

LLMs and Code Readability: Theories

A Research Agenda

- Today is all discussion.
- We've seen several different approaches regarding readability.
- Today, we will try to propose theories of readability.

Theoretical Development: Ideation

- First, hypothesize relationships.
- You can use any information at your disposal.

Theoretical Development: Skepticism

- Some initial theories may lack "face validity": you can rule them out by reasoning about them.

Theoretical Development: Gathering Evidence

- What experiments could you do to falsify your theory?
- Design some experiments.

- "Theory construction in the social sciences faces a series of difficulties, or different circumstances, from those faced in the physical sciences, with the result that theories in the social sciences, despite some surface similarities, have significant differences from the natural sciences. Among the differences are the massive causal complexity of the subject matter; the fact that the topics of interest to social scientists and the users of social science are generated from folk, normative, or common sense concerns, and cannot be completely separated from ordinary language; the fact that even the most successful models work only in limited contexts whose boundaries are poorly understood; and the fact that successful prediction often, if not always, results from simplifications known to be false or inadequate as explanations."

- "This means that social science theories are better understood as models which work, either to predict or explain, in limited settings, rather than laws of science which hold and apply universally."

Prediction Vs. Explanation

- Explanation requires causation; prediction does not
- Smoking is a *predictor* of risk for adolescent pregnancy
 - But smoking doesn't *cause* pregnancy, so it doesn't explain it
 - Smoking only predicts adolescent pregnancy in some contexts (e.g. not in a society in which smoking was universal)