



Report no. 3.4

Interim report

Version: 1.0

Language: English

Project name: **Intelligent avatars for digital heritage applications**

Project acronym: **IntAvaVR**

Project no: **09I03-03-V04-00495**

Authors:

Lead researcher: **Prof. Dr. Selma Rizvić**

Other authors: **RNDr. Ján Lacko, PhD.**

Date: **30.8.2025**



Financované
Európskou úniou
NextGenerationEU

PLÁN [OBNOVY]



ÚRAD PODPRESEDU VLÁDY
SLOVENSKEJ REPUBLIKY
PRE PLÁN OBNOVY
A ZNALOSTNÚ EKONOMIKU



VÝSKUMNÁ
AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495.”



INTERIM PROGRESS REPORT: IntAvaVR (MONTHS 1–12)

Executive Summary

The IntAvaVR research project explores the integration of artificial intelligence (AI), extended reality (XR), and interactive digital storytelling (IDS) to enrich cultural heritage education and museum engagement. Traditional virtual heritage exhibitions often rely on static visual guides or pre-recorded video narrators that lack real-time conversational adaptability. IntAvaVR addresses this limitation by developing interactive, intelligent avatars—comparing 2D deepfake-driven video agents and 3D Gaussian splatting primitives—to enable dynamic, spoken interaction within virtual spaces.

During its first 12 months, the project completed **Work Package 1 (SOTA Research)**, concluded core methodology development in **Work Package 2 (XR Methodology)**, advanced pilot application implementations in **Work Package 3 (Avatar Case Studies)**, and initiated academic dissemination under **Work Package 4 (Dissemination)**. This interim report details achievements across all four work packages, outlines technical bottlenecks and mitigation strategies, and provides a strategic roadmap for months 13 to 23.

Overall Project Progress by Work Package

Work Package 1: SOTA Research (Months 1–3) — Status: Completed

Objective: Establish the theoretical foundation, review state-of-the-art implementations, compile historical site datasets, and set up research laboratory infrastructure at PEU.

Task T1.1 (State-of-the-Art Review): The team completed a comprehensive literature analysis on AI-driven virtual humans, Large Language Models (LLMs), neural rendering, and ethical frameworks (delivering **D1.1: State-of-the-art report**). The review synthesized the Sarajevo Charter for IDS, advanced gamification paradigms (A-IDS), embodied conversational agents (ECAs), and legal/ethical governance concerns regarding deepfake technologies.

Task T1.2 (Historical Information Collection): Completed in collaboration with regional heritage stakeholders, producing **D1.2: Historical information for case studies**. The report documented historical narratives, timelines, and visual sources for three selected sites:

1. *Devín Castle (Bratislava City Museum):* Historical context spanning the 9th-century Great Moravian fortress under Prince Rastislav, late Gothic expansion by the Garay family, Renaissance adjustments by the Báthory line, and its 1809 destruction by Napoleonic troops.



2. *Stummer's Forest Railway (Topolčany region)*: Archival documentation of the 760 mm industrial narrow-gauge railway built in 1913 by Baron Leopold Haupt-Stummer, covering rolling stock (e.g., Orenstein & Koppel locomotives) and routes across Kulháň and Duchonka.
3. *Pharmacy Museum at the Red Crab (Bratislava City Museum)*: Interior layout, 16th-century Paracelsus publications, and pharmaceutical artifact inventories in Bratislava.

Task T1.3 (Building Infrastructure): Formed the cross-disciplinary research team and cataloged laboratory resources (delivering **D1.3: List of people and equipment for the new research laboratory**), utilizing PEU's Virtual and Augmented Reality Laboratory, Usability Testing Laboratory, and Media Center facilities.

Work Package 2: XR Methodology (Months 4–8) — Status: Completed

Objective: Formulate an extended A-IDS methodology incorporating AI conversational pipelines, establish guidelines for avatar creation, and benchmark early prototypes.

Task T2.1 (Methodology Development): Delivered **D2.1: New extended A-IDS methodology description**. The framework integrates gameplay elements with narrative-first structures to address the "narrative paradox" in VR, detailing a 4-stage pipeline: Multimodal Input Acquisition (ASR/gestures), Intent Understanding & Context Integration (LLM reasoning & RAG retrieval), Response Generation (TTS and neural motion synthesis), and Real-Time Engine Rendering. It also introduced audiological text perturbation models to simulate avatar hearing impairments in accessibility research.

Task T2.2 (Intelligent Avatars Creation): Created **D2.2: Guidelines for creation of intelligent avatars**. This document details standardized setups for:

1. *2D Video Avatars*: Standardized chroma-key guidelines (diffused 3-point lighting, matte non-reflective attire, 4K/30fps capture) and background extraction pipelines.
2. *3D Gaussian Splatting Avatars*: Multi-view synchronized 3-camera configurations (frontal and $\pm 45^\circ$ angles), intrinsic/extrinsic calibration, point cloud initialization, and neural texture optimization.
3. *Deepfake Video Chatbot Architecture*: Integrating Meta LLaMA 3 via Groq API for sub-second text reasoning, Deepgram TTS (aura-arcaas-en) for real-time speech generation, and Wav2Lip for audio-to-viseme lip synchronization.



Task T2.3 (Testing and Adjustments): Benchmarked prototype avatar generation routines (published in Deliverable **D2.3: Testing results**). The team established that inference latency depends primarily on synthesized audio duration rather than source video length, and confirmed that reusing pre-calculated facial tracking coordinates (`use_previous_tracking_data`) significantly accelerates generation.

Work Package 3: Avatar Case Studies (Months 3–18) — Status: On Track (M12 Interim Review)

Objective: Implement deployable XR applications, execute on-site museum installations, and evaluate user experiences.

Task T3.1 (Production of Digital Cultural Heritage Applications):

Site Photogrammetry: Conducted comprehensive multi-angle photographic surveys of Devín Castle's architectural remains, processing dense point clouds and textured meshes using RealityScan and RealityCapture.

Transparent WebM Pipeline: Resolved OpenCV's default alpha-channel stripping by engineering an HSV masking workflow that outputs BGRA image sequences, followed by an FFmpeg re-encoding pass (libvpx-vp9, yuva420p) to maintain background transparency for in-engine compositing.

Interactive Unity Prototype: Built a functional prototype inspired by interactive portrait mechanics (resembling the Map Chamber concept). A C# runtime script (LoadVideo.cs) receives user text input, submits queries via HTTP to a Flask/Ngrok backend running on a remote GPU, downloads generated WebM files, and streams transparent, lip-synced video responses to an in-engine display plane.

Tasks T3.2 & T3.3 (Installation, Testing, and UX Evaluation): Preparation underway for hardware deployment and kiosk installation across partner venues, followed by empirical UX testing. The completion of this interim report fulfills **Deliverable D3.4 / Milestone MS3**.

Work Package 4: Dissemination (Months 8–23) — Status: In Progress

Objective: Communicate research outcomes through peer-reviewed open-access journals, conference presentations, research seminars, and museum-industry workshops.



Task T4.1 (Scientific Publications): Research manuscripts detailing the extended A-IDS framework, the transparent 2D deepfake pipeline, and the comparative evaluation of Gaussian splatting versus 2D sprites are currently being drafted .

Task T4.2 (Promotion and Knowledge Sharing): Promoted preliminary findings across academic seminars, public web channels, and university research platforms. Direct engagement with local heritage groups and the Bratislava City Museum continues to align project outputs with public visitor needs.

Deliverable and Milestone Tracking

Code	Title	WP	Due Month	Target Verification	Status
D1.1	State-of-the-art report	WP1	M3	Evaluated by GLAM institution representatives	Completed
D1.2	Historical information for case studies	WP1	M3	Project Coordinator verification	Completed
D1.3	List of people & equipment for research lab	WP1	M3	Project Coordinator verification	Completed
MS1	State-of-the-art analysis completed	WP1	M3	Internal document verification	Achieved
D2.1	New extended A-IDS methodology description	WP2	M8	Project Coordinator verification	Completed
D2.2	Guidelines for creation of intelligent avatars	WP2	M8	Project Coordinator verification	Completed
D2.3	Testing results	WP2	M8	Project Coordinator verification	Completed



MS2	New method for intelligent avatars completed	WP2	M8	Project verification	Coordinator	Achieved
D3.4	Interim report (WP3 status / overall update)	WP3	M12	Project verification	Coordinator	Delivered
MS3	Interim report completed	WP3	M12	Internal document verification		Achieved
MS4	Journal publication 1	WP4	M14	Journal manuscript copy		<i>In Preparation</i>
D3.1	XR applications	WP3	M18	Museum verification	directors	<i>In Development</i>
D3.2	Testing results	WP3	M18	Project verification	Coordinator	<i>In Progress</i>
D3.3	UX evaluation study description and results	WP3	M18	Project verification	Coordinator	<i>In Preparation</i>
MS5	XR applications production completed	WP3	M18	Museum verification	directors	<i>Pending (M18)</i>

Preserving Alpha Transparency in Video Generation:

Standard video generation routines in OpenCV strip alpha transparency by converting frames into standard 3-channel BGR arrays. To allow avatar sprites to blend into virtual environments without rectangular artifacts, the team built a custom extraction script that parses frames through an inverted HSV green-screen mask, stores individual BGRA PNG frames, and uses an FFmpeg command (-c:v libvpx-vp9 -pix_fmt yuva420p) to generate transparent WebM video files.

Optimizing Synthesis Latency:

Sequential frame masking originally took 203.44 seconds for 954 frames. Applying ThreadPoolExecutor parallelization reduced processing time to **31.41 seconds**. In addition, reusing facial landmark coordinates via use_previous_tracking_data eliminated redundant face



detection passes, while downscaling intermediate generation resolution to 640×360 halved WebM compilation times.

Server-Side Runtime Stabilization:

Running the generative avatar pipeline on standalone consumer hardware proved unfeasible. The project adopted a decoupled client-server architecture, delegating speech and video synthesis to high-performance GPU instances while Unity handles scene rendering. The team is currently replacing temporary Colab/Ngrok routing with an on-premise, persistent server hosted at PEU to eliminate cold-start module loading delays.

Risk Management Matrix Update

Risk Description	Probability & Severity	WPs Impacted	Status & Implemented Mitigation
Museum Inertia / Scheduling Delays	Prob: Medium Sev: Low	WP1, WP3	Established active communication with the Bratislava City Museum. Initial testing uses modular, self-contained interactive displays that do not interfere with daily gallery operations.
Rapid Hardware / API Deprecations	Prob: High Sev: Medium	WP2, WP3	Decoupled client scripts from specific external APIs. If commercial cloud endpoints change, local open-source models (such as offline Whisper ASR and local LLaMA/TTS weights) can be hosted on PEU's GPU infrastructure.
Scope Expansion & Latency vs. Fidelity	Prob: Low Sev: Low	WP2, WP3	Retained lightweight 2D deepfake video avatars for real-time interactive kiosks to preserve system responsiveness, while utilizing Gaussian splatting for pre-rendered or static visual demonstrations.
Academic Publication Lead Times	Prob: High Sev: Low	WP4	Prepared manuscripts ahead of the M14 milestone for open-access journals with established peer-review turnaround times. Preprint repositories will be used to ensure timely dissemination.



Strategic Roadmap and Next Steps (Months 13–23)

1. **Backend Infrastructure Migration (Months 13–14):** Migrate the conversational synthesis server from Google Colab to PEU's on-site high-performance computing environment, deploying persistent memory daemons to eliminate library import latencies.
2. **First Peer-Reviewed Journal Submission (Milestone MS4, Month 14):** Submit research documenting the extended A-IDS framework and real-time deepfake avatar architectures to a journal.
3. **On-Site Museum Installation and Staff Training (Task T3.2, Months 15–16):** Install physical interactive kiosk stations at Bratislava City Museum venues (Devín Castle exhibition context), and conduct technical training workshops for museum operational staff.
4. **Comprehensive User Experience Evaluations (Task T3.3, Months 16–17):** Conduct structured empirical testing with museum visitors, measuring narrative immersion, historical knowledge retention, perceived latency, and avatar engagement.
5. **Completion of WP3 Deliverables (Milestone MS5, Month 18):** Submit **D3.1 (XR Applications)**, **D3.2 (Testing Results)**, and **D3.3 (User Experience Evaluation Study Description and Results)**.
6. **WP4 Culmination & Follow-Up Grants (Months 19–23):** Host the final academic and industry dissemination workshop, publish the second journal paper (MS6, M23), and submit follow-up European research grant applications leveraging project results (D4.4).