

Experiments in Artificial Intelligence — Continuing, Winter 2026

Session 3 | February 25, 2026 | OLLI at Southern Oregon University

Terms & Concepts

Alignment

Whether an AI's underlying motivation actually serves the user's interest, or some other interest (like keeping you engaged, or generating revenue for the company that runs it).

Token

A unit of text (roughly a word or part of a word) that an AI processes. "Context window" size is often measured in the number of tokens an AI can hold in mind at once.

Context window

How much conversation history or document content an AI can "remember" and refer back to at one time.

Reinforcement Learning from Human Feedback (RLHF)

A training process where human reviewers rate an AI's responses for accuracy and appropriateness, and the AI is adjusted based on that feedback.

Backpropagation

The mathematical process by which a neural network adjusts its internal weights after each training example, based on how far off its answer was.

Weights & bias (in a neural network)

Internal numerical values a neural network adjusts during training; together they determine how strongly each input influences the network's output.

Cost function

A mathematical measure of how wrong a neural network's output is, compared to the correct answer, during training.

Frontier model

The most advanced, cutting-edge AI models currently available — a term several companies now use directly in their marketing.

Agent (recap)

An AI system that can take multi-step action on its own — for example, researching a topic and producing a finished, cited document with minimal further input.