

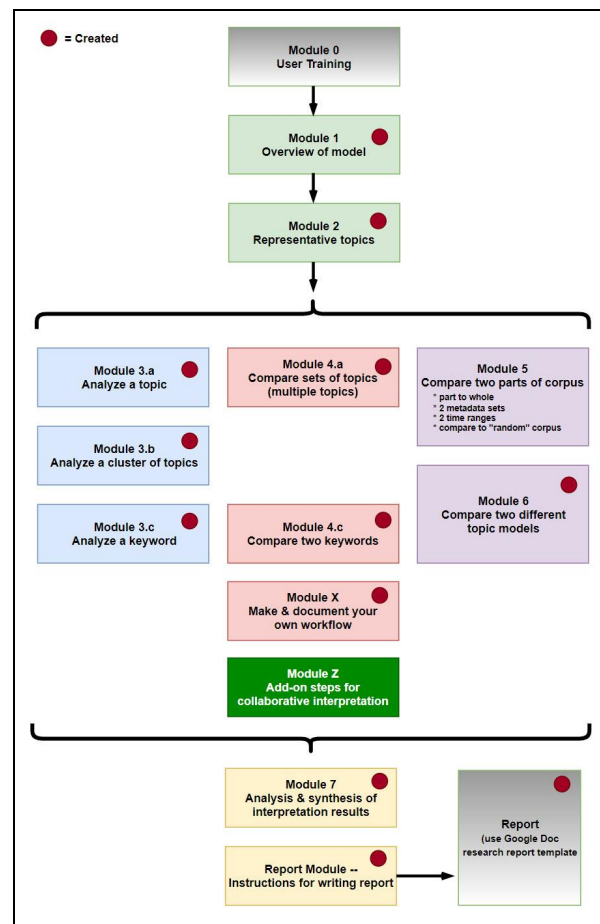
WE1S pioneers an “interpretation protocol” for guiding how humans learn from machine learning.

Machine-learning, artificial intelligence, and other algorithmic analysis methods today can be mysterious “black boxes.” How did the computer do that? “Interpretability” or “explainability” studies in computer science is thus a booming field.

But just as important is the human side of the problem: how humans learn from--and interpret or explain--machine learning. In much research using machine learning today, it all seems like an *abracadabra* magic trick. The researcher looks into the black box and says: “Trust me. This is the pattern or meaning I see in the results. And, by the way, here is a cherry-picked example and one graph I’ve pulled out of the hat.”

Of course, there is no certain or all-purpose way to regularize how humans learn from machine learning. But WE1S innovates an [Interpretation Protocol](#) for reading topic models as a paradigm for declaring *some* method that can then be followed or revised by others. The WE1S Interpretation Protocol guides readers of topic models through observation waypoints (features of a model to look at), analysis procedures (making use of visualization tools), topic labeling, and iterative note taking—culminating in the writing of research reports. The Protocol consists of individual modules (e.g., “Analyze a topic” or “Analyze a keyword”) that can be combined into flexible workflows for particular research questions.

A precedent in the social sciences--which has long wrestled with closing the gap between quantitative and qualitative analysis--is the “grounded theory” method for studying data and procedurally drawing conclusions from it.



Current modules in the WE1S topic modeling [Interpretation Protocol](#) (versions available as Qualtrics surveys and Word documents)

Resources

Interpretation & Interpretability research: See WE1S Bibliography on [Interpretability & Explainability in Machine Learning](#) and [Topic Model Interpretation](#).

“Grounded Theory” method: See WE1S Bibliography on [“Grounded Theory”](#).