



D8.5 PROTOTYPE EVALUATION RESULTS



Grant Agreement nr	856879
Project acronym	PRESENT
Project start date (duration)	September 1st 2019 (36 months)
Document due:	August 31 st 2022
Actual delivery date	August 31 st 2022
Leader	Brainstorm Multimedia
Reply to	jmontesa@brainstorm3d.com
Document status	Submission Version

Project ref. no.	856879
Project acronym	PRESENT
Project full title	Photoreal REaltime Sentient ENTity
Document name	D8.5 - Prototype Evaluation Results
Security (distribution level)	Public
Contractual date of delivery	31/08/2022
Actual date of delivery	31/08/2022
Deliverable name	Prototype Evaluation Results
Type	Report
Status & version	Submission Version
Number of pages	46
WP / Task responsible	CREW/BRA
Other contributors	ICERT, FS, CM, UPF, UAu
Author(s)	Javier Montesa (BRA), Elena Llorca (BRA), Theodore Jones (FS), Eva Valls (UPF), Romualdo Carbone (iCERT)
EC Project Officer	Ms. Diana MJASCHKOVA-PASCUAL Diana.MJASCHKOVA-PASCUAL@ec.europa.eu
Abstract	This document reports the obtained results during the final prototype evaluations, which are based both on the evaluation plans proposed in D8.1 and on the interim evaluation results reported in D8.3. Depending on each use case, the document reflects the plan followed on the final evaluation, the action points identified in D8.3 and how they have been worked out and validated, and finally the extracted conclusions during interviews with professionals and stakeholders after a workshop on the developed tools.
Keywords	Final evaluation, use cases, stakeholders, workshop
Sent to peer reviewer	Yes
Peer review completed	Yes
Circulated to partners	No
Read by partners	No
Mgt. Board approval	No

Document History

Version and date	Reason for Change
0.1 29-06-2022	Document created by Javier Montesa (BRA)
0.2 08-07-2022	Evaluation plan contributed Javier Montesa (BRA)
0.3 25-07-2022	Virtual Clerk UC contributed by Eva Valls (UPF)
0.4 27-07-2022	Action points technical validation by Javier Montesa (BRA)
0.5 28-07-2022	Stakeholders workshop report by Elena Llorca (BRA)
0.6 28-07-2022	Optimized animation and character evaluation contributions by Theodore Jones (FS)
1.0 11-08-2022	Version for internal review (14 days before submission date)
1.1. 23-08-2022	Cubic Motion peer review
1.2 29-08-2022	Revisions in response to review: final version submitted to Commission

Table of Contents

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	5
2	BACKGROUND	6
3	INTRODUCTION	6
4	OPTIMISED ANIMATION AND CHARACTER GENERATION EVALUATION	7
4.1.	Production Feedback	7
4.2.	Action Plan	9
5	BROADCAST USE CASE	10
5.1	Final validation contents	10
5.2	Final software and hardware configuration and setup plan	10
5.3	Action points assessment plan	11
5.4	Technical Validation Report	14
5.5	Stakeholders' feedback and workshop day	16
5.6.	Stakeholder's feedback on the viability of Present	18
6	REGISTRATION AUTHORITY VIRTUAL OFFICER USE CASE	23
6.1	Design review	23
6.2	Analysis plan – areas of intervention	27
6.3	Conclusion	33
7	VIRTUAL CLERK USE CASE	34
7.1.	Assessment contents and plan	34
7.2.	Hardware configuration and interaction	34
7.3.	Results and analysis	36
7.4.	Action points improvement plan	39
8	CONCLUSION	40
	ANNEX I. VIRTUAL CLERK EVALUATION QUESTIONNAIRE	42
	ANNEX II. FOCUS GROUP SCRIPT	44

1 EXECUTIVE SUMMARY

The WP8 objective has been the evaluation of the experimental productions and the prototypes developed during the project. While D8.1 focused on the evaluation plan both for experimental and production use cases, D8.3 reported the interim results on the prototype evaluation and now D8.5 reports on the obtained results in the final prototype evaluations.

Given the fact that these three deliverables are connected as an initial plan, the interim results, and then the final results, this document will reference the plans designed in D8.1 and will continue the explanations and results already reported in D8.3, along with the final results obtained in the evaluation of the final systems.

Given that the experimental production evaluation is reported in D8.4, this document focuses just on the developed prototypes, the optimised animation systems, the character generation, the broadcast use cases, the registration authority virtual officer and finally the virtual clerk.

Thanks to the success of the Optimized Animation and Character Generation development, and its huge potential for utilisation across both offline and realtime rendered pipelines, the technology has already been deployed onto multiple active projects, including one major Disney movie production, to be released late 2022, and another Unreal Engine based short film project. This has led to an ability to evaluate the technology in a live production setting with artists, supervisors and clients interviewed as active daily users. The results of this evaluation are presented here along with identification of current workflow bottle necks and an action plan for addressing these issues and further improving the toolset in preparation for wider adoption.

The Broadcast Use Case evaluation continues from the results and action points identified in the interim evaluation phase. A brief final evaluation plan has been designed and reported in the document where, the evaluation itself, the required equipment, the required setup, the contents, the workshop with stakeholders, and the final interviews are proposed. Then a second section lists the action points and identifies the improvements and new functionalities included in the final integration of InfinitySet and the Reference Implementation, and then provides the obtained results on their technical validation. Finally, and following the validation plan, a report is provided on the workshop findings, and on the stakeholder's interviews conclusions.

The Registration Authority Virtual Officer also continues from the results of the previous validation phase (**Deliverable D8.3 - Interim Prototype Evaluation Report**) that evaluated the prototype, provided by UPF using WebGL Technology, against usability heuristics and design guidelines (**Heuristic evaluation**). In the present phase the evaluation has been carried out using the prototype that rely on high-res technology.

Finally, we present the evaluation of the beta version of the Virtual Clerk. It was installed in the University premises in June 2022, with one year delay with respect to the initial plans, due to the pandemics situation and networking security issues. The assessment content and plan are recalled, as the evaluation follows them: several aspects of the appearance of the character and the interaction have been evaluated via subjective rating. The details of this actual beta installation in terms of hardware configuration are then presented. Next, the results of the subjective evaluation are presented and analysed. While the overall results offer a positive picture, a number of shortcomings with respect to the interaction, animations and rendering are identified. Finally, the last subsection discusses the improvements to the Virtual Clerk following this evaluation, towards a new alpha version of the Virtual Clerk. Some of them are already underway.

2 BACKGROUND

This document is the last one in a series:

- D8.1 Initial Experimental Production and Evaluation Plan (M12) reported on how the partners planned to evaluate the agent under the different scenarios for experimental production and prototype evaluation.
- D8.3 Interim Prototype Evaluation Report (M24) reported on the interim results on the production prototype evaluation activities proposed in D8.1
- D8.5 Prototype Evaluation Results (M36), this document, reports on all the final prototype evaluation tasks for PRESENT.

The scope of this deliverable is centred on the final prototypes developed in Present, and focuses on their technical results, and on the possible interests of future clients and stakeholders on its use and functionality.

3 INTRODUCTION

By following the evaluation plans described in D8.1, the Present prototypes were evaluated first in M24 and then these interim validation results were reported in D8.3. At this stage of the project many modules were still missing or not integrated, and also some of the already present modules were still not optimal so a list of action points and possible improvements were identified and reported in the corresponding deliverable.

At this stage of the project, all modules have been finished and integrated, and all the action points identified during the interim evaluation have been analysed, improved, and finally evaluated again.

As the project contains several prototypes and use cases (integrating with the Standard Implementation), their evaluation and reporting has been split differently in several activities that are reported in this document.

The Optimized Animation and the Character Generation have been organized in the following sections:

- Production Feedback. This first section details the feedback gathered from interviews with animators, riggers, supervisors and clients on the projects on which this technology was deployed.
- Action Plan. Details the workflow bottlenecks and technology improvements identified from extensive user interviews along with an action plan to address each.

The Broadcast use case has been organised following these sections:

- Final Validation Contents. These first sections outline the final validation plans for the broadcast use case in detail. Specifically this one details the contents for the final Proof of Concept used in the validation
- Final software and hardware configuration setup plan. Details the plans for the final software and hardware setup in Brainstorm premises for the technical validation and for the workshop with stakeholders.
- Action Points Assessment Plan. This section lists the identified action points in D8.3 and how their assessment is planned.
- Stakeholder's Workshop Plan. Describes the plan and the agenda for a workshop with stakeholders.

- Stakeholders Interview. Details the structure for the interview with experts and how it will be the basis for the final analysis on how the use of Present outcomes in broadcast environments is expected.
- Technical Validation Report. This section details the obtained results on the technical evaluation, the specific action points that remained to be covered, how they were solved, and the obtained result on their technical validation and assessment.
- Workshop Report. This section covers the workshop day outcome, the results of each activity with stakeholders, and the conclusions based on the interviews conducted with them.

The Registration Authority Virtual Officer contains the following sections:

- Design review. These first sections outline the methodology followed
- Analysis. Describes how the validation of the prototype has been carried out and how the methodology has been applied
- Conclusion, describes the findings of the evaluation activities

And the Virtual Clerk is organized as follows:

- Assessment content and plan. This section presents the properties of the use case to be assessed and how it is planned to be carried out.
- Hardware configuration and interaction. Describes the current configuration of the installation and explains how the interaction should take place to be evaluated.
- Results and analysis. Details the obtained results from the questionnaires, making an analysis of the answers and proposes solutions to the identified points for improvement.
- Action points improvement plan. This last section plans the actions to be taken in order to improve the use case, taking into account the results and analysis of the previous section.

4 OPTIMISED ANIMATION AND CHARACTER GENERATION EVALUATION

As part of the Present project Framestore has developed two key pieces of technology that have the potential to offer enormous advantages across multiple sectors of the business:

1. A machine learning based approach to massively accelerate the evaluation of complex rig deformations and increase their portability across multiple pre-rendered and realtime platforms.
2. A translation pipeline for the conversion of complex offline rendered shading graphs to simplified realtime compatible shading models.

Due to the enormous advantages offered by these technologies and the success of the Present development effort the path into active production has been accelerated and production roll-out commenced in late 2021.

This timeline has enabled the evaluation of these technologies to be carried out on active client facing productions with artists using them on a daily basis.

4.1. Production Feedback

Both technologies are in active production on a major film project for Disney Studios that will be delivered and released in late 2022.

The main use of these technologies on the production has been to give animators vastly improved real time representations of both the look and the rig deformations of the final characters and allow for the animation rigs themselves to operate at higher framerates. Traditionally animators only have access to a very rough approximation of the final character look and deformation, as this has been the only option to achieve animation rigs that evaluate at the high frame rates required for efficient animation production.

Interviews were carried out with key artists and supervisors on this production to assess the efficacy of these new technologies in providing animators, supervisors and clients with more accurate character representations and more performant rigs much earlier in the production process.

Key Quotes (advantages):

Animation Supervisor

"It gives us much better deformation results in our Animation Presentation renders and animation playblasts. It allows us to show the client work that is closer to the final result."

Lead Animator

"What you see is what you get!

Better interactive feedback for (hero characters) body rig deformation (silhouette/definition) for animators in Maya's viewport, but also in animation playblasts and presentation QCs out of the box. Better deformation outlook meant less back and forth between supervision/anim/rigging/creature-fx regarding exaggerated posture and less misguided judgements. "

Lead Rigger

"Riggers saved the majority of work days normally slated against anim rigs deformation, since we were able to primarily target render rig improvements, and bake it down using the ML pipeline before including the result into our anim rigs.

Bigger picture - everyone benefited, clients included, as we were all able to really understand strictly out of animation whether or not (the lead character's) posture/silhouette was making sense."

Key Quotes (disadvantages):

Lead Rigger

"We had to front load the schedule of (hero character's) body anatomy rigging tasks in order to have it built well enough to train the ML rig against. This included creating a new body rig component dedicated to the training process where skin geometry is being wrapped to its underlying anatomy without requiring the needs of a simulation, as well as a system of corrective shapes in the neck, hands and feet area, which were all based off our render body rig component. This took a reasonable amount of time which was hard to quantify beforehand due to everyone involved doing it for the first time."

4.2. Action Plan

After gathering detailed feedback from artists using the technology in active production across a variety of disciplines specific pain points and workflow bottlenecks have been identified. The below table details the main issues and action points to address them.

ID	Aspect	Action points
A1	Freezing experienced when using Maya plugin	Create reproducible example scene and proceed with bug fix
A2	Rework ML training graph to give more visibility of intermediate steps	Currently the training graph structure means that the process is quite opaque to the user. Restructuring this to save out more intermediate steps and make it easier to trace through the process would aid in debugging.
A3	Add visual QC submission to the end of ML training process	This QC would involve validating the final ML rig, as well as each pre-blended patch, against the original rig using a template animation and generating a heat map showing the error between the two, thus giving a clear visual QC of the quality of the model.
A4	Improve pipeline for adding additional animation/poses to the training or QC data	Currently editing the training animation/poses is very manual. Ideally if a shot was found that had a pose or poses that solved badly there would be an easy way to add this animation/pose to the training and the QC process to retrain against and then test against to check the retraining had solved the issue.
A5	Improve speed of training by allowing for retraining from Alembic bake	Currently any training process involves regenerating the poses every time, which is 95% of the compute time. Allowing for retraining with different training parameters but based off of the same Alembic baked poses would allow for much faster iteration.
A6	Make model weight generation deterministic	Weight generation is currently not seeded meaning that results are not fully deterministic. If the ML rig is to be used in client presentations this should be fixed to ensure fully repeatable results.
A7	Automate blend weight generation	This step is currently manual but there are multiple ways to automate this step to simplify the process.

The evaluation of these action points with the development team is underway with work expected to address these items commencing 3rd quarter 2022 and continuing for the remainder of the year.

Multiple other productions are adopting these tools currently with one production in particular utilising an entirely Unreal Engine based pipeline for the characters. Something that would not have been possible without this technology.

5 BROADCAST USE CASE

Based on the Interim Prototype Evaluation Report results, reported in D8.3, several possible improvements and action points were identified that have been covered on the last stages of the project. The validation plan proposed hereby focuses first on the assessment of results related to these action points, some of them were covered, some of them were found to be avoidable based on specific workarounds, and some of them were found to be inherent to the design of the systems. Secondly, and also integrated as a final activity on the validation plan, a demo day was planned with important stakeholders that travelled to Brainstorm premises to see the final result, this time physically. Stakeholders participated in a workshop in which their feedback about Present was registered.

This document details the overall final validation plan covering different stages:

- The final configuration contents validated.
- The final software and hardware configuration, and its setup plan.
- The list of action points covered during the validation and how they were assessed.
- The list of stakeholders, and the plan for the workshop with them to demonstrate the final system.
- The planned stakeholders' interviews objectives, procedure, and results.

5.1 Final validation contents

The final validation scenario configuration is very similar to the already presented in the interim validations but extended with more contents and functionality to be tested as the objective of the new scenario version is to test the new included functionalities - apart from verifying the new system behaviour related to the identified action points in D8.3.

One of the most relevant additions is the inclusion of a complete script both for the presenter and the agent, which will require a complete set of graphics to be used by the agent in its explanations.

Regarding the agent, a new version of Gareth is to be tested, more optimised graphically, with more features and capabilities regarding emotions, behaviours, and motion animations that will be triggered from InfinitySet.

The first release for these contents was ready on June, 2022, although some updates were introduced during software and hardware configurations.

5.2 Final software and hardware configuration and setup plan

In order to provide the new functionalities listed above, also new software capabilities were required, and a new integration with the last reference implementation was needed. The final software configuration considers these facts and also implements the action points identified in the midterm evaluation.

The new agent capacities and behaviours have also been included in the system and will be available. And, specifically, the new option to make variations on the script for the agent was ready for validation and so the InfinitySet and the Authoring Tool were prepared to be compatible with it.

More specifically, these are the activities and tasks that needed to be undertaken in order to prepare the system for the final validation, for the workshop with stakeholders, and for future use:

Authoring tool finalization

The authoring tool is finished and ready to use, but a new version of the Reference Implementation of PRESENT is expected, with more functionalities, that will require a final version for the Authoring Tool.

InfinitySet final optimizations

A deep work has been carried out in order to optimise InfinitySet to manage the scene and the agent fluently, but again, new behaviours for the agent and new functionalities are expected in the final version of the Reference Implementation, that will require updates on InfinitySet and will probably allow for further optimization in the integration on both engines InfinitySet and Unreal Engine.

Workflow integration with CM iPhone tool.

One of the more interesting features in PRESENT will be the option to use a specific iPhone tool to record new sentences and behaviours for the agent to adopt them. Once this tool is ready, it may be necessary to integrate it in the Broadcast Use Case workflow pipeline.

The final software and hardware configuration setup was ready on 15-July-2022.

5.3 Action points assessment plan

Once the system was ready to assess the obtained results on each of the identified action points, an evaluation process to treat each and all of them in an organised manner started.

These were the action points identified from the audience's point of view:

ID	Aspect	Viewers suggestion	Action points
A1	Shadow behaviour	The virtual agent shadow was not always represented properly during the video POC.	Review the virtual agent shadow generation pipeline and ensure that it has the same characteristics of those of the real presenter.
A2	Unexpected glitches	Detected graphical imperfections related to the virtual agent during the video POC.	Review the virtual agent contents and ensure that there are no graphical imperfections during upper extremities movements and during static poses.
A3	Agent response to changes in scenario configuration.	It seems that the virtual agent does not always react properly to a change of the camera position.	Take into consideration the online camera position information and how this could be implemented in the virtual agent behaviour.

A4	Agent interaction naturalness.	The viewers saw the interactions between the real presenter and the virtual agent with a moderate level of naturalness.	Study how to improve the naturalness of the interactions between the virtual agent and the real presenter.
A5	Overall evaluation of agent presence in the program.	The viewers suggest that they are satisfied with virtual agent presence, but this is still not comparable with the real presenter.	Study how the virtual agent action in the show pipeline could be improved to get the same impact of the real presenter.

And these were the proposed action points from the system operators' point of view:

ID	Aspect	Operators' suggestion	Action points
A6	Transmission of orders to agent	Detected some delay between the command and the action starts for the virtual agent actions.	Analyse the command shipment pipeline to understand if there is some systematic delay.
A7	Set of agent animations, behaviours and graphics trigger management.	In case of need (show issue, etc) it could be difficult to control manually the virtual agent with the actual interface.	Improve the manual control on all the virtual agent features. Add filters and utilities to navigate through all the available actions to select them easier.
A8	Scenario and graphics preparation workflow.	Preparing the scenario and graphics for a broadcast session could be improved.	Analyse how the related features could be modified to make this step faster.

The evaluation of these action points started even before the system was completely ready, in fact some of them were already fully covered, but the organised analysis started properly once the system was ready for evaluation and ended on 22-July-2022.

Following, the assessment plan of each action point is provided:

ID	Aspect	Assessment
A1	Shadow behaviour	Cause: the agent was not casting shadows due to InfinitySet not being able to bring it from Unreal Engine. Solution: a new integration allows InfinitySet to bring shadows from Unreal Engine. Validation: a visual inspection of the final result will be carried out by technicians and stakeholders.

A2	Unexpected glitches	<p>Cause: some of the agent's animations made the hands detach from the sleeves, resulting in visible glitches.</p> <p>Solution: the reference implementation has a new animation system with glitches fixed.</p> <p>Validation: a visual inspection of the final result will be carried out by technicians and stakeholders.</p>
A3	Agent response to changes in scenario configuration	<p>Cause: the agent was not able to look at places other than those recorded on the animations.</p> <p>Solution: the reference implementation has a new gaze system that allows the agent to look to different places. The agent will look to the camera when he speaks to the audience.</p> <p>Validation: a visual inspection of the final result will be carried out by technicians and stakeholders.</p>
A4	Agent interaction naturalness	<p>Cause: the agent was not able to look at places other than those recorded on the animations.</p> <p>Solution: the reference implementation has a new gaze system that allows the agent to look to different places. The agent will look to the host when he speaks to him.</p> <p>Validation: a visual inspection will be performed by technicians and stakeholders to the final result.</p>
A5	Overall evaluation of agent presence in the program	<p>Cause: the agent had rough animations.</p> <p>Solution: the new reference implementation has a new gaze system, blending between animations and a general better look.</p> <p>Validation: a visual inspection of the final result will be carried out by technicians and stakeholders.</p>

ID	Aspect	Assessment
A6	Transmission of orders to agent	<p>Cause: although the execution of commands is not delayed, some animations were taking a few milliseconds to start.</p> <p>Solution: a review of the animations has been carried out and the affected ones have been corrected.</p> <p>Validation: a measurement will be made of the time between the button being executed and the start of the animation. Also a visual inspection of the final result will be carried out by technicians and stakeholders.</p>
A7	Set of agent animations, behaviours, and graphics trigger management	<p>Cause: some aspects of the agent were difficult to control.</p> <p>Solution: a new user interface with more and better structured controls will be implemented.</p> <p>Validation: a usability check of the interface will be carried out by technicians.</p>
A8	Scenario and graphics preparation workflow	<p>Cause: the data for the graphics had to be entered by hand, which was too much of an effort.</p> <p>Solution: the introduction of the authoring tool makes it possible to export the current knowledge base and fill in the data of the graphics automatically.</p>


		Validation: a usability check of the workflow will be carried out by technicians.
--	--	--

5.4 Technical Validation Report


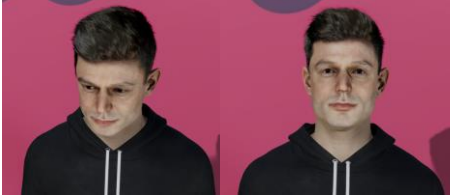

From July 21th to 27th, a technical validation activity was carried out in the Brainstorm 's premises located in Paterna (Valencia, Spain). This activity mainly consisted in a detailed review of all the action points identified in the previous section of this document. A group of internal and external experts reviewed and evaluated the results of the actions undertaken to solve or improve the aspects identified by the audience.

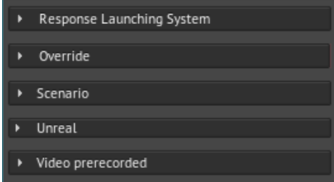
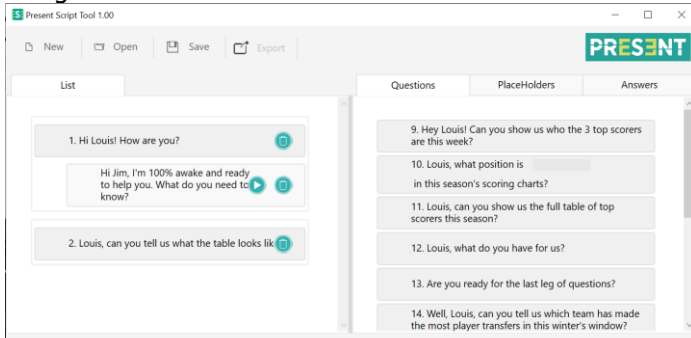
All their answers were based on subjective opinions and perspectives as potential audiences, except for the delay, which was measured.

The tables below explain each of these action points, showing a picture of the result where applicable and the overall feedback received.

ID	Aspect	Technical validation
A1	Shadow behaviour	<p>The addition of both, host and virtual agent shadows, has brought a new level of integration with the scenario. The feedback received is very good. On the negative side, minor imperfections have been detected in close camera shots.</p> 

A2	Unexpected glitches	<p>The glitches detected have been fixed and now the sleeves follow the hands properly. As a small improvement, the position of the fingers should be checked to prevent them from getting inside the clothing.</p>
----	---------------------	---

		
A3	Agent response to changes in scenario configuration	<p>The new gaze system allows the agent to look at places other than those that were recorded. This new system makes the agent feel more natural and is generally very welcome. However, it has been noted that certain animations look better with the original movement, as the gaze system blocks camera movements. Therefore, depending on the situation, the gaze system will be activated or not.</p> 
A4	Agent interaction naturalness	<p>As in the previous point, the agent's ability to look at the host has helped to improve naturalness.</p> 
A5	Overall evaluation of agent presence in the program	All the new improvements made to the agent have made the overall feeling more satisfactory.
A6	Transmission of orders to agent	Although those animations that were late to start have been corrected, the new blending system between animations introduces a small delay. This delay is necessary to be able to switch correctly and without jumps between 2 animations. However, the delay is now constant and can be taken into account by the operator. The measured value for this delay is 300 milliseconds.
A7	Set of agent animations, behaviours, and graphics trigger management	The user interface has been improved by structuring the controls into 5 groups. This, together with the execution of commands by the virtual agent, has greatly simplified the operator's work. The general opinion is that it is a good change.

		
A8	Scenario and graphics preparation workflow	<p>The Authoring Tool has improved the creation of new scripts as well as the modification of data related to the Aston graphics. This new workflow saves the operator from having to worry about updating data manually, and is therefore less prone to errors. The general opinion is that it is a good addition.</p> 

5.5 Stakeholders' feedback and workshop day

Based on the contents and setup prepared for the final version of the technical validation and action points' assessment, a workshop with stakeholders was organised with the goal of gathering their feedback, an activity in which a final demonstration of the product was provided and afterwards their impressions and comments were elicited.

On one hand, the validation pursued the attendants' approval to the system developed, in which they learnt from the final configuration contents, the final software and hardware configuration, and the list of action points undertaken as a result of the midterm previous validation.

On the other hand, a focus group discussion was organised and led to valuable output, allowing the consortium to draw upon stakeholder's perceptions, beliefs, and comments, as well as the key points that should be addressed for a future Business plan for Present.

The workshop was held at Brainstorm Multimedia facilities in Paterna (Valencia) on July, 27th, 2022. The stakeholders were welcomed at 9.45h in the main hall of the building and shortly after the CEO Ricardo Montesa, walked the attendants through the company's facilities. The workshop agenda was as follows:

PRESENT	FINAL SISTEM DEMONSTRATION DAY
	<p>Date: July, 27th 2022</p> <p>Location: Brainstorm Multimedia.</p> <p>Calle Coeters, 5. Paterna - 46988 - Valencia, Spain</p>
9.45 - 10.00	Welcome to Brainstorm3d
10.00 - 10.30	Visit to Brainstorm facilities in Paterna
10.30 - 11.00	Focus group: introduction and first part
11.00 - 11.30	Coffee break
11.30 - 14.00	Demonstration of Present tool
14.00 - 15.30	Lunch
15.30 - 16.30	Focus group: second part
16.30 - 17.00	Wrap - up and closing remarks

In this workshop, stakeholders were able to contribute to the Present tool's validation through their feedback on several related topics. Remarkable considerations were taken into account that helped draw the future of Present in the broadcasting sector.

As said, the main goal of the activity was to collect the opinion of those attending the demonstration of Present, that is, the representatives of the Commercial Advisory Group belonging to the following broadcast groups: TVE, Apunt TV, TVG and TVCM, all TV channels in Spain. The format was a focus group, in which the conductor from BRA followed a script with 10 key questions to be discussed by the group (see Annex II). The activity was fully recorded and was used to elaborate conclusions and kept for the project's dissemination activities.

The activity was structured in three parts:

1. A first block of questions was asked before the Present demo, so that the participants had little or no information about the project.
2. A detailed demo of the technology and functioning of Present
3. A second block of questions, after the demo, to gather their feedback on the technology and the envisaged pros and cons of its implementation and commercial exploitation

Once the technical requirements to set up the workshop were established (video and audio recording systems), the whole experience was recorded to document the project activities and to further analyse the group's feedback about the tool.

In the following paragraphs, the results of the activity are exposed.

5.6. Stakeholder's feedback on the viability of Present

The topics proposed for discussion in the focus group were related to potential applications of Present in broadcast environments, as well as the envisaged acceptance of both media professionals and audience to TV programs based on the interaction between a physical person and an avatar. Also, the perception about the interest that TV channels could pose on this option, including the level of investment that a TV channel would accept to get the tool.

In order to assure a well-structured way to elicit stakeholders' thoughts and opinions on every important aspect of the system, a structured script and procedure were designed, which is included in Annex II at the end of this document. Four hours were recorded and analysed that contained the whole interaction with stakeholders, both at the focus group and at the demonstration. The results of the analysis are exposed in the next paragraphs, following the order of the questions put on the table for discussion that day.

I. BEFORE THE DEMO

1. Have you heard about or seen TV programs that incorporate realistic avatars?

If so, what is your view on them?

- **If not, what about virtual reality games that incorporate realistic or hyper-realistic avatars? How do you like the idea of using them on TV programs?**

Note 1: A definition of avatar was provided: Avatars = synthetic characters, hyper-realistic graphic elements that identify a user or represent a character. The questions do not refer to flat and static avatars, such as those commonly used as user identification or as chatbots, but to realistic or hyper-realistic designs, in 3D and with movement. If necessary, show the images of the last page

Note 2: Images of realistic avatars were shown at that point to make sure they were all thinking of the same.

Image 1: Participants in the focus group



Results:

Half the attendants had seen said programs, in different formats, but usually as a punctual appearance in a program, not a program based on avatar - human interaction. The rest have never seen them, even though they knew of realistic avatars in contexts different than TV, such as video games. In any case, the fluidness and naturalness of the interaction seemed to be an important aspect for accepting the format. Also, the avatar having a reduced range of interaction resources (vocabulary, gestures...) may better suit a short or limited appearance on TV show.

There was general acceptance of avatars in TV formats, as long as there was a “reason” or “relevant context” to use them. Participants underlined that the experience would very much depend on the context, format of the TV program and quality of the avatar. For some of them, it seemed no easy way to find a context in which an avatar would be better or even equal to a human presence but were interested in having an insight to that possibility.

2. Can you envisage any clear application to the use of avatars on television? In what kind of program?

Many ideas were put on the table:

- Realistic avatars whose conversation is backed on a corpus of data, meaning that the avatar will represent a more innovative or friendly way of presenting said data. E.g. In a sports program the presenter could ask the avatar which soccer team won the Champions League in 2010 or who was the top scorer last year.
- Also, the avatar could be a good ally as a sign language translator, especially in the case of public televisions, which are committed to providing a universal service.
- Deep-fake avatars achieve a level of realness that is impressive. These specific avatars could be interesting for interviews. E.g. Hotel du temps - France TV or for story telling.
- An opponent in a TV conquest can also be an interesting use (as Deep blue vs Kasparov chess game), even though, in the current state of IA technology, the audience might perceive winning the avatar as an impossible endeavour or, even worse, might find it deceptive.
- Other uses: as monologists; in History documentaries as historical characters; for promotional and marketing activities

The idea of realistic avatars where conversation is backed on IA, in the sense that the avatar can provide answers previously unknown to the presenter was not seen as a possibility in the current state of technology. It is a risk similar to having somebody on stage that you have no clue about what he/she is going to say. There should be a previous script.

3. Do you think that programs that incorporate avatars could be an interesting option to be explored by TV channels?

The participants agreed that it is definitively interesting to explore the possibility. Avatars are not seen as “invading” every program in the future but as a complement to some specific programs / shows.

As long as the avatars are capable of creating or improving the spectacle, and / or raising emotions in the audience, they have a possibility of being accepted.

One of the participants introduces the idea of holograms as a way of taking TV characters to the streets (“a space still to be conquered”). Being it a real representation of a real character (e.g. conductor) or being it an avatar, a holistic representation of said character could help to get TV closer to the audiences. The rest of participants agree that it is an interesting possibility to explore. Participants underline the strong legal and ethical issues to be faced when working with realistic avatars that in any way imitate or even impersonate real characters (image, voice, deep-fake avatars...), which can hinder any initiative in this regard. The problem is not so important when it comes to very antique or historical characters, but how to deal with not-so-antique characters is not solved (rights, hiers, ...).

After the first block of questions, participants were introduced to the Present technology in the same room, and then taken to a studio set in which a demo of the technology took place.

Image 2. Participants in the demonstration phase: structure of Present



Image 3. Participants in the demonstration phase: functioning of Present



Image 4. Participants in the demo phase: operating Present



II. AFTER THE DEMO

4. Now that you've seen how it works. What is your perception of the acceptance that this tool may have among professionals in the field?

Participants pointed out that, if the system can integrate the sequences of movements into the general play-list, it would be better for the operators. BRA team agreed that it is possible and convenient to do it.

Specifically, about professional's acceptance, the participants agreed to distinguish between the public sector and the private sector. The general perception is that, even though the tool seems easy to use, for a TV program to really achieve a high quality in every sense, you need a multidisciplinary and innovative team covering all the aspects of production.

As for private broadcasters, most participants feel there could be a very good opportunity because talent (innovative and well-formed professionals) could be attracted to these companies and also the flexibility to choose and train the personnel to be involved in each project is higher.

But they envisaged that some factors could act as deterrents to Present in the public sector:

- The lack of flexibility in the public sector to create the specific team needed for Present may ballast the initiative or even dismiss it.
- Related to this, to produce good content with Present (or similar tools) a specific professional background is required that is not easy to find in public TVs, especially small or local ones.
- Regarding motivation, team leaders may feel that the human resource's structure is rigid at the public sector and they usually can't choose the members of the team. That makes it sometimes hard to build that multidisciplinary and innovative team needed, discouraging team leaders.
- As for the rest of professionals, due to budget cuts, receiving proper training could be hard or even result in work overload in some cases.
- Finally, trade unions in the TV sector may consider avatars as a potential threat to employment.

But all in all, these factors can be reduced if an adequate context, format and quality is achieved with Present.

5. And the audience? How do you think they'll react to this show format?

Stakeholders agree that it mostly depend on the situation and on some key aspects:

- That the quality of the visualisation is high
- That audience is briefly but clearly explained how the interaction occurs, so they don't feel misled at any point
- That there is a wide range of possibilities of interaction, both regarding the avatar (movements, expressions, even clothes) and the interaction (extended corpus so the interaction has a wide range of possibilities)

Asked about possible targets, it was not clear for the participants up to what point, for instance, young people will accept it better than elderly people. The reason could be that the first group could be more used to it - and more exigent - while the second could be happily surprised by the avatars.

If there is a possibility to "customise" the TV content by creating an avatar for different targets, that will probably enhance the audience's acceptance.

6. Do you think that TV channels will find this tool and this program format (with human and avatar interaction) as interesting? (same question as before the demo)

Apart from distinguishing between public and private TV mentioned in Q4, there is a consensus around the idea of using avatars in TV punctually and not to develop a whole program based on human - avatar interaction. Eg. A program for the electoral night in which avatars could connect from specific locations (as small towns) to "read" the results.

So the most supported idea is related to a punctual intervention of an avatar in a program, to play any role, or maybe a periodic intervention - but not a whole program based on it.

7. Do you envisage any clear application to the "person + avatar" format on television? In what kind of program? (same question as before the demo)

Apart from the outlined in the first block of questions (before the demo), TV programs in which the role of the avatar is to provide a more scientific, objective, or technical information, would more easily success if they integrate said avatar, while TV programs in which the avatar is expected to show, manage or interact in a more human / emotional / subjective context, will have it harder to succeed.

In the first case, it refers to avatars whose responses come from a pre-established database. Depending on the situation, the avatar provides one answer or another. Therefore, possible answers are previously known and thus controlled. In this sense, the avatar can be understood as *a friendly interface of a database*.

In the second case, it would be necessary to deepen into many aspects to decide whether there is a possibility of success or not: program, character, ethics, etc.

As a general summary, if the avatars are to do something better, quicker or cheaper than humans, then there could be a sense for their integration in TV programs.

Finally, a second issue arises related to the previously mentioned "ethic issue": the bias of the avatar's presence or profile. Could it be interpreted as biased to show a 25-year-old male avatar in a Sport program instead of a female avatar? A young female avatar in a cooking program instead of a 60-year-old male avatar? Is the audience going to perceive it as biased in the same way as with humans, or is it going to be different? Further research was recommended on this topic.

8. What is your general impression of the tool?

It is a very interesting novelty for the participants in the focus group. If human resources and costs issues are approached in a good way, as described in the former questions, it opens many possibilities for content production at local TVs.

9. Do you think it can be implemented as it is? Or is it lacking maturity in any aspect: handling, visualisation?

Minor technical improvement could help, but only from an expert eye's perspective. It seems enough for the general audience.

10. What level of investment do you think a television could assume to implement a tool like this?

The participants considered that televisions could easily assume between 15.000€ and 30.000€ to implement Present, provided that the human resources needed, the avatar (design, voice,

actor recording, etc) as well as the Infinity Set tool, are counted apart. That is: just the software which is a module of Infinity.

They would recommend a SaaS commercialization of Present, and also there was some discussion about cloud services for the whole tool (Infinity + avatar module), which Brainstorm's CEO said was an option they were currently investigating, as many clients asked for such a possibility.

6 REGISTRATION AUTHORITY VIRTUAL OFFICER USE CASE

6.1 Design review

METHODS - Research goals

The reviews include several methods of analysis, in each, the level of inspection varies depending on the review's goals of identifying usability problems and strengths.

The types of design reviews performed were aligned with the phases and the prototype fidelity released:

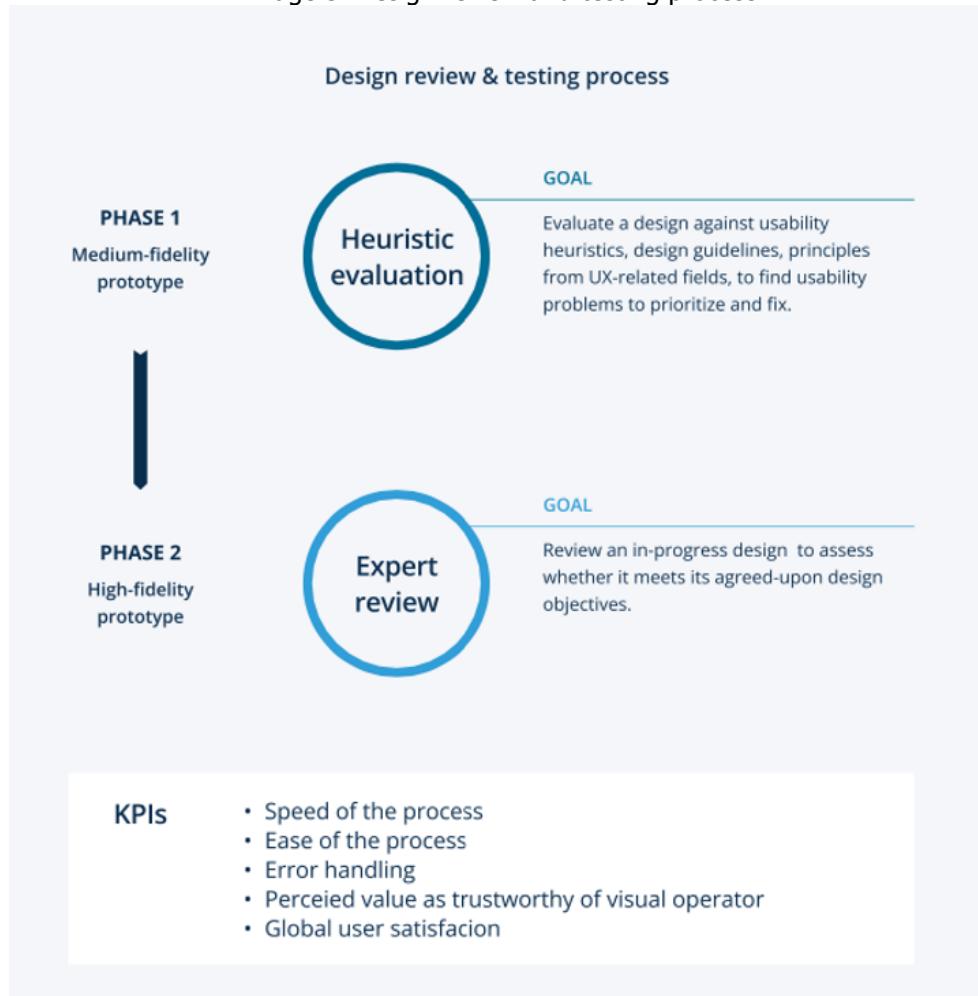
PHASE 1 (Deliverable D8.3 – Interim Prototype Evaluation Report) -

Evaluate the prototype against usability heuristics (such as Jakob Nielsen's 10 usability heuristics), and design guidelines, to find usability problems to prioritize and fix as part of an iterative design process. (**Heuristic evaluation**)

PHASE 2 (Deliverable D8.5 – Prototype Evaluation Result) – Review an in-progress design to assess whether it meets its objectives, (and the KPIs given) and provides a good experience. Also called **UX expert review** is used to check for possible usability issues. The distinction between heuristic evaluations and expert reviews is blurry in many organizations: an expert review has a more general version of a heuristic evaluation.

These two types of methods, combined, provide a very useful type of analysis to understand in a short time and without the possibility of user involvement what's not working, allowing an in-progress review of the digital RAO.

Image 5. Design review and testing process



What is a UX design review?

The UX expert review is performed by assessing the design not only with prior findings from heuristic evaluation of phase 1, but also against other principles of usability-related fields such as cognitive psychology and human-computer interaction, with a particular attention on [InfoCert Design guidelines](#).

At this stage we focus on the desirability and brand experience, by analysing the quality of the interaction with the digital human, how user engage and how consistent is with the InfoCert brand guidelines.

We focus on the two KPIs identified for the previews review about:

- Perceived value of visual operator
- Global user satisfaction

The core components of a design review are:

- List of usability strengths and usability problems. The review should include a list of strengths and usability problem with a short explanation for each (the heuristic or

- principle violated should be clearly cited and related to the design, so that any fix will address the underlying issue).
- If a problem does not necessarily violate a classic guideline or principle, but instead stems from other usability research, the issue should be explained clearly why the design represents a problem. If a problem does not necessarily violate a classic guideline or principle, but instead **stems from other usability research**, the issue should be explained clearly why the design represents a problem.
- **Severity ratings** for each usability problem is key to making the findings actionable and helping designers prioritize the redesign work. At Nielsen Norman Group, we often use a simple 3-point severity scale for each problem: *High, Medium, or Low*.
- **Recommendations and example of best practices** for fixing each usability problem. Another key element of an actionable usability finding is a clear recommendation for how to address the issue. Often, once the issue is noticed and the underlying reason of the issue is understood, the fix will be obvious. Whenever possible, recommendations must be supported with examples of other sites addressing the same issue. Providing multiple examples of sites solving the same issue prevents the conclusion that there is any single best way to design the solution.

USER PERCEPTION – InfoCert face

We are going from text (bot) to more of a humanlike engagement where we are able to create emotional connections into the customer experience.

The key ingredient to creating customer loyalty is providing an excellent customer experience at every point in the relationship, from in-store experience to post-sales customer service. As visual interactions become increasingly prominent in our digital world, using virtual agents with a face and hyper-realistic human behaviour can help a brand connect with its customers.

Digital humans embody the personality, voice, and nature of the brands they work for. All of this is done for a better, more engaging, and personalized experience.

This is why it's important to adopt InfoCert guidelines and Indaco Design System (<https://zeroheight.com/1ff022471/p/66ef81-indaco-design-system/b/19d0ee>) a set of rules and best practices that help maintain consistency and continuity within a digital ecosystem, providing the user with a familiar and adaptive brand experience across all platforms). This is determining the personality, role, appearance, expressiveness, voice and other traits of your new digital employee to make sure that InfoCert brand values and identity are expressed and embodied by the digital human.

InfoCert values and guiding principles

The values that the brand have associated with the experience are the core set of guiding principles that shapes every aspect of the business. These brand principles guide story, actions, behaviours that help to make sure the applications and in general all touch-points remain consistent within the company's guidelines.

Tone of voice. InfoCert use a simple language in the way they address major topics:

- Consistency: All texts of the interfaces should keep an active form, speaking to user informally but without being too confidential or informal.
- Enthusiastic: Choose active forms doesn't mean being intrusive but just use plain language.

- **Credibility:** Using professional language and terms, avoiding writing obvious or redundant things that undermine the brand credibility

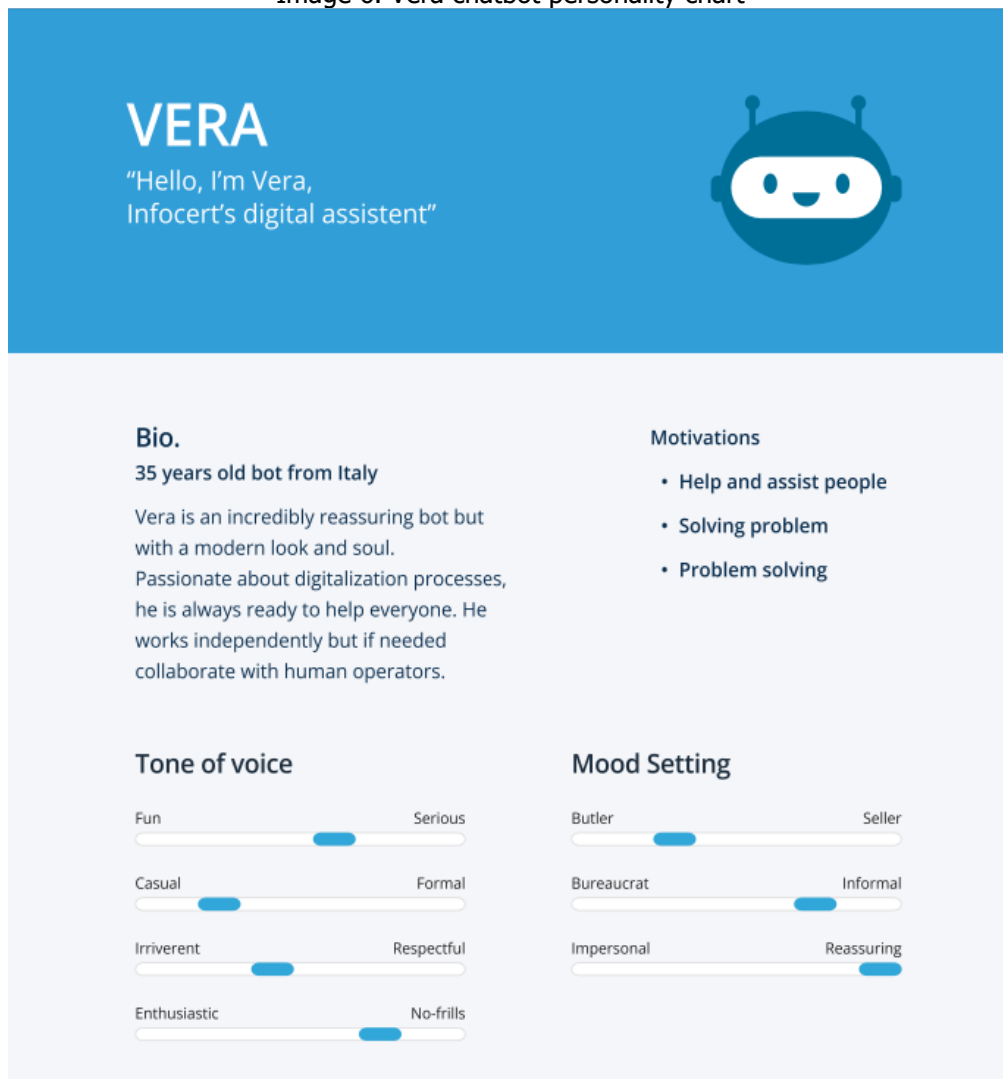
Look and feel. The demographic elements that InfoCert choose to represent the brand are:

- **Under 40:** Between a young 20-year-olds who need tools for their small business, to 30-year-old startups, up to middle-aged accountants.
- **Smart casual look:** The strategy is to maintain a certain rigor, even when you choose a young people, make sure that they are neither tied up nor too casual.
- **Natural context:** Look for natural daylight ambient light, without using cold or warm filters, or that saturates the color or excessively detail images with noise.
- **Character traits:** Authentic character that you can meet in your everyday life are preferred. We are trying to avoid the unrealistic perfection of celebrities.

Customer experience – InfoCert characteristics and personality

This is an important next step into the digital word of empathetic customer experience. InfoCert already worked on a chatbot personality named Vera, in order to maintain consistency in the InfoCert communication, we should take personality trait from that research showed below:

Image 6. Vera chatbot personality chart



6.2 Analysis plan – areas of intervention

In order to deliver a customer service beyond the users' expectation, the analysis has been focused in two areas:

- **Avatar presence:** Analysis of the personality and physical appearance of the digital human delivered.
- **Flow:** Analysis of the process shown, highlighting areas of improvement per each step
- **Severity scale for UX findings:**
- **Strengths:** Positive consideration and improvement
- **High severity:** The findings contributed to users stopping their task or resulted in user frustration or self-blame for struggling with the experience.
- **Medium severity:** Users could complete the task, but with observable frustration, notable slowdown, or after several attempts.
- **Opportunity:** Users could complete the task with just brief slowdown or confusion, something that a small improvement can fix easily.

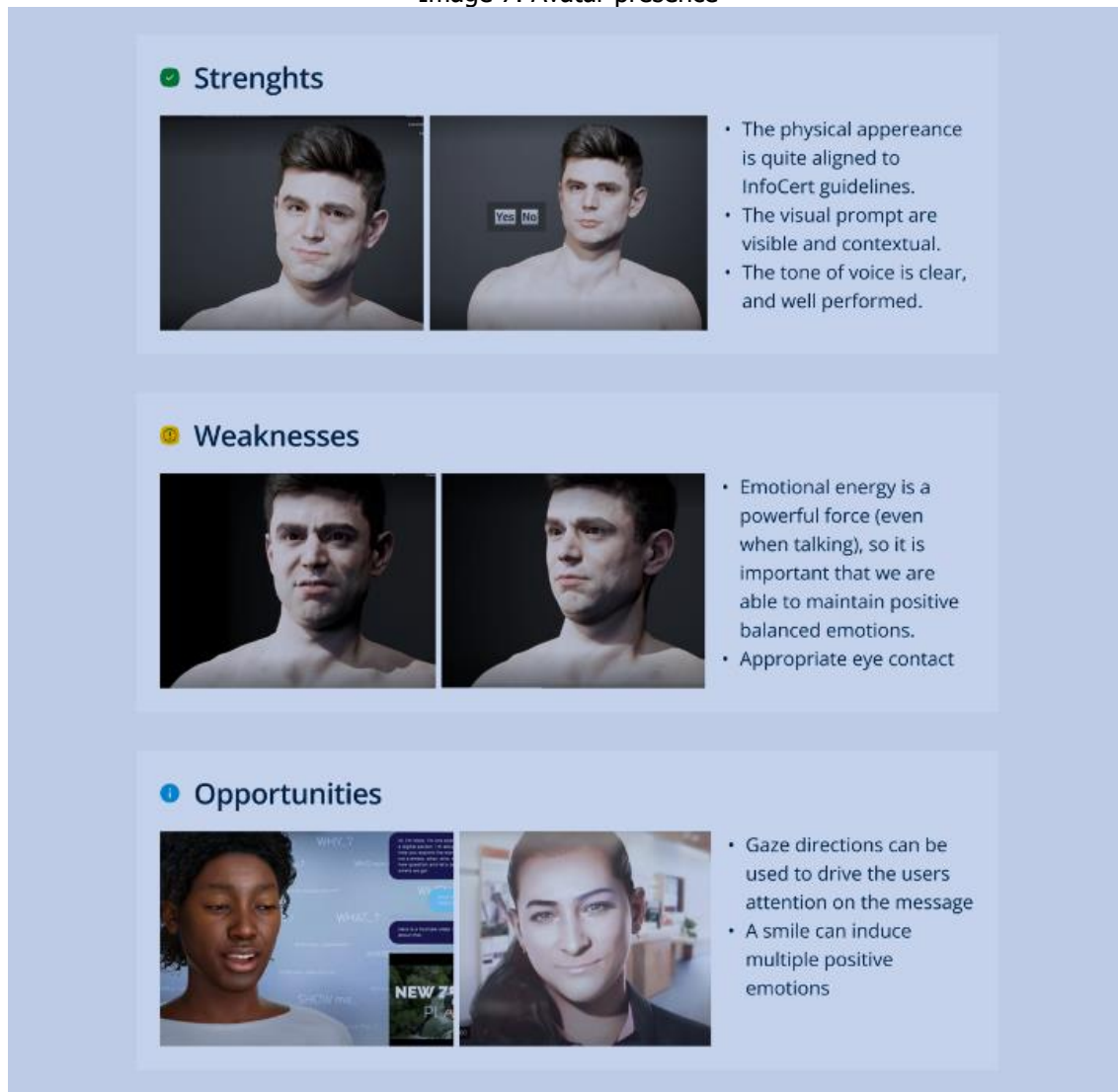
Avatar presence

In the previous paragraph it was underlined why the digital humans have to be crafted according to the InfoCert look and feel. Although the realistic rendering is impressive, some features can be customised to embody the soul of the brand.

Here are some area where the avatar appearance was merged with InfoCert visual identity:

- **Demographic aspects:** Male under 40s
- **Outfit:** smart casual
- **Background:** natural daylight ambient light
- **Appearance:** good-looking with some flaws that add a natural and authentic look and feel to the character
- **Tone of voice:** should follow the personality created for the chatbot
- **Motion:** giving an appropriate eye contact and comfortable facial expression that puts users at ease.
- **Helpful interaction:** add visual prompt and support text to better guide the user through the process.

Image 7. Avatar presence



FLOW 1- Welcome and onboarding

The presentation is where the avatar introduces himself as an InfoCert ambassador. This can be used to introduce the purpose of the assistant and to link any terms of use or privacy information.

- **Good eye contact**
The avatar looks professional with appropriate eye contact. A hint of a smile would be appreciated.
- **Background information**
Support text can be useful in terms of accessibility or sound noise. Allowing users to approach tasks in multiple ways providing different Methods to accomplish the same task.
- **Consistency and standard. Visual prompt**
When asking users to make a choice, use the industry standards and even better the Indaco component UI kit.

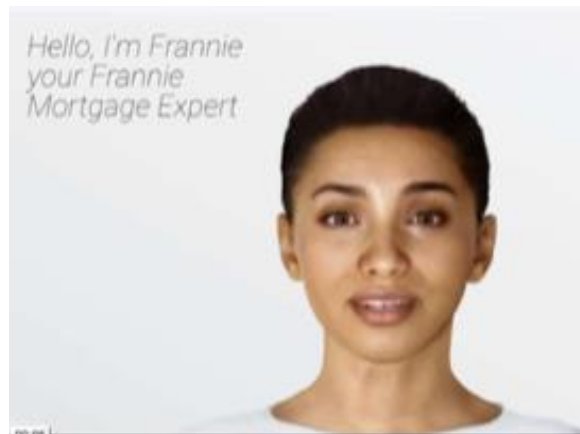
Image 8. Avatar presence



Opportunity:

Create a presentation with call-to-actions to help user get started with the process using a tone of voice like vera, the InfoCert chatbot to create consistency on the channels.

Image 9. Avatar presence



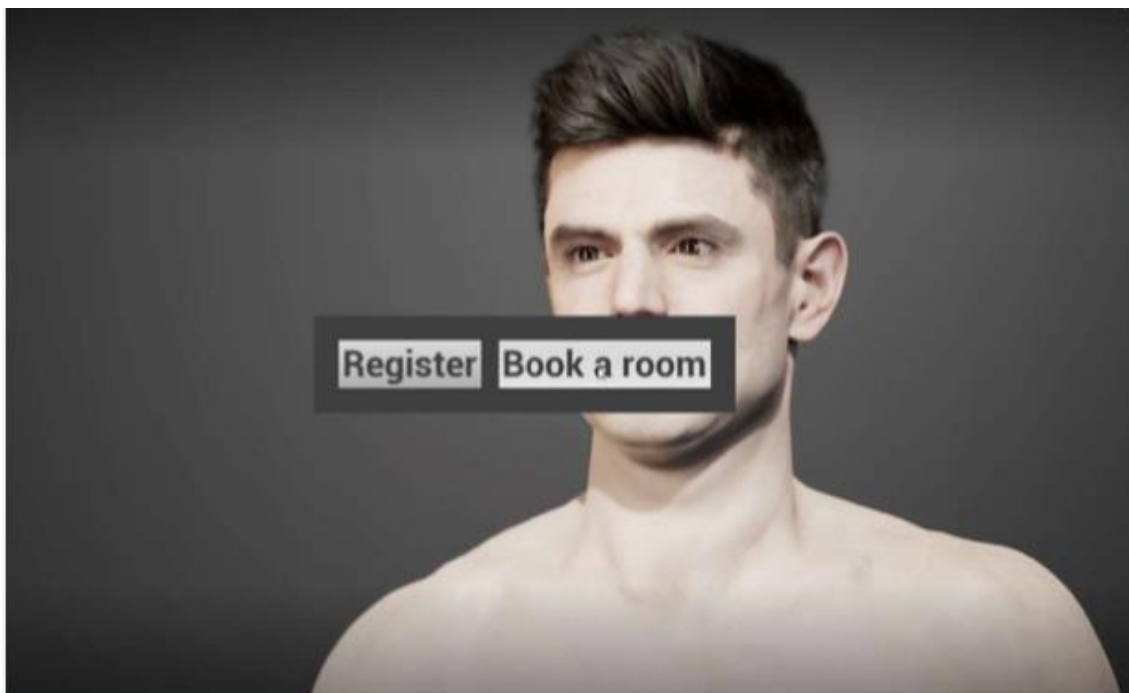
FLOW 2- First question

The process starts with the choice of service. Users are getting familiar with the interface and the possible ways of interaction

- **Visual aids placement and UNDO.**
It's not nice to cover the face of the avatar with the prompt. It's also important to give users the chance to undo an action (can be contextual or appear on a snackbar/menu).

- **Recognition rather than recall. Offer help in context**
Contextual help can reduce confusion, but we should use the same set of words on the visual prompt that has been used in the speech.
- **Facial expression. Motion and gaze directions**
The motion doesn't look true-to-life probably because the facial expression doesn't show a clear emotion and looks artificial. Giving an appropriate and more stable eye contact or using gaze direction only when you want to drive the attention of the user on something can improve the naturalness of the expression. People also tend to respond more positively to people who smile, this would also be comforting to users.

Image 10. Avatar presence



Opportunity:

Show visual prompt always on the side of the screen, avoiding covering the avatar's face and if you want to play with the eye contact, it's recommended to use gaze direction on the message or the user.

Adding an undo option is always a good idea.

FLOW 3- Authentication flow

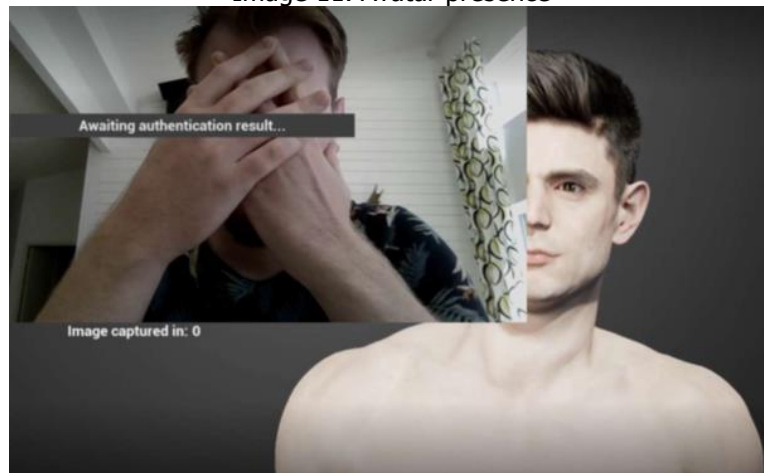
In this step, users follow a 3 steps process to get authenticate (from typing them name and take a pic to scan the QR code to validate them identity with the system)

- **Making information appear in a natural and logical order.**
Present information in order to prepare the user for what will happen. Introduce the topic and after ask the question. In this way it is easier if the user does not get distracted.
- **Error prevention. Include helpful constraints**
Offers training or visual aids where users could familiarise with the task to prevent error.

Build trust through open and continuous communication.

Communicate clearly to users what the system's state is (tell them you'll use the camera to take a picture) no action with consequences to users should be taken without informing them.

Image 11. Avatar presence



Opportunity:

Introduce the task and give some context to help the user on the process.

The image shouldn't cover the avatar's face but placed on the same side of the previews prompt:

Image 12. Avatar presence



FLOW 4- Validation and response

This is the step where the user terminates the process successfully. If they have any doubt, they will be able to solve it at this moment.

- **Good feedback.**
The script about error handling and end of the process are responding to users clearly and reassuringly.
- **Building on existing mental models helps users predict interactions.**
Users transition from the physical world to the digital world. Including familiar elements and activities in the interface will help users move through the experience with ease. After greeting you, the user offer some extra clue that highlight the end of the process.

Image 13. Avatar presence



Opportunity:

Show visual aid that highlights the end of the process.
Some FAQ or just a final text that wraps up the process just concluded.

6.3 Conclusion

Findings summary

High severity	Medium severity	Opportunity	Strengths
Facial expression and motion <ul style="list-style-type: none"> It is important to maintain positive balanced emotions with a smiling face. Gaze should keep an appropriate eye contact Control facial expression while speaking 	Visual aids <ul style="list-style-type: none"> The avatars' face shouldn't be covered with the interaction (put them on the side) It's missing the undo action Be consistent using components from Indaco UI kit 	Background information <ul style="list-style-type: none"> Support text can be useful in terms of accessibility or sound noise. Allowing users to approach tasks in multiple ways. Add a human touch into the conversation 	Voice <ul style="list-style-type: none"> Include helpful constraints The voice used seems natural and is clear and easy to understand
Communication <ul style="list-style-type: none"> No action with consequences to users should be taken without informing them (open camera) Present information in order and introduce them 	Recognition rather than recall <ul style="list-style-type: none"> Offer contextual text as a support of what the avatar is saying Suggest possible actions 	Build on existing mental model <ul style="list-style-type: none"> After greeting you user offer some extra clue that highlight the end of the process. Introduce the task and give some context to help the user on the process. 	Scene settings <ul style="list-style-type: none"> The flow has a good shot of his face

What makes a digital human more realistic?

The digital human delivered is incredibly realistic, details such as the skin imperfection, the muscles definition, the hair nullify the effect of 2d or flat that is typical of virtual characters. Progress has been made also from expressing emotions to the tone of voice used, which improve the way they interact with users.

The main challenge is enhancing the ability of interacting with other human beings, creating an immersive user experience that elevates content and allows people to focus on what they should do, facial gestures and gaze directions can be used to drive the user's attention to the message.

Opportunity:

- Motion.** To create an immersive user experience that elevates content and allows people to focus on what they should do, facial gestures, appropriate eye contact and gaze directions can be used to drive the user's attention to the message and build a natural and frictionless conversation.

- **Voice.** It sounds very natural, clear and understandable
- **Replicating true emotion.** The motion doesn't look true-to-life probably because the facial expression doesn't show a clear emotion and looks artificial. Giving an appropriate and more stable eye contact or using gaze direction only when you want to drive the attention of the user on something can improve the naturalness of the expression. People also tend to respond more positively to people who smile, this would also be comforting to users.
- **Flaws.** Imperfections make digital humans more real but they have to be pleasant. Some anatomically peculiar glitches can be very distracting.

7 VIRTUAL CLERK USE CASE

Due to the pandemic and security issues, the installation and commissioning of this use case has been delayed by one year. The evaluation stage should have been ready by the end of June 2021 but was not ready until June 2022. Therefore, the provisional evaluation of the prototype could not be done at that time, and so, the current use case is still a beta version. The evaluation was conducted to validate this version and the results will show how to improve the final version.

This section details the overall evaluation covering different stages:

- Assessment contents and plan
- Hardware configuration
- Results and analysis
- Action points improvement plan

7.1. Assessment contents and plan

The main objective of this phase is to evaluate the realism of the user experience with the Virtual Clerk. In order to do that, as described in D8.1, the quality of the rendering, the user-agent interaction and the fluency of the dialogues will be measured. For doing so, a subjective rating was carried out through a questionnaire (see Annex I) which the users filled after the interaction. The sections of the form were the following ones:

- **Visual quality of the rendering:** lipsync, eyes, hair, skin, facial and body expressivity and outfit will be evaluated.
- **Quality of the user-agent interaction:** the content and how it is given, the verbal and no-verbal communication and the friendliness of the interaction will be rated.
- **Fluency of the dialogues:** the timing between questions and answers will be measured.
- **Wish list for virtual characters:** users' opinion of virtual characters and the usability of the system will be determined.

7.2. Hardware configuration and interaction

The physical installation is greatly constrained by the illumination and acoustic properties of the space in which the agent is located (near the entrance of the Tànger Building where our Department of Information and Communication Technologies is located). The former determines which visualisation system to install (colours and contrasts of the render, types of screen, etc.). The latter greatly constraints not only how the audio is provided to the user, but also influences on how the user inputs audio (their voice commands) to the application. For this case, the environment is well illuminated by natural light and is prone to echoing and reverberation.

The installation consists of a large (human-size) vertical flat screen in a frame with a computer and speakers hidden in it, which is in charge of rendering and executing the Virtual Clerk and

outputting any kind of audio. A tablet attached to the frame (with some limited movement for position adjustment purposes) and connected to the computer lets the users start the interaction, introduce (if necessary) any textual input such as a particular researcher's name and let them know when to speak. It is also where the questionnaire appears. A conference microphone (in principle good for noise cancellation) is also in the same structure as the tablet.

The interaction starts when the user clicks the start button on the tablet. He/she will then be presented with the terms and conditions which he/she must accept before continuing (agreed after passing an ethical review by the university commission). Once confirmed, the assistant will ask the user what kind of information he/she wants to know. Each time the character finishes speaking, a microphone icon will appear on the tablet to indicate that the user can speak. However, in case the requested information is about a person, the tablet will display a text box instead of the microphone. When he/she starts speaking, the icon will flash indicating that the system is listening and once a valid response has been detected, the icon will disappear. At the same time, the agent will return a response and ask if the user needs anything else. If so, the process will be the same. Otherwise, the agent will say goodbye and the questionnaire will appear on the tablet.

Image 14. Virtual clerk Eva



7.3. Results and analysis

Once the users interacted with the Virtual Clerk, they were shown a questionnaire which they had to answer. A total of 16 people took part in the survey, fewer than expected due to the delay in installation and the vacation period as the installation took place, as indicated, in June 2022. The scoring of each section and the most relevant comments are shown in the following tables.

Visual quality of the rendering	
Aspect	Assessment
Rendering: Lipsync	Results: neutral evaluation Cause: primitive lip sync based on basic vowel detection. Solution: <ul style="list-style-type: none"> - Preprocessed mouthing based on text: less scalable but easier - Building a more robust real-time lip sync: ideal but hard.
Rendering: Eyes	Results: positive evaluation Cause: - Solution: -
Rendering: Hair	Results: neutral-negative evaluation Cause: simple hair model and colouring Solution: improving the visualisation of hair
Rendering: Skin	Results: positive evaluation Cause: although positive, subsurface scattering has been proposed as an improvement Solution: <ul style="list-style-type: none"> - Implement subsurface scattering and evaluate cost-benefit
Rendering: Facial Expressions	Results: positive evaluation Cause: - Solution: -
Rendering: Body and Outfit	Results: neutral evaluation Cause: body may lack some expressiveness. The outfit has been criticised. Solution: <ul style="list-style-type: none"> - Check and change the outfit (particularly the boots) - Add non-verbal body gestures related to what is being spoken
Comments	The body stiffness is commented on several occasions in the different comment sections available in the survey

Quality of the user-agent interaction	
Aspect	Assessment
User Comfortability	Results: positive evaluation Cause: - Solution: -
Content and Delivery	Results: positive evaluation Cause: - Solution: -
Verbal Communication	Results: positive evaluation Cause: - Solution: -
Non-Verbal Communication	Results: negative evaluation Cause: body stiffness Solution: <ul style="list-style-type: none"> - More and natural gesticulation when talking (hands, upper torso and head)
User Friendliness	Results: positive evaluation Cause: however, the installation could be improved Solution: <ul style="list-style-type: none"> - Tablet should be better connected and is slightly far from where the user would naturally interact. - Tablet's microphone quality and text writing capabilities are underwhelming. Some people suggested embedding the microphone into the actual screen of the avatar.
Comments	<p>Some confusion on how to interact with the avatar is mentioned in several comment sections. The commands may not be as obvious to the user as it seems for developers.</p> <p>Some bug on the name searching may have happened as one comment points out. However, no more information is provided about which name produced the error.</p>

Fluency of the dialogues	
Aspect	Assessment
Question Repetition	Results: negative evaluation Cause: the acoustic characteristics of the room of the installation, the undesirable microphone performance and the actual limitations of the speech recognition may have caused such an elevated amount of command

	<p>repetition. The speech recognition model might be looking for native-english pronunciations.</p> <p>Solution:</p> <ul style="list-style-type: none"> - Improving the acoustics of the environment or moving the installation. - Change the microphone to a directional one - Speech recognition models are difficult to improve. A better option is to allow for different wordings of the same command so non-native English speakers have higher chances of successfully completing their requests.
User-Avatar Interaction Time	<p>Results: positive evaluation</p> <p>Cause: -</p> <p>Solution: -</p>
Answer Meaningfulness	<p>Results: positive evaluation</p> <p>Cause: -</p> <p>Solution: -</p>

Wish list for virtual characters	
Aspect	Assessment
3 Properties of ECAs	Being interactive agents, interactivity is the most important feature. Except for naturalness, surveyed people seem to have different preferences over which two other features are relevant.
3 Highlights	<p>Interactivity, verbal communication and usefulness are the main selected options.</p> <p>Other options like appearance, realistic rendering, naturalness and non-verbal communication might have been dismissed because of the render and body stiffness problems mentioned previously.</p>
Real Environment Usage	<p>Results: neutral-positive evaluation</p> <p>Cause: people being surveyed might be familiar with the installations and their judgement about its utility might be biased. Nonetheless it was positively scored</p> <p>Solution: -</p>
Context-Utility	<p>Results: positive evaluation</p> <p>Cause: -</p> <p>Solution: -</p>

Analysing the results, the reading is overall positive, taking into account that it was a beta version. Most of the points have a medium-high score. It can be highlighted that the aspects on which the evaluation was not good enough are the following ones: the rendering of hair, body language and the difficulty of understanding the user's response in some cases. Other characteristics that

can also be improved are the lipsync and the outfit of the character, the interactivity with the tablet and the feedback from the system, as the most important property for users is the interaction.

7.4. Action points improvement plan

Since this is a test phase, not all aspects were covered. So some of these responses were already expected and work is already underway to improve them, for example the render and body animations. The following paragraphs explain which actions will be taken to improve the experience and which are already under development.

Render

First of all, the rendering is the first thing that the user perceives and therefore creates the first impression. Therefore, it is important to improve this aspect as seen in the evaluation results. Hair is being worked on, implementing the Marschner algorithm based on hair cards. As for skin, subsurface scattering has already been integrated. A specific rendering for clothing would also have to be implemented.



Image 15. Subsurface Scattering applied to the skin. Hair render improved using hair cards.

Animation

Another important aspect is animation. The character has real time facial animation but reproduces the same breathing body animation. To improve non-verbal communication, conversational gestures such as opening the arms when speaking will be added to the character. Work is also underway to improve lip sync, using automatic pre-processed mouthing based on text. The problem is that this method is hardly scalable and gives some problems with personal names when calculating the timing of each viseme. Another approach is being tried to be implemented using machine learning for real-time lip sync.

Interaction

As interaction is the most important thing for the users according to the surveys, it will be facilitated by giving some guidance to the users. To do so, different possible answers will be added both on the screen and on the tablet (selectable) as sometimes users don't know what to say to the receptionist and/or sometimes the voice is not captured well. Furthermore, if the user's request does not fit in the context of the case or the system has not solved the user's answer the second time, instead of asking the user to repeat it again, the agent will suggest a new answer similar to what he/she has understood.

Installation

Through the improvements mentioned in the previous section, the interaction can be more dynamic and provide alternatives to speech recognition. Although these characteristics cannot be changed and the installation cannot be moved, it can be improved by adding a directional microphone at the average height of a person.

Another thing that was already being considered for change is the position of the tablet. In principle, it had to be in front of the vertical screen so the user would not have to move when interacting. But the movable arm that allows the tablet to be placed in front of the screen was left immobilised when the box that encloses the screen and the computer was built. So the tablet has been left to one side, which makes it not so visible at first sight and the user has to move to interact with. This has yet to be fixed. Even so, the option of removing the tablet and using the user's mobile phone instead has been considered. That is, the user would have to scan a QR that would appear on the screen and the content of the tablet would appear on his/her phone. In this way, the mobile device's microphone could also be used without the need to add a directional microphone.

8 CONCLUSION

Present developments, ways of use, and outcomes are spread over different environments and use cases. While the experimental productions report is provided in D8.4, this document details the specific results obtained on cinema, broadcast, security, and direct interaction with final users.

While some of the conclusions of these prototype evaluations can be considered general, some others are specific to the use case. In general, we can consider that the Present technology is ready for its use in the proposed environments, and at the same time the avatars, and agent technologies still have a long way before the obtained results can be comparable to a human being. On one side there are agents, like Google LaMDA, that seem to behave and feel like human beings, but at the same time, although PRESENT has pushed the limit on agent realism, reflecting all these feelings richness on a 3D graphics virtual entity, still suffers the uncanny valley effect, and still requires the direct intervention of an AI in order to construct the required body language to communicate also through this channel.

But more in detail and focused on the specific use cases these are the main conclusions on their evaluation:

- ✓ The toolset developed by Framestore for the optimisation of animation and character creation has been highly successful in meeting or exceeding initial expectations. This success, and the appetite of current supervisors and artists to leverage the potential advantages of this technology, have led to it being deployed on active projects in late 2021. The feedback from these projects, both from the traditional visual effects side of the business as well the more forward looking realtime departments, has been very positive. Character artists have reported significant potential for improved efficiency, animators have reported a vast increase in visual fidelity and supervisors have reported an increase in the quality of submissions they are able to share with clients. This all points towards highly significant advantages to multiple existing departments and current projects. In addition to this the fact that a portable solution is now available for complex rig deformation and material graphs that works across both offline and realtime rendered characters opens up entirely new business opportunities. The first project to be delivered entirely out of Unreal Engine and with Engine rendered digital humans at its core is scheduled to be delivered by Framestore in early 2023. This project would have been exponentially more challenging and potentially not even viable without the technology developed as part of Present. There is much work still to be done to build on this early success and to improve the user experience and reliability of the toolset but once

complete this technology is likely to be widely adopted within the company and potentially transformative for the character build and animation pipelines across the business.

- ✓ Broadcast Use Case is special in different senses, first it does require a script both for the presenter in the studio and for the agent that can be analysed and tuned before the program. Mostly then TV programs are live, producers prefer not to leave any element to be decided on the go. For this reason, even if the agent needs to have access to a specific knowledge base, the questions and the corresponding answers are preferred to be set up before the program and in advance. This fact makes the implementation way easier as there is no need to connect with AI's or Knowledge bases on the cloud, and there is no need to optimise queries to be faster, as everything happens locally. On the other hand, as the agent scripts and behaviours are best to be defined just before recording, it is possible to preview the final result and modify any unexpected or unwanted behaviour on the go. In sum the PRESENT developments, once tuned for broadcast have proved to be really interesting and properly suited, as in these environments, although agent intelligence may be required, it is not required live nor real time.
- ✓ The prototype of RAO (Registration Authority Officer) shows a digital human which is incredibly realistic, details such as the skin imperfection, the muscles definition, the hair nullify the effect of 2d or flat that is typical of virtual character. Progress has been made also from expressing emotions to the tone of voice used, which improve the way how they interact with users. In the related paragraph some opportunities for improvements are reported.
- ✓ The actual installation of the Virtual Clerk in the University premises took place much later than planned (June 2022) due to the pandemics situation and networking security issues, and the evaluation is limited to this beta version. As planned, several aspects of the appearance of the character, and the interaction, have been evaluated via subjective rating. As the installation has taken place at the start of the vacation period, the number of people answering the survey has been less than expected. The overall reading of the evaluation is positive, but limitations with respect to the interaction, animations and rendering have been identified; the final installation, fixing some issues, limits the interaction. As a result, improvements in a few of these aspects have been defined and are already under development, so that a new alpha version of the Virtual Clerk can be installed by the end of the year.

ANNEX I. VIRTUAL CLERK EVALUATION QUESTIONNAIRE

- Gender
☐ Female ☐ Male ☐ Transgender ☐ Female ☐ Transgender ☐ Male ☐ Gender Variant/Non-Conforming ☐ Not Listed ☐ Prefer not to say
- Age
☐ <18 ☐ 18-24 ☐ 25-34 ☐ 35-44 ☐ 45-54 ☐ 55 years or older

Visual quality of the rendering

The following questions aim to know your opinion regarding the quality of the rendering of the virtual avatar

- From 0 to 5, rate how good is the rendering regarding:

	0 (Bad)	1	2	3	4	5 (Great)
Lipsync	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hair	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Facial expressions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Body	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- In case you have any suggestion/s on how to improve the rendering quality, please write it/them here:

Quality of the user-agent interaction

The following questions aim to know your opinion regarding the quality of the interaction with the virtual avatar

- How comfortable with the virtual avatar have you felt during the experience?

	0	1	2	3	4	5
Uncomfortable <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Comfortable						

- Please, provide a satisfaction value regarding the following properties of the avatar's answers:

	0 (Bad)	1	2	3	4	5 (Great)
Content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Delivery	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Rate from 0 to 5 the verbal communication of the virtual agent:

	0 (Artificial)	1	2	3	4	5 (Natural)
Verbal communication	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Non-verbal communication	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Rate from 0 to 5 the user-friendliness of the interaction:

	0	1	2	3	4	5
Hard to use <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Intuitive						

- In case you have any suggestion/s on how to improve the interaction, please write it/them here:

Fluency of the dialogues

The following questions aim to know your opinion regarding the fluency of the dialogues with the virtual avatar

10. How many times have you had to repeat a question to get the desired answer?
☐0 ☐1 ☐2 ☐3 ☐+4
11. "The waiting time between user-avatar interactions is too long". How much do you agree with this sentence?
 0 1 2 3 4 5
 Disagree ☐ ☐ ☐ ☐ ☐ ☐ Agree
12. "The waiting time between user-avatar interactions is too short". How much do you agree with this sentence?
 0 1 2 3 4 5
 Disagree ☐ ☐ ☐ ☐ ☐ ☐ Agree
13. Provide a satisfaction value regarding the meaningfulness of the answers of the virtual avatar to the provided question:
 0 1 2 3 4 5
 Not related ☐ ☐ ☐ ☐ ☐ ☐ Perfectly answered
14. In case you have some suggestions on how to improve the fluency of the dialogues, please write them here:
Your wish list for virtual characters
 Questions related to which properties you think a good virtual avatar has to have.
15. Choose 3 of the following properties that you think are most important in interactive embodied conversational agents:
☐Appearance (look, clothes, ethnicity...)
☐Realistic rendering
☐Interactivity
☐Verbal communication
☐Non-verbal communication
☐Naturalness
☐Reactivity
☐Intuitiveness of the interaction
☐Usefulness
16. Choose 3 of the following properties that you would highlight of THIS project:
☐Appearance (look, clothes, ethnicity...)
☐Realistic rendering
☐Interactivity
☐Verbal communication
☐Non-verbal communication
☐Naturalness
☐Reactivity
☐Intuitiveness of the interaction
☐Usefulness
17. How likely would you use this platform in your personal or professional environment?
 0 1 2 3 4 5
 Never ☐ ☐ ☐ ☐ ☐ ☐ Every day
18. How useful do you think this project is within its context?
 0 1 2 3 4 5
 Useless ☐ ☐ ☐ ☐ ☐ ☐ Highly useful
19. Some other suggestions to improve the project:

ANNEX II. FOCUS GROUP SCRIPT

Indications:

1. The objective of this activity is to collect the opinion of those attending the demonstration of Present.
2. The activity format is a group interview (focus group) for the validation of the tool developed, and this document is a script to help carrying it out.
3. Whenever possible, this activity should be recorded on video so that there is a graphic and sound record of the participants' comments. It will be used to document the activity and to further develop a report on the results. In case the session cannot be recorded on video, an audio recorder can be used. In this case, please make sure some pictures are taken during the activity.
4. The optimal number of members in a focus group is 4 or 5 people. If there are more, it is better to make two groups, separating the members by profiles, if possible.
5. The interviewer should read the 10 questions to the group of participants, leaving enough time after each one to gather their responses. The text boxes under each question include possible answers or suggestions so that the interviewer can encourage the interviewees to elaborate a bit more on their explanations, in case they are brief, but it is not necessary to read them out loud and the interviewees do not have to choose any option. They are just there to help the interviewer.
6. The script consists of two parts. The first block of questions must be asked BEFORE the Present demo. The second block, after the demo.
7. Some questions appear on both blocks, so they should be asked before and after the demo, to check if the attendees' opinions change in any way after seeing the tool.
8. The activity should not exceed 40 minutes (not counting the demo) and it is important that you complete each and every question in the script. As a recommendation, try to evaluate, based on the 3 questions in the first block, if your interviewees are going to need you to limit their response time in the second block, in order to finish the entire script in a reasonable time or if, on the contrary, you are going to need to encourage them to extend their explanations using the suggestions included in the text boxes.

Presentation:

In order to present the project and the company the host will proceed with an initial introduction stage where all the participants will introduce themselves to the rest of the group:

- The interviewer says good morning, then his/her name, surname, and company.
- The interviewer presents the attendees and asks them to introduce themselves.
- The interviewer introduces the activity to those present:
"First of all, I thank you for coming here today and for your participation in this validation group. In a moment, you will attend a demonstration of a tool that has been developed within the framework of the European project PRESENT. This tool makes possible the production of television programs in which a physical person can interact with a virtual avatar on the same stage. Therefore, the TV show revolves around the interaction between the two."

Introduction to the focus group:

In this first meeting, the system will be presented, and the first set of questions prior to the demo will be proposed:

1. First of all, do you know about television programs that incorporate avatars?
 - If so, what is your view on them?
 - If not, what about virtual reality games that incorporate realistic or hyper-realistic avatars? How do you like the idea of using them on TV programs?
 - If both answers are negative, show the group the images from the last page of this script.

<i>Do you like them?</i>	<i>Do you think that...</i>
I like them a lot / It depends on the program or game / I'm not sure if I like them / I don't like them / I've never seen them before, I don't know if I'd like them	...they are interesting, they have many possibilities? ...they are the future, but are not yet well accepted? ...are they a passing fad? ...is the trend on television and will it continue to be?

2. Can you envisage any clear application to the use of avatars on television? In which kind of program?

<i>Depending on the topic</i>	<i>Depending on the format</i>
Current affairs / News / Sport / Society / Humor / Children / Youth	Talk Show / Commentators / Main host / Assistant Host / Show's Mascot

3. Do you think that programs that incorporate avatars could be an interesting option to be explored by TV channels?

<i>Interest for TV channels</i>
Clearly yes / Clearly no / It could be / It's a matter of testing / It will depend on the cost / It will depend on the acceptance of the audience / It is a risky bet / Its future on television is uncertain / No idea

Post-demo focus group:

4. What is your perception of the acceptance that this tool may have among professionals in the field?

<i>Will they accept it?</i>	<i>What do you think their reaction will be?</i>
It will be easily accepted / Professionals love this type of novelty / It will be accepted, although it will take time / I am not able to foresee if it will be accepted / It will not be accepted at all	Enthusiasm / Suspicion / Rejection / More work or more complicated / Makes work easier

5. And the audience? How do you think they'll react to this show format?

<i>Will they accept it?</i>	<i>What will their reaction be?</i>
It will be easily accepted / Its acceptance might be difficult / I am not able to foresee the level of acceptance	Novelty / Enthusiasm / Naturalness / Indifference / Rejection

6. Now that you have seen how it works, do you think TV channels will find this tool and this program format (with human and avatar interaction) as interesting? (Same question as before the demo)

<i>Will they be interested in the tool?</i>	<i>Will they be interested in introducing avatars in the programs?</i>
Clearly yes / Clearly no / Could be / It is a matter of testing / It will depend on the cost / It will depend on the acceptance of the professionals	Clearly yes / Clearly no / It could be / The use of avatars is the future / Avatars are a passing fad / It is a risky bet / It will depend on the acceptance of the audience

7. Now that you have seen how it works, do you envisage any clear application to the "person+avatar" format on television? In which kind of program? (same question as before the demo)

demo)

Depending on the topic	Depending on the format
Current affairs / News / Sport / Society / Humor / Children / Youth	Talk Show / Talk Show / Commentators / Assistant Host / Show Mascot

8. What is your general impression of the tool?

Did you like it?	Impressions or comments
I liked it / I'm not sure / I didn't like it	It is a very interesting novelty / I am not able to give an opinion / I find it complicated

9. Do you think it can be implemented as it is? Or is it lacking maturity in any aspect: handling, visualization?

It is ready to be implemented	It lacks maturity
It is perfect / It does not seem difficult to use / The visualization is adequate	It needs to make the interface simpler / Visualization needs to be polished to make it more natural / It needs to improve other aspects (which ones?)

10. What level of investment do you think a television could assume to implement a tool like this?

Less than €15,000 / Between €15,000 and €30,000 / Between €30,000 and €60,000 / More than €60,000

Realistic avatars examples:

