

GLOSSARY



AI Accountability Essential Training

With Barton Poulson

Use these terms and definitions below to understand concepts taught in the course.

Transcript Search: note that you can search for terms directly within the course. To search video text, switch to the *Transcripts* tab, then press Cmd/Ctrl + F on your keyboard to run a search within the active transcript.

Term	Definition
data leakage	When artifacts in the data provide unintended clues to the outcome
evasion attack	An attack involving new algorithms that fool the model
harassment by AI	A tool used to identify the extent of variability in a process, determine whether the level of variability is acceptable, and improve the process
interrater reliability bias	The case in which judgment of the data category is made by a person, and thus can be different
narrow AI	AI that is more focused on a topic or a specific problem
supervised learning	When learning puts new observations into existing criteria