

# The future of archives with generative AI

An opinion article by  
Stefan Kollinger, ORF

Generative AI is quietly rewriting the role and value of multimedia archives. What once appeared to be a vault for yesterday's material has become even more valuable for tomorrow's products and services. Images, sound, video, and reliable metadata now act as the foundation for a new class of AI-enabled applications. Archives are no longer cost centers but engines for discovery, personalization, and format innovation. Yet the scale of this opportunity amplifies the responsibility to govern it. Rapid deployment of generative systems across entire collections risks degrading trust, infringing rights, and erasing the provenance that makes archives authoritative in the first place. The path forward must therefore combine ambition with restraint, pairing technical acceleration with an ethic of care. At the center of this balance stand archivists, whose expertise transforms data into knowledge and knowledge into trusted experiences.

## **The role of archivists is emerging in the age of generative AI**

The role of the archivist evolves. Archivists have always curated and contextualized, implementing standards that make collections findable and comprehensible. In the age

of generative AI, they also become architects who decide what is learnable and what should remain off limits. They design the taxonomies, ontologies, and controlled vocabularies that turn unstructured media into machine comprehensible knowledge, they set the quality thresholds that distinguish ground truth from conjecture, and they enforce provenance, ensuring that every transformation leaves an auditable trace. Just as importantly, they bridge the agendas of editorial, legal, privacy, engineering, and product teams, translating constraints into capabilities and preventing the short-termism that can arise when models are optimized without a mission.

### **Context-rich archives empower generative AI**

The reason archives are gaining value is simple. Generative AI thrives on context, variety, and data. A well described multimedia corpus offers multimodal inputs like text, audio, images, and video, augmented by metadata about people, places, events, rights, and technical characteristics. This density of context unlocks superior performance in retrieval, reasoning, and generation. Forgotten footage becomes discoverable when semantic links surface themes and entities beyond keyword matching. Editorial teams can compose narratives more efficiently when search spans dialogue, visuals, music, and ambient sound while respecting rights and sensitivities. New product categories become visible, from adaptive educational modules and interactive timelines to dynamic trailers and highlight reels generated under editorial supervision. All of this elevates the archive from repository to active knowledge infrastructure.

### **Context-rich archives empower generative AI**

Metadata sits at the heart of this transformation. Without it, archives are opaque to machines, with it, they become navigable knowledge graphs and vector spaces. Structural descriptions, shot boundaries, scene segmentation, speaker diarization, music and effects detection, timecodes, version histories, let systems understand the internal anatomy of an asset. Descriptive annotations like entities, locations, events, topics, mood, connect material to broader contexts and identifiers. Rights and usage metadata like license terms, geographic and temporal constraints, consent and sensitivity flags, constrain what can be proposed or generated. The immediate gains are visible across existing workflows.

## **Building trustworthy AI-ready archives**

A sustainable operating model for AI-ready archives also depends on architecture and governance. Institutions need explicit policies clarifying who can apply which models to which data for what purpose, along with versioning and lifecycle management for both datasets and models. Generative AI elevates the value of multimedia archives because it can unlock, recombine, and contextualize knowledge at scale. But the difference lies in the craft of archival work. Archivists provide the standards, context, rights management, and provenance that allow AI to be not just powerful but trustworthy.