



Report no. 2.3

Testing results

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: **6.10.2026**



PLÁN [OBNOVY]



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



Pre-Implementation Expert Evaluation and Validation of XR Applications for Digital Cultural Heritage

1. Introduction

As part of the development of the project, two extended reality (XR) applications were proposed for the presentation and interpretation of cultural heritage. The first application focuses on **Devín Castle**, while the second presents the historical heritage associated with the **Stummer family, the forest railway near Topolčany, the sawmill, and parquet production**.

Before proceeding to the final implementation, a structured expert evaluation was conducted at the level of the proposed application concepts, scenarios, interaction mechanisms, and educational objectives. The purpose of this stage was not to evaluate an already completed software product, but to **validate the proposed design and identify potential usability, narrative, educational, and interaction risks before implementation**.

This approach makes it possible to incorporate expert recommendations into the development process at a stage when modifications to the application architecture, interaction model, narrative structure, and content remain feasible and relatively cost-effective.

The evaluation was performed by **four experts**, who assessed both applications independently using a structured evaluation framework. The evaluation combined quantitative ratings with qualitative assessment and recommendations.

2. Objectives of the Pre-Implementation Evaluation

The principal objective was to determine whether the proposed applications provide an appropriate basis for the subsequent implementation of immersive digital cultural heritage experiences.

The evaluation focused on the following objectives:

- validation of the overall application concept;
- assessment of the logical structure of the proposed scenes;
- evaluation of the historical narrative;
- assessment of historical relevance and authenticity;
- evaluation of educational potential;
- assessment of the proposed interaction mechanisms;



PLÁN [OBNOVY]



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

- evaluation of the planned 2D video avatar;
- assessment of expected immersion and engagement;
- identification of potential usability risks;
- identification of possible narrative or pedagogical weaknesses;
- formulation of recommendations for the final implementation.

The evaluation therefore represents a **design validation and requirements-verification activity** rather than a conventional software usability test.

3. Methodological Framework

3.1 Expert Evaluation

Four experts were selected to assess the proposed applications from a professional perspective. The evaluation was structured so that each expert independently assessed the defined criteria and provided qualitative observations.

Expert evaluation was selected because it is particularly suitable during the pre-implementation phase. At this stage, it is possible to evaluate storyboards, scene descriptions, interaction concepts, educational objectives, and interface principles without requiring the complete software implementation.

The assessment combined two complementary approaches:

1. **structured quantitative evaluation**, using a five-point rating scale;
2. **qualitative evaluation**, based on expert comments, observations, and recommendations.

The combination of these methods provides both an overall numerical indication of the quality of the proposed design and a more detailed explanation of the strengths and weaknesses identified by the evaluators.

3.2 Scenario-Based Walkthrough

The experts evaluated the applications through a scenario-based walkthrough. The proposed user experience was considered sequentially from the beginning of the application to the completion of the individual scenarios.

Particular attention was paid to:

- what information the user receives at each stage;

- what the user is expected to do;
- whether the purpose of an interaction is understandable;
- whether the interaction has a clear historical meaning;
- whether the transition between scenes is logical;
- whether the avatar contributes meaningfully to the experience;
- whether the proposed activities support the educational objectives.

This approach allows potential problems to be identified before implementation rather than after the complete application has been developed.

3.3 Think-Aloud Principle

The planned evaluation framework also incorporates the **Think-Aloud principle**, in which evaluators verbalize their observations while progressing through the proposed interaction scenarios.

In the pre-implementation context, this method can be applied to prototypes, storyboards, interaction mock-ups, scene descriptions, or other representations of the planned application.

The purpose is to identify situations in which an evaluator might ask:

- What am I supposed to do here?
- What is interactive?
- Why am I performing this action?
- What is the purpose of this object?
- Is the avatar currently speaking to me or providing background information?
- What should happen after I complete the task?

Such observations are particularly useful for identifying **discoverability and cognitive-load problems** before implementation.

4. Evaluation Criteria

Both applications were evaluated using the same twelve criteria to enable methodological consistency and comparison.

The criteria were:



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



1. *clarity of the application concept;*
2. *logical structure of scenes;*
3. *clarity of historical narrative;*
4. *historical relevance;*
5. *educational potential;*
6. *expected interaction intuitiveness;*
7. *quality of the avatar concept;*
8. *expected usefulness of avatar communication;*
9. *expected immersion;*
10. *expected user engagement;*
11. *appropriateness of interactive tasks;*
12. *suitability for cultural heritage education.*

A five-point ordinal scale was used:

Score	Interpretation
1	Very poor
2	Poor
3	Satisfactory
4	Good
5	Excellent

The use of the same scale and criteria for both applications provides a consistent basis for comparative evaluation.

5. Quantitative Processing of Expert Evaluation

For each criterion, the mean score was calculated using:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

where:

- $\bar{x}$ is the arithmetic mean;
- x_i is the rating assigned by expert i ;
- N is the number of experts, in this case $N = 4$.

The standard deviation can subsequently be calculated as:

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$

The standard deviation provides information about the degree of agreement among the experts. A low standard deviation indicates that the evaluators assessed the criterion similarly, whereas a higher value indicates greater variation in expert opinion.

For the overall application assessment, the arithmetic mean of the twelve criterion means was calculated:

$$\bar{x}_{overall} = \frac{\sum_{j=1}^{12} \bar{x}_j}{12}$$

The percentage representation of the maximum possible score was calculated as:

$$P = \frac{\bar{x}_{overall}}{5} \times 100$$

The resulting percentage should be interpreted as a normalized indicator of the achieved rating relative to the maximum available score, rather than as a probability of application success.

6. Expert Evaluation of the Devín Castle XR Application

6.1 Application Concept

The proposed Devín Castle application consists of three interconnected scenes:

1. an introductory historical scene;
2. a scene focused on the construction of the castle well;
3. a scene presenting the defence of the castle against the Ottomans.

A 2D video avatar accompanies the user throughout all three scenes. At the beginning of each scene, the avatar provides a monologue containing relevant historical information. Following the



introductory narration, the user can interact with the avatar in Slovak and ask questions concerning the development of the castle during different historical periods.

The concept was evaluated as a combination of immersive historical reconstruction, interactive storytelling, task-based learning, and conversational digital guidance.

6.2 Quantitative Results

The four experts provided the following ratings:

Criterion	Expert 1	Expert 2	Expert 3	Expert 4	Mean
Concept clarity	5	5	4	5	4.75
Scene structure	5	4	5	5	4.75
Narrative clarity	4	5	4	5	4.50
Historical relevance	5	5	5	5	5.00
Educational potential	5	5	4	5	4.75
Interaction intuitiveness	4	4	4	5	4.25
Avatar concept	5	5	4	5	4.75
Avatar usefulness	4	5	4	5	4.50
Expected immersion	5	5	4	5	4.75
Expected engagement	5	5	5	4	4.75
Interactive task suitability	5	4	5	5	4.75
Cultural heritage suitability	5	5	5	5	5.00

The overall mean was:

$$\bar{x}_{overall} = \frac{56.75}{12} = 4.73$$

Thus, the proposed Devín Castle application received an overall expert evaluation of:

$$\boxed{4.73/5}$$

Relative to the maximum possible score, this corresponds to:



$$\frac{4.73}{5} \times 100 = 94.6\%$$

The result indicates a highly positive expert assessment of the proposed application concept.

6.3 Interpretation

The highest scores were obtained for **historical relevance** and **suitability for cultural heritage education**, both receiving a mean score of 5.00.

This result indicates strong agreement that the application provides an appropriate framework for presenting Devín Castle through immersive and interactive means.

The high score for educational potential (4.75) indicates that the experts considered the combination of historical information and practical interaction to be appropriate for experiential learning.

The three-scene structure also received a high score (4.75). The structure establishes a meaningful progression from historical introduction through construction activity to a defensive historical scenario.

The lowest-rated criterion was **interaction intuitiveness**, with a score of 4.25. Although this remains a positive result, it represents the most important area requiring refinement before implementation.

7. Scene-Specific Recommendations for Devín Castle

7.1 Introductory Scene

The introductory scene should provide sufficient historical context to establish the user's understanding of Devín Castle and the purpose of the experience.

However, experts recommended avoiding excessive information at the beginning of the experience. The information should instead be distributed progressively across the scenes.

The introduction should clearly establish:

- the historical importance of Devín Castle;
- the chronological context;
- the purpose of the user's visit;
- the relationship between the three scenes;



- the role of the avatar.

7.2 Castle Well Construction

The construction of the castle well was evaluated as a particularly suitable interactive activity because it transforms a historical process into a direct user experience.

The final implementation should provide:

- clearly identifiable interactive objects;
- concise instructions;
- immediate feedback;
- visible task progression;
- historical explanations associated with important actions.

The interaction should have a clear educational purpose. Users should understand not only **what** they are doing, but also **why the activity was historically important**.

7.3 Defence Against the Ottomans

The defensive scene has significant potential to increase engagement and immersion.

Experts nevertheless recommended maintaining an appropriate balance between entertainment and historical interpretation. The activity should not become primarily a combat-oriented game.

The historical context should remain visible throughout the scenario, with the avatar or environmental elements explaining the relevance of the events being experienced.

8. Avatar Evaluation – Devín Castle

The avatar concept received a mean rating of 4.75, while its expected usefulness received 4.50.

The experts considered the avatar a valuable mechanism for connecting the different scenes and providing historical explanations.

The final implementation should clearly distinguish between:

- avatar monologue;
- user interaction;
- conversational mode;

- completion of the dialogue.

Visual, auditory, or environmental cues should communicate when the user is expected to listen and when interaction is possible.

The avatar should function as a **context-sensitive historical guide**, rather than merely as an animated information display.

9. Expert Evaluation of the Stummer Forest Railway XR Application

9.1 Application Concept

The second proposed application presents the historical heritage associated with the Stummer family and the forest railway near Topoľčany.

The application contains three scenes:

1. the interior of the mansion, including a historical painting incorporating the 2D video avatar and providing information about the Stummer family;
2. a forest railway scene in which the user assists with timber handling and controls the loading of wood onto the train;
3. a scene depicting the first arrival of the train at the sawmill, where information about the sawmill and parquet factory is presented.

The three scenes establish a historical and economic relationship between the Stummer family, forestry, railway transport, timber processing, and parquet production.

9.2 Quantitative Results

Criterion	Expert 1	Expert 2	Expert 3	Expert 4	Mean
Concept clarity	5	4	5	5	4.75
Scene structure	4	5	5	5	4.75
Narrative clarity	4	4	5	5	4.50
Historical relevance	5	5	5	5	5.00
Educational potential	5	4	5	5	4.75
Interaction intuitiveness	4	4	4	4	4.00



Avatar concept	5	5	5	4	4.75
Avatar usefulness	4	5	4	5	4.50
Expected immersion	5	5	5	4	4.75
Expected engagement	5	5	5	5	5.00
Interactive task suitability	5	4	5	4	4.50
Cultural heritage suitability	5	5	5	5	5.00

The overall mean was:

$$\bar{x}_{overall} = \frac{56.25}{12} = 4.69$$

Therefore, the proposed Stummer Forest Railway application received an overall expert rating of:

4.69/5

This corresponds to:

$$\frac{4.69}{5} \times 100 = 93.8\%$$

of the maximum possible score.

The result indicates a highly positive expert assessment of the proposed application.

10. Interpretation of the Stummer Forest Railway Evaluation

The highest scores were obtained for **historical relevance**, **expected engagement**, and **cultural heritage suitability**, each receiving a mean score of 5.00.

The experts considered the application particularly promising because it presents industrial and family heritage through active participation rather than relying exclusively on conventional textual or audiovisual presentation.

The forest railway scene was considered especially valuable because users can participate in timber handling and loading activities. This creates a direct relationship between the historical process and the user's actions.



The lowest score was obtained for **interaction intuitiveness**, with a mean value of 4.00. Consequently, interaction design represents the primary area requiring additional attention before implementation.

11. Scene-Specific Recommendations for the Stummer Forest Railway Application

11.1 Stummer Mansion

The use of a historical painting containing the 2D avatar was considered innovative and potentially effective for storytelling.

However, the experts identified a possible discoverability issue: users may initially perceive the painting as a purely static historical object.

The final implementation should therefore introduce subtle cues indicating that the painting contains an interactive element.

The narrative should establish the causal relationship:

Stummer family → economic activity → forest railway → timber processing

This relationship provides the conceptual foundation for the subsequent scenes.

11.2 Forest Railway and Timber Loading

The second scene provides the strongest task-based interaction within the application.

Experts recommended structuring the activity into clearly identifiable stages:

1. identify the timber;
2. manipulate or pick up the timber;
3. transport it to the loading area;
4. position it appropriately;
5. load it onto the train;
6. receive confirmation of task completion.

Each action should generate immediate feedback.



At the same time, the historical meaning of the activity should be explained so that interaction is not perceived as an isolated game mechanic.

11.3 Sawmill and Parquet Factory

The arrival of the train at the sawmill provides a logical conclusion to the preceding scenario.

The final scene should make the complete historical process explicit:

Forest → Timber → Railway → Sawmill → Parquet production

This relationship helps users understand the economic and technological significance of the railway.

12. Avatar Evaluation – Stummer Forest Railway

The avatar concept achieved a mean score of 4.75, while its expected usefulness achieved 4.50.

The experts considered the avatar particularly suitable for presenting information about the Stummer family and for providing additional historical explanations.

The avatar should, however, be integrated naturally into the historical environment. The user should understand why the avatar is present and how it relates to the painting, the mansion, and the subsequent railway scenes.

The conversational functionality should enable users to obtain additional information without disrupting the main narrative sequence.

13. Comparative Evaluation of Both Applications

The results of the expert evaluation are summarized below.

Evaluation Dimension	Devín Castle	Stummer Railway
Concept clarity	4.75	4.75
Scene structure	4.75	4.75
Narrative clarity	4.50	4.50
Historical relevance	5.00	5.00



Educational potential	4.75	4.75
Interaction intuitiveness	4.25	4.00
Avatar concept	4.75	4.75
Avatar usefulness	4.50	4.50
Expected immersion	4.75	4.75
Expected engagement	4.75	5.00
Interactive task suitability	4.75	4.50
Cultural heritage suitability	5.00	5.00
Overall mean	4.73	4.69
Percentage of maximum	94.6%	93.8%

The difference between the two overall scores is:

$$\Delta = 4.73 - 4.69 = 0.04$$

The difference of only 0.04 points indicates that the experts assessed the two proposed applications at a very similar level.

The Devín Castle application received a marginally higher overall rating, primarily because of its slightly higher assessment of interaction intuitiveness and interactive task suitability.

The Stummer Forest Railway application achieved a higher score for expected engagement, reflecting the potential of its task-oriented interaction involving timber and railway operations.

14. Risk-Based Analysis

The expert evaluation also provides a basis for prioritizing implementation risks.

Risk Area	Devín Castle	Stummer Railway	Priority
Interaction discoverability	Identified	Identified	High
Avatar state clarity	Identified	Identified	Medium
Narrative overload	Identified	Identified	Medium



Historical authenticity	Must be maintained	Must be maintained	High
Task instructions	Requires refinement	Requires refinement	High
Scene continuity	Requires attention	Requires attention	Medium
Educational integration	Positive	Positive	Low
Engagement potential	Positive	Very positive	Low

The most important common risk is interaction discoverability. Users must be able to determine which objects can be manipulated and what action is expected without requiring unnecessary trial and error.

A second important risk concerns the distinction between educational content and game mechanics. Interactive tasks should always contribute to the historical learning objective.

15. Recommendations for Final Implementation

Based on the combined expert evaluation, the following recommendations should be incorporated into the final implementation of both applications.

15.1 Interaction Design

Interactive elements should be visually and contextually identifiable. Users should receive sufficient guidance to understand what can be manipulated and what the expected action is.

15.2 Progressive Guidance

Instructions should be provided progressively. Presenting all information simultaneously should be avoided because it may increase cognitive load.

15.3 Historical Context

Each interactive activity should be explicitly connected with its historical meaning. Interaction should therefore function as part of the storytelling mechanism.

15.4 Avatar Design

The avatar should have clearly distinguishable states for:

- narration;
- listening;

- conversational interaction;
- response;
- completion.

This distinction is particularly important because the avatar is intended to support both passive information delivery and active dialogue.

15.5 Narrative Structure

The scenes should form a coherent chronological and causal sequence. Transitions between scenes should reinforce the relationship between historical events, locations, and activities.

15.6 Educational Design

The applications should prioritize historical understanding over entertainment. Gamified mechanisms should be used only where they contribute directly to engagement with the cultural heritage content.

15.7 Historical Authenticity

Historical reconstructions, dialogue, environmental elements, and interactive scenarios should be based on reliable historical sources and should avoid introducing fictional elements that could be interpreted as historical facts.

15.8 Feedback

Users should receive immediate feedback following important interactions. Feedback can be visual, auditory, textual, or provided through the avatar.

15.9 Conversational Exploration

The avatar should support exploratory questioning where technically feasible. This can allow users to obtain additional historical information according to their individual interests while maintaining the main narrative path.

16. Recommended Development Process

The results support a staged development process:

Concept → Prototype → Expert Evaluation → Analysis → Redesign → Final Implementation
→ Technical Testing → End-User Evaluation → Deployment

This sequence is important because the expert evaluation should not be considered the final validation of the applications.



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



Instead, it represents an intermediate quality-assurance stage. The recommendations generated during expert assessment should first be incorporated into the prototypes and final implementation. The completed applications should then undergo empirical testing with representative users.

This enables a distinction between:

- **expert-based design validation**, performed before implementation;
- **technical verification**, performed during software testing;
- **empirical user evaluation**, performed after implementation.

Such separation strengthens the methodological rigor of the overall evaluation process.

17. Expected Outcomes of the Evaluation

The pre-implementation evaluation is expected to produce the following outputs:

1. *a validated conceptual structure for both XR applications;*
2. *identification of potential interaction problems;*
3. *recommendations for improving the avatar interface;*
4. *recommendations for strengthening historical narratives;*
5. *identification of potential educational weaknesses;*
6. *prioritization of implementation risks;*
7. *recommendations for scene-specific interaction design;*
8. *requirements for historical authenticity;*
9. *recommendations for improving user guidance;*
10. *a structured basis for subsequent prototype development and user testing.*

The expert assessment therefore functions as a decision-support mechanism for the implementation team.

18. Conclusion

The pre-implementation expert evaluation demonstrated a high level of conceptual suitability for both proposed XR cultural heritage applications.



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



The Devín Castle application achieved an overall expert score of **4.73/5**, corresponding to **94.6% of the maximum possible rating**. The Stummer Forest Railway application achieved **4.69/5**, corresponding to **93.8%**.

Both applications received particularly strong evaluations in terms of historical relevance, cultural heritage suitability, educational potential, and expected immersion. The results indicate that the proposed combination of immersive environments, interactive historical activities, and 2D conversational video avatars provides a suitable foundation for digital cultural heritage interpretation.

At the same time, the evaluation identified interaction intuitiveness as the most important area requiring refinement. The final implementation should therefore place particular emphasis on the discoverability of interactive objects, clarity of instructions, feedback mechanisms, and clear differentiation between avatar narration and conversational interaction.

The expert evaluation should consequently be understood as a **formative validation stage** rather than as evidence of final application usability. Its primary value lies in identifying and resolving potential design weaknesses before implementation. Following the incorporation of the recommendations, the completed applications should undergo technical verification and subsequent empirical evaluation with representative users.

The proposed methodological sequence—**expert validation followed by implementation and end-user evaluation**—provides a rigorous framework for developing immersive digital cultural heritage applications and increases the likelihood that the final systems will satisfy their intended usability, educational, narrative, and cultural heritage objectives.