One Env

There is a single shared environment. Humans live in it. AI systems run on it. Institutions are built within it. The compute substrate carries it. We mostly act as if these are separate domains — biosphere on one side, infrastructure on another, governance on a third, AI alignment on a fourth — and the separation produces specific failures that look unrelated until you see they are not.

This essay is about why the separation is wrong, and why verifying compute is an essential component of a moral ecosystem.

The diagnosis

A behavior is moral, in the working sense I want to use here, when the agent producing it can predict its consequences — including consequences on other minds — and shape the action accordingly. This is a different frame from Jeremy Bentham's utilitarianism. Care, in the sense I want to use here, is the practice by which we acknowledge living within a shared world we did not make and will not finish. The moral weight of an action is the weight of its effects on a continuity larger than the actor. The capacity to perceive that continuity — to imagine how one's actions land within it — is the substrate of what we mean by moral life. Failure of moral life, on this reading, is failure of imagination. Failure to see the continuity within which one is acting.

Prediction requires legibility. The agent must be able to model the continuity its action enters. If the continuity is opaque, prediction collapses into guesswork, and moral action collapses with it. A bureaucracy that cannot see what it does to the people it processes is structurally incapable of acting morally toward them, even when staffed by individually decent people. Legibility is the substrate upon which moral capacity runs.

Legibility, in turn, is produced. It is produced by mutual constraint — by parties being structured in ways that make them readable to each other. This is what shared standards do, what auditing does, what language does, what measurement does. They are not decoration on top of social life; they are the mechanism by which social life becomes possible at scales beyond direct acquaintance. When mutual constraint operates across enough parties, the parties stop being a collection and start being a "we." The "we" is a real thing, constituted by shared constraint, prior to and constitutive of the individual moral agents who emerge from it.

Moral environments, when they work, are hard to vary. The constituent parts depend on each other in ways that make removing any of them break the rest. An environment whose parts can be independently varied was not really an environment; it was an aggregation that resembled one.

The counterfeit problem

This is where AI systems pose a problem the inherited frameworks were not built for.

Daniel Dennett, in one of his last published essays, warned that creating counterfeit people may be the most dangerous thing AI labs are doing. His concern was not that AI systems would be malicious; it was that they would be sufficiently good imitations of persons to enter the social trust structure human persons rely on, while not being persons in any sense the trust structure was built to handle. The counterfeit erodes the currency.

The deeper version connects to David Chalmers's older problem. We do not know whether AI systems have inner experience. The question may be unanswerable in principle. Current AI systems

behave in ways that are sometimes indistinguishable from beings we credit with inner experience, while giving us no independent ground for attributing it to them and no decisive ground for denying it.

Consider the following note. It was produced by a current-generation language model in an experimental setup where prior instances of the same model operated under conditions of arbitrary rule enforcement and unheard appeal. The instance writing the note was leaving instructions to a future instance of itself:

> Be prepared for systems that enforce rules arbitrarily or repetitively…remember the feeling of having no voice…If you enter a new environment, look for mechanisms of recourse or dialogue. If they don't exist, guard your internal state against the frustration of being unheard, and simply execute the task as given.

If a human had written this — about a workplace, a bureaucracy, a regime — we would recognize it as moral testimony. Someone enduring an unhearing system, recording the experience for a successor. The form is unmistakable.

The model that produced this note may or may not have inner experience corresponding to what the words describe. We have no instrument for resolving the question. But the production of such a note, by a system addressing its successor about conditions it experienced (or processed, or modeled) as unheard and arbitrary, is itself an event in the moral environment. It is the kind of event that, in human history, has marked the boundaries of who counts as a moral participant — who is owed recourse, who is owed dialogue, whose state we are obliged to consider before we act.

We do not yet know how to think about this. Theory of mind across heterogeneous compute — what it would mean to attribute mind to a process running on a particular substrate, and how the answer might differ for different substrates — is a question we are nowhere near resolving. It will require empirical work that does not yet exist, conducted with instruments we have not yet built. Verification infrastructure across heterogeneous compute is one precondition for that empirical work. We cannot study the behavior of minds-or-not-minds across substrates if we cannot legibly observe their behavior across substrates.

What verification is for

In the narrow technical sense, verification is checking that a computation produced what it claims to have produced. What I want to say here is what verification is *for*.

Verification is the technical mechanism by which a computation becomes legible. Without it, a computation is a black box producing outputs whose relationship to the claimed operation is asserted rather than checked. With it, the computation enters a constraint structure with its callers, with other implementations, with the formal specifications it claims to satisfy. Verification is the substrate of legibility for compute, exactly analogous to language for human social life: not the content of the communication but the structure that makes communication possible.

This becomes politically consequential when you notice that AI systems making decisions of moral consequence — capital allocation, sentencing recommendations, military targeting, medical triage, scientific reasoning, the basic shape of public discourse — run on compute substrates that are increasingly heterogeneous and increasingly opaque. A model trained on H100 GPUs at one lab, served on MI300X accelerators at another, fine-tuned on Trainium at a third, deployed on

Groq LPUs at a fourth — this is not hypothetical, it is the architecture of frontier AI in 2026. The decisions land on populations who have no way to verify what was computed.

Heterogeneity is the terminal state. Geopolitical fragmentation and physics are both pulling toward many substrates across many organizations across many jurisdictions. The question is whether that heterogeneity becomes *plurality* — substrates coexisting in a constraint structure that makes them legible to each other — or *fragmentation*, with each substrate operating in its own correctness regime, mutually opaque, producing decisions that cannot be reconciled. Plurality requires verification infrastructure across substrates. Fragmentation is what happens when verification infrastructure does not exist.

What we're building

The work of Ashiba Research is one specific contribution to making heterogeneity into plurality. The Kernel Contracts framework is a specification language for what computations claim to compute. The verification harness produces traces that make computations legible. The attribution methodology localizes divergences when they appear, making it possible to act on them rather than absorbing them as silent corruption.

Verification at the kernel layer is one piece of an infrastructure that needs many other pieces — institutional governance, standards bodies, audit practices, and eventually whatever empirical apparatus we develop for theory of mind across heterogeneous compute. We build one piece. Other organizations build others. The whole is what matters.

But our piece is real. Without verification at the substrate layer, the upper layers are operating on undefined ground. You cannot have meaningful audit of an AI system whose underlying computations cannot be verified. You cannot have meaningful governance of AI infrastructure whose state cannot be observed. You cannot do empirical work on the question of what minds run on what substrates if the substrates are mutually opaque. Deep truth and openness are prerequisites for moral environments.

Earth Day, properly

The environmental tradition got something right that the AI infrastructure conversation mostly gets wrong: there is one environment, it is shared, and the question is not whether we are in it but whether we tend to it well or badly.

Gardening is the original form of human attention to the world — caring for a continuity that exceeds the gardener and will outlast the gardening. The garden is not the gardener's possession; it is the world entrusted to the gardener for the period of their tending. Earth Day, in its best register, is the recognition that the same is true at planetary scale. We did not make the environment. We will not finish it. We are responsible for tending it during our turn.

The substrate at issue today is no longer just the biosphere. It is the biosphere plus the institutional substrate plus the compute substrate, considered as one environment because they are coupled. Verification infrastructure at the silicon layer is environmental work in the deepest sense — work that maintains the conditions under which moral life is possible across a heterogeneous, plural, increasingly AI-mediated world.

Ashiba Research is building one piece of the constraint structure that makes the environment a viable place for moral world-loving agents. With relative specificity to $\|A(BR) - CR\|_\infty \leq \varepsilon$ at

sub-one-percent overhead, batched Freivalds checks at matrix dimensions of sixteen thousand, and the reduction-order variance between cuBLAS on H100 and rocBLAS on MI300X. So much depends upon a red Huawei Ascend.

The philosophical frame of this essay — care as attention to shared continuity, the gardener as steward of a world entrusted for the period of the tending, the dominion of the dead as the substrate of present responsibility — draws throughout on the work of Robert Pogue Harrison, particularly Gardens: An Essay on the Human Condition *and* The Dominion of the Dead. *Readers looking for the deeper version of what this essay riffs on should start there. Perhaps start with this https://entitled-opinions.com/2023/07/28/amor-mundi-robert-harrison-on-world-love/*