WHITE PAPER

From Executive to Digital Double

Capture, Realism, Reliability and Consent for MetaHuman Digital Humans in a Business Context

Literature Review, Legal Framework and Method

Thomas Varin

Chief Technology Officer

VP Capture Studio

Montréal (Québec), Canada

info@sosie3d.com · sosie3d.com · kiosk3d.com

Version 1.4 — November 17, 2026

Suggested citation: Varin, T. (2026). *From executive to digital double: Capture, realism, reliability and consent for MetaHuman digital humans in a business context* [White paper]. VP Capture Studio.

Foreword

While developing our capture technology in Montréal, we kept running into the same three questions that any organisation asks before entrusting its image to a digital double. Will the result look strange? Can the character say just anything? And is it lawful to use a person's face and voice?

I wanted to answer them with sources rather than promises. This white paper brings together what research, law and technical documentation say today about these three questions, and how we turned that into our working method. Where data are lacking, we say so; section 10 describes what we will measure ourselves.

It is written for executives, communications teams and legal counsel considering a digital double. I hope it helps them ask the right questions, of us and of any other provider.

Your comments and criticism are welcome at info@sosie3d.com: they will shape the next version.

Thomas Varin
Chief Technology Officer, VP Capture Studio
Montréal, November 17, 2026

Abstract

Photorealistic digital humans have moved quickly from the laboratory into business use: executives already present financial results through avatars, and creation tools such as MetaHuman are now open to commercial use and partly released as open source. This white paper examines three questions: (1) does a digital double's visual realism foster public trust? (2) under what conditions is a conversational double reliable and perceived as such? (3) what legal and ethical framework governs reproducing a real person's face and voice?

A narrative review of primary sources published between 2022 and November 2026 — peer-reviewed studies, legal texts, decisions and official technical documentation — leads to three findings. First, high realism is associated with greater trust, while intermediate realism is the most unsettling. Second, the perceived reliability of a conversational agent depends above all on its ability to abstain when its sources are insufficient and to mask latency through natural behaviour. Third, despite the diversity of Québec, European and U.S. regimes, written, specific, separate and revocable consent is emerging as the common standard. From this we derive a seven-step production method, ten good-practice principles, a supplier evaluation grid and a measurement programme.

Keywords: *digital human; digital double; MetaHuman; photogrammetry; embodied conversational agent; cloned voice; trust; uncanny valley; Québec Law 25; consent; content provenance*

Notice

The factual findings in this document refer to verified public sources, cited in APA style (7th edition). Market estimates are reproduced as published by their authors, without endorsement. The state of the law is described as of November 17, 2026; this document is not legal advice. MetaHuman and Unreal Engine are trademarks of Epic Games, Inc.; VP Capture Studio uses these tools under licence and is not affiliated with Epic Games.

Contents

1. Introduction 6
1.1 Context 6
1.2 Research questions 6
1.3 Contribution and outline 6
2. Definitions and scope 7
3. Review method 7
4. State of the technology 8
4.1 Face and head capture 8
4.2 From scanned model to animatable character 8
4.3 Face, body and voice animation 9
4.4 Rendering 9
4.5 Conversational pipeline 9
5. Literature review: perception, voice and conversation 10
5.1 The uncanny valley, from hypothesis to data (RQ1) 10
5.2 Photorealism, trust and detection 10
5.3 The cloned voice 10
5.4 Social presence 11
5.5 Embodied agents in real settings (RQ2) 11
5.6 Latency and natural fillers 11
5.7 Personalised video 11
5.8 Synthesis 12
6. Legal and ethical framework (RQ3) 12
6.1 Québec: privacy, image and voice 12
6.2 Personal information and biometrics 12
6.3 Liability for the agent's statements 13
6.4 Canada 13
6.5 International 13
6.6 Professional norms 14
6.7 Provenance and fraud prevention 14
6.8 Ethical issues 14
7. Market and business uses 15
7.1 Published estimates 15
7.2 Market signals 15
7.3 Families of use 16
8. The VP Capture method 16
8.1 Scoping and consents 16
8.2 Capture 17

8.3 Reconstruction, conforming and validation 17
8.4 Reliability of the conversational double 18
8.5 Deployment, transparency and end of life 18
9. Discussion 19
9.1 Answers to the research questions 19
9.2 Ten good-practice principles 19
9.3 Supplier evaluation grid 20
10. Limitations and measurement programme 20
11. Conclusion 21
Glossary 21
References 22

1. Introduction

1.1 Context

In May 2025, two listed companies had their chief executive's avatar present quarterly results. Klarna's opened an 83-second video by introducing itself as the CEO's AI avatar; Zoom's delivered the opening remarks before the real executive answered questions live (Business Today, 2025; TechCrunch, 2025). The reception was mixed: observers noted a fixed gaze and mechanical blinking (TechSpot, 2025). These two episodes sum up the issue: the use case exists, but its credibility depends on execution quality.

The same technology is also used for fraud. In January 2024, an employee in the Hong Kong office of engineering firm Arup made 15 transfers totalling HK$200 million after a video conference in which the chief financial officer and several colleagues were digital imitations built from public videos (CFO Dive, 2024; Dezeen, 2024). A legitimate digital double must therefore stand clearly apart from a deepfake: through the consent of the person represented, transparency towards the public and traceability of content.

Between these two realities, the tools have changed scale. MetaHuman, Epic Games' environment for creating and animating digital humans, left early access in 2025 with a licence that permits commercial use outside Unreal Engine (Thacker, 2025), and its version 5.8 releases part of its core libraries as open source (Epic Games, 2026c). What used to require a visual-effects studio is becoming accessible to mid-sized organisations. The question is no longer whether a credible double can be produced, but how to do so reliably, lawfully and in a trustworthy way.

1.2 Research questions

This white paper pursues three questions:

1. **RQ1.** Does a digital double's visual realism foster or undermine public trust?
2. **RQ2.** Under what conditions is a conversational double, answering out loud, reliable and perceived as such?
3. **RQ3.** What legal and ethical framework governs reproducing a real person's face and voice, in Québec and among its main partners?

1.3 Contribution and outline

Our contribution is twofold: an up-to-date synthesis of technical, scientific and legal knowledge; and an operational method derived from it, applied in our Montréal studio. Section 2 sets out definitions; section 3, the review method; section 4 describes the state of the technology; section 5 reviews the literature on perception and conversation; section 6 analyses the legal and ethical framework; section 7 examines the market and uses; section 8 presents our method; section 9 discusses the results

and proposes good-practice principles; section 10 sets out limitations and the measurement programme.

2. Definitions and scope

The field's vocabulary is unstable: "avatar", "virtual human", "digital human", "double", "replica" and "deepfake" refer to related but distinct objects. Since legal regimes rely on precise definitions, we adopt those in Table 1.[1]

Table 1

Definitions adopted

Term	Definition	Source
Digital human	Photorealistic synthetic character, animated and, where applicable, conversational.	Common usage
Digital double	Digital human reproducing the appearance, movement and, where applicable, voice of a real person, with their consent.	This document
Digital replica	Computer-generated, highly realistic electronic representation readily identifiable as a person's voice or visual likeness.	California, AB 2602 and AB 1836
Deepfake	AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear authentic.	EU, AI Act, Art. 3(60)
Embodied conversational agent	Software agent with a visible body that converses through speech, gaze and gesture.	Cassell et al. (2000)

Note. The AI Act and California definitions are summarised; see Zolynski (2026) and Davis+Gilbert (2024) for the full text.

The central distinction is consent: a digital double and a deepfake can be technically identical; only the person's consent, and transparency towards the public, separate them. Our scope covers doubles of real people in a business context — executives, spokespersons, trainers — and excludes entirely invented characters, deceased persons and political uses.

3. Review method

We carried out a targeted narrative review in November 2026, focusing mainly on 2022–2026 and supplemented by older foundational work where it structures the field (Mori et al., 2012; Reeves & Nass, 1996). Three families of sources were retained: (a) empirical studies that are peer-reviewed or presented at refereed conferences (PNAS, Psychological Science, Scientific Reports, Frontiers in Psychology, Journal of Science Communication, ICLR, CVPR, ACM CUI, ISPRS); (b) legislative texts,

[1] We use "digital double" for the representation of a real, consenting person, and "digital human" for any photorealistic representation, including invented characters.

decisions and documents from public authorities; (c) vendors' official technical documentation. Trade press was used only to date events and was systematically cross-checked.

Two exclusion criteria were applied: any figure whose primary source could not be traced, and any legal statement whose legislative stage could not be verified. The second criterion proved necessary: several publications present as adopted texts that are still only bills (section 6.5). This is a narrative, not a systematic, review; its limitations are discussed in section 10.

Literature searching and drafting were carried out with the help of generative artificial intelligence tools. The author verified the sources cited and takes full responsibility for the content.

4. State of the technology

Producing a digital double chains four operations: capturing the person, building an animatable character, animating it (face, body, voice) and rendering it. When it is conversational, a dialogue pipeline is added. We examine each link.

4.1 Face and head capture

Photogrammetry reconstructs a 3D surface from many photographs taken from different angles: structure-from-motion estimates the camera positions, then multi-view stereo densifies the geometry (Schönberger & Frahm, 2016). For a human face, capture must be simultaneous, since the slightest movement between shots degrades the reconstruction; hence the use of camera arrays fired together, often with a flash to freeze the expression.

These arrays no longer require cinema hardware. Giuliani et al. (2024) evaluated a multi-view system based on cameras with a Sony IMX477 sensor (12.3 megapixels) and 16 mm lenses to produce 3D face models: with 27 viewpoints over three rows and ten scale measurements, the standard deviation reaches 0.9 mm and 80% of points lie within ±1 mm of a reference model.[2] The authors nonetheless present their work as a feasibility study: sub-millimetre accuracy holds for only 80% of points, and trials on human volunteers were evaluated only qualitatively. Kersten et al. (2024) describe a synchronous 24-camera system of the same type. These studies establish that rigorous multi-view capture, with careful scale calibration, approaches millimetre accuracy at controlled cost.

4.2 From scanned model to animatable character

A point cloud is not a character. To speak and express emotion, the model must be "conformed" to a standard topology equipped with a rig — a skeleton and a set of facial controls. Modern facial rigs inherit from the Facial Action Coding System (FACS), which breaks expressions down into muscular action units (Ekman & Friesen, 1978). MetaHuman provides such a standardised rig: Mesh to

[2] In the study, only three cameras were moved to nine positions to simulate the 27 viewpoints; a real system with simultaneous triggering also eliminates subject movement between shots.

MetaHuman conforms a scanned model to the MetaHuman topology, and version 5.8 extends this conforming to the whole body, which precisely suits characters built from scans (Epic Games, 2026b).

The licensing framework evolved in parallel. Since version 5.6 (June 2025), use of MetaHuman falls under the Unreal Engine licence; characters may be used in other engines and software, including in commercial projects, and only studios earning over one million U.S. dollars a year must acquire per-seat licences (Epic Games, n.d.; Thacker, 2025). The MetaHuman Creator web application was discontinued in March 2026, creation now being built into Unreal Engine (Epic Games, 2026a). Finally, Epic released under the MIT licence the RigLogic and DNA libraries, which evaluate the MetaHuman facial rig outside Unreal Engine (Epic Games, 2026c) — a guarantee of longevity for a character whose useful life exceeds that of a software version. Unreal Engine 5.8 is presented as the last planned major UE5 release before UE6 (Epic Games, 2026d).

4.3 Face, body and voice animation

Three animation sources coexist. The first is captured performance: MetaHuman Animator can now capture the full performance — face, body or both — from a single camera, and real-time facial animation accepts a streamed video input (Epic Games, 2026b). The second is multi-camera markerless motion capture: open-source software such as Pose2Sim triangulates 2D pose detections to obtain full-body 3D kinematics (Pagnon et al., 2022). The third is voice-driven animation: in September 2025, NVIDIA released as open source the models, SDK and training framework of Audio2Face, which produces facial animation and lip-sync from audio (Animation World Network, 2025).

The voice itself can be cloned from recordings of the person. As we shall see (section 5.3), the public struggles to tell a cloned voice from the real one; the voice must therefore be treated as an attribute as identifying as the face, both for data security and for consent.

4.4 Rendering

Real-time rendering, at the heart of Unreal Engine, enables interaction but imposes a per-frame compute budget; pre-rendering reaches higher quality but rules out improvisation. A third path is emerging: Gaussian splatting represents a scene with millions of 3D Gaussians rendered in real time (Kerbl et al., 2023), and its variants produce relightable photorealistic heads, including hair and fine reflections (Saito et al., 2024; Zhang et al., 2025). In the short term these approaches complement the rigged character rather than replace it: the standardised rig remains the asset for directed animation and interoperability.

4.5 Conversational pipeline

A conversational double chains speech recognition, information retrieval, answer drafting, speech synthesis and animation. Retrieval-augmented generation supplies the language model with

passages retrieved from a document base at answer time, rather than relying on its parameters alone (Lewis et al., 2020). This architecture reduces fabrication without eliminating it: when the retrieved context is insufficient, models tend to produce an answer rather than abstain (Joren et al., 2025). Each link also adds latency, whose effect on perception we examine below (section 5.6). Figure 2 (section 8) shows the architecture we derive from this.

5. Literature review: perception, voice and conversation

5.1 The uncanny valley, from hypothesis to data (RQ1)

Formulated by roboticist Masahiro Mori in 1970, the uncanny valley hypothesis predicts that affinity for a humanoid figure rises with its human likeness, then drops sharply when the figure is almost, but not quite, human (Mori et al., 2012). The hypothesis long served as an argument against photorealism. Recent data strongly qualify it.

A pre-registered network meta-analysis of 13 studies and 2,343 participants published between 2015 and 2025 concludes that avatars with high realism are rated more attractive and more trustworthy than those with medium or low realism, and that medium realism causes the most unease (Tao et al., 2025).[3] An experiment with 491 people, in a science-communication context, points the same way: the more realistic the avatar, the more trustworthy it is rated, with no uncanny valley effect; familiarity with AI and trust in the institution moderate this perception (Baake et al., 2025).

5.2 Photorealism, trust and detection

For still images, the line has already been crossed. Faces generated by adversarial networks proved indistinguishable from real faces and were rated on average 7.7% more trustworthy (Nightingale & Farid, 2022). Miller et al. (2023) even observe "hyperrealism" — synthetic faces judged more human than real ones — but only for white faces, which they attribute to the over-representation of such faces in training data; the participants most deceived were also the most confident. Two lessons follow: perceived authenticity is no longer a reliable safeguard, and rendering quality can vary with the characteristics of the person represented.

5.3 The cloned voice

Barrington et al. (2025) show that voices cloned from short recordings with a consumer tool are perceived as having the same identity as the real voice about 80% of the time, and are recognised as artificial only about 60% of the time; performance improves for longer clips. The voice is therefore a full identity attribute, calling for the same level of consent and protection as the face.

[3] The meta-analysis ranks realism levels using the surface under the cumulative ranking curve (SUCRA); the protocol was pre-registered on the Open Science Framework.

5.4 Social presence

Media equation theory holds that people spontaneously apply to media and computers the social rules they apply to humans (Reeves & Nass, 1996). A character that looks, speaks and gestures triggers these rules more strongly than text. That is the digital double's asset — a presence text cannot provide — and its responsibility: an error spoken by a familiar face commits more than an error shown on screen.

5.5 Embodied agents in real settings (RQ2)

Embodied conversational agents have been studied in real settings for over fifteen years, notably as museum guides: the virtual twins Ada and Grace have welcomed visitors to Boston's Museum of Science since 2009 (Traum et al., 2012). A recent evaluation of a conversational avatar deployed in a science museum, with 118 visitors, reports that visitors often ask the avatar about itself and that its answers are of good quality about half the time; the authors stress the importance of keeping the underlying technology at a high level (InformalScience, 2025). That rate is a reminder that engagement is worthless without accuracy.

5.6 Latency and natural fillers

In a within-subjects study of 54 participants interacting in virtual reality with agents powered by a large language model, in store, hotel and museum scenarios, Maslych et al. (2025) compared delays of 1.5, 4 and 6.5 seconds. Delay degrades most measures, and the authors conclude that latency above 4 seconds worsens the experience. "Natural fillers" — a thinking gesture with a short phrase — clearly improve perceived response time at medium and long delays, whereas an artificial indicator (loading icon and sound) brings no significant improvement. Participants say they are more willing to use a slow system when fillers are present, while still preferring fast answers.

5.7 Personalised video

In a randomised field experiment with 21,328 customers of a retailer, personalised video ads produced with generative AI achieved an engagement rate 6 to 9 percentage points higher than personalised images and generic videos, with large production savings (Kapoor & Kumar, 2025). Dennis and Kizer (2025) also observe more downloads, rewatches and clicks for personalised videos. These results concern consumer marketing and a single exposure; their transfer to internal communications — a president's greetings to each employee, for example — remains to be measured.

5.8 Synthesis

Table 2

Summary of findings and design implications

Question	Main finding	Implication
Realism (RQ1)	High realism = more trust; medium realism = most unease (Tao et al., 2025; Baake et al., 2025).	Aim for high realism or abstain.
Detection	The public struggles to tell real from synthetic, for faces and voices (Nightingale & Farid, 2022; Barrington et al., 2025).	Transparency is the producer's duty.
Reliability (RQ2)	Without sufficient context, models invent rather than abstain (Joren et al., 2025).	Explicit abstention and approved sources.
Latency (RQ2)	Beyond 4 s, experience degrades; natural fillers help (Maslych et al., 2025).	Budget < 4 s and an "inhabited" wait.
Personalisation	+6 to 9 points of engagement in marketing (Kapoor & Kumar, 2025).	Promising; to be measured internally.

6. Legal and ethical framework (RQ3)

6.1 Québec: privacy, image and voice

In Québec, the right to privacy is protected by the Charter of Human Rights and Freedoms (s. 5) and by the Civil Code, which notably considers it an invasion to use a person's name, image, likeness or voice for any purpose other than the legitimate information of the public (C.C.Q., arts. 35-36). A digital double used for commercial purposes therefore requires, in principle, the consent of the person represented.

6.2 Personal information and biometrics

The Act respecting the protection of personal information in the private sector, modernised by Law 25, requires consent to be manifest, free, enlightened and given for specific purposes, and to be requested for each such purpose (s. 14; Usercentrics, 2026). The Act to establish a legal framework for information technology further requires express consent to verify or confirm a person's identity by biometrics, and the declaration to the Commission d'accès à l'information of any bank of biometric characteristics no later than 60 days before it is put into service (ss. 44-45; Gowling WLG, n.d.). The Commission enforces these rules strictly, as shown by its decision on a retailer's facial-recognition project (Torys, 2025).

A digital double is not used to identify a person; it is nonetheless built from precise measurements of their face. As a precaution, we treat this data as sensitive information, and the application of sections 44 and 45 to each project must be validated by legal counsel.[4]

6.3 Liability for the agent's statements

In Moffatt v. Air Canada (2024), the carrier was held liable for wrong information on bereavement fares given by the conversational agent on its website; the tribunal rejected the argument that the agent was a separate entity responsible for its own actions, and found negligent misrepresentation.[5] For a digital double, the consequence is direct: the business is liable for what the face of its president or spokesperson says, as for any other official communication.

6.4 Canada

Federally, the proposed Artificial Intelligence and Data Act, part of Bill C-27, died on the Order Paper when Parliament was prorogued on January 6, 2025; no text has replaced it, and Canada still has no general federal AI law (CASRAI, 2026). Oversight therefore rests on existing privacy, civil liability and consumer-protection law.

6.5 International

In the European Union, the transparency obligations of Article 50 of the AI Act have applied since August 2, 2026: deployers must disclose that a deepfake has been artificially generated or manipulated, and providers must mark synthetic content in a machine-readable way; a code of practice published in June 2026 and guidelines from July 2026 set out implementation (Mishcon de Reya, 2026). In the United States, several states already regulate digital replicas, while a federal law is still under consideration (Table 3). In Denmark, a bill granting everyone a right in their face and voice was tabled on October 8, 2026; several publications wrongly presented it as adopted (The AI Musicpreneur, 2026).

[4]The Moffatt decision comes from the British Columbia Civil Resolution Tribunal (small claims); it is not binding precedent in Québec, but illustrates the reasoning applicable to negligent misrepresentation.

[5]Whether sections 44 and 45 of the Act apply depends on purpose: they cover verifying or confirming identity and banks of biometric characteristics. Classifying each project is a matter for legal advice.

Table 3

Main regimes governing digital replicas as of November 17, 2026

Jurisdiction	Instrument	Status	Main requirement
Québec	C.C.Q. arts. 35-36; Law 25; IT Act ss. 44-45	In force	Consent; biometrics declared to the CAI
European Union	AI Act, Art. 50	In force since Aug. 2, 2026	Deepfake disclosure; marking
United States (federal)	NO FAKES Act (S.4591)	Approved in committee; not passed	Federal right in voice and likeness
California	AB 2602 ; AB 1836	In force since Jan. 1, 2025	Specific description of uses in contract; deceased persons
New York	S7676B	In force since Jan. 1, 2025	Same contractual requirement
Tennessee	ELVIS Act	In force (2024)	Bans non-consensual voice imitation
Denmark	Bill L 49	Tabled Oct. 8, 2026; not passed	Right in one's face and voice (target: 2027)

Note. Sources: Davis Wright Tremaine (2025); Mishcon de Reya (2026); Morrison Foerster (2025); Paubox (2026); The AI Musicpreneur (2026); U.S. Congress (2026).

6.6 Professional norms

The film industry has set a standard useful to business. The 2023 SAG-AFTRA agreement requires informed, clear and conspicuous consent — not buried in standard terms — to create and use a performer's digital replica; it provides for compensation, and separate consent for each new project (Perkins Coie, 2024; SAG-AFTRA, 2023). The California and New York laws likewise require, for validity, that digital-replica clauses describe the intended uses with reasonable specificity (Davis Wright Tremaine, 2025).

6.7 Provenance and fraud prevention

The C2PA standard (Content Credentials) lets a cryptographically signed provenance history be attached to an image, video or sound; it counts over 6,000 members and affiliates (TrueScreen, 2026), version 2.3 extended provenance to live video and version 2 is on the ISO fast track (PPC Land, 2026). Provenance certifies a content's history, not its truth; combined with consent and internal protocols — for example, never authorising a transfer on the strength of a video alone — it helps distinguish the authorised double from the fraudulent imitation illustrated by the Arup case.

6.8 Ethical issues

Dignity and control. The person must be able to approve each piece of content, refuse a use and withdraw consent; their double must never be made to say what they would not have said.

Ownership and dependency. A digital-human pioneer, placed in receivership in February 2026, contractually reserved ownership of its clients' custom avatars (BERI, 2026; UneeQ, 2026). The

question “who owns my double, and what happens to it if the provider disappears?” must be settled in the contract.

Fairness. The biases observed in the perception of synthetic faces (Miller et al., 2023) require checking that the double’s quality does not vary with the person’s skin colour, age or gender.

After the person. Death, departure from the company or the end of a mandate must be provided for: California already regulates replicas of deceased persons (AB 1836), and the SAG-AFTRA agreement then requires the consent of the estate (Davis Wright Tremaine, 2025; SAG-AFTRA, 2023).

7. Market and business uses

7.1 Published estimates

Market size estimates diverge by an order of magnitude, mainly because firms do not define the market the same way.[6] We reproduce them for reference.

Table 4

Published market estimates

Firm	Scope	Published estimate
Market Glass	Digital human avatars	$14.1B (2024) → $116.7B (2030)
Technavio	Digital human avatars	$7.47B (2025); 43% annual growth (2026-2030)
MarketsandMarkets	AI avatars	$0.80B (2025) → $5.93B (2032)

Note. Sources: Market Glass (2025); MarketsandMarkets (n.d.); Technavio (2026).

7.2 Market signals

Two signals are more reliable than these projections. On the one hand, investment: a British leader in avatar video was valued at four billion U.S. dollars in January 2026 and is steering its growth towards conversational agents (HLTH, 2026). On the other, fragility: a New Zealand digital-human pioneer went into receivership in February 2026 (UneeQ, 2026). The market rewards production at scale but punishes models that do not find lasting uses; for a client, the provider’s durability and the double’s portability become purchasing criteria.

[6] Amounts are in U.S. dollars, as published. Scopes differ: some firms include virtual influencers and video games, others only business avatars.

7.3 Families of use

Table 5

Business uses of the digital double: value, risks and safeguards

Use	Expected value	Main risk	Safeguard
Leadership communication (results, announcements)	The executive's presence without a shoot	Loss of credibility if rendering is approximate	High realism; explicit disclosure; approval
Welcome and advice kiosk (store, hotel, trade show)	Continuous, multilingual welcome with social presence	Wrong answer binding the business	Approved sources; abstention; human hand-off
Web version of the answers	Same service online, without per-visitor compute	Outdated content	Validated, dated bank; periodic review
Virtual reality (events, launches)	Strong sense of presence	Discomfort if realism is medium	Careful rendering; short sessions
Personalised greetings and messages	Named message at scale	Impression of manipulation	Transparency; script approved by the person
Training	Gesture shown by the right person, in every language	Gap between gesture and procedure	Validation by the subject-matter expert

8. The VP Capture method

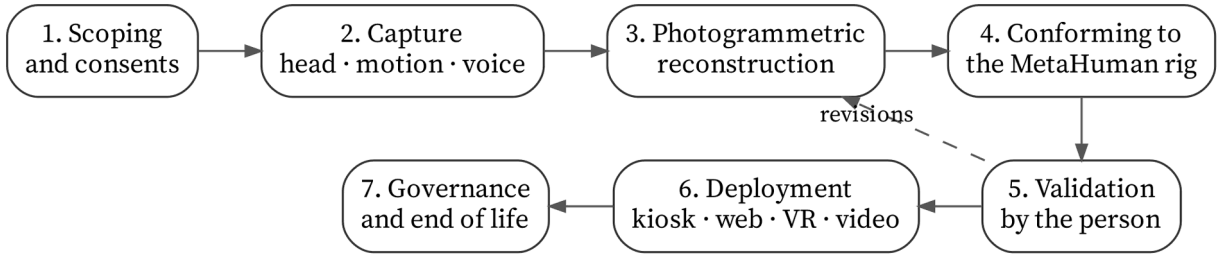
The method we apply turns the findings of sections 4 to 6 into practice. It is described as designed; its performance measurements will be the subject of a separate publication (section 10). It is organised in seven steps (Figure 1) and four cross-cutting requirements: realism, reliability, consent and governance.

8.1 Scoping and consents

Before any capture, the intended uses are described precisely and the person signs two separate consents: one for their digital double, the other for their voice. Each lists the authorised uses, can be withdrawn at any time and provides for erasure of the data within 30 days of withdrawal. This practice follows the spirit of section 14 of Québec's private-sector act, the SAG-AFTRA agreement and the California and New York laws (sections 6.2 and 6.6).

Figure 1

Seven-step digital double production pipeline



Note. Validation by the person may send the model back to the reconstruction step (dashed arrow).

8.2 Capture

- **Head and face.** Columns of cameras with Sony IMX477 sensors on stands, fired simultaneously with flash while the person turns on a seat. Two options: Regular (84 photographs) and High quality (132 photographs, with a pause before expressions). The multi-view approach with scale calibration follows the conditions that, in the literature, approach millimetre accuracy (Giuliani et al., 2024).
- **Motion.** Suit-free, marker-free capture by several calibrated cameras, triangulated with open-source multi-view kinematics software (Pagnon et al., 2022), for the person's walk, gestures and way of greeting.
- **Voice.** Guided recording with prepared scripts and every take checked; cloning is processed on our own servers rather than at a third party.

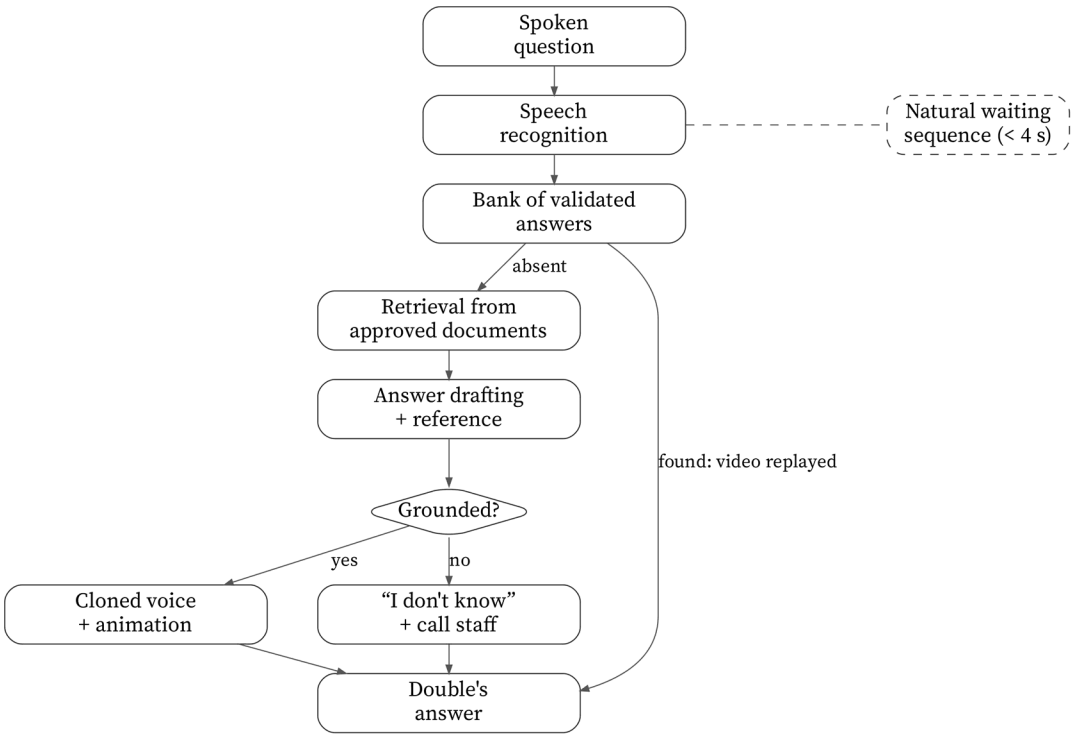
8.3 Reconstruction, conforming and validation

The reconstructed model is conformed to the MetaHuman rig, giving a standardised character that can be animated and moved between tools (section 4.2). The explicit goal is high realism: an intermediate result is not delivered, since that is the zone the literature associates with unease (Tao et al., 2025). The person validates their double before any use; their comments may send the model back to reconstruction. Quality is checked with the same care whatever the person's skin colour, age or gender (section 6.8).

8.4 Reliability of the conversational double

Figure 2

Conversational architecture with a bank of validated answers and abstention



Note. The waiting sequence starts as soon as the question ends, so that the perceived wait stays under the 4-second threshold reported by Maslych et al. (2025).

- **Abstain rather than invent.** The character answers only from the client's approved documents, with the official reference; otherwise it says "I don't know" and offers to call a staff member (Joren et al., 2025; Moffatt v. Air Canada, 2024).
- **Bank of validated answers.** Frequent questions are verified by the client, rendered once at film quality, then replayed. This is also what makes the web version possible without graphics compute per visitor; each new question is rendered once then added to the bank.
- **Inhabited latency.** Pre-prepared sequences — thinking gestures, movements — cover processing time, in line with the natural fillers studied by Maslych et al. (2025).

8.5 Deployment, transparency and end of life

- **Transparency.** The public is told it is addressing a digital double, in keeping with the spirit of Article 50 of the AI Act.
- **Approval of each piece of content.** The person approves each video before release.
- **Data minimisation at the kiosk.** No facial recognition of visitors; photo-booth photographs and videos are erased automatically after 7 days.

- **End of life.** Departure, withdrawal of consent or the end of the contract trigger erasure within the agreed period.

9. Discussion

9.1 Answers to the research questions

RQ1. Recent literature does not confirm the fear of excessive realism: high realism fosters trust, and approximation destroys it (Tao et al., 2025; Baake et al., 2025). The corollary is ethical: since the public struggles to tell real from synthetic (Nightingale & Farid, 2022), the duty of transparency falls on the producer. **RQ2.** Reliability rests less on the model's intelligence than on two disciplines: knowing when to abstain for lack of sources, and inhabiting the wait with natural behaviour. **RQ3.** Despite different regimes, Québec, the European Union and several U.S. states converge on written, specific, separate and revocable consent, supplemented by transparency towards the public.

9.2 Ten good-practice principles

1. **Aim for high realism or don't do it.** Medium realism is the zone of unease.
2. **Capture before generating.** A calibrated multi-view scan of the real person beats an approximate likeness.
3. **Treat the voice like the face.** It is just as identifying and requires the same consent.
4. **Consent in writing, use by use.** Specific description of uses; separate consent for each new project.
5. **Keep the person in control.** Approval of each piece of content, withdrawal at any time, erasure within a set period.
6. **Settle ownership of the double.** The contract states who owns the data and the character, and what happens if the provider disappears.
7. **Prefer "I don't know" to error.** Answers grounded in approved sources, with a reference.
8. **Stay under 4 seconds.** And inhabit the wait with natural behaviour, not an icon.
9. **Tell the public it is talking to a double.** Transparency protects trust and meets emerging obligations.
10. **Make content traceable.** Signed provenance helps tell the authorised double from the deepfake.

9.3 Supplier evaluation grid

The principles above translate into questions an organisation can put to any digital-double provider:

Table 6

Evaluation grid for a digital-double provider

Area	Question to ask
Capture	Is the double built from a calibrated multi-view capture of the real person?
Realism	Can I see the double in close-up, moving and speaking, before signing?
Consent	Are the image and voice consents separate, specific and revocable?
Approval	Does the person approve each piece of content before release?
Ownership	Who owns the data and the character? What happens if you cease operations?
Data	Where are face and voice data processed and stored, and for how long?
Reliability	How does the character abstain when it does not know? Do its answers cite their source?
Latency	What is the measured response time, and how is the wait filled?
Transparency	How is the public told it is addressing a double?
Provenance	Do the videos produced carry verifiable provenance?

10. Limitations and measurement programme

This review has limitations. It is narrative, not systematic; other relevant work may have escaped our search. Several perception studies concern still images or short videos in the laboratory; their conclusions must be confirmed for animated, conversational characters, in real settings and over time. The personalised-video results come from consumer marketing. The capture accuracy cited comes from a feasibility study. Finally, the law is moving fast: several texts cited are still only bills, and the state described here will need updating.

To close these gaps, VP Capture Studio will measure its own pilot deployments and publish the results according to the indicators in Table 7.

Table 7

Measurement programme indicators

Indicator	Definition	Method
Geometric accuracy	Deviation between the double and a measured reference	Point-cloud comparison on a calibration object and volunteers
End-to-end latency	Delay between the end of the question and the start of the answer	Time-stamped logging at the kiosk
Accuracy	Share of grounded answers, abstentions and errors	Review of a sample by the client
Perceived trust	Attributed expertise, integrity and benevolence	Validated questionnaire after the interaction
Internal reception	Viewing and reactions to personalised messages	Delivery statistics and anonymous survey

11. Conclusion

Digital double technology has matured faster than the practices that govern it. The tools are powerful and increasingly open; the public struggles to tell real from synthetic; and the law, still fragmented, is converging on a simple requirement: the person's informed consent and transparency towards the public. In this context, the advantage goes not to whoever produces fastest, but to whoever produces most rightly: uncompromising realism, a character that knows to stay silent when it does not know, and a real person who keeps control of their image and voice from start to finish. That is the method we apply, and the one we invite organisations to demand from any provider.

Glossary

Abstention. Reasoned refusal to answer when available sources do not support a grounded answer.

Scale calibration. Operation that gives the 3D model its real dimensions from known distances.

Conforming. Fitting a scanned model to a standard topology and rig.

C2PA. Open provenance standard that attaches a signed history to digital content.

FACS. Facial Action Coding System, which breaks expressions into action units.

Retrieval-augmented generation. Architecture that supplies the language model with document passages at answer time.

Photogrammetry. 3D reconstruction from photographs taken from several angles.

Rig. Skeleton and controls that make a character animatable.

Uncanny valley. Drop in affinity for a figure that is almost, but not quite, human.

References

Act respecting the protection of personal information in the private sector, CQLR c P-39.1, s. 14.

Act to establish a legal framework for information technology, CQLR c C-1.1, ss. 44-45.

Animation World Network. (2025, September). *NVIDIA open sources Audio2Face animation model*. https://awn.com/news/nvidia-open-sources-audio2face-animation-model

Baake, J., Schmitt, J. B., & Metag, J. (2025). Balancing realism and trustworthiness: AI avatars in science communication. *Journal of Science Communication, 24*(02), A03. https://doi.org/10.22323/2.24020203

Barrington, S., Cooper, E. A., & Farid, H. (2025). People are poorly equipped to detect AI-powered voice clones. *Scientific Reports, 15*, 11004. https://doi.org/10.1038/s41598-025-94170-3

BERI. (2026). *AppDirect buys Soul Machines, which owns your custom avatar*. https://www.beri.net/article/appdirect-soul-machines-acquisition-receivership-custom-avatar-ownership-contract-terms

Business Today. (2025, 23 May). *Tech CEOs use AI avatars for earnings calls*. https://www.businesstoday.in/technology/news/story/tech-ceos-use-ai-avatars-for-earnings-calls-signalling-a-shift-in-corporate-communication-477455-2025-05-23

CASRAI. (2026). *AIDA (Bill C-27): What it would have required*. https://casrai.org/guides/aida-bill-c27-what-it-would-have-required

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. MIT Press.

CFO Dive. (2024). *Scammers siphon $25M from engineering firm Arup via AI deepfake CFO*. https://www.cfodive.com/news/scammers-siphon-25m-from-engineering-firm-arup-via-ai-deepfake-cfo/716501/

Charter of Human Rights and Freedoms, CQLR c C-12, s. 5.

Civil Code of Québec, CQLR c CCQ-1991, arts. 35-36.

Davis Wright Tremaine. (2025). *Lights, camera, legislation: Are your entertainment contracts AI ready?* https://dwt.com/insights/2025/03/state-laws-regulating-ai-in-entertainment-industry

Davis+Gilbert. (2024). *California passes new legislation prohibiting unauthorized AI replicas*. https://www.dglaw.com/california-passes-new-legislation-prohibiting-unauthorized-ai-replicas/

Dennis, M., & Kizer, T. H. (2025). *Lights, camera, engagement! The role of personalized video as a marketing strategy in capturing consumer attention*. https://scholarship.rollins.edu/as_facpub/420

Dezeen. (2024, 17 May). *Arup victim of multimillion-dollar deepfake video scam in Hong Kong*. https://www.dezeen.com/2024/05/17/arup-victim-deepfake-video-scam/

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Consulting Psychologists Press.

Epic Games. (n.d.). *Unreal Engine — Licensing FAQ*. https://gateway.unrealengine.com/en-US/faq

Epic Games. (2026a). *MetaHuman news*. https://www.metahuman.com/en-US/news

Epic Games. (2026b). *MetaHuman 5.8 release notes in Unreal Engine*. https://dev.epicgames.com/documentation/metahuman/metahuman-5-8-release-notes-in-unreal-engine

Epic Games. (2026c). *OpenRigLogic* [Software, MIT licence]. GitHub. https://github.com/EpicGames/openriglogic

Epic Games. (2026d). *Unreal Engine 5.8 is now available*. https://www.unrealengine.com/en-US/news/unreal-engine-5-8-is-now-available

Giuliani, S., Tosti, F., Lopes, P., Ciampini, C., & Nardinocchi, C. (2024). Design of a multi-view photogrammetric system based on low cost cameras for the 3D forensic face recognition. *ISPRS Archives, XLVIII-2/W8-2024*, 177–183. https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-177-2024

Gowling WLG. (n.d.). *Biometrics: Québec legal framework*. https://gowlingwlg.com/getmedia/7ec810e8-6b8e-4d3d-bda6-ee3b568d60dd/Gowling-WLG-Quebec-Biometrie-Legal-Framework-Fr-v2.pdf.xml

HLTH. (2026, 27 January). *Synthesia raises $200M, with $4B valuation*. https://hlth.com/insights/news/synthesia-raises-200m-with-4b-valuation-2026-01-27

InformalScience. (2025). *The impact of a conversational agent avatar on the experience of museum visitors* [Evaluation]. https://informalscience.org/evaluation/the-impact-of-a-conversational-agent-avatar-on-the-experience-of-museum-visitors/

Joren, H., Zhang, J., Ferng, C.-S., et al. (2025). Sufficient context: A new lens on retrieval augmented generation systems. *ICLR 2025*. https://arxiv.org/abs/2411.06037

Kapoor, A., & Kumar, M. (2025). *Frontiers: Generative AI and personalized video advertisements*. https://ide.mit.edu/insights/personalized-ai-video-ads

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics, 42*(4), 139. https://doi.org/10.1145/3592433

Kersten, T. P., Timm, F., & Zobel, K. (2024). Development of a photogrammetric 3D measurement system for small objects using Raspberry Pi cameras as low-cost sensors. *ISPRS Archives, XLVIII-2/W8-2024*, 219–225. https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-219-2024

Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*. https://arxiv.org/abs/2005.11401

Market Glass. (2025, November). *Digital human avatars: Global market report*. MarketResearch.com. https://www.marketresearch.com/Market-Glass-Inc-v1039/Digital-Human-Avatars-42950110

MarketsandMarkets. (n.d.). *AI avatar market*. https://www.marketsandmarkets.com/Market-Reports/digital-human-market-212510229.html

Maslych, M., Katebi, M., Lee, C., et al. (2025). Mitigating response delays in free-form conversations with LLM-powered intelligent virtual agents. *CUI '25*. ACM. https://arxiv.org/abs/2507.22352

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science, 34*(12), 1390–1403. https://doi.org/10.1177/09567976231207095

Mishcon de Reya. (2026). *AI Act transparency obligations: Code of practice and draft guidelines*. https://www.mishcon.com/news/ai-act-transparency-obligations-code-of-practice-and-draft-guidelines

Moffatt v. Air Canada, 2024 BCCRT 149 (CanLII). https://www.canlii.org/en/bc/bccrt/doc/2024/2024bccrt149/2024bccrt149.html

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morrison Foerster. (2025). *Digital avatars deep dive series: Navigating the legal and regulatory landscape in 2025*. https://www.mofo.com/pdf/resources/insights/250922-digital-avatars-deep-dive-series-navigating

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, 119*(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

Pagnon, D., Domalain, M., & Reveret, L. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software, 7*(77), 4362. https://doi.org/10.21105/joss.04362

Paubox. (2026). *Senate committee advances NO FAKES Act*. https://www.paubox.com/blog/senate-committee-advances-no-fakes-act

Perkins Coie. (2024). *Generative AI in movies and TV: How the 2023 SAG-AFTRA and WGA contracts address generative AI*. https://perkinscoie.com/insights/blog/generative-ai-movies-and-tv-how-2023-sag-aftra-and-wga-contracts-address-generative

PPC Land. (2026). *C2PA*. https://ppc.land/c2pa/

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

SAG-AFTRA. (2023). *Digital replicas 101: What you need to know about the 2023 TV/Theatrical contracts*. https://sagaftra.org/files/sa_documents/DigitalReplicas.pdf

Saito, S., Schwartz, G., Simon, T., Li, J., & Nam, G. (2024). Relightable Gaussian codec avatars. *CVPR 2024*. https://arxiv.org/abs/2312.03704

Schönberger, J. L., & Frahm, J.-M. (2016). Structure-from-motion revisited. *CVPR 2016*, 4104–4113. https://doi.org/10.1109/CVPR.2016.445

Tao, Z., Liu, Y., Qiu, J., & Li, S. (2025). Impact of virtual avatar appearance realism on perceptual interaction experience: A network meta-analysis. *Frontiers in Psychology, 16*, 1624975. https://doi.org/10.3389/fpsyg.2025.1624975

TechCrunch. (2025). *After Klarna, Zoom's CEO also uses an AI avatar on quarterly call*. https://techcrunch.com/?p=3011302

Technavio. (2026, May). *Digital human avatar market analysis, size, and forecast 2026-2030*. https://www.technavio.com/report/digital-human-avatar-market-industry-analysis

TechSpot. (2025). *The CEOs of Zoom and Klarna have presented earnings calls using AI avatars*. https://www.techspot.com/news/108068-ceos-zoom-klarna-have-presented-earnings-calls-using.html

Thacker, J. (2025, 4 June). *You can now sell MetaHumans, or use them in Unity or Godot*. CG Channel. https://www.cgchannel.com/2025/06/you-can-now-sell-metahumans-or-use-them-in-unity-or-godot/

The AI Musicpreneur. (2026). *Denmark digital replica bill L 49*. https://www.aimusicpreneur.com/ai-music-news/denmark-digital-replica-bill-ai-voice-likeness-consent-2027/

Torys. (2025, April). *The CAI and biometric information*. https://www.torys.com/fr-ca/our-latest-thinking/publications/2025/04/cai-renseignements-biometriques

Traum, D., Aggarwal, P., Artstein, R., Foutz, S., Gerten, J., Katsamanis, A., Leuski, A., Noren, D., & Swartout, W. (2012). Ada and Grace: Direct interaction with museum visitors. In *Intelligent Virtual Agents (IVA 2012)*. Springer. https://doi.org/10.1007/978-3-642-33197-8_25

TrueScreen. (2026). *C2PA standard: History, promises and structural limitations*. https://truescreen.io/articles/c2pa-standard-history-limitations/

U.S. Congress. (2026). *S.4591 — NO FAKES Act of 2026*. https://www.govinfo.gov/content/pkg/BILLS-119s4591is/xhtml/BILLS-119s4591is.html

UneeQ. (2026, 9 February). *UneeQ vs. Soul Machines: The best digital human platforms compared*. https://www.digitalhumans.com/blog/uneeq-vs-soul-machines-best-digital-human-platforms-compared

Usercentrics. (2026). *Consent and Law 25*. https://usercentrics.com/knowledge-hub/consentement-loi-25/

Zhang, D., Liu, Y., Lin, L., Zhu, Y., Chen, K., Qin, M., Li, Y., & Wang, H. (2025). HRAvatar: High-quality and relightable Gaussian head avatar. *CVPR 2025*, 26285–26296. https://doi.org/10.1109/CVPR52734.2025.02448

Zolynski, C. (2026). Deepfakes and transparency obligations. *Columbia Journal of Law & the Arts, 49*(4), 773.