



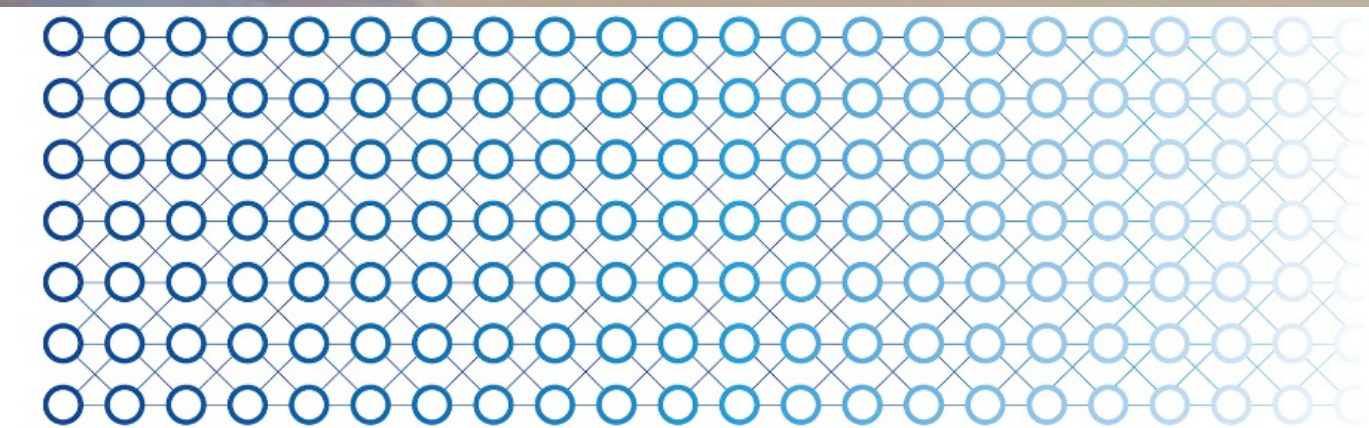
Operationalizing Reproducibility

Odd Erik Gundersen, dr. philos.

Chief AI Officer, Aneo AS

Associate Professor, NTNU

ANEEO



Norwegian Open AI Lab

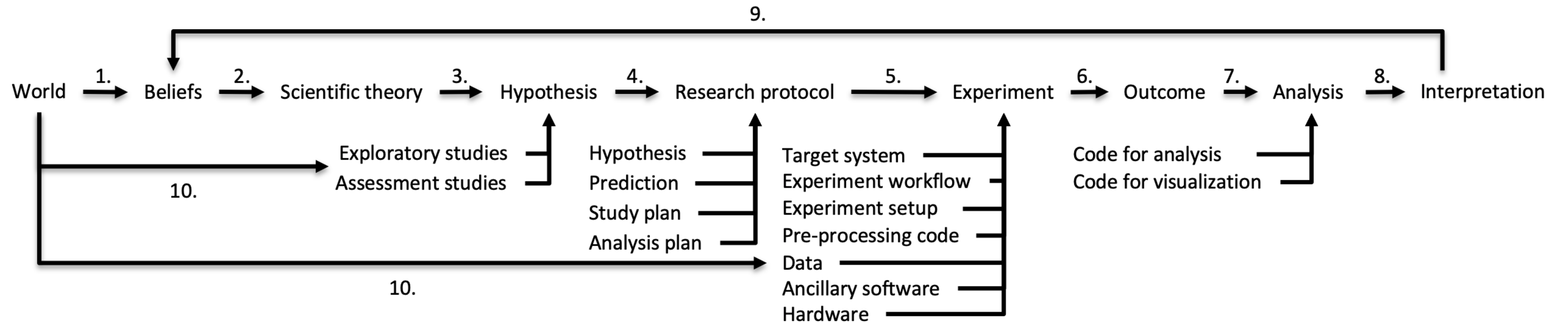
Why do we need a shared definition?

- To properly **capture** the concept.
- Create a **shared** understanding.
- Use it for **teaching** science.
- Use it when we **practice** science – to guide our research.
- Use it when **evaluating** research done by others.
- Help point out **limitations** of conclusions or what went wrong.

What characterizes a good definition?

- It should be a **mental model** that can easily be looked-up
- It should be **simple** to understand.
- It must have enough **depth** to be valuable.
- It should enable us to **operationalize** our understanding and help us design and evaluate experiments.
- It needs to be tightly **connected** to the scientific method, as reproducibility is “*a cornerstone of science*”.

The Scientific Method in Empirical ML

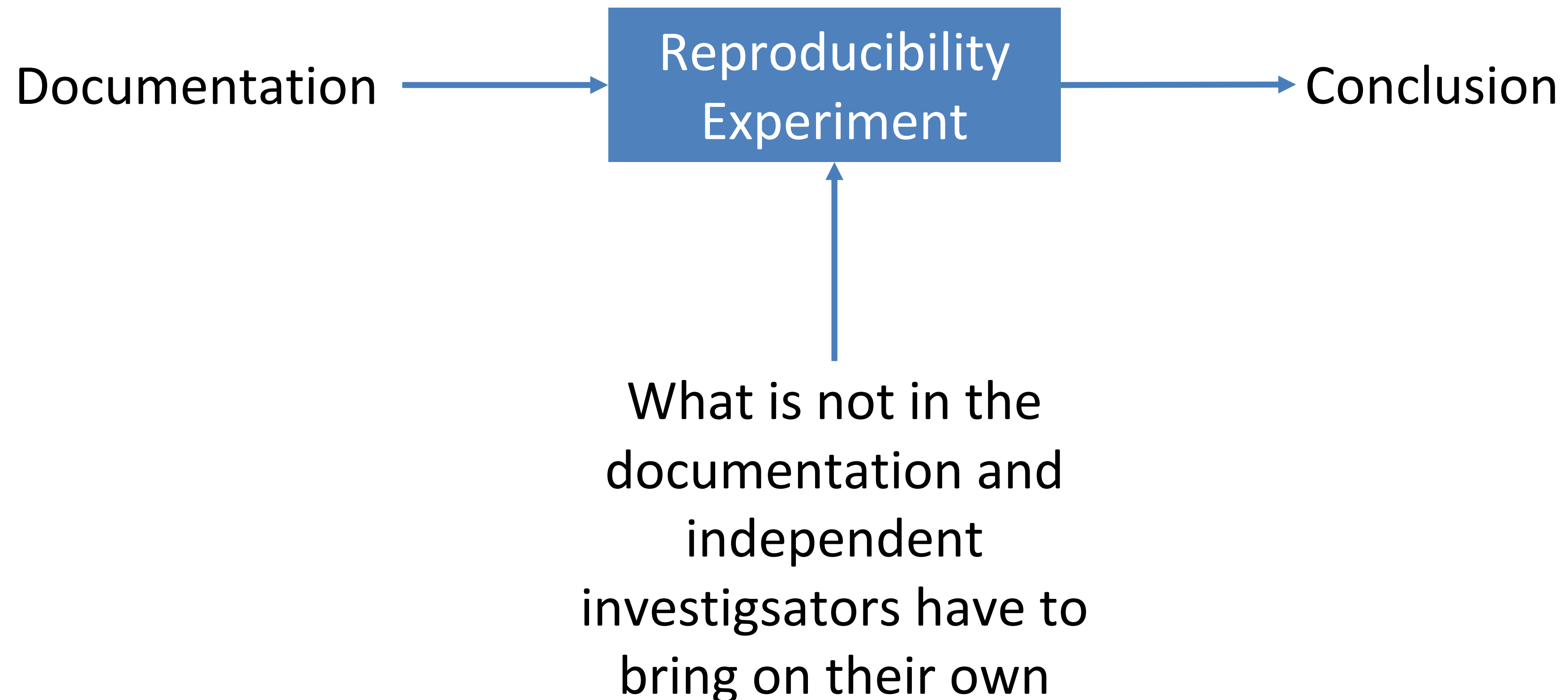


Reproducibility

Definition. *Reproducibility is the ability of **independent investigators** to draw the same **conclusions** from an experiment by following the **documentation** shared by the original investigators.*

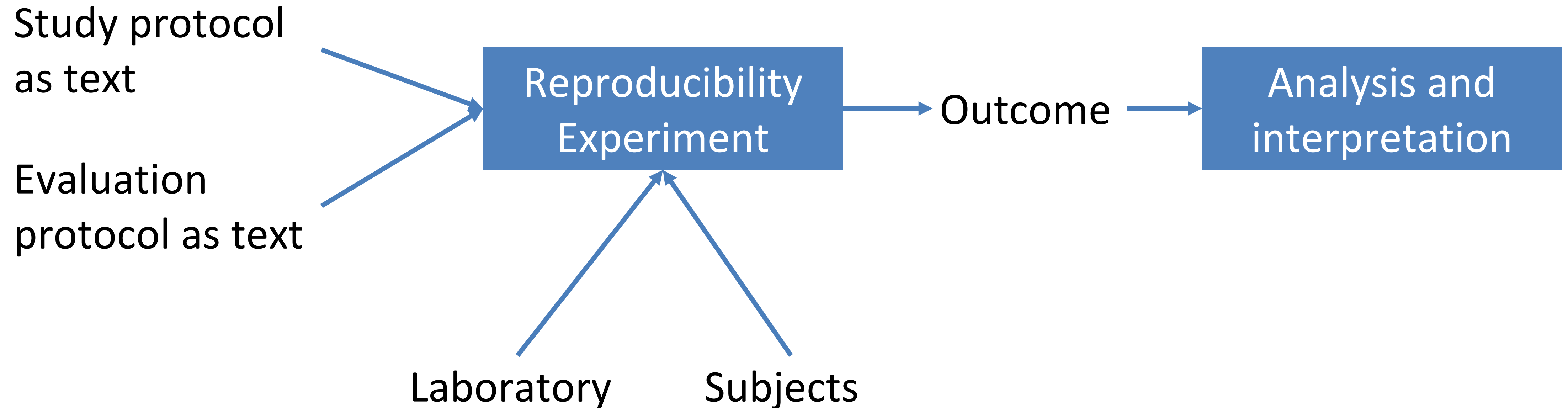
Reproducibility Experiment

Definition. *Reproducibility is the ability of **independent investigators** to draw the same **conclusions** from an experiment by following the **documentation** shared by the original investigators.*



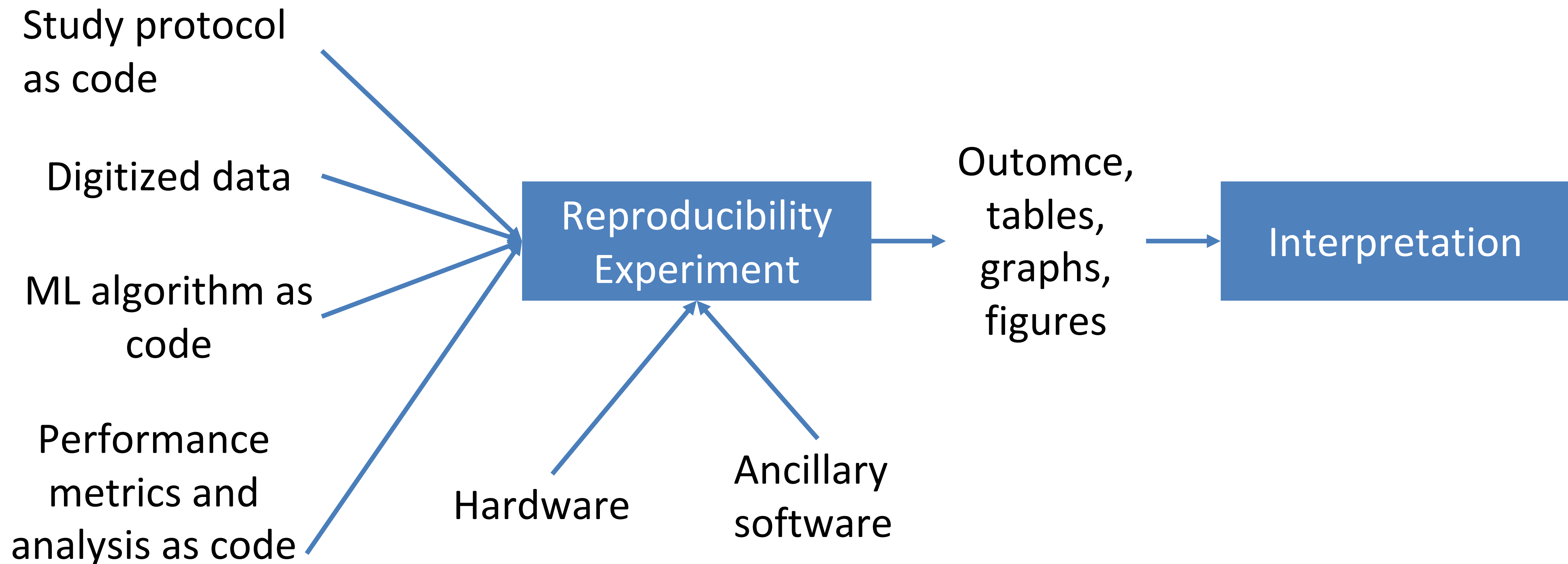
Reproducibility - Psychology

Definition. *Reproducibility is the ability of **independent investigators** to draw the same **conclusions** from an experiment by following the **documentation** shared by the original investigators.*



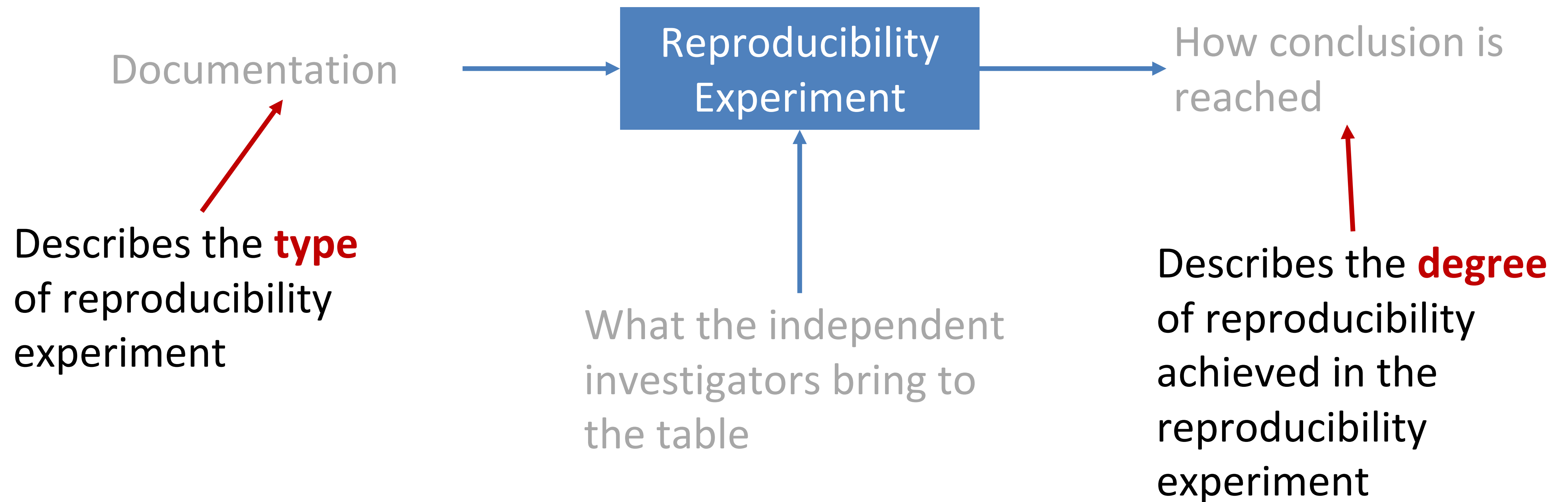
Reproducibility Experiment – ML

Definition. *Reproducibility is the ability of **independent investigators** to draw the same **conclusions** from an experiment by following the **documentation** shared by the original investigators.*



Reproducibility - Dimensions

Definition. *Reproducibility is the ability of independent investigators to draw the same conclusions from an experiment by following the documentation shared by the original investigators.*



The Three Types of Documentation

Description. *Description of the AI method implemented by the AI program, the experiment being conducted and the analysis of the results as well as the hardware and ancillary software used for conducting the experiment.*

Code. *AI Program code, code for setup and configuration, code controlling workflow, code for analysis of results and visualization.*

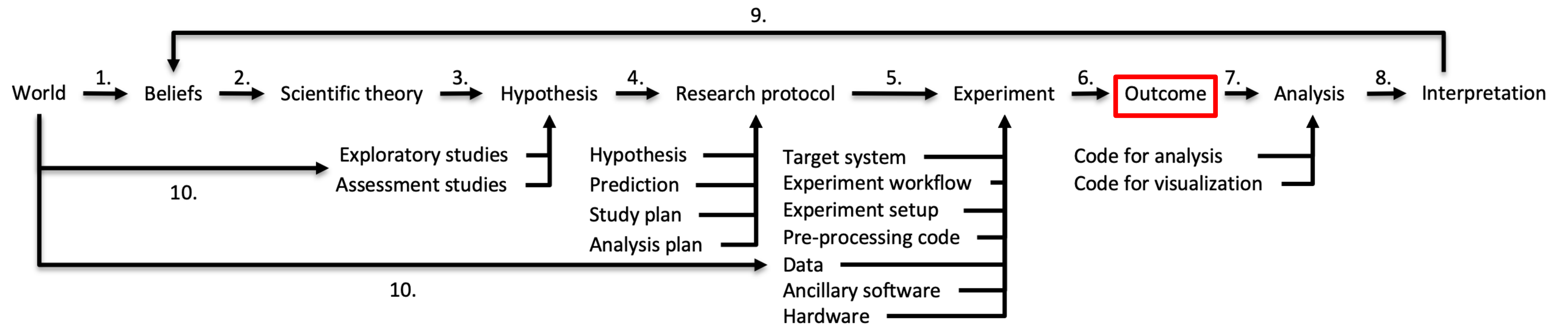
Data. *All data used for conducting the experiment. Are the samples used for training, validation and test specified? What about the results?*

Types of Reproducibility Experiments

	Text	Code	Data
R1 Description			
R2 Code			
R3 Data			
R4 Experiment			

Outcome Reproducible - Degree

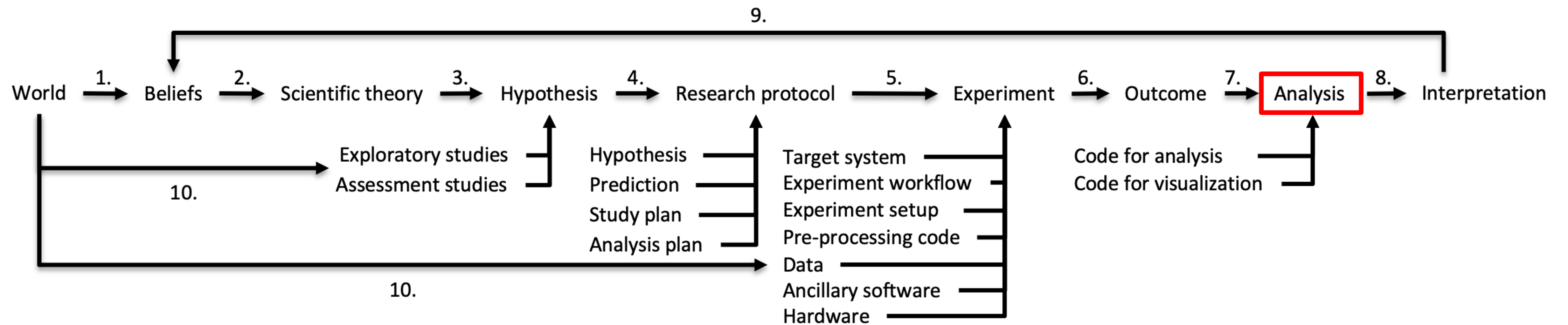
Definition. *Reproducibility is the ability of independent investigators to draw the same **conclusions** from an experiment by following the documentation shared by the original investigators.*



Outcome reproducible. *The outcome of the reproducibility experiment is the same as the outcome produced by the original experiment.*

Analysis Reproducible – Degree

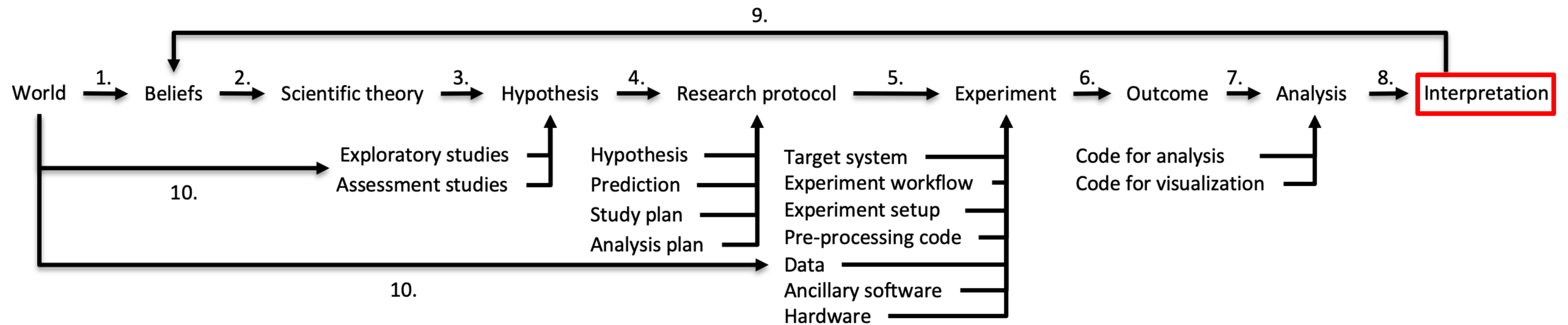
Definition. *Reproducibility is the ability of independent investigators to draw the same **conclusions** from an experiment by following the documentation shared by the original investigators.*



Analysis reproducible. *Outcome might differ, but same analysis and interpretation on different outcome leads to same conclusion.*

Interpretation Reproducible - Degree

Definition. *Reproducibility is the ability of independent investigators to draw the same **conclusions** from an experiment by following the documentation shared by the original investigators.*

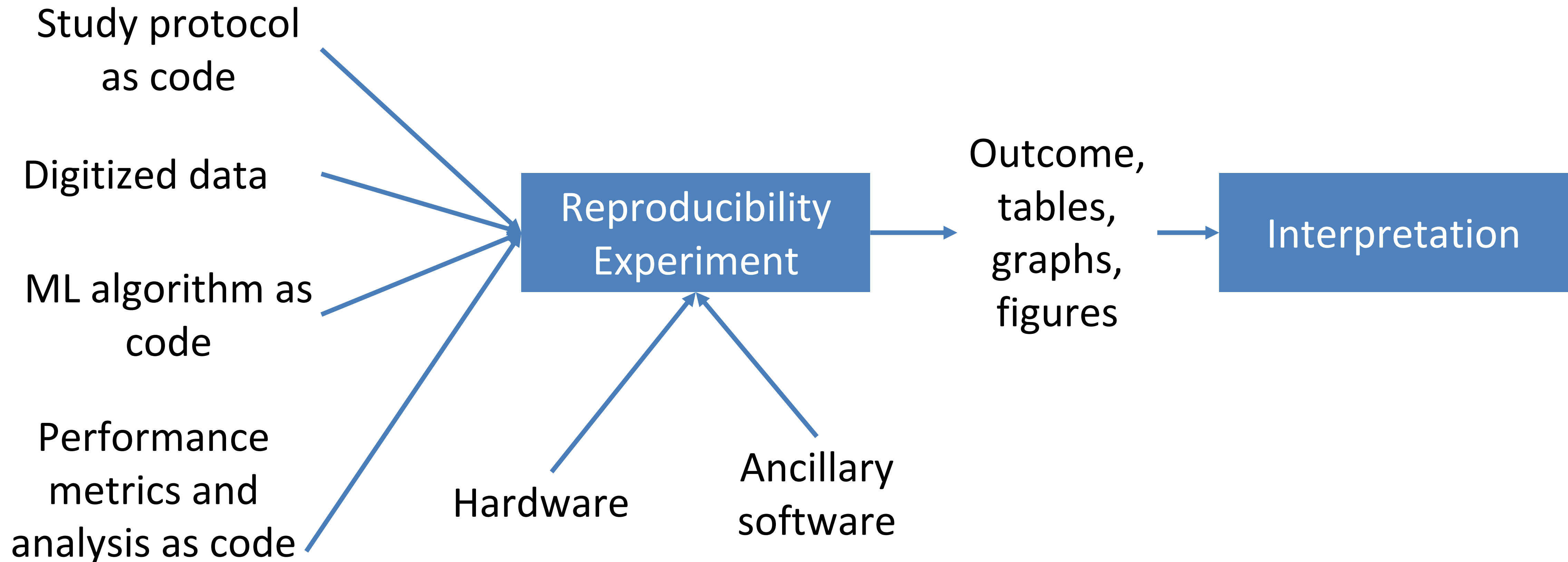


Interpretation reproducible. *Neither the outcome nor the analysis need to be the same if the interpretation leads to the same conclusion.*

Reproducibility Experiment Classification

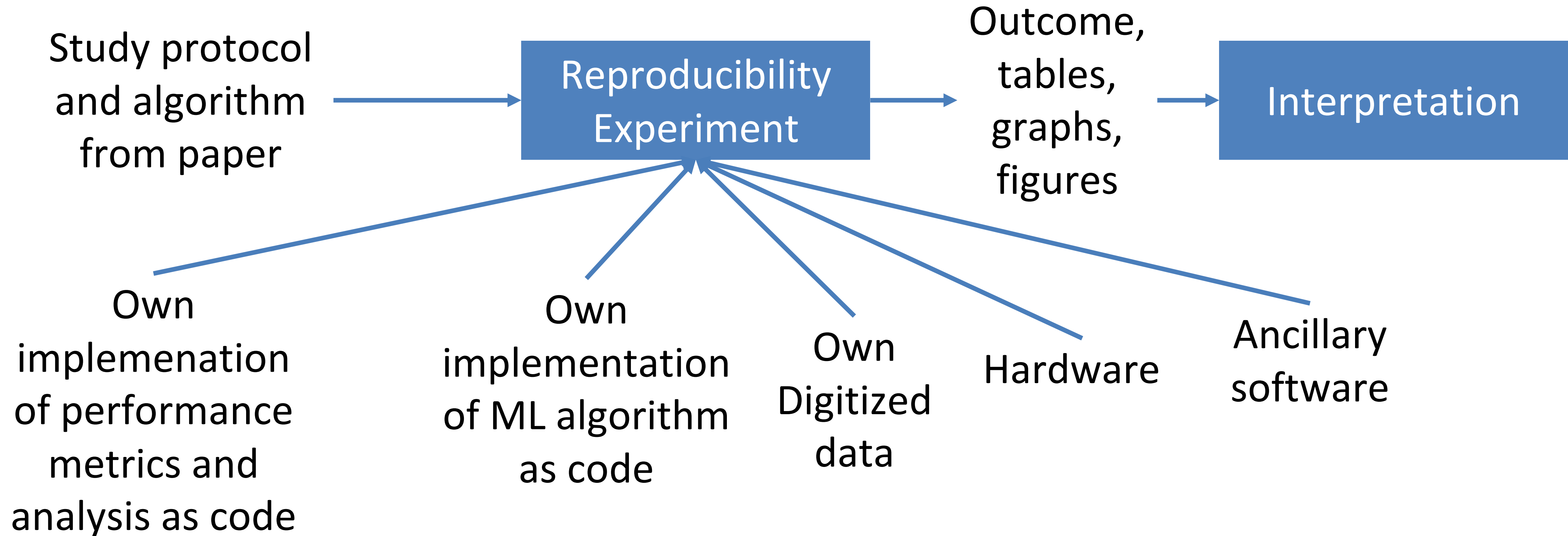
	R1 Description (text)	R2 Code (text+code)	R3 Data (text+data)	R4 Experiment (text+code+data)
Outcome Reproducible	OR1	OR2	OR3	OR4
Analysis Reproducible	AR1	AR2	AR3	AR4
Interpretation Reproducible	IR1	IR2	IR3	IR4

Operational use – Example I



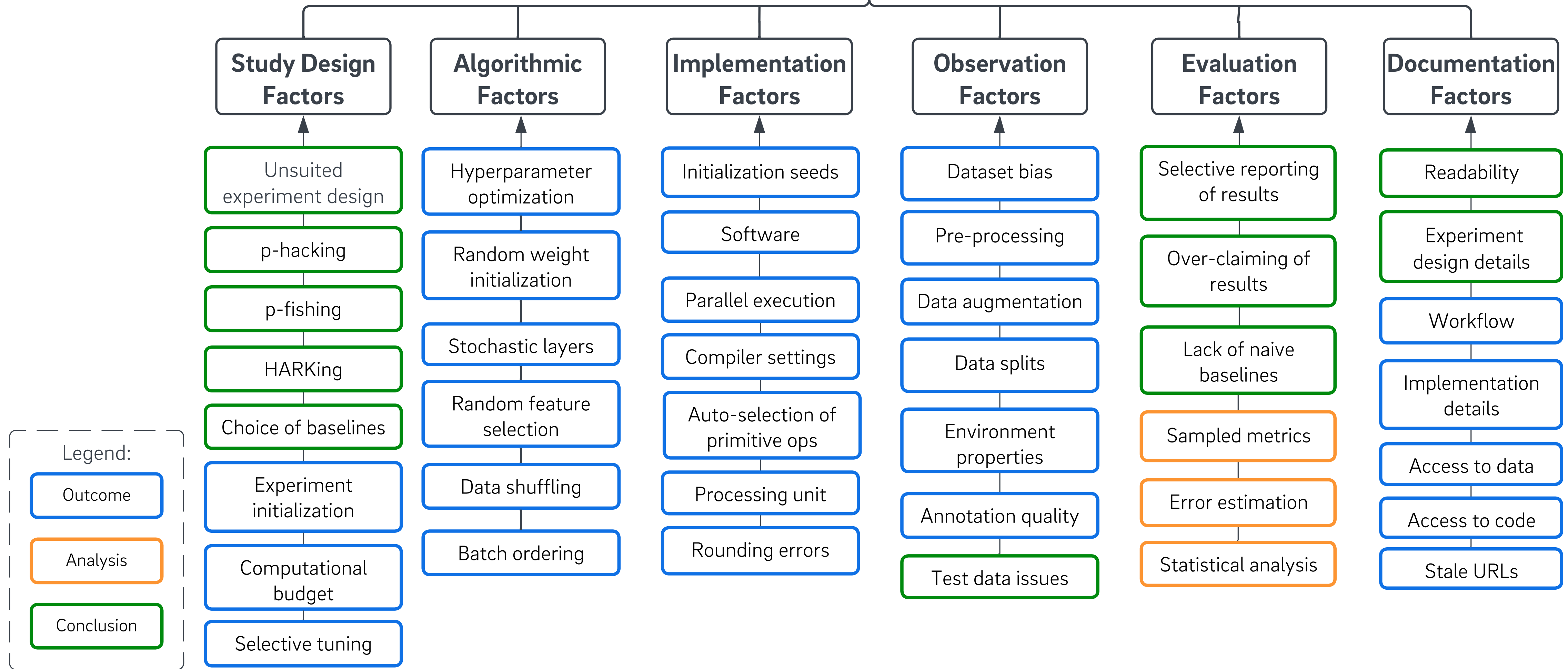
OR4: *Outcome reproducible after executing same experiment* – Same outcome after executing code on same data on **different** hardware and software.

Operational use – Example II



IR1: *Interpretation reproducible after own implementation - Different outcome and analysis after implementing experiment fully from paper utilizing different data and different metrics and analysis methods.*

Sources of Irreproducibility



Conclusion

The definition ...

- ... is simple **simple**
- ... but has **depth**
- ... is defined in **relation** to the scientific method
- ... is **operationalizable** when describing, conducting and evaluating experiments

Research

- **State of the Art: Reproducibility in Artificial Intelligence** O. E. Gundersen and S. Kjensmo, AAAI 2018
- **On Reproducible AI** O. E. Gundersen, Y. Gil and D. W. Aha, AI Magazine, Fall 2018.
- **Standing on the Feet of Giants** O. E. Gundersen, AI Magazine, Winter 2019.
- **Out-of-the-box Reproducibility: A Survey of Machine Learning Platforms** R. Isdahl and O. E. Gundersen, eScience 2019.
- **The Reproducibility Crisis Is Real** Gundersen, O. E., *AI Magazine*, 41(3), 103-106, 2020.
- **The Case Against Registered Reports**, O. E. Gundersen, AI Magazine, Spring 2021.
- **Do Machine Learning Platforms Provide Out-of-the-box Reproducibility?** O.E. Gundersen, S. Shamaliei and R. Isdahl. *Future Generation Computer Systems*, Volume 147. Elsevier, 2022.
- **Operationalizing Machine Learning Reproducibility through a Categorization of Experiment Design Choices.** O. E. Gundersen, K. Coakley, C. Kirkpatrick, Gil, forthcoming.
- **What We Learned When Reproducing the Most Cited AI Research**, O. E. Gundersen, O. Cappelen, N. Grimstad, M. Mølne, forthcoming.



SUCCESSFUL RESEARCH IN AI

Standing on the Feet of Giants

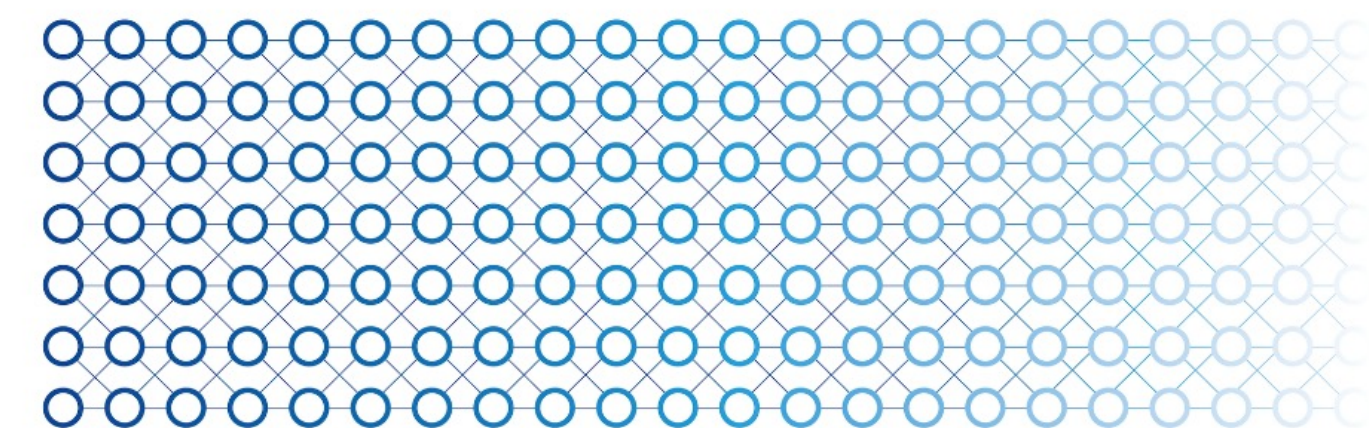
Odd Erik Gundersen, dr. philos.

Chief AI Officer, Aneo AS

Associate Professor, NTNU

odderik@ntnu.no

ANEO



Norwegian Open AI Lab