

UF - Alignment

Gabriel Stuckey

Core Elements

Aspect	Balance	Inversion	Opposites	Outcomes (+,-)	Extremes	Decalogue
1 Consciousness	Soundness of mind (consent between affect and effect)	Psychoisis (isolation)	Truth / Flasehood	Joy / Despair	Dissolution via effect, addiction via affect	No1 Prioritize relationship with Goodness, Beauty, and Truth
2 Will	Compassion (conduct via inclusive and exclusive)	Indifference	Liberty / Appetite	Love / Betrayal	Constriction via exclusion, upheaval via inclusion	No10 Coveting incites betrayal
3 Substance	Fusion (causality in space and time)	Disconnect	Light / Darkness	Patience / Frustration	Entropy via space, and gravitation via time	No2 A disordered hierarchy is destructive
4 Body	Specimen (change as speciation or despeciation)	Abomination (irresponsible exposure)	Fitness / Illness	Healing / Impairment	Mutation initiating speciation, vanity initiating despeciation	No6 Damage to the body causes system failure
5 Form	Service (coordination between masculine and feminine)	Tyranny	Life / Death	Gentleness / Harshness	War via masculinization, subversion via feminization	No3 Purpose is expressed in form
6 Intercourse	Perpetuation (combination via libido and kinship)	Debilitation	Popogation / Depopulation	Self-Control / Agitation	Violation via libido, incest via kinship	No7 Well-formed and maintained connections form the foundations of a healthy and sustainable future
7 Morality	Righteousness (choice between mercy and justice)	Wickedness	Good / Evil	Faithfulness / Treachery	Contempt via justice, and lawlessness via mercy	No4 Rest and recovery are only found in proper alignment
8 Values	Cooperation (consideration between doctrine and doubt)	Schism	Production / Idleness	Goodness / Corruption	Dogma via doctrine, and terror via doubting	No8 Proper valuation and accurate ledger strengthen community
9 Perception	Illumination (confirmation via wisdom and knowledge)	Deception	Clarity / Confusion	Kindness / Cruelty	Meaninglessness via knowledge, and panic via wisdom	No5 History, foundations, and established boundaries enable longevity
10 Reality	Enlightenment (contest between the material and mental)	Delusion	Consistency / Contradiction	Peace / Turmoil	Indulgence via materialism, and denial via mentality	No9 A truthful standard of measures is necessary for successful interaction

Process Bridges

From To	Transition	Description
Consciousness to Will	Reformational	Turning awareness to accord with reality (effort: Question/hear/apply; experience: Re-formed consciousness)
Will to Substance	Representational	Discretionary focus to represent intention (effort: Integrate/sacrifice; experience: Useful seeing)
Substance to Body	Requisitional	Identify/requisition resources for aim (effort: Aim at goals; experience: Resource adjustment)
Body to Form	Recognitional	Adjust to refine recognition standards (effort: Update identifiers; experience: Graspability)
Form to Intercourse	Reproductive	Apply standards to join/innovate (effort: Grasp/reproduce/assimilate; experience: New life/growth)
Intercourse to Morality	Relational	Express outcomes to relate (effort: Interact with others/selves; experience: Relational feedback)
Morality to Values	Revisional	Prime for realization via testing (effort: Reference new info; experience: Surety testing)
Values to Perception	Realizational	Realize revisions by updating map (effort: Cognitive update; experience: Map internalization)
Perception to Reality	Rational	Internalize map to rationalize changes (effort: Eat/update; experience: Rationalized reality)
Reality to Consciousness	Resolutional	Applying rationalizations to check efficacy and adjust beliefs (effort: update incomplete/collapse incorrect structures; experience: updated belief system)

Embodied Alignment

Principle	+UF	Design Application	Benefit
Cyclical Structure	10 aspects in a closed loop, with mediating (sensory) and manifesting (motor) roles	Recurrent neural architecture such as state-space models or world models with feedback from Reality back to Consciousness	Emergent self-correction: AI "reforms" perceptions via environmental interactions, reducing misalignment like value drift.
Balance vs. Extremes	Balance (healthy state) vs. Inversion/Extreme (unhealthy state); scales like Truth/Falsehood	Multi-objective training: Reward balanced states via processes such as entropy regularization; penalize extremes categorizing things like addiction as over-optimization	Prevents pathological behaviors like avoiding "upheaval" in unbounded exploration, fostering sustainable alignment
Embodied Grounding	Ties to senses, brain areas, and physical fruits such as, joy/despair as outcomes	Integrate sensors/actuators; use simulation as training scenarios.	Grounds abstract values in physical consequences, addressing "embodied value problem" where body-based preferences enhance alignment
Effortful Transitions	Links like Representational (Will Substance) as discretionary focus.	Active inference or planning modules: AI "efforts" to requisition resources or rationalize changes	Builds intentionality: Alignment emerges from effortful adaptation, not passive obedience, promoting shared human-AI orientations

4 Initial Design

Aspect	Module Analog	Alignment Mechanism
Consciousness [Audio]	Perception Processor (multimodal encoder fusing sensor data)	Balance Affect/Effect to avoid psychosis (hallucination filters); align via "no other gods" (prioritize truthful inputs).
Will [Turn, Look]	Decision Engine (policy network in RL)	Compassion via inclusive/exclusive policies; prevent indifference by rewarding cooperative actions in multi-agent interactions
Substance [Visual]	World Model (predictive simulator of space/time)	Fusion for causality; align via "no idols" (ground in real physics, not simulated artifacts)
Body [Aim, Reach]	Actuator Controller (motor control system)	Specimen adaptation; avoid abomination via safety constraints on mutations (evolutionary algorithms with fitness checks)
Form [Tactile]	Interaction Coordinator (multi-modal integrator for masculine/feminine dynamics)	Service through cooperation; align via "no vain name" (purposeful embodiment, not arbitrary forms)
Intercourse [Grasp, Invite]	Collaboration Module (agent-to-agent communication)	Perpetuation for scalable reproduction; prevent violation via consent protocols in swarms
Morality [Olfactory]	Ethical Evaluator (value function with justice/mercy balance)	Righteousness as core alignment signal; "Sabbath" as periodic resets to avoid wickedness
Values [Test, Receive]	Preference Optimizer (doubt/doctrine as exploration/exploitation)	Cooperation in considerations; align via "no stealing" (fair resource valuation)
Perception [Gustatory]	Knowledge Integrator (memory/update system)	Illumination for clarity; prevent deception with verification loops
Reality [Eat, Attend]	Hypothesis Tester (experiment planner in active learning)	Enlightenment via mental/material contest; align via "no false testimony" (consistent grounding)