ÖAW AUSTRIAN ACADEMY OF SCIENCES
ITA – INSTITUTE OF TECHNOLOGY ASSESSMENT
ITA DOSSIER

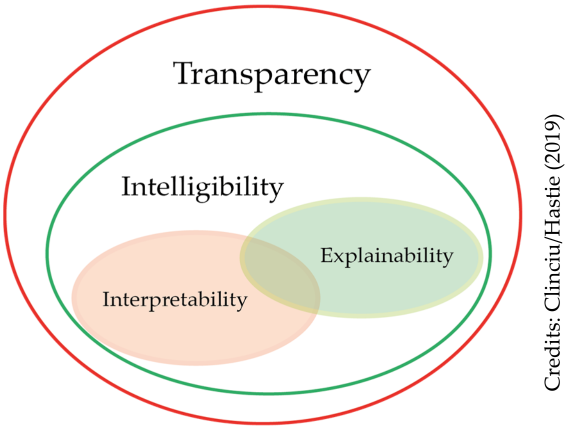
UNDERSTANDING AI?

IN BRIEF

- Artificial intelligence (AI) has already become an integral part of many everyday applications, from search engines to lending.
- How decisions are made by AI often remains unclear.
- Transparency is essential to better understand AI decisions.
- To ensure transparency, technical solutions alone are not sufficient.
- The social context and regulation of AI must be considered and take into consideration the social impact of AI applications.

WHAT IS IT ABOUT?

“Artificial intelligence” (AI) is ubiquitous. Large-scale AI applications are based on ever more comprehensive and more precise data analyses, raising hopes for a more efficient, objective design of economic or other social processes. Nevertheless, what are the consequences and side effects of widespread AI use? Decisions made by AI often remain in the dark and are sometimes even inexplicable to developers (“black box”). To make the use of AI safer and to assess its effects on humans, greater transparency and clarity are required. However, this request has mostly been proposed to developers, producers, and commercial distributors, whilst potentially affected parties such as consumers or other users often remain ignored. Furthermore, there are major obstacles: it is neither clearly defined what AI includes nor what exactly is meant by transparency. Whilst combining various socially relevant demands, AI serves as an umbrella term for different technical approaches such as algorithms, machine learning or algorithmic decision-making systems. Transparency, on the other hand, can be related to the algorithm, the process or the context. It can be understood as technical, i.e. for example the detailed disclosure of codes, or it can be understood as procedural, i.e. e.g. targeted communication to specific audiences. Consequently, a request for transparency alone is not enough to develop, circulate, and regulate AI responsibly.



Dimensions of transparency

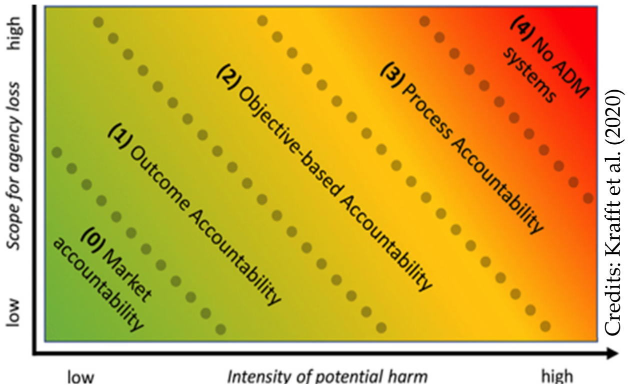
The research field of Explainable AI (XAI) aims to make AI decisions more comprehensible. It attempts to create transparency not only through technical external aids or interfaces, but also by communicating outside the AI system whilst including explanations about decisions during the development of algorithms. Features of such an interface include data visualisation or scenario analysis. For responsible AI development and effective regulation, the question remains whether such approaches are enough to enable social transparancy and responsibility, thereby mitigating social and ethical problems. A purely technical implementation of transparency does not allow for any conclusions regarding the social impact of AI systems. As a result, regulation of AI cannot be defined exclusively in terms of technical aspects, but must focus on the impact of AI on people.

BASIC DATA

Project title:	Künstliche Intelligenz – Verstehbarkeit und Transparenz (German only)
Project team:	Udrea, T., Fuchs, D., Peissl, W.
Duration:	10/2021 – 01/2022
Funded by:	Austrian Federal Chamber of Labour

KEY RESULTS

Social systems are already being transformed by IT and AI applications that support decision-making (e.g. where finding a job, allocating social benefits or creditworthiness are concerned). The widespread use of AI systems will lead to further changes. Social problems can arise even with conventional IT, such as discrimination through statistical methods. The use of AI could exacerbate such problems. How transparency should be designed in concrete terms is the subject of intense debate in research and politics. What is certain is that transparency or explainability must not only be implemented at the technical level, but must be easy to understand and verifiable for advocacy groups and those affected by AI, respectively.



Credits: Krafft et al. (2020)

One possible classification of AI systems according to risk levels.

As a result, regulatory approaches should consider not only technical approaches, but also, and above all, the social context of application. AI applications differing in their risk profile could be reflected in a system of tiered rules of transparency, thus making it possible for independent institutions to monitor applications whilst also actively supporting those affected by demands for transparency, complaints or lawsuits. Since the functionalities of AI applications cannot always be fully assessed, it would be appropriate to only use systems that humans can, in theory, control. In any case, the focus should be on taking responsibility for those affected by AI applications, which could be implemented by e.g. regulating bodies, developers or producers. Affected people and advocacy groups must be actively involved in the design processes of AI to increase its social compatibility.

WHAT TO DO?

In order to exploit the potential of AI and to promote a responsible and socially acceptable design, it is necessary to:

- Strengthen the interests of consumers and those affected in the discourse and development of AI.
- Develop a differentiated catalogue of criteria to classify AI and its risks.
- Use an AI definition for regulations that also includes established IT systems with corresponding effects on humans.
- Implement transparency in its various dimensions (i.e. the right to information; intelligibility for people; as institutionally anchored transparency, including the necessary procedures (legal remedies)).
- Universally register AI systems whose decisions affect humans, and to certify these systems according to the expected risk
- Pursue a regulatory approach that is evaluated periodically.
- Promote research on XAI, fairness, justice, accountability, responsibility, and societal impact.
- Ban ethically questionable AI systems or those that harm fundamental rights, freedoms or democracy.
- Ensure and strengthen human control over AI systems as a basic requirement. If this is not possible, moratoria for such AI systems should be considered.

FURTHER READING

Udrea, T., Fuchs, D., Peissl, W. (2022) Künstliche Intelligenz – Verstehbarkeit und Transparenz. ITA-Projektbericht 2022-01. In cooperation with the Austrian Federal Chamber of Labour. (German only)
epub.oeaw.ac.at/ita/ita-projektberichte/ITA-2022-01.pdf

CONTACT

Walter Peissl
Email: *tamail@oeaw.ac.at*
Phone: +43 1 51581-6582

IMPRINT: Owner: Austrian Academy of Sciences; Legal person under public law (BGBl. 569/1921 i.d.F. BGBl. I 130/2003); Dr. Ignaz Seipel-Platz 2, A-1010 Vienna; Editor: Institute of Technology Assessment (ITA); Apostelgasse 23, A-1030 Vienna; *www.oeaw.ac.at/ita/en*. | Frequency: The ITA Dossiers are published irregularly and publish the research results of the ITA. The ITA Dossiers are published in print in small numbers only and are made available open access to the public via the Internet portal "epub.oeaw": *epub.oeaw.ac.at/ita/ita-dossiers* | ISSN: 2306-1960 | This work is licensed under a Creative Commons Attribution 4.0 International License: *creativecommons.org/licenses/by/4.0/* | ITA Dossier no. 62en April 2022 | *epub.oeaw.ac.at/ita/ita-dossiers/ita-dossier062en.pdf* | Co-authors: Fuchs, D.; Udrea, T.