en/news/ethic-guidelines-trustworthy-ai>.
- [6] AI HLEG, "Ethics Guidelines for Trustworthy AI", European Commission, Brussels, 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, pp. 6-14.
- [7] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems", IEEE, Version 2, 2017, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf
- [8] President Donald J. Trump, "Accelerating America's Leadership in Artificial Intelligence", February 2019
- [9] Government of the Republic of Korea Interdepartmental Exercise, "Mid- to Long-term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution", 2019

- [10] C.-J. Chen, "Information Technology, Organizational Structure, and New Product Development—The Mediating Effect of Cross-Functional Team Interaction", In: IEEE Transactions on Engineering Management, vol. 54, October 2007, pp. 687-698.
- [11] M. Lenglet. "Ambivalence and ambiguity: The interpretive role of compliance officers.", In: Finance: The Discreet Regulator, Palgrave Macmillan, London, 2012, pp. 59-84.
- [12] V. Dhar, "Data Science and Prediction", In: Communications of the ACM, vol. 56 no. 12, December 2013, pp. 64-73.
- [13] K. Zweig, "Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können", Heyne Verlag, 2019
- [14] M. Mitchell, et al. "Model cards for model reporting", In: Proceedings of the Conference on Fairness, Accountability and Transparency, ACM, January 2019, pp. 220-229
- [15] N. Funke, "Human In The Loop: Interactive labeling of training data", agile im, 2019
- [16] Dr. L. Elliott, "Human-In-The-Loop Vs. Human-Out-Of-The-Loop in AI Systems: The Case of AI Self Driving Cars", the AI Trends Insider, 2019
- [17] Unknown Author, "Advance of artificial intelligence brings advantages- and concerns", CIO von IDG, 2017