AI Authorship Evaluator

OVERVIEW

The AI Authorship Evaluator is an interactive assessment tool designed to measure human control in AI-assisted work. Most conversations about AI focus on output quality. This project focuses on a different question: who actually directed the work?

Who defined the problem? Who performed the reasoning? Who selected the evidence? Who shaped the final structure and meaning? A polished result can appear fully authored while critical decisions were quietly delegated along the way. This tool was built to make that distinction visible.

WHAT IT MEASURES

Agency	Who directed the work?
Cognitive Engagement	Who performed the thinking?
Evidentiary Control	Who selected and verified evidence?
Authorship Integrity	Who shaped meaning and structure?

DESIGN PROBLEM

AI can assist with production. It can also absorb portions of the thinking process. The final product often conceals that distinction. A central challenge of this project was designing an assessment that reflects the conditions of the task, not just the characteristics of the output.

HOW IT WORKS

The evaluator begins by establishing the task context—specifically, the level of human control the task demands. It then measures observed authorship behavior across four domains. The two are compared, and the result is an Authorship Fit score. A low-stakes brainstorming session does not require the same level of human control as a policy recommendation or research report. The score reflects that difference.

DESIGN DECISION

The evaluator does not function as an AI detector, a grading system, or a quality assessment. It measures authorship behavior—nothing more, nothing less. This constraint was deliberate. The goal is not to maximize human involvement. The goal is to determine whether the level of human control matched the demands of the task.

PROJECT STATUS

This evaluator is a working prototype—part assessment tool, part research artifact. It is an attempt to make invisible delegation visible.