

Developing a Human Centric Architecture

25-11-2021

ASSISTANT, being executed by 12 partners from 7 nations, has received funding from European H2020 programme



Content

- Where do we come from?
- Developing a Human Centric Architecture
- ART-principles
- Reflections

ASSISTANT

ens European New School
of Digital Studies



ASSISTANT, being executed by 12 partners from 7 nations, has received funding from European H2020 programme



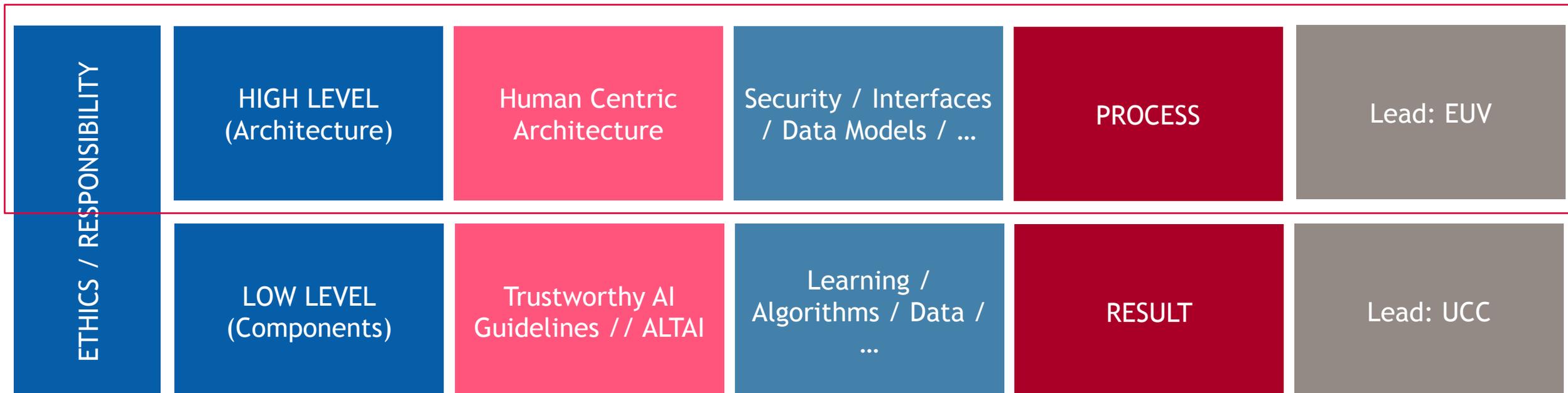
Human-centric architecture and responsible AI

- Calls for human-centric, responsible AI cannot be overheard
 - e.g. > 170 frameworks/guidelines, Algorithmwatch 2021)
- **Challenge: Translate abstract values and into concrete practices in the development process.**
 - Rather abstract approaches are hardly translated into concrete results in the development process, as they are not "put into practice" (Hagendorff, 2020a, p. 1)



Different Approaches within ASSISTANT

- focus on the result
 - e.g. evaluation using an abstract framework (Trustworthy AI, High-Level Expert Group on AI. 2019)
- **focus on the process**
 - **e.g. using ART principles (Dignum 2019) for responsible AI (Aler Tubella et al 2019) design and development**



Human Centric Architecture

- Human Centric: Interfaces of the machine with focus on humans (Nowak et al., 2018)
- vs Human Centered: Including humans in the design process to ensure that products are built for humans (Giacomin, J. 2014)



Human Centric Architecture

- Ethics / Guidelines / Frameworks -> focus on Responsibility (Responsible Research and Innovation)
- Issue: Translating abstract frameworks into concrete results in the development process, as they are not “put into practice” (Hagendorff2020: 1)



Success criteria

Our expectation towards an approach within the ASSISTANT project is defined by the following criteria. Our approach has to

- enable explicit deliberations about values.
- offer potential to actually improve processes in which AI systems are produced.
- enable reflections about potential biases and different perspectives of the people involved.
- be connected to broader discussions in the field.
- be concrete and offer tangible instructions for the actors working in the ASSISTANT project.
- reflect specificities of the manufacturing sector in connection to artificial intelligence.
- function as a blueprint for others that want to adopt our work for their projects.

ART-Principles

- One specific attempt to translate these rather abstract dimensions from the frameworks into concrete steps during the development phase are the **ART-principles** (Dignum 2019a/2019b).
- These principles allow us to address issues of accountability, responsibility, and transparency within the development process of ASSISTANT.

“Accountability refers to the requirement for the system to be able to explain and justify its decisions to users and other relevant actors.” This means that the system needs to be able to be held accountable in relation to humans that are interacting with and affected by it. Therefore, decisions need to be explainable *after* they have been taken.

“Responsibility refers to the role of people themselves in their relation to AI systems.” Responsibility is different from Accountability in the way that it is focusing on the people involved and not related to the content of the decision: It links to questions of liability on the one hand but also to who is capable of behaving morally. Questions of responsibility could be: Who is delegating which decisions to the system and how are decisions supervised? The responsibility dimension encourages reflections about the role of different persons within the process of decision making and system development.

“Transparency indicates the capability to describe, inspect and reproduce the mechanisms through which AI systems make decisions [...]” (Dignum 2019). It is therefore a precondition to determine responsibilities and to hold the responsible people accountable. Transparency increases trust, as people do not only have to trust but can ground their faith on a sophisticated understanding of how algorithms work. Making the algorithms transparent allows stakeholders to criticize what is going on. Transparency is different from accountability in the way that is not necessarily linked to one specific situation that is evaluated ex-post but includes a more general need for openness.

EXAMPLE: Responsibility and Problem of Many Hands

- “*Responsibility* refers to the role of people themselves in their relation to AI systems.”
- When many people are involved in an activity, it is often difficult, if not impossible, to pinpoint who is morally responsible for what, a phenomenon known as the ‘problem of many hands.’ (Van de Poel et al 2015)
- → Develop a Responsibility map with the partners (workshop) that tells about the different responsibilities of the different components
- → Need for new methods and workshop concepts to produce concrete inputs for the Human Centric Architecture

Reflections

- Not applying but developing frameworks and approaches of Human Centric Design
- Which questions to ask at which level (level of abstraction)?
- Synchronisation of requirements and human centric design - when is the right moment for which “interventions”?
- (COVID-situation and difficulty of in-person-meetings)

Thank you!

ASSISTANT, being executed by 12 partners from 7 nations, has received funding from European H2020 programme



More examples

- **Interoperability as negotiation** → *It is therefore necessary that the deliberations and decisions about the architecture and technical implementation are done in a moderated environment to ensure the fair exchange of arguments and to ensure that the implementation is chosen that works best altogether for the entire system and especially for its users.*
- **The early nature of requirements** → *It is therefore necessary to stress that the requirements that are collected in a very early phase of the document remain flexible. This way, insights from the human-centric architecture and the discussion it facilitates can be included as requirements that are equally to be fulfilled as the functional requirements.*
- **Making decisions of the system explicit** → *It is therefore necessary to develop clear decision trees and visualizations of decision processes that are then delegated to algorithms in the further course of the project.*
- **Operationalizing transparency by keeping privacy** → *A strategy for understanding how a concrete decision was taken by the system needs to be developed. The output of this strategy could be a visualization that is not only traceable in cases of conflicts but also presented to the user.*