

Experiments in Artificial Intelligence — Continuing, Winter 2026

Session 5 | March 11, 2026 | OLLI at Southern Oregon University

Terms & Concepts

Digital Twin

A data-driven model of a specific person (or group) built from real responses, used to predict how they'll react to something — an ad, a medicine, a policy — without testing it on them directly.

Embedding

The numerical representation a transformer model assigns to each token, positioned so that its relationship to other embeddings captures meaning.

Attention (in a transformer)

The mechanism that lets a model use surrounding words to refine the meaning of a given word or token, rather than looking it up in isolation.

Reinforcement Learning from Human Feedback (RLHF)

The process where human reviewers rate an AI's responses as appropriate or not, which is then used to adjust and constrain the model's behavior.

Confidence paradox

The pattern where the more a person trusts an AI's output, the less critically they examine it — increasing the risk of relying on an error.

Skill atrophy

The gradual weakening of a person's own thinking or professional skill from over-relying on AI to do that thinking for them.

MCP connector

A specific integration that lets an AI tool interact with an outside service — a travel site, a file drive, a payment processor — through the Model Context Protocol.

Prepared with AI assistance from a transcript of the recorded session.