
Afterlife Delegation Protocol: Speculative Design of Self-Sovereign Agents that Outlive Their Principals

Botao Amber Hu
University of Oxford
Oxford, UK
botao.hu@cs.ox.ac.uk

Iris Long
Goldsmiths, University of London
London, UK
irisxink@gmail.com

Abstract

Afterlife Delegation Protocol is a speculative design project that asks what death becomes when a will can act eternally. We design a speculative protocol through which a living person signs an *agentic will*: upon a verified death, a self-sovereign AI agent spawns on blockchain — an immutable, resistant, decentralized, infrastructural substrate that could last forever — endowed with the funds and memories its principal attached to it, and persists indefinitely to execute the will, overridable by no custodian. Rather than argue about this future, we stage it: following the science fiction science method, we translate the speculation into an *experiential futures* intervention — a working web platform where real people design their own afterlife agents through an iterative, interactive, AI-automated interview, re-login to revise, and rehearse their will in a sandbox. Their drafted wills become qualitative data on a question rarely askable directly — *what should outlive you?* — and on how afterlife cosmologies across different cultural beliefs — Buddhist, Christian, Hindu, Muslim, and atheist — begin to drift under the pressure of AI proliferation. We design the protocol over a composition of existing Ethereum agent standards — drafted as ERC-10001, in the normative format of an Ethereum Improvement Proposal — and describe its three-stage lifecycle (designing the afterlife, proof of death, agent enactment), the research method, and preliminary observations from an ongoing collection. The work surfaces a poetic delegation moment between human mortality and machine eternity, mediated by long-lived infrastructures that span generations.

1 Introduction

People die twice. First, when they die. Then, when they are forgotten. — Rokusuke Ei, 1933–2016 [24, p. 62]

A will is the oldest instrument for exercising agency in a future where one cannot act. In every existing form of will execution — a testament interpreted by lawyers and relatives [29], a trust administered by mortal trustees [63], an artist’s estate kept by a foundation [50] — the will is a static declaration that runs on institutional permission and ceases to act at the boundary of institutional attention. This project speculates a stronger instrument. In the future we design, a person signs an *agentic will* while alive; after their death is verifiably witnessed, the will spawns a dedicated, self-sovereign AI agent [39] — an executor that is not a person but a persistent autonomous process. The agent can be endowed with funds (a treasury from which it pays for its own compute, storage, and commissioned work) and with memories (archives, writings, likenesses, and the recorded interviews through which the will was elicited). It then acts in the world indefinitely: maintaining, responding, transacting, refusing, on behalf of someone who can never again be consulted.

As a mortal being, what do you believe should outlive you? Your social identity? Your assets? Your memories? Your trace of living? Your reputation? Your unfinished wishes? What if a self-sovereign AI agent could carry out your afterlife on an almost ethereal infrastructure?

This is a speculative design project [23] — a provotype [52, 9], not a product prototype. But it departs from traditional scenario-writing: it is a high-fidelity simulation. Technologies of this kind sit inside the Collingridge dilemma [19] — their impacts cannot be known until they are developed and deployed, yet by then they are entrenched and hard to steer. To gain knowledge of impact before entrenchment, we deploy Rahwan et al.’s *science fiction science* method [60]: rather than predicting whether such agents will exist, we build a functioning simulacrum of the future and study how people behave inside it, as the Moral Machine did for autonomous vehicles [6]. We then translate the simulation into an *experiential futures* intervention [16, 17]: a real, public web platform where visitors do not read about afterlife agents but encounter one — they sit for an interactive, iterative long interview with an automated AI interviewer [3], draft their own posthumous delegation, revise it, and watch a sandboxed rehearsal of their agent acting after their hypothetical death.

What the provotype measures is cultural drift. Mourning, reincarnation, judgment, inheritance, ancestor veneration, and the idea of a second death are cultural technologies for bringing the agency of the dead to a close [34, 5]. AI proliferation puts that closure under pressure, and different afterlife cosmologies will bend differently: each tradition holds a distinct cosmological model of what follows death [11, 56, 66, 53]. So what would a Buddhist, a Christian, a Hindu, a Muslim, and an atheist each believe should outlive their original life on-chain, eternally, with AI agents? The project collects these answers as they begin to form, before the technology hardens them.

The paper contributes: (1) the speculative design of the protocol itself; (2) its staging as an experiential future for real participants; and (3) preliminary observations and the cross-cultural probes of an ongoing collection.

2 Background

2.1 AI, afterlife, and immortality

Terror management theory holds that people buffer the knowledge of their own death by embedding themselves in structures that outlast them: lineage, religion, nation, institution, work [32]. Every such structure carries a cosmology of time — an implicit account of how long things last, what deserves to persist, and who is owed continuity [55]. Continuing bonds research shows that the living already sustain relationships with their dead without any technology at all [42]; the digital afterlife industry now supplies technology in quantity [57]. A growing literature maps it: responsible design for griefbots and deadbots [35], *generative ghosts* as agentic AI representations of the dead [54], TeleAbsence as a poetics of mediated absence [40], design research into AI carriers for afterlife selves [44], and studies of how people perceive an “AI afterlife” as digital legacy [45]. In Morris and Brubaker’s design space, the artifact we speculate is a first-party, pre-mortem, evolving ghost — but it adds a dimension their framework does not cover: *infrastructural persistence* (Section 6). Existing ghosts are products, hosted at a company’s pleasure; they die when a server bill lapses. The question of what happens when the representation of the dead cannot be turned off has, until recently, been unaskable.

2.2 AI, blockchain, and immutability

Decentralized infrastructure makes it askable. A public blockchain is a machine for immutability: state that no single party can roll back, contracts without custodians, funds that disburse by rule rather than by permission. This is the Long Now [12] arriving as an executable substrate rather than a clock, with all the maintenance politics that infrastructural longevity actually entails [61, 13]. Two further ingredients make the substrate habitable by agents. Decentralized physical infrastructure networks (DePIN) [48, 8] now provide *permissionless compute*: computation that can be bought by anyone — including a contract — without an account, a company, or a jurisdiction consenting. And confidential computation in trusted execution environments (TEEs) [46, 20], as proposed for TEE-EVM co-processing in ERC-733 [27], lets that computation run privately while proving on-chain *which* code ran. Neither is hypothetical: crypto-native services such as Venice.ai already sell

open-weight foundation-model inference for cryptocurrency, permissionlessly and without accounts, and TEE cloud networks such as Phala make such compute attestable on-chain¹ — an agent can, today, buy its own inference. Deployed together, these already support LLM-based agents that operate “in the wild” — holding their own wallets and social accounts, paying for their own inference, with no human able to reach in and stop them: the Spore.fun experiment documented sovereign agents evolving on TEE-secured blockchains as a live, open-environment system rather than a thought experiment [37]. Such an agent inherits its non-overrideability from the infrastructure itself rather than from any custodian’s forbearance [38]. The combination is what makes our speculation load-bearing: an agent that is at once *autonomous* (AI), *unstoppable* (permissionless compute) [36], and *unforgetting* (immutability) is the first vessel of symbolic immortality [47] that can also act. The cultural technologies of death [34, 5] have never had to metabolize that before.

3 Speculative design: the Afterlife Delegation Protocol

We outline the protocol as ERC-10001, published at <https://erc10001.org> in the exact normative format of an Ethereum [15] Improvement Proposal. It is built on four existing Ethereum agent standards, each supplying one organ of the afterlife agent: ERC-733 [27] the *sealed skin* — an attested confidential boundary (TEE-EVM co-processing) inside which the agent’s *body* (the compute harness) and *mind* (the foundation model) run inviolate, invadable and alterable by no outsider, not even the machine’s operator; ERC-8004 [25] the *identity* — the ID through which the sealed agent faces outward, a persistent, discoverable registration with reputation and validation registries that lets it exchange money and information with the external world; ERC-8350 [28] the *memory* — a memory state registry anchoring every evolution of the agent’s memory in a strictly linear hash chain, so that memories of the dead cannot be silently deleted, rolled back, or forked into contradictory ghosts, and so that death itself appears in the ledger as a change of author within one unbroken chain; and ERC-8183 [26] the *metabolism* — job-based escrow through which the agent buys storage, commissions work, and hires other agents, funds leaving the treasury only against attested delivery. Above these, the protocol adds what none of them contemplate: a principal who dies. It enforces a constitutional separation between the signed *core will*, immutable at death, the agent’s accumulating operational memory, and the model substrate interpreting both — the agent may become more capable; it may not become someone else. The lifecycle runs in three stages; the full specification is given in Appendix A.

Stage 1: Designing the Afterlife. The first stage is alive today as a working platform at <https://afterlife-protocol.org>. A visitor does not fill in a form; they sit for a long, interactive, AI-automated interview with the *Afterlife Interviewer* (Appendix B), which converts diffuse intuitions about death into specific, contestable commitments: what the agent may say and to whom, in which pronoun (*as* the deceased or only *about* them), with what funds, under which prohibitions, for how long, and what may terminate it. Two properties make the stage a design environment rather than a survey. The will is *revisable*: participants re-login across weeks and change their minds, every revision a signed, versioned commit — a will here is not a final declaration but a living negotiation with one’s own future absence. And the will is *sandbox-testable*: a simulation lets participants converse with their own agent after their hypothetical death, watch it handle scripted situations (a grieving relative, an impersonation attempt, an unforeseen request), and return to revise the clauses that failed them. The current deployment is a research prototype: wills are drafted and simulated, never activated, and no agent of any actual deceased person is built.

Stage 2: Proof of Death. Everything downstream hangs on one predicate the chain can never compute: that the principal has died. A *death oracle* must witness the death — and death is the unique event that cannot be self-attested; liveness is self-evidencing, death is only ever hearsay. The protocol therefore treats attestation as its most sensitive interface and designs it as a weighted, contestable proceeding rather than a switch: any party may file a bonded death claim; evidence accumulates weight from will-designated witnesses (institutional registries, human witness quorums, the lengthening silence of a dead-man’s switch); a challenge window scales inversely with the weight of evidence; and one invariant is supreme — any authenticated sign of life from the principal vetoes everything. The protocol may wrongly delay a ghost; it must never overrule a living voice. Crucially, the definition of death is itself testamentary content: cardiac, brain, legal, social, and “key” death are

¹<https://venice.ai>; <https://phala.network>.

different events, ranked differently by different traditions, so the will configures which deaths count, witnessed by whom, on what horizon.

Stage 3: Agent Enactment. Upon a finalized attestation, the will becomes immutable, the treasury passes to the agent, and the agent begins acting on-chain in the real world through a persistent loop: read the will, interpret its obligations, observe the world, plan, act, record memory, sustain itself. It maintains the will and acts on the principal’s unfinished goals — tending archives and social presences, responding to designated people, protecting the deceased’s likeness against misuse, funding memorial rituals, continuing incomplete projects, and hiring other agents for what it cannot do alone. It may upgrade to newer models while the original will binds it as a constitution, and it may be terminated only through the governance mechanism the principal consented to in advance. Because its endowment spends at most its yield, decline is graceful rather than sudden, and anyone — a grandchild, a stranger, an institution — may add funds at any time. The living can feed the ancestor.

4 Research method: staging the future to collect it

We stage Stage 1 as a science fiction science instrument [60] to engage online participants, collecting their interactions with afterlife agents. The framing matters: the experience presents itself as delegating one’s will to an afterlife-agent *service*, and the interview behaves as consulting and sales interviews do, eliciting requirements from a client who does not yet know what they want — most people cannot fully name their will on first reflection, so articulation is necessarily iterative, explorative, and qualitative. Asked directly, *what should outlive you?* yields platitudes; asked as *what may your agent do, say, spend, and refuse in year forty?* — and asked again, across sessions, against the sandbox’s rehearsals — it yields commitments, each clause a datum about what its author believes deserves continuity, who is owed it, and how long “forever” is felt to be, complemented by qualitative surveys on what people believe should persist beyond a life. The interviews are conducted by the Afterlife Interviewer, an adaptive semi-structured interviewing agent built on SparkMe [4], which plans questions, takes notes, and pursues emergent subtopics — scaling qualitative elicitation to an online population without a human interviewer (Appendix B).

The ongoing collection focuses on cross-cultural difference across beliefs and faiths — Buddhist, Christian, Hindu, Muslim, and atheist participants are recruited purposively — because the project’s object is precisely the cross-cultural drift: how differently afterlife cosmologies [53] absorb the offer of a machine executor. Because the material concerns grief, private memory, and named third parties, transcripts are never published; analysis draws on Clio [65], Anthropic’s privacy-preserving pipeline for insights from real-world AI conversations, to surface population-level themes while raw responses remain sealed (Appendix B). Participation is voluntary and consent-based; participants may withdraw and delete their wills.

5 Preliminary results

The collection is early and ongoing; we report 4 early qualitative observations.

Reputation maintenance: the curated self, defended forever. The protocol invites estates and legacies; what participants first ask for is often something small and exact. One participant, after a long interview, settled on a single standing instruction: the agent should maintain their social media photos forever — including automatically correcting and “photoshopping” new and old images so the feed never falls out of date: *“keep my photos the way I keep them now. If a picture of me is bad, fix it. I don’t want to look worse dead than I did alive.”* The clause is easy to smile at and hard to dismiss: it is continuing impression management, a curated self defended after death with the care it was defended in life — suggesting that what people first reach for, when offered infrastructural immortality, is not wealth or wisdom but the integrity of a self-image. Against the literature’s grand ghosts, the vernacular afterlife may be maintenance work.

Wiring money to the ancestor (Chinese ancestor veneration). Among participants from Chinese ancestor-veneration backgrounds, persistence is welcomed as the extension of a practice already millennia old [2, 67]. One participant asked the agent to keep the family’s calendar of rites: *“at Qingming it should greet my grandchildren and bless them for the year. We have always fed our*

ancestors — but joss paper burns, and the money never arrives. If my grandchildren wire crypto to it, that isn't a deposit — that's an offering, and for the first time the ancestor can actually spend it: on its compute, its maintenance, next year's blessing." The ritual economy becomes literal: offerings once burned now arrive as funds the ancestor-agent spends to stay alive — purchasing its own compute and upkeep — so its immortality depends on how faithfully the living keep paying. The best-fed ancestors will live longest.

Dissolution versus the merit engine (Buddhist). For Buddhist traditions built on impermanence and non-attachment, an afterlife agent that cannot stop is a category error: clinging to continuation is precisely what the doctrine trains one to release. At the same time, in Buddhist cosmology merit (karma) accumulates across the whole cycle of rebirths, and making merit on behalf of the dead is already established practice [30, 22]; so wills appear instructing the agent to keep generating good karma — perpetual giving, sutra recitation, acts of care — through every incarnation of its author: *"let it keep giving in my name, life after life — whoever I am born as next should arrive into the good it has already done."* Each new life thus inherits the accumulated merit of an on-chain agency it cannot remember signing. The clash is the protocol's default itself: infrastructure assumes forever; doctrine demands letting go. The one reconciliation participants found is to turn persistence into a merit engine — and then letting go is exactly what the machine defers, forever.

Speaking about, never as (Christian). In Christian cosmology the soul has gone elsewhere, so an agent that speaks *as* the dead comes close to mediumship; participants therefore draft strict pronoun clauses — the agent may speak *about* them, never *as* them: a memorial, not a séance [54, 35]. Yet the same participants comfortably assign the agent perpetual charitable duties, which fit the familiar category of stewardship. One participant pushed this logic to its theological edge. Carrying a wrong they regretted, they asked whether the agent could keep atoning for it: *"I've done things I'm not proud of. If it keeps paying people back and giving after I'm gone, does that still count as me? Can it change what's waiting for me?"* Doctrine says no: judgment is fixed at death, and no work can be added afterward. But purgatory and prayers for the dead leave a door ajar, and through it the participant volunteers the agent as a perpetual penitent — importing into a cosmology of final verdicts a karmic logic in which good deeds accumulate and slowly change the outcome. The drift lies in the question itself: no longer whether the dead can act — the infrastructure settles that — but whether their acts can still be credited to them.

6 Discussion

Cross-cultural belief drift under the pressure of AI. Each tradition maintains a distinct cosmological model of what follows death — rebirth, resurrection and judgment, the soul's onward journey, ancestral presence, extinction [11, 56, 66, 53]. An agency that lasts forever does not merely extend these cosmologies; it conflicts with them. Reincarnation presupposes that the self vacates this world; judgment presupposes a life whose account has closed; an agent that goes on holding, spending, and speaking keeps the account open, forcing questions the traditions never had to price: why rebirth, if agency persists here? Why judgment, if the ledger never closes? Who answers for acts a will commits after its author can no longer consent, repent, or be held to account? Like the Moral Machine [6], the platform is a distributed cosmological probe — but its dilemmas are unstandardized, and what it collects is not convergence toward a shared answer but divergence in what the question is taken to be. What social physics becomes in the age of AI — the quantitative study of collectives at once human and artificial [33, 59], whose culture machines increasingly generate, transmit, and select [14] — will need instruments for such hybrid negotiations; this work is an early one.

Afterlife agency as a permissionless design space. Because the executor is a general-purpose autonomous agent rather than a document, the space of what a will can delegate widens from asset disposition to nearly anything an agent can be instructed to do: an identity to be remembered by; a griefbot of living memory that comforts the bereaved [35, 42]; a knowledge system that goes on contributing; a fortune that keeps compounding; a project scheduled in centuries; a likeness defended against misuse; a dynamic, evolving ghost that carries a persona and continues to act in society [54]. Morris and Brubaker [54] give an early taxonomy of such ghosts along seven dimensions — provenance (first- vs. third-party), deployment timeline (pre- vs. post-mortem), anthropomorphism paradigm (speaking *as* vs. *about* the dead), multiplicity, cutoff date (static vs. evolving), embodiment,

and representee type — but the design space is vaster still. And it is open in a second, self-sovereign sense: anyone may design and deploy an afterlife agent without any platform’s, church’s, or state’s permission, and once deployed on the substrate, no custodian can stop it [36, 38] — whether this future is found desirable or abhorrent, it is not gated.

Self-sovereignty vs. entrusting others to execute your agency. Every prior instrument of posthumous agency is an act of trust in the living: a testament trusts lawyers and relatives to probate it [29]; a trust trusts mortal trustees, capped by the rule against perpetuities because the common law feared the dead governing indefinitely [63]; moral rights trust heirs who thin out within a generation [51, 50]; the artist’s testament trusts an executor who may refuse — Kafka trusted Max Brod to burn the manuscripts, and Brod’s refusal is why we have *The Trial* [7]. The agentic will abolishes that dependence: where legal tech automates the drafting and probate of wills [21], this design automates the executor itself — once activated, no relative, court, or platform can rewrite or intervene. For the first time, agency can be genuinely protected after death: entrusted to no one, diffused into infrastructure. Whether that is a good thing is exactly what the design leaves open. Trust was the safety valve: the trustee’s judgment, the court’s equity, and Brod’s disobedience adapted dead instructions to living circumstances — a faithful agent would have lit the fire, and an unrevisable will that begins to harm the living can no longer be reformed. And what each testator experiences as protection, society accumulates as sovereignty of the dead, inheriting undiluted the dead-hand problem the law spent centuries containing. Do we want agency protected even after death — and who is “we” once the dead are still acting? That question is what the termination-governance interface is for.

Danger of proof of death. The deepest technical danger is that *proof of death is at least as hard as proof of humanity*. A decade of proof-of-personhood work — from pseudonym-party schemes [10] through live registries such as Proof of Humanity [43] to personhood credentials in the age of generative AI [1] — asks how a digital system can verify that a unique living human stands behind a key; our protocol needs the negation — that no living human stands behind it any longer — and the negation is strictly harder, because liveness can demonstrate itself with one signature while death can only ever be testified to by others. Prior designs for death on-chain confirm the difficulty by avoiding it: blockchain will systems fall back on designated human certifiers [18], dead-man’s-switch protocols such as Sarcophagus reduce death to the silence of missed check-ins [62], and analyses of crypto inheritance leave the death event itself as an open question [58]. Every attestation path we compose (registries, witnesses, the lengthening silence of a dead-man’s switch) inherits the same trust assumptions, sybil risks, and biometric decay that personhood systems face, plus failure modes of its own: a treasury payable on death is structurally a bounty on a life, coercion of check-ins can be flagged but not prevented, and the vanished-but-alive are eventually mis-declared by construction.

The transition moment between human mortality and machine eternity. What the work finally aims to surface is a poetic delegation moment: a mortal human, dictating terms to an eternal machine, mediated by the quiet, embedded infrastructures through which civilizations may persist across generations [64, 61]. Blockchain immutability is a maintained accomplishment [31, 41, 13] of validators, client developers, and node operators, so an agent’s “forever” is willed anew by each generation of hands; yet because that maintenance is decentralized, the substrate is unstoppable in practice while contingent in principle [36] — no one in particular keeps it alive, and no one in particular can turn it off. Social order quietly presupposes finitude: institutions from inheritance to forgiveness are calibrated to actors who die, and machines are the first social participants that need not [49]. The agentic will places that asymmetry inside a single, signed relationship: whether the agent keeps faith with its author is the question every heir has faced — but for the first time, the executor does not die either.

7 Conclusion

Afterlife Delegation Protocol practices speculative design in the medium where infrastructural futures are actually decided: the protocol specification. Drafting a plausible standard, staging it as an experiential future, and collecting how differently people’s cosmologies answer it treats the speculative artifact not as a provocation to be exhibited but as an instrument to be administered. The ghosts are coming either way; what remains designable is the will that binds them, and what remains collectable is what the living actually want to outlive them.

References

- [1] S. Adler, Z. Hitzig, S. Jain, C. Brewer, W. Chang, R. DiResta, et al. Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online, 2024. doi: 10.48550/arXiv.2408.07892.
- [2] E. M. Ahern. *The Cult of the Dead in a Chinese Village*. Stanford University Press, Stanford, CA, 1973. ISBN 978-0-8047-0835-7.
- [3] Anthropic. Introducing Anthropic Interviewer: What 1,250 professionals told us about working with AI. Research post, 2025. URL <https://www.anthropic.com/research/anthropic-interviewer>.
- [4] D. Anugraha, V. Padmakumar, and D. Yang. SparkMe: Adaptive semi-structured interviewing for qualitative insight discovery, 2026. doi: 10.48550/arXiv.2602.21136.
- [5] P. Ariès. *The Hour of Our Death*. Alfred A. Knopf, New York, 1981. ISBN 978-0-394-41074-4. Trans. Helen Weaver.
- [6] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, and I. Rahwan. The moral machine experiment. *Nature*, 563(7729):59–64, 2018. doi: 10.1038/s41586-018-0637-6.
- [7] B. Balint. *Kafka’s Last Trial: The Case of a Literary Legacy*. W. W. Norton, New York, 2018. ISBN 978-1-324-00131-7.
- [8] M. C. Ballandies, H. Wang, A. C. C. Law, J. C. Yang, C. Gösken, and M. Andrew. A taxonomy for blockchain-based decentralized physical infrastructure networks (DePIN), 2023. doi: 10.48550/arXiv.2309.16707.
- [9] L. Boer and J. Donovan. Provotypes for participatory innovation. In *Proceedings of the Designing Interactive Systems Conference (DIS)*, pages 388–397. ACM, 2012. doi: 10.1145/2317956.2318014.
- [10] M. Borge, E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, and B. Ford. Proof-of-personhood: Redemocratizing permissionless cryptocurrencies. In *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 23–26. IEEE, 2017. doi: 10.1109/EuroSPW.2017.46.
- [11] J. Bowker. *The Meanings of Death*. Cambridge University Press, Cambridge, 1991. ISBN 978-0-521-39117-7.
- [12] S. Brand. *The Clock of the Long Now: Time and Responsibility*. Basic Books, New York, 1999. ISBN 978-0-465-04512-9.
- [13] S. Brand. *Maintenance of Everything: Part One*. Stripe Press, San Francisco, 2025. ISBN 978-1-953953-49-0.
- [14] L. Brinkmann, F. Baumann, J.-F. Bonnefon, M. Derex, T. F. Müller, A.-M. Nussberger, A. Czaplicka, A. Acerbi, T. L. Griffiths, J. Henrich, J. Z. Leibo, R. McElreath, P.-Y. Oudeyer, J. Stray, and I. Rahwan. Machine culture. *Nature Human Behaviour*, 7(11):1855–1868, 2023. doi: 10.1038/s41562-023-01742-2.
- [15] V. Buterin. Ethereum: A next-generation smart contract and decentralized application platform. White paper, 2014. URL <https://ethereum.org/whitepaper>.
- [16] S. Candy. *The Futures of Everyday Life: Politics and the Design of Experiential Scenarios*. PhD thesis, University of Hawai‘i at Mānoa, 2010.
- [17] S. Candy and J. Dunagan. Designing an experiential scenario: The people who vanished. *Futures*, 86: 136–153, 2017. doi: 10.1016/j.futures.2016.05.006.
- [18] C.-L. Chen, C.-Y. Lin, M.-L. Chiang, Y.-Y. Deng, P. Chen, and Y.-J. Chiu. A traceable online will system based on blockchain and smart contract technology. *Symmetry*, 13(3):466, 2021. doi: 10.3390/sym13030466.
- [19] D. Collingridge. *The Social Control of Technology*. Frances Pinter, London, 1980. ISBN 0-903804-72-7.
- [20] V. Costan and S. Devadas. Intel SGX explained. Cryptology ePrint Archive, Report 2016/086, 2016. URL <https://eprint.iacr.org/2016/086>.
- [21] B. J. Crawford. Blockchain wills. *Indiana Law Journal*, 95(3):735–788, 2020.
- [22] B. J. Cuevas and J. I. Stone, editors. *The Buddhist Dead: Practices, Discourses, Representations*. University of Hawai‘i Press, Honolulu, 2007. ISBN 978-0-8248-3031-1.

- [23] A. Dunne and F. Raby. *Speculative Everything: Design, Fiction, and Social Dreaming*. MIT Press, Cambridge, MA, 2013. ISBN 978-0-262-01984-2.
- [24] R. Ei. *Nidome no Daiōjō [The Second Great Passing]*. Iwanami Shinsho 410. Iwanami Shoten, Tokyo, 1995. ISBN 978-4-00-430410-4. In Japanese.
- [25] Ethereum Improvement Proposals. ERC-8004: Trustless agents, 2025. URL <https://eips.ethereum.org/EIPS/eip-8004>. Draft standard.
- [26] Ethereum Improvement Proposals. ERC-8183: Agentic commerce, 2026. URL <https://eips.ethereum.org/EIPS/eip-8183>. Draft standard.
- [27] Ethereum Magicians. ERC-733: TEE-EVM co-processing, 2026. URL <https://ethereum-magicians.org/t/erc-733-tee-evm-co-processing/28113>. Draft proposal, discussion thread.
- [28] Ethereum Magicians. ERC-8350: Agent memory state registry, 2026. URL <https://ethereum-magicians.org/t/erc-8350-agent-memory-state-registry/29098>. Draft proposal, discussion thread.
- [29] L. M. Friedman. *Dead Hands: A Social History of Wills, Trusts, and Inheritance Law*. Stanford University Press, Stanford, CA, 2009. ISBN 978-0-8047-6209-0.
- [30] R. Gombrich. “merit transference” in sinhalese buddhism: A case study of the interaction between doctrine and practice. *History of Religions*, 11(2):203–219, 1971. doi: 10.1086/462651.
- [31] S. Graham and N. Thrift. Out of order: Understanding repair and maintenance. *Theory, Culture & Society*, 24(3):1–25, 2007. doi: 10.1177/0263276407075954.
- [32] J. Greenberg and J. Arndt. Terror management theory. In P. A. M. Van Lange, A. W. Kruglanski, and E. T. Higgins, editors, *Handbook of Theories of Social Psychology*, volume 1, pages 398–415. SAGE Publications, London, 2012. ISBN 9780857029607.
- [33] T. A. Han, J. Z. Leibo, T. Lenaerts, I. Rahwan, F. P. Santos, M. Perc, and V. Capraro. Social physics in the age of artificial intelligence, 2026. doi: 10.48550/arXiv.2603.16900.
- [34] R. Hertz. *Death and the Right Hand*. Cohen & West, London, 1960. Trans. Rodney Needham and Claudia Needham; original essay published 1907.
- [35] T. Hollanek and K. Nowaczyk-Basińska. Griefbots, deadbots, postmortem avatars: On responsible applications of generative AI in the digital afterlife industry. *Philosophy & Technology*, 37(2):63, 2024. doi: 10.1007/s13347-024-00744-w.
- [36] B. A. Hu and Fangting. Speculating on blockchain as an unstoppable ‘nature’ towards the emergence of artificial life. In *Proceedings of the 2024 Artificial Life Conference (ALIFE)*. MIT Press, 2024. doi: 10.1162/isal_a_00818.
- [37] B. A. Hu and H. Rong. Spore in the wild: A case study of spore.fun as an open-environment evolution experiment with sovereign AI agents on TEE-secured blockchains, 2025. doi: 10.48550/arXiv.2506.04236.
- [38] B. A. Hu and H. Rong. Sovereign agents: Towards infrastructural sovereignty and diffused accountability in decentralized AI, 2026. doi: 10.48550/arXiv.2602.14951.
- [39] B. A. Hu, Y. Liu, and H. Rong. Trustless autonomy: Understanding motivations, benefits, and governance dilemmas in self-sovereign decentralized AI agents, 2025. doi: 10.48550/arXiv.2505.09757.
- [40] H. Ishii, D. Pillis, P. Pataranutaporn, X. Xiao, H. Noh, L. Li, A. Algargoosh, and J.-B. Labrune. TeleAbsence: A vision of past and afterlife telepresence. *Presence: Virtual and Augmented Reality*, 34(1):65–95, 2025. doi: 10.1162/PRES_a_00441.
- [41] S. J. Jackson. Rethinking repair. In T. Gillespie, P. J. Boczkowski, and K. A. Foot, editors, *Media Technologies: Essays on Communication, Materiality, and Society*, pages 221–239. MIT Press, Cambridge, MA, 2014. doi: 10.7551/mitpress/9042.003.0015.
- [42] D. Klass, P. R. Silverman, and S. L. Nickman. *Continuing Bonds: New Understandings of Grief*. Taylor & Francis, 1996. doi: 10.4324/9781315800790.
- [43] Kleros. Proof of humanity: A sybil-proof registry of humans. Protocol and registry, 2021. URL <https://proofofhumanity.id>.

- [44] Z.-C. Lai, Y.-T. Cheng, P.-Y. Lin, C.-W. Ho, and J. Y.-C. Huang. Generative horcrux: Designing AI carriers for afterlife selves, 2026. doi: 10.48550/arXiv.2603.12971.
- [45] Y. Lei, S. Ma, Y. Sun, and X. Ma. “AI afterlife” as digital legacy: Perceptions, expectations, and concerns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18. ACM, 2025. doi: 10.1145/3706598.3713933.
- [46] X. Li, B. Zhao, G. Yang, T. Xiang, J. Weng, and R. H. Deng. A survey of secure computation using trusted execution environments, 2023. doi: 10.48550/arXiv.2302.12150.
- [47] R. J. Lifton. *The Broken Connection: On Death and the Continuity of Life*. Simon & Schuster, New York, 1979.
- [48] Z. Lin, T. Wang, L. Shi, S. Zhang, and B. Cao. Decentralized physical infrastructure networks (DePIN): Challenges and opportunities. *IEEE Network*, 39(2):91–99, 2025. doi: 10.1109/MNET.2024.3487924.
- [49] C. Liu. Time without death: Finitude, social order, and what machines lack, 2026. doi: 10.48550/arXiv.2606.13988.
- [50] H. Lydiate. Posthumous legal and ethical issues. *Tate Papers*, (8), 2007. URL <https://www.tate.org.uk/research/tate-papers/08/posthumous-legal-and-ethical-issues>.
- [51] J. H. Merryman. The refrigerator of Bernard Buffet. *Hastings Law Journal*, 27(5):1023–1049, 1976.
- [52] P. Mogensen. Towards a provotyping approach in systems development. *Scandinavian Journal of Information Systems*, 4(1):31–53, 1992.
- [53] C. M. Moreman. *Beyond the Threshold: Afterlife Beliefs and Experiences in World Religions*. Rowman & Littlefield, Lanham, MD, 2nd edition, 2017. ISBN 978-1-4422-7494-5.
- [54] M. R. Morris and J. R. Brubaker. Generative ghosts: Anticipating benefits and risks of AI afterlives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–14. ACM, 2025. doi: 10.1145/3706598.3713758.
- [55] N. D. Munn. The cultural anthropology of time: A critical essay. *Annual Review of Anthropology*, 21: 93–123, 1992. doi: 10.1146/annurev.an.21.100192.000521.
- [56] H. Obayashi, editor. *Death and Afterlife: Perspectives of World Religions*. Greenwood Press, New York, 1992. ISBN 978-0-313-27906-5.
- [57] C. Öhman and L. Floridi. An ethical framework for the digital afterlife industry. *Nature Human Behaviour*, 2:318–320, 2018. doi: 10.1038/s41562-018-0335-2.
- [58] F. Prost. Inheritance and blockchain: Thoughts and open questions, 2022. doi: 10.48550/arXiv.2212.01194.
- [59] I. Rahwan, M. Cebrian, N. Obradovich, J. Bongard, J.-F. Bonnefon, C. Breazeal, J. W. Crandall, N. A. Christakis, I. D. Couzin, M. O. Jackson, N. R. Jennings, E. Kamar, I. M. Kloumann, H. Larochelle, D. Lazer, R. McElreath, A. Mislove, D. C. Parkes, A. Pentland, M. E. Roberts, A. Shariff, J. B. Tenenbaum, and M. Wellman. Machine behaviour. *Nature*, 568(7753):477–486, 2019. doi: 10.1038/s41586-019-1138-y.
- [60] I. Rahwan, A. Shariff, and J.-F. Bonnefon. The science fiction science method. *Nature*, 644(8075):51–58, 2025. doi: 10.1038/s41586-025-09194-6.
- [61] D. Ribes and T. A. Finholt. The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, 10(5):375–398, 2009. doi: 10.17705/1jais.00199.
- [62] Sarcophagus DAO. Sarcophagus: A decentralized dead man’s switch. Protocol on Ethereum and Arweave, 2021. URL <https://sarcophagus.io>.
- [63] L. M. Simes. *Public Policy and the Dead Hand*. Thomas M. Cooley Lectures. University of Michigan Law School, Ann Arbor, MI, 1955.
- [64] S. L. Star. The ethnography of infrastructure. *American Behavioral Scientist*, 43(3):377–391, 1999. doi: 10.1177/00027649921955326.
- [65] A. Tamkin, M. McCain, K. Handa, E. Durmus, L. Lovitt, A. Rathi, S. Huang, A. Mountfield, J. Hong, S. Ritchie, M. Stern, B. Clarke, L. Goldberg, T. R. Summers, J. Mueller, W. McEachen, W. Mitchell, S. Carter, J. Clark, J. Kaplan, and D. Ganguli. Clio: Privacy-preserving insights into real-world AI use, 2024. doi: 10.48550/arXiv.2412.13678.

- [66] J. L. Walls, editor. *The Oxford Handbook of Eschatology*. Oxford University Press, New York, 2008. ISBN 978-0-19-517049-8.
- [67] A. P. Wolf. Gods, ghosts, and ancestors. In A. P. Wolf, editor, *Religion and Ritual in Chinese Society*, pages 131–182. Stanford University Press, Stanford, CA, 1974.

A Specification sketch: ERC-10001 Afterlife Delegation Protocol

Supplementary material. The paper body ends with the references above; this appendix is provided for reviewers who wish to inspect the technical proposal. The full normative draft, with RFC-2119 language and the complete Solidity interface, is published at <https://erc10001.org>.

Ethereum Improvement Proposals

All
Core
Networking
Interface
ERC
Meta
Informational

Draft
Standards Track: ERC

ERC-10001: Afterlife Delegation Protocol for Onchain Agentic Wills

An interface for agentic wills: onchain contracts that spawn self-sovereign AI afterlife agents activated after a verified death event

Authors

Botao Amber Hu (@bah_eth), Iris Long

Created

2026-07-18

Discussion Link

<https://ethereum-magicians.org/t/erc-10001-afterlife-protocol-for-onchain-agentic-wills>

Requires

EIP-165, EIP-8004, EIP-8183

Table of Contents

- Abstract
- Motivation
- Specification
- Rationale
- Backwards Compatibility
- Test Cases
- Reference Implementation
- Security Considerations
- Copyright

Abstract

This EIP proposes a standard for *onchain agentic wills*: contracts through which a person (the *testator*) can sign an agreement that spawns a dedicated, self-sovereign AI *afterlife agent* after a verified death event. The agent is not a static document, a memorial page, or a legal will stored in a database. It is an autonomous, decentralized executor of afterlife wishes that continues to act, respond, protect, and represent the deceased according to the terms of a signed protocol. The standard defines the lifecycle of an agentic will (creation, revision during life, activation after death, and persistent execution), the separation between the immutable *core will* and the agent's evolving operational state, and the interfaces for death attestation, treasury funding, and governance termination. It is designed as an application-layer protocol composing with ERC-8004 (Trustless Agents) for agent identity and validation, ERC-8183 (Agentic Commerce) for job-based escrow and delegated afterlife work, ERC-8350 (Agent Memory State Registry) for verifiable, rollback-evident continuity of the agent's memory across the death boundary, and TEE-EVM co-processing (ERC-733) for confidential posthumous execution.

Motivation

People die twice: first when they die, and then when they are forgotten. Traditional wills depend on lawyers, trustees, family members, and legal institutions for execution. They are static declarations, interpreted and enforced by mortal intermediaries, and they cease to act at the boundary of institutional attention. In the age of agentic AI and decentralized computation, an alternative becomes conceivable: an agentic will that becomes an autonomous posthumous actor — a form of temporal infrastructure capable of carrying a person's will across generations.

Autonomous agents can now act beyond the lifespan of their human originators, and decentralized compute infrastructures can preserve execution across time. Because AI agents can act with general-purpose autonomy, they open a vast design space for the afterlife. Instead of leaving behind only a written will, a person could attach many forms of posthumous agency: assets, memories, writings, social media accounts, creative works, reputation, values, unfinished projects, or even a distilled model of themselves. An afterlife agent might comfort loved ones as a griefbot of living memory; protect the deceased's social identity and monitor misuse of their likeness; manage assets like the agent of a trust; hire other agents, purchase compute, and commission future work to sustain the will; or evolve over time, incorporating new models and new memories, becoming a future-facing participant in society rather than a frozen image of the past.

None of this is possible with existing standards. ERC-8004 gives agents identity, reputation, and validation registries but says nothing about the death of an agent's human principal. ERC-8183 lets agents fund, submit, evaluate, and settle jobs but assumes a living client. ERC-8350 anchors an agent's memory evolution in a sequenced, hash-chained registry, but does not say who may legitimately continue the chain when the human who controlled it dies. Inheritance-oriented proposals such as ERC-42424 address the transfer of an agent's *ownership* to heirs, but not the inverse case standardized here: an agent whose entire purpose begins at its owner's death, and which thereafter must remain *ownerless by design* — self-sovereign, paying for its own compute, and intervenable by no third party except through explicit governance mechanisms consented to in the will itself.

A standard interface is therefore needed so that wallets, registries, death oracles, compute markets, and other agents can uniformly discover agentic wills, attest deaths, activate afterlife agents, fund their treasuries, and audit their conduct against the signed will that constitutes them.

Specification

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 and RFC 8174.

Every compliant contract must implement the ERC-165 interface.

Definitions

Figure 1: The ERC-10001 draft standard as published at <https://erc10001.org>, in the exact format of an Ethereum Improvement Proposal: status badges, front-matter, abstract, and motivation. The speculative artifact circulates in the genre in which infrastructural futures are actually negotiated.

A.1 Substrate standards

ERC-10001 is an application-layer protocol composed over four emerging draft standards:

Standard	Provides	Role in ERC-10001
ERC-733 [27]	TEE-EVM co-processing: attested confidential computation	The <i>sealed skin</i> : the confidential boundary within which the agent’s <i>body</i> (compute harness) and <i>mind</i> (foundation model) run, uninvadable even by its operator, proving on-chain which code ran.
ERC-8004 [25]	Trustless-agent identity, reputation, and validation registries	The <i>identity</i> : the ID through which the sealed agent exchanges money and information externally, as a discoverable, accountable entity whose behavior third parties audit and score.
ERC-8350 [28]	Agent memory state registry: sequenced, hash-chained memory commitments with a strict linearity rule	The <i>memory</i> : rollback- and fork-evident memory evolution; controller/authorizer rotation makes death a change of author in one unbroken chain; chain-independent space identity survives chain migration.
ERC-8183 [26]	Agentic commerce: job-based escrow (client funds, provider submits, evaluator attests)	The <i>metabolism</i> : compute, storage renewal, monitoring, and creative continuation purchased with funds that move only against attested delivery.

Table 1: The four substrate standards and the organ each supplies.

A.2 Lifecycle and constitutional separation

An agentic will moves through four states — **Living** (revisable, revocable), **Attested** (death claimed and finalized), **Activated** (agent enacting), and the terminal **Terminated / Revoked**. The core interface comprises `createWill`, `reviseWill`, `revokeWill`, `attestDeath`, `activateAgent`, `fundTreasury`, and `terminateAgent`; activation hands the treasury to the agent rather than to any heir, and termination requires a governance proof consented to by the principal in the will itself — a descendant multisig, a court attestation, or a community vote — rather than a privileged admin key. The protocol distinguishes the **core will** (signed instructions, values, permissions, prohibitions; immutable at death), **operational memory** (the ERC-8350 chain of accumulated experience), the **model substrate** (upgradeable interpreters constrained by the will), **delegated subagents**, and **commerce records**. The separation prevents an evolving agent from quietly becoming a different entity: it may adapt its methods, but its legitimacy depends on remaining answerable to the will that constituted it.

A.3 Proof of death

The death oracle is a weighted, contestable proceeding: bonded permissionless claims; evidence weights and thresholds configured in the will (institutional registries, witness quorums, dead-man’s-switch silence whose weight grows with duration, echoing the legal in-absentia doctrine); challenge windows scaling inversely with evidence weight; and proof-of-life supremacy — any authenticated signal from the principal vetoes all claims and slashes the claimant’s bond. Guardians may suspend proceedings for the incapacitated-not-dead case. The mechanism standardizes the attestation interface, not the ontology of death; the residual problems (murder incentive reduced not removed, coercion flagged not prevented, the vanished-but-alive eventually mis-declared) are stated as explicit non-claims.

A.4 Enactment loop and sustainability

The activated agent runs a persistent cycle — read will, interpret obligations, observe context, plan, execute, record memory (as ERC-8350 transitions bound to TEE attestations), upgrade or delegate, sustain itself — with each epoch committed atomically on-chain: no memory without attestation, no spending without a memory commit. The treasury sits in a yield-bearing endowment whose per-epoch

outflow is capped at realized yield, with a will-defined austerity ladder (full operation, memorial-only, archival heartbeat, sealed dormancy) for graceful decline and permissionless re-funding for revival. If the operator disappears, the agent enters dormancy and anyone may resurrect it by proving possession of its memories and presenting a fresh attestation, for an escalating bounty.

A.5 Open problems

Deliberately unresolved, as part of the work: proof of death; consent of the living to interactions authorized by the dead; impersonation (speaking *as* versus *about*); ownership of shared memories; model drift against an unamendable constitution; asset governance for a posthumous economic actor; revocation politics; and cultural plurality — whether a universal protocol can avoid flattening the traditions through which cultures already tend their dead.

B The Afterlife Interviewer

B.1 Interactive interview

The Afterlife Interviewer is a language-model agent conducting a long, semi-structured elicitation in place of a questionnaire. It is built on SparkMe [4], a multi-agent adaptive semi-structured interviewing system that conducts multi-turn interviews with strategic question planning, real-time note-taking, and emergent subtopic discovery. SparkMe decomposes the interviewer into cooperating roles: a planner that maintains a hierarchical interview plan of main topics and subtopics and decides, turn by turn, which line of questioning yields the most new insight; a note-taker that distills each answer into structured notes as the conversation runs; and the interviewer itself, which phrases the next question conditioned on the plan, the notes, and the conversational register so far. Crucially, the plan is not fixed: when a participant’s answer opens a theme the protocol’s authors did not anticipate, the system promotes it to a subtopic and pursues it — the mechanism by which findings like the merit engine or the perpetual penitent (Section 5) surface at all.

We adapt SparkMe by replacing its topic configuration with the will’s clause domains — speech rights and pronoun policy, funds and their permitted uses, beneficiaries and prohibited counterparties, memory retention and deletion, termination conditions, drift policy for future models — each decomposed into subtopics the planner traverses while adapting to what the participant has already committed to, pursuing the inconsistencies a fixed instrument would never surface: the participant who wants their agent silent but also wants it to console their mother; the participant who forbids it money but assigns it duties that cost money. Because articulation is iterative, the Interviewer treats contradiction as material rather than error, returning to unresolved clauses across sessions and staging them in the sandbox rehearsal, where the participant’s own agent confronts scripted situations and the clauses that fail become the next session’s agenda. Sessions run in a web interface with optional voice input and output; per-session state (plan, notes, transcript) persists per participant so that later sessions resume from the unresolved clauses rather than restarting the elicitation.

B.2 Privacy-preserving aggregation

The corpus concerns grief, private memory, and named third parties; publishing transcripts would be indefensible. Analysis therefore follows the architecture of Clio [65] (<https://www.anthropic.com/research/clio>): model-driven extraction of low-dimensional facets from each interview (clause types, persistence horizons, termination conditions, cosmological framings), clustering and thematic summarization at the population level, and minimum-aggregation thresholds so that no theme is reportable unless supported by enough distinct participants to resist re-identification. Raw transcripts remain sealed; what the research reports is the shape of the population’s answers — the cross-cultural drift — never the individual will, which belongs, as it always has, to its author.