Accepted poster at ISRII 2026, October 20-22, 2026 | Newcastle, Australia

Designing Chatbot-Delivered Digital Coaching with Configurable Levels of Generative AI Autonomy

Tobias Kowatsch[1,3,4*], **Alexandre de Montleau**[1], **Marcia Nißen**[1,2,3]

[1]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[2]School of Health Sciences, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland
[3]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[4]School of Medicine, University of St. Gallen, St. Gallen, Switzerland
**Corresponding author: tobias.kowatsch@uzh.ch*

Context: Chatbots are central to scalable digital health interventions, and large language models (LLMs) promise personalized, engaging, human-like coaching. Yet generative AI remains largely unregulated across many jurisdictions, leaving researchers between strong scientific interest in its potential and the need for safety, transparency, and accountability when working with humans. Authoring tools often force a binary choice: fully scripted sessions that are safe and transparent but rigid, or fully generative sessions that are flexible but hard to control and audit.

Methods: Within the PRECIOUS project, University Medicine Zurich's technology initiative aiming to build a design and trial service for precision digital therapeutics, we built a fully functional system with an implementation partner that replaces this binary choice with configurable levels of generative AI autonomy. For each chatbot-delivered coaching session element, researchers choose how much is delegated to generative AI.

Intervention: The system offers three levels: (1) Rule-based scripted sessions are fully deterministic and transparent, with all messages and response options authored in advance and no generative content; (2) One-shot LLM sessions generate only the delivered chatbot message at runtime, while response options stay pre-defined, keeping a constrained and inspectable flow; (3) Conversational LLM sessions support prompt-driven coaching across multiple turns with free-text input; stop criteria and typed, validated variable extraction keep even open dialogue within researcher-defined bounds before control returns to the rule-based flow. AI-generated messages are marked to participants, and prompts, guardrails, and model choice are researcher-defined. Both commercial and local LLMs can be configured, and self-hosting local models on private infrastructure addresses the regulatory, safety, and data privacy constraints that limit generative AI in many settings.

Results: A single, disease-agnostic authoring environment spans the full range from transparent rule-based coaching to open generative dialogue without switching tools, so the autonomy of any AI-generated content is explicit and auditable by design.

Conclusions: Configurable autonomy levels reframe generative AI in digital coaching as a controllable design parameter rather than an all-or-nothing choice, letting researchers study what generative AI can contribute while retaining full oversight of its autonomy, including in cautious or regulated settings.

Implications: The concept enables controlled comparisons across the three autonomy levels, dose-response studies of personalization, and systematic safety evaluation of generative coaching. As a hands-on demonstrator, ISRII attendees can try out the system, switch between autonomy levels and compare scripted, one-shot, and conversational delivery in real time.

Suggested submission type: Demonstration; **Suggested topics:** User Engagement & Design, Machine Learning and Artificial Intelligence, Mobile health, Next generation interventions & Technology, Passive Sensing, Wearables, and Implementation Science.