WHAT DOES TRANSPARENT AI MEAN?

When talking about ***responsible and trustworthy data driven and AI systems****,* ***transparency*** *is often mentioned as an important value to aspire to.*

However, it's not always as clear what transparency means or how you can make sure your system is transparent. It could for example refer to the ***legal aspects*** *and* ***communication of transparency****. Are the algorithms explainable? Is it clear what happens with personal data? On the other hand, it could refer to* ***an ethical value*** *that helps in improving trust in the system.*

Both are important to ***mitigate bias*** *in AI systems,* ***diminish damaging assumptions*** *about the technology and* ***guarantee responsible systems****.*

However, too much transparency can also lead to ***information overload****,* ***more questions*** *and even* ***distrust*** *among users. The development of transparent systems is therefore a difficult balancing act.*

In this brAInfood we will elaborate on a few ***different types of transparency*** *and how you best approach this* ***balancing act*** *of developing transparent systems.*

brAInfood of the Knowledge Centre Data & Society

TYPES OF TRANSPARENCY

The **different types of transparency** are **instrumental** in enabling users to adopt a technology or exercise their rights. Below, we briefly discuss **four perspectives on transparency**, from the point of view of **three different domains**:

1 GENERAL DATA PROTECTION REGULATION (GDPR)

The GDPR is a European regulation that contains rules for companies and organisations on the collection, storage and management of **personal data**. The third chapter (Articles 12 to 23) must promote transparent information and communication about **what data is processed**, **by whom and why**. This enables data subjects to exercise their **rights**, including the right to information, the right to inspection, and more. Companies and organisations must communicate to data subjects, among other things, the purpose of the data collection and the recipients with whom the data is shared.

2 HUMAN-COMPUTER INTERACTION (HCI)

HCI is a field of study that examines the **interaction between people and machines** and the **design of computer technology**. Research on transparency in HCI includes two types of transparency:

- **Prospective transparency** (information obligations, such as GDPR)
- **Retrospective transparency** (post-hoc statements and elements such as inspectability, traceability and auditability)

HCI has a strong focus on the **potential end users**. It studies their wishes and needs regarding system transparency. This is because the perception and attitude towards transparency differs from one end-user to another, and according to the type of technology or service and the user context.

3 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HLEG)

The European Commission's AI HLEG recommends pursuing transparency through three core concepts:

- **Traceability:** by documenting the process of data collection, data labelling and the algorithms used in this process.
- **Explainability:** via three components, namely
  - Technical explainability (i.e. the functioning of the AI system).
  - User-centric focus (i.e. an explaination of the decision-making process of the AI system).
  - Explainability of the business model (i.e. the design choices and purpose of the system).
- **Communication:** by informing users when they interact with an AI system and giving them the freedom to choose whether to interact with a computer or with a human. In addition, users are given information about the capabilities and limitations of the AI system.

According to the AI HLEG expert group, **trust** in an innovation can be created by **transparent communication** between technology provider and end user. When communication is too vague or too specific, it often creates distrust.

METHODS

How can you gain insight into the context and the users for whom transparency is important? With the help of **personas,** or other **user research methods** (e.g. interviews with end users).

In a persona, you describe a potential user (name, age, profession, characteristics, etc.) and his/her transparency needs. What information does he/she want to receive and why? Test these personas against your innovation during the development process and answer the transparency needs of your end users. Good luck!

Knowledge Centre Data & Society (2021). What does transparent AI mean? brAInfood of the Knowledge Centre Data & Society. Brussels: Knowledge Centre Data & Society.