



# POLICY BRIEF

15 05 2024

12

## TRANSPARENCY IN THE AI ACT

STEFANO TRAMACERE



Transparency in the AI Act	
BACKGROUND AND FIELD OF APPLICATION	<p>Transparency is a <b>key principle</b> and an overarching obligation in the whole EU legislation and in particular within the Digital Strategy, but it is also an important <b>legal requirement</b> provided by national laws in different sectors to solve the <b>problem of information asymmetries</b> between parties.</p> <p>For example, this problem is evident in the healthcare sector, where transparency refers to the access to information in a comprehensible way to <b>empower patient</b> and <b>healthcare professionals</b> and to enable them to make <b>informed consent</b>.</p> <p>The new AI Act provides for specific measures to ensure <b>transparency in the design, development, deployment</b> phases for both (1) <b>high-risk AI systems</b> regulated by the Chapter III; and (2) <b>low and minimal risk systems</b> ruled by Chapter IV; and (3) <b>General Purpose AI</b> models regulated by Chapters V.</p>
HIGHLIGHTS	<p>Transparency of AI is an <b>overarching legal principle</b> that can be implemented through measures such as: <i>e.g.</i>, information provision, record keeping and documentation, auditability, traceability, explainability, interpretability.</p> <p>The AI Act concerns <b>three different categorizations</b> of transparency for different types of AI systems:</p> <p><b>Interpretability</b> for high-risk AI systems (Chapter III – Section 2, more specifically Articles 13, 14 and 86);</p> <p><b>Information/communication</b> requirements when interacting with low and minimal risk systems (Chapter IV, Article 50); and</p> <p><b>Technical documentation</b> of the model (including its training and testing process and the results of its evaluation) and <b>information</b> to enable providers of AI system <b>to have a good understanding</b> of the capabilities and limitation of <b>General-Purpose AI models</b> (Chapter V, Section 2, Article 53).</p> <p>Regarding the first category, the first paragraph of Article 13 states that <i>“High-risk AI systems shall be <b>design and develop in such a way to ensure that their operation is sufficiently transparent to enable deployers to interpret the system’s output and use it appropriately</b>”</i>. Here, the concept of transparency and the related concept of interpretability of the final output matches the techniques of <b>explainable AI</b> (XAI): if the models used are opaque, such as deep learning (DL), it is necessary to use XAI methods to explain and enable the deployer (<i>e.g.</i>, the doctor) to interpret and use them properly.</p> <p>In this regard, the paragraph 3 states <i>“that the instruction of use shall contain (iiiia) the technical capabilities characteristics of the AI system <b>to provide</b></i></p>



	<p><b>information that is relevant to explain its output</b>". In addition, to ensure human oversight of the system as also provided for in Article 14, it is necessary (among others) to <i>"put in place technical measures to facilitate the interpretation of the outputs of AI system by the deployers</i> (Article 13 (3)(d)).</p> <p>The second part of the first paragraph of Article 13 states that <i>"An appropriate type and degree of <b>transparency shall be ensured with a view to achieving compliance with the relevant obligation</b>". This implies that transparency is instrumental in ensuring an effective compliance system, and indeed the same Section provides for specific measures that contribute to respect the legal principle of transparency: to ensure proper <b>data governance</b> in the training, validation and testing phases (Article 10); an adequate <b>technical documentation</b> system (Article 11 which refers to Annex IV); a <b>record keeping</b> system to ensure the traceability of the system for the purposes for which it was developed (Article 12).</i></p> <p>In addition to <b>ex ante protective</b> measures - listed above - to ensure transparency and an <b>overall accountability system</b> to minimize the risks to health, safety and to safeguard the fundamental rights, the AI Act provides an <b>ex-post remedy</b>, heavily inspired by Article 22 of the General Data Protection Regulation (EU, GDPR 2016/679). Indeed, the regulation states that the deployer shall have <b>the right to obtain clear and meaningful explanations</b> (1) of the <b>role</b> of the AI system in the automated decision-making and (2) of the <b>main elements</b> considered (Recital 171, Article 86).</p> <p>In this respect, as well, XAI techniques can play an important role if certain opaque models are used.</p> <p>Regarding the second category which includes <b>low and minimal risk systems</b> (Chapter IV, Article 50) the AI Act states that providers shall ensure that if the AI system interacts with a natural person, that <b>person is informed that s/he is interacting with an AI system</b>, unless this is obvious by virtue of the specific context of use.</p> <p>Regarding the third category, which concerns <b>General-Purpose AI models</b>, the AI Act establishes in Article 53 (1)(a) and (1)(b)(i) that the providers shall draw up and keep up to date <b>technical documentation of the model</b>, including its training and testing process and the result of its evaluation; and to <b>make information and documentation</b> available to providers <b>who intend to integrate</b> the General-Purpose AI model into their AI systems, thus <b>to facilitate a good understanding</b> of the capabilities and limitations of the General-Purpose AI.</p>
<b>CASE LAW</b>	<p>One of the most recent and important cases concerned algorithmic transparency, which creates a direct connection between the AI Act and</p>



	<p>the data protection Regulation is <b>Shufa case of the CJEU</b> (C-634/21) regarding Article 22 GDPR. In paragraphs 44-45 of the judgement, the Court states that <i>“the concept of ‘decision’ (...) refers not only to acts which produce legal effects concerning the person at issue but also to acts which similarly significantly affect him or her”</i>, thus giving a broad interpretation of the scope of “decision”, in line with what Recital 71 stated.</p> <p>In this regard, the judgement is strictly connected with AI Act <i>ex-post</i> and <i>individual empowering</i> remedy: <b>the right to obtain clear and meaningful explanations from the deployer</b> of the role of the AI system in the decision-making procedure and the main elements of the decision taken.</p> <p>Indeed, under the GDPR, Article 13 (2)(f) states that the data controllers have to inform data subjects on the <i>“existence of automated decision-making, including profiling, referred to in Article 22 (1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject”</i>. Therefore, if the concept of a “decision” is interpreted in a similar way that the CJEU did in the <i>Shufa</i> case, the <b>Article 86 will have a wider scope</b> and will complement the GDPR’s obligations on automated decision-making.</p> <p>By virtue of this, <b>transparency and human oversight</b> obligations will have a more concrete scope, thus ensuring that <b>deployers should be better able to understand and oversee the proper functioning</b> of the decision-making process, given the <i>direct or significantly similar legal effects</i> the decision will have on the person.</p>
<p><b>IMPACT ON PROJECT</b></p>	<p>The AI Act will have a major impact on all projects, especially when using high-risk AI systems. For example, an AI system used for <b>medical research purposes</b> within the <b>BRIEF</b> infrastructure and <b>may be commercialized or put into service</b> on the European market at a later stage <b>will have to comply with all the requirements of the Chapter III, Section 2</b>, including the transparency rules. In addition, General-Purpose AI models used in a high-risk context for health, safety and fundamental rights will likely have to comply with the same rules for high-risk system, since only technical documentation of the model laid down in Article 53 may not be sufficient to guarantee adequate protection. Therefore, this could imply that to ensure AI transparency under the new regulatory framework, for instance an AI decision support system used for diagnostic purposes in the healthcare sector would have to meet the requirements for high-risk AI systems before being placed on the market or put into service.</p>