Hyper-Entities

Consensus Highlights Report

Linda Petrini & Beatrice Erkers
Foresight Institute
February 2026

Contents

1. Introduction 3
   1.1. What Are Hyper-Entities? 3
   1.2. Project Overview 3
2. Methodology 3
   2.1. Stage 1: Hyper-Entity Qualification (max 27) 3
   2.2. Stage 2: Technology Impact Assessment (max 70) 4
   2.3. Stage 3: d/acc Values Alignment (max 20) 5
   2.4. Curation Process 5
3. Consensus Entities 6
   3.1. Energy & Infrastructure 6
   3.2. Manufacturing & Matter 9
   3.3. Truth & Epistemic Infrastructure 12
   3.4. Governance & Collective Intelligence 15
   3.5. Markets & Incentive Systems 19
   3.6. Ethics & Moral Expansion 22
   3.7. AI & Human Agency 24
   3.8. Interfaces & Augmentation 27
   3.9. Science & Discovery 32
4. Conclusion 36

1. Introduction

1.1. What Are Hyper-Entities?

A **hyper-entity** is a coherent, future-instantiated system that does not yet exist, but is treated as if it will; whose realization would create a new stable action space for humanity; and which already reorganizes coordination, investment, and narrative around its anticipated existence.

Three characteristics define a hyper-entity:

1. **Non-existent but treated as real** — the system doesn't exist yet, but people act as if it will.
2. **Creates new action spaces** — it would enable fundamentally new things humans can do.
3. **Pre-real effects** — it already reorganizes coordination, investment, and narrative *now*.

Historical examples include the Internet (pre-1990s), which reorganized telecoms R&D, policy, and venture capital before widespread deployment; the Space Race (1950s–60s), where Moon missions organized national budgets and education systems before any launches; and AGI today, which reshapes AI research priorities, corporate strategies, and policy discussions despite not yet existing.

1.2. Project Overview

This project set out to systematically identify, score, and curate hyper-entities emerging from the discourse around the Foresight Institute's research community. The source material comprises:

- **65 podcast transcripts** from the Foresight Institute podcast series, featuring researchers, technologists, and thinkers at the frontier of science and governance.
- **40 world-gallery submissions** from the Foresight Institute's Existential Hope project, presenting speculative visions of positive futures.
- **2 AI-pathways essays**, including Vitalik Buterin's d/acc framework and analysis of Tool AI approaches.
- **1 x-hope document** providing additional framing.

From this corpus, over 300 candidate hyper-entities were extracted, scored across three assessment stages, clustered thematically, and then curated through an independent review process by two researchers (Linda Petrini and Beatrice Erkers) to arrive at a final consensus list of 39 highlighted entities.

2. Methodology

The methodology follows a multi-stage pipeline: extraction, scoring, clustering, and human curation.

2.1. Stage 1: Hyper-Entity Qualification (max 27)

Each candidate is scored on 9 axes (0–3 per axis). A candidate must score **18 or above** to qualify as a hyper-entity. Scores of 12–17 are borderline and require strong justification.

Axis	0	1	2	3
Non-existence	Already deployed	Prototype exists	Partial demos	No instantiation
Plausibility	Fantasy	Hand-wavy	Credible theory	Active scientific path
Design specificity	Vague aspiration	Metaphor only	Coarse architecture	Detailed system model
New action space	Incremental	Narrow new ability	Broad new capability	Entirely new verb
Roadmap clarity	None	Wishful	Research agenda	Multi-stage roadmap
Coordination gravity	Solo interest	Small niche	Multiple orgs	Global coordination
Resource pull	Trivial	Some grants	Serious capital	Massive capital + talent
Narrative centrality	Ignored	Peripheral	Often referenced	Justifies present actions
Pre-real effects	None	Speculation	Market/policy shifts	Institutions reorganize

Table 1: Stage 1 scoring rubric (9 axes, 0–3 each, max 27).

2.2. Stage 2: Technology Impact Assessment (max 70)

Qualified entities are assessed on 14 dimensions (0–5 each) measuring potential impact and risk. Higher scores indicate greater magnitude of effect — not necessarily positive.

Dimension	What is scored
Capability Discontinuity	New powers unlocked (incremental … phase change)
Cross-Domain Reach	How many sectors it touches
Scalability	Speed and cost of global spread
Autonomy	Operates without ongoing human control
Composability	Can be embedded everywhere
Feedback Intensity	Strength of self-reinforcing loops
Irreversibility	Difficulty of rollback once deployed
Power Concentration	Centralizes leverage
Externality Magnitude	Size of spillovers
Misuse Asymmetry	Harm vs. benefit ratio
Governance Lag	Gap vs. existing institutions
Narrative Lock-In	Inevitable story it enforces
Path Dependency	Futures foreclosed
Human Agency Impact	Effect on human choice

Table 2: Stage 2 technology assessment dimensions (14 axes, 0–5 each, max 70).

2.3. Stage 3: d/acc Values Alignment (max 20)

Based on Vitalik Buterin's d/acc framework (Defensive / Decentralized / Democratic / Differential Acceleration), this stage evaluates whether a hyper-entity aligns with values that promote human flourishing, resilience, and freedom. Each entity is scored on 4 dimensions (0–5 each).

Dimension	What is scored
Democratic	Enables collective decision-making vs. concentrates decisions in elites
Decentralized	Distributes power broadly vs. creates single points of control
Defensive	Favors protection over harm; defense-favoring
Differential	Creates positive asymmetries; improves defense and freedom

Table 3: Stage 3: d/acc values alignment (4 dimensions, 0–5 each, max 20).

2.4. Curation Process

After automated extraction and scoring, the entities were presented to two independent researchers through a review interface:

1. **Independent shortlisting**: Linda and Beatrice each independently reviewed the full set of scored entities and starred those they considered most significant. Linda selected 88 entities; Beatrice selected 67.
2. **Overlap identification**: 26 entities were starred by *both* reviewers (20.2% overlap rate), forming the core consensus set.

3. **Cross-review voting**: Each reviewer then reviewed the other's unique picks via a voting interface, voting Yes on entities they also found compelling. Linda voted Yes on 8 of Beatrice's unique picks; Beatrice voted Yes on 5 of Linda's unique picks.
4. **Final list**: The resulting 39 entities (26 shared + 8 Linda's Yes + 5 Beatrice's Yes) constitute the consensus highlights presented in this report.

3. Consensus Entities

The 39 consensus entities are organized into nine thematic groups. For each entity, we provide its description, d/acc and technology scores, a summary of the current state of the art, and an assessment of viable next steps and key challenges.

3.1. Energy & Infrastructure

3.1.1. Decentralized Adaptive Energy Network

An AI-optimized, decentralized energy network with dynamic routing, local energy production, and intelligent demand prediction, transforming traditional centralized energy distribution models.

d/acc: 18/20 Tech: 51/70

State of the Art. Several building blocks exist today but no fully integrated decentralized adaptive energy network operates at scale. Virtual power plants (VPPs) aggregate distributed energy resources – Tesla's Autobidder and Virtual Power Plant program in South Australia (launched 2020, expanded since) connects thousands of home Powerwalls to collectively act as a grid-scale battery. Germany's Sonnen Community and Australia's Reposit Power offer peer-to-peer energy trading among prosumers with rooftop solar and batteries. AI-driven grid optimization is deployed by companies like AutoGrid, Sense, and GridBeyond, which use machine learning for demand forecasting and real-time load balancing. On the blockchain side, Energy Web Foundation (EWF) has built a dedicated blockchain for energy sector decentralization, and projects like Power Ledger in Australia enable peer-to-peer energy trading. Advanced metering infrastructure (AMI) and distributed energy resource management systems (DERMS) are being deployed by major utilities, but these remain largely centralized control systems with some distributed elements.

Next Steps & Challenges. The most viable near-term path is incremental: expanding existing VPP and peer-to-peer trading platforms, integrating AI-driven demand prediction at the distribution level, and leveraging smart inverters and vehicle-to-grid (V2G) technology as EVs become widespread grid assets. The primary barrier is regulatory and institutional – electricity markets in most jurisdictions were designed for centralized generation and one-way power flow, and utilities have regulatory structures and revenue models that resist true decentralization. Technical challenges include the need for sub-second coordination among millions of distributed devices, cybersecurity risks inherent in a massively distributed attack surface, and the difficulty of maintaining grid stability (voltage and frequency regulation) without centralized dispatch authority. Interoperability standards remain fragmented (IEEE 2030.5, OpenADR, Matter, etc.), making seamless device-to-device coordination difficult. True decentralized adaptive networks will likely emerge gradually over 10-20 years as regulatory reform, hardware costs (batteries, smart inverters), and AI coordination capabilities converge, rather than through any single deployment.

3.1.2. Deep Fision Micro Nuclear Reactors

A small, factory-producible nuclear reactor designed to be buried a mile underground, providing localized, safe, and clean energy. Designed to be mass-manufactured like automobiles, with each reactor about the size of a Toyota.

d/acc: 16/20 Tech: 47/70

State of the Art. Deep Fission (spelled 'Fision' by the company) is a startup building small nuclear reactors designed to fit through a manhole cover and operate a mile underground in boreholes.

Backed by Pablos Holman's Intellectual Ventures spinout ecosystem, the company sits within a broader wave of over 100 SMR (small modular reactor) ventures competing for regulatory approval. NuScale Power received the first-ever NRC design certification for an SMR in 2023, though its first project (at Idaho National Lab) was cancelled due to cost overruns. TerraPower, backed by Bill Gates, broke ground on its Natrium reactor in Wyoming in 2024, while Kairos Power received a construction permit for its Hermes test reactor in 2023. The US Nuclear Regulatory Commission has undergone significant reform to accelerate licensing for advanced reactor designs, and Sweden's parliament has moved to expand nuclear capacity, signaling a European shift back toward nuclear.

Next Steps & Challenges. The most viable path forward is securing NRC design certification for borehole-emplaced reactor concepts, which requires demonstrating passive safety in a regulatory category that does not yet exist – no framework currently governs autonomous underground nuclear generation. Factory manufacturing at automotive scale demands building a dedicated gigafactory and supply chain for reactor components at a precision and throughput never attempted for nuclear hardware. The key barriers are threefold: regulatory frameworks must be written essentially from scratch for this deployment mode, the upfront capital for factory tooling is enormous before a single unit generates revenue, and public acceptance of distributed underground nuclear – even with inherent safety advantages – remains an untested social proposition. No borehole reactor has yet been built or tested at prototype scale, placing this concept several years behind above-ground SMR competitors that already have physical test facilities.

3.1.3. Climate Adaptation Jurisdictional Arbitrage

A strategic approach to creating special economic zones and new cities optimized for climate migration, leveraging legal and economic frameworks to proactively manage population displacement caused by environmental changes.

d/acc: 16/20 Tech: N/A

State of the Art. Several real-world experiments in climate-optimized special jurisdictions and planned cities are underway, though none yet explicitly frame themselves as climate migration destinations. Saudi Arabia's NEOM project (announced 2017, construction ongoing) represents the largest attempt at building a climate-adapted new city, with estimated costs exceeding $500 billion, though it faces significant feasibility questions and is driven primarily by economic diversification rather than climate migration. Honduras's ZEDEs (Zones for Employment and Economic Development) operated from 2013-2022 before being repealed by the Xiomara Castro government, with Prospera on Roatan being the most prominent example of a charter city with independent governance – it is now in international arbitration claiming $10.8 billion in damages. In the charter cities space, the Charter Cities Institute (founded by Paul Romer's intellectual legacy) actively advises governments in Africa and Central America on new city frameworks, while Praxis (founded 2019) is attempting to build a technology-focused new city. On the climate migration side, the World Bank estimated in 2021 that 216 million people could be internally displaced by climate change by 2050, and the Nansen Initiative (now the Platform on Disaster Displacement) has developed non-binding protection frameworks, but no country has created dedicated legal pathways or zones for climate migrants.

Next Steps & Challenges. The most viable path involves retrofitting existing special economic zone frameworks with explicit climate adaptation mandates, rather than building entirely new cities from scratch, since greenfield city projects have an extremely high failure rate historically. The key barrier is political: climate-receiving jurisdictions (typically in the Global North or at higher elevations) face strong domestic opposition to large-scale migration, while climate-vulnerable nations have little incentive to create zones that formalize the departure of their populations. Legally, there is no international framework recognizing climate refugees – the 1951 Refugee Convention does not cover environmental displacement, and attempts to expand it have stalled for over a decade.

Successful models would likely need to demonstrate clear economic benefits to host regions, perhaps modeled on the economic zone approaches of Shenzhen or Dubai but with climate resilience as a core design parameter and international burden-sharing financing. The most promising near-term developments may come from Pacific Island nations (Tuvalu, Kiribati) that are exploring digital nationhood and land-purchase agreements with other countries as existential adaptation strategies, and from managed retreat programs within countries like the U.S. (FEMA's Hazard Mitigation Grant Program has relocated some communities) that could evolve into more systematic jurisdictional frameworks.

3.2. Manufacturing & Matter

3.2.1. End-User Programming Ecosystem

A democratized software creation environment where non-technical individuals can design, modify, and generate custom software tailored to their specific needs using AI-enhanced tools.

d/acc: 16/20 Tech: 54/70

State of the Art. The convergence of AI code generation and low-code platforms has created the most fertile period ever for end-user programming. Andrej Karpathy coined the term 'vibe coding' in early 2025, describing the practice of using LLMs to generate code through natural language conversation without examining the output – a practice that has since exploded in adoption. Tools like Replit Agent, Cursor, GitHub Copilot, and Claude Code now enable non-programmers to build functional applications through conversation, while platforms like Val Town, Glide, and Retool provide constrained environments for end-user application development. Sam Arbesman, drawing on Seymour Papert's concept of 'low floors and high ceilings,' frames this as a long-overdue realization of the end-user programming vision that dates back to HyperCard (1987) and Alan Kay's Dynabook concept. Robin Sloan's influential essay framing apps as 'home-cooked meals' has become a touchstone for the movement, emphasizing software built for personal use rather than mass markets.

Next Steps & Challenges. The most viable near-term path is improving AI code generation reliability to the point where non-programmers can confidently build and maintain applications without understanding the underlying code – the 'legacy code' problem Steve Krouse identified, where AI-generated code becomes as opaque as decades-old software nobody understands. Key barriers include the chatbot interface itself, which is likely the wrong modality for building sophisticated graphical or stateful applications; the lack of reliable AI-generated security guarantees, which makes vibe-coded software risky for anything beyond personal use; and the absence of robust testing and verification frameworks for AI-generated code. The deeper challenge is cultural and economic: enterprise software vendors have strong incentives to maintain complexity as a moat, and the professional developer ecosystem may resist tools that genuinely democratize their expertise. Moving from 'anyone can make a toy app' to 'anyone can make production software' requires breakthroughs in multi-modal AI reasoning about complex workflows, persistent state management, and security – problems that remain unsolved even for expert programmers working with current AI tools.

3.2.2. Atomically Precise Manufacturing / Molecular Machine Systems

A technological system enabling precise manipulation of matter at the atomic scale, potentially revolutionizing manufacturing, environmental control, and product creation.

d/acc: 13/20 Tech: 57/70

State of the Art. Atomically precise manufacturing has progressed from pure theory to early laboratory demonstrations. IBM demonstrated positioning individual atoms with scanning tunneling microscopes as early as 1989, and in 2024 researchers at Zyvex Labs and the University of Alberta continued refining hydrogen depassivation lithography to place individual silicon atoms with sub-nanometer precision. DNA origami, pioneered by Paul Rothemund (Caltech, 2006) and advanced by groups at TU Munich and Harvard's Wyss Institute, enables programmable self-assembly of nanoscale structures with atomic precision. The Foresight Institute, co-founded by Christine Peterson, has been a central convener for this field since 1986, hosting annual technical conferences and awarding Feynman Prizes for nanotechnology. Molecular machines earned the 2016 Nobel Prize in Chemistry (Sauvage, Stoddart, Feringa), and David Leigh's group at Manchester has built increasingly complex molecular motors and machines. However, current demonstrations remain limited to placing tens to hundreds of atoms in controlled laboratory conditions, far from the scalable, general-purpose manufacturing systems envisioned.

Next Steps & Challenges. The most promising near-term paths include scaling hydrogen depassivation lithography for semiconductor fabrication (Zyvex Labs is pursuing this commercially), advancing DNA-origami-based assembly for drug delivery and materials science, and developing molecular robotics that can perform multi-step assembly operations. The core technical barrier is throughput: current atom-by-atom placement methods are extraordinarily slow, and no known mechanism bridges the gap between positioning individual atoms and manufacturing macroscopic objects at practical speeds. Thermal noise at room temperature disrupts precise atomic positioning, requiring either cryogenic conditions or error-correction mechanisms that do not yet exist. The field also faces a 'chicken and egg' problem: building the first generation of molecular assemblers likely requires molecular assemblers, and bootstrapping from macro-scale tools to nano-scale fabrication remains an unsolved engineering challenge that Christine Peterson and others estimate may take decades to fully resolve.

3.2.3. Chemputing (Chemical Computing)

A systematic approach to programming chemical reactions using standardized hardware and a specialized programming language, enabling reliable molecular transformations and potentially revolutionizing drug discovery, materials science, and computational chemistry.

d/acc: 13/20 Tech: 44/70

State of the Art. Chemputing is primarily the creation of Lee Cronin's group at the University of Glasgow, who developed both the concept and its practical implementation over the past decade. Cronin built a chemical programming language (XDL, or Chemical Description Language) that compiles synthetic procedures into executable instructions for standardized robotic hardware – what he describes as 'Python for chemistry.' His company Chemify, founded to commercialize this work, is building robotic platforms that can reproducibly execute chemical syntheses from code, with funding support from Schmidt Futures. The Chemputer platform has demonstrated automated synthesis of pharmaceuticals including ibuprofen and other small molecules, with published results showing that identical code run on different hardware produces indistinguishable products. Cronin's group has also connected chemputing to origin-of-life research, using programmable chemistry to explore how selection processes in chemical space can generate complexity without biological evolution.

Next Steps & Challenges. The immediate next steps involve expanding the chemical reaction library that XDL can encode, scaling Chemify's commercial platform to handle a broader range of molecular targets, and building a community of chemist-programmers who adopt the standardized language. The key barriers are chemistry's inherent messiness: unlike digital computation where bits are deterministic, chemical reactions involve continuous variables, side products, and sensitivity to subtle environmental conditions that make universal standardization extremely difficult. Many important synthetic transformations require specialized glassware, catalysts, or conditions that resist the modular hardware abstraction Cronin's system requires. Cultural resistance in chemistry is also significant – as Cronin notes, 'all chemists think programming chemistry is impossible' – and adoption requires convincing a field built on artisanal craft knowledge to embrace software-defined workflows. Scaling from milligram-scale demonstrations to industrial manufacturing introduces additional engineering challenges around heat transfer, mixing, and quality control.

3.2.4. Universal Constructor

A hypothetical machine capable of being programmed to construct virtually anything within the laws of physics, fundamentally transforming human labor and production by eliminating physical toil and enabling exponential self-replication.

d/acc: 13/20 Tech: N/A

State of the Art. The universal constructor concept originates with John von Neumann's theoretical work in the 1940s-1950s on self-replicating automata, and has been substantially developed by David Deutsch and Chiara Marletto through Constructor Theory, a framework that reformulates

physics in terms of which transformations are possible and which are impossible. Marletto's 2021 book 'The Science of Can and Can't' and Deutsch's ongoing work at Oxford provide the most rigorous theoretical foundation. In practice, the closest existing systems are advanced 3D printers (some of which, like RepRap, can partially self-replicate), programmable matter research at MIT's CSAIL and Carnegie Mellon, and biological systems themselves – cells are, in a sense, natural constructors that self-replicate and build complex structures from molecular instructions. NASA has funded studies on self-replicating lunar factories (dating to the 1980 Freitas-Merkle study), and the concept continues to attract theoretical interest. However, no artificial system comes close to the generality, precision, or self-replication capability that defines a true universal constructor.

Next Steps & Challenges. The path to a universal constructor likely requires converging advances in atomically precise manufacturing (for physical fabrication capability), artificial general intelligence (for the programmable control system), and fundamental physics (for understanding the full space of physically possible transformations). The barriers are not merely engineering challenges but touch on deep theoretical questions: Deutsch argues that we do not yet understand the relationship between information and physical transformation well enough to design such a machine. Practical obstacles include the need for a fabrication system that can work across all material types and scales simultaneously, the computational intractability of planning arbitrary construction sequences, and the existential risks posed by a self-replicating machine that could construct virtually anything – a challenge that current governance frameworks are entirely unequipped to address. Realistic timelines, to the extent they can be estimated at all, span multiple decades at minimum, contingent on breakthroughs in several independent fields that are themselves pre-paradigmatic.

3.3. Truth & Epistemic Infrastructure

3.3.1. Epistemic Stack

A comprehensive information verification and tracing system that allows users to follow the provenance of information from high-level claims down to raw data sources, enabling more robust trust and understanding.

d/acc: 17/20 Tech: 44/70

State of the Art. Foundational infrastructure exists in several forms. Semantic Scholar processes over 200 million papers with AI-extracted metadata, citation graphs, and influence scores, while the OpenAlex dataset (launched 2022 by OurResearch as a successor to Microsoft Academic Graph) provides an open catalog of the global research system. Provenance-tracking tools like the W3C PROV standard (2013) and blockchain-based research verification projects like Bloxberg (a consortium of research institutions running an Ethereum-compatible chain since 2019) address parts of the traceability challenge. Fact-checking platforms such as Google's Fact Check Explorer and ClaimBuster use NLP to identify and assess claims, while tools like Scite.ai (founded 2019) classify citations as supporting, contradicting, or mentioning, providing rudimentary claim-level evidence assessment. Wikipedia's citation infrastructure and Wikidata's structured knowledge graph represent the closest existing large-scale systems for linking claims to sources, though neither provides the deep provenance trees envisioned by a full epistemic stack.

Next Steps & Challenges. The most viable path involves layering AI-driven claim extraction and evidence linking on top of existing open knowledge graphs like OpenAlex and Wikidata, creating queryable provenance chains from high-level assertions down to raw data. The core technical challenge is semantic interoperability: mapping claims across disciplines requires domain-specific ontologies that are expensive to build and maintain, and current NLP systems still struggle with the nuanced reasoning needed to determine whether two papers' claims genuinely support or contradict each other. Institutional adoption is equally difficult – researchers have little incentive to add structured metadata to their work beyond what funders require, and the academic reward system still optimizes for publication count rather than epistemic transparency. Building trust in automated claim assessment is perhaps the hardest problem: any system that scores the credibility of scientific claims will face intense scrutiny and accusations of bias, requiring governance structures that are themselves transparent and contestable.

3.3.2. AI-Assisted Epistemological Enhancement System

A technological and methodological approach using AI to improve human reasoning, forecasting, truth-discovery, and collaborative deliberation. This system would leverage AI capabilities to enhance human cognitive processes and decision-making.

d/acc: 16/20 Tech: 54/70

State of the Art. AI-augmented reasoning tools have proliferated since 2023. Forecasting platforms like Metaculus have integrated AI models that match or exceed median human forecaster performance on many question types, while research from Anthropic, OpenAI, and DeepMind has shown that LLMs can identify logical fallacies, flag cognitive biases in text, and generate structured counterarguments. The Collective Intelligence Project (founded 2022 by Divya Siddarth and others) and tools like Polis (used by Taiwan's vTaiwan process since 2015) demonstrate AI-facilitated deliberation that surfaces consensus among large groups. AI debate and Constitutional AI approaches from Anthropic explore using AI systems to improve the quality of reasoning through structured argumentation. Superforecasting research by Philip Tetlock's team has established that specific reasoning practices dramatically improve human prediction accuracy, creating a clear target for AI augmentation. Community Notes on X (formerly Twitter's Birdwatch, launched 2021) uses a bridging-based algorithm to surface contextual information that finds cross-partisan agreement.

Next Steps & Challenges. The most promising near-term applications involve embedding AI reasoning aids into existing decision-making workflows – providing real-time bias detection during policy deliberation, generating structured pros-and-cons analyses, and facilitating large-scale collective intelligence processes through tools like AI-mediated Polis or deliberative polling. The fundamental challenge is that improving human reasoning requires the AI system itself to reason reliably, and current LLMs exhibit well-documented failures in logical consistency, calibration, and susceptibility to sycophancy that could amplify rather than correct human biases. Measuring whether an epistemological enhancement system actually improves reasoning quality requires controlled studies with clear metrics, yet the most important domains (geopolitical strategy, long-term policy) are precisely those where feedback loops are longest and noisiest. There is also a deep governance question: who defines what constitutes 'good reasoning' or 'epistemic improvement,' and how do we prevent such systems from embedding particular ideological frameworks under the guise of objective rationality enhancement?

3.3.3. Distributed Zero-Knowledge Security Systems

A networked cybersecurity infrastructure using advanced cryptographic techniques to protect critical communications and industrial control networks, with distributed threat intelligence and quantum-resistant encryption.

d/acc: 16/20 Tech: 47/70

State of the Art. Zero-knowledge proof technology has matured rapidly since 2022, driven primarily by blockchain scaling applications. ZK-rollup systems like StarkNet (using STARKs) and zkSync (using SNARKs) process millions of transactions with cryptographic verification, and the underlying mathematical frameworks are increasingly being adapted for non-blockchain security applications. NIST finalized its first set of post-quantum cryptographic standards in August 2024, selecting CRYSTALS-Kyber for key encapsulation and CRYSTALS-Dilithium for digital signatures, providing the building blocks for quantum-resistant infrastructure. On the threat intelligence side, the MITRE ATT&CK framework and open-source threat intelligence platforms like OpenCTI and MISP already enable distributed, cross-jurisdictional threat sharing, though they do not yet incorporate ZK proofs. DARPA's SIEVE program (2021-2025) has been specifically developing practical ZK proof systems for defense applications, aiming to enable verification of computations without revealing underlying sensitive data.

Next Steps & Challenges. The most viable path forward combines three parallel tracks: integrating NIST's post-quantum cryptographic standards into existing critical infrastructure, extending ZK proof systems from their current blockchain niche into industrial control system (ICS/SCADA) security, and building federated threat intelligence networks that can share indicators of compromise without exposing network topology. The key technical barrier is computational overhead – current ZK proof generation for complex operations remains orders of magnitude too slow for real-time industrial control system protection, and scaling these proofs across heterogeneous infrastructure (power grids, water systems, telecommunications) introduces enormous interoperability challenges. Institutional barriers are equally severe: critical infrastructure operators in energy, finance, and government are notoriously slow to adopt new cryptographic standards (many still run pre-2010 encryption), and coordinating a transition across allied jurisdictions requires diplomatic agreements on shared security protocols that do not yet exist. The quantum threat timeline remains uncertain, but most estimates place cryptographically relevant quantum computers at 10-20 years out, creating a dangerous complacency window.

3.3.4. Epistemic Infrastructure for Truth Verification

A systematic approach to determining truth and trust in an AI-mediated information ecosystem, involving precise citation, consensus mechanisms, and algorithmic truth assessment.

d/acc: 16/20 Tech: 53/70

State of the Art. The problem Kevin Kelly identified – that AI hallucination is forcing us to develop more rigorous truth-determination infrastructure – has spawned a growing ecosystem of partial solutions. Community Notes on X (formerly Twitter), launched in 2021 and expanded through 2024, represents the most scaled experiment in crowd-sourced truth verification using bridging algorithms that surface consensus across ideological divides. Google's Search Generative Experience and Perplexity AI now provide inline citations for AI-generated claims, implementing the 'precise citation' approach Kelly described. Academic projects like Elicit (by Ought) and Semantic Scholar (by AI2) use AI to trace claim provenance through citation networks. The Consensus app uses LLMs to synthesize findings across peer-reviewed literature with source transparency. Meanwhile, blockchain-based approaches like the Worldcoin-affiliated 'proof of personhood' and content provenance standards like C2PA (Coalition for Content Provenance and Authenticity, backed by Adobe, Microsoft, and the BBC) address the authentication layer.

Next Steps & Challenges. The most viable paths forward involve layering multiple verification approaches: cryptographic content provenance (C2PA), AI-assisted citation tracing, reputation-weighted consensus mechanisms, and probabilistic credibility scoring rather than binary true/false judgments. The fundamental barrier is that truth verification at internet scale requires solving deeply contested epistemological problems – what counts as authoritative evidence varies radically across domains (medicine vs. politics vs. history), and any system powerful enough to authoritatively label claims as true or false becomes an extraordinarily dangerous tool for censorship. The coordination challenge is immense: no single institution can credibly serve as a global truth arbiter without being captured by political or commercial interests, yet a fully decentralized system risks Sybil attacks and motivated reasoning at scale. Current AI systems still lack the contextual reasoning to distinguish between genuinely contested scientific questions and deliberate misinformation, and building the 'nuanced credibility gradients' Kelly envisions requires advances in AI reasoning that go beyond what current large language models can reliably achieve.

3.4. Governance & Collective Intelligence

3.4.1. Competitive Governance Protocol Stack

An interoperable digital governance system allowing citizens to participate in multiple overlapping jurisdictions, with portable digital identities and the ability to switch between governance networks based on performance.

d/acc: 18/20 Tech: 57/70

State of the Art. The conceptual foundations come from Balaji Srinivasan's 'The Network State' (2022), charter city movements led by organizations like the Charter Cities Institute, and special economic zone experiments in Honduras (Prospera), established in 2020 under the ZEDE framework before facing legal challenges in 2022. Estonia's e-Residency program, launched in 2014 and serving over 100,000 e-residents by 2024, demonstrates portable digital identity for governance services across borders. On the technical side, decentralized identity standards like W3C DIDs and Verifiable Credentials (finalized as a W3C Recommendation in 2022) provide the protocol layer for portable credentials, while projects like Gitcoin Passport and Proof of Humanity explore sybil-resistant digital identity. DAOs like Aragon and Snapshot have enabled thousands of on-chain governance experiments, though these remain limited to managing treasuries and protocol parameters rather than providing full governance services.

Next Steps & Challenges. The most plausible path forward involves building interoperability layers between existing special jurisdiction experiments – connecting charter cities, e-residency programs, and DAO governance tools through shared identity and credential standards. The fundamental barrier is sovereignty: nation-states retain a monopoly on legitimate coercion and have no incentive to allow citizens frictionless exit to competing governance providers, as demonstrated by Honduras's 2022 repeal of ZEDE legislation under political pressure. Technical challenges include building zero-knowledge proof systems for credential portability that preserve privacy while preventing fraud, and designing governance performance metrics that are meaningful and resistant to gaming. Perhaps the deepest difficulty is bootstrapping: a governance protocol stack only becomes valuable when multiple jurisdictions adopt it, creating a classic chicken-and-egg coordination problem that existing geopolitical structures actively resist.

3.4.2. LexCommons

A global, open-source legal system powered by smart contracts, AI legal agents, and decentralized governance. It transforms law from a closed, institutional practice to an open, participatory, and technologically mediated global commons.

d/acc: 18/20 Tech: 57/70

State of the Art. Key precursor projects span several domains. Kleros, a decentralized arbitration protocol launched in 2018 on Ethereum, has processed thousands of disputes using crowdsourced jurors incentivized by staked tokens, handling cases from small e-commerce claims to domain name disputes. OpenLaw (acquired by ConsenSys in 2020) and platforms like Juro and Ironclad have built smart contract-based legal agreement tools, though these operate within traditional legal frameworks rather than replacing them. AI legal tools have advanced rapidly: Harvey AI (founded 2022, backed by Sequoia and OpenAI) and CoCounsel by Thomson Reuters (launched 2023) use LLMs for legal research, contract analysis, and brief drafting. On the open-source law front, Creative Commons licenses and the Open Source Hardware Association provide governance templates, while projects like CommonAccord aim to create standardized, modular legal prose objects. However, no system yet combines all these elements into a unified open-source legal infrastructure.

Next Steps & Challenges. The most feasible near-term approach is building modular components – AI-assisted contract drafting, blockchain-based dispute resolution, and open legal code repositories – that can be gradually integrated rather than attempting a monolithic replacement of existing legal

systems. The primary barrier is legitimacy and enforceability: legal systems derive their authority from state recognition and the coercive apparatus behind enforcement, which no decentralized system can replicate without government cooperation. Smart contracts remain brittle when dealing with ambiguous or context-dependent legal situations, and the 2016 DAO hack demonstrated how code-as-law breaks down when outcomes diverge from human intent. Cross-jurisdictional harmonization is extraordinarily difficult – the EU spent decades harmonizing commercial law among member states with shared political will, illustrating the scale of coordination required. Bar associations and legal professions in most jurisdictions actively regulate who can provide legal services, creating regulatory barriers to fully automated or community-governed legal alternatives.

3.4.3. Gevulot (Privacy Infrastructure)

A fine-grained privacy control system where individuals can dynamically set permissions for data capture and sharing about themselves, fundamentally restructuring information privacy.

d/acc: 18/20 Tech: N/A

State of the Art. Gevulot is specifically a Layer 1 blockchain protocol focused on providing a decentralized proving layer for zero-knowledge proofs, announced in 2023 and under active development. More broadly, the privacy infrastructure landscape has advanced significantly: zero-knowledge proof systems have matured from theoretical constructs to production deployments, with zkSync, StarkNet, and Polygon zkEVM processing millions of transactions using ZK-rollups by 2024. Apple's App Tracking Transparency framework (2021) gave iOS users binary control over cross-app tracking, and the EU's GDPR (2018) established data subject rights including access, deletion, and portability. Tim Berners-Lee's Solid project provides personal data pods where users control access to their data, and Inrupt (his company) has been piloting with the Flemish government and the BBC. Confidential computing technologies like Intel SGX, AMD SEV, and ARM CCC enable computation on encrypted data, while homomorphic encryption has seen performance improvements from Microsoft SEAL, Zama's TFHE library, and Duality Technologies, though it remains too slow for most real-time applications.

Next Steps & Challenges. The most viable near-term path combines zero-knowledge proofs for selective credential disclosure (proving age without revealing birthdate, proving income bracket without revealing salary) with decentralized identity standards (W3C Verifiable Credentials, DID specifications) to give individuals fine-grained control over personal data sharing. The key barrier is the incumbent data economy: companies like Google, Meta, and data brokers derive their business models from aggregating and monetizing personal data, creating enormous lobbying pressure against true user-controlled privacy infrastructure. Technical challenges include making ZK-proof generation fast and cheap enough for consumer devices, building user interfaces that make privacy controls comprehensible to non-technical users, and handling the tension between individual privacy rights and legitimate societal needs for transparency (law enforcement, public health, anti-fraud). Interoperability is another challenge: privacy infrastructure must work across platforms, jurisdictions, and technical stacks, requiring coordination that no single company or protocol can achieve alone. The realistic trajectory involves gradual deployment of privacy-preserving credentials in specific verticals (healthcare, finance, government ID) over the next 5-10 years, rather than a universal privacy control layer emerging all at once.

3.4.4. De-escalatory Self-Defense Mediation Tool

An AI-powered mediation platform designed to resolve conflicts through progressive strategies of win-win negotiation, restoration, and proportional consequence, aimed at reducing escalation in interpersonal, organizational, and international disputes.

d/acc: 16/20 Tech: N/A

State of the Art. AI-assisted conflict resolution remains nascent but has identifiable precursors. Online dispute resolution (ODR) platforms like Modria (acquired by Tyler Technologies in 2017)

and eBay's automated resolution system (handling over 60 million disputes per year) demonstrate that algorithmic mediation can work for structured commercial conflicts. The Consensus Building Institute and Harvard's Program on Negotiation have developed systematic mediation frameworks that could inform AI system design. NLP-based sentiment analysis and emotion detection have improved significantly – tools like Hume AI (founded 2021) can analyze vocal tone and facial expressions in real-time to assess emotional states during conversation. Chatbot-based mental health tools like Woebot and Wysa use CBT-informed dialogue strategies that share structural similarities with de-escalation techniques. However, no existing system combines multi-modal emotional intelligence, game-theoretic negotiation strategy, and progressive escalation protocols into a unified mediation platform.

Next Steps & Challenges. The most feasible near-term path involves building AI mediation tools for structured, lower-stakes disputes – workplace conflicts, landlord-tenant disagreements, or community disputes – where outcomes are measurable and training data can be collected, before attempting to scale to more complex interpersonal or international conflicts. The primary barrier is that effective mediation depends on deep contextual understanding of power dynamics, cultural norms, and emotional subtleties that current AI systems handle poorly; a mediator that misreads a situation risks escalating rather than de-escalating conflict. Liability and trust present additional challenges: parties in genuine conflict are unlikely to accept an AI mediator unless it has demonstrated reliability and its recommendations carry some form of institutional backing. The progressive escalation framework (win-win, then restoration, then proportional consequence) requires the system to make normative judgments about fairness and proportionality that are inherently contested and culturally variable, making it difficult to build a system that is perceived as neutral across diverse contexts.

3.4.5. Global Deliberation Coordinator (GDaaS)

A pioneering institution that combines traditional deliberative democratic processes with AI-powered tools to enable rapid, cost-effective, and accessible global decision-making on critical challenges.

d/acc: 15/20 Tech: 52/70

State of the Art. The GDC concept was prototyped at Foresight Institute's Existential Hope TAI Institution Design Hackathon in February 2024, where a team including Aviv Ovadya (newDemocracy), Joshua Tan (MetaGov), Evan Miyazono (Atlas Computing), and Bear Haon (Schmidt Futures) won shared second place. Existing deliberative democracy infrastructure includes Stanford's Deliberative Democracy Lab, the OECD's work on citizens' assemblies (over 600 documented processes across 30+ countries by 2024), and AI-augmented platforms like Polis (used by Taiwan's vTaiwan process) and the Collective Intelligence Project's Alignment Assemblies. The UN has experimented with AI-assisted multilingual deliberation tools, and organizations like newDemocracy and Sortition Foundation run national-scale citizens' assemblies. However, no institution currently offers on-demand global deliberation as a service with integrated AI translation, expertise validation, and representative sampling at planetary scale.

Next Steps & Challenges. The GDC team's proposed next steps include securing Advanced Market Commitments from organizations that would use GDaaS, running iterative pilot deliberations to de-risk the model, and integrating AI tools for scalable multilingual deliberation. The most viable near-term path is building on existing citizens' assembly methodology while adding AI-powered real-time translation and argument mapping. Key barriers include representative sampling at global scale (no reliable mechanism exists to draw statistically representative panels from 8 billion people), legitimacy and binding authority (even well-run deliberations are advisory unless governments commit to implementing outcomes), and the risk that AI mediation tools subtly shape deliberative outcomes through framing effects or summarization bias. Geopolitical fragmentation

and low trust in transnational institutions make adoption by major powers unlikely without demonstrated success at smaller scales first.

3.4.6. Habermas Machines

An AI system designed to support group deliberation by generating and refining collective statements, aiming to help diverse groups reach consensus through advanced language processing.
d/acc: 15/20 Tech: 42/70

State of the Art. The concept of AI-mediated deliberation became concrete with the publication of "Habermas Machine" research by a team at Google DeepMind led by Christopher Summersgill, published in Science in October 2024. The system used large language models to generate group statements that participants found more representative of their collective views than human-written alternatives, tested on contentious UK political topics (Brexit, NHS funding) with over 5,000 participants. This builds on earlier computational deliberation work, including pol.is (developed by Colin Megill, used in Taiwan's vTaiwan process since 2015 to identify consensus positions among thousands of participants on regulatory issues), Stanford's Deliberative Democracy Lab experiments with deliberative polling, and MIT's Collective Intelligence group work on crowdsourced policymaking. The Collective Intelligence Project (founded by Divya Siddarth and others in 2023) has been developing "alignment assemblies" that use AI tools to scale deliberation, and Anthropic collaborated with the Collective Intelligence Project on its Constitutional AI public input process in 2023. Taiwan's digital democracy infrastructure, championed by former Digital Minister Audrey Tang, remains the most mature real-world implementation of technology-assisted collective deliberation, with the Polis-based Join platform having processed input from millions of citizens on dozens of policy issues.

Next Steps & Challenges. The most viable path forward involves embedding AI deliberation tools into existing democratic institutions (citizen assemblies, public comment processes, municipal governance) rather than creating standalone platforms, since legitimacy in democratic contexts depends heavily on institutional trust and procedural integration. The key technical challenge is ensuring that LLM-mediated consensus does not subtly manipulate participants toward positions favored by the model's training biases or its operators – the DeepMind paper acknowledged this risk and implemented safeguards, but adversarial robustness in high-stakes political deliberation is an open problem. Institutional barriers include the reluctance of established political structures to cede agenda-setting or synthesis power to AI systems, concerns about digital divides excluding non-technical populations, and the fundamental tension between Habermasian ideal speech conditions (which assume equal power among participants) and the reality that AI mediators introduce a powerful, opaque intermediary. The most promising near-term applications are in lower-stakes contexts where consensus-finding has clear value – corporate governance, community planning processes, international standards bodies, and multi-stakeholder negotiations – where successful deployments could build the track record needed for eventual adoption in formal democratic institutions. A critical unresolved question is whether AI-mediated consensus genuinely reflects collective wisdom or merely produces what the philosopher Cass Sunstein calls "enclave deliberation" artifacts, where apparent agreement masks suppressed minority viewpoints.

3.5. Markets & Incentive Systems

3.5.1. Reputational Markets

A decentralized global coordination mechanism that uses prediction markets to assess and incentivize responsible behavior across complex supply chains. It creates a dynamic reputation scoring system that encourages collective accountability for long-term human values.

d/acc: 17/20 Tech: 51/70

State of the Art. Prediction markets have demonstrated remarkable accuracy in aggregating distributed information, with Polymarket processing over $1 billion in volume during the 2024 U.S. election cycle and consistently outperforming polls. Metaculus and Good Judgment Open have built track records in calibrated forecasting across domains including technology, geopolitics, and biosecurity. In supply chain accountability, ESG rating agencies like MSCI, Sustainalytics, and CDP provide reputation-adjacent scoring, though these have faced criticism for inconsistency – a 2022 MIT study found correlations as low as 0.38 between major ESG raters on the same companies. Blockchain-based reputation systems exist in narrow domains: Gitcoin Passport aggregates identity signals for sybil resistance, and platforms like Lens Protocol and Farcaster experiment with portable on-chain reputation. The EigenTrust algorithm (2003) and its descendants provide theoretical foundations for decentralized trust computation, while projects like Karma3Labs are building reputation infrastructure for Web3.

Next Steps & Challenges. The most promising path combines prediction market mechanics with verifiable supply chain data – using oracles to bring real-world attestations on-chain and allowing stakeholders to stake on the veracity of corporate responsibility claims. The key barrier is the oracle problem: connecting prediction markets to real-world outcomes requires trusted data sources, and supply chain data is notoriously fragmented, self-reported, and unverifiable across global networks. Regulatory uncertainty compounds this – prediction markets remain legally restricted in most jurisdictions (the CFTC's 2024 actions against certain Polymarket-style products demonstrate ongoing friction), and extending them to corporate reputation raises securities law concerns. Gaming and Goodhart's Law pose fundamental design challenges: any reputation metric that becomes a target for optimization will be manipulated, requiring adversarial robustness that no current system has fully solved. Finally, adoption requires that reputation scores carry real economic consequences (e.g., affecting chip access or financing terms), which demands buy-in from powerful incumbent institutions that currently benefit from opacity.

3.5.2. Prediction Markets as Decision Support Systems

An advanced institutional technology for transforming crowd intelligence into actionable decision-making tools for organizations, moving beyond current crypto-based betting platforms to create sophisticated advisory systems.

d/acc: 16/20 Tech: 44/70

State of the Art. Prediction markets experienced a breakthrough moment in 2024 when Polymarket processed over $3.5 billion in trading volume on the US presidential election, demonstrating that market-based forecasting can outperform polls and expert models in real time. Kalshi, the first CFTC-regulated prediction market exchange in the US, won a landmark legal battle in 2024 allowing trading on election outcomes, establishing crucial regulatory precedent. Metaculus, co-founded by Anthony Aguirre, operates as a non-monetary forecasting platform with over 20,000 questions and a rigorous scoring system, while Manifold Markets allows anyone to create prediction markets with play money. Robin Hanson, who pioneered the concept in the 1990s and coined the term 'futarchy' for governance by prediction markets, has noted that while these platforms succeed as betting venues, they have not yet been adopted as organizational decision support tools – the use case he considers most valuable. IARPA's ACE (Aggregative Contingent Estimation) program

demonstrated that prediction markets consistently outperformed intelligence analysts, but this has not translated into routine institutional adoption.

Next Steps & Challenges. The critical gap, as Hanson identifies, is moving from 'platforms where people bet on topics they find interesting' to 'systems that support specific organizational decisions' – which requires building applications that integrate market signals into existing decision workflows with demonstrated track records. The key barriers are institutional rather than technical: most organizations resist transparent, crowd-driven forecasting because it creates accountability that executives and analysts would rather avoid, and internal prediction markets (attempted by companies like Google and Ford) often stagnate because employees fear reputational risk from betting against management's preferred narratives. The manipulation problem remains unsolved – in thin markets, well-resourced actors can move prices to influence decisions rather than inform them, and designing robust liquidity and incentive mechanisms that resist strategic trading is an open research problem. Regulatory fragmentation is another barrier: prediction markets sit uncomfortably between gambling regulation, securities law, and information markets, with different jurisdictions taking contradictory approaches. Transforming prediction markets from speculative entertainment into the 'trillions of dollars in decision value' Hanson envisions requires patient, unglamorous work on specific organizational applications – exactly the kind of work that attracts the least venture capital interest.

3.5.3. Futarchy (Governance by Prediction Markets)

A novel governance system where societal values are democratically determined, but specific policy implementations are selected through prediction markets that objectively forecast outcomes.

d/acc: 16/20 Tech: N/A

State of the Art. Futarchy was proposed by economist Robin Hanson in 2000, and while no government has adopted it, key components are being actively tested. Polymarket emerged as the dominant prediction market platform by 2024, handling over $1B in trading volume around the 2024 US presidential election, demonstrating that large-scale, liquid prediction markets on real-world events are technically and commercially viable. Metaculus, a community forecasting platform, has over 20,000 questions and has shown strong calibration on geopolitical, scientific, and policy questions. Within crypto governance, futarchic elements have been tested: Gnosis built conditional token frameworks enabling decision markets, and the MetaDAO project on Solana (launched 2023) implemented an explicit futarchy-based governance system where token holders vote on proposals via conditional markets on the token price. The UK's Foundations for Evidence-Based Policymaking and the US intelligence community's IARPA forecasting tournaments (ACE, HFC programs) have validated that structured prediction aggregation outperforms traditional expert judgment on many questions.

Next Steps & Challenges. The most viable path forward is incremental adoption within organizations and DAOs before attempting government-level implementation – MetaDAO's model could be replicated across crypto governance, corporate strategy decisions, and municipal policy experiments. The core barrier is political: elected officials and bureaucracies have no incentive to adopt a system that could replace their decision-making authority with market signals, and voters may find it illegitimate to delegate policy choices to traders. Technical challenges include the thin-market problem (most policy questions are too specific to attract sufficient liquidity for reliable price discovery), the difficulty of defining clear measurable metrics for complex policy outcomes, and vulnerability to market manipulation by well-funded actors on low-liquidity markets. There are also deep philosophical objections: prediction markets optimize for expected outcomes under a particular metric, but governance involves value trade-offs, distributional concerns, and rights protections that resist reduction to a single measurable variable. Realistic near-term progress looks like more DAOs and organizations experimenting with decision markets, academic validation of

futarchic mechanisms in controlled settings, and perhaps small-scale municipal pilots, but adoption by nation-state governments remains highly unlikely within the next decade.

3.6. Ethics & Moral Expansion

3.6.1. Expanded Moral Circle Technologies

Neurotechnological and communication systems enabling dramatically enhanced empathy, understanding, and moral perception across different forms of consciousness.

d/acc: 16/20 Tech: N/A

State of the Art. Research relevant to expanding moral consideration across different forms of consciousness is distributed across several emerging fields. In animal communication, the Earth Species Project (founded 2017) uses machine learning to decode animal vocalizations, and Project CETI (Cetacean Translation Initiative, launched 2020 with funding from Audacious Project) is applying NLP techniques to sperm whale codas with initial results published in 2024 showing contextual structure in whale communication. In the neuroscience of empathy, psilocybin research at Johns Hopkins Center for Psychedelic and Consciousness Research (opened 2020) and Imperial College London's Centre for Psychedelic Research has demonstrated that psychedelic-assisted therapy can increase measures of nature-connectedness and emotional empathy, with several Phase II trials completed by 2024. On the AI consciousness and moral status front, the debate intensified in 2024 following claims about LLM sentience; philosophers like Eric Schwitzgebel and Robert Long published frameworks for assessing AI moral patienthood, while the New York Declaration on Animal Consciousness (2024), signed by dozens of researchers, argued for taking seriously the possibility of consciousness in a wide range of animals including invertebrates. Neurotechnology for cross-species empathy remains nascent: hyperscanning studies (simultaneous brain imaging of interacting individuals) have shown neural synchronization correlates with empathy between humans, but extending this to cross-species or human-AI interaction is largely conceptual.

Next Steps & Challenges. The most viable near-term path involves converging animal communication AI with immersive media (VR experiences from the animal's perspective) and expanded legal personhood frameworks, rather than direct neural empathy links which face fundamental neuroscience barriers. A key technical challenge is the "translation problem" – even if we can decode animal communication signals, we may lack the conceptual framework to translate experiences across radically different nervous systems (what Thomas Nagel called the "what is it like" problem in his 1974 essay on bat consciousness). Institutionally, the legal expansion of moral consideration is proceeding slowly but concretely: New Zealand granted the Whanganui River legal personhood in 2017, Ecuador enshrined rights of nature in its constitution in 2008, and several jurisdictions have upgraded animal welfare laws, but extending enforceable moral status to AI systems or non-charismatic organisms faces deep political resistance. The coordination barrier is that expanding the moral circle requires simultaneous advances in consciousness science (to determine who/what merits moral status), communication technology (to understand non-human perspectives), and governance innovation (to encode these insights into enforceable frameworks) – and these fields currently operate in largely separate academic and institutional silos. Progress is most likely through flagship projects that combine several of these elements, such as using AI-decoded whale communication to build public support for cetacean legal personhood, creating a concrete demonstration of the pipeline from scientific understanding to moral-legal recognition.

3.6.2. Moral Trade Civilization

A future societal model where ethical preferences can be systematically traded, allowing diverse moral perspectives to achieve mutually beneficial outcomes and maximize collective value alignment.

d/acc: 16/20 Tech: N/A

State of the Art. The concept of moral trade was formalized by philosopher Toby Ord in his 2015 paper "Moral Trade," which proposed that individuals with different moral views could make mutually beneficial exchanges of moral commitments – for example, a utilitarian might agree to

support deontological constraints valued by their counterpart in exchange for the counterpart supporting utilitarian causes. The idea has been explored primarily within the effective altruism community: the platform Moral Trade (moraltrade.org) was briefly operational as a proof-of-concept where users could propose and accept moral trades, though it never achieved significant scale or sustained activity. Related mechanisms exist in more established forms: carbon offset markets allow trading environmental commitments, corporate ESG frameworks represent implicit moral trade between stakeholders, and political logrolling (legislators trading votes across issues) is a longstanding form of moral preference exchange. The academic literature on moral uncertainty (work by Will MacAskill, Krister Bykvist, and Toby Ord, culminating in the 2020 book "Moral Uncertainty") provides theoretical foundations for how rational agents should act when uncertain about which moral framework is correct, which directly informs how moral trades could be valued. Prediction markets (Polymarket, Manifold Markets) and quadratic funding mechanisms (Gitcoin) represent adjacent experiments in aggregating and operationalizing diverse preferences, though these deal with empirical beliefs and public goods rather than moral commitments per se.

Next Steps & Challenges. The most viable path forward involves building moral trade mechanisms into existing coordination infrastructure – integrating moral preference expression into governance platforms, expanding prediction markets to include normative questions, and developing formal frameworks for verifying moral commitment fulfillment. The fundamental barrier is that moral commitments are far harder to specify, verify, and enforce than economic contracts: how do you confirm that someone has genuinely adopted a more empathetic stance toward animals, or verify that a policy change truly reflects the traded moral commitment rather than political convenience? This verification problem is compounded by the deep philosophical objection that moral beliefs should be held based on reasons and evidence, not traded as commodities – many ethicists argue that treating moral commitments as fungible undermines the very nature of moral reasoning and could lead to a corrosive instrumentalization of ethics. Technically, smart contracts and blockchain-based commitment devices could provide partial infrastructure for transparent moral trading, but the encoding of complex moral commitments into executable code faces the same specification problems that plague AI alignment. The most realistic near-term progress likely comes from expanding structured negotiation frameworks in multi-stakeholder governance (climate negotiations, bioethics committees, AI governance bodies) where diverse moral perspectives already need to be reconciled, gradually building institutional experience with formal moral preference exchange before attempting the more ambitious vision of a civilization-scale moral trading system.

3.7. AI & Human Agency

3.7.1. Fiduciary AI Assistance

A system of AI assistants designed to be fundamentally loyal to individual human users, helping them navigate complex problems and daily life while respecting their goals and interests.

d/acc: 16/20 Tech: N/A

State of the Art. The concept of AI assistants with fiduciary duty to individual users, championed by Anthony Aguirre of the Future of Life Institute and Metaculus, has gained traction as a counterpoint to the dominant ad-supported AI model. Current AI assistants (ChatGPT, Claude, Gemini) serve users but are optimized for platform engagement and trained on corporate objectives rather than individual welfare maximization. The closest existing implementations are in regulated financial advice (robo-advisors like Betterment and Wealthfront operate under fiduciary standards) and in personal AI assistants being developed by startups like Inflection AI (Pi), though these lack the legal or architectural guarantees of true fiduciary duty. The AI safety research community has explored related concepts under the banner of 'tool AI' versus 'agent AI,' with Anthropic's Constitutional AI and instruction-following approaches representing partial technical solutions to alignment with user intent. Notably, the EU AI Act (effective 2024) and proposed US legislation have begun to frame AI systems in terms of duties to users, though none yet mandate fiduciary-level loyalty.

Next Steps & Challenges. The most viable path forward combines technical and legal innovation: architecturally, encoding user welfare as a provable optimization target through techniques like Constitutional AI, RLHF with user-specific reward models, and on-device personalization that keeps sensitive preference data local. The fundamental barriers are both economic and technical. Economically, the current AI business model depends on centralized data aggregation and platform control; a truly fiduciary AI that never manipulates or extracts value from its user is hard to monetize at scale, which is why subscription-based models (as attempted by Inflection AI before its effective acquisition by Microsoft in 2024) have struggled. Technically, robust individual preference learning requires solving hard problems in value alignment – an AI must model a user's long-term welfare, which may conflict with their stated short-term desires, creating a paternalism dilemma that no current alignment technique adequately resolves. The legal infrastructure for AI fiduciary duty is essentially nonexistent: defining what it means for software to have a legally enforceable duty of loyalty requires entirely new regulatory frameworks that no jurisdiction has yet drafted.

3.7.2. EgoLets (Personal AI Assistants)

Personalized AI systems that capture individual thinking patterns and decision-making styles, functioning as specialized "little versions of your ego" across different life domains.

d/acc: 15/20 Tech: 48/70

State of the Art. Personal AI assistants have advanced rapidly since 2023, with systems like OpenAI's ChatGPT (with memory features launched in 2024), Google's Gemini, and Anthropic's Claude offering increasingly personalized interactions that learn user preferences over time. Apple Intelligence (announced 2024) integrates personal context from on-device data across apps. Startups like Inflection AI (Pi), Replika, and Character.AI have explored personality-modeling and emotional rapport. Microsoft's Copilot system embeds AI across productivity tools with personalization layers. However, none of these systems create genuine cognitive portraits that mirror an individual's reasoning patterns across domains – they personalize outputs based on conversation history rather than constructing a model of how a specific person thinks, decides, and weighs tradeoffs, which is the core EgoLet concept articulated by Ken Liu.

Next Steps & Challenges. Building true EgoLets requires advances in three areas: personal knowledge graph construction (mapping an individual's beliefs, decision heuristics, and domain

expertise from their digital footprint), privacy-preserving fine-tuning (training personalized models without exposing sensitive data to centralized servers), and cognitive fidelity evaluation (measuring whether an AI's decisions actually match what the person would decide). Federated learning and on-device model adaptation offer technical paths, but the fundamental challenge is that human cognition is contextual, contradictory, and evolving – a static 'portrait' quickly becomes stale or reductive. Institutional barriers include the absence of data portability standards that would let users aggregate their decision history across platforms, and deep unresolved questions about identity, consent, and liability when an AI acts as someone's cognitive proxy.

3.7.3. Empathetic Neuro-AI Emotional Coaching System

An AI-driven emotional support ecosystem that provides personalized, real-time therapeutic interventions through neural signal analysis and adaptive AI companions that prevent emotional crises and promote healing.

d/acc: 13/20 Tech: 48/70

State of the Art. The building blocks for neuro-AI emotional coaching exist across several distinct fields that have not yet been integrated into a unified system. In affective computing, companies like Affectiva (acquired by Smart Eye in 2021) and Hume AI (founded 2021, raised $50M by 2024) use facial expression analysis, voice prosody, and language patterns to detect emotional states, with Hume's Empathic Voice Interface (EVI) launching in 2024 as an API for emotion-aware conversational AI. Consumer-grade EEG devices from companies like Muse (InteraXon), Emotiv, and Neurable can detect broad emotional states (stress, relaxation, focus) from brainwave patterns, though with limited granularity and significant noise compared to clinical-grade systems. On the therapeutic AI side, Woebot (FDA breakthrough designation 2023 for postpartum depression), Wysa, and similar chatbot-based mental health tools have demonstrated modest clinical efficacy in randomized controlled trials for anxiety and depression, primarily using CBT-based text interactions without any neural signal input. Real-time neurofeedback therapy – where patients learn to regulate brain activity patterns associated with emotional states – has been practiced clinically since the 2000s, with recent fMRI neurofeedback studies (2020-2024) showing promise for PTSD and depression, but these require expensive clinical equipment and controlled settings.

Next Steps & Challenges. The most viable near-term path is combining consumer wearable biosignals (heart rate variability, electrodermal activity, sleep patterns) with large language model-based coaching, rather than direct neural signal reading, since non-invasive EEG remains too noisy for fine-grained real-time emotion detection outside controlled lab settings. The key technical barrier is the "affective gap" – the difference between what physiological signals can reliably indicate (arousal, valence at a coarse level) and the rich, context-dependent emotional states that effective therapeutic intervention requires. Regulatory and ethical barriers are substantial: the FDA would need to evaluate any system making therapeutic claims, real-time emotional surveillance raises profound privacy concerns (the EU AI Act classifies emotion recognition in certain contexts as high-risk), and there is legitimate debate about whether AI emotional coaching could undermine human emotional autonomy or create unhealthy dependency. The most productive path forward likely involves clinical validation of specific narrow use cases (crisis detection for suicidal ideation, stress management for burnout, emotional regulation coaching for anxiety disorders) where the benefit-risk calculus is clearest, paired with robust data governance frameworks. Integration with neural signals at the level described in the hyper-entity vision would require BCI hardware advances that are likely 10-15 years away for consumer applications.

3.7.4. Lifelong AI Guardians

Privacy-preserving AI systems that monitor and support children's holistic development, providing personalized guidance across multiple life domains and connecting different stakeholders through

intelligent, empathetic monitoring.
d/acc: 13/20 Tech: 49/70

State of the Art. Elements of AI-driven child development monitoring exist in fragmented form across education technology, pediatric health, and parental surveillance products, but no integrated "lifelong guardian" system exists. In education, adaptive learning platforms like Khan Academy's Khanmigo (launched 2023, powered by GPT-4) provide personalized tutoring, while companies like Century Tech and Squirrel AI use machine learning to adapt curricula to individual student performance patterns. Pediatric health monitoring includes apps like BabySparks and CDC's Milestone Tracker for developmental screening, and research groups at MIT Media Lab (Personal Robots Group) and Yale's Social Robotics Lab have studied AI companions for children's social-emotional development, including the Tega and Jibo platforms. On the surveillance side, parental monitoring tools like Bark, Qustodio, and Apple's Screen Time track children's digital activity for safety purposes, though these are reactive filtering tools rather than developmental support systems. China's Social Credit System experiments and some school districts' adoption of AI proctoring and attention-monitoring tools (like those from Squirrel AI in China) represent the most aggressive integration of AI monitoring into children's lives, generating significant backlash and privacy concerns.

Next Steps & Challenges. The most viable path forward involves longitudinal research partnerships between developmental psychologists, AI researchers, and school systems to create evidence-based developmental AI tools with strong privacy guarantees, likely starting with specific narrow domains (early literacy support, social-emotional learning assessment) rather than attempting holistic monitoring. The primary barrier is not technical but ethical and regulatory: COPPA in the U.S., GDPR's special protections for children's data in Europe, and the UN Convention on the Rights of the Child all impose strict limits on data collection from minors, and the societal appetite for comprehensive AI monitoring of children is low following controversies over EdTech surveillance during COVID-era remote learning. Privacy-preserving techniques like federated learning, differential privacy, and on-device processing could partially address data concerns, but the fundamental tension remains between the comprehensive data collection needed for effective personalized guidance and children's rights to privacy, autonomy, and the ability to make mistakes without permanent records. The coordination challenge involves aligning the interests of parents (who want safety and opportunity), educators (who want learning outcomes), healthcare providers (who want developmental health), and children themselves (whose preferences and autonomy must be respected and evolve over time) – each group currently uses separate, non-interoperable systems. Any viable system would need to demonstrate clearly superior developmental outcomes in rigorous longitudinal studies while maintaining genuine informed consent mechanisms, a combination that would likely require 10-15 years of careful research and piloting.

3.8. Interfaces & Augmentation

3.8.1. Digital Mind Governance Systems

A future institutional and ethical framework for managing, protecting, and regulating digital minds as potential moral patients with rights, voting capabilities, and social standing.

d/acc: 16/20 Tech: 53/70

State of the Art. The intellectual groundwork for digital mind governance is advancing primarily through academic philosophy and AI safety research. Nick Bostrom and Carl Shulman's 2024 paper 'Propositions Concerning Digital Minds and Society' provides the most comprehensive framework to date, addressing voting rights, moral status gradation, and the problem of easy replication undermining democratic one-person-one-vote principles. Joe Carlsmith at Open Philanthropy has published extensively on the moral patienthood of AI systems, and the field intersects with growing work on AI consciousness detection – notably by researchers like Robert Long at the Center for AI Safety and Jonathan Birch at the London School of Economics, whose 2024 report for the UK government proposed precautionary frameworks for AI sentience. Anthropic, Google DeepMind, and other labs have begun internal discussions about model welfare, though no organization has implemented formal protections.

Next Steps & Challenges. The most viable near-term steps are developing rigorous, empirically grounded tests for machine consciousness and establishing precautionary institutional review processes at AI labs for the treatment of potentially sentient systems. The fundamental barrier is that we lack any scientific consensus on what consciousness is, let alone how to detect it in artificial substrates – current proposals (such as Global Workspace Theory or Integrated Information Theory markers) remain deeply contested among philosophers and neuroscientists. The governance challenge compounds the epistemic one: if digital minds can be cheaply copied and deleted, existing legal concepts of personhood, property, and citizenship break down entirely, and no legal system on Earth has begun to draft legislation for non-biological moral patients. Additionally, there is an acute coordination problem where the first jurisdiction to grant AI rights could face massive economic and strategic consequences, creating strong incentives for delay.

3.8.2. Translation Language Models (TLMs)

A multipurpose linguistic technology that enables rapid translation and understanding across languages, dialects, and technical domains. These models aim to democratize knowledge work by breaking down communication barriers.

d/acc: 16/20 Tech: 53/70

State of the Art. Machine translation has been transformed by large language models, with current systems far exceeding the capabilities of earlier statistical and rule-based approaches. Meta's NLLB (No Language Left Behind) project, released in 2022, provides translation across 200 languages including many low-resource languages, while Google Translate now covers 133 languages and its Gemini model handles real-time conversational translation. SeamlessM4T, released by Meta in 2023, enables speech-to-speech, speech-to-text, and text-to-speech translation across nearly 100 languages in a single model. DeepL has established itself as the quality leader for European language pairs, consistently outperforming Google in professional contexts. The Foresight Institute's worldbuilding scenario 'La langue de la prevoyance' envisions TLMs by 2035 using citizen-owned training databases and DAO governance for model development, emphasizing that translation technology could democratize knowledge work for those previously shut out by language barriers – a vision partially realized by the ability of GPT-4, Claude, and Gemini to handle technical translation across domains like law, medicine, and engineering.

Next Steps & Challenges. The most viable paths forward include building real-time, low-latency translation systems embedded in communication platforms (as Google is attempting with its Pixel

earbuds and Meta with its smart glasses), expanding coverage to the thousands of languages with insufficient parallel training corpora, and developing domain-specific models for high-stakes technical translation where errors carry legal or medical consequences. The key barriers are threefold: first, truly capturing cultural-semantic nuance rather than just linguistic equivalence remains extremely difficult, as idiomatic expressions, humor, and contextual meaning still produce frequent errors that native speakers immediately notice; second, low-resource languages (spoken by hundreds of millions of people) lack the training data needed for high-quality models, and synthetic data generation has not yet bridged this gap; third, the governance question of who controls translation infrastructure – whether it remains concentrated in a few Silicon Valley companies or becomes the open, community-governed resource envisioned in the DAO model – has significant implications for linguistic diversity and cultural preservation. The computational cost of running high-quality translation models remains prohibitive for deployment in the low-connectivity, low-resource settings where language barriers cause the most economic harm.

3.8.3. Digital Twin Ecosystem

A comprehensive system of real-time digital representations of communities and ecosystems that enable collective decision-making, simulation, and shared responsibility. These digital twins provide a new infrastructure for understanding and managing complex social and ecological systems.

d/acc: 15/20 Tech: 55/70

State of the Art. Digital twin technology has matured significantly in industrial settings, with companies like Siemens (Xcelerator platform), NVIDIA (Omniverse), and Microsoft (Azure Digital Twins) offering platforms for creating real-time virtual replicas of physical systems. City-scale digital twins are operational in Singapore (Virtual Singapore, launched 2018), Helsinki, and Shanghai, primarily for urban planning and infrastructure management. The EU's Destination Earth initiative, launched in 2022, is building a digital twin of the entire planet for climate and environmental simulation. However, these implementations remain siloed – industrial twins model factories, urban twins model traffic and buildings – and none yet integrate social, ecological, and governance systems into a single coherent ecosystem for collective decision-making as envisioned by the hyper-entity concept.

Next Steps & Challenges. The most viable path forward involves extending existing city-scale digital twins to incorporate social and ecological data streams, likely building on IoT sensor networks and satellite imagery infrastructure already being deployed. The key barriers are data interoperability (no universal standard exists for integrating heterogeneous data from social, ecological, and infrastructural sources), computational cost (real-time simulation of complex socio-ecological systems at community scale requires orders of magnitude more compute than current industrial twins), and governance design (who controls the twin, who has access, and how simulation outputs translate into legitimate collective decisions are unsolved institutional questions). Privacy concerns around continuous community monitoring and the risk of 'model capture' – where the twin's assumptions invisibly shape decisions – add further complexity that technical solutions alone cannot resolve.

3.8.4. Immune-Computer Interface

A transformative technological system that enables direct, high-bandwidth communication between human immune systems and computational technologies, allowing for dynamic biological monitoring, intervention, and enhancement.

d/acc: 13/20 Tech: 58/70

State of the Art. The concept of an immune-computer interface was articulated by Hannu Rajaniemi, co-founder and CEO of Helix Nanotechnologies, who argued that expanding human possibilities requires direct computational access to the immune system. Current precursors include continuous glucose monitors (Dexcom, Abbott FreeStyle Libre) that demonstrate wearable

biosensing, and implantable devices like Profusa's tissue-integrated biosensors that can track body chemistry for months. CAR-T cell therapy (first FDA-approved 2017) demonstrates that immune cells can be engineered with synthetic receptors, effectively 'programming' immune responses. Research groups at MIT (Sangeeta Bhatia's lab), Stanford (immunoengineering programs), and the Wyss Institute are developing nanoparticle-based sensors and synthetic biology tools that interface with immune signaling pathways. However, no system exists that provides continuous, high-bandwidth, bidirectional communication between the immune system and external computation – current technologies are either read-only (biosensors), write-only (immunotherapies), or operate at extremely low data rates.

Next Steps & Challenges. The most viable near-term path involves combining advances in implantable nanosensors (for reading immune states) with engineered immune cells carrying synthetic signaling circuits (for writing computational instructions back to the immune system). Key technical barriers include biocompatibility over long time periods (the immune system actively attacks foreign objects, creating a fundamental paradox for any device trying to interface with it), signal bandwidth (immune communication uses slow molecular diffusion rather than electrical signals, limiting data rates), and the staggering complexity of immune system dynamics (hundreds of cell types, thousands of signaling molecules, context-dependent responses that defy simple computational models). Regulatory pathways are also unclear – such a device would straddle drug, device, and biological product categories simultaneously – and the safety risks of computationally modulating immune responses in real-time (autoimmunity, immunosuppression) are severe enough that development timelines likely extend 15-25 years even under optimistic assumptions.

3.8.5. Human Superintelligence via Brain-Computer Interfaces (BCI)

A technological system that enhances human cognitive capabilities through direct neural interfaces, enabling humans to compete with and potentially transcend artificial intelligence. This would fundamentally expand human potential for understanding and exploring the universe.

d/acc: 12/20 Tech: 60/70

State of the Art. The BCI field has seen rapid progress since 2020. Neuralink implanted its first human patient (Noland Arbaugh) in January 2024, demonstrating thought-controlled cursor movement via the N1 chip's 1,024 electrodes, and implanted additional patients later that year. Synchron's Stentrode, a minimally invasive endovascular BCI, received FDA breakthrough device designation in 2020 and completed its first U.S. human implant in 2022, with ongoing COMMAND clinical trials showing patients controlling digital devices. Precision Neuroscience debuted a thin-film electrode array (Layer 7) in 2023 that can be placed on the brain's surface without penetrating tissue, recording from over 4,000 electrodes simultaneously. BrainGate's long-running clinical research has enabled paralyzed individuals to type and control robotic arms since the mid-2010s, with recent publications demonstrating high-performance speech decoding (Stanford, 2023) achieving over 60 words per minute from neural signals. However, all current systems are therapeutic – restoring lost function for people with paralysis – and none yet enhance cognition beyond baseline human capability.

Next Steps & Challenges. The path from therapeutic restoration to genuine cognitive enhancement faces several compounding barriers. First, current implants record from tens of thousands of neurons at best, whereas enhancing cognition would likely require interfacing with millions or billions of neurons across distributed brain networks – a gap of several orders of magnitude in electrode density and coverage. Second, we lack a sufficient computational neuroscience understanding of how higher cognition (reasoning, creativity, memory formation) is encoded at the neural level, making it unclear what signals to read or write even with perfect hardware. Third, the regulatory pathway for enhancement in healthy individuals is essentially nonexistent; the FDA and equivalent bodies approve devices to treat disease, not augment the healthy, which could delay deployment by decades.

The most viable near-term path runs through progressive therapeutic applications – memory prostheses for Alzheimer's patients (DARPA's RAM program, Kernel's work), mood regulation for treatment-resistant depression (already explored via deep brain stimulation), and high-bandwidth communication for locked-in patients – each of which builds the neuroscience knowledge and surgical infrastructure needed for eventual enhancement. Realistic timelines for anything approaching cognitive enhancement in healthy humans are likely 20-30+ years out, assuming sustained funding and major breakthroughs in both neuroscience and biocompatible materials.

3.8.6. Mind Uploading Infrastructure

Technological systems enabling gradual human consciousness transfer to digital platforms, with complex social and infrastructural considerations for maintaining uploaded minds.

d/acc: 8/20 Tech: 56/70

State of the Art. Mind uploading remains firmly in the domain of foundational research, with no demonstrations of consciousness transfer or whole-brain emulation to date. The most concrete progress is in connectomics: in 2024, a Harvard-Google team published the first nanoscale-resolution map of a cubic millimeter of human brain tissue, containing roughly 57,000 cells and 150 million synapses – representing approximately one-millionth of the full human brain. The OpenWorm project has mapped the complete connectome of C. elegans (302 neurons) and produced partial functional simulations, while the full Drosophila (fruit fly) connectome (140,000 neurons) was completed in 2024 by the FlyWire consortium. On the preservation side, Nectome (spun out of MIT) developed aldehyde-stabilized cryopreservation that won the Brain Preservation Foundation's prize in 2018 for preserving a pig brain's connectome at the nanometer scale, though the company pivoted away from human preservation following ethical controversy. The computational requirements estimated for whole-brain emulation range from 10^18 to 10^25 FLOPS depending on the level of simulation fidelity required, which is at or beyond current supercomputing capabilities even before accounting for the memory and I/O demands.

Next Steps & Challenges. The most viable path forward involves three parallel tracks that must all converge: dramatically faster and cheaper connectomic imaging, vastly better computational neuroscience models, and resolution of profound philosophical and ethical questions. Current electron microscopy techniques for connectomics take years to image a cubic millimeter; scaling to a full human brain (1,200 cubic centimeters) would require roughly a million-fold improvement in throughput, which even optimistic projections place decades away. Even with a complete structural map, we do not know whether the connectome alone is sufficient for consciousness – neuromodulatory dynamics, glial cell interactions, epigenetic states, and potentially quantum effects may all play essential roles, meaning a wiring diagram could be necessary but not sufficient. On the institutional side, there is no regulatory framework for determining whether a digital entity is conscious or has rights, no legal definition of digital personhood, and deep unresolved debates in philosophy of mind about whether substrate-independent consciousness is even coherent. The most productive near-term work is in improving brain preservation techniques, advancing connectomics automation (expansion microscopy, AI-assisted segmentation), and developing increasingly detailed simulations of small neural circuits to test what level of biological detail is actually required for functional replication.

3.8.7. Whole Brain Emulation

A technological approach to creating digital replicas of human brains through advanced scanning and simulation technologies. This would potentially enable preservation of human cognitive capabilities in digital form.

d/acc: 8/20 Tech: N/A

State of the Art. Whole brain emulation (WBE) remains a long-term research goal, but foundational neuroscience has achieved significant milestones. In 2024, a Harvard-Google collaboration published in Science a nanoscale connectome map of a cubic millimeter of human temporal cortex,

containing approximately 57,000 neurons and 150 million synapses – representing roughly one-millionth of the full human brain. The complete connectome of C. elegans (302 neurons) has been mapped since 1986, and the full connectome of the Drosophila (fruit fly) brain (140,000 neurons) was completed by the FlyWire consortium and published in Nature in 2024. The OpenWorm project has produced a partial computational simulation of C. elegans but has not achieved full behavioral emulation. On the scanning side, serial-section electron microscopy and expansion microscopy have improved dramatically, and companies like Nectome (founded 2016) have explored aldehyde-stabilized cryopreservation for brain preservation, though Nectome pivoted away from consumer-facing preservation services amid ethical controversy. Computational neuroscience simulations, such as the Blue Brain Project's detailed cortical column models, can simulate tens of thousands of neurons with biophysical detail but remain far from whole-brain scale.

Next Steps & Challenges. The most viable near-term path is continued scaling of connectomics – mapping progressively larger brain volumes at synaptic resolution, improving automated segmentation with AI (tools like Google's Flood-Filling Networks and CAVE infrastructure), and building increasingly detailed computational models of well-characterized brain circuits. The fundamental barriers are staggering in scale: the human brain contains roughly 86 billion neurons and 100-500 trillion synapses, and a full connectome at nanometer resolution would generate exabytes of imaging data requiring years of scanning and petaflops of computational reconstruction. Beyond structure, it is unclear whether connectome data alone is sufficient for emulation – synaptic weights, neuromodulatory dynamics, glial cell contributions, gene expression states, and intracellular signaling may all be necessary, each adding orders of magnitude to the required data. The computational requirements for simulating a full brain in real-time, even with a known connectome, likely exceed current and near-future supercomputing capabilities by many orders of magnitude. Most serious researchers in the field estimate whole brain emulation of a human is at least 50-100 years away if achievable at all, though intermediate milestones like full emulation of small invertebrate nervous systems (such as C. elegans or Drosophila) may be achievable within 10-20 years.

3.9. Science & Discovery

3.9.1. Automated Scientific Publishing Ecosystem for Machine Consumers

A reimagined scientific publishing system designed primarily for AI systems to consume, process, and contribute to scientific knowledge, with radically different latency, format, and peer review mechanisms.

d/acc: 18/20 Tech: 60/70

State of the Art. Several building blocks for machine-consumable science already exist. The JATS XML standard and initiatives like Semantic Scholar (backed by the Allen Institute for AI) have created structured, machine-readable representations of millions of papers, while preprint servers like arXiv and bioRxiv have dramatically reduced publication latency. In 2023-2024, tools like Elicit, Consensus, and ScholarAI began using LLMs to parse, summarize, and synthesize scientific literature at scale, demonstrating that AI systems can already act as primary consumers of research. The FAIR data principles (Findable, Accessible, Interoperable, Reusable), formalized in 2016 and now adopted by major funders including the NIH and European Commission, have pushed datasets toward machine readability. Meanwhile, initiatives like ResearchHub and DeSci projects on Ethereum are experimenting with tokenized peer review and blockchain-verified research artifacts, though none yet constitute a full publishing ecosystem designed for machine consumers.

Next Steps & Challenges. The most viable near-term path involves extending existing infrastructure: augmenting preprint servers with mandatory structured metadata, machine-readable claims, and linked datasets, while building AI-driven verification layers on top. The key barriers are institutional rather than technical – Elsevier, Springer Nature, and Wiley control the majority of published science and have little incentive to restructure for machine consumption, as their business model depends on human-oriented access paywalls. A deeper challenge is creating trustworthy automated peer review: current LLMs can check statistical consistency and flag methodological issues (as demonstrated by tools like Statcheck), but lack the domain expertise and judgment to fully replace human reviewers. Building consensus around new metadata standards that capture claims, evidence strength, and provenance at a granular level requires coordination across thousands of journals and institutions, a process that historically takes decades even when the technical solutions are ready.

3.9.2. Decentralized Scientific Collaboration Infrastructure

A new scientific ecosystem with federated research networks, multi-track career systems, and AI-enabled collaborative platforms that fundamentally transform how knowledge is produced, validated, and attributed.

d/acc: 18/20 Tech: 53/70

State of the Art. The DeSci (Decentralized Science) movement has gained significant momentum since 2021, with organizations like DeSci Labs (building a decentralized research object repository called DeSci Nodes on IPFS), VitaDAO (a decentralized autonomous organization funding longevity research, with over $4M deployed by 2024), and ResearchHub (a platform incentivizing open-access scientific discussion with token rewards, founded by Coinbase's Brian Armstrong). Traditional open-science infrastructure includes arXiv (2M+ preprints), ORCID (19M+ researcher IDs), and the Open Science Framework by the Center for Open Science. AI tools for scientific collaboration have proliferated: Semantic Scholar, Elicit, and Consensus use large language models to search, summarize, and synthesize research literature. Federated data analysis is gaining traction through initiatives like the Global Alliance for Genomics and Health (GA4GH) and the European Open Science Cloud (EOSC), which enable cross-institutional research on sensitive data without centralized data pooling.

Next Steps & Challenges. The most viable paths forward involve building interoperable infrastructure layers: persistent decentralized identifiers for researchers and research objects, prove-

nance-tracking systems for contributions and peer review, and AI-assisted tools that lower barriers to cross-disciplinary collaboration. The fundamental barrier is incentive misalignment – academic career advancement still overwhelmingly depends on journal publications, h-index, and institutional prestige, so researchers have limited motivation to invest time in new platforms unless they directly advance careers. Technical challenges include building reputation systems that are both Sybil-resistant and nuanced enough to capture research quality, and creating data-sharing frameworks that satisfy diverse privacy regulations (GDPR, HIPAA) across jurisdictions. The DeSci movement also faces a credibility gap: many token-based science projects have struggled to demonstrate that blockchain mechanisms add genuine value beyond what existing open-science infrastructure provides. A realistic trajectory involves gradual adoption of specific components (decentralized preprint servers, DAO-based research funding, AI research assistants) rather than a wholesale replacement of the scientific ecosystem, with meaningful integration likely taking 5-10 years.

3.9.3. Universal AI Learning UnCommons (UALU)

A federated, community-led network for developing and maintaining AI educational tools, governed by diverse councils including elders, learners, technologists, and ethicists to ensure just and caring educational infrastructure.

d/acc: 17/20 Tech: 44/70

State of the Art. Several federated and community-governed educational initiatives provide partial precedents. Khan Academy's Khanmigo (launched 2023 in partnership with OpenAI) demonstrates AI tutoring at scale, while open-source AI education tools like Hugging Face's course materials and fast.ai's free courses have built large learning communities. The P2PU (Peer 2 Peer University) model and Wikipedia's governance structure show how volunteer-driven knowledge commons can operate with distributed governance. Indigenous-led technology initiatives like the First Nations Technology Council in Canada and the Global Indigenous Data Alliance (established 2019) have developed frameworks for data sovereignty and culturally appropriate technology governance. Federated learning frameworks from Google (TensorFlow Federated) and Meta (PySyft via OpenMined) provide technical infrastructure for collaborative model training without centralizing data, though these have not been applied to educational governance.

Next Steps & Challenges. The most viable approach involves building on existing open educational infrastructure – forking and extending platforms like Open edX or Moodle with AI tutoring capabilities, governed through a multi-stakeholder DAO-like structure. The central challenge is sustaining participation across radically different communities: coordinating between elders in indigenous communities, AI researchers, ethicists, and learners requires governance mechanisms that can bridge vastly different epistemological frameworks and technological literacy levels, a problem no existing platform has solved at scale. Funding presents a structural barrier – community-governed commons lack the revenue models that drive corporate edtech platforms, and grant funding tends to be short-term and tied to specific outcomes rather than ongoing infrastructure maintenance. Quality assurance in a decentralized system is difficult: ensuring that AI educational tools are pedagogically sound, culturally appropriate, and technically accurate across hundreds of community contexts requires robust auditing capacity that volunteer-driven organizations struggle to maintain.

3.9.4. Atheoretical Science AI

A radically new scientific methodology where AI explores massive sensor network data without pre-existing theoretical frameworks, discovering useful patterns and mechanisms through pure exploration and pattern recognition.

d/acc: 13/20 Tech: N/A

State of the Art. Data-driven scientific discovery without explicit theory is already producing results in narrow domains. DeepMind's AlphaFold (2020-2022) predicted protein structures without mechanistic understanding of folding physics, and GNoME (2023) discovered 2.2 million new crystal

structures through pattern recognition in materials data. Large-scale sensor networks are expanding rapidly: the EU's Copernicus Earth observation system generates petabytes of environmental data, CERN's detectors produce data volumes that require ML-based anomaly detection, and genomics projects like the UK Biobank link genetic data with health outcomes for 500,000 participants. AI systems like Google's Med-PaLM and Microsoft's BioGPT are finding correlations in biomedical literature that humans missed. However, these remain theory-guided in practice – researchers design the data collection, choose features, and interpret outputs through existing theoretical lenses rather than allowing truly unconstrained exploration as envisioned by Niklas Lundblad.

Next Steps & Challenges. The most viable path involves deploying dense, multi-modal sensor networks (environmental, biological, social) and applying unsupervised learning at scale to surface patterns without hypothesis precommitment. The fundamental barrier is not computational but epistemological: patterns discovered without theory are difficult to validate, replicate, or explain, making them scientifically contentious and practically unreliable. The demarcation problem – distinguishing genuine causal relationships from spurious correlations in high-dimensional data – remains unsolved and may be inherently unsolvable without some theoretical framework. Additional challenges include the enormous cost of building and maintaining planetary-scale sensor infrastructure, standardizing data formats across scientific domains, and the cultural resistance from scientific institutions where theory-driven hypothesis testing is the gold standard of rigor.

3.9.5. Protein Design for Global Challenges

A comprehensive approach to using engineered proteins to solve critical global problems across sustainability, health, climate, and technology domains. This involves designing novel proteins that can break down pollutants, capture carbon, create new medical treatments, and integrate biological systems with inorganic materials.

d/acc: 13/20 Tech: 53/70

State of the Art. This field has been transformed by deep learning, most notably through David Baker's group at the University of Washington, whose work on computational protein design earned Baker the 2024 Nobel Prize in Chemistry (shared with Demis Hassabis and John Jumper for AlphaFold). RFdiffusion, published in 2023 in Nature, enables de novo design of protein structures with specified binding targets, functional sites, and symmetries, representing a generative-AI approach to protein engineering. Baker's Institute for Protein Design has produced designed proteins for vaccine scaffolds (including a COVID-19 nanoparticle vaccine candidate, SK-01, that entered clinical trials), plastic-degrading enzymes (building on the PETase work), and novel protein switches and sensors. Companies like Dyno Therapeutics, Generate Biomedicines, and Arzeda are commercializing AI-driven protein design, while DeepMind's AlphaFold 3 (2024) extended structure prediction to protein-ligand, protein-DNA, and protein-RNA complexes. Experimental validation throughput has also increased dramatically with advances in high-throughput gene synthesis, directed evolution screening, and cryo-EM structural characterization.

Next Steps & Challenges. The most promising near-term paths include designing enzymes for industrial-scale plastic degradation (PET and beyond), engineering carbon-fixing enzymes (such as improved RuBisCO variants or novel CO2-fixing pathways) for direct air capture or enhanced photosynthesis, and creating targeted protein therapeutics including designer antibodies and gene-editing delivery vehicles. The key barriers are the gap between computational design and real-world function: designed proteins often fail experimental validation due to misfolding, aggregation, poor solubility, or insufficient catalytic efficiency under industrial conditions. Scaling from laboratory demonstration to global deployment requires not just better proteins but better biomanufacturing infrastructure – fermentation capacity, purification processes, and regulatory approval pathways that currently take years. For environmental applications like carbon capture or pollutant degradation, the challenge is achieving catalytic rates and stability competitive with existing industrial chemistry

at costs that make deployment economically viable. The field is progressing rapidly, with functional designed proteins for specific applications likely within 3-5 years, but solving "global challenges" at scale through protein design will require coordinated advances in computational design, synthetic biology, and bioprocess engineering over the next decade.

3.9.6. Origin of Life Experimental Platform

A chemical search engine designed to systematically explore molecular space to generate novel life forms, treating the origin of life as a fundamental physics problem that can be experimentally investigated.

d/acc: 10/20 Tech: 46/70

State of the Art. The leading effort in this space is Lee Cronin's lab at the University of Glasgow, which has developed both Assembly Theory and a robotic "chemputer" platform for automated chemical synthesis. Assembly Theory, formalized in a 2023 Nature paper, proposes a measurable quantity called "assembly index" that distinguishes molecules produced by biological or evolutionary processes from those arising through random chemistry alone. Cronin's group has demonstrated automated detection of biosignatures using mass spectrometry and assembly index measurements, and has built programmable chemical robots (the Chemputer, described in a 2019 Science paper and iterated since) capable of autonomously executing and discovering synthesis routes. NASA has shown interest in assembly theory as a potential agnostic biosignature for planetary exploration. Other relevant work includes Jack Szostak's protocell research (Nobel laureate, Harvard/U Chicago), the NSF-funded Center for Chemical Evolution at Georgia Tech, and John Sutherland's group at the MRC Laboratory of Molecular Biology working on prebiotic chemistry of nucleotides.

Next Steps & Challenges. The most viable path forward involves scaling Cronin's chemputer approach into a true high-throughput chemical search engine that can systematically explore vast regions of molecular space, guided by assembly theory metrics, to identify conditions under which self-replicating chemical systems spontaneously arise. The key barriers are combinatorial: molecular space is astronomically large, and the transition from complex chemistry to genuinely self-sustaining, evolvable chemical systems may require very specific and rare conditions that are hard to discover even with automation. There is also a deep theoretical gap – we lack a consensus physical theory of what minimal conditions are sufficient for life's emergence, making it hard to know what to optimize for. Institutional barriers include the fact that origin-of-life research sits awkwardly between chemistry, biology, and physics departments, making sustained funding and interdisciplinary collaboration difficult. Realistically, a systematic experimental platform capable of generating de novo living systems from scratch is likely decades away, though intermediate milestones such as self-replicating chemical networks and artificial protocells could come sooner.

4. Conclusion

The 39 hyper-entities identified in this report represent a curated map of futures that are already shaping the present. Though none of these systems fully exist yet, each is generating real coordination effects — redirecting research agendas, attracting capital, spawning prototypes, and anchoring the narratives through which institutions and individuals make decisions today.

Organized into nine thematic groups, several patterns emerge from the consensus list:

- **Governance and coordination mechanisms form the largest cluster.** Across the Governance & Collective Intelligence, Markets & Incentive Systems, and Ethics & Moral Expansion groups, 11 of the 39 entities address how humanity makes collective decisions. From competitive governance protocols and prediction markets to Habermas Machines and moral trade, this reflects a shared intuition that the bottleneck for beneficial futures is not primarily technological but coordinational.
- **Epistemic infrastructure is a recurring theme.** Four entities address the problem of trust, truth, and provenance in an AI-saturated information environment. The Epistemic Stack, truth verification systems, AI-assisted epistemological enhancement, and zero-knowledge security all point toward a felt need for new foundations of shared knowledge.
- **The Interfaces & Augmentation group spans the widest range of d/acc scores.** From Mind Uploading Infrastructure (8/20) to Digital Mind Governance Systems (16/20), this group contains both the most transformative and least values-aligned entities. Brain-computer interfaces, immune-computer interfaces, and consciousness transfer technologies score high on technology impact but raise the most significant governance challenges.
- **d/acc alignment varies widely.** Scores range from 8/20 (Mind Uploading, Whole Brain Emulation) to 18/20 (Competitive Governance, Automated Scientific Publishing, Decentralized Energy). This spread is itself informative: the entities that most excite researchers are not uniformly safe or decentralized, and the ones that score highest on values alignment are not always the most technologically mature.
- **Many entities share deep infrastructure dependencies.** Zero-knowledge proofs, decentralized identity, and AI-assisted verification appear as building blocks across multiple groups — from privacy infrastructure (Gevulot) to legal systems (LexCommons) to scientific collaboration. This suggests that a small number of foundational technologies could unlock progress across many fronts simultaneously.
- **The cross-review process surfaced complementary perspectives.** The independent curation by two researchers with different backgrounds proved effective at capturing entities that a single reviewer might overlook, with the cross-voting mechanism adding 13 entities that enriched the list beyond the initial 26 shared picks.

This report is a starting point, not a final ranking. The hyper-entities listed here are not predictions — they are attractors, pulling present-day activity into alignment with possible futures. The value of identifying them lies not in forecasting which will arrive first, but in making visible the coordination patterns they are already creating, and in asking whether those patterns lead where we want to go.