Nataliya YAKYMETS
nataliya.yakymets@cea.fr

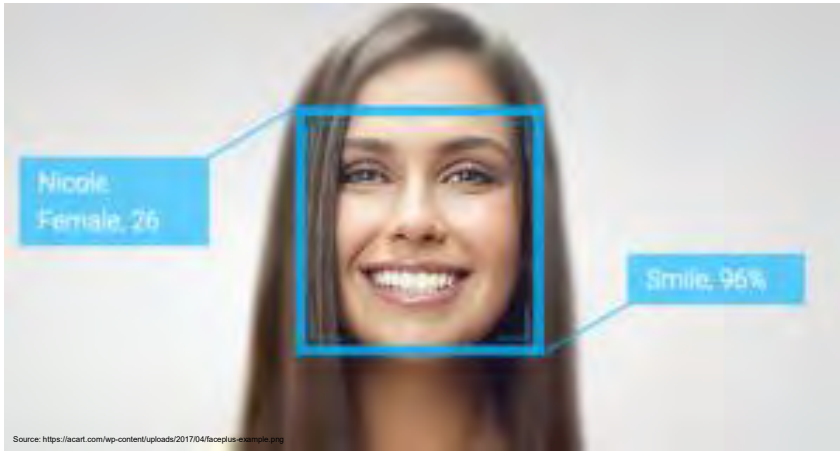
TOWARDS SAFE-BY-DESIGN ARTIFICIAL INTELLIGENCE SYSTEMS

SafeComp2019

Safe-by-Design is the concept of applying methods to minimize hazards early in the design process.

CONTEXT

- AI-based systems help to solve challenging issues from cancer detection to image understanding and natural language processing.



- On 8 April 2019, the High-Level Expert Group on AI (HLEG AI) presented Ethics Guidelines* for Trustworthy Artificial Intelligence: **“AI systems need to be resilient and secure. They need to be safe…”**

* https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

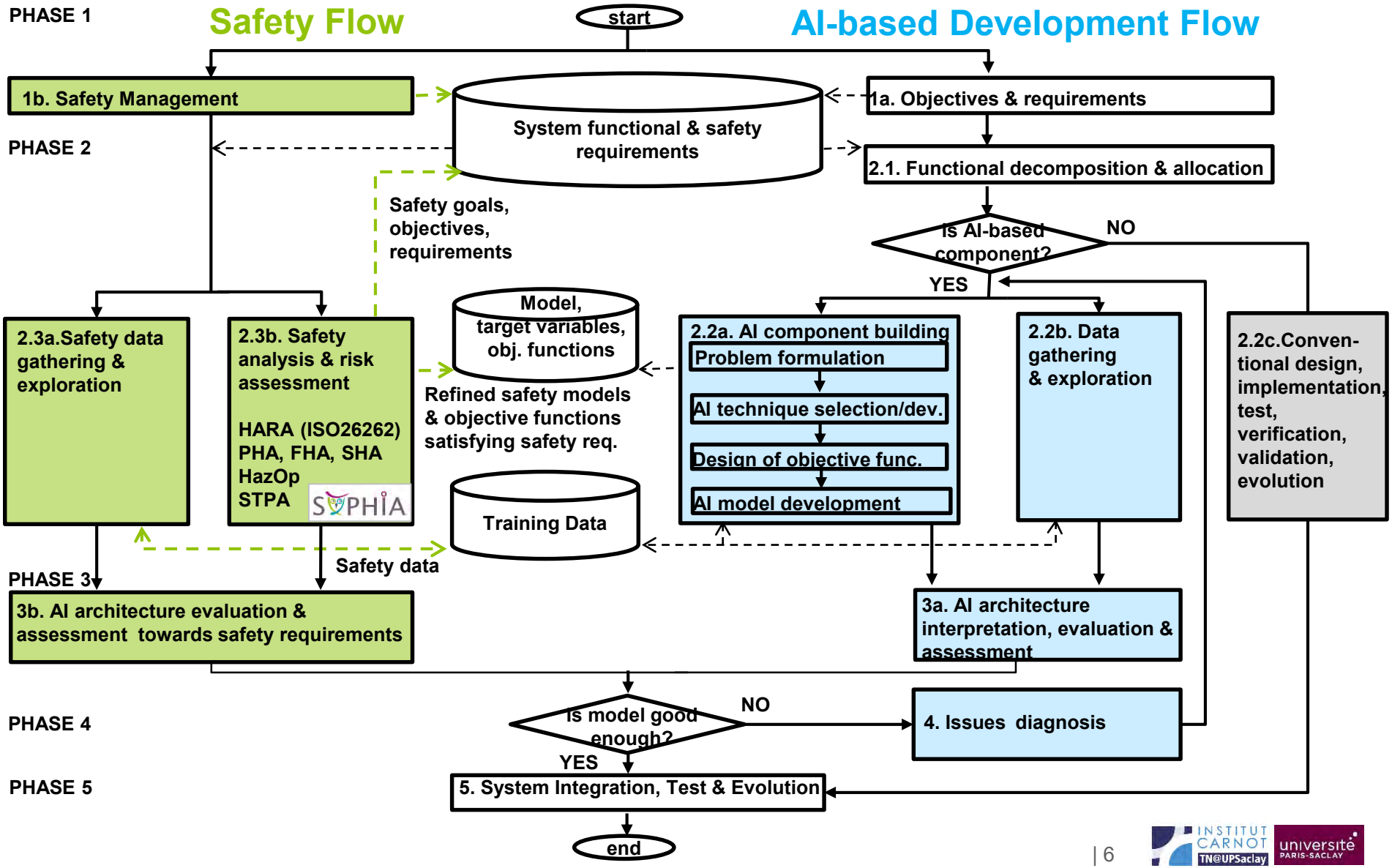
CHALLENGES

- **How to develop AI-based systems?**
  - Lack of standards describing development life-cycle of AI-based systems

- **How to ensure safety of AI-based systems?**
  - Lack of methodological support for safety analysis of AI-based systems

Propose a methodology for coupling the development of AI-based systems with safety analysis.

METHODOLOGY



OPEN QUESTIONS

- **Accuracy** and **completeness** of safety analysis:
  - It is performed on data that may not be representative of the infinite number of scenarios the systems can face.

- **Controllability** of AI systems is limited as the human is no longer in the loop.

FURTHER WORK

- Refine the methodology to more particular AI techniques (e.g. deep learning, cognitive computing).
- Adapt the classical safety methods (PHA, HARA, SHA, HazOp, etc.) to the context of specific AI techniques.
- Validate the methodology through application on industry-relevant case studies.