Experiments in Artificial Intelligence — Continuing, Winter 2026

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

Session Outline

I. Opening: Claude's Stated Principles

- No autonomous offensive weapons
- No surveillance of Americans

II. Global AI Adoption Patterns

- Northern European countries lead in business AI adoption
- Possible explanations discussed in class and in the source article

III. Digital Twins

- Example: CVS Health and ad targeting
- Example: personalized medicine and biologic drug response
- Class discussion: other possible uses (e.g. polling and policy testing)

IV. How a Transformer Understands Language

- Tokens and embeddings
- Worked example: the many meanings of “mole”
- Attention: how surrounding words refine meaning

V. Reinforcement Learning from Human Feedback, in Practice

- The global human-review workforce behind AI guardrails
- Past incidents where review gaps let biased outputs through

VI. AI's Effect on Human Thinking

- The confidence paradox and skill atrophy
- Real example: lawyers sanctioned for AI-fabricated legal citations
- Brief note on a chatbot-safety lawsuit involving self-harm (sensitive topic, discussed only generally)

VII. Live Demo — Claude's Newest Features

- Downloadable desktop app (Windows, Mac, Linux)
- Cowork mode
- Touring the MCP connector marketplace (travel, file storage, payments, government data lookups)

VIII. Closing: Presenter's Personal Recommendations

- Continuing Claude and Google Gemini; keeping Microsoft for Office apps
- A stated personal, political view on ChatGPT and AI safety — the presenter's own opinion, not a class consensus

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