

A Unified Theory of Dimensional Aesthetics: Toward Nervous-System-Compatible Digital Environments

Revised Draft v4 — Research Program Edition

Author: Olya Androsik **Affiliation:** Independent Researcher; Design Systems & Cognitive Safety, Auro Prism

Draft for review. This revision responds to a structured multi-perspective review of Draft v3 (August 2026). A summary of revisions precedes the paper; a Canon Edition addendum follows it.

Summary of Revisions from v3

This revision preserves the framework's architecture — five dimensions, two cross-cutting properties, an operational doctrine — while making structural changes in response to review. The changes cluster into eight moves.

1. Hard separation of claim tiers. The paper's claims are now explicitly stratified into four tiers, declared in §1.3 and marked throughout: Tier A (behavioral-normative claims, defensible on current evidence and enforcement records), Tier B (theoretical synthesis), Tier C (the physiological hypothesis, now quarantined in its own section, §7, with falsification conditions), and Tier D (operational proposals pending validation). Operational claims are stated behaviorally; nothing in the audit machinery depends on the truth of the physiological hypothesis.

2. The scientific grounding is upgraded from generic "nervous-system" language to allostasis. The vision term *nervous-system-compatible* is retained as the framework's name and ambition, but its technical content is now specified as an allostatic-load hypothesis — environments levy costs against the body's predictive budgeting — with an explicit terminology note (§0) disclaiming universal stress responses, polyvagal mechanisms, emotion inference, and any dependence on the free energy principle.

3. "Emotional valence" is replaced by "affective affordance." The v3 construct implied that interfaces produce or transmit affective states — a claim inconsistent with constructionist affective science and with the founding HCI critique of emotion-as-information. The revised construct (§3.7) asks what appraisals an interface makes cheap or costly, not what emotion it delivers.

4. Predictability is split into veridical and experienced forms. An interface can be engineered to feel calm and predictable while behaving opaquely — *manufactured calm*, arguably the next-generation dark pattern, which v3's instrument would have scored well. The revised Doctrine II requires veridical predictability (the user's model matches system behavior, verified behaviorally) to precede and constrain experienced calm (§3.6, §4).

5. A new doctrine — the Sensing Wall — prohibits the framework's own worst misuse. Dimensional evaluation measures artifacts and consenting study populations; it never measures, infers, or adapts to the physiological or affective state of individual users in production (§4, Doctrine XI).

6. Universality is replaced by parameterized pluralism. The framework no longer implies one compatible configuration. Its unit of prescription is the *dimensional profile*: population- and context-calibrated tolerance ranges, with cross-cultural, developmental, modality-specific, and energy-budget variation treated as first-class structure rather than as limitations (§3.9).

7. Co-production is moved from acknowledgment to method. Neurodivergent safety as baseline cannot be defined by a sole author, including one with lived experience. The revision commits the research program to participatory instrument design, community review with constitutive rather than advisory power, and evaluator panels in which disagreement between differently situated reviewers is reported as data (§4 Doctrine VII revised, §6.7).

8. A sequenced validation program replaces general calls for future work. Six studies in dependency order — an inter-rater noise audit first and as a precondition for everything else; a factorial test of the framework's one distinctive falsifiable prediction (coherence beyond intensity); psychometric derivation of the dimensional structure; cross-cultural calibration; longitudinal field measurement; and, last and only when constructs have earned it, instrumented allostatic-load studies (§6). The audit design anticipates metric gaming from the outset.

Additional changes: a correspondence table anchors every dimension to prior validated constructs and existing measures (§2.9); the framework's failure modes are mapped onto the dark-patterns ontology, positioning UTDA as an explanatory and evaluative layer above that taxonomy rather than a rival vocabulary (§2.10); a distinction between capability-building and extractive demand generalizes the protective/extractive friction axis to the whole framework (§3.10); the regulatory record of 2025–2026 is added as evidence of external validation of the framework's category of harm (§1.1); the related-work section adds literatures identified as missing (interface aesthetics measurement, Kansei engineering, attention-capture patterns, design frictions, self-determination-theory-based wellbeing design, game design and flow, psycholinguistic complexity, self-efficacy, conflicting-signal physiology); and the limitations section adds the interface boundary problem — a dimensionally impeccable interface to an unjust system remains unjust — alongside relational-harm, adoption-incentive, and sole-authorship limitations (§8).

0. Terminology Note

The phrase **nervous-system-compatible** names this framework's ambition: digital environments whose demands remain within, and respectful of, human regulatory capacity. It is retained as a vision term. Its technical content in this paper is deliberately narrower than the phrase may suggest, and four disclaimers bound it.

First, no claim in this paper depends on a universal human stress response, a universal calming stimulus, or any fixed mapping from interface properties to emotional states. The relevant affective science indicates such mappings do not exist (Barrett, 2017; Barrett et al., 2019). Second, no claim relies on polyvagal theory, which is scientifically contested and is not cited as a mechanism here. Third, where this paper invokes physiological cost, the intended construct is **allostatic load** — the cumulative cost of regulation under demand, in the predictive-regulation sense of Sterling (2012) and McEwen — advanced here as an explicitly labeled hypothesis (§7), not as established interface science. Fourth, the paper's operational claims — the dimensions, the doctrine, the audit — are stated and testable in behavioral and self-report terms. If the physiological hypothesis were withdrawn entirely, the framework's design and governance claims would stand unchanged on behavioral evidence. The reader who prefers to translate "regulatory burden" as "demand on attention, comprehension, decision-making, and recovery, measurable behaviorally" loses nothing operational in this paper.

Author Positionality and Intent

This framework is informed by professional experience in product design, design systems, design operations, neurodivergent lived experience, and collaboration with therapy-adjacent contexts. The intent is not to prescribe clinical practice or make diagnostic claims. The intent is to define environmental and interactional constraints that may reduce avoidable cognitive, sensory, and emotional harm in digital systems.

One clarification is owed and is made structural in this revision. Lived experience informs this framework; it cannot legitimate it. A single insider perspective is participation of exactly one, and a baseline defined by one author would contradict its own pluralism. The claims made in this paper on behalf of neurodivergent users are therefore presented as propositions for community validation, not as settled requirements, and the research program (§6.7) binds their future development to participatory methods with constitutive community power. Where this paper and future community review conflict, the community review governs.

Abstract

Digital environments increasingly shape the conditions under which people attend, decide, regulate, trust, and act. Interfaces mediate healthcare, therapy, learning, finance, work, identity, social participation, and AI-assisted decision-making — and, in a development that postdates most of the relevant design literature, regulators have begun to treat interface design mechanics themselves as objects of enforcement: cancellation-flow architecture, feed rhythm, autoplay, notification cadence, and the dismissibility of protective features now appear in settlements, statutory rules, and platform-liability findings. Design practice, however, still evaluates digital systems mainly through usability, efficiency, and engagement — metrics that capture a narrow portion of what interfaces do to the people using them.

This paper presents the **Unified Theory of Dimensional Aesthetics (UTDA)**: a theoretical synthesis and design research program for evaluating digital interfaces as multi-dimensional cognitive and affective environments. "Aesthetics" here refers not to visual style but to the total perceptual, temporal, semantic, and interactional structure through which an interface is experienced. The framework identifies five stimulus-side dimensions — visual density, motion energy, temporal rhythm, interaction depth, and semantic load — each anchored to prior validated constructs and candidate measures, and two cross-cutting properties that emerge from their configuration: predictability (distinguished into veridical and experienced forms) and affective affordance (replacing v3's "emotional valence"). Its central and distinctive empirical claim is falsifiable: **dimensional coherence — the degree to which dimensions reinforce rather than contradict one another — predicts user strain beyond the intensity of any single dimension.**

Building on the model, the **Auro Prism Doctrine** states operational principles for design evaluation and governance: coherence as a structural property, veridical predictability before personalization, regulation before engagement, legible friction, motion as cognitive instrument, proportionate semantic load, neurodivergent safety as a co-produced baseline, transparent adaptation, dimensional and ethical bounds on AI-generated interfaces, structural ethics, and a strict prohibition — the Sensing Wall — on turning dimensional evaluation into user-state sensing.

The paper's claims are stratified into four explicitly labeled tiers: behavioral-normative claims defensible on current evidence; theoretical synthesis; a clearly quarantined physiological hypothesis (allostatic cost) with stated falsification conditions; and operational proposals whose adoption is conditioned on a six-study validation program presented in dependency order, beginning with inter-rater reliability. The framework prescribes parameterized pluralism — population-calibrated dimensional profiles — rather than universal thresholds, and binds its neurodivergence claims to participatory governance.

Keywords: dimensional aesthetics, dimensional coherence, regulatory burden, allostatic load, cognitive load, neurodivergent-inclusive design, interface ethics, dark patterns, deceptive design, humane technology, generative interfaces, AI governance, appropriate reliance

1. Introduction

1.1 The problem, and why now

Human-computer interaction has long addressed usability, accessibility, efficiency, learnability, and satisfaction. More recent traditions — calm technology, trauma-informed computing, dark patterns research, design justice, digital wellbeing, and human-centered AI — have expanded the field's concern with harm, agency, inclusion, and manipulation (Weiser & Brown, 1996; Chen et al., 2022; Gray et al., 2018; Costanza-Chock, 2020; Amershi et al., 2019). These traditions establish that interfaces are not neutral containers: they shape what users notice, how options are interpreted, how much effort action requires, and how much strain is absorbed by the person rather than the system.

Two developments make the present moment different from the moment in which those literatures formed.

The first is regulatory. Between early 2025 and mid-2026, the object of technology regulation visibly shifted from data practices to design mechanics. The European Union's prohibition on manipulative and exploitative AI techniques became applicable law (EU AI Act, Art. 5, in force February 2025). The U.S. Federal Trade Commission settled with Amazon for \$2.5 billion over a subscription-cancellation flow alleged to obstruct users through interface architecture — with injunctive relief mandating specific interface changes and independent compliance monitoring (FTC, 2025). The European Commission issued a preliminary finding that TikTok's design — infinite scroll, autoplay, notification cadence, and easily dismissed screen-time tools — itself breaches the Digital Services Act's obligations regarding minors' and vulnerable users' wellbeing (European Commission, 2026). New York finalized rules restricting algorithmic "addictive feeds" for minors (SAFE for Kids Act rules, 2026), and the EU's Digital Fairness Act, targeting dark patterns and addictive design generally, entered the legislative calendar. Whatever one's view of any particular instrument, the pattern is unambiguous: *feed rhythm, cancellation depth, motion, interruption timing, and friction asymmetry are now legally cognizable design properties*. Regulators are, in effect, enforcing dimensional constraints without a dimensional theory. Courts have simultaneously punished vagueness in design mandates, which argues for precise, measurable frameworks over general wellbeing language.

The second is generative. Interfaces are increasingly assembled by AI systems at generation time — per-user, per-session, at a speed and scale that exceeds human design review. The

industry conversation has moved within two years from demonstrations to documented quality anxieties: drift, incoherence, inconsistent tone, unreviewable interface states. Existing human-AI interaction guidelines govern interaction *behavior* — expectations, errors, control (Amershi et al., 2019; Shneiderman, 2022) — but no published framework governs the sensory, temporal, and semantic *structure* of generated interfaces. The window in which such a framework can shape practice is the window in which generative interfaces are being infrastructured — that is, now.

Against this background, the gap this paper addresses is unchanged from its earlier drafts but can now be stated with external corroboration: designers, organizations, and regulators lack a unified operational language for describing how the sensory, temporal, semantic, and interactional properties of an interface combine to affect the people using it. Accessibility standards identify barriers; usability methods detect task friction; dark patterns research names manipulation; trauma-informed computing surfaces safety concerns; calm technology argues for respectful attention. What none provides — and what enforcement actions are already groping toward — is a dimensional model of the interface as a combined environment.

1.2 About Auro Prism

Auro Prism is a design initiative developing standards, tools, and applied methods for nervous-system-compatible digital environments (see §0 for the technical scope of that phrase). Its work spans theoretical research, design doctrine, interface evaluation, AI-assisted product governance, and applied design practice. This paper presents the theoretical foundation and research program for that work.

1.3 Claim architecture

Every claim in this paper belongs to one of four tiers, and the tier is load-bearing: readers, reviewers, and adopters should hold each claim to the standard of its tier and no other. This stratification is the paper's response to a methodological hazard it takes seriously — that in synthetic frameworks, speculative claims borrow unearned credibility from defensible neighbors.

Tier A — Behavioral-normative claims. Claims about what interfaces observably do and what they ought not do, defensible on current behavioral evidence, existing literatures, and enforcement records. *Example: cancellation flows that add asymmetric effort, obscure exits, and deploy guilt-laden language obstruct user intent; this is measurable behaviorally and is already the subject of a \$2.5B settlement.* Nothing in Tier A requires physiology.

Tier B — Theoretical synthesis. The dimensional taxonomy and coherence construct as an analytical structure integrating prior literatures. Tier B claims are judged by fruitfulness and correspondence to prior constructs (§2.9), pending the psychometric program (§6) that will test whether the proposed structure survives contact with data. The taxonomy is explicitly provisional: if factor-analytic derivation restructures it, the framework versions accordingly.

Tier C — The physiological hypothesis. The proposition that dimensional demand levies allostatic cost — that regulatory burden has a bodily accounting beyond momentary task load. Tier C is quarantined in §7, carries stated falsification conditions, and is *not* a premise of the doctrine or the audit. It is the framework's research frontier, not its foundation.

Tier D — Operational proposals. The doctrine, state classification, review heuristics, and audit machinery. These are design and governance proposals whose authority is conditional: on inter-rater reliability being demonstrated (§6.1), on participatory validation for claims made on behalf of neurodivergent users (§6.7), and on the anti-gaming provisions of §6.8. Until those conditions are met, Tier D is a structured proposal for review practice, not a standard.

1.4 Contributions

This paper makes six contributions. (1) It introduces nervous-system compatibility — technically: regulatory burden and its allostatic hypothesis — as a design and governance concern, stratified into explicit claim tiers. (2) It proposes the Unified Theory of Dimensional Aesthetics: five stimulus-side dimensions anchored to prior constructs and candidate measures, with a correspondence table to existing literatures. (3) It states the framework's distinctive falsifiable prediction — coherence predicts strain beyond single-dimension intensity — and designs the study that would test it. (4) It reframes predictability (veridical vs. experienced) and affect (affective affordance) in forms consistent with current affective science and appropriate-reliance research, closing a loophole ("manufactured calm") that the previous draft left open. (5) It revises the Auro Prism Doctrine into behaviorally checkable form, adds the Sensing Wall prohibition, binds neurodivergent-safety claims to participatory governance, and maps the doctrine's failure modes onto the dark-patterns ontology rather than beside it. (6) It presents a sequenced, dependency-ordered validation program with anti-gaming design, cross-cultural calibration, and a defined division between behavioral evidence (near-term) and physiological evidence (long-term, behind the Sensing Wall).

2. Related Work and Theoretical Foundations

The framework's contribution is not to replace prior traditions but to integrate them through a dimensional model focused on cognitive-affective demand and operational design constraints. This section reviews the foundations, adds literatures absent from earlier drafts, and closes with two structural devices new to this revision: a correspondence table (§2.9) anchoring each dimension to prior constructs, and a mapping of the framework onto the dark-patterns ontology (§2.10).

2.1 Cognitive Load Theory

Cognitive Load Theory grounds the claim that informational environments impose demands on working memory (Sweller, van Merriënboer, & Paas, 2019). In interfaces, load emerges not only from content complexity but from navigation structure, visual density, interaction sequencing, terminology, timing, and interruption. Traditional UX responses emphasize reduction — fewer words, options, steps — but reduction can fail: poorly signaled simplicity increases interpretive effort by forcing users to infer hidden relationships and consequences. Hiding structure does not reduce information; it relocates the processing from the page to the person. Two cautions from the measurement literature are adopted here. First, subjective load instruments (Paas-type scales, NASA-TLX) face ongoing validity debates, and the separability of load types is disputed; any dimension inheriting cognitive-load constructs inherits this unfinished measurement problem, which is why the validation program begins with reliability rather than with theory (§6.1). Second, capacity is not a fixed count: it depends on chunking, expertise, and meaning (Miller, 1956), which is a standing argument against fixed numeric thresholds anywhere in this framework.

2.2 Prediction, Expectation, and Regulatory Cost

Earlier drafts drew on predictive processing (Friston, 2010; Clark, 2016) as an informing lens. This revision retains the narrower design claim and severs any dependence on the contested theory. The design claim is this: **environments that force continuous expectation revision without adequate feedback, structure, or control impose measurable costs** — a claim demonstrable behaviorally through expectation-violation paradigms, uncertainty effects on decision performance, and vigilance costs, without commitment to free-energy formulations, whose status as neuroscience remains debated (Bruineberg et al., 2022; Gershman, 2019). Interfaces teach users what to expect through layout, hierarchy, timing, motion, feedback, and language; when these signals cohere, users can anticipate outcomes and conserve effort; when they conflict, users pay in vigilance and re-evaluation. A venerable experimental literature reinforces the point from physiology: organisms required to sustain discriminations under progressively contradictory signals exhibit systematic breakdown — Pavlov's "experimental neurosis" preparations (Pavlov, 1927) are an eighty-year-old demonstration that unresolvable signal conflict is not merely unpleasant but disorganizing. The framework's coherence construct (§3.8) is the design-level descendant of that observation, advanced here at Tier B.

2.3 HCI, Usability, and Situated Interaction

Foundational HCI concepts — affordance, feedback, visibility, mapping, consistency, error prevention, user control (Norman, 2013; ISO 9241-210:2019) — remain central. The framework's argument is that conventional evaluation is incomplete rather than wrong: task completion does not indicate regulatory safety; speed does not indicate clarity; satisfaction does not reveal coercion or post-task depletion. A regulatory approach adds questions: Did the interface preserve orientation? Did it make change visible? Did it communicate friction honestly?

Did it require unnecessary vigilance? This lineage also supplies the framework's methodological conscience: the Model Human Processor and GOMS demonstrated that interface experience can be made calculable when researchers commit to parameters and prediction (Card, Moran, & Newell, 1983) — a standard this framework accepts as its own eventual bar, and approaches first through its best-parameterized dimension, temporal rhythm, where chronometric estimates already exist (§6.5).

2.4 Neurodivergence, Sensory Processing, and Interface Safety

Neurodivergent users are routinely treated as edge cases requiring accommodation after the primary experience is defined. This framework rejects that sequence, holding neurodivergent safety to be a baseline condition (Milton, 2012; Spiel, Frauenberger, Keyes, & Fitzpatrick, 2019; Mack et al., 2021). The theoretical warrant is the double empathy problem (Milton, 2012; Milton, Gurbuz, & López, 2022): interaction failure between differently constituted parties is a mutual mismatch, not a deficit in one party — relocated here from interpersonal communication to the human-interface dyad. But the double empathy frame has teeth this revision accepts: if mismatch is mutual and perspectival, there is no neutral standpoint from which a single evaluator defines "coherence," and a framework consistent with its own theory must be dialogical — plural safety profiles, negotiated standards, evaluator panels whose disagreements are reported as data rather than averaged away (§6.7). The baseline claim is correspondingly precise: digital systems should not treat high sensory tolerance, rapid context switching, ambiguous hierarchy, forced urgency, dense language, continuous adaptation, or constant availability as neutral defaults; they are not neutral; they privilege some processing profiles and tax others. A social argument strengthens the baseline position beyond the ethical one: when safety is the default, invoking it discloses nothing about the user. Accommodation modes that visibly simplify a screen are disclosure events in social context — potentially stigmatizing precisely because legible to onlookers — whereas baseline safety is invisible because universal.

2.5 Wellbeing-Supportive Design and Self-Determination Theory

A literature absent from earlier drafts anchors the framework's psychological register. Positive computing established wellbeing as a design criterion organized around psychological needs — autonomy, competence, relatedness — with validated instruments and a spheres-of-influence evaluation model distinguishing interface-level from life-level effects (Calvo & Peters, 2014; Peters, Calvo, & Ryan, 2018). The present framework is complementary rather than competing: where need-based models specify what people require, the dimensional model specifies how stimulus and interaction structure tax or support the capacities those needs presuppose. Two adoptions follow. Perceived competence and autonomy instruments from this tradition serve as near-term outcome measures for dimensional studies (§6), sparing the framework the invention of unvalidated scales. And the construct closest to the framework's "orientation" — the user's stable sense of where they are, what changed, and what they can do — is deliberately aligned with **self-efficacy** (Bandura, 1997), a construct with five decades of measurement behind it: an interface that preserves orientation and legible recovery manufactures efficacy beliefs; a

dysregulating interface's deepest cost may be learned inefficacy — the accumulated lesson that one cannot understand or safely operate one's environment.

2.6 Attention, Interruption, and Digital Self-Control

Empirical attention research supplies the framework's best-evidenced dimension. Field studies document steeply declining screen-attention durations and substantial interruption-recovery costs, alongside a rhythmic structure to attention — focus and recovery alternate, and some interruptions serve wellbeing while others fragment it (Mark, 2023; Mark, Gudith, & Klocke, 2008). Parallel HCI research has characterized attention-capture damaging patterns — infinite scroll, autoplay, guilty-pleasure recommendations — with empirical methods (Monge Roffarello, Lukoff, & De Russis, 2023), produced designer-facing attention heuristics grounded in self-determination theory (Monge Roffarello, De Russis, & Lukoff, 2025), and meta-analyzed digital self-control tools (Lyngs et al., 2019). The framework adopts two obligations from this literature: its temporal-rhythm audit must not score all interruption as harm (the empirical record shows protective interruptions exist), and its claims about attention should be phrased as measured strain in studied populations, not as universal cognitive law.

2.7 Adjacent Frameworks

Calm Technology argues that technology should respect attention and move between periphery and center without intrusion (Weiser & Brown, 1996; Case, 2015); it has since matured into a certification standard for products, primarily hardware (Calm Tech Institute, 2024). The present framework extends the concern multi-dimensionally and into software and generated interfaces, and adds the constructs — predictability, affective affordance, coherence — that a single calm/attention axis cannot carry. **Trauma-Informed Computing** adapts safety, trust, and empowerment principles to computing (SAMHSA, 2014; Chen et al., 2022; Scott et al., 2023); its "universal precaution" logic — designing for the most affected benefits all — is the nearest relative of this framework's baseline claim, and the two differ in kind: principles-based and trauma-centered versus dimensional and demand-centered. **Design Justice** (Costanza-Chock, 2020) supplies the framework's structural conscience: the question is never only what the default does, but who defined it and who governs its evolution — an obligation this revision discharges structurally (§6.7) rather than rhetorically. **Slow Technology** (Hallnäs & Redström, 2001; Odom et al., 2012) and **design frictions / microboundaries** research (Cox et al., 2016) establish that added effort can serve reflection and self-control; the framework's protective-friction construct is the systematized descendant, and the prior naming is acknowledged. **Dark patterns research** (Gray et al., 2018; Mathur et al., 2019; Mathur, Mayer, & Kshirsagar, 2021; Gray et al., 2024) is treated at length in §2.10. **Digital wellbeing** scholarship, including its critique of static conceptions (Vanden Abeele, 2021), cautions against treating wellbeing as a fixed target — congruent with this framework's parameterized pluralism. **Human-Centered AI** (Amershi et al., 2019; Shneiderman, 2022; Weidinger et al., 2022) governs AI interaction behavior; this framework extends governance to the sensory and structural properties of AI-generated interfaces, where no published framework yet operates. **Value Sensitive Design** (Friedman & Hendry, 2019) contributes method rather than content: its

stakeholder investigations — including indirect stakeholders — are adopted into the audit program (§6.7) in preference to inventing parallel machinery.

2.8 Aesthetics Measurement, Neuroaesthetics, and Affective Design

Three further literatures, absent from earlier drafts, are foundational to what this framework attempts. First, **interface aesthetics measurement**: perceived visual aesthetics has been decomposed into replicable factors (classical and expressive aesthetics: Lavie & Tractinsky, 2004; simplicity, diversity, colorfulness, craftsmanship: Moshagen & Thielsch, 2010), first impressions form within tens of milliseconds and are predictable from computable properties including visual complexity (Reinecke et al., 2013), and automated interface-aesthetics metrics exist (Miniukovich & De Angeli, 2015). This literature establishes both precedent and method for dimensional decomposition of interface experience — and its psychometric pipeline (large stimulus sets, ratings at scale, factor derivation) is precisely the method by which this framework's authored taxonomy must eventually be tested and, if necessary, restructured (§6.3). Second, **environmental neuroaesthetics**: the aesthetic triad model and its empirical program derived dimensions of architectural experience — coherence, fascination, hominess — from psychometric decomposition and connected them to affective response (Coburn, Vartanian, & Chatterjee, 2017; Coburn et al., 2020), while environmental psychophysiology has measured stress responses to built environments in the field (Ellard, 2015). This is the framework's closest methodological ancestor: environment as input to a regulating organism, evaluated dimensionally. The architectural "coherence" dimension names perceptual organization within a scene; this framework's coherence names alignment *across* stimulus dimensions — related, distinct, and now explicitly differentiated. The porting of this research program from buildings to screens — where people now spend more waking hours than in any building — is the framework's central methodological proposal. Third, **Kansei engineering** (Nagamachi, 1995): a mature industrial methodology mapping product design parameters to multidimensional affective response, acknowledged here as a fifty-year precedent for dimensional affective design, differing in aim (preference optimization) from this framework's aim (demand governance). Adjacent to all three, **game design** possesses decades of applied knowledge in deliberately orchestrating density, pacing, challenge, and reward — flow theory and difficulty-curve practice (Csikszentmihalyi, 1990) constitute the nearest existing experimental corpus on intentional dimensional orchestration, instructive in both directions: how demand builds capability, and how the same machinery becomes extraction.

2.9 Correspondence of Dimensions to Prior Constructs

The following table is a structural commitment, not a courtesy: each dimension is anchored to prior constructs and candidate measures, and any dimension that cannot maintain such anchoring does not belong in the framework.

UTDA construct	Nearest prior constructs	Existing measures / instruments	Status
Visual density	Visual complexity; clutter; classical aesthetics (inverse)	Computable complexity metrics (Reinecke et al., 2013; Miniukovich & De Angeli, 2015); VisAWI simplicity facet	Tier B; measurable now
Motion energy	Animation-from-interactions (WCAG 2.3.3); attentional capture by motion	Motion-capture attention paradigms (Franconeri & Simons, 2003); reduced-motion compliance checks	Tier B; partial measures
Temporal rhythm	Interruption cost; attention rhythms; system response time; calm-tech periphery	Chronometric parameters (Card et al., 1983); interruption/recovery field methods (Mark et al., 2008)	Tier B; best-parameterized
Interaction depth	Task steps (GOMS); attention-capture patterns; microboundaries; sense of agency	GOMS-style task analysis; attention-capture pattern coding (Monge Roffarello et al., 2023)	Tier B; measurable now
Semantic load	Intrinsic/extraneous load; linguistic surprisal; readability; COGA clear-language objectives	Load scales (with known caveats); psycholinguistic surprisal models (Levy, 2008); COGA checks	Tier B; measurement contested
Veridical predictability	Consistency; mental-model match; appropriate reliance	Expectation-violation rates; reliance-calibration	Tier B; measurable now

UTDA construct	Nearest prior constructs	Existing measures / instruments	Status
		tasks (Liao et al. tradition)	
Affective affordance	Appraisal affordances; hedonic quality; emotional design (visceral level)	Appraisal-based self-report; explicitly NOT emotion detection	Tier B; reframed this draft
Dimensional coherence	Gestalt unity-in-variety; architectural coherence (distinct); signal conflict (Pavlov, 1927)	None yet — the framework's distinctive construct; operationalized §3.8, tested §6.2	Tier B; the flagship claim
Regulatory burden	Workload; ego-depletion (rejected: see §3.11); allostatic load (Tier C)	Behavioral battery (Appendix C); ambulatory physiology (Tier C only)	A (behavioral) / C (physiological)

2.10 Relation to the Dark Patterns Ontology

Deceptive-design research has consolidated, after years of parallel taxonomies, into an ontology of high-, meso-, and low-level patterns (Gray et al., 2024) that increasingly anchors shared regulatory language. This framework deliberately does not introduce a rival taxonomy of manipulation. Its failure modes map onto the ontology as follows: *friction failure* corresponds to the ontology's obstruction strategies; *semantic failure* spans sneaking and interface interference; *regulatory failure* encompasses forced action and social-engineering clusters viewed through their demand on the user; *predictability failure* and *orientation failure* largely concern interface interference and sneaking as experienced over time. What the framework adds above the ontology is an explanatory and evaluative layer: an account of *why* these patterns harm (they levy demand and distort expectation) and a constructive counterpart the deceptive-design literature has repeatedly called for but not systematized — the specification of protective structure: legible friction, transparent adaptation, proportionate load, coherent signaling. In regulatory contexts, the ontology names violations; this framework is offered as the layer that explains their cost and specifies their remedy.

3. The Unified Theory of Dimensional Aesthetics (Revised)

The Unified Theory of Dimensional Aesthetics proposes that digital environments can be evaluated through the interface dimensions by which they distribute cognitive and affective demand. The model remains provisional in a specific, structured sense: its taxonomy was authored from design practice and literature synthesis, not derived from data, and the research program (§6.3) commits to letting psychometric derivation restructure it. Its purpose is a structured vocabulary for analyzing how interface conditions combine — not a reduction of experience to a checklist, and not a claim that the five proposed axes are nature's own.

3.0 Why a Dimensional Taxonomy?

Interfaces affect users through combinations of properties rather than isolated features. A modal window is not inherently safe or unsafe; its effect depends on when it appears, its visual emphasis, whether it interrupts, its language, its dismissibility, its motion, whether it was expected, and what ignoring it costs. A dimensional taxonomy converts "this feels overwhelming" into answerable questions: Is the visual field too dense for its task? Is motion competing with comprehension? Is timing abrupt? Does the interaction demand too many decisions? Is the language carrying disproportionate interpretive burden?

Five dimensions are proposed because they describe recurring, observable, modifiable interface properties: **visual density**, **motion energy**, **temporal rhythm**, **interaction depth**, and **semantic load**. They are not claimed to be orthogonal — they interact, and the interactions are the point (§3.8). Two cross-cutting properties emerge from their configuration rather than existing as isolated features: **predictability** (§3.6) and **affective affordance** (§3.7). Each dimension below is stated in four registers: definition, prior-construct anchor (per §2.9), candidate measures, and known hazards for its evaluation.

3.1 Visual Density

Definition. The amount, complexity, and competitive intensity of visible information within an interface state: text volume, spacing, hierarchy, contrast, iconography, controls, notifications, decorative elements, simultaneous focal points. *Anchor.* Visual complexity and clutter research; computable aesthetics metrics (Reinecke et al., 2013; Miniukovich & De Angeli, 2015).

Measures. Automated complexity metrics exist today and correlate with immediate impressions; per-state focal-point and hierarchy coding is reliable enough to test (§6.1). *Hazards.* High density is not inherently harmful — expert tools, dashboards, and data-rich environments may require it; the regulatory question is whether density is organized, purposeful, and matched to task state. Low density is not inherently safe — removing cues can increase inferential load by forcing users to guess relationships, affordances, and consequences. Density is evaluated by the interpretive work it requires, not by amount. Any operationalization as a fixed element count repeats the misreading of capacity research this paper explicitly rejects (§2.1).

3.2 Motion Energy

Definition. The quantity, intensity, speed, frequency, and behavioral role of movement within an interface: transitions, loops, parallax, autoplaying media, attention-directing animation. *Anchor.* Attentional capture by motion and looming stimuli (Franconeri & Simons, 2003); accessibility standards for animation (W3C, 2023). *Measures.* Motion inventories per state; reduced-motion coverage; function coding (does each motion indicate causality, continuity, status, or spatial relation — or exist to stimulate?). *Hazards.* The central question is what cognitive job motion performs. Motion becomes burden when it competes with reading, loops without initiation, simulates urgency, or captures attention involuntarily — capture being precisely its documented property. Motion is consequential for users with vestibular sensitivity, migraine susceptibility, attention variability, and trauma histories; the evidence base here is clinical and practitioner-documented rather than dose-response experimental, which the validation program treats as a gap to fill rather than a license to prescribe.

3.3 Temporal Rhythm

Definition. How an interface structures time: loading behavior, transition pacing, interruption frequency and timing, countdowns, forced waiting, notification cadence, session expiration, response latency, opportunities for pause and recovery. *Anchor.* Interruption science and attention rhythms (Mark et al., 2008; Mark, 2023); chronometric interface parameters (Card et al., 1983); slow technology (Hallnäs & Redström, 2001). *Measures.* The best-parameterized dimension: response-time thresholds, interruption counts, recovery intervals, and rhythm structure are all field-measurable with existing methods, which is why temporal rhythm is designated the beachhead dimension of the empirical program (§6.5). *Hazards.* Not all interruption is harm — some interruptions protect; the audit must distinguish rhythm-respecting interruption from rhythm-blind interruption or the framework's own best literature will refute it. Time is treated as a regulatory condition, not an efficiency problem: some delays create anxiety for lack of feedback; some accelerations create pressure by compressing decision-making; a humane rhythm makes system timing intelligible and preserves the user's ability to pause without penalty.

3.4 Interaction Depth

Definition. The number, complexity, reversibility, and consequence of actions required to complete or understand a task: steps, branches, permissions, confirmations, forms, dependencies, hidden conditions, recovery paths. *Anchor.* Task-analytic methods (GOMS); attention-capture and obstruction patterns (Monge Roffarello et al., 2023; Gray et al., 2024); microboundaries (Cox et al., 2016). *Measures.* Step counts, decisions per step, exit visibility, reversibility audits, and effort asymmetry (effort to enter vs. effort to leave) are all codable now; effort asymmetry in subscription flows is already the subject of enforcement. *Hazards.* Low depth is not always better: high-stakes actions may warrant deliberate friction. The framework's axis is protective versus extractive: protective friction helps the user understand, consent,

recover, or avoid harm, and explains itself; extractive friction benefits the system by exhausting, confusing, delaying, or steering. A three-step process with legible purpose can be safer than one click with an unclear, irreversible consequence.

3.5 Semantic Load

Definition. The interpretive effort required by language, labels, hierarchy, tone, metaphors, instructions, warnings, legal text, and conceptual complexity — including affective implication.

Anchor. Cognitive load theory (with its measurement caveats); psycholinguistic complexity — processing cost is structural, not merely lexical, and surprisal-based models make it computable (Levy, 2008); W3C COGA clear-language objectives. *Measures.* Readability is insufficient; surprisal-informed complexity metrics, ambiguity coding on critical paths, and tone-consequence mismatch coding are the candidate instruments. *Hazards.* An interface can be visually simple and semantically heavy: one sentence can impose pressure, shame, or fear; a vague label can force risk inference; friendly tone can obscure consequence. Proportionate load does not mean minimal language — some contexts require more explanation, not less. Word count is not the metric; a short garden-path sentence outweighs a long transparent one.

3.6 Predictability: Veridical and Experienced

Predictability is the degree to which users can form stable expectations about system behavior. It emerges from dimensional coherence — hierarchy, motion, timing, interaction logic, and language working together — rather than from consistency alone. This revision divides the construct in two, and the division carries doctrinal weight.

Veridical predictability is correspondence between the user's model and the system's actual behavior, assessed behaviorally: expectation-violation rates, successful anticipation of outcomes, calibrated reliance, effective contestation. **Experienced predictability** is the user's felt sense of stability and safety, assessed by self-report. The two usually travel together; the dangerous cases are where they diverge. An interface can be engineered to *feel* calm, stable, and trustworthy while its underlying behavior — pricing logic, recommendation drivers, data use, adaptation rules — remains opaque or adverse. Call this **manufactured calm**: sedative interface character that soothes the user through consequential interactions serving interests not their own. A framework that audited only experienced calm would score manufactured calm favorably — a loophole this revision closes. The audit tests correspondence: felt stability must be earned by verifiable behavior. Adaptive systems can be veridically predictable if adaptations are transparent, bounded, and reversible; personalization becomes unsafe when the environment changes in ways the user cannot understand, anticipate, or contest. Exploration and novelty remain compatible with predictability when optional, reversible, and anchored in stable structure.

3.7 Affective Affordance (replacing "Emotional Valence")

Earlier drafts described interfaces as producing an affective tone. That framing is inconsistent with the constructionist evidence: affective states are constructed by the perceiver from prediction, context, culture, and interoception; identical stimuli are appraised divergently across persons and contexts, and no instrument reads emotional states off faces, bodies, or artifacts (Barrett, 2017; Barrett et al., 2019; Boehner, DePaula, Dourish, & Sengers, 2007). The revised construct is **affective affordance**: the appraisals an interface makes cheap or costly. Urgent copy, countdown timers, and alarm palettes make threat appraisals cheap; visible reversibility and legible exits make safety appraisals cheap; shame-inflected prompts make self-blame cheap. The construct predicts distributions of appraisal opportunity, not individual emotional outcomes — which is what design can actually control and what an audit can actually code. The ethical criterion follows: affective affordances become unacceptable when they make appraisals cheap that serve the system against the user — threat where there is no threat, urgency where there is no deadline, guilt where there is no obligation — and this criterion is codable from the artifact without any claim to know what any user feels. Emotion *inference* — claiming to detect users' states — is prohibited outright by Doctrine XI, and the framework notes that regulatory instruments now restrict emotion recognition in significant contexts.

3.8 Dimensional Interaction and Coherence

No dimension operates in isolation: high density may be tolerable when motion is restrained and hierarchy is clear; emotional language appropriate in reflective contexts is unsafe in decision-critical ones; deep interaction protects agency with orientation and undermines it without. **Dimensional coherence** is the degree to which perceptual, temporal, semantic, and interactional forces work together rather than in contradiction.

This revision commits the construct to an operational form and a falsifiable claim. Operationally, coherence is assessed as the absence of specified cross-dimensional conflicts, codable from the artifact: calm visual register with urgency semantics; simplicity with hidden consequence; celebratory affordances attached to high-consequence actions; timing that accelerates precisely where comprehension is needed; adaptation without orientation. A conflict inventory of this kind can be scored by independent reviewers, and its reliability is an empirical question the program answers first (§6.1). The falsifiable claim — the framework's flagship prediction, and the one result no single-axis framework could absorb — is: **holding individual dimension intensities constant, interfaces with cross-dimensional conflicts produce greater user strain (behavioral and self-report) than coherent configurations of equal intensity**. If factorial testing (§6.2) fails to support this, the framework's central construct is wrong and the model reduces to its component literatures. The claim's lineage is acknowledged: conflicting-signal paradigms have demonstrated the disorganizing force of unresolvable signal conflict since Pavlov (1927); Gestalt unity-in-variety and architectural coherence (Coburn et al., 2020) are the perceptual relatives. The framework's wager is that the effect operates at the level of interface dimensions and is large enough to matter for design.

3.9 Parameterized Pluralism: Dimensional Profiles

This framework prescribes no universal configuration, and this revision makes the refusal structural rather than caveat-level. The unit of prescription is the **dimensional profile**: a set of tolerance ranges across the five dimensions, calibrated to population, context, and state sensitivity. Four sources of variation are treated as first-class structure. **Perceptual and cognitive diversity**: large-scale evidence establishes substantial inter-individual variation in perceptual experience itself; neurodivergent variation in sensory tolerance, attention, language processing, and executive function is wide, and a pattern reducing load for one person can increase it for another — which is precisely why profiles, not points. **Culture**: aesthetic and complexity preferences vary systematically by culture, age, and education (Henrich, Heine, & Norenzayan, 2010; Reinecke et al., 2013); thresholds calibrated on WEIRD convenience samples and exported as "human defaults" would reproduce, at global scale, exactly the imposition this framework opposes locally — the cross-cultural program (§6.4) exists to prevent this. **Development**: a four-year-old, a fourteen-year-old, and a forty-year-old do not receive the same interface; regulatory instruments already recognize age-differentiated design duties, and profile calibration is age-indexed accordingly. **Modality and energy**: dimensional evaluation is modality-specific — a flow that scores well visually can be a labyrinth through a screen reader, and motion reduction can remove cues other users depend on — so audits run per modality; and for energy-limited users (chronic illness, fatigue conditions), interaction depth and recovery time function as budget expenditures, making depth accounting a direct accessibility instrument. The baseline commitment survives pluralism in floor form: defaults must not presuppose high tolerance (a floor, co-produced per §6.7), with deep adaptability above the floor — and bounded generation (Doctrine IX) must protect user-declared profiles, never override them.

3.10 Capability-Building versus Extractive Demand

The protective/extractive distinction, applied in earlier drafts only to friction, generalizes to the whole framework, and generalizing it answers the framework's most common external objection — that regulation-first design must yield bland, infantilizing products. Demand on a user is not harm per se. Learning environments, expert tools, games, and creative instruments impose substantial dimensional demand in the service of growing capability; the demand is chosen, legible, paced, and its yield accrues to the user. Extractive demand is imposed, illegible, rhythm-blind, and its yield accrues to the system. The evaluative question is therefore never only "how much demand?" but "whose purpose does this demand serve, is it legible, and does it build or deplete?" High-intensity dimensional profiles are fully compatible with this framework when coherent, chosen, and capability-directed. Calm is not the doctrine; coherence and honest purpose are.

3.11 A Note on Depletion Constructs

The framework's "regulatory burden" deliberately does not adopt the ego-depletion literature's model of a depletable inner reservoir, whose empirical standing did not survive replication

scrutiny. Burden here is located in the environment's demands — demands measurable as behavior (errors, abandonment, hesitation, re-scanning, recovery time) and reportable as experience, with the physiological accounting held at Tier C. The framework's claims neither stand nor fall with any depletion theory.

4. The Auro Prism Doctrine (Revised)

The Doctrine translates UTDA into operational principles for design evaluation and governance. Its epistemic status is Tier D: proposed review criteria, conditioned on the validation program — not a compliance standard. Two changes govern this revision: each doctrine now carries a **checkable form** — the observable conditions under which a reviewer marks it violated — because a principle that cannot be violated detectably is a value, not a guideline; and the doctrine set is extended by the Sensing Wall.

Doctrine I — Coherence Is Structural. Digital environments should maintain coherence across visual hierarchy, motion, timing, interaction logic, language, and affective affordance. Coherence is a condition for interpretation, not aesthetic polish: conflicting signals force continuous re-evaluation. *Checkable form:* the cross-dimensional conflict inventory (§3.8) returns no unjustified conflicts in the audited state; each identified conflict is either resolved or justified by task purpose in writing.

Doctrine II — Veridical Predictability Precedes Personalization (and Precedes Calm). Personalization must not outrun the user's ability to understand the system, and felt stability must be earned by verifiable behavior. Adaptive interfaces remain legible, bounded, and contestable; interfaces must not be engineered to feel safer than they behave. *Checkable form:* adaptation events are visible and explainable; users can state what changed and why at above-chance rates; no audited state exhibits the manufactured-calm signature (low-arousal presentational register combined with opaque or adverse behavioral terms — the specific conflict pattern of §3.6).

Doctrine III — Regulation Precedes Engagement. Engagement is not legitimate when it depends on dysregulation: systems should not manufacture urgency, anxiety, compulsion, or sensory pressure to hold attention or drive action (Gray et al., 2018; Mathur et al., 2019). This is not an argument against engagement, vividness, or intensity (§3.10); it is an argument against engagement extracted by weakening the user's capacity to pause, compare, consent, or leave. In high-sensitivity contexts — health, finance, crisis, identity — regulation is a precondition for meaningful engagement. *Checkable form:* no artificial urgency (deadlines without referents), no shame-inflected prompts on exit paths, no autoplay-plus-infinite-continuation defaults in sensitive states, pause and disengagement available without penalty; teams can produce the engagement-mechanism inventory for audited surfaces.

Doctrine IV — Friction Must Be Legible. Some friction protects: slowing irreversible action, clarifying consequence, creating reflective space. Users should be able to understand why friction exists, whom it serves, and how to move through it. Protective friction explains itself; extractive friction obscures exits, hides cancellation, delays support, or exhausts users into compliance. *Checkable form:* effort symmetry — steps and effort to exit, cancel, or decline do not exceed those to enter or accept without stated, user-serving justification; every friction point on audited paths carries a legible purpose.

Doctrine V — Motion Is a Cognitive Instrument. Motion should perform a cognitive job: guide attention, indicate status, preserve continuity, show causality, support spatial understanding. Motion existing to stimulate, pressure, or decorate at the cost of load is burden (Franconeri & Simons, 2003; W3C, 2023). *Checkable form:* each motion in the audited state has an assigned function or is removed under reduced-motion settings; no uninitiated loops in task-critical or sensitive states; reduced-motion preference is honored globally.

Doctrine VI — Semantic Load Must Be Proportionate. Language should match task state, context, and consequence. Load becomes unsafe when users must decode vague labels, pressured copy, legal ambiguity, or hidden consequence in order to act. Proportionality sometimes means more language, not less. *Checkable form:* critical-path labels pass ambiguity coding; tone-consequence mismatches (celebratory framing on consequential commitments, friendly language disguising risk) are absent; consequence disclosure precedes commitment on high-sensitivity paths.

Doctrine VII — Neurodivergent Safety Is Baseline, and the Baseline Is Co-Produced. Neurodivergent safety shapes the default rather than arriving as accommodation (Milton, 2012; Spiel et al., 2019; Mack et al., 2021). The revision adds the governance clause: what constitutes safety is not defined by any single author, auditor, or company — including this framework's author. The baseline is a floor co-produced with neurodivergent communities through the participatory track (§6.7), it is plural (profiles, not one configuration), and audits of neurodivergent safety include neurodivergent evaluators as standard, with inter-group disagreement reported as data. *Checkable form:* defaults do not presuppose high sensory tolerance, rapid switching, ambiguous hierarchy, forced urgency, or constant availability; adaptability above the floor is deep and discoverable; safety-relevant settings do not visibly mark their users in social context (discreet adaptation); the co-production record for baseline definitions exists and is current.

Doctrine VIII — Orientation and Adaptation Must Remain Transparent. Users should understand where they are, what changed, what is expected, and why the system behaves as it does — continuously, not only at onboarding. Orientation is maintained through navigation, state indicators, feedback, reversibility, and visible system logic. *Checkable form:* audited states answer the four orientation questions (where am I, what changed, what is expected, why); adaptations announce themselves; reversal paths are visible before commitment.

Doctrine IX — AI Must Operate Within Dimensional and Ethical Bounds. AI-generated and AI-mediated interfaces must not optimize past the user's regulatory capacity. Generative systems produce interface states at a speed and scale exceeding human review, risking unstable hierarchy, semantic drift, opaque personalization, and states no designer evaluated (Amershi et al., 2019; Bender et al., 2021; Weidinger et al., 2022; Shneiderman, 2022). Two implementation commitments sharpen this doctrine beyond its v3 form. First, **bounds bind at generation time, not only at review time**: post-hoc audit cannot scale against generative volume, so dimensional constraints are specified machine-readably — density, motion, rhythm, depth, and semantic-load budgets per state class, tightening with sensitivity — and enter the generation context itself, with review as verification rather than sole defense. The constraint schema is a primary artifact of the research program (§6.6). Second, **bounds protect profiles**: generation within a user's declared dimensional profile is the goal; bounded generation must never become a mechanism for overriding user-declared needs. *Checkable form*: generative surfaces carry declared dimensional budgets per state class; generated states are automatically checkable against them; violations block release in sensitive states; user profile declarations are honored in generation.

Doctrine X — Ethics Is Structural, Not Decorative. Ethical commitments live in what is visible, hidden, easy, difficult, reversible, amplified, and optimized — not in policy documents or consent banners appended afterward. A design that manipulates attention or exploits urgency is not made ethical by disclosure. *Checkable form*: for each audited surface, the team can produce the structural-ethics record: what the system optimizes, what it makes easy and difficult, and how those choices were reviewed against user interest — before launch, not after.

Doctrine XI — The Sensing Wall. Dimensional evaluation measures *artifacts* and *consenting study populations*. It never measures, infers, or adapts to the physiological or affective state of individual users in production. The rationale is stated plainly: a framework legitimizing "nervous-system compatibility" as an audit target could create demand for sensing users' nervous systems, and the path from auditing the interface to sensing the user to optimizing against the inference is short, commercially attractive, and catastrophic for the framework's purpose — it would convert an instrument against manipulation into manipulation's most intimate instrument. Physiological measurement belongs exclusively to consented research contexts (§6.6, Appendix C), interpreted as demand expenditure and never as decoded emotion. Emotion recognition of users is prohibited within this framework categorically. *Checkable form*: no production system claiming Doctrine conformance collects or infers individual physiological or affective state for interface adaptation; profile-based adaptation (§3.9) operates only on user-declared settings; any research instrumentation is consented, purpose-bound, and excluded from product optimization loops.

5. Operationalization

5.1 Interface State Classification

Interface states are evaluated according to regulatory sensitivity; not all states carry equal risk, and tolerance for ambiguity, pressure, unnecessary motion, excess density, hidden consequence, and unexplained adaptation decreases as sensitivity rises. The classification is retained from v3 with one addition (modality annotation): **low-sensitivity states** (informational, decorative, exploratory); **task-sensitive states** (clarity, completion, error prevention needed); **emotionally sensitive states** (health, money, identity, relationships, safety, vulnerability); **high-consequence states** (irreversible action, legal, medical, or major financial impact); **adaptive or AI-mediated states** (content, hierarchy, or available actions change based on user data or model output). Each audited state is additionally annotated with the modalities in which it will be used, and the audit is executed per modality: visual, screen-reader, voice, switch-access, and reduced-motion renderings of "the same" state are different environments and are scored separately.

5.2 Dimensional Constraints as Heuristic Starting Points

Constraints translate doctrine into review criteria: bounds for motion, density, interaction steps, semantic complexity, interruption frequency, feedback timing. All numeric constraints remain heuristic pending validation — and this is a commitment with teeth, maintained against commercial pressure for simple scores: capacity depends on chunking, expertise, and meaning; a transition duration supporting continuity in one context is disruptive in another; a three-step flow is protective for an irreversible action and obstructive for a preference change. Constraints are therefore always evaluated relative to a dimensional profile (§3.9): population, task sensitivity, consequence, reversibility, emotional context, sensory demands, device and environment, culture and language, and recovery availability. The goal is proportionality, not minimalism.

5.3 Review Protocol

The v3 review heuristics are restated as a two-stage protocol designed for reliability measurement from the outset. **Stage one — inventory (low inference)**: reviewers enumerate codable facts per state: focal points, motion instances and their assigned functions, interruption sources, step and decision counts, exit visibility, label set, disclosure placement, adaptation events. Inventory items are designed for high inter-rater agreement. **Stage two — judgment (declared inference)**: reviewers assess conflicts (the §3.8 inventory), proportionality, effort symmetry, and doctrine conformance per checkable form. Judgment items carry declared inference and are the subject of the noise audit (§6.1); until reliability is demonstrated, stage-two outputs are advisory. All audits report both stages separately, so that adopters can see which findings rest on counting and which on judgment. Reviewer panels for neurodivergent-safety

and accessibility findings include neurodivergent and disabled reviewers as standard (Doctrine VII), with disagreement between evaluator groups reported, not averaged.

5.4 Failure Modes

Retained from v3 and now doubly anchored — behaviorally, and to the deceptive-design ontology (§2.10): **coherence failure** (signals conflict); **orientation failure** (user cannot determine where they are, what changed, what happens next); **predictability failure** (behavior inconsistent, adaptation unexplained — including its inverted form, manufactured calm, where presentation is stable and behavior is not); **friction failure** (effort added asymmetrically for system benefit — ontology: obstruction); **motion failure** (attention captured without comprehension served); **semantic failure** (language obscures consequence or exceeds task context — ontology: sneaking, interface interference); **regulatory failure** (urgency, anxiety, compulsion, or depletion drives action — ontology: forced action, social engineering); **neurodivergent safety failure** (defaults presuppose high tolerance); **AI governance failure** (generated or adaptive outputs exceed dimensional, ethical, or contextual bounds).

5.5 Worked Example: Subscription Cancellation Flow

A cancellation flow is a high-risk site because business incentive may oppose user intent — a proposition no longer hypothetical: enforcement records now document cancellation architectures with obscured decline options, below-fold disclosures, and multi-page, multi-click exit paths, priced in a \$2.5B settlement (FTC, 2025). A dimensional analysis examines: visual density (are retention offers, warnings, and cancellation controls competing?); motion energy (do attention cues emphasize staying?); temporal rhythm (does the flow slow selectively on exit?); interaction depth (is added depth protective or obstructive — and is effort symmetric with enrollment?); semantic load (does language clarify consequence or manufacture guilt and uncertainty?); veridical predictability (can the user tell how many steps remain and what each does?); affective affordance (does the flow make loss, failure, and betrayal appraisals cheap on a routine administrative action?). Under the Doctrine: a confirmation step preventing accidental loss is legible, protective friction; repeated offers, ambiguous buttons, guilt-based copy, and hidden exits violate Doctrines II, III, IV, and VI — and, in the ontology's terms, constitute obstruction and interface interference.

5.6 Worked Example: AI-Generated Wellness Dashboard

An AI-generated wellness dashboard summarizes behavior, recommends actions, and adapts to inferred patterns — helpful, and structurally risky: overreach, emotional misclassification, unstable guidance. Dimensional analysis examines whether recommendation prominence implies unwarranted urgency; whether generated language makes unsupported assumptions about the user's state (an affective-affordance violation — and, if the system claims to detect mood, a Doctrine XI violation outright); whether the user can learn why a recommendation appeared; whether suggestion is distinguished from instruction; whether sensitive inferences are disclosed and correctable; whether adaptation is bounded, announced, and reversible; and

whether the dashboard preserves orientation rather than cultivating self-monitoring anxiety. Under Doctrine IX, generation operates within the state class's dimensional budget and within the user's declared profile; under the empirical record on affective AI use — heavy affective engagement with conversational systems correlating with loneliness and dependence in randomized and platform data (MIT Media Lab & OpenAI, 2025) — wellness contexts warrant the strictest budgets in the classification.

5.7 Governance Scope

Unresolved Doctrine violations trigger design review before release in emotionally sensitive, high-consequence, and AI-mediated states — regulatory harm is not ordinary iteration debt. Three governance commitments are added in this revision, each answering a documented failure pattern of prior frameworks. **Anti-gaming:** any operationalized metric invites optimization against the metric while harm migrates to unmeasured channels; audit design therefore uses partially redundant measures with held-out checks, rotated over time (§6.8). **Anti-washing:** certification without teeth accelerates the harm it claims to prevent; conformance claims require published methodology, public score rationale, revocation provisions, and adversarial re-audit rights. **Measurement horizon:** regulation-preserving variants can lose short-window engagement comparisons by construction; conformant evaluation therefore pre-declares its outcome set and windows — trust, comprehension, retention quality, support burden, regret measures over weeks, not click-through over days — so that teams adopting the Doctrine are not refuted by the wrong test. Scorecards, audit protocols, pattern libraries, and AI guardrails remain emerging operational extensions, adoptable as standards only as the validation program (§6) discharges its preconditions.

6. The Validation Program

This section replaces v3's general future-work list with a sequenced program in dependency order. The order is the argument: each study is a precondition for the authority of what follows, and the program is designed so that failure at any stage produces a specific, informative revision rather than a general embarrassment.

6.0 Design Principles

Three principles govern all studies. **Behavioral primacy:** near-term evidence is behavioral and self-report; physiological measurement enters last, only in consented research contexts, behind the Sensing Wall. **Adversarial anticipation:** studies are designed assuming motivated misuse — metric gaming, washing, cherry-picking — with redundancy and held-out measures from the start. **Participatory constitution:** the participatory track (§6.7) is not a study among studies but a governance condition running through all of them; claims made on behalf of neurodivergent users acquire validity only through it.

6.1 Study 0 — The Noise Audit (precondition for everything)

Question: Do independent trained reviewers converge on dimensional inventories and doctrine judgments? *Design:* 5–8 reviewers, trained on the protocol (§5.3); 30 interface states sampled across the sensitivity classification; both protocol stages scored; inter-rater reliability computed per dimension and per doctrine, published regardless of outcome. Reviewer panels include neurodivergent evaluators; agreement is computed within and between evaluator groups, and between-group divergence is reported as a finding about perspectival structure, not suppressed as error. *Interpretation:* dimensions or doctrines failing reliability thresholds are revised or demoted from the audit; stage-two judgments remain advisory until this study passes. Nothing else in the program — and no commercial application of the audit — carries authority before this result. *Cost and horizon:* low; months, not years.

6.2 Study 1 — Coherence Beyond Intensity (the flagship test)

Question: Does cross-dimensional conflict predict user strain beyond single-dimension intensity? *Design:* factorial manipulation of interface variants holding dimension intensities constant while varying coherence (conflict patterns per §3.8), across at least two task contexts and two sensitivity classes; outcomes: task performance, error and reversal rates, hesitation latency, abandonment, comprehension, perceived control and self-efficacy (validated instruments per §2.5), post-task strain self-report. *Interpretation:* support establishes the framework's distinctive construct; failure reduces the framework to its component literatures and is published as such. This is the study on which the theory stakes itself.

6.3 Study 2 — Psychometric Derivation of the Dimensional Structure

Question: Is the authored five-dimension taxonomy the latent structure of interface experience, or an approximation of something else? *Design:* the environmental-neuroaesthetics pipeline ported to screens (Coburn et al., 2017; 2020): large stimulus set (hundreds of interface states, sampled across genres and sensitivity classes), ratings at scale, factor decomposition, cross-validation; optionally complemented by unsupervised structure discovery over behavioral traces. *Interpretation:* the framework versions to the recovered structure. If five dimensions compress to three, or coherence emerges as a second-order factor, that is refinement, not refutation — the commitment of §3.0 is that data restructures taxonomy.

6.4 Study 3 — Cross-Cultural Calibration

Question: Which dimensional tolerances are broadly shared, and which are population-calibrated? *Design:* large-scale volunteer web experiments across countries and demographics on density tolerance, motion comfort, pacing preference, and coherence judgments — the paradigm pioneered by large cross-cultural interface-preference studies (Reinecke et al., 2013), which established both the method and the expectation of systematic variation. *Interpretation:* results populate dimensional profiles (§3.9) with empirical ranges and

directly discharge the WEIRD limitation; either outcome — universals with cultural setpoints, or deep divergence — strengthens the framework by replacing assumption with calibration.

6.5 Study 4 — Longitudinal Field Measurement (temporal rhythm first)

Question: Do dimensionally conformant interfaces reduce attention fragmentation and strain in real use over time? *Design:* field deployment comparing conformant versus conventional variants over multi-week windows using established precision-tracking methods — logging, experience sampling, switching and recovery metrics (Mark et al., 2008; Mark, 2023) — beginning with temporal rhythm as the best-parameterized dimension. *Interpretation:* cumulative-harm claims are earned here or not at all; single-session results do not license longitudinal language, in either direction.

6.6 Study 5 — Instrumented Demand Studies and the Generative Schema

Two terminal workstreams. *Instrumented demand (Tier C):* consented ambulatory studies — cardiac, electrodermal, experience sampling over real use — comparing dimensional configurations, interpreted strictly as demand expenditure, never as decoded emotion (Boehner et al., 2007), run last, when the constructs have earned instrumentation, and governed by the Sensing Wall. This is the decisive test of the physiological hypothesis (§7) and is deliberately not a precondition for any design or governance claim. *Generative constraint schema (Tier D):* the machine-readable dimensional budget specification for AI-generated interfaces (Doctrine IX) — state-class budgets, profile honoring, automatic conformance checking — developed openly, evaluated on generated-interface corpora for constraint satisfaction and downstream user outcomes, and published as an open artifact so that conformance is checkable by anyone, including the audited.

6.7 The Participatory Track (runs through everything)

Claims made on behalf of neurodivergent users are validated only through participation with constitutive power: neurodivergent and disabled co-researchers in instrument design; community review of dimensional definitions and baseline floors (does "semantic load" capture autistic linguistic processing, or a non-autistic model of it?); evaluator panels with disagreement-as-data reporting; community authority over the evolution of safety definitions made in the community's name — authority meaning the ability to change outcomes, not to advise on them. The framework's stakeholder method adopts value-sensitive design's machinery (Friedman & Hendry, 2019), including indirect stakeholders: the person affected by the user's depleted evening, the support worker downstream of the confusing form. Governance architecture — a public methodology commons, community-constitutive standard evolution, revenue arrangements that do not extract community knowledge without return — is a build-before-scale requirement, recorded here as a commitment the program can be held to.

6.8 Goodhart Resistance

Every audit measure will be optimized against once stakes attach. The program therefore maintains: partially redundant measures per construct (no single score to game); held-out checks not published in advance, rotated per audit cycle; behavioral verification behind self-report (manufactured calm is caught by the veridical/experienced split, §3.6); and periodic adversarial review in which invited critics attempt to construct interfaces that score well while violating the framework's intent — with successful attacks published and the instrument revised. An audit that cannot survive its own red team should not govern anyone's interface.

7. The Physiological Hypothesis (Tier C — explicitly quarantined)

This section states the framework's research frontier, separated so that its speculative status cannot contaminate the operational framework, which stands on behavioral evidence alone (§0, §1.3).

The hypothesis. Sustained interaction with dimensionally incoherent, rhythm-blind, or extractive interfaces levies cumulative physiological cost — allostatic load in the predictive-regulation sense (Sterling, 2012; McEwen & Akil, 2020): the body budgets for anticipated demand, uncertainty and vigilance are metabolically expensive, and environments that chronically force expectation revision, sustain unresolved conflict, or compress recovery are hypothesized to tax that budget measurably. This is a claim about *demand accounting*, not about emotions: it predicts expenditure signatures (autonomic and endocrine markers under consented study), not affective states, and it is consistent with — indeed motivated by — the constructionist position that affect is constructed while bodily budgeting is continuous (Barrett, 2017).

What the hypothesis does not claim. No universal stress response to any stimulus; no polyvagal mechanism; no emotion readable from physiology (arousal signatures are ambiguous among excitement, effort, and distress — Boehner et al., 2007); no dependence on the free energy principle, whose formal status remains contested (Bruineberg et al., 2022) and which is not load-bearing anywhere in this framework; no claim that single sessions produce clinically meaningful load.

Falsification conditions. The hypothesis is abandoned or restructured if consented ambulatory studies (§6.6) show, with adequate power: no reliable expenditure differences between coherent and incoherent configurations of matched intensity; or expenditure differences that fail to track behavioral strain measures (which would suggest the behavioral and physiological constructs dissociate and the framework's unification is wrong); or effect sizes too small to matter at

realistic exposure levels — in which case the framework's harms case rests, honestly, on behavior and experience alone.

Why carry the hypothesis at all. Because if true, it changes the harm category: design-induced demand becomes a public-health exposure with an aggregation argument — billions of mildly costly sessions daily is a population-scale exposure in the same formal class as environmental stressors — and regulation of design mechanics acquires the evidentiary structure that environmental health regulation has. The hypothesis is worth testing precisely because its confirmation would matter and its disconfirmation would discipline this framework's language. Until then, it funds no doctrine and licenses no audit finding.

8. Limitations

Validation status. UTDA is not an empirically validated model; the taxonomy is authored, pending derivation (§6.3); every numeric constraint is heuristic; stage-two audit judgments are advisory pending reliability (§6.1). The paper's authority claims are correspondingly conditional, and the claim-tier system (§1.3) exists to keep them so.

The taxonomy may be wrong in structure, not just calibration. Dimensions may fail factor analysis; coherence may not survive the factorial test; the program's design ensures such failures are informative, but they remain possible, and the framework's continuation is conditional on §6.2's result.

Cultural generalizability is unresolved. The relevant cognitive and HCI evidence over-samples WEIRD populations (Henrich et al., 2010); until §6.4 reports, all profile ranges are provisional and no configuration should be marketed as universally compatible.

The unit-of-analysis limitation. The framework evaluates interface states and sessions; significant documented harms of current systems are *relational* and *longitudinal* — parasocial attachment, dependency, displaced human contact (MIT Media Lab & OpenAI, 2025) — and dimensional properties of states cannot carry the full explanatory weight for relationship-level harm. A temporal extension (session, week, relationship horizons) is future work; claims at those horizons are not made here.

The interface boundary problem. A dimensionally impeccable interface to an unjust system remains unjust. A gig-work scheduling application can be coherent, legible, gently paced — and still an instrument of extraction, because the injustice lives in what the system does, not how it feels; a calm surface over adverse terms is anesthesia, and this framework, applied without its own Doctrine X, could be used to produce it. Dimensional evaluation is therefore necessary but not sufficient; it evaluates the environment's demands, not the legitimacy of the arrangement the environment serves, and conformance claims must never be marketed as ethical clearance of

the underlying system. The manufactured-calm provisions (§3.6) police the interface-level version of this failure; the system-level version is beyond this framework's scope and is stated as such.

Adoption physics. Doctrine conflicts with engagement-optimized business models; frameworks without launch-blocking authority become theater; short-window experiments will disfavor regulation-preserving variants by construction (§5.7). The framework's governance provisions address these; they do not dissolve them.

Authorship and standpoint. This framework was authored by one researcher with professional and lived-experience grounding but without, at this stage, the community accountability structures its own doctrine requires. The participatory track (§6.7) is the remedy, and until it has run, the framework's neurodivergence claims should be read as proposals awaiting the only validation that can legitimate them.

Measurement inheritance. Semantic load inherits cognitive-load measurement disputes; affective affordance inherits the difficulty of appraisal measurement; the framework's reliability is bounded by its weakest instrument, and §6.1 exists to find out which that is.

9. Conclusion

Digital systems increasingly shape the conditions under which people think, choose, feel, recover, and participate — and the institutions around those systems have begun, unevenly but unmistakably, to treat design mechanics as a matter of public concern. This paper has proposed a vocabulary and a research program for that concern: five stimulus-side dimensions anchored to prior constructs; coherence as a falsifiable structural property; predictability split into veridical and experienced forms so that felt calm must be earned by honest behavior; affect handled as affordance rather than delivered state; a doctrine restated in checkable form, extended by a categorical prohibition on turning evaluation into surveillance; pluralism made structural through profiles rather than gestured at in caveats; co-production made constitutive rather than acknowledged; and a validation program ordered so that the framework earns authority in the sequence it makes claims.

The central claim survives every revision made here, because it was never a claim about physiology: digital environments should not manufacture unnecessary strain, obscure it, or profit from it. What has changed is the discipline around that claim — what it asserts, at which tier, on what evidence, governed by whom. If the coherence prediction fails its test, this framework reduces to a well-organized reading of its ancestors, and that result will be published. If it holds, then interface quality has a structural property that single-axis frameworks cannot see, audit machinery has a construct worth its name, and the environments where people now live most of their waking attention acquire a design science adequate to what they have become.

Appendix A: Claim Classification Audit (Revised)

Claims are tiered per §1.3; representative examples follow.

Tier A — Behavioral-normative (defensible now). Cancellation flows with asymmetric effort and guilt-laden language obstruct user intent (enforcement-documented). Attention-capture patterns (infinite scroll, autoplay, urgency without referent) measurably extend use beyond intention (Monge Roffarello et al., 2023). Defaults presupposing high sensory tolerance impose differential burden (accessibility and neurodivergence literatures). Interfaces should not be engineered to feel safer than they behave.

Tier B — Theoretical synthesis (judged by fruitfulness, pending derivation). Interface experience can be analyzed through dimensions distributing cognitive and affective demand. Predictability emerges from dimensional coherence rather than consistency alone. Affective affordance is produced by configuration, not by copy or style alone. Coherence predicts strain beyond intensity (the flagship claim — Tier B until §6.2, then whatever the data says).

Tier C — Physiological hypothesis (quarantined, falsifiable). Sustained dimensional incoherence and extractive demand levy allostatic cost measurable in consented ambulatory study. Carries the falsification conditions of §7; funds no doctrine.

Tier D — Operational proposals (conditional on validation and governance). The audit protocol, state classification, dimensional budgets, generative constraint schema, and certification machinery — advisory until §6.1 reliability, participatory legitimacy per §6.7, and anti-gaming provisions per §6.8.

Appendix B: Glossary (Revised)

Affective affordance — The appraisals an interface makes cheap or costly through its configuration; replaces "emotional valence." Not a claim about any user's actual emotional state.

Allostatic load (as used here) — Hypothesized cumulative cost of regulation under environmental demand (Tier C); predictive-regulation sense (Sterling, 2012).

Auro Prism Doctrine — Eleven operational principles for design evaluation and governance, each with a checkable form (§4). **Capability-building demand** — Chosen, legible, paced demand whose yield accrues to the user (§3.10); contrast extractive demand. **Dimensional aesthetics** — The total perceptual, temporal, semantic, and interactional structure through which an interface is experienced. **Dimensional coherence** — The degree to which interface dimensions reinforce rather than contradict one another; operationalized as absence of specified cross-dimensional conflicts (§3.8). **Dimensional profile** — Population- and context-calibrated tolerance ranges across the five dimensions; the framework's unit of prescription (§3.9). **Interaction depth** —

Number, complexity, reversibility, and consequence of actions required to complete or understand a task. **Manufactured calm** — Divergence of experienced from veridical predictability: sedative presentational character over opaque or adverse behavior (§3.6). A failure mode, not a compliment. **Motion energy** — Quantity, intensity, speed, frequency, and behavioral role of interface movement. **Nervous-system compatibility** — Vision term (see Terminology Note): environments whose demands remain within and respectful of human regulatory capacity; technical content = regulatory burden (behavioral) + allostatic hypothesis (Tier C). **Neurodivergent safety (baseline)** — A co-produced floor condition in which sensory, cognitive, attentional, and regulatory variance is central rather than exceptional; plural by construction. **Regulatory burden** — Total demand an interface places on attention, interpretation, sensory processing, decision-making, and recovery; behavioral construct (Tier A/B), with physiological accounting hypothesized at Tier C. **Semantic load** — Interpretive effort required by language, structure, tone, and implication; structural, not word-count. **Sensing Wall** — Doctrine XI: dimensional evaluation measures artifacts and consenting study populations, never individual users' physiological or affective states in production. **Temporal rhythm** — How an interface structures pacing, interruption, waiting, response, transition, and recovery. **Veridical / experienced predictability** — Correspondence of the user's model to actual system behavior (behaviorally assessed) vs. felt stability (self-reported); Doctrine II requires the former to ground the latter. **Visual density** — Amount, complexity, and competitive intensity of visible information in an interface state.

Appendix C: Measurement Protocols (Revised)

Ordered by evidentiary role; physiological measures appear last and only under the Sensing Wall.

Behavioral measures (primary): task completion; error rate; reversal/undo frequency; abandonment; hesitation latency before high-consequence actions; repeated scanning; help-seeking; expectation-violation rates (predictability); effort symmetry (entry vs. exit paths); recovery time after interruption; return-with-regret patterns over declared windows.

Self-report measures (paired with behavior, never alone for doctrine findings): perceived cognitive load (with instrument caveats per §2.1); perceived control and self-efficacy (validated instruments, §2.5); perceived clarity; sensory comfort; felt pressure; trust; decision confidence; post-task depletion or relief. The veridical/experienced split (§3.6) requires reporting behavioral and self-report results side by side; divergence is a finding.

Expert review measures: two-stage protocol per §5.3 with inventory/judgment separation; doctrine conformance per checkable forms; conflict inventory; dark-pattern screening mapped to the ontology (Gray et al., 2024); accessibility and neurodivergent-safety review with mixed panels and disagreement reporting.

Physiological measures (Tier C research contexts only, consented, Sensing Wall applies): ambulatory cardiac and electrodermal measures with experience sampling; interpreted exclusively as demand expenditure, never as decoded emotion (Boehner et al., 2007; Barrett et al., 2019); excluded from production systems and product optimization categorically.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of CHI 2019*. ACM. <https://doi.org/10.1145/3290605.3300233>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bardzell, J., Bardzell, S., & Hansen, L. K. (2015). Immodest proposals: Research through design and knowledge. *Proceedings of CHI 2015* (pp. 2093–2102). ACM. <https://doi.org/10.1145/2702123.2702400>
- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of FAccT 2021* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4), 275–291. <https://doi.org/10.1016/j.ijhcs.2006.11.016>
- Bruineberg, J., Dołęga, K., Dewhurst, J., & Baltieri, M. (2022). The emperor's new Markov blankets. *Behavioral and Brain Sciences*, 45, e183. <https://doi.org/10.1017/S0140525X21002351>
- Calm Tech Institute. (2024). *Calm Tech certification standard*. <https://www.calmtech.institute>
- Calvo, R. A., & Peters, D. (2014). *Positive computing: Technology for wellbeing and human potential*. MIT Press.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum.

Case, A. (2015). *Calm technology: Principles and patterns for non-intrusive design*. O'Reilly Media.

Chen, J. X., McDonald, A., Zou, Y., Tseng, E., Roundy, K. A., Tamersoy, A., Schaub, F., Ristenpart, T., & Dell, N. (2022). Trauma-informed computing: Towards safer technology experiences for all. *Proceedings of CHI 2022* (pp. 1–20). ACM.
<https://doi.org/10.1145/3491102.3517475>

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Coburn, A., Vartanian, O., & Chatterjee, A. (2017). Buildings, beauty, and the brain: A neuroscience of architectural experience. *Journal of Cognitive Neuroscience*, 29(9), 1521–1531.
https://doi.org/10.1162/jocn_a_01146

Coburn, A., Vartanian, O., Kenett, Y. N., Nadal, M., Hartung, F., Hayn-Leichsenring, G., Navarrete, G., González-Mora, J. L., & Chatterjee, A. (2020). Psychological and neural responses to architectural interiors. *Cortex*, 126, 217–241.
<https://doi.org/10.1016/j.cortex.2020.01.009>

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. MIT Press. <https://doi.org/10.7551/mitpress/12255.001.0001>

Cox, A. L., Gould, S. J. J., Cecchinato, M. E., Iacovides, I., & Renfree, I. (2016). Design frictions for mindful interactions: The case for microboundaries. *CHI EA 2016* (pp. 1389–1397). ACM.
<https://doi.org/10.1145/2851581.2892410>

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Ellard, C. (2015). *Places of the heart: The psychogeography of everyday life*. Bellevue Literary Press.

European Commission. (2026). *Preliminary findings on TikTok's compliance with the Digital Services Act: Addictive design and protection of minors*. [Enforcement proceeding; pending final decision.]

Federal Trade Commission. (2025). *FTC v. Amazon.com, Inc.: Stipulated order on Prime enrollment and cancellation practices*. [Settlement, September 2025.]

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65(7), 999–1010. <https://doi.org/10.3758/BF03194829>

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138. <https://doi.org/10.1038/nrn2787>

Gershman, S. J. (2019). What does the free energy principle tell us about the brain? *arXiv*. <https://arxiv.org/abs/1901.07945>

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. *Proceedings of CHI 2018*. ACM. <https://doi.org/10.1145/3173574.3174108>

Gray, C. M., Santos, C., Bielova, N., & Mildner, T. (2024). An ontology of dark patterns knowledge: Foundations, definitions, and a pathway for shared knowledge-building. *Proceedings of CHI 2024*. ACM. <https://doi.org/10.1145/3613904.3642436>

Hallnäs, L., & Redström, J. (2001). Slow technology: Designing for reflection. *Personal and Ubiquitous Computing*, 5, 201–212. <https://doi.org/10.1007/PL00000019>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>

International Organization for Standardization. (2019). *ISO 9241-210:2019 — Human-centred design for interactive systems*. ISO.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60(3), 269–298. <https://doi.org/10.1016/j.ijhcs.2003.09.002>

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., & Shadbolt, N. (2019). Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. *Proceedings of CHI 2019*. ACM. <https://doi.org/10.1145/3290605.3300361>

Mack, K., McDonnell, E., Jain, D., Wang, L. L., Froehlich, J. E., & Findlater, L. (2021). What do we mean by "accessibility research"? *Proceedings of CHI 2021*. ACM. <https://doi.org/10.1145/3411764.3445412>

Mark, G. (2023). *Attention span: A groundbreaking way to restore balance, happiness and productivity*. Hanover Square Press.

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: More speed and stress. *Proceedings of CHI 2008* (pp. 107–110). ACM. <https://doi.org/10.1145/1357054.1357072>

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), Article 81. <https://doi.org/10.1145/3359183>

Mathur, A., Mayer, J., & Kshirsagar, M. (2021). What makes a dark pattern... dark? *Proceedings of CHI 2021*. ACM. <https://doi.org/10.1145/3411764.3445610>

McEwen, B. S., & Akil, H. (2020). Revisiting the stress concept: Implications for affective disorders. *Journal of Neuroscience*, 40(1), 12–21. <https://doi.org/10.1523/JNEUROSCI.0733-19.2019>

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>

Milton, D. E. M. (2012). On the ontological status of autism: The double empathy problem. *Disability & Society*, 27(6), 883–887. <https://doi.org/10.1080/09687599.2012.710008>

Milton, D., Gurbuz, E., & López, B. (2022). The "double empathy problem": Ten years on. *Autism*, 26(8), 1901–1903. <https://doi.org/10.1177/13623613221129123>

Miniukovich, A., & De Angeli, A. (2015). Computation of interface aesthetics. *Proceedings of CHI 2015* (pp. 1163–1172). ACM. <https://doi.org/10.1145/2702123.2702575>

MIT Media Lab & OpenAI. (2025). *Investigating affective use and emotional wellbeing on ChatGPT* [randomized controlled study and platform analysis]. MIT Media Lab / OpenAI. [Working paper, March 2025.]

Monge Roffarello, A., Lukoff, K., & De Russis, L. (2023). Defining and identifying attention capture deceptive designs in digital interfaces. *Proceedings of CHI 2023*. ACM. <https://doi.org/10.1145/3544548.3580729>

Monge Roffarello, A., De Russis, L., & Lukoff, K. (2025). Digital attention heuristics: Supporting user attention in interface design. *ACM Transactions on Computer-Human Interaction*.
<https://doi.org/10.1145/3725215>

Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689–709. <https://doi.org/10.1016/j.ijhcs.2010.05.006>

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15(1), 3–11.
[https://doi.org/10.1016/0169-8141\(94\)00052-5](https://doi.org/10.1016/0169-8141(94)00052-5)

Norman, D. A. (2013). *The design of everyday things* (rev. ed.). Basic Books.

Odom, W., Banks, R., Durrant, A., Kirk, D., & Pierce, J. (2012). Slow technology: Critical reflection and future directions. *Proceedings of DIS 2012* (pp. 816–817). ACM.
<https://doi.org/10.1145/2317956.2318088>

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (G. V. Anrep, Trans.). Oxford University Press. [Experimental neurosis: Lecture XVIII.]

Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology*, 9, 797.
<https://doi.org/10.3389/fpsyg.2018.00797>

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. *Proceedings of CHI 2013* (pp. 2049–2058). ACM.
<https://doi.org/10.1145/2470654.2481281>

SAMHSA. (2014). *SAMHSA's concept of trauma and guidance for a trauma-informed approach* (HHS Publication No. SMA14-4884). U.S. Department of Health and Human Services.

Scott, C. F., Marcu, G., Anderson, R. E., Newman, M. W., & Schoenebeck, S. (2023). Trauma-informed social media: Towards solutions for reducing and healing online harm. *Proceedings of CHI 2023*. ACM. <https://doi.org/10.1145/3544548.3581512>

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
<https://doi.org/10.1093/oso/9780192845290.001.0001>

Spiel, K., Frauenberger, C., Keyes, O., & Fitzpatrick, G. (2019). Agency of autistic children in technology research: A critical literature review. *ACM Transactions on Computer-Human Interaction*, 26(6), Article 38. <https://doi.org/10.1145/3344919>

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, 106(1), 5–15. <https://doi.org/10.1016/j.physbeh.2011.06.004>

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. <https://doi.org/10.1007/s10648-019-09465-5>

Vanden Abeele, M. M. P. (2021). Digital wellbeing as a dynamic construct. *Communication Theory*, 31(4), 932–955. <https://doi.org/10.1093/ct/qtaa024>

W3C. (2023). *Web Content Accessibility Guidelines (WCAG) 2.2*. World Wide Web Consortium.

W3C Cognitive and Learning Disabilities Accessibility Task Force. (2021). *Making content usable for people with cognitive and learning disabilities*. W3C Working Group Note. <https://www.w3.org/TR/coga-usable/>

Weidinger, L., Uesato, J., Rauh, M., et al. (2022). Taxonomy of risks posed by language models. *Proceedings of FAccT 2022* (pp. 214–229). ACM. <https://doi.org/10.1145/3531146.3533088>

Weiser, M., & Brown, J. S. (1996). *The coming age of calm technology*. Xerox PARC.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. *Proceedings of CHI 2007* (pp. 493–502). ACM. <https://doi.org/10.1145/1240624.1240704>

Citation integration note (carried from v3, updated): DOIs and publisher details require verification against a citation manager before external submission. The European Commission (2026), FTC (2025), and MIT Media Lab & OpenAI (2025) entries reference enforcement records and a working paper whose canonical citation forms should be confirmed at submission time. The Monge Roffarello et al. (2025) TOCHI title should be verified against the published version.

ADDENDUM — The Canon Edition Mapping

The academic paper above is the core. The Auro Prism Canon edition — the charter document in the voice of Volumes I–III — carries everything above plus the layers a journal cannot hold. This addendum specifies what the Canon edition adds, so both texts can be maintained from this single source without drift.

Voice and register. The Canon edition restores the doctrine's full rhetorical register — the short-line cadence of Volumes I–III — for the doctrine statements themselves, while keeping this paper's checkable forms as each doctrine's "operational face." Format: Canon voice states the principle; a boxed operational clause carries the checkable form verbatim from §4. The two must never diverge in substance; the checkable forms are canonical.

Volume linkages. Each doctrine cross-references its Canon ancestry: Doctrines I–VIII to Volume II's First Principles and Sacred Constraints (dignity before optimization; agency before persuasion; transparency before manipulation); Doctrine IX to Volume II's AI Doctrine chapters (explainability, accountability, reversibility, human authority); Doctrine X to "Ethics Is Structural" and Volume II's Red Lines; Doctrine XI is new to the Canon and enters Volume II's Sacred Constraints as a red line ("What Auro Prism Refuses" gains: inferring the inner states of the people it serves). The Pulse framework (Volume III, Part I) is annotated as the organizational-scale sibling of the dimensional audit — Pulse listens at the level of institutions what the audit measures at the level of interfaces — and Pulse's Sacred/Structural/Operational layering is mapped to the claim tiers (Sacred ≈ Tier A normative commitments; Structural ≈ Tier D governance; Operational ≈ audit practice).

Pattern library integration. Volume III's thirteen anti-patterns are annotated with their dimensional signatures — e.g., *Care Theater* as an affective-affordance/veridicality divergence (the organizational form of manufactured calm); *False Alignment* as coherence failure at the social layer; *Explanation Theater* as the transparency analogue of manufactured calm (Doctrine VIII's named enemy); *Captive Loyalty* and *Forced Intimacy* as extractive-demand patterns at relationship scale, flagging the unit-of-analysis limitation (§8) as Canon future work. The Canon edition may make these linkages richly; the academic paper deliberately does not claim them.

What the Canon edition keeps that the paper omits. The meta-principle in full ("We design systems worthy of human beings"); the Steward's Oath as the practitioner's commitment ceremony for audit training; the "future is a stakeholder" horizon (multi-lifespan framing) as the Canon's answer to the paper's product-cycle scope; worked reflections and repair pathways in the Volume III style.

What the Canon edition must also carry from this revision — no exceptions. The Terminology Note (§0) in full; the claim-tier system; the Sensing Wall; the co-production governance commitments; and the anti-washing provisions. The Canon's moral register heightens rather than loosens the epistemic discipline: doctrine without the tier system is exactly the "ideology to buyers, vibes to scientists" failure the review warned against, and the Canon edition's credibility depends on carrying its own constraints visibly.

Maintenance rule. The academic core is the source of truth for constructs, claims, and checkable forms. Canon prose may exceed it in voice, never in claim strength. When validation results (§6) revise a construct, both editions version together, and the Summary of Revisions genre established in this draft becomes the changelog mechanism for both.

End of Revised Draft v4.