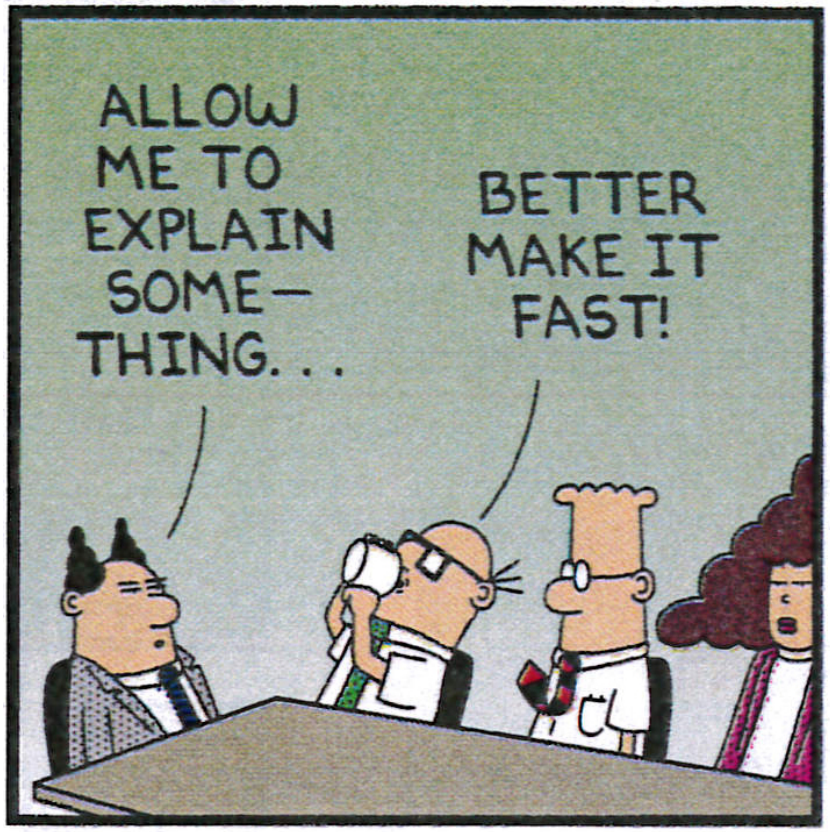
Conceptual Model and Measurement Issues for Explainable AI

October 2019
HFES

Outline

1. Key Concepts
2. Measurement Scales
3. Conceptual Models
4. Measurement Concepts



ihmc

Literature Review

Classical Philosophy (Causation)
Cognitive Psychology
Developmental Psychology
Social Psychology
Philosophy of Science
Psychology of Science
Human Factors

591

then

152 more

ihmc

Key Concepts

Importance of the “Mental Model” Concept

The system needs a good model of the user (knowledge/experience, goals, context, minimum necessary information)

Explaining as an Interactive Process vs. Explanations as Statements

The user needs a good mental model of the AI; the AI must be aware of user’s goals.

The XAI must help users identify/highlight boundary conditions and thereby engender appropriate trust and reliance.

ihmc

Psychometrics

Explanation Goodness - Criteria in the Literature

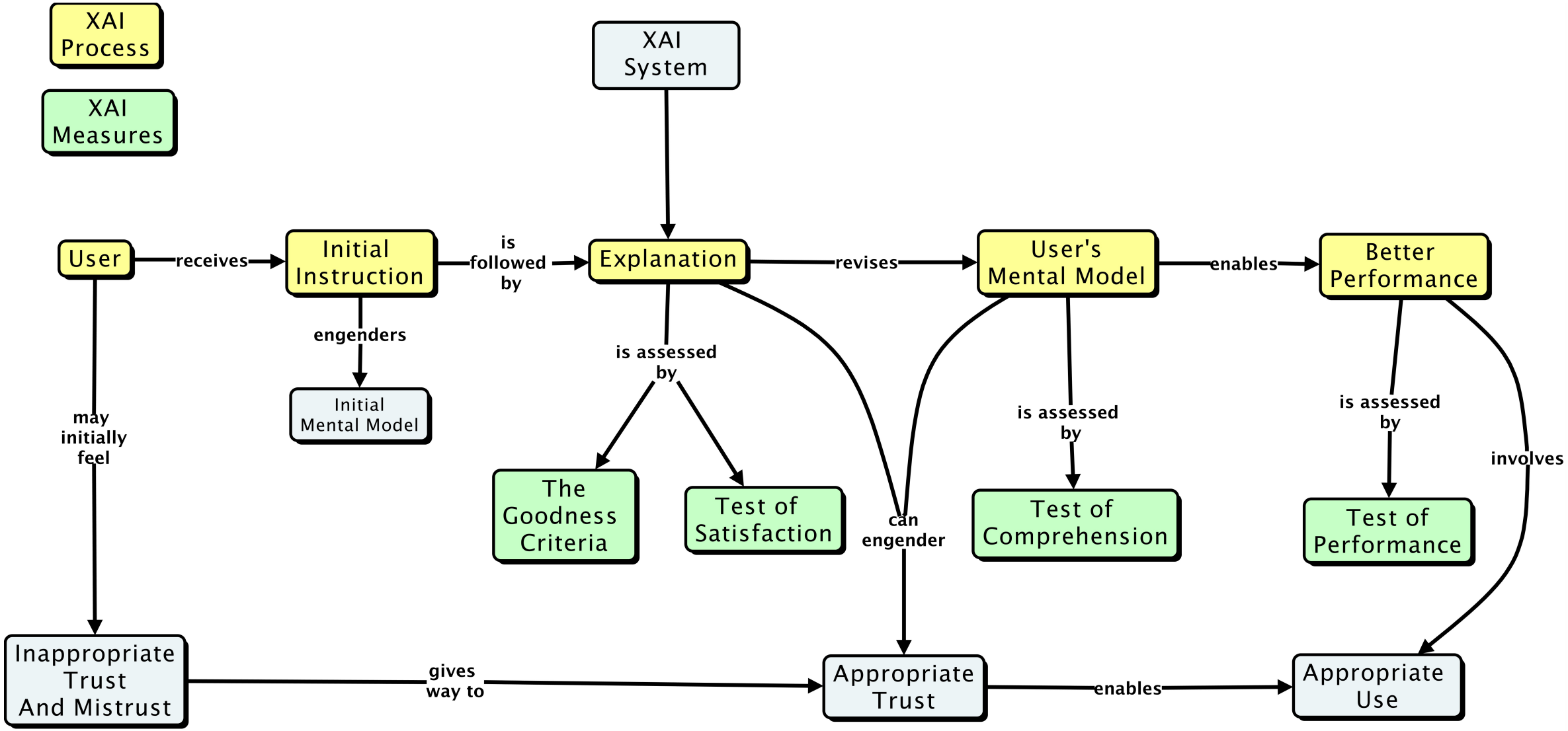
Veridicality or Fidelity
Completeness
Succinctness
Dimensions of Variation
Predictability

Explanation Satisfaction – User Judgments

Understandable
Satisfying
No Irrelevant Details
Usefulness
Instills Trust

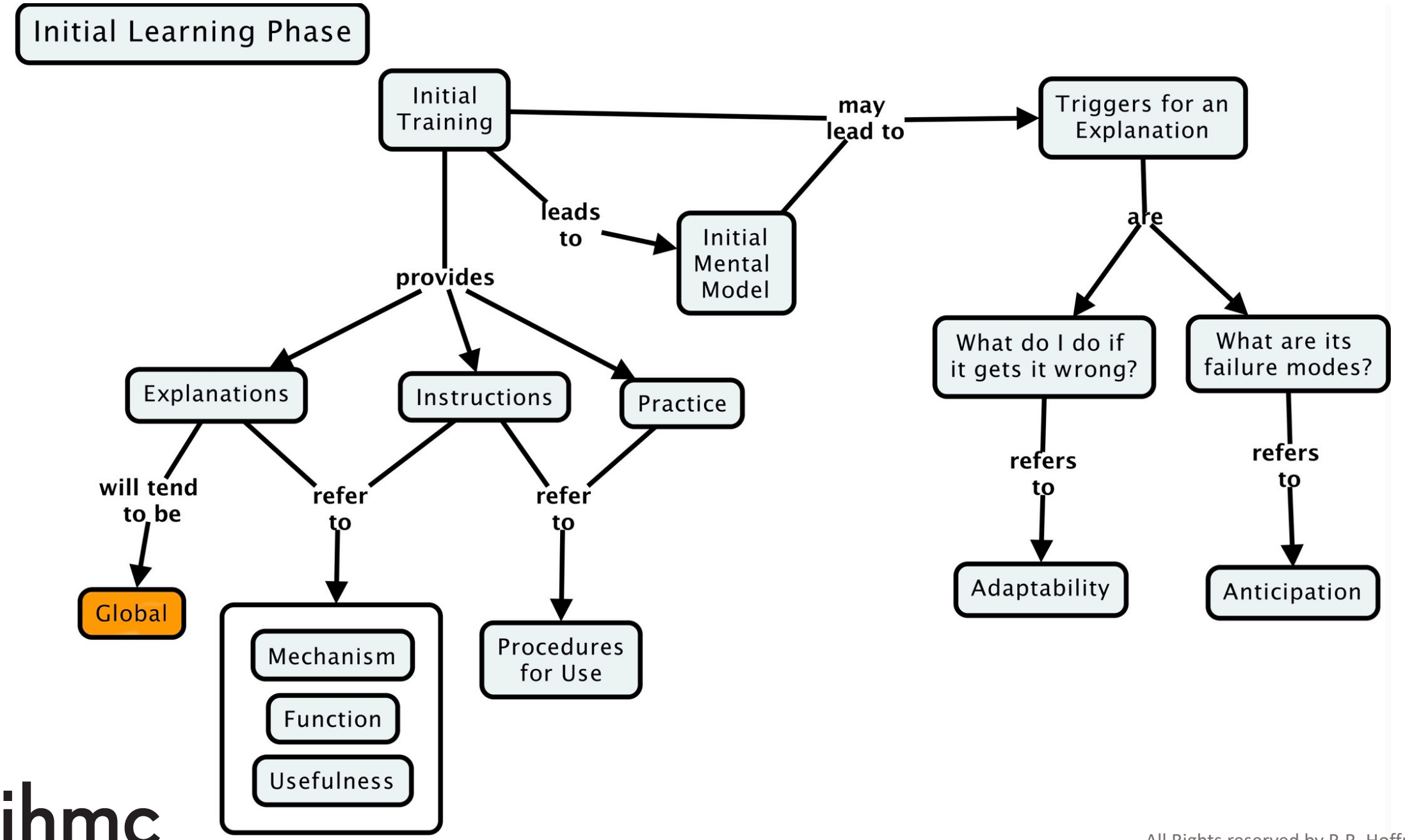
ihmc

How the Classes of Measures Fit into the XAI Program

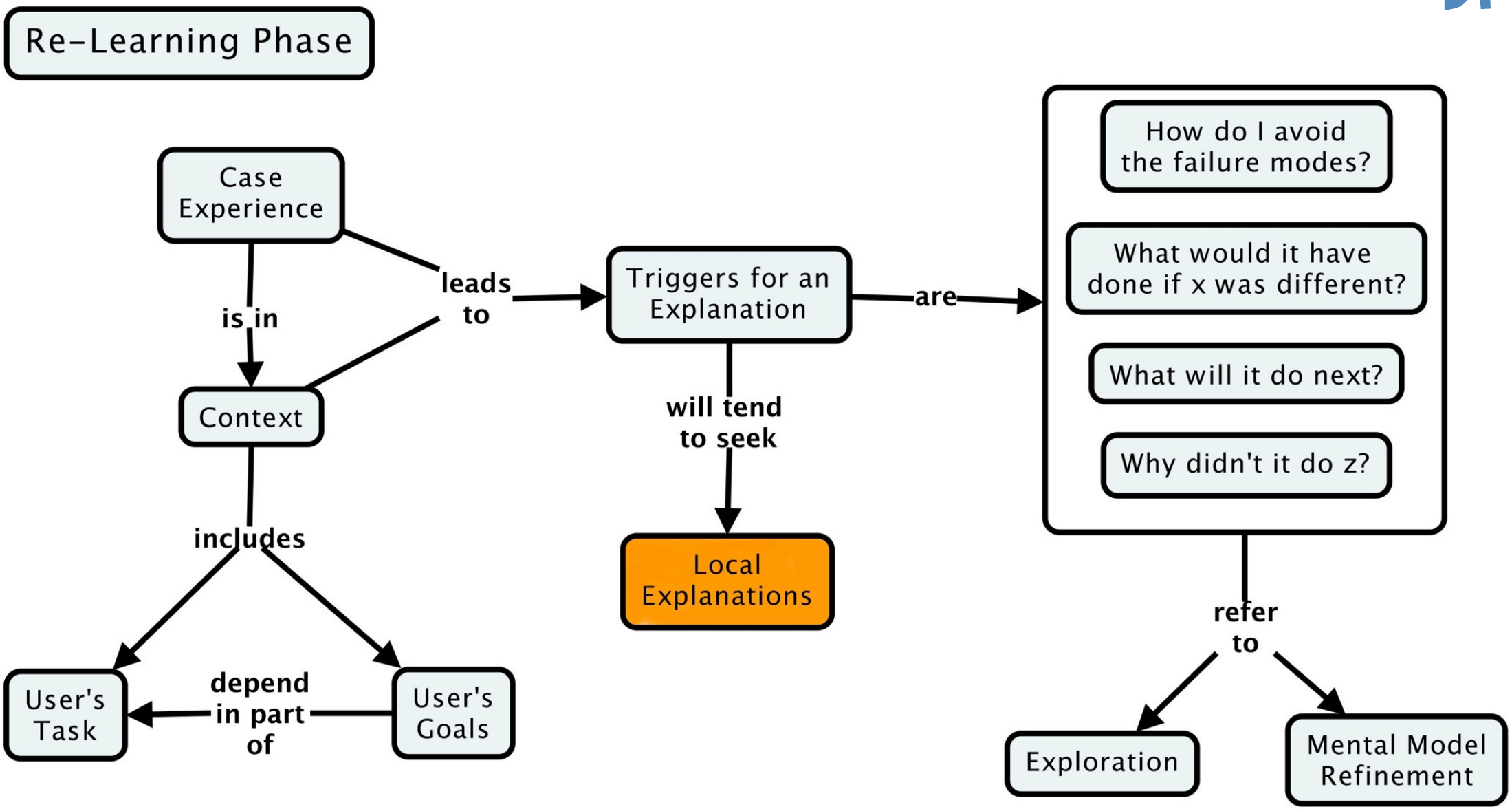


Conceptual Process Models

ihmc



ihmc



Some Concerns

- AI ≠ DN and ML
- Heat Maps have nothing to do with heat
- Data visualizations do now show what the machine “pays attention to”

ihmc

Some Concerns

- XAI was already addressed in the second generation of Expert Systems and Intelligent tutoring Systems. Those lessons learned have been forgotten.

- The focus of current XAI systems on local explanation does not address the core problem:

  Robust, defeasible, context-sensitive inference over world knowledge.

Good luck with that !

ihmc

Thank you!

rhoffman@ihmc.us