

NeurIPS 2025 Education Track – Cover Page — Elizaveta Gavrilova

Target Audience: Non-technical professionals, students, and researchers interested in understanding how modern AI assistants actually work in real-world environments.

INSIDE AI ASSISTANT PART I: Why Your AI Assistant Isn't a Know-It-All

Format & Duration: Video lecture (~6 minutes).

Abstract / Description:

This short educational video explains how an AI assistant processes a user's request step by step. We introduce the key architectural components — interface, API gateway, orchestrator, preprocessing, retrieval with embeddings and vector databases, prompt construction, LLM inference, and postprocessing. The video uses visual metaphors and simple language to show that an assistant is not “a brain that knows everything,” but a system that searches, assembles, and generates answers. The material is designed for a broad audience and highlights both technical depth and conceptual clarity.

INSIDE AI ASSISTANT PART II: Five Ways AI Assistants Fail And What That Teaches Us

Format & Duration: Video lecture (~3 minutes).

Abstract / Description:

This video continues from the architecture overview and focuses on real-world challenges of production AI assistants. We illustrate common failure modes — latency, hallucinations, outdated knowledge, security concerns, and lack of observability. Each challenge is explained through an example highlighting why building a production-ready assistant is not just about having a strong model, but about addressing engineering trade-offs. The material is accessible for a general audience yet relevant for those designing or studying AI systems.