SAELIA - DATA COLLECTION

A. Executive summary

Saelia does not require continuous physiological, biometric, psychological, or neural monitoring to create a useful initial Digital Replica. The scientifically defensible core is a consent-governed representation built from information deliberately supplied, reviewed, and corrected by the user.

The Minimum Viable Dataset should contain account and consent records; autobiographical narratives; user-approved memories and source materials; knowledge and expertise; declared values, preferences, intentions, and boundaries; language and communication samples; optional voice samples where voice interaction is offered; relationship context supplied with appropriate third-party protections; user corrections and approval history; provenance and confidence metadata; and Digital Legacy instructions if Legacy Mode is activated.

Physiological signals such as heart rate, HRV, temperature, SpO2, sleep, and activity may add contextual information but do not reliably reveal a person's emotions, meanings, values, personality, memories, intentions, or identity. They should remain optional and purpose-specific, not mandatory inputs to the Digital Replica.

Automated emotion recognition from voice, face, physiology, or EEG is scientifically uncertain, context-sensitive, culturally variable, and legally sensitive. Saelia should prioritise emotional self-report and user interpretation over automated emotional labels.

Consumer EEG and other neurophysiological data should not be collected for the initial product. Current systems may support controlled research on attention, workload, or motor intent, but they do not provide reliable access to private thoughts, autobiographical memories, subjective meaning, subjective experience, consciousness, or a complete identity.

Faggin's distinction between symbolic information and lived meaning is directly relevant as an ethical limit: a Digital Replica may preserve authorised symbolic expressions of a person's life but should not be described as transferring or recreating consciousness.

Saelia should use a phased strategy: build the prototype from direct, user-controlled narrative and knowledge data; validate interaction quality and correction workflows during the pilot; introduce carefully bounded optional contextual features at commercial launch; and confine neurotechnology to separately governed research until scientific validity, necessity, and proportionality are demonstrated.

B. Data collection

Legend: Direct = provided by the user; Observed = captured by a device or service; Generated = created by the system; Inferred = algorithmically estimated from other data.

Table 1. Data collection

Category	Examples	Purpose	Collection	Contribution/science	Origin/ persons	Risks	Regulation	Legal basis	Retention	User control	Priority
Identity and account data	Name/ pseudonym, age confirmation, email, account ID, authentication factors, verification result.	Create and secure account; verify eligibility; prevent impersonation.	User entry, authentication and proportionate verification. Prefer verification tokens over ID copies.	Identity anchor and anti-impersonation; no psychological fidelity.	Direct/observed; user only.	Identity theft, function creep, biometric compromise.	GDPR Arts. 5, 6 and 9 where biometric identification is used; DPIA may be required; AI Act transparency/risk controls.	Contract/ legitimate interest for basic security; explicit consent or other Art. 9 condition for	Account life plus necessary legal/security periods; raw ID imagery deleted promptly.	Access, correction, export and deletion subject to legal retention.	Essential for initial product

								biometrics.			
Consent, permissions and governance records	Consent receipts, purpose toggles, sharing/model-training/voice/image/legacy choices, withdrawal history.	Prove and enforce lawful choices; support audits and withdrawal.	Granular interface, versioned notices, periodic reconfirmation.	Defines what the Replica may know, say, share and preserve.	Direct/ generated; user and authorised representatives.	Consent fatigue, dark patterns, bundling.	GDPR consent and accountability; AI Act transparency and human oversight.	Explicit consent for special-category and optional high-risk features; contract where appropriate.	As long as processing continues plus limitation periods; preserve withdrawal evidence minimally.	Always reviewable and modifiable; withdrawal must propagate.	Essential for initial product
Biographical profile	Education, employment, languages, places, roles, interests, milestones.	Stable factual context and disambiguation.	Structured onboarding and user-approved imports.	Improves grounding and contextual responses.	Direct; may mention third parties.	Outdated facts, over-disclosure, third-party privacy.	GDPR principles of accuracy, data minimisation, purpose limitation and, where third-party personal data are included, applicable GDPR obligations relating to third-party data.	Contract/ consent depending on feature.	User-controlled while active; legacy only by opt-in.	Review, correct, restrict, export, delete.	Essential for initial product
Autobiographical narratives and memories	Life stories, events, relationships, reflections, photos with captions, diary excerpts.	User-authored memory archive and source-grounded recall.	Guided interviews, voice notes, text, uploads and prompts.	Core of narrative identity and legacy; not objective event reconstruction.	Direct; often concerns third parties.	Reconstructive memory, sensitive third-party content, conflicting accounts.	GDPR lawful basis, special-category content, accuracy and third-party safeguards where applicable, obligations relating to third-party personal data.	Contract/ consent; explicit consent for sensitive content.	User-set retention; legacy only by explicit opt-in; preserve source/version metadata.	Review, annotate, dispute, restrict, export and delete.	Essential for initial product
Knowledge and expertise	Documents, publications, projects, instructions, lessons, professional insights.	Accurate retrieval and transmission of user knowledge.	Uploads, scoped connectors and knowledge interviews.	Supports semantic-memory functions through RAG.	Direct/observed; may contain third-party/employer data.	Copyright, intellectual property rights, confidentiality and trade secrets.	GDPR, copyright and contractual duties; connector DPAs.	Contract/ consent and rights warranties.	While authorised; source-level deletion and revocation.	Review, classify, export and delete.	Essential for initial product
Declared	Value	Align	Structured	Values and	Direct;	Processin	GDPR Art. 9	Explicit	Current	Full review,	Essential,

values, beliefs, preferences and boundaries	statements, moral priorities, voluntary religious/philosophical views, personal preferences, communication style, behavioural boundaries, and preferred or prohibited topics.	the Digital Replica's responses with the user's explicitly expressed values, beliefs, preferences, and personal boundaries while preventing inaccurate or unwanted representation.	reflection, guided scenarios, direct user input, editing, and periodic reconfirmation.	preferences influence decision-making, communication, and behavioural consistency, but should be explicitly provided and periodically updated by the user rather than covertly inferred by AI.	primarily concerns the user.	g of special-category personal data, outdated or changing values, oversimplification of complex beliefs, manipulation, and misrepresentation.	where special-category data such as religious or political beliefs are processed; GDPR principles of data minimisation and accuracy; AI Act safeguards concerning human autonomy, transparency, and protection against manipulation.	consent for special-category data; contract where applicable for non-sensitive preference data.	versions plus limited version history where necessary for transparency, auditability, and user review; periodic reconfirmation to reflect changes over time.	update, correction, restriction, withdrawal of consent, export, and deletion.	but granular and entirely voluntary
Intentions, goals and future instructions	Goals, wishes, unfinished projects, legacy messages and conditions.	Support planning and future/legacy functions without treating predictions or informal statements as the user's legally binding will.	Direct input and periodic confirmation.	Adds temporal orientation and enables authorised future action.	Direct; may affect beneficiaries, designated representatives, relatives, or other third parties.	Outdated or ambiguous instructions, conflicts with formal legal instruments, disputes between beneficiaries, and the risk of treating informal statements as legally binding wishes.	GDPR principles of accuracy, purpose limitation, transparency, and accountability; applicable succession, inheritance, personality-rights, consumer-protection, and electronic-document requirements; AI Act transparency, human-oversight, and accountability requirements where AI interprets or executes user instructions.	Contract or consent for service-related instructions; explicit consent where special-category personal data are involved; formal succession or representation requirements must be assessed under the law of the	Expiry and periodic reconfirmation; superseded versions retained only where necessary for auditability, dispute resolution, or legal compliance and otherwise restricted or deleted.	Review, update, revoke, replace, restrict, export, and delete.	Recommended; Essential for Legacy Mode

								relevant jurisdiction.			
Communication style and linguistic patterns	Vocabulary, syntax, idioms, humour, formality, response length and multilingual use.	Personalise text/dialogue while preserving AI disclosure.	User-authored text, approved conversation samples and corrections.	Style is partly stable but context-dependent; simulation, not identity equivalence.	Direct/ observed/ inferred; may include third-party messages.	Authorship ambiguity, context collapse, copying private communication.	GDPR lawful basis and third-party minimisation; transparency.	Contract/ consent; separate consent for unrelated model training.	Derived parameters linked to deletable sources; periodic review.	Review exemplars, correct rules, disable imitation, export/delete.	Essential for initial product
Voice recordings and acoustic features	Voice samples, pronunciation, rate, prosody, pauses, accent and authorised voice model.	Voice interaction, recognition and optional synthetic voice.	Controlled smartphone recordings and explicit enrolment.	Recognisable communication; does not reveal consciousness, personality, or reliable inner emotional states.	Direct/ observed/ generated; may capture others if uncontrolled.	Voice cloning, fraud, biometric identification, emotion inference and deepfakes.	GDPR Art. 9 where voice data are processed as biometric data for the purpose of uniquely identifying a natural person; AI Act disclosure obligations for AI-generated or synthetic content.	Explicit consent, separate for synthesis and post-life use.	Minimum source samples; protect embeddings; delete/deactivate on withdrawal.	Preview, approve, restrict audiences, export and delete/deactivate.	Recommended; Essential only for voice product
Images, video, appearance and gestures	Portraits, video, facial expression, posture, gestures and avatar preferences.	Avatar and visual communication; memory archive.	User uploads or controlled sessions.	Visual cues support social presence but do not reliably establish emotion or identity.	Direct/ observed/ generated; may contain bystanders.	Deepfakes, biometric templates, bias and non-consensual synthesis.	GDPR and AI Act synthetic-media transparency.	Art. 9 GDPR where images are processed as biometric data for the purpose of uniquely identifying a natural person.	Selected assets only; separate avatar from raw footage.	Preview, approve, label, restrict, export and delete.	Optional / Recommended
Social and relational data	Relationship labels, shared milestones, user-authored relational context and access permissions.	Audience adaptation, shared memories and legacy access.	User-created maps and invite-based shared spaces; avoid bulk message scraping.	Identity and communication are relational and context-sensitive.	Direct; strongly concerns third parties.	Secrets, conflicting memories, surveillance and unauthorised access.	GDPR third-party lawful basis, notices, minimisation and access controls.	Consent/ contract does not automatically justify broad third-party processing.	Only necessary descriptors; remove when relationship/access ends.	User editing plus third-party complaint/correction channels.	Recommended
Interactio	Queries,	Alignme	Configurabl	User	Observed/	Sensitive	GDPR purpose	Contract/	Short default	Review/	Essential

n history and feedback	outputs, ratings, corrections, rejected answers, approved examples and escalations.	nt, error detection, continuity and safety.	e logging and correction tools.	correction is stronger than passive inference.	generated; may include third-party data.	conversations, profiling and indefinite surveillance.	separation; consent for unrelated training; AI governance logging.	legitimate interest for quality/security; consent for unrelated training.	raw-log retention; longer only for approved memories or safety evidence.	delete interactions, opt out of training, export and restrict.	for initial product
Provenance, confidence and model metadata	Source IDs, timestamps, consent status, citations, model version and approval state.	Prevent hallucinations being presented as memories; enable audit.	Automatically attached to every record and output.	Supports epistemic transparency and traceability.	Generated; relates to user and system.	Sensitive metadata and misleading confidence scores.	GDPR accountability/accuracy; AI Act transparency/logging/risk management.	Contract/ legitimate interest for governance.	With underlying record and necessary audit periods; deletion linkage.	User-visible source labels; correction and deletion linkage.	Essential for initial product
Contextual and temporal data	Time, timezone, device state, coarse location, calendar context, current task and audience.	Disambiguate events and adapt interaction.	User input, scoped integrations and optional coarse context.	Behaviour and language are context-dependent.	Observed/ direct/inferred; may concern event participants.	Location/ calendar surveillance and boundary collapse.	GDPR data minimisation, connector permissions and, where applicable, international data transfer safeguards.	Consent/ contract with granular permissions.	Ephemeral by default; store only if converted into memory.	Review permissions, disconnect, delete and export.	Optional
Behavioural data	App use, navigation, frequency, timing, completion and choice patterns.	Usability, personalisation and safety monitoring.	Privacy-preserving product telemetry.	Useful for interface design; a weak and context-dependent proxy for identity or mental state.	Observed/ inferred; mainly user.	Profiling, manipulation and addictive optimisation.	GDPR legitimate-interest balancing or consent for non-essential analytics; AI Act manipulation safeguards.	Legitimate interest after balancing; consent for non-essential analytics.	Aggregate or short-lived events; avoid permanent behavioural dossiers.	Opt out where feasible; delete and restrict.	Recommended for product improvement, not identity
Emotional self-reports	User-labelled feelings, meaning, triggers, confidence and reflective journals.	Self-reflection and contextualised memories.	Voluntary check-ins and event annotations.	First-person report is more defensible than automated emotion labels.	Direct; may mention third parties.	Mental-health sensitivity, crisis disclosures and over-medicalisation.	Potential health/special-category data; clear non-medical positioning.	Explicit consent for sensitive processing.	User-controlled; avoid default permanent retention.	Review, annotate, restrict, export and delete.	Optional / Recommended
Inferred emotions from	Estimated sadness, anger, stress,	Only experimental	Machine-learning inference	Correlates with expressions/	Inferred; user and potentially third parties.	Misclassification, discrimina	AI Act restrictions may apply; workplace/educati	Exceptional consent cannot	Prefer no retention; transient	Contest and disable; no consequenti	Should not be collected/

voice, face or behaviour	engagement or confidence.	adaptive interfaces or research.	from audio/video/ behaviour.	arousal, not private subjective emotion; context and culture decisive.		tion, manipulation and false authority.	on emotion recognition is particularly restricted; GDPR profiling.	cure prohibited or disproportionate use; legal/scientific review required.	research only; never permanent identity labels.	al action from label alone.	inferred for MVP
Personality and psychological assessments	Validated trait questionnaires, narrative-identity measures and cognitive-style tasks.	Optional self-understanding and personalisation.	Validated self-report instruments with explanations.	Measures constructs with error; not a complete person model.	Direct/ generated/ inferred; user only.	Labelling, self-fulfilling profiles and mental-health inference.	GDPR profiling and possibly health/special-category data; test licensing.	Explicit consent.	Time-limited results with instrument/date /uncertainty.	Review/ contextualise, restrict, export and delete.	Optional; validated instruments only
Cognitive-task data	Reaction time, working memory, attention, language and decision tasks.	Accessibility adaptation or research.	Short validated tasks under controlled conditions.	Measures task performance, not general intelligence, identity or consciousness.	Observed/ generated/ inferred; user only.	Fatigue, device variability, unfair scoring and medical interpretation.	GDPR profiling; high-impact uses may trigger AI Act obligations.	Consent and separate governance.	Short retention; summaries only if requested; record task conditions.	Review limitations, delete, restrict and export.	Optional / Experimental
Health and psychophysical data	Diagnoses, medication, symptoms, heart rate, blood pressure, temperature, SpO2, stress and sleep indicators.	Only clearly defined well-being or research functions; not core Replica.	Direct health input or connected devices.	May contextualise physical state but not identity, emotion or meaning.	Direct/ observed/ inferred; user and providers.	Special-category sensitivity, medical-device boundary and discrimination.	GDPR Art. 9, DPIA, medical-device/consumer-health assessment; AI Act depends on intended purpose.	Explicit consent or another Art. 9 condition.	Purpose-specific and generally shorter than identity archive; minimise raw streams.	Access, correction, restriction, portability and deletion where possible.	Optional; not required for Digital Replica
Wearable-device data	RHR, HRV, sleep, activity, skin temperature, SpO2 and device-quality metadata.	Optional longitudinal context and user-facing trends.	Wearable APIs with device-specific validation and quality flags.	Some devices can measure nocturnal RHR/HRV reasonably well under specific conditions; there is no validation of emotion	Observed/ generated; user only.	Continuous surveillance, false precision and vendor transfers.	GDPR special-category processing, vendor DPAs, transfer safeguards and DPIA.	Explicit consent.	Prefer daily summaries over raw streams; user-selected periods; delete on disconnect unless archived.	Review source/device, disconnect, export and delete.	Optional

				or personality inference.							
Biometric identity data	Face template, voiceprint, liveness score or fingerprint.	Authentication and prevention of unauthorised replicas.	Specialised provider; local/device processing where possible.	Useful for authentication only; no improvement to personality/ memory fidelity.	Observed/ inferred; user only.	Irreversible compromise, surveillance and function creep.	GDPR Art. 9, strict necessity, DPIA and security segregation.	Explicit consent or other valid Art. 9 condition.	Delete raw captures promptly; retain protected result/template only as necessary.	Challenge errors; delete/re-enrol where possible.	Recommended only where proportionate
EEG and neurophysiological data	Scalp EEG, ERP, motor-intention, workload/attention markers; future invasive BCI data.	Controlled research or accessibility interfaces.	Research-grade supervised protocols; consumer EEG only low-stakes experiments.	Measures aggregate electrical activity and limited task correlates; cannot directly decode complete identity, private meaning, autobiographical memory, or consciousness.	Observed/ inferred; user only.	Mental privacy, incidental findings, bias, security and exaggerated claims.	Explicit consent, ethics review, DPIA, research governance and possible medical-device rules; neurorights considerations.	Explicit research consent.	Research-specific coded retention, predefined deletion and no uncontrolled secondary use.	Access to results/limitations, withdrawal, export/deletion subject to research law.	Future development / Experimental
Post-life and Digital Legacy instructions	Activation status, delegates, access tiers, permitted functions, prohibited outputs, duration, revocation and memorial messages.	Ensure post-life use follows prior autonomous choices.	Separate addendum, explicit opt-in, periodic reconfirmation and verified trigger.	Governance data essential to ethical legacy use, not psychological fidelity.	Direct/ generated; affects relatives/beneficiaries.	Conflicts with heirs, grief vulnerability, stale consent, impersonation and exploitation.	GDPR where applicable, succession/personality/consumer law and AI disclosure; jurisdiction-specific review.	Explicit, specific and revocable consent.	While Legacy Mode is active and defensible; periodic reconfirmation; suspend if uncertain.	Review, amend, revoke and export; delegates get scoped rights.	Essential only for Legacy Mode
Third-party communications and data	Emails, chats, photos, recordings, contacts and stories involving non-users.	Potential relational context and memories.	Prefer user-authored summaries; narrowly selected imports only.	May improve context but creates disproportionate privacy risk.	Direct/observed; concerns third parties.	Lack of consent or another lawful basis, confidentiality risks, children's data,	GDPR principles of lawfulness, fairness, transparency, data minimisation, purpose limitation, accuracy, and data-subject rights; user	A specific lawful basis under GDPR Art. 6 and, where special-category	Default to no retention or strictly bounded retention; redact and minimise; retain only where	Provide appropriate third-party complaint, access, correction, objection, restriction, and	Should not be bulk-collected

						sensitive disclosures, and post-life disputes.	ownership or access to an account does not by itself establish a lawful basis for processing third-party data.	data are involved, an applicable condition under Art. 9; consent from the Saelia user alone is not automatically sufficient for broad processing of third-party data.	necessary for a defined purpose and lawful basis.	deletion pathways where applicable.	
Inferred sensitive traits or mental states	Predicted religion, politics, sexuality, ethnicity, criminality, health or mental condition.	No necessary core purpose.	Latent profiling from language, networks, face, voice, physiology or behaviour.	Weak proxies can produce confident but false sensitive profiles.	Inferred; user and third parties.	Severe discrimination, manipulation and chilling effects.	GDPR special-category/profiling restrictions; AI Act prohibitions/high-risk rules may apply.	Generally no defensible basis for broad generation.	Do not retain; should not be generated except approved research.	Contest/ delete if generated accidentally; block persistence.	Should not be collected

Source: study based on analyzed sources.

C. Minimum Viable Dataset

1. Account and consent layer: Minimal account identifiers, authentication, eligibility confirmation, consent receipts, data-purpose permissions, sharing settings and withdrawal history.

2. User-authored biographical layer: Core biographical facts and relationship labels selected by the user.

3. Autobiographical memory layer: Narratives, memories, timelines, photographs or recordings with source, date, confidence and privacy labels.

4. Knowledge layer: User-authorised documents, expertise and instructional content organised in a source-grounded knowledge base.

5. Values and boundaries layer: Directly declared values, preferences, communication boundaries and prohibited uses; no automated inference of sensitive personal characteristics.

6. Communication layer: Textual style samples and user corrections; voice only where voice functionality is part of the product.

7. Alignment and correction layer: Approved outputs, rejected outputs, user edits, model version and provenance.

8. Legacy governance layer: Only for users opting into Legacy Mode: designated persons, permitted functions, duration and revocation rules.

The MVP should not require physiological monitoring, emotion recognition, psychological testing, wearable integration, biometric identity beyond proportionate anti-impersonation measures, or EEG.

D. Extended Dataset

Optional voice synthesis with separate enrolment, approval and audience controls.

Optional avatar appearance and gesture samples with synthetic-media disclosure.

User-authored relationship maps and shared-memory spaces based on invitation and scoped permissions.

Voluntary emotional journals and reflective check-ins, clearly separated from medical or diagnostic functionality.

Validated optional psychometric instruments for self-reflection, with uncertainty and time stamps.

Contextual integrations such as calendar, coarse location or activity state, processed ephemerally where possible.

Wearable summaries such as nocturnal resting heart rate, HRV, sleep and activity trends, offered as a separate opt-in contextual service rather than identity-defining evidence.

E. High-risk and experimental data

Automated emotion inference: High scientific uncertainty and contextual bias; use only in separately validated research or narrowly bounded functions.

Health and psychophysical streams: Require a clear independent purpose, Article 9 basis, DPIA and medical-device boundary analysis.

Biometric templates: Use only for proportionate authentication, segregated from the generative system.

Personality, mental-state and sensitive-trait inference: Avoid covert inference; direct self-description is more accurate and defensible.

EEG and neural data: Research-only until scientific validity, necessity, proportionality and secure governance are demonstrated.

Post-life interactive simulation: Requires exceptional consent, grief-sensitive design, disclosure, capability limits and human escalation.

F. Data-use map

Table 2. Data-use map

Functionality	Data	Use operations	Part of Replica?
Account security / anti-impersonation	Account data, verification result, authentication logs	Verification, security, access control	No
Memory archive and recall	Narratives, media, timelines, provenance	Storage, retrieval, source-grounded generation	Yes, as authorised representation
Knowledge transmission	Documents, expertise, projects and instructions	Retrieval, summarisation and question answering	Yes
Communication personalisation	Language samples, style rules, optional voice/visual model	Personalisation, generation and communication	Yes
Values alignment	Declared values, preferences, prohibitions and corrections	Constraint application, response selection and review	Yes
Relationship-aware interaction	Relationship context and access permissions	Audience adaptation, sharing and legacy access	Yes, with safeguards
Self-reflection	Emotional self-reports and optional wearable summaries	Visualisation and longitudinal reflection, not diagnosis	Optional and separable
Quality and safety	Feedback, rejected outputs, logs and model versions	Evaluation, monitoring and incident reconstruction	Indirectly
Digital Legacy	Legacy instructions, delegates, permitted archives and outputs	Activation control, access and memorial/interactive mode	Yes, explicit opt-in
Research / future neurotechnology	EEG, cognitive tasks and experimental sensors	Research inference under separate protocol	No automatic incorporation

Source: study based on analyzed sources.

G. Data that should not be collected

Continuous raw audio or ambient recording by default.

Bulk imports of private communications involving third parties without selection, redaction and a defensible legal basis.

Physiological data collected solely to claim that Saelia can determine authentic emotions, personality, values or consciousness.

Automated sensitive-trait profiles concerning religion, politics, sexuality, ethnicity, criminality, health or mental condition derived from unrelated behaviour.

Emotion labels generated from face, voice or physiology and stored as permanent identity facts.

EEG or neural data marketed as reading private thoughts, memories, meaning or consciousness.

Unnecessary copies of identity documents and permanent raw biometric captures after verification.

Data collected for undefined future purposes, speculative future AI capabilities, unrelated commercial profiling, or third-party advertising.

Children's or vulnerable persons' intimate data for experimental personalisation without heightened necessity and safeguards.

Any data source that the user cannot inspect, challenge, restrict and disconnect from the Digital Replica.

H. Scientific and technical limitations

Consciousness cannot currently be copied or verified as present in a Digital Replica. The uploaded sources disagree about ontology and mechanism, but none offers a validated transfer method.

Stored memories are representations, not the original memory experience. Autobiographical memory is reconstructive and source content reflects a perspective at a particular time.

Language models perform pattern-based generation, not demonstrated human understanding. They may generate coherent identity-consistent language while hallucinating or combining sources.

Emotion cannot be read directly from a single signal. Heart rate, prosody, facial action and behaviour have multiple causes and require context and self-report.

Voice similarity or behavioural similarity does not establish identity continuity. Synthetic voice remains generated output.

Personality measures are probabilistic and incomplete. Traits vary by context and instrument.

Consumer wearables vary by device, algorithm, and measurement conditions. RHR/HRV validation does not validate emotional or identity inference.

EEG has limited spatial resolution and high artefact sensitivity. Consumer systems cannot reconstruct a private mental life or identity.

Social presence may encourage anthropomorphism and attachment. Users may experience AI as a social partner despite knowing it is not a person.

Legacy interactions may change memories of the deceased. Generated content can diverge from sources unless provenance and audits are enforced.

I. Ethical, legal, and governance recommendations

Purpose registry and necessity test: Map every field to a function, lawful basis, retention rule, security class and user control.

Data quality management: Implement procedures to ensure that personal data remain accurate, up to date, complete where necessary, appropriately sourced, and versioned. Users should be encouraged to periodically review and confirm information used to build their Digital Replica.

Layered consent: Separate core processing from voice, biometrics, health, research, model improvement, public sharing and Legacy Mode.

Privacy by design: Separate identity/consent vaults from content and generative systems; minimise raw streams and use deletion lineage.

Third-party data governance: Prefer user-authored summaries over mass ingestion; add notices, redaction and complaint pathways.

Human oversight: Users must pause, inspect, correct and deactivate the Replica; escalate disputed or safety-sensitive cases.

Provenance and uncertainty: Distinguish quotation, verified fact, approved interpretation and model speculation.

AI transparency: Disclose AI interaction and label synthetic voice, image and video.

Risk management and testing: Red-team impersonation, grief manipulation, hallucination, bias, prompt injection and model extraction.

Vulnerable persons: Use stricter defaults, age/capacity-appropriate consent and avoidance of dependency design.

Security and incident response: Use MFA, encryption, least privilege, logging, vendor due diligence, penetration tests and breach procedures.

DPIA and records: Complete DPIA before biometric, health, psychophysical, extensive behavioural or neural processing.

AI Act classification: Document intended purpose; avoid prohibited manipulation and high-impact decision uses.

Research governance: Use ethics review, validated instruments, preregistered objectives and separation of research/commercial data.

Legacy governance: Use a separate addendum with opt-in, reconfirmation, verified triggers, scoped delegates and suspension mechanisms.

J. Final recommendation: phased data strategy

Prototype: Use account/consent data, biography, selected memories, knowledge, values/boundaries, text style and feedback. No continuous sensors, emotion recognition, health inference or EEG.

Pilot: Add optional voice, relationship-aware access, provenance dashboards and living-user Legacy Mode tests. Evaluate fidelity, hallucination, anthropomorphism and emotional safety. Complete DPIA and security testing.

Commercial launch: Introduce only validated optional features, secure connectors and user-facing controls. Maintain separate consents, retention dashboards, incident response and legal review. Launch Legacy Mode only with a dedicated addendum.

Future research and neurotechnology: Create a separate research programme for wearables, cognitive tasks, EEG or BCIs with a specific hypothesis, ethics approval, validation and proof that less intrusive data cannot achieve the purpose.

References

Berger, S., & Rossi, F. AI and Neurotechnology: Learning from AI Ethics to Address an Expanded Ethics Landscape.

Berger, S., & Rossi, F. The Future of AI Ethics and the Role of Neurotechnology.

Chalmers, D. J. Consciousness and its Place in Nature.

Dial, M. B., et al. (2025). Validation of nocturnal resting heart rate and heart rate variability in consumer wearables. Physiological Reports, 13, e70 527.

Duch, W. (2019). Artificial Intelligence and Neurocognitive Technologies.

Elmaresa, M. V., & Irwansyah (2025). Human and Artificial Intelligence Interaction from the Perspective of Social Construction of Technology.

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act): https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Faggin, F. (2024). Irreducible: Consciousness, Life, Computers, and Human Nature. Essentia Books.

Ferreira, J. J., & Monteiro, M. (2021). The human-AI relationship in decision-making.

Harley, T. A. (2021). The Science of Consciousness. Cambridge University Press.

Harris, L. T. (2024). The Neuroscience of Human and Artificial Intelligence Presence. Annual Review of Psychology, 75, 433-466.

McCullum, T., & Contreras Mancera, L. C. (2025). Neurotechnology Combined with Artificial Intelligence and Neurorights.

Roberts, R. (2019). Neurotechnologies: Connecting Human Brains to Computers and Related Ethical Challenges.

Saelia Constitution and Digital Identity Charter, working draft, 19 June 2026.

Searle, J. R. Consciousness.

Zeman, A. (2001). Consciousness. Brain, 124, 1263-1289.