EvoTegra GmbH
Kreuzwingert 11 • D – 55296 Gau-Bischofsheim

EVOTEGRA
CONCEPTS & INTEGRATION

Position paper by Evotegra GmbH on the current discussion on Artificial Intelligence (AI) ethics and explainability in Germany

EvoTegra GmbH is a provider of customer specific image processing solutions based on artificial intelligence and deep learning. With our individual solutions, we specifically target small and medium-sized companies that do not have the opportunity to build up their own AI knowledge.

Ethics is a "philosophical discipline or single doctrine that deals with the moral behavior of *humans*." Through ethics a collective subjective valuation becomes a social norm, but not an objectifiable quantity. Among other things this can be seen in the fact that ethical valuations might depend on culture and are subject to temporal change.

The "weak artificial intelligence" used today in practice is a self-optimizing mathematical approximation to an unknown complex function. The optimization is based on observations, which are made available to artificial intelligence in the form of data during the *training phase*. During the subsequent *application phase*, a weak AI will no longer learn.
Since today's artificial intelligence does not develop an awareness of what it learns, AI in principle is completely objective. Therefore the risks of using AI are related to the data and not in the AI itself.

The current discussion of AI and ethics aims to prevent artificial intelligence systems from discriminating people. To achieve this goal these systems should be designed to minimize the risk of unintended consequences. For example, a company would need to test upfront whether artificial intelligence discriminates or devalues certain groups of people. (Link)

In what context can ethics and AI be seen?
1.) Integration of ethics in AI
Since AI does not develop awareness of the learned content and ethics is not objectifiable, ethics can not be meaningfully integrated into artificial intelligence.
2.) Ethics of the data
The result of an AI is based on the data. Therefore the responsible use of data is an ethical approach to AI.
3.) Ethics of the results
This raises the question of how to assess the result of an AI and how AI and humans interact to maintain ethical principles.

If you compare artificial intelligence with the current methods, it becomes evident that AI basically fulfills the same purpose as "electronic data processing" for around 60 years. Data processing and evaluation. The two approaches differ only in the type of models used to evaluate the data.

Anschrift:
EvoTegra GmbH
Kreuzwingert 11
D – 55296
Gau-Bischofsheim

Kontakt:
Tel.: +49 171 / 402 4242
E-Mail: info@evotegra.de
Homepage: www.evotegra.de

Gerichtsstand: Mainz
HR-Nr.: B 43519
USt.-ID Nr.: DE 217667197
Steuer-Nr.: 44/654/70280

Geschäftsführer:
Tobias Manthey

Bankverbindung:
Deutsche Bank 24 AG
IBAN DE46 5507 0024 0024 0820 00
BIC DEUTDEDBMAI

Humans are limited in their ability to recognize complex relationships by Miller's number (7+-2). Therefore in human models complex relationships are simplified and e.g. evaluated in the form of a decision tree.

Therefore results are explicable *within the model*. The question of whether the assumptions of the model are statistically valid or (implicitly) discriminating is usually neither asked nor validated. It is largely socially accepted to get a *justification* whose *explanatory value*, i.e "relevance with respect to the decision", is typically not verifiable.

Models of artificial intelligence recognize relationships directly in complex data. Instead of simplified assumptions, they reflect the statistical reality of the data. As the evaluation is basically based on statistical correlations, this enables a fundamentally fairer assessment. However this becomes problematic when data is incomplete or contains apparent relationships that are not transferable to reality.

Due to the implied legal risks, it is a fundamental goal to develop non-discriminating systems.
In order to prevent discrimination in AI system several techniques can be used such as:
1. Removal of potentially discriminatory features from the data
2. Add more of the underrepresented data features
3. Higher weighting of under-represented data features
Overall dealing with this kind of data issues is a common but usually well manageable problem. Yet AI does not provide a universal solution for every task. If no sufficient amount of data can be made available, AI usually does not represent a good solution approach.

Human intuition and ethics complement perfectly with the speed and objectivity of AI. Since AI quickly uncovers false or incomplete human assumptions, not only the AI learns from humans, but humans can learn from AI. Since human intuition can not be replaced by artificial intelligence in the medium term, we generally recommend the use of AI as an assistance system. This can not just increase quality and efficiency of a company. In times of shortage of skilled workers, companies can benefit by using AI to increase the productivity of key employees.

Legal risks related to a lack of nationwide guidelines in dealing with the GDPR already present a significant obstacle to the introduction of artificial intelligence in Germany. From our point of view the use of AI in a *production* environment is still an isolated case within the industry or the public sector. The requirement that systems of artificial intelligence fundamentally must correspond to ethical principles and must be explainable means an additional hurdle that is not generally met by today's classical data processing systems.

Regardless of the technology used, the GDPR prohibits people being subject to fully automatic decisions, given the decision has a significant impact on the person. As therefore compliance related to ethical standards and explainability is an implied requirement for relevant decisions, we recommend to not differentiate in this respect between the classical data processing systems an AI.

In order not to further increase Germany's enormous backlog related to artificial intelligence, the goal for the further discussion about AI should be from our point of view on how to reduce legal risks by creating legal guidelines related to the GDPR, how to reduce bureaucracy related to public funding as well as raising awareness within the industry and public sector about the specific uses of AI.