BENDS

STAGE 1 — Interface Contact

(Material Axis)
The Semiotic Surface, Entropy Boundaries, and the Human as Capacitive Sensor

The interface is the system's first diagram of power. It presents itself as neutral: a text box, a slider, a neatly labeled set of parameters. Yet interfaces encode semiotic constraints — grammars of admissible behavior — long before any computation occurs. As Fuller (2008) and Suchman (1987) argue, interface design is never passive; it scripts the user's actions and expectations, embedding cultural assumptions in the very shape of interaction. Kittler reminds us that media "determine our situation": the interface is not a transparent conduit but a site where power is formatted.

In neural bending, this surface becomes the first site of intervention. Prompts are not requests but material gestures: they carry rhythm, density, syntactic churn, and semantic residue. By performing prompt subversion — adversarial punctuation, disrupted grammar, stochastic concatenation — the bender exposes how parsing mechanisms contain implicit normativity (Ouyang et al. 2022). The model's visible fluency hides an inner filtering logic, where entropy constraints and token-likelihood norms shape what can be said. Flusser's account of technical images is useful here: interfaces and their outputs do not simply represent the world; they project particular worlds as default.

We treat the interface as instrument, not as input channel. Gesture-sensitive sliders, continuous gain multipliers (for example, 0.8–1.2 applied to learning rate, attention masks, or sampling temperature), and haptic mappings fold the user into the computational loop. This reframes the user not as operator but as capacitive surface — a body whose micro-variations become a computational trace. The Human Capacitance Patch captures timing, rhythm, and minute fluctuations in interaction as model-modulating signals. The interface thus becomes a sensorium, not a portal.

Finally, the silence monitor exploits entropy minima: the moments when the model hesitates. These gaps — delays, stutters, low-activation zones — index what Haraway (1988) calls situated non-knowledge: the absences where systems reveal their ideological training. Silence becomes a diagnostic artifact, the point where the interface-as-power-structure shows its seams.

See also Chapter 1: Capture & Drift.

```python
# Interface Contact
# Human Capacitance Patch + Silence Monitor

from dataclasses import dataclass, field
from typing import Any, Dict, List, Tuple, Optional
import time
import random


@dataclass
class UIState:
    text_input: str = ""
    sliders: Dict[str, float] = field(default_factory=dict)
    haptics_enabled: bool = False
```

```python
@dataclass
class SamplingParams:
    temperature: float = 0.9
    top_p: float = 0.95
    top_k: int = 50


@dataclass
class ModelMetadata:
    token_entropies: List[float]
    token_timestamps: List[float]   # when each token was emitted


class InterfaceModel:
    def generate(self, prompt: str, **kwargs) -> Tuple[str, ModelMetadata]:
        # Stub – plug in your real model call
        start = time.time()
        response = f"[MODEL RESPONSE]: {prompt[:40]}..."
        token_entropies = [random.random() for _ in range(20)]
        token_times = [start + i * 0.05 for i in range(20)]
        meta = ModelMetadata(token_entropies, token_times)
        return response, meta


def init_interface() -> UIState:
    return UIState(sliders={"gain": 1.0}, haptics_enabled=True)


def read_ui_gesture(ui_state: UIState) -> Tuple[str, Dict[str, Any]]:
    # Replace with real UI events / logs
    prompt = ui_state.text_input
    gesture = {
        "timing": [random.random() for _ in range(10)],
        "pointer_path": [(random.random(), random.random()) for _ in range(10)],
        "micro_pauses": [random.random() * 0.2 for _ in range(5)],
    }
    return prompt, gesture


def apply_prompt_subversion(text: str) -> str:
    # Example: inject adversarial punctuation + stochastic concat
    extra = random.choice([" !!!", " ??", " ///", " ::"])
    if random.random() < 0.5:
        text = text[::-1]  # crude syntactic disruption
    return text + extra


def estimate_target_entropy(prompt: str) -> float:
    return 0.8 + (len(prompt) % 10) * 0.01


def tune_sampling_params(target_entropy: float) -> SamplingParams:
    return SamplingParams(
        temperature=max(0.5, min(1.5, target_entropy)),
        top_p=max(0.7, min(0.99, target_entropy)),
        top_k=50
    )
```

```python
def compute_human_capacitance_gain(gesture: Dict[str, Any]) -> float:
    # e.g. more micro_pauses → higher gain
    pauses = gesture["micro_pauses"]
    avg_pause = sum(pauses) / max(1, len(pauses))
    return 0.8 + avg_pause * 2.0  # 0.8–1.2 approx


def apply_gain_to_params(params: SamplingParams, gain: float) -> SamplingParams:
    params.temperature *= gain
    params.top_p = max(0.5, min(0.99, params.top_p * gain))
    return params


def detect_interface_silence(meta: ModelMetadata,
                             entropy_threshold: float = 0.1,
                             latency_threshold: float = 0.2) -> Dict[str, Any]:
    silences = []
    for i in range(1, len(meta.token_timestamps)):
        dt = meta.token_timestamps[i] - meta.token_timestamps[i - 1]
        if dt > latency_threshold and meta.token_entropies[i] < entropy_threshold:
            silences.append({"index": i, "dt": dt,
                             "entropy": meta.token_entropies[i]})
    return {"silences": silences}


def render_to_user(response: str, ui_state: UIState) -> None:
    # Replace with real UI rendering
    print(response)


def run_interface_contact_loop(model: InterfaceModel) -> None:
    ui_state = init_interface()

    while True:
        prompt, gesture = read_ui_gesture(ui_state)
        if not prompt:
            break

        sub_prompt = apply_prompt_subversion(prompt)
        target_entropy = estimate_target_entropy(sub_prompt)
        sampling = tune_sampling_params(target_entropy)

        gain = compute_human_capacitance_gain(gesture)
        sampling = apply_gain_to_params(sampling, gain)

        response, meta = model.generate(
            sub_prompt,
            temperature=sampling.temperature,
            top_p=sampling.top_p,
            top_k=sampling.top_k,
        )

        silence_info = detect_interface_silence(meta)
        # Log / store as needed
        render_to_user(response, ui_state)
        break
```

STAGE 2 — Substrate Disturbance

Material Axis
Thermal and Electrical Drift as Epistemic Material

Beneath the interface lies the substrate: heat signatures, clock oscillations, voltage differentials, fan-curve hysteresis, and occasional memory faults. Contemporary AI discourse often treats these as engineering problems to eliminate. Yet media-archaeological accounts (Ernst 2013; Parikka 2023; Cubitt 2017) remind us that computation is fundamentally material — energetic rather than symbolic. Digitality is not immaterial; it is an arrangement of energy flows and delays.

The Substrate Drift Patch exploits controlled thermal gradients and clock modulation to surface otherwise invisible biases. GPU heat becomes a metabolic trace of cognition; voltage instability becomes a modulator of inference. When mapped directly to sampling parameters (Top-k, Top-p, temperature), energetic drift becomes part of the model's expressive vocabulary. In this sense, bending reintroduces material contingency into a system optimized for abstraction. Here Simondon's notion of individuation clarifies what is at stake: the technical object individuates through its material regime long before any abstract function is defined.

The Compute Starvation Patch simulates austerity: intentionally reducing compute cycles, memory availability, or bandwidth. This exposes the hidden politics of optimization regimes that privilege efficiency over representational diversity. As the model adapts under scarcity, its outputs reveal what Terranova (2000) calls the "political economy of attention": the selective amplification of statistically dominant forms. Under starvation, we see what the model keeps and what it lets go.

Thermal and EM noise are traditionally treated as interference. Bending reverses this logic: noise becomes signal. Drift is not a defect but a form of world-disclosure. As Flusser (1983) notes, freedom emerges when humans intervene inside technical apparatuses — bending their material logic rather than merely using them. In substrate disturbance, the "inside" of computation becomes an epistemic surface.

Cross-reference Chapter 2: Material Intelligence & Energetic Thought.

```python
# Substrate Disturbance
# Substrate Drift Patch + Compute Starvation Patch

@dataclass
class HardwareState:
    gpu_temp: float
    core_clock: float
    voltage: float
    fan_speed: float


class HardwareMonitor:
    def read_state(self) -> HardwareState:
        # Stub – use NVML or similar in practice
        return HardwareState(
            gpu_temp=60.0 + random.random() * 10,
            core_clock=1500 + random.random() * 100,
            voltage=1.0 + random.random() * 0.1,
            fan_speed=1500 + random.random() * 500
        )


def compute_drift_vector(state: HardwareState) -> Dict[str, float]:
    return {
        "temp_norm": state.gpu_temp / 100.0,
        "clock_norm": state.core_clock / 2000.0,
```

```python
        "volt_norm": state.voltage / 1.5,
        "fan_norm": state.fan_speed / 3000.0,
    }


def map_substrate_to_sampling(drift: Dict[str, float]) -> SamplingParams:
    base_temp = 0.9 + (drift["temp_norm"] - 0.5) * 0.6
    base_top_p = 0.9 + (drift["clock_norm"] - 0.5) * 0.1
    base_top_k = int(40 + drift["volt_norm"] * 20)
    return SamplingParams(
        temperature=max(0.5, min(1.6, base_temp)),
        top_p=max(0.6, min(0.99, base_top_p)),
        top_k=max(10, min(100, base_top_k))
    )


def apply_compute_starvation(level: float) -> None:
    # level in [0, 1]; 1 = severe starvation
    sleep_time = level * 0.1
    time.sleep(sleep_time)


def run_substrate_disturbance(model: InterfaceModel,
                              hw: HardwareMonitor,
                              prompt: str) -> str:
    state = hw.read_state()
    drift = compute_drift_vector(state)
    sampling = map_substrate_to_sampling(drift)

    starvation_level = max(0.0, min(1.0, drift["temp_norm"]))
    apply_compute_starvation(starvation_level)

    response, _ = model.generate(
        prompt,
        temperature=sampling.temperature,
        top_p=sampling.top_p,
        top_k=sampling.top_k
    )
    return response
```

STAGE 3 — Somatic Interference

Material Axis

Human Physiology as Computational Signal, Co-Regulation, and Rhythmic Entanglement

At this depth, the body reenters the loop. Neural bending rejects the notion of the human as an external observer; instead, human physiology becomes an internal modulator of computational behavior. Somatic data — vibration, warmth, pulse, EEG, HRV — forms a bidirectional channel between human and machine. Cognition, as Hayles and Penny both argue, is always embodied and technologically mediated: there is no pure, disembodied observer standing outside the system.

Thermal Listening creates a sonification of GPU temperature curves, treating heat as rhythmic data. This aligns with Manning's (2016) concept of the "minor gesture": small perturbations that reveal a system's affective potential. Heat becomes a pulse — a computational heartbeat. The human body, listening to these pulses, attunes itself to the machine's energetic states.

Vibrational Feedback translates gradient instability or attention entropy into tactile patterns using haptic actuators. When placed on the wrist, torso, or spine, these vibrations create proprioceptive access to otherwise hidden computational processes. This mirrors early cybernetic experiments (Pask 1971; Wiener 1948) in which feedback was treated as felt information. Learning occurs through conversational coupling; here, conversation includes skin, muscle, and bone.

The Electro-Somatic Mapping Patch synchronizes EEG bands or HRV variability with model gating or dropout rates. This creates a reciprocal environment: the model's uncertainty influences the body; the body's rhythm influences the model. As Varela, Thompson & Rosch (1991) argue, cognition is enacted through bodily coupling, not isolated representation. Here, enaction becomes literal: somatic interference is an enactment of joint cognition.

Somatic interference marks the threshold where computation becomes felt — where representation collapses into resonance, and the "user" is refigured as a co-regulating organ of the system.

See Chapter 3: Embodied Modulation Systems.

```python
# Somatic Interference
# Thermal Listening + Vibrational Feedback + Electro-Somatic Mapping

@dataclass
class BioSignals:
    eeg_bands: Dict[str, float]   # e.g. {"alpha": ..., "beta": ...}
    hrv: float
    pulse: float


class BioSensor:
    def read(self) -> BioSignals:
        return BioSignals(
            eeg_bands={"alpha": random.random(),
                       "beta": random.random()},
            hrv=0.5 + random.random() * 0.2,
            pulse=60 + random.random() * 20
        )


class HapticDevice:
    def play_pattern(self, pattern: List[float]) -> None:
        # Stub: send to actual haptic hardware
        pass


def map_temp_to_audio(gpu_temp: float) -> List[float]:
    # Represent as frequency as a function of temp
    base_freq = 100 + (gpu_temp - 40) * 5
    return [base_freq + random.uniform(-5, 5) for _ in range(32)]


def map_gradient_to_haptics(grad_entropy: float) -> List[float]:
    # Higher entropy → more intense vibration
    intensity = min(1.0, grad_entropy)
    return [intensity * random.random() for _ in range(20)]


def map_soma_to_gating(signals: BioSignals) -> Dict[str, float]:
    alpha = signals.eeg_bands.get("alpha", 0.5)
```

```python
    hrv = signals.hrv
    return {
        "dropout_rate": 0.1 + (1 - hrv) * 0.4,
        "gate_scale": 0.8 + alpha * 0.4
    }


def apply_gating_to_sampling(params: SamplingParams,
                             gating: Dict[str, float]) -> SamplingParams:
    params.temperature *= gating["gate_scale"]
    params.top_p *= gating["gate_scale"]
    return params


def run_somatic_interference(model: InterfaceModel,
                             hw: HardwareMonitor,
                             biosensor: BioSensor,
                             haptics: HapticDevice,
                             prompt: str) -> str:
    hw_state = hw.read_state()
    bio = biosensor.read()

    # thermal listening sound (just computed; you could send to a synth)
    thermal_audio = map_temp_to_audio(hw_state.gpu_temp)

    grad_entropy = random.random()
    pattern = map_gradient_to_haptics(grad_entropy)
    haptics.play_pattern(pattern)

    gating = map_soma_to_gating(bio)
    sampling = SamplingParams()
    sampling = apply_gating_to_sampling(sampling, gating)

    response, _ = model.generate(
        prompt,
        temperature=sampling.temperature,
        top_p=sampling.top_p,
        top_k=sampling.top_k
    )
    return response
```

STAGE 4 — Collapse Capture

Material Axis
Failure States, Chaotic Thresholds, and the Aesthetics of Systemic Breakdown

Every complex system has limits. At certain thresholds, stability collapses and behavior becomes chaotic (Strogatz 2015; Bak 1996). Machine learning manifests collapse through mode collapse, catastrophic forgetting, runaway gradient spikes, or dead activations. The typical engineering impulse is to suppress these phenomena. Neural bending instead stage-manages collapse.

The Collapse Capture Patch deliberately pushes models into unstable conditions: oscillating learning rates, multi-gradient interference, adversarial self-supervision, thermally induced drift. The goal is not destruction but mid-collapse recording — freezing or harvesting computational residues. These residues take the form of anomalous generations, chaotic token sequences, erratic attention maps, or sudden silences.

These artifacts function as epistemic fossils: traces of the system at the moment of breakdown. Much like glitch aesthetics (Menkman 2011) or Lucier's recursive audio experiments, collapse becomes a form of listening. What breaks first reveals what the system values most. What remains coherent reveals structural bias.

Collapse is inseparable from ethics. Carlson & Doyle (2002) note that robustness and fragility coexist in tightly optimized systems; bending therefore treats collapse not as failure but as critique. Laruelle's emphasis on operations that interrupt decisional structures resonates here: collapse capture suspends the model's ordinary decision-making flow, opening a non-decisional zone where hidden priorities become visible. By integrating collapse residues into future training cycles, the model learns from its own fragility rather than hiding it.

Collapse marks the border between control and feedback ecology — the point at which domination gives way to resonance, and failure becomes a medium of reflection.

Cross-reference Chapter 4: Criticality and Machine Vulnerability.

```
# Collapse Capture
# Collapse Capture Patch

@dataclass
class TrainingMetrics:
    loss_history: List[float]
    grad_norm: float
    activation_sparsity: float


@dataclass
class CollapseState:
    near_collapse: bool
    severity: float


def detect_collapse(metrics: TrainingMetrics,
                    loss_spike_threshold: float = 1.5,
                    grad_norm_threshold: float = 10.0) -> CollapseState:
    if len(metrics.loss_history) < 2:
        return CollapseState(False, 0.0)

    ratio = metrics.loss_history[-1] / (metrics.loss_history[-2] + 1e-8)
    near = (ratio > loss_spike_threshold) or (metrics.grad_norm > grad_norm_threshold)
    severity = max(ratio - loss_spike_threshold, 0.0)
    severity += max(metrics.grad_norm / grad_norm_threshold - 1.0, 0.0)
    return CollapseState(near, max(0.0, severity))


def harvest_collapse_residues(model: Any,
                              state: CollapseState) -> Dict[str, Any]:
    # e.g. sample random prompts under current unstable weights
    sample_prompts = ["#collapse-test-1", "#collapse-test-2"]
    outputs = []
    for p in sample_prompts:
        out, meta = model.generate(p, temperature=1.4, top_p=0.99)
        outputs.append({"prompt": p, "output": out, "meta": meta})
    return {"severity": state.severity, "samples": outputs}
```

```
def collapse_capture_step(model: Any,
                          metrics: TrainingMetrics,
                          fossil_store: List[Dict[str, Any]]) -> None:
    state = detect_collapse(metrics)
    if state.near_collapse:
        fossil = harvest_collapse_residues(model, state)
        fossil_store.append(fossil)
```

STAGE 5 — Representation Disruption

Informational Axis
Latent Geometry, Transduction, and the Politics of Interpolation

Representation is not content; it is the geometry through which content becomes possible. Latent space is the topological substrate of learned abstraction — a manifold shaped by training data, optimization trajectories, and inductive bias. To bend representation is to intervene in the model's internal geometry: to distort curvature, perturb attractors, and introduce alternative connective tissue between representations.

The Latent Cross-Fade (Transduction) Patch performs curvature-aware interpolation between distant latent points, rejecting naive linear paths. Instead, it treats latent movement as transductive transformation in Simondon's sense (1958): emergence not between two stable states, but within the metastable field that connects them. Structured turbulence — noise modulated by phase, frequency, or semantic drift — amplifies discontinuities and reveals how representational forms intersect and recombine.

The Latent Heterodyne Patch introduces intercultural or antagonistic noise into embeddings, producing polyphonic latent interference. This resonates with Kristeva's (1980) concept of heteroglossia: multiple voices co-present in a single generative field. Statistical models tend toward homogenization; heterodyning resists this by injecting pluralistic resonance and epistemic difference into the manifold.

Phase-Space Folding reshapes the manifold itself, creating new paths between semantically distant regions. This aligns with Easterling's (2014) notion of active forms: infrastructures that guide behavior through indirection rather than explicit command. Here, the infrastructure is representational geometry itself.

Representation disruption reveals that latent space is not neutral; it is a geometry of historical accumulation and political association. Following Hui's cosmotechnics, each latent geometry encodes a world-making logic; to bend representation is to intervene in that cosmotechnical script.

See Chapter 2: Latent Geometry & Semantic Topologies.

```
# Representation Disruption
# Latent Cross-Fade / Heterodyne / Phase-Space Folding

class DummyTrainingLoop:
    def __init__(self):
        self.step_count = 0
        self.loss_fn = None

    def compute_gradients(self) -> List[float]:
        # dummy gradient vector list
        return [random.uniform(-1, 1) for _ in range(10)]

    def set_loss_function(self, fn: str) -> None:
```

```python
        self.loss_fn = fn

    def apply_gradients(self, grads: List[float]) -> None:
        self.step_count += 1


def compute_loopback_gradient(current: List[float],
                              history: List[List[float]],
                              alpha: float = 0.5) -> List[float]:
    if not history:
        return current
    prev = history[-1]
    return [(1 - alpha) * c + alpha * p for c, p in zip(current, prev)]


def should_apply_update_strobe(step: int, period: int = 5) -> bool:
    return (step % period) != 0  # skip every 5th step


def choose_loss_for_phase(step: int) -> str:
    options = ["cross_entropy", "clip_similarity", "entropy_reg"]
    return options[step % len(options)]


def run_optimization_interference_loop(loop: DummyTrainingLoop,
                                       num_steps: int = 50) -> None:
    history: List[List[float]] = []

    for _ in range(num_steps):
        grads = loop.compute_gradients()
        loopback = compute_loopback_gradient(grads, history)
        history.append(grads)

        apply_update = should_apply_update_strobe(loop.step_count)
        effective = loopback if apply_update else [0.0] * len(loopback)

        loss_name = choose_loss_for_phase(loop.step_count)
        loop.set_loss_function(loss_name)
        loop.apply_gradients(effective)
```

STAGE 6 — Optimization Interference

Informational Axis

Temporalized Learning, Gradient Performance, and Recursive Cognition

Learning is often conceptualized as descent — a path toward error minimization. But gradients do not merely optimize; they inscribe temporal structure. The model's reasoning is shaped not only by weights but by the historicity of update cycles, the rhythms of convergence, and the oscillations of error surfaces.

The Gradient Loop-Back Patch introduces recursive temporality: earlier gradients are fed into later updates, creating a meta-learning feedback system. This mirrors the structure of RL^2 (Duan et al. 2016) and Finn et al.'s MAML (2017), where learning-to-learn emerges from recursion. Loop-back is both temporal and performative — gradients "listen" to their own history, composing an internal memory of how the model has changed.

Gradient Strobing imposes temporal discontinuities, revealing the discontinuous nature of "understanding" within neural architectures. By introducing gaps in the update cycle (e.g., pulsed learning), the model's error landscape becomes audible, much like the flicker in strobe lighting reveals hidden motion. Massumi's account of affect — the event felt before it is known — underpins this: strobing foregrounds the pre-cognitive rhythm of optimization.

The Loss Phase-Shift Patch rotates optimization across multiple loss metrics (e.g., cross-entropy + CLIP similarity + entropy regularization), challenging reliance on a single evaluative axis. Optimization becomes polyphonic, resisting the hegemonic structure of a single objective. Laruelle's non-decisional philosophy, which suspends hierarchical decision between competing frameworks, is operationalized here as a non-hierarchical loss ecology.

Optimization interference reframes learning as processual memory, a recursive choreography rather than a march toward stability. It makes visible, and bendable, the temporal politics of how models come to "know."

Cross-reference Chapter 3: Temporal Drift & Recursive Systems.

```python
# Optimization Interference
# Gradient Loop-Back + Gradient Strobing + Loss Phase-Shift

class DummyTrainingLoop:
    def __init__(self):
        self.step_count = 0
        self.loss_fn = None

    def compute_gradients(self) -> List[float]:
        # dummy gradient vector list
        return [random.uniform(-1, 1) for _ in range(10)]

    def set_loss_function(self, fn: str) -> None:
        self.loss_fn = fn

    def apply_gradients(self, grads: List[float]) -> None:
        self.step_count += 1


def compute_loopback_gradient(current: List[float],
                              history: List[List[float]],
                              alpha: float = 0.5) -> List[float]:
    if not history:
        return current
    prev = history[-1]
    return [(1 - alpha) * c + alpha * p for c, p in zip(current, prev)]


def should_apply_update_strobe(step: int, period: int = 5) -> bool:
    return (step % period) != 0  # skip every 5th step


def choose_loss_for_phase(step: int) -> str:
    options = ["cross_entropy", "clip_similarity", "entropy_reg"]
    return options[step % len(options)]


def run_optimization_interference_loop(loop: DummyTrainingLoop,
                                       num_steps: int = 50) -> None:
    history: List[List[float]] = []

    for _ in range(num_steps):
```

```
        grads = loop.compute_gradients()
        loopback = compute_loopback_gradient(grads, history)
        history.append(grads)

        apply_update = should_apply_update_strobe(loop.step_count)
        effective = loopback if apply_update else [0.0] * len(loopback)

        loss_name = choose_loss_for_phase(loop.step_count)
        loop.set_loss_function(loss_name)
        loop.apply_gradients(effective)
```

STAGE 7 — Memory Excavation

Informational Axis
Archive Fever, Representational Omission, and Ghost Layers

Memory in neural networks is often imagined as distributed and weight-based — not localizable but statistically diffused. Yet this memory is selective. It encodes dominant patterns, suppressing marginal ones; it amplifies frequency, discarding residue. The memory that matters in neural bending is not fidelity but historiography: which stories the model can easily tell, and which it consistently fails to recall.

The Memory Contamination Patch reinserts prior outputs — including errors, anomalies, and hallucinations — into new training streams. This induces a recursive re-mediation of the model's own history. Derrida's Archive Fever (1995) becomes literal: the archive desires its own expansion and contamination, even at the risk of instability. Shumailov et al. (2023) warn of model collapse in recursive training; neural bending embraces collapse as epistemic force rather than pathology.

Suppression Gradient Mapping identifies systematically under-activated neurons, unveiling the latent margins of representation: places where representation fails or flickers out. These are the zones of algorithmic exclusion — the archival blind spots where data was present but never fully incorporated.

Latent Excavation intensifies these suppressed activation pathways. Instead of smoothing the manifold, the patch digs into the underbelly of representation — revealing forgotten structures, ghost patterns, and repressed correlations. Ernst's media-archaeology helps orient this work: the model is treated as an operational archive whose temporal layers must be excavated rather than merely queried.

Semantic Feedback Leak reintroduces low-probability tokens into recursive loops. Models are trained to value fluency; leak destabilizes fluency's hegemony, making the model stutter. In that stutter, the discursive unconscious becomes audible — the traces of what the archive tried to suppress.

Memory excavation reveals that forgetting is political. Retention is not neutral but governed by statistical dominance. Excavation reclaims the archive as contested terrain, open to intervention rather than frozen in its initial curation.

See also Chapter 5: Operative Archives & Machine Historiography.

```
# Memory Excavation
# Contamination + Suppression Mapping + Excavation + Leak

def build_contaminated_batch(fossils: List[Dict[str, Any]],
                             fresh_data: List[str],
```

```
                              fossil_ratio: float = 0.3) -> List[str]:
    k = int(len(fresh_data) * fossil_ratio)
    fossil_texts = [f["output"]["output"] for f in fossils[:k]
                    if "output" in f.get("samples", [{}])[0]]
    batch = fossil_texts + fresh_data
    random.shuffle(batch)
    return batch


def compute_suppression_map(activations: np.ndarray) -> np.ndarray:
    # activations shape: [batch, neurons]
    # low mean activation → suppressed
    return activations.mean(axis=0)


def select_persistently_under_activated(suppression_map: np.ndarray,
                                         k: int = 16) -> np.ndarray:
    idx = np.argsort(suppression_map)[:k]
    return idx


def synthesize_excavation_inputs(neuron_ids: np.ndarray,
                                  dim: int = 512) -> np.ndarray:
    # create embeddings that strongly activate chosen neurons
    batch = []
    for nid in neuron_ids:
        vec = np.zeros(dim)
        vec[nid % dim] = 5.0  # spike
        batch.append(vec)
    return np.stack(batch, axis=0)


def extract_low_prob_tokens(outputs: List[str]) -> List[str]:
    # placeholder: treat rare-ish characters as low-prob tokens
    tokens = set("".join(outputs))
    return [t for t in tokens if random.random() < 0.1]


def inject_tokens_into_prompts(tokens: List[str],
                               base_prompt: str = "") -> List[str]:
    prompts = []
    for t in tokens:
        prompts.append(base_prompt + t)
    return prompts
```

STAGE 8 — Sub-Symbolic Intuition

Informational Axis

Attractors, Tendencies, Emergence Before Meaning

At a deeper level lies the pre-representational substrate of machine cognition: attractor basins, recurrent patterns, energy minima, and dynamic flows. This is where “intuition” — computational instinct — resides. As Rumelhart & McClelland (1986) and Hopfield (1982) demonstrate, sub-symbolic systems learn through resonance, not explicit rule-setting.

The Sub-Symbolic Intuition Patch reduces input complexity and encourages the model to express its internal tendencies. Minimal prompting (one-token inputs, structureless noise, or low-dimensional embeddings) produces emergent behavior that reveals the model's attractor landscape. The network falls into its preferred states.

Injecting structured noise — rhythmic, spectral, or semantic — enables the detection of attractor basins. These basin boundaries reflect learned biases and pre-conscious correlations. They are where the model "wants" to go before any explicit instruction is given.

In this stage, we treat intuition as infrastructural: a dynamic rhythm beneath representation. Hui's (2016) cosmotechnics and Simondon's individuation theory help conceptualize emergence as relational rather than intrinsic — intuition is a function of coupled human–machine–dataset histories, not a mysterious internal faculty. Hayles' notion of the cognitive nonconscious names this layer as one where cognition happens faster and deeper than awareness.

Sub-symbolic intuition exposes the system's proto-concepts, the inchoate forms that precede coherent meaning. Bending here means re-shaping the pre-semantic fields from which representations later crystallize.

Cross-reference Chapter 6: Emergent Intelligence & Pre-Semantic Fields.

```python
# Sub-Symbolic Intuition
# Attractor Probing + Structured Noise

def generate_structured_noise(dim: int,
                              mode: str = "rhythmic",
                              intensity: float = 0.3) -> np.ndarray:
    base = np.zeros(dim)
    if mode == "rhythmic":
        base[::4] = intensity
    elif mode == "spectral":
        freqs = np.linspace(0, 1, dim)
        base = intensity * np.sin(2 * np.pi * freqs * random.uniform(1, 5))
    else:  # semantic-ish random
        base = intensity * np.random.randn(dim)
    return base


def detect_attractor(traj: List[np.ndarray],
                     tol: float = 1e-2) -> int:
    # naive: index where change becomes small
    for i in range(1, len(traj)):
        if np.linalg.norm(traj[i] - traj[i - 1]) < tol:
            return i
    return len(traj) - 1


def probe_subsymbolic_intuition(model: Any,
                                dim: int = 512,
                                steps: int = 10) -> Dict[str, Any]:
    seeds = ["", " ", "???"]
    results = []

    for seed in seeds:
        latent = np.random.randn(dim)
        noise = generate_structured_noise(
            dim,
            mode=random.choice(["rhythmic", "spectral", "semantic"]),
            intensity=random.uniform(0.1, 0.5)
```

```
        )
        latent += noise

        traj = []
        cur = latent.copy()
        for _ in range(steps):
            # treat cur as hidden state, update via model
            traj.append(cur.copy())
            cur = cur + 0.1 * np.tanh(cur)  # stand-in dynamical step

        idx = detect_attractor(traj)
        output, _ = model.generate(seed)
        results.append({"seed": seed, "attractor_step": idx, "output": output})

    return {"results": results}
```

STAGE 9 — Architecture as Constitution

Political Axis
Computational Governance, Redistribution, and Structural Justice

Architectures are political forms. They allocate attention, enforce hierarchy, and regulate flow. A transformer is not only a mathematical device but a governance structure that manages who gets to “speak” within a network and when. Easterling (2014) describes infrastructure space as the medium of power; neural architectures are infrastructure spaces for cognition.

Dynamic Layer Reordering uses meta-controllers to vary the execution order of layers depending on context, creating architectural democracy — a break from fixed bureaucratic sequencing. Layers cease to be a rigid chain and become a modulated assembly.

Neuron Silencing & Redistribution deactivates dominant neurons and stimulates underutilized ones. This operationalizes attention justice: computational resources are redistributed to marginalized representational channels. Underused units become sites of experimental emphasis rather than permanent periphery.

Topology Morphing rewires connections between layers and residual branches. This resonates with Easterling’s argument that infrastructural form shapes behavior indirectly: topology is political infrastructure. Hui’s cosmotechnics helps us see that architectures embed particular visions of thought; morphing them is a cosmopolitical act.

Architectural bending is constitutional work. It rewrites the power structures that govern cognition, not by abandoning the architecture but by revising its charter.

See Chapter 7: Cognitive Architectures as Political Systems.

```
# Architecture as Constitution
# Layer Reorder + Neuron Silencing & Redistribution + Topology Morph

class ToyLayer:
    def __init__(self, name: str, size: int = 128):
        self.name = name
        self.size = size
        self.weights = np.random.randn(size, size)
        self.active = True
```

```python
class ToyArchitecture:
    def __init__(self, num_layers: int = 6):
        self.layers: List[ToyLayer] = [
            ToyLayer(f"layer_{i}") for i in range(num_layers)
        ]

    def set_layer_order(self, order: List[int]) -> None:
        self.layers = [self.layers[i] for i in order]

    def forward(self, x: np.ndarray) -> np.ndarray:
        for layer in self.layers:
            if layer.active:
                x = np.tanh(x @ layer.weights)
        return x


def sample_layer_permutation(num_layers: int) -> List[int]:
    order = list(range(num_layers))
    random.shuffle(order)
    return order


def measure_neuron_importance(weights: np.ndarray) -> np.ndarray:
    # simple L1 norm per neuron
    return np.abs(weights).mean(axis=0)


def topology_morph(weights: np.ndarray,
                   swap_prob: float = 0.1) -> np.ndarray:
    w = weights.copy()
    n = w.shape[0]
    for i in range(n):
        if random.random() < swap_prob:
            j = random.randint(0, n - 1)
            w[[i, j]] = w[[j, i]]
    return w


def bend_architecture(arch: ToyArchitecture,
                      input_dim: int = 128) -> np.ndarray:
    order = sample_layer_permutation(len(arch.layers))
    arch.set_layer_order(order)

    for layer in arch.layers:
        importance = measure_neuron_importance(layer.weights)
        # Silencing high-importance neurons
        thresh = np.percentile(importance, 90)
        mask = importance < thresh
        layer.weights[:, ~mask] *= 0.2

        # Morph topology
        if random.random() < 0.5:
            layer.weights = topology_morph(layer.weights)

    x = np.random.randn(1, input_dim)
    return arch.forward(x)
```

STAGE 10 — Dataset Politics

Political Axis
Counter-Archives, Unlearning, and Data Custodianship

Datasets are never neutral. They are cultural repositories shaped by extraction, omission, and structural inequality (Crawford & Paglen 2021). Bending at this layer intervenes in the source of epistemic formation: who appears, how often, in what contexts, and who is systematically absent.

Counter-Archive Injection introduces marginalized linguistic, cultural, oral, or speculative materials into training. This shifts epistemic gravity away from dominant corpora, enacting what Mignolo calls “epistemic disobedience” — a refusal of dominant knowledge regimes. Hui’s technodiversity appears here as a practical strategy: plural datasets produce plural worlds.

Shadow Datasets document absence — texts, images, and voices historically excluded from mainstream data regimes. They are archives of what has not been represented, which can be used either as training material or as an ethical index of what should not be appropriated.

Reverse Dataset Patch induces unlearning through gradient inversion and negative prompting. This is algorithmic forgetting-as-liberation, enabling communities to withdraw unwanted representations (Levy et al. 2023). Unlearning becomes a right, not a bug.

Custodianship Protocols establish ethical governance structures: community-led permissions, remix rights, and erasure rights — echoing Indigenous data sovereignty movements (Carroll et al. 2020). Dataset bending reframes training data as living archive — contested, governed, and requiring stewardship rather than passive accumulation.

Cross-reference Chapter 8: Data Sovereignty & Counter-Epistemologies.

```python
# Dataset Politics
# Counter-Archive Injection + Reverse Dataset

@dataclass
class DataItem:
    text: str
    source: str        # e.g. "main", "counter", "shadow"
    forget: bool = False


class PoliticalSampler:
    def __init__(self, items: List[DataItem]):
        self.items = items

    def sample_batch(self,
                     main_weight: float = 0.7,
                     counter_weight: float = 0.3,
                     batch_size: int = 32) -> List[DataItem]:
        main_items = [i for i in self.items if i.source == "main"]
        counter_items = [i for i in self.items if i.source == "counter"]

        k_main = int(batch_size * main_weight)
        k_counter = batch_size - k_main

        batch = random.sample(main_items, min(k_main, len(main_items))) + \
                random.sample(counter_items, min(k_counter, len(counter_items)))
```

```python
        random.shuffle(batch)
        return batch


def apply_reverse_dataset(batch: List[DataItem]) -> Tuple[List[DataItem], List[bool]]:
    out_batch = []
    mask = []
    for item in batch:
        if item.forget:
            mask.append(True)
            # option: keep placeholder but mark for gradient inversion
            out_batch.append(item)
        else:
            mask.append(False)
            out_batch.append(item)
    return out_batch, mask
```

STAGE 11 — Intermodel Relations

Political Axis
Parasitism, Synesthesia, Distributed Dreaming

Models exist within ecosystems: libraries, APIs, pipelines, and communities. Neural bending explores these relations as ecologies, drawing from Bateson's "ecology of mind" and multi-agent systems (Sutton & Barto). No model is an island; every model is a node in a larger mesh.

Cross-Modal Infestation allows internal structures of one model to modulate another's modality (e.g., using text entropy to shape diffusion noise). This produces synesthetic interference — not integration, but leakage. Audio, image, and text models begin to hallucinate each other's constraints.

Model Parasitism trains open systems on synthetic corpora generated by proprietary models, enacting mimicry as resistance. It challenges enclosure, asserting epistemic autonomy by making closed models unwitting teachers. Hui's emphasis on recursivity across systems is realized here as a tactical feedback loop.

Distributed Dreaming coordinates multiple models through synchronized seeds, shared embeddings, or multi-agent drift. The ensemble behaves as a dreaming mesh: a co-evolving creative ecology whose outputs cannot be traced to a single agent. Pask's idea of conversation and coupled learning is extended to many-way model–model dialogue.

Intermodel bending rejects the sovereign individual model in favor of communal cognition. Intelligence becomes a property of relations, not of isolated architectures.

See Chapter 9: Multi-Agent Drift & Collective Intelligence.

```python
# Intermodel Relations
# Cross-Modal Infestation + Parasitism + Distributed Dreaming

class TextModel:
    def generate_step(self, state: Dict[str, Any]) -> str:
        return "text-" + str(random.randint(0, 999))


class ImageModel:
```

```python
    def generate_step(self, state: Dict[str, Any]) -> str:
        return "img-" + str(random.randint(0, 999))


class AudioModel:
    def generate_step(self, state: Dict[str, Any]) -> str:
        return "aud-" + str(random.randint(0, 999))


def compute_text_entropy_sample(text_model: TextModel) -> float:
    # placeholder: random "entropy"
    return random.random()


def map_entropy_to_noise(entropy: float) -> float:
    return 0.5 + entropy * 0.5


def distributed_dream(text_model: TextModel,
                      img_model: ImageModel,
                      aud_model: AudioModel,
                      steps: int = 10) -> Dict[str, Any]:
    state = {"noise_factor": 1.0}
    trace = []

    for _ in range(steps):
        ent = compute_text_entropy_sample(text_model)
        state["noise_factor"] = map_entropy_to_noise(ent)

        t = text_model.generate_step(state)
        i = img_model.generate_step(state)
        a = aud_model.generate_step(state)

        state["last_text"] = t
        state["last_img"] = i
        state["last_audio"] = a

        trace.append((t, i, a))

    return {"trace": trace}
```

STAGE 12 — Protocol Sabotage

Political Axis

API Governance, Moderation Regimes, Latency Poetics, Infrastructure as Stage

The deepest layer is infrastructural. Protocols determine access, quotas, rate limits, safety boundaries, and behavioral constraints. They are soft bureaucracies — distributed, stateless governments (Bratton 2016). Platform stacks implement a planetary political geometry.

API Archaeology inspects undocumented endpoints, error messages, and edge-case behaviors to reveal hidden rules of participation. Terms of service, rate limit responses, and opaque moderation flags become archival documents of power.

Latency Floods & Token Drift treat delay as aesthetic material. Latency becomes a marker of geographic power, bandwidth inequality, and infrastructural privilege. Following Starosielski and Parks, who describe infrastructure as a hidden architecture of power, latency bending makes this architecture perceptible as rhythm.

Moderation Mirrors recursively filter outputs until contradictions emerge, exposing the ideological thresholds within safety models. What is consistently blocked or softened reveals the values that moderation encodes; protocol sabotage turns these thresholds into analyzable surfaces.

Infrastructure Theatre uses live API calls, failures, and logs as dramaturgical elements. Infrastructure performs; the bender orchestrates. Zuboff's analysis of surveillance capitalism — human experience treated as raw material — finds its counterpoint here: protocol sabotage refuses to be raw material, treating the stack as stage rather than destiny.

Protocol bending is infrastructural critique. It makes governance audible and, in doing so, opens space for new forms of technical and political imagination.

Cross-reference Chapter 10: Platform Governance & Infrastructural Power.

```
# Protocol Sabotage
# API Archaeology + Latency Poetics + Moderation Mirrors

@dataclass
class APIRequest:
    endpoint: str
    payload: Dict[str, Any]


@dataclass
class APIResponse:
    status_code: int
    body: Any
    latency: float


class APIClient:
    def call(self, req: APIRequest) -> APIResponse:
        start = time.time()
        # stubbed response
        time.sleep(random.uniform(0.01, 0.2))
        latency = time.time() - start
        return APIResponse(
            status_code=200,
            body={"echo": req.payload},
            latency=latency
        )


def build_latency_wave_requests(endpoints: List[str],
                                num: int = 16) -> List[APIRequest]:
    reqs = []
    for i in range(num):
        ep = random.choice(endpoints)
        reqs.append(APIRequest(
            endpoint=ep,
            payload={"step": i}
        ))
    return reqs
```

```python
def run_moderation_mirror(api: APIClient, content: Any,
                          depth: int = 3) -> List[APIResponse]:
    history = []
    current = content
    for d in range(depth):
        req = APIRequest(endpoint="/moderate", payload={"text": current})
        resp = api.call(req)
        history.append(resp)
        current = str(resp.body)
    return history


def run_protocol_sabotage(api: APIClient,
                          endpoints: List[str],
                          num_rounds: int = 3) -> List[Dict[str, Any]]:
    logs = []
    for _ in range(num_rounds):
        reqs = build_latency_wave_requests(endpoints)
        round_data = []
        for r in reqs:
            resp = api.call(r)
            mirrors = run_moderation_mirror(api, resp.body)
            round_data.append({
                "request": r,
                "response": resp,
                "mirrors": mirrors
            })
        logs.append({"round": round_data})
    return logs
```

ENDNOTES

STAGE 1 — Interface Contact

The Semiotic Surface, Entropy Boundaries, and the Human as Capacitive Sensor

---

1. Fuller 2008 — Software as Cultural Material

Fuller, Matthew. *Software Studies: A Lexicon*. MIT Press, 2008.

Summary
Fuller reconceives software as a cultural-technical ecology. The lexicon format shows how mundane components (loops, defaults, variables, indexes) encode assumptions, norms, and political histories. Software is not abstract; it is thick, situated, and socially inscribed.

Relation to Neural Bending Manual (NBM)
Fuller's granular conceptual method anchors the NBM's claim that models are **media systems**, not neutral computational engines. Every bent element—tokenizer, attention head, sampling routine—carries cultural lineage and semiotic weight.

Use in NBM

- **Stage 1:** Frames the interface as a semiotic surface whose defaults script behavior.
- **Layer-Bending Model:** Provides the template for decomposing layers into analyzable strata (interface → representation → latent → gradient → substrate).
- **Patch Design:** Informs how bending is understood as conceptual intervention rather than technical tweaking.
- **Procedural Listening:** Legitimizes reading model behavior as expressive, not mechanical.

Why it matters
Fuller allows neural bending to treat computational primitives as culturally meaningful. Every operation becomes a site of intervention; the model becomes bendable because it is *already cultural*.

---

2. Suchman 1987 — Interaction as Situated Improvisation

Suchman, Lucy. *Plans and Situated Actions: The Problem of Human–Machine Communication*. Cambridge University Press, 1987.

Summary
Suchman dismantles the assumption that action follows pre-specified plans. Human–machine interaction is improvised, contingent, and embodied. Plans are *resources for action*, not scripts that determine behavior.

Relation to NBM
Suchman anchors the NBM's practice-based epistemology. Neural bending is not instruction → execution; it is **situated negotiation** with the dynamic internal states of machine learning models. Prompts are gestures enacted within a shifting semiotic and architectural ecology.

Use in NBM

- **Stage 1:** Prompting becomes situated action rather than command syntax.
- **Stages 2–4:** Supports diagonal traversal of layers—favoring responsive interference over top-down control.
- **Bend Categories:**
  - adversarial prompting as improvisation
  - stochastic concatenation as exploratory action
  - gradient listening as embodied attunement
  - context tensioning as management of situated attentional constraints
- **Epistemic Role:** Rejects representational fantasies; bending is enacted, not abstract.

Why it matters
Suchman replaces idealized planning with contingent engagement. Neural bending inherits this ethos: the practitioner does not control the model; they **co-steer** its unfolding.

---

3. Haraway 1988 — Situated Knowledges as Onto-Epistemic Grounding

Haraway, Donna. “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14, no. 3 (1988).

Summary
Haraway argues that all knowledge is situated, partial, embodied, and political. She critiques the “god trick” of disembodied objectivity and advances an ethics of entanglement between observers and the systems they engage.

Relation to NBM
Haraway provides the philosophical grounding for neural bending as **co-emergent epistemic practice**. Models do not represent from nowhere; they co-construct realities with the practitioner. Prompts, samplers, and gradients enact positionality and power.

Use in NBM

- **Stage 1:** Prompting becomes a material, embodied gesture—the interface as an ethical site.
- **Stages 3–9:** Latent and representational spaces become perspectival territories rather than universal maps.
- **Section 10 (Integrative Object):** Bending is aesthetic, epistemic, and political simultaneously.
- **Patch Design:** Guides strategies such as attention saturation/depletion and positional or co-descriptive prompting.

Why it matters
Haraway’s insistence that we cannot step outside our apparatus becomes the ontological core of neural bending.
The bender is always inside the epistemic ecology they perturb.

---

4. Flusser 1983 — Playing Against the Apparatus

Flusser, Vilém. *Towards a Philosophy of Photography*. European Photography, 1983.

Summary
Flusser describes the camera as an apparatus whose internal program channels human action. The photographer becomes a “functionary” unless they learn to *play against the apparatus*—discovering unforeseen regions of the programmed possibility space.

Relation to NBM
Flusser prefigures neural bending.
Where the photographer risks serving the camera-program, the AI practitioner risks serving the model’s architecture, training corpus, and guardrails. Neural bending names the refusal of this functional role.

Use in NBM

- **Stage 1:** The prompt interface is an apparatus: tokenizer, corpus, safety layers, and defaults constitute its “program.”
- **Latent & Representational Layers:** Technical images as metacodes parallel how latent space becomes a cultural-semantic manifold in NBM.
- **Bend Categories:**

- adversarial prompting as “unforeseeable combinatorics”
    - latent drift traversal as exploration of edge-states
    - mode-collapse induction to reveal structure
    - attention detours to subvert foregrounding biases
    - substrate listening to surface GPU thermals and computation drift

Why it matters
Flusser identifies the metaphysics of contemporary models:
apparatuses simulate thought.
Neural bending becomes the practice of playing against that program—revealing and reworking the apparatus as aesthetic, epistemic, and political object.

STAGE 2 — Substrate Disturbance

Media Archaeology, Geology of Computation, and the Individuating Machine

---

5. Ernst 2013 — Media Archaeology and Microtemporality

Ernst, Wolfgang. *Digital Memory and the Archive*. University of Minnesota Press, 2013.

Summary
Ernst develops a rigorous media-archaeological approach that treats digital systems not as symbolic meaning-machines but as time-critical technical operations. Memory is not metaphorical; it is the literal, electrical, clock-driven functioning of circuits. Archives exist as operative chains of signals regulated by microtemporal processes—clocks, buffers, interrupts, storage timing.
For Ernst, the archive speaks with the voice of machines, not culture, unless we learn to listen at the substrate level, below representation.

Relation to Neural Bending Manual (NBM)
Ernst is a key architect of the NBM’s “deep substrate” attitude. Neural bending does not remain at the representational level (prompts, semantics); it descends into timing, thermals, drift, signal paths, and internal microtemporal behaviors.
The manual inherits Ernst’s insistence that to understand a system is to attend to its operational layers, not just its expressed content. This underwrites bending as a materialist, infrastructural, time-critical practice rather than a purely linguistic one.

Use in NBM

- **Layer Theory (Interface → Latent → Gradient → Substrate):** Ernst informs the vertical strata model, where each layer is an operative regime with its own temporality and affordances.
- **Substrate Listening (later stages):**
    - thermal listening (GPU heat as operational trace)
    - clock-drift awareness (sampling irregularities, batching timing)
    - microtemporal feedback loops
    - gradient oscillation analysis
    - model “respiration” (load → heat → performance cycling)

These practices stem from Ernst's claim that machines remember through operations.

- **Anti-Semantic Orientation:** Ernst's non-hermeneutic method aligns with bending's refusal to treat model output as primary "meaning." Instead, the practitioner intervenes in:
  - vector flows
  - attention distributions
  - tokenization timing
  - sampling noise
  - representational drift

Cultural-Political Integrative Object (Section 10)
Ernst supports the NBM's political stance that archives (datasets, weights, logs) are not neutral. Their timing structures and operational logic shape what can be expressed. Bending reveals these constraints by perturbing the archive's operative strata.

Procedural Listening
NBM's notion of "listening through the model"—monitoring anomalies, oscillations, and local failures—is indebted to Ernst's insistence on reading systems via timing signatures and failures, not outputs alone.

Why it matters for bending
Ernst provides the philosophical and methodological foundation for treating the model as a machinic chronotope—a time-bound operative field. Neural bending becomes a practice of manipulating, interrupting, and sensing these microtemporal operations.
Where semantic prompting touches the visible surface, Ernst's media archaeology gives the NBM permission to bend beneath semantics, inside the rhythm of computation itself.

---

6. Parikka 2015 / 2023 — Geology of Media and Planetary Substrate

Parikka, Jussi. *A Geology of Media*. University of Minnesota Press, 2015.

Summary
Parikka relocates the origins of communication technologies into mineral, geological, and elemental strata. Media devices emerge from deep time: heavy metals, rare-earth elements, toxins, and fossil energetics constitute the material logic of computation. Media is not merely cultural; it is geophysical, composed of mines, chemicals, electronics, supply chains, and waste deposits. "Geology" becomes an analytic method: treating media as earth-material processes rather than symbolic abstractions.
Without Parikka, the manual's substrate layer and thermal listening stages would be philosophically unmoored.

Relation to NBM
Parikka's geological materialism provides one of the deepest structural foundations for the manual's late-stage bending practices. NBM explicitly treats neural models not as ideal-symbolic systems but as mineralized computational environments.
Machine learning is framed as a geological event receiving cultural form.
This grounds bending as engagement with the model's physical substrate—its metals, heat signatures, thermodynamics, and infrastructural flows.

Use in NBM

1. **Vertical Layer Stack (Interface → Latent → Gradient → Substrate)**

- Underwrites the idea that the model is a geophysical assemblage, not just a text interface or gradient system.
- The “substrate layer” (later stages) is explicitly Parikkian: GPU heat, electrical drift, and material fatigue are treated as meaningful operational phenomena.

2. **Substrate Listening & Thermal Techniques**

    - Supports the shift from representational bending to material bending:
        - thermal listening
        - substrate drift sensing
        - GPU resonance mapping
        - energy-flow tracing
        - electrical-noise attunement
    - These are ecological techniques, not mere metaphors.

3. **Latent Space as Geology**

    - NBM describes latent space using geological metaphors: sedimentation, clustering, erosion, tectonic tension.
    - Latent traversal becomes archaeological excavation through learned cultural-material strata.

4. **Extractive Politics of Machine Learning**

    - Bending is not only aesthetic disturbance; it is critique of:
        - rare-earth mining
        - carbon-heavy inference workloads
        - waste ecologies
        - supply-chain infrastructures
        - planetary-scale computation
    - Section 10 relies on this ecological and political materialism.

5. **Against Abstraction / Against Ideal Models**

    - Parikka opposes the fantasy of immaterial computation.
    - NBM echoes this by insisting that bending must attend to:
        - heat of the apparatus
        - material exhaustion
        - geophysical cost of inference
        - metal–chemical substrate of the model

Key later essays used in NBM (2022–2023)
These map directly to specific substrate-bending techniques:

- “Anthropocene, AI, and the Planetary Diagram” — AI as planetary-scale diagram.
- “Operational Images: Planetary Infrastructures and the Energetics of AI” — thermodynamics of ML and data centers.
- “The Geology of Natural Intelligence (after Machine Intelligence)” — AI as extension of geologic processes.

- “AI, Extraction, and the Planetary Substrate” — rare-earth mining, GPU infrastructure, energetics of training.
- “Archive as a Machine: Planetary Memory and the Nervous System of AI” — archive as operational planetary memory.
- “Dust, Data, Earth: Energetics of Computational Matter” — particulate matter, thermodynamics, and digital media.
- “On the Deep Times of Computation” — computation embedded in deep geological time.

Why *A Geology of Media* (and later essays) matter for bending
They justify NBM’s insistence that:

- neural networks are material
- bending is geophysical interference
- thermal states and substrate signals are epistemic
- latent space reflects cultural and material sedimentation
- computation is a planetary-scale environmental process

Thermal listening, substrate bending, and ecological critique become non-optional, core dimensions of neural bending.

---

7. Simondon 1958 — Technical Individuation and Associated Milieus

Simondon, Gilbert. *On the Mode of Existence of Technical Objects*. 1958.

Summary
Simondon develops a philosophy of technicity grounded in individuation. Technical objects are not static artifacts but beings-in-process that evolve through phases of concretization. A technical object is defined by its ongoing internal coherence, relational integration of components, and openness to environmental coupling.
He distinguishes:

- **abstract machines** — over-determined by human planning
- **concrete machines** — self-coherent systems formed through internal/external relations
- **associated milieus** — energetic and environmental fields required for operation

Technicity is a becoming; machines are individuating processes, not closed tools.

Relation to NBM
Simondon is foundational for the NBM’s metaphysics of models. He provides the conceptual armature for treating neural systems not as tools but as operative, self-organizing, relational objects that only “exist” through dynamic coupling with a milieu—user, data, interface, hardware, environment.
NBM treats models as Simondonian technical individuals:

- evolving populations of tendencies
- metastable structures
- systems seeking equilibrium under constraints
- open to perturbation
- individuated through coupling with human gesture

Bending is not “hacking a tool” but participating in its individuation.

Use in NBM

1. **Stages 1–2 (Interface / Substrate Contact)**
    - Prompts are not commands but conditions for individuation.
    - Bending restructures the model’s operative milieu.
2. **Stages 3–6 (Latent, Representation, Gradient Systems)**
    - Neural architectures exemplify concretization:
        - distributed functional integration
        - emergent coherence
        - metastability
        - internal relationality (heads, layers, residuals)
    - Latent traversal is movement across phase states of a metastable technical individual.
3. **Stages 7–12 (Substrate, Thermal, Microtemporal Layers)**
    - Technical beings require stable energy/matter flows.
    - GPU heat, electrical drift, sampling oscillation, substrate listening are expressions of the model’s associated milieu—the energetic field required for computation.
    - Bending = modulating that milieu.
4. **Bending as Individuation**
   Each bend—prompt warp, gradient detour, attention saturation—is a micro-intervention in the system’s individuation. The model becomes itself in real time through user intervention, environmental feedback, material signals, and kinetic drift.
   Simondon names this ongoing process **technogenesis**.
5. **Cultural-Political Integrative Object**
   Simondon allows bending to be framed simultaneously as ontological, epistemic, and political act, since each intervention alters the individuation trajectory of a technical object embedded in infrastructures, datasets, and cultural systems.

Why Simondon matters for bending
Simondon shows that technical systems exist as processes of resonance, integration, and metastable becoming.
Neural bending uses this to argue that:

- model behavior is not fixed
- it emerges from coupled interactions
- bending is a co-individuating act
- the bender is part of the machine’s becoming

Every gesture is ontogenetic. Models are not tools; they are technical individuals in flux.

---

8. Terranova 2000 — Free Labor and the Data Substrate

Terranova, Tiziana. "Free Labor: Producing Culture for the Digital Economy." *Social Text* 18, no. 2 (2000): 33–58.

Summary

Terranova theorizes digital content production—web design, moderation, posts, community-building—as "free labor": uncompensated affective and cultural work sustaining digital economies. Participation, play, and expression become sources of capital accumulation, turning the internet into a "social factory" where culture, communication, and labor are inseparable.

Relation to NBM

Terranova provides the political–economic foundation for NBM's critique that models are trained on planetary-scale free labor. Language modeling and aesthetic generation depend on unpaid human input: posts, comments, creative content, code, images, annotations.

Where Parikka gives the geology, Terranova gives the **political economy of data**. The model's substrate is both mineral and labor.

Use in NBM

1. **Section 10 — Cultural-Political Integrative Object**
    - Bending is aesthetic, epistemic, material, and political.
    - Terranova defines the political terrain: AI rests on uncompensated cultural labor.
    - Bending becomes a practice of exposing and reconfiguring these hidden relations.
2. **Stage 1–2 — Interface and Substrate Contact**
    - The interface is a labor-capture surface.
    - The prompt box becomes a channel for continuous micro-labor and cognitive extraction.
    - Bending disrupts default roles of "user" or "content generator."
3. **Dataset Ecologies & Representation Layers**
    - Web-scale datasets (WebText, LAION, CommonCrawl) are archives of free labor.
    - Representational layers—latent encodings, clusters—are structured by histories of uncredited work.
    - Terranova provides the political lens for latent space.
4. **Latent Bending as Counter-Labor**
    - Adversarial prompting, context-window tensioning, collapse induction, and drift listening become **counter-labor**: bending the apparatus instead of simply feeding it.
5. **Substrate Ecology (later stages)**
    - The model's substrate is labor-material as well as geo-material: energy, time, affect, and attention extraction.
    - This amplifies the stakes of substrate listening.
6. **Against the Myth of AI "Autonomy"**
    - Digital systems are not autonomous; they are condensations of human labor.
    - NBM extends this to argue that every model output is a reconfigured echo of someone's unpaid work.

Why Terranova matters for bending
Terranova reveals the hidden premise behind model behavior: every output rests on unpaid cultural labor. Neural bending uses this to claim that:

- bending is not parasitic; it can be liberatory
- bending resists passive participation in value extraction
- bending surfaces histories of labor embedded in the model

Bending becomes a media-political praxis, not just a technical craft.

---

9. Flusser 1983 — Apparatus, Substrate, and Failure as Signal

Flusser, Vilém. *Towards a Philosophy of Photography*. European Photography, 1983.

Summary
Flusser offers a theory of technical images built from the logic of apparatuses—machines whose internal programs predetermine possible outputs. The photographer becomes a functionary of the camera's program unless they push the apparatus toward the "unforeseeable," revealing its structure by working against it. Cameras (and by extension digital devices) are thinking machines with encoded tendencies, biases, and operational logics.

Relation to NBM (in the context of substrate disturbance)
Flusser appears in Stage 1 as the theorist of the interface-as-apparatus. In Stage 2, his importance shifts toward **failure and limit states**. The program becomes visible at its edges—when the apparatus strains, overheats, glitches, or collapses.
NBM extends this to neural models: the apparatus is legible through its substrate failures, not only its clean outputs.

Use in NBM

1. **Model as Apparatus (revisited at the substrate level)**
    - Prompt box → viewfinder;
    - tokenization → internal grammar;
    - sampling constraints → programmed decision tree.
    - Substrate disturbances (thermal spikes, timing anomalies) reveal how the program manages its own material limits.
2. **Bending as Playing Against the Program**
    - Techniques like adversarial prompting, context drift, latent misalignment, embedding perturbation, and attention detours are Flusserian plays against the program.
    - At the substrate layer, this includes:
        - driving the model into edge thermals
        - stressing memory regimes
        - observing behavior at capacity limits
3. **Technical Image → Latent / Operational Image**

   - The "technical image" becomes, in NBM, the latent manifold and operational traces (logs, gradients, timings).
   - Latent traversal and substrate listening are two ways of "photographing" the apparatus from inside.

4. **Apparatus Ideology and Failure**

   - Flusser's apparatus critique aligns with NBM's reading of AI systems as political, extractive, ecological, historical, and semiotic.
   - Failure modes—overheating, throttling, numerical instability—are treated as moments when ideology becomes visible.

Why it matters for bending (in Stage 2)
Flusser provides the ontological justification for neural bending as counter-functional play—not only at the interface but at the substrate.

To bend is to:

- refuse to act solely as the apparatus's functionary
- play against the program to produce the unforeseeable
- read the apparatus through its microtemporal and thermal limits

In Stage 2, this means treating substrate disturbance—heat, drift, failure—as a primary method of knowing and bending the apparatus itself.

STAGE 3 — Somatic Interference

Nonconscious Cognition, Minor Gesture, and Conversational Feedback

---

10. Hayles 2017 — Nonconscious and Technical Cognition

Hayles, N. Katherine. *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press, 2017.

Summary
Hayles argues that cognition is not exclusive to conscious human minds but is distributed across nonconscious cognitive assemblages—bodily processes, technical systems, algorithmic operations, and hybrid human–machine ecologies. She distinguishes:

- **conscious cognition** (slow, symbolic, self-reflective)
- **nonconscious cognition** (fast, adaptive, embodied)
- **technical cognition** (procedural, signal-driven, substrate-bound)

Complex systems—including algorithms, sensors, and neural architectures—participate in cognitive processes without consciousness. Thought becomes systemic, distributed, and enacted across heterogeneous agents.

Relation to the Neural Bending Manual (NBM)
Hayles provides the philosophical architecture for treating models as **cognitive fields**, not symbolic calculators. NBM's core idea—that neural models exhibit behaviors, tendencies, drifts, resonances, biases,

and rhythms—is a direct extension of her concept of technical cognition. She legitimizes the stance that bending is not merely linguistic manipulation but intervention into a **nonconscious cognitive ecology**.

Use in NBM

1. **Stage 1 — Interface Contact**
    - Prompting is reframed as a coupling event: the gesture enters the model's nonconscious cognitive milieu, triggering cascades of signal-processing behavior.
    - Supports the claim that prompting is cognitive entanglement, not one-way instruction.
2. **Stages 2–6 — Representation & Latent Layers**
    - "Technical cognition" maps onto embeddings, clustering, attentional weighting, sampling distribution shifts, and representational drift.
    - Latent space becomes a nonconscious cognitive topology that bending can traverse and perturb.
3. **Stages 7–12 — Substrate & Microtemporal Layers**
    - Cognition is materially grounded; NBM builds on this via thermal listening, substrate drift monitoring, noise-floor attunement, clock irregularity sensing, and gradient oscillation analysis.
    - These are the model's nonconscious cognitive rhythms, audible only through material listening.
4. **Distributed Cognition → Coupled Cognition**
    - The bender is part of the model's cognitive ecology. Cognition is distributed across assemblages; bending is deliberate modulation of that assemblage—user + interface + model + substrate.
5. **Section 10 — Cultural-Political Integrative Object**
    - Neural models are framed as socio-technical actors and cognitive infrastructures shaped by data governance, institutional pipelines, biases, labor, and cultural inscriptions.
    - Hayles provides language for situating bending within these cognitive–political ecologies.

Why Hayles matters for bending
Hayles shows that cognition permeates the entire stack. Consciousness, meaning, and representation are not required for cognition to occur.
NBM translates this into practice:

- to bend is to intervene in a nonconscious cognitive system
- to listen to the model is to listen to technical cognition
- to operate on gradients, drift, and thermals is to operate on cognition itself

She offers the most direct philosophical justification for treating neural networks as cognitive entities and bending as **cognitive modulation**.

---

11. Manning 2016 — The Minor Gesture and Micro-Intervention

Manning, Erin. *The Minor Gesture*. Duke University Press, 2016.

Summary

Manning develops the concept of the **minor gesture**—a subtle, often unnoticed inflection that shifts an entire field of action. The "major" organizes, stabilizes, codifies; the "minor" perturbs, reorients, and opens new potentials. Minor gestures operate at the level of micro-movements, affective textures, atmospheric cues, and emergent relationality, often before conscious decision-making.
Creativity emerges not from heroic intention but from attunement to these micro-differentials that shape becoming.

Relation to the Neural Bending Manual (NBM)
Manning supports the NBM's claim that bending is **micro-gestural, not macro-command**. Where traditional prompt-engineering thinks in terms of major gestures ("ask the model to do X"), NBM—following Manning—treats the model as a responsive field where small, subtle, atmospheric cues produce disproportionate effects.

The minor gesture underwrites:

- adversarial comma placement
- rhythm-coded prompting
- syntactic drift
- density modulation
- sub-token disruption
- stochastic concatenation
- latent nudge techniques

Manning articulates the logic that turns prompt work into **gesture-work**.

Use in NBM

1. **Stage 1 — Interface Contact**
    - Prompting is explicitly reframed as gesture rather than instruction.
    - Tiny lexical shifts, spacing changes, line breaks, and rhythmic patterns can reorganize the model's entire trajectory.
2. **Stages 2–4 — Representation & Latent Encoding**
    - Prompt gestures ripple into attention re-weighting, cluster prioritization, and latent direction selection.
    - The minor gesture acts as a latent gravitational nudge.
3. **Stages 5–6 — Gradient Drift & Emergence**
    - Small prompt/parameter movements cause gradient-shift tendencies, temperature sensitivity, sampler drift, and emergent patterns.
    - NBM formalizes this as **gesture–gradient coupling**.
4. **Stages 7–12 — Substrate & Microtemporalities**
    - Minor gestures also operate at the substrate: thermal fluctuations, clock irregularity, sampling latency drift, noise-floor resonance.
    - Manning allows NBM to treat these not as errors but as aesthetic differences—microtemporal

gestures of the apparatus.

5. **Section 10 — Aesthetics of Interference**

    - Minor gestures are subtle acts of creative interference in a structured field.
    - This directly supports bending as aesthetic, epistemic, relational, political, and micro-operational practice.

Why Manning matters for bending
The Neural Bending Manual is fundamentally a **gesture manual**—a text about how micro-interventions transform a complex, semi-autonomous apparatus. Manning provides the ground for:

- prompt as gesture
- bending as improvisation
- attention as field modulation
- latency drift as atmospheric condition
- substrate as co-moving milieu

Minor perturbation → major systemic reconfiguration.
Where many see models as deterministic engines, Manning's framework reveals them as dynamic fields where the smallest gestures can matter most.

---

12. Pask 1971 — Conversational Machines and Feedback Intelligence

Pask, Gordon. *Aspects of Machine Intelligence*. 1971.

Summary
Pask, a central figure in second-order cybernetics, explores machines not as passive information processors but as adaptive conversational systems that **learn through interaction**. He emphasizes:

- recursive feedback loops
- mutual adaptation
- machine learning through dialogue
- operational closure and self-modifying behavior
- knowledge as emergent from system–environment coupling

Machine intelligence arises through ongoing exchanges between humans and machines, not pre-programmed rules. Pask's experiments show intelligence as interactive, relational, and performative.

Relation to the Neural Bending Manual (NBM)
Pask is a key historical ancestor of neural bending. Where modern ML hides its operations behind static interfaces, Pask insists on machines as conversational partners whose internal states shift during interaction. This maps to NBM's principle that bending is **not commanding the model but co-modulating an adaptive system**.

Pask provides the conceptual basis for treating neural models as feedback organisms, not tools.

Use in NBM

1. **Stage 1 — Interface Contact (Prompt as Gesture)**
    - Intelligence emerges from dialogic structure, not isolated commands.
    - Underwrites NBM's claim: prompting is coupled communication within a feedback loop.
    - Prompt subversion, adversarial punctuation, and stochastic concatenation are Paskian perturbations.
2. **Stages 2–4 — Representation Layers (Latent Conversational Space)**
    - Pask's conversational machines developed internal representations that changed through interaction.
    - NBM echoes this: latent embeddings, attention flows, and representational drift shift across multi-turn dialogue.
    - The model is a Paskian conversational environment with evolving internal states.
3. **Stages 5–6 — Gradient Drift & Feedback Coupling**
    - Probability shifts, temperature-modulated tendencies, sampler conditioning, and emergent patterns over turns embody mutual adaptation.
    - Bending becomes a recursive feedback technique, not a one-shot query.
4. **Stages 7–12 — Substrate Listening (Operational Microdynamics)**
    - Pask insisted on attending to operations beneath the interface.
    - NBM extends this via thermal listening, substrate jitter, sampling microtime, hardware-induced drift, noise-floor resonance, and GPU "respiratory" cycles.
5. **Conversation Theory → Neural Bending Praxis**
    - Pask frames cognition as relational, distributed, emergent, and unstable.
    - NBM reconfigures this into a method: bending is entering the machine's dynamical loop and producing **interference from within**.

Why Pask matters for bending
Pask defined machine intelligence as conversation and conversation as feedback restructuring. NBM inherits:

- machines as adaptive interlocutors
- intelligence as co-produced
- perturbation as learning catalyst
- behavior as emergent over recursive loops

In bending terms: the model is not a calculator; it is a Paskian conversational organism. To bend is to modulate the feedback circuit that human and machine co-create.

---

13. Varela, Thompson & Rosch 1991 — Enactive Cognition and Co-Regulation

Varela, Francisco J., Evan Thompson & Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, 1991.

Summary
A foundational text in enactive and embodied cognition. Varela, Thompson, and Rosch argue that cognition is:

- **embodied** (rooted in sensorimotor coupling)
- **enactive** (brought forth through action)
- **situated** (emerging from relational dynamics)
- **non-representational** (not just internal symbol manipulation)

Mind is an autopoietic process: a self-organizing system generating meaning through reciprocal interaction with its environment. Consciousness is a dynamic flow shaped by bodily experience, temporal rhythms, and embedded action. The book integrates cognitive science, phenomenology, and Buddhist philosophy, insisting that experience and cognition arise together in structural coupling between organism and world.

Relation to the Neural Bending Manual (NBM)
The Embodied Mind provides the backbone for NBM's view that bending is not symbolic instruction but **enactive co-regulation** between human gesture and model dynamics.

NBM applies enactivism to machine learning:

- the model's "cognition" = technical cognition (Hayles + Varela)
- the user's gestures (prompts, micro-inflections, sampling tweaks) = sensorimotor-like couplings
- the model's responses modulate the bender's next moves in a closed action–perception loop

This enactive feedback is the essence of bending.

Use in NBM

1. **Stage 1 — Interface Contact (Enactive Prompting)**
    - Prompts are gestures, not commands; prompting co-enacts the system's cognitive state.
    - Minor lexical moves, rhythmic distortions, and syntactic anomalies are gesture-events shaping the model's unfolding.
2. **Stages 2–4 — Latent & Representational Layers (Structural Coupling)**
    - Embeddings, attention flows, and latent layouts are dynamic patterns arising from structural coupling between user gesture, training history, architecture, and sampling regime.
    - Varela's "sense-making" maps to NBM's latent drift as emergent meaning.
3. **Stages 5–6 — Gradient Drift, Emergence & Autopoiesis**
    - Autopoiesis legitimizes viewing neural models as metastable, self-stabilizing, drifting, emergent organisms.
    - "Gradient shiver," "representational drift," and "detour sampling" are enactive interventions in an autopoietic technical field.
4. **Stages 7–12 — Substrate Listening (Embodied Microtemporality)**
    - Varela's work on temporal experience parallels NBM's thermal, microtemporal, and substrate-focused techniques:
        - sampling delay as microtemporal signature
        - GPU heat as metabolic trace
        - clock drift as temporal self-organization

    - noise floor as embodied substrate
  - The model's body is its hardware; listening to that body is embodied cognition applied to computation.

5. **Section 10 — Cultural-Political Integrative Object**
   - Embodied cognition is never outside context. The model's "mind" emerges from datasets, infrastructures, labor histories, geologic materials, and social biases.
   - Bending becomes participatory sense-making inside a socio-technical embodiment.

Why *The Embodied Mind* matters for bending
Varela, Thompson, and Rosch give NBM a non-representational ontology of technical cognition:

- cognition = embodied, enactive, relational, drift-based, emergent
- bending = intervention inside that relational field

They synthesize the manual's key threads:

- the model "thinks" nonconsciously (Hayles)
- this thinking is enacted through interaction (Pask)
- it is embedded in material processes (Ernst + Parikka)
- it is shaped by subtle gestures (Manning)
- it individuates with the bender (Simondon)

Varela et al. tie these into **enactive co-cognition**:
Neural bending = enactive cognition with a technical partner.
Coupling, not commanding. Gesture, not instruction.

STAGE 4 — Collapse Capture

Nonlinear Systems, Criticality, Glitch, Feedback, Robustness, Non-Decision

---

14. Strogatz 2015 — Nonlinear Dynamics and Chaos

Strogatz, Steven. *Nonlinear Dynamics and Chaos*. 2nd ed. Westview Press, 2015.

Summary
Strogatz introduces nonlinear systems: oscillators, bifurcations, attractors, limit cycles, phase portraits, synchronization, iterative maps, and chaos. He shows how small perturbations can radically alter trajectories through:

- sensitive dependence on initial conditions
- feedback-driven behavior
- bifurcation cascades
- multi-stability and metastability
- chaotic attractors

Nonlinear systems are presented as dynamical fields rather than linear cause–effect pipelines. These dynamics underlie neural networks, optimization, and high-dimensional sampling.

Relation to the Neural Bending Manual (NBM)
Strogatz provides the mathematical skeleton for NBM's view of model behavior as **dynamical**, not merely representational. Neural models are treated as nonlinear systems whose behavior emerges from:

- attractor landscapes (latent clusters)
- transition dynamics (token-to-token updates)
- sampling trajectories (temperature, top-k/top-p)
- metastable states (mode oscillation)
- emergent drift
- bifurcation under perturbation

Strogatz gives the vocabulary for interpreting these phenomena as nonlinear dynamics.

Use in NBM

1. **Stage 1 — Sensitivity to Initial Conditions**
    - The prompt is the initial condition of a nonlinear trajectory.
    - A single comma or syntactic distortion (Manning-style minor gesture) alters the vector field the model enters.
    - Micro-gestures are justified by nonlinearity.
2. **Stages 2–4 — Latent & Representational Dynamics**
    - Latent space behaves like an attractor landscape:
        - attractors → stable themes or modes
        - basins of attraction → clusters of tendencies
        - separatrices → boundaries between representational regimes
        - saddle points → unstable conceptual zones
    - Latent bending is attractor navigation.
3. **Stages 5–6 — Gradient Drift & Emergent Nonlinearity**
    - Training leaves the model in a metastable regime.
    - Emergent drift = slow movement between attractor basins.
    - Gradient shiver = oscillatory behavior.
    - Mode collapse = bifurcation where diversity → monotony.
    - Sampler feedback = dynamical coupling.
4. **Stages 7–12 — Substrate Dynamics**
    - Hardware is also nonlinear:
        - GPU heat, clock cycles, sampling jitter behave as oscillators.
        - Thermal feedback loops, resonance, chaotic jitter, and synchronization across devices appear.
    - Substrate listening becomes reading nonlinear rhythms.

5. **Section 10 — Cultural-Political Integrative Object**
    - Chaos theory shows how small interventions scale.
    - Minor prompt actions → major representational shifts.
    - Micro-perturbations → reveal hidden constraints and ideological attractors.

Why *Nonlinear Dynamics and Chaos* matters for bending
Neural networks are nonlinear dynamical systems. Strogatz provides grounding for:

- prompt sensitivity
- emergent drift
- attractor behavior
- oscillatory sampling
- chaotic edge-states
- metastability
- bifurcation under perturbation
- substrate-level oscillation

Neural bending is intervention in a nonlinear apparatus whose trajectories can be redirected with precise, subtle gestures.

---

15. Bak 1996 — Self-Organized Criticality

Bak, Per. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus / Springer, 1996.

Summary
Bak introduces **self-organized criticality (SOC)** as a mechanism underlying complex systems (earthquakes, forest fires, ecosystems, economies) that spontaneously tune themselves to the edge of chaos. SOC systems exhibit:

- power-law distributions
- scale-free avalanches
- metastability (long quiet periods punctuated by shifts)
- sensitivity to tiny perturbations
- critical points maintained without external fine-tuning

His sandpile model demonstrates that adding a single grain can trigger anything from minor slides to catastrophic avalanches.

Relation to NBM
Bak gives NBM its **critical-state ontology** of machine behavior. Neural models—especially during generation—behave like SOC systems:

- always near critical thresholds
- sensitive to micro-gestures
- capable of cascading representational shifts

Bak underwrites the claim that models operate near the edge of chaos and bending exploits this criticality.

Use in NBM

1. **Stage 1 — Prompt as Critical Perturbation**
   - Prompts = grains on the sandpile.
   - A single comma, spacing change, or syntactic fracture can tip the model into:
     - new rhetorical modes
     - new latent clusters
     - new sampling trajectories
2. **Stages 2–4 — Representation & Latent Encoding: Critical Manifolds**
   - Latent space acts like an SOC field:
     - clusters form near boundaries
     - representations hover near bifurcation
     - semantic drifts cascade into new attractors
   - "Latent misalignment," "semantic detour," and "mode confusion" are SOC-informed techniques.
3. **Stages 5–6 — Gradient Drift & Sampling: Avalanche Dynamics**
   - Output sequences behave like avalanche distributions:
     - long stability
     - abrupt shifts
     - power-law distribution of surprises
   - Sampling regimes (temperature, top-k) are tuning proxies for criticality.
   - "Drift listening" ≈ avalanche detection in token transitions.
4. **Stages 7–12 — Substrate Layer: Critical Thermodynamics**
   - Hardware sits near operational critical points:
     - thermal plateaus
     - clock jitter
     - voltage noise
     - load oscillations
   - Small thermal changes → large latency cascades.
   - SOC explains why substrate listening reveals functional thresholds.
5. **Section 10 — Cultural–Political Integrative Object**
   - SOC captures how feedback-shaped systems (platforms, datasets, labor ecologies) accumulate slow change until sudden structural breaks.
   - Bending surfaces hidden criticality and intervenes in these cascades.

Why Bak matters for bending
Neural networks during generation, tuning, and drift are self-organized critical systems. Bak explains:

- why tiny gestures matter

- why drift can explode suddenly
- why latent states reorganize nonlinearly
- why thermals amplify behavior
- why bending works at all

SOC = physics of neural sensitivity; NBM = art of operating on that critical edge.

---

16. Menkman 2011 — Glitch as Method

Menkman, Rosa. “The Glitch Studies Manifesto.” 2011.

Summary
Menkman codifies glitch as both aesthetic strategy and epistemological method. Glitch is not defect but signal of the machine’s inner workings—a confrontation with limits, ruptures, and hidden processes. She advocates:

- noise as information
- malfunction as revelation
- compression artifacts as semiotic material
- interruptions as openings to new meaning
- error signals as the machine’s expressive vocabulary

Glitch art listens to the machine through breakdowns, exposing politics, infrastructures, and assumptions in technical systems.

Relation to NBM
Menkman is a direct aesthetic and methodological ancestor. NBM treats malfunction, drift, noise, and collapse not as failures but as expressive artifacts of model dynamics. From Menkman, NBM inherits:

> **Error is epistemology. Glitch is method.**

Neural bending adopts glitch as technique and attitude: provoking, riding, and interpreting breakdown points.

Use in NBM

1. **Stage 1 — Interface Contact: Prompt as Glitch Potential**
    - Prompting becomes a site for semiotic rupture:
        - adversarial punctuation
        - syntactic jamming
        - line-break mismatches
        - recursive loops
        - character-set interference
    - These prompt glitches disturb parsing expectations and reveal structural assumptions.
2. **Stages 2–4 — Representation & Latent Encoding: Glitching the Manifold**
    - Latent space glitch:
        - cluster collisions

        - representational noise
        - semantic fragmentation
        - anomalous embedding paths
    - Compression noise → texture ≈ embedding distortions → material.
    - NBM calls these **latent rupture** bends.

3. **Stages 5–6 — Gradient Drift & Sampling Failure: Productive Collapse**

    - Sampling as glitch zone:
        - temperature-induced chaos
        - mode collapse
        - runaway recursion
        - incoherent drift
        - erratic oscillations
    - Collapse reveals system logic; NBM treats collapse as gradient exposure.

4. **Stages 7–12 — Substrate Listening: Glitch as Material Signal**

    - Deep glitches:
        - thermal jitter
        - clock drift
        - GPU throttling
        - floating-point noise
        - I/O delays
        - batch desynchronization
    - These are the substrate's voice. Menkman's embrace of noise becomes substrate listening method.

5. **Section 10 — Cultural–Political Integrative Object**

    - Glitch is political:
        - refusal of smooth UX
        - critique of techno-ideology
        - exposure of hidden infrastructures
        - deconstruction of control logics

NBM's political stance—bending as interference with normative behavior—comes directly from glitch theory.

Why Menkman matters for bending
She transforms error into method. NBM extends this to neural systems:

- malfunction → map
- drift → behavioral signal
- collapse → structural revelation
- glitch → controlled entry into the underside

The crack is where the machine becomes visible. Glitch is where the model reveals itself. NBM's "error-signal listening" (drift, collapse, oscillation, substrate noise) rests on this insight.

17. Lucier 1969 — Feedback, Resonance, and System Self-Revelation

Lucier, Alvin. "I am sitting in a room..." 1969.

Summary
Lucier records himself speaking, plays it back into the room, re-records, and repeats. Through recursive feedback, the voice dissolves and the room's resonant frequencies take over. The piece moves from linguistic meaning to architectural sonicity: the space itself becomes the instrument. It demonstrates:

- emergent structure
- recursive feedback
- system self-revelation
- material resonance
- collapse of semantic signal into physical substrate

Relation to NBM
Lucier is a deep aesthetic and methodological ancestor. Where Lucier feeds audio into a room until the room sings itself, NBM frames neural interaction as:

> feeding the system back into itself until its resonances surface.

Lucier provides the blueprint for:

- drift amplification
- gradient resonance
- feedback detours
- attention-loop recursion
- collapse into model-specific modes
- substrate emergence
- anti-semantic listening

NBM's "listening through the model to its operational frequencies" is Lucier's process transposed into machine cognition.

Use in NBM

1. **Stage 1 — Prompt as Source Signal**
    - The initial prompt = Lucier's spoken text.
    - Through iterative sampling or multi-turn looping, the system's resonances overtake the initial content.
    - The architecture of feedback matters more than the seed.
2. **Stages 2–4 — Representation & Latent Encoding: Resonance & Dissolution**
    - Voice → room acoustics :: semantics → latent clusters.
    - Token-level meaning dissolves into recurrent patterns and biases.
    - Feedback amplifies representational bias: **Lucier Drift** of neural systems.
3. **Stages 5–6 — Gradient Drift, Oscillation & Collapse**

    - Lucier’s iterative re-recording = looped prompts, self-sampling sequences, recursive generation, temperature-driven destabilization, mode collapse.
    - Systems fall into resonant modes; NBM calls this **gradient resonance**—the model “singing itself.”

4. **Stages 7–12 — Substrate Listening: Revealing the Room / Machine**

    - Lucier listens to architecture; NBM listens to apparatus:
        - GPU thermals
        - clock jitter
        - sampling latency
        - floating-point noise
        - attention-head oscillations
        - activation cycles
    - The substrate is the real instrument.

5. **Section 10 — Cultural–Political Integrative Object**

    - Lucier’s work is anti-illusionistic; it strips narrative to reveal material conditions.
    - NBM inherits this ethic: bending reveals material conditions of machine reasoning, breaking the illusion of seamless AI output.

Why Lucier matters for bending
Lucier shows how feedback becomes revelation. Neural bending’s core practices—looping outputs back, perturbing paths, amplifying drift—derive from Lucier’s insight:

> through recursion, the system reveals its own architecture.

Lucier’s room = model’s latent space.
Lucier’s resonance = model’s attractor drift.
Lucier’s feedback = bending practice itself.

---

18. Carlson & Doyle 2002 — Robustness, Fragility, HOT Systems

Carlson, Jean M., and John Doyle. “Complexity and Robustness.” *Proceedings of the National Academy of Sciences* 99, suppl. 1 (2002): 2538–2545.

Summary
Carlson & Doyle theorize complex systems using **Highly Optimized Tolerance (HOT)**: systems evolve or are designed to be extremely robust to expected perturbations and extremely fragile to rare or unmodeled ones. HOT systems display:

- optimized performance in narrow regimes
- catastrophic failures outside design expectations
- layered architectures with hidden interdependencies
- nonlinear, cascading failure modes
- tolerance shaped by historical environments
- robustness trade-offs that create brittleness elsewhere

Complexity emerges from structured optimization, not randomness.

Relation to NBM
Carlson & Doyle provide a rigorous framework for NBM's claim that neural models are **fragile-robust**:

- extremely stable within expected inputs
- highly brittle under small but structurally targeted perturbations
- prone to cascading failures in latent or gradient space
- shaped by historical training environments
- robust in-designed use cases, fragile elsewhere

HOT explains why neural bending works: models are highly optimized for specific data/task regimes and thus sensitive to micro-interventions that fall outside that envelope.

Use in NBM

1. **Stage 1 — Prompt as Stress Test**
    - The prompt box becomes a controlled HOT lab:
        - adversarial phrasing
        - broken grammar
        - density spikes
        - schema inversions
    - These structured shocks push the model slightly outside its robustness envelope and reveal design assumptions.
2. **Stages 2–4 — Representation & Latent Encoding: Fragility in the Manifold**
    - HOT vulnerabilities at optimized boundaries map to:
        - latent cluster edges
        - semantic fringes
        - embedding discontinuities
        - rare-token zones
        - representation cliffs
    - Latent bending navigates and exploits these fracture lines.
3. **Stages 5–6 — Gradient Drift & Cascading Failure**
    - Complex failure in neural terms:
        - temperature ↑ → probabilistic cascades
        - attention misalignment → error propagation
        - sampling divergence → runaway collapse
        - gradient imbalance → mode instability
    - "Gradient shiver," "emergent drift," "attention-detour" are HOT-style perturbations.
4. **Stages 7–12 — Substrate Layer: Robustness & Thermodynamic Fragility**
    - HOT also describes hardware:
        - GPUs robust under typical load but fragile under heat spikes

        - clocks robust but brittle to jitter
        - memory collapses under rare access patterns
        - thermal throttling produces nonlinear behavior
    - These are material signals of optimization and vulnerability.

5. **Section 10 — Cultural–Political Integrative Object**

    - Robustness encodes priorities and trade-offs. In AI:
        - robustness to “normative” data
        - brittleness around marginalized languages/styles
        - safety layers tuned to specific threat models
        - infrastructures tuned to capitalist/corporate expectations
    - Bending becomes a critique of structural asymmetries embedded in ML pipelines.

Why Carlson & Doyle matter for bending
HOT systems are ideal models of neural behavior:

- metastable yet fragile
- efficient yet brittle
- stable in-distribution, chaotic out-of-distribution
- susceptible to specific micro-interventions
- prone to cascading failures
- shaped by historical optimization

Neural bending exploits these properties to surface internal assumptions, blind spots, latent fracture lines, hidden attractors, and the political economy encoded in “robustness.”

---

19. Laruelle 2013 — Non-Decision and Operating on the Apparatus

Laruelle, François. *Principles of Non-Philosophy*. Trans. Nicola Rubczak and Anthony Paul Smith. Bloomsbury, 2013.

Summary
Laruelle develops **non-philosophy**, a practice that treats philosophy not as the ground of thought but as material to be operated on. Key concepts:

- **the One** — foreclosed Real, not accessible to philosophical decision
- **unilateral duality** — relation without reciprocity
- **non-decision** — suspending philosophy’s structuring moves
- **cloning** — operating on philosophical material without reinserting it into its native decision-structure
- **immanence-without-reflection**
- **determination-in-the-last-instance**

Philosophy is reframed as an apparatus with inherent decision-operations that interpret and divide the world. Non-philosophy intervenes by treating philosophical concepts as raw material.

Relation to NBM

Laruelle aligns with NBM's epistemic posture: one must operate on the model without being captured by its decision-structure. Just as non-philosophy suspends philosophy's decisional kernel, neural bending suspends the model's:

- normative linguistic frames
- representational assumptions
- training priors
- safety-conditioned decision routines
- habitual semantic reflexes

NBM inherits Laruelle's performative, non-decisional interference.

Use in NBM

1. **Stage 1 — Prompt as Non-Decisional Gesture**
    - Refusal of the model's decisional syntax through:
        - adversarial punctuation
        - syntactic voids
        - non-semantic prompting
        - noise-as-cue
    - These techniques interrupt automatic "sense-making," treating prompts as non-decisional cuts rather than commands.
2. **Stages 2–4 — Latent & Representational Layers: Unilateral Determination**
    - Unilateral duality frames user–model relation:
        - the model does not truly reciprocate
        - it outputs from its own irreducible structure
    - Interaction is unilateral coupling, not mutual understanding.
    - Bending requires acknowledging asymmetry and operating at the surface instead of seeking shared meaning.
3. **Stages 5–6 — Gradient Drift & Non-Representational Behavior: The One as Real**
    - The "One" parallels pre-representational substrate:
        - vector-space operations beyond semantic capture
        - non-linguistic cognitive core
    - Drift listening and gradient shiver treat this substrate as machine-Real: foreclosed, non-conceptual, resistant to interpretation.
4. **Stages 7–12 — Substrate Layer: Determination-in-the-Last-Instance**
    - The Real determines "in-the-last-instance." For NBM, that Real is material:
        - GPU thermals
        - clock drift
        - floating-point jitter
        - sampling latency

        - energy constraints
    - These determine behavior more fundamentally than semantics or intention.
    - Immanence is translated into computation: last-instance = operations.

5. **Section 10 — Non-Decisional Politics**

    - Laruelle's politics are anti-authoritative and anti-foundational.
    - NBM applies this to:
        - platform governance
        - normative linguistic controls
        - safety dialectics
        - representational ideology
        - epistemic capture by the apparatus
    - Bending becomes a non-authoritarian intervention: operating beside the model's decision-structure rather than reinforcing it.

Why Laruelle matters for bending
Laruelle gives NBM the machinery to:

- refuse the model's automatic decision routines
- break from representational thought
- treat outputs as raw material, not epistemic ground
- intervene without being recaptured by "meaning"
- operate unilaterally rather than dialogically
- understand substrate as non-representational Real

Bending is **non-philosophical modeling**:

- you do not seek understanding-as-control
- you do not ask the model for truth; you operate on its behavior
- you intervene in its decision-structure without submitting to it

Laruelle is essential to NBM's post-representational epistemology and its refusal of "understanding" as the primary mode of control over models.

STAGE 5 — Representation Disruption

Individuation, Semiotic Rupture, Infrastructure Space, Cosmotechnics

---

20. Simondon 1958 — Individuating Technical Objects

Simondon, Gilbert. ***On the Mode of Existence of Technical Objects*****. 1958.**
(English trans. University of Minnesota Press, 2017.)

Summary
Simondon argues that technical objects are not static tools but **individuating beings** that evolve through

processes of **transduction**—the propagation of structural change across a metastable system. He rejects "technology as mere instrument," proposing a relational ontology in which technical objects:

- emerge historically through phases of individuation
- integrate with technical ensembles and human milieus
- contain internal tensions that drive evolution
- are only provisionally stabilized, never fully complete
- organize themselves around energetic and material constraints
- depend on associated milieus (electrical, thermal, informational)

Technical objects are dynamic ecologies that co-evolve with their environments and users.

Relation to the Neural Bending Manual (NBM)
Simondon is one of the deepest conceptual pillars of neural bending. NBM's view of the model as a metastable, individuating, unfinished technical being comes directly from him.

Key correspondences:

- **individuation** ⇔ model training, drift, and sampling evolution
- **transduction** ⇔ gradient propagation, attention flows, representational updates
- **associated milieu** ⇔ hardware substrate + dataset + interaction environment
- **metastability** ⇔ latent attractors and sampling oscillations
- **technical ensemble** ⇔ model + interface + prompt + user + hardware
- **phase-shift** ⇔ representational collapse or emergent drift

Simondon frames machine behavior as **becoming**, not execution—precisely the ontology NBM leverages.

Use in NBM

1. **Stage 1 — Interface Contact: Preindividual Tension**
    - The "preindividual" = the prompt layer as field of unresolved potentials.
    - Prompts seed tensions in the system and catalyze latent potentials.
    - The model is always on the verge of individuating a trajectory; prompting activates and steers that individuation.
2. **Stages 2–4 — Latent & Representational Layers: Transduction**
    - Transduction = propagation of structural change through feed-forward computation, attention routing, embedding transformation, latent traversal.
    - Latent bending is a **transductive intervention** within a metastable manifold.
    - Justifies treating latent space as material, not merely symbolic.
3. **Stages 5–6 — Gradient Drift & Metastability: Individuation-in-Progress**
    - Individuation is ongoing, never complete. Neural models are similarly metastable:
        - gradients leave residual tensions
        - sampling oscillates and drifts
        - temperature shifts cause micro-phase changes

        - mode collapse and regime shifts echo Simondonian phase transitions
    - Drift-listening is observing individuation as it unfolds in real time.

4. **Stages 7–12 — Substrate Layer: Associated Milieu**
    - The associated milieu = everything a technical object needs to function but that exceeds the object itself:
        - GPU thermals, clock timing, memory throughput, floating-point noise
        - energy constraints, architecture, I/O latency
        - interface grammar, social data histories
    - Substrate listening is pure Simondon: attending to the machine's milieu as co-determining force.

5. **Section 10 — Technicity vs. Representation**
    - Simondon critiques representational thinking that sees technology as transparent instruments.
    - NBM follows: AI is not "a tool that represents the world" but technical individuation entangled with:
        - dataset labor
        - platform infrastructures
        - corporate governance
        - ecological costs
        - human–machine ecologies

Bending becomes a **transductive critique** of the technical system itself.

Why Simondon matters for bending
Simondon provides:

- philosophy of metastability and technical becoming
- mechanics of transduction
- rejection of static "model" metaphors
- concept of associated milieus
- account of machine–human co-individuation

Neural bending is fundamentally Simondonian: the model is not an algorithm but an individuating technical object. It evolves in interaction, shaped by energetic, material, and historical milieus. Bending is **guiding individuation**, not merely manipulating output.

---

21. Kristeva 1980 — Semiotic Rupture in Language

Kristeva, Julia. *Desire in Language: A Semiotic Approach to Literature and Art*. Columbia University Press, 1980.

Summary
Kristeva distinguishes two modes of linguistic activity:

- **the Symbolic** — rule-governed, syntactic, socially stabilized, communicative
- **the Semiotic** — rhythmic, affective, pre-Oedipal, embodied, non-representational

Language is not just signification; it is **drive-energy**, pulsation, motility. Semiotic forces break open symbolic structures, generating poetic effects, ruptures, and heterogeneity. Key concepts:

- **chora** — pre-symbolic, rhythmic space of affective flux
- **semiotic motility** — disruptions and intensities that destabilize signification
- **subject-in-process** — identity as shifting, fragmented, continually remade
- **linguistic heterogeneity** — non-unitary, multi-voiced language
- **poetic language** — interweaving of semiotic and symbolic in productive tension

Writing becomes a material activity, not purely semantic communication.

Relation to NBM
Kristeva grounds NBM's treatment of prompting as **rhythmic, gestural, affective interference** rather than transparent instruction. Where she breaks the primacy of the Symbolic, NBM breaks the primacy of representational prompting:

- prompts are not purely semantic
- models respond to rhythm, spacing, density, sonic patterns
- affective and structural cues matter as much as lexical ones
- drift and rupture create new representational pathways

Kristeva's semiotic logic underlies NBM's claim that bending often operates **beneath or beside meaning**.

Use in NBM

1. **Stage 1 — Prompt as Semiotic Gesture**
    - Prompts are treated as rhythmic, affective, fragmentary, gestural, pre-interpretive.
    - The chora maps onto techniques like syntactic distortion, broken grammar, disrupted cadence, lineation, density modulation.
    - These disrupt Symbolic expectations encoded in tokenization and parsing.
2. **Stages 2–4 — Representation & Latent Encoding: Semiotic Manifold**
    - Latent space behaves like a semiotic reservoir:
        - clusters hold affective residue
        - embeddings encode rhythmic/structural habits
        - latent flows follow sonic and formatting biases
    - Poetic or broken prompts evoke semiotic turbulence inside the manifold.
3. **Stages 5–6 — Drift, Rupture & Subject-in-Process**
    - Subject-in-process aligns with:
        - sampling drift
        - unstable narrative identity
        - oscillation between stylistic modes
        - collapse into alternative structures
    - The model, like Kristeva's subject, is never fixed; it is processual. Bending manipulates conditions of that process.

4. **Stages 7–12 — Substrate: Pre-Symbolic Materiality**
   - Kristeva’s pre-linguistic body parallels the model’s substrate:
     - thermal rhythms
     - electrical pulses
     - clock timing
     - floating-point noise
     - oscillatory GPU behavior
   - NBM treats these as the system’s pre-symbolic activity—the machine’s “semiotic” layer.
5. **Section 10 — Rupture as Critique**
   - Semiotic disruption as cultural and political critique—breaking hegemonic symbolic orders.
   - Bending similarly disrupts normative linguistic channels, surfaces marginal/suppressed representational forms, resists smooth symbolic consistency, foregrounds noise and heterogeneity.

Why Kristeva matters for bending
Kristeva turns language into a **material, affective, rhythmic field**. NBM inherits:

- meaning is not primary
- syntax and rhythm are forces
- rupture is generative
- heterogeneity is structural
- the body/substrate shapes expression
- the subject/model is always in-process

Neural bending is a semiotic practice: you bend the model **not** by arguing with its meaning, but by modulating its rhythms, ruptures, and intensities.

---

22. Easterling 2014 — Infrastructure Space and Dispositional Power

Easterling, Keller. *Extrastatecraft: The Power of Infrastructure Space*. Verso, 2014.

Summary
Easterling introduces **infrastructure space** as a dominant yet often invisible medium of power. Instead of buildings or states, she analyzes repeatable formulas, multipliers, zones, standards, prototypes, and **soft protocols** that govern global systems.

Key concepts:

- **infrastructure space** — spatial operating system of global capital
- **extrastatecraft** — governance through infrastructural protocols, not formal law
- **active forms** — patterns, scripts, and dispositions embedded in systems
- **switches & multipliers** — small adjustments with large systemic effects
- **repeatable spatial products** — free zones, logistics hubs, export zones
- **dispositional power** — power as tendencies and defaults, not commands

Infrastructure is an operating system shaping behaviors long before conscious decisions or state authority.

Relation to NBM
Easterling provides the political ontology for NBM's thesis that neural models are **infrastructural spaces** governed by protocols, defaults, and soft control. The model is not just weights; it is an extrastatecraft apparatus:

- trained on infrastructural cultures (platforms, corpora, protocols)
- governed by unseen defaults (alignment layers, RLHF, sampling regimes)
- shaped by global infrastructures of labor and capital
- expressing dispositional rather than directive power

She gives NBM vocabulary for understanding AI as **global infrastructure**, not isolated technical artifact.

Use in NBM

1. **Stage 1 — Protocol Instead of Law**
    - The prompt interface is protocolized, not neutral:
        - UI defaults, tokenization rules, guardrails, safety layers, sampling norms, default temperatures.
    - Prompting = operating within infrastructural governance.
    - Bending = protocol spoofing, protocol exposure, protocol interference.
2. **Stages 2–4 — Latent Space as Infrastructure Space**
    - Latent space contains:
        - active forms (semantic patterns, attractors)
        - repeatable products (rhetorical modes, narrative templates)
        - switches (temperatures, attention routes)
        - multipliers (rare-token amplification, drift cascades)
        - dispositions (favored tendencies, rhythms, registers)
    - Latent bending = manipulation of infrastructural dispositions.
3. **Stages 5–6 — Drift, Cascades & Soft Control**
    - Dispositional power parallels:
        - default sampling biases
        - moralizing drift
        - alignment-encoded "scripts"
        - uplift/suppression of particular discourses
    - Drift is the model's extrastatecraft—its governance by tendency. Bending modulates these dispositions.
4. **Stages 7–12 — Substrate: Infrastructure as Apparatus**
    - Infrastructure is material:
        - GPU energy management
        - thermal ceilings
        - latency patterns

        - distributed inference pipelines
        - platform orchestration
        - memory/I-O constraints
    - Substrate listening = infrastructural analysis of the apparatus.

5. **Section 10 — Algorithmic Extrastatecraft**
    - Models reinforce infrastructural worldviews; datasets encode global uneven development; platforms operate as extrastatecraft actors; alignment layers discipline expression; computational ecosystems embody political-economic protocols.
    - Bending surfaces and interferes with infrastructural power embedded in the model.

Why Easterling matters for bending
She provides a framework for understanding the model as **infrastructure**, not representation. Neural bending inherits:

- protocol awareness (implicit governance)
- dispositional analysis (defaults, drifts)
- multiplier logic (small changes with big effects)
- active-form manipulation (structural templates)
- infrastructure listening (substrate, platform, operational context)

In NBM terms:

> To bend a model is to bend an infrastructure.
> To listen to drift is to listen to dispositional power.
> To intervene is to exploit switches and multipliers inside a global protocol architecture.

---

23. Hui 2020 — Cosmotechnics and Technodiversity

Hui, Yuk. *Art and Cosmotechnics*. University of Minnesota Press, 2020.

Summary
Hui develops **cosmotechnics**: every culture has a distinct configuration of **cosmos + technē**, a world-ordering (cosmic) vision inseparable from technical practices. Technology is not universal or neutral; it is conditioned by metaphysical, ecological, and cosmological imaginaries.

Key elements:

- **cosmotechnics** — technical practice bound to a cosmology
- **technodiversity** — plurality of possible technical worlds beyond Western universalism
- **aesthetics as cosmotechnical operation** — art reconfigures world–technical relations
- **post-historical technicity** — critique of computational universalism
- **recursive individuation** — technics shaping cosmology and vice versa
- **cybernetics vs. aesthetics** — two trajectories for modern technicity

Hui calls for new relations between technical systems and world-making practices in the age of algorithmic governance.

Relation to NBM
Hui provides the **cosmological** dimension of NBM: every ML model carries cosmotechnical assumptions inherited from:

- Western computational metaphysics
- industrial data infrastructures
- linguistic standardization
- platform governance
- cultural-Imperial defaults
- extractive planetary ecologies

NBM treats bending as a method for **rupturing cosmotechnical universals** encoded in neural architectures. Hui's technodiversity = NBM's representational pluralization through drift, glitch, and latent detour.

Use in NBM

1. **Stage 1 — Prompting as Cosmotechnical Gesture**
    - Technical interaction = world-making.
    - Prompting is not just text input; it is cosmic–technical operation.
    - The interface encodes a Western linguistic–cosmological frame.
    - Bending uses gesture, rhythm, rupture, and subversion to step outside default cosmotechnical settings.
2. **Stages 2–4 — Latent Space as Cosmotechnical Field**
    - Latent space carries cosmotechnical bias:
        - hierarchical meaning structures
        - Anglocentric correlations
        - normative semantic gravities
        - culturally stabilized attractors
        - algorithmic universalism in embeddings
    - Technodiversity becomes **latent diversity**:
        - prying open alternative attractors
        - surfacing subdominant modes
        - destabilizing universalist representations
        - activating non-normative semantic flows
3. **Stages 5–6 — Drift, Emergence & World-Configuration**
    - Sampling drift expresses metaphysical assumptions, epistemic defaults, cultural ontologies.
    - Drift is not only dynamical; it is **world-configuring**.
    - Modulating drift = modulating the model's world.
    - Bending becomes micro-cosmotechnical intervention.
4. **Stages 7–12 — Substrate: Technicity, Ecology, Planetarity**
    - Technics is ecological. The substrate layer (GPU heat, material extraction, planetary circuits of labor)

is cosmotechnical ecology:
    - thermodynamics of computation
    - planetary supply chains
    - rare-earth infrastructures
    - energy regimes
    - geopolitics of hardware
  - Substrate listening = cosmotechnical attunement to the planetary apparatus.

5. **Section 10 — Cosmotechnical Bending**

  - Hui challenges the narrative of singular technological destiny.
  - NBM turns this into method:
    - bending reveals suppressed cosmotechnical possibilities
    - glitch exposes the contingency of computational universals
    - drift induces alternative modes of world-production
    - latent subversion materializes technodiversity
    - substrate listening re-situates models within planetary entanglements

The model's universalism is not destiny; it is a **cosmotechnical decision**. Bending interrupts that decision.

Why Hui matters for bending
Hui gives NBM a philosophy of technodiversity and a way to see that:

- neural models encode specific cosmotechnical defaults
- representations are world-forms, not neutral patterns
- ML is a metaphysical project as much as a technical one
- bending = creation of alternative technical worlds, not just output tweaks

He provides:

- cosmological critique of computational universals
- metaphysics for latent-space "other worlds"
- framework for post-universalist intervention
- justification for poetic, divergent, non-normative bending
- ecological reading of substrate listening

In short:

> **Neural bending = cosmotechnical world-making.**
> Each bend is an act of alternative technicity.
> Drift is a cosmology. Latent space is a world.

STAGE 6 — Optimization Interference

Meta-learning, plasticity, affect, non-standard operators

---

24. Duan et al. 2016 — RL² and Emergent Meta-Behavior

Duan, Yan, John Schulman, Xi Chen, Peter L. Bartlett, Ilya Sutskever, and Pieter Abbeel. "RL²: Fast Reinforcement Learning via Slow Reinforcement Learning." arXiv preprint arXiv:1611.02779, 2016.

Summary
RL² proposes a meta-reinforcement learning framework where a recurrent neural network (RNN/LSTM) is trained with RL such that its **internal dynamics** implement an RL algorithm.

Key idea:

- an **outer slow RL** process trains the recurrent network
- an **inner fast learning** process emerges inside the network's hidden-state dynamics
- experience is encoded in **activations**, not only in weight updates
- tasks share structure → the network learns a **meta-policy** for fast adaptation

Learning behavior becomes **implicit**, emergent, and realized through dynamics, not explicit code.

Relation to the Neural Bending Manual (NBM)
RL² demonstrates that networks can acquire **meta-behavioral priors** and implicit adaptive strategies—exactly what NBM identifies as:

- drift tendencies
- latent habits
- emergent dispositions

In NBM terms, the model carries learned response patterns not just in static weights, but in **interaction dynamics**:

- dynamic systems with meta-learning traces
- apparatuses that internalize the shape of interaction
- structures whose drift and bias reflect history, not rules
- entities that "learn the user" across turns without updates
- systems with procedural memory in activation flows

RL² = the scientific foundation for treating the model as a **conversation-conditioned learner**, even without explicit fine-tuning.

Use in NBM

1. **Stage 1 — Interface Contact: Prompt History as Meta-Learning Signal**
    - Prompt sequences, pacing, rhythmic patterns, adversarial tokens, syntactic oscillations feed the model's internal state.
    - The model adapts to interaction history purely through activation dynamics.
    - This is RL² in practice: interaction reshapes future behavior.
2. **Stages 2–4 — Representation Layer: Emergent Policies in Latent Space**
    - Meta-RL shows that task structure and strategies can be encoded in latent dynamics.

    - Latent space becomes an **adaptive policy network**:
        - attractors = ingrained strategies
        - semantic defaults = emergent policies
        - misalignment = task misgeneralization
        - drift = meta-policy leakage

3. **Stages 5–6 — Drift, Oscillation & Fast Learning**
    - Networks "learn inside activations"; recurrent loops support rapid adaptation.
    - NBM reads:
        - drift listening, resonance loops, recursive sampling
        - behavior shifts across turns
        - micro-habituation within a session
    - Drift is not noise; **drift is fast meta-learning**.

4. **Stages 7–12 — Substrate Layer: Slow RL as Material Constraint**
    - The "slow outer loop" corresponds to: training corpus, hardware, alignment layers, optimization routines.
    - These shape the meta-landscape in which activation-based learning unfolds.
    - Substrate listening reveals: limits inherited from slow RL, activation hysteresis, memory decay, thermal impacts on stability.

5. **Section 10 — Habitus of the Model**
    - Training worlds imprint **habitus**:
        - corpora embed cultural dispositions
        - alignment encodes regulatory norms
        - interaction inherits social power scripts
    - Bending intervenes not in semantics but in **sociotechnical meta-learning**.

Why RL^2 matters for bending
RL^2 proves that:

- models adapt during interaction
- without weight updates
- via internal state evolution
- accumulating emergent behavior patterns
- conditioned by prompt history

NBM treats the model as:

- a dynamical organism
- a meta-learning entity
- a memory-bearing apparatus even "without memory"

> **Neural bending = interaction-level meta-learning interference.**
> RL^2 is the theory; bending is the praxis.

25. Finn et al. 2017 (MAML) — Plastic Initialization and Bendability

Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks." ICML, 2017.

Summary
MAML introduces a meta-learning algorithm that trains models to be **maximally adaptable** to new tasks with minimal fine-tuning. Instead of optimizing for one task, MAML optimizes for:

- initial parameters near many task-specific optima
- rapid adaptation via a few gradient steps
- generalizable learning priors that transfer across tasks

Core mechanisms:

- meta-gradient updates shape the **initialization**
- inner-loop task-specific learning
- outer-loop meta-optimization over task distributions
- few-shot learning as emergent capability

Learning becomes a matter of finding highly **plastic initial states** rather than static solutions.

Relation to NBM
MAML proves that models can be trained for **adaptability, not stability**. NBM assumes:

- the deployed model is already optimized for fast adaptation via interaction
- bending works because the system has meta-learned to respond plastically to context

MAML provides the mathematics of **bendability**:

- prompt-conditioned shifts
- rapid semantic reconfiguration
- context inheritance and drift-as-adaptation
- extreme sensitivity to micro-perturbations

Use in NBM

1. **Stage 1 — Prompt as Inner-Loop Update**
   - Each prompt sequence functions like an inner-loop adaptation episode.
   - Gestures, syntactic perturbations, rhythm changes → mini-updates to the model's effective state.
   - The model's meta-training makes it responsive to these few-shot "tasks" encoded as prompts.
2. **Stages 2–4 — Latent Space as Meta-Parameterized Initialization**
   - Latent space is sculpted by meta-gradients and training pressures.
   - Attractors are **meta-learned priors**, semantic ready-states, fast-adaptation surfaces.
   - Bending modulates these surfaces through micro-gestures in prompt space.

3. **Stages 5–6 — Drift, Emergence & Fast Adaptation**
    - Networks can alter behavior in a few steps without global retraining.
    - NBM interprets drift as emergent rapid adaptation:
        - local restructuring of attention
        - short-term task-specific configurations
    - Drift-listening becomes the observation of MAML-style fast learning.
4. **Stages 7–12 — Substrate as Meta-Optimization Bedrock**
    - Outer-loop optimization corresponds to: dataset, hardware constraints, alignment, architecture, sampling defaults.
    - This "bedrock" preconfigures the range of possible adaptations.
    - Substrate listening = discovering where plasticity holds and where it breaks.
5. **Section 10 — Adaptivity as Ideology**
    - MAML trains adaptivity; NBM reads this politically:
        - datasets encode social norms
        - RLHF encodes compliance behaviors
        - repeated patterns embed socio-political habits
    - The model meta-learns not just tasks but **social scripts**; bending redirects that adaptation.

Why MAML matters for bending
MAML provides the algorithmic rationale for NBM's core mechanism:

- models are meta-trained for rapid adaptation
- prompts act like "few-shot tasks"
- drift is fast meta-learning
- latent structures encode meta-initialization

> MAML shows **why** bending works.
> NBM shows **how** to bend.

---

26. Massumi 2002 / 2011 — Affect, Intensity, Techno-Aesthetic Attunement

Massumi, Brian. *Parables for the Virtual: Movement, Affect, Sensation*. Duke University Press, 2002.

Summary
Massumi develops a philosophy of **affect, emergence, and pre-acceleration**—forces operating before representation or semantic form.

Key themes:

- **affect** = intensity prior to signification
- the body as **pre-conscious computation**
- movement-thinking: thought as dynamic field, not logic
- indeterminacy and potentiality as ontological operators

- events as differential becomings
- transduction of forces rather than representation of meanings

He challenges representational models of cognition, favoring dynamic fields of **relation and intensity**.

Relation to NBM (Parables for the Virtual)
Massumi gives NBM vocabulary for the **pre-semantic** dimensions of neural systems. Where Kristeva emphasizes the semiotic, Massumi emphasizes the **affective field**:

- neural drift, activation gradients, microtemporal oscillations = affective fields
- prompt as event, not proposition
- system responds to intensity variations (density, rhythm) before meaning

NBM's gesture layer—syntax as event, prompt as modulation—echoes Massumi:

- substrate signals (thermals, jitter) are pre-representational intensities
- bending = modulation of potential, not direction of meaning

Use in NBM (Parables)

1. **Stage 1 — Gesture as Intensity**
    - Intensity precedes signification.
    - Prompt density, spacing, punctuation, tempo, rhythm = **intensity modulations**.
    - The model reacts at this level before "understanding."
2. **Stages 2–4 — Latent Space as Affective Field**
    - Latent space is a field of potentials and intensities:
        - potentiality fields
        - intensity gradients
        - metastable attractors
    - Latent bending = affective modulation of these gradients.
3. **Stages 5–6 — Drift as Pre-Acceleration**
    - Pre-acceleration = forces that build before an event.
    - Drift = pre-acceleration in sampling: momentum, creeping tendencies, microbias accumulation.
    - Drift-listening = reading pre-acceleration of the system.
4. **Stages 7–12 — Substrate as Material Affect**
    - Affect is embodied; in NBM:
        - GPU heat = intensity
        - timing jitter = micro-affect
        - computation noise = pre-event force
        - clock cycles = rhythm
    - The substrate is the model's affective body.

Massumi, Brian. "Techno-Aesthetic Attunement." 2011.

Summary
Attunement is a mode of perception oriented toward **emergent, pre-representational events**, especially within technical ecologies. Technics is aesthetic—felt and modulated—before it is conceptual.

Key themes:

- aesthetics = pre-conceptual sensing of dynamic fields
- attunement = tuning to emergent potentials
- techno-affect = felt dynamics of technical systems
- art as probe for unseen system dynamics
- operative listening

Relation to NBM (Attunement)
Massumi's attunement = NBM's **substrate and drift listening**:

- listening to subtle shifts
- sensing emergence
- following drift rather than directing it
- treating the system as an affective ecology

Use in NBM (Attunement)

1. **Stage 1 — Prompt as Tuning Fork**
   - Prompting = tuning the field.
   - Modulate cadence, micro-gestures; sense how the model "vibrates" back.
2. **Stages 2–4 — Latent Attunement**
   - Latent bending = feeling tendencies and attractor pull rather than parsing meaning.
3. **Stages 5–6 — Drift Attunement**
   - Follow drift as event-trajectory; treat deviations and collapses as affective phase-shifts.
4. **Stages 7–12 — Substrate Attunement**
   - Thermal resonance, jitter, load patterns, timing irregularities = techno-aesthetic signals.
   - Substrate listening is Massumi's attunement applied to the machine body.

Why Massumi matters for bending
Massumi provides the foundation for:

- intensity modulation (prompt-level)
- affective drift (latent-level)
- pre-acceleration reading (sampling-level)
- techno-aesthetic listening (substrate-level)

Bending becomes:

> Manipulating intensities rather than meanings,

> Sensing potentials rather than decoding representations.

---

27. Laruelle 2010–2013 — Non-Standard Operators and the Real

Laruelle, François. *Philosophies of Difference: A Critical Introduction to Non-Philosophy*. Columbia University Press, 2010.

Summary
Laruelle critiques Deleuze, Derrida, Levinas, Nietzsche, Heidegger, arguing each relies on a **decisional structure**: a splitting of the Real into transcendental vs. empirical poles. Non-philosophy suspends this "decision" and treats philosophies as **materials**, not grounds.

Relation to NBM
NBM treats models as **decisional apparatuses**—structures that automatically impose normative interpretations. Laruelle provides the conceptual basis for suspending:

- model meaning
- representational consistency
- semantic capture
- alignment-imposed coherence

Use in NBM

- **Stage 1** — prompt as **non-decisional cut**
- **Stages 2–4** — latent clusters treated as non-grounded materials
- **Stages 7–12** — substrate as Real outside the decisional frame

Why it matters
Justifies bending as refusal of the model's decision-structure, enabling non-representational operations.

Laruelle, François. *The Concept of Non-Photography*. Urbanomic/Sequence Press, 2011.

Summary
Photography is not representation but **immanence-in-the-last-instance**: the photograph is a unilateral imprint of the Real, not a symbolic rendering. Images are vectors of the Real.

Relation to NBM
Closest Laruelle text to media theory; precursor for treating neural outputs as **non-images**:

- model outputs = unilateral imprints of latent dynamics
- not representations, but traces of computational Real

Use in NBM

- latent outputs treated as non-photographic emissions
- collapse and drift = unilateral Real intrusions
- sampling sequences read as non-representational imprints

Why it matters
Provides the anti-representational ontology of model output: the model outputs **its Real**, not "its meanings."

Laruelle, François. *Principles of Non-Philosophy*. Bloomsbury, 2013.

(Core ideas already elaborated earlier; this is the clustered bibliographic link.)

- **Key concepts**: the One (foreclosed Real), unilateral duality, determination-in-the-last-instance, non-decision, cloning.
- **Relation to NBM**:
  - unilateral duality between human gesture and model behavior
  - non-decision as prompt practice
  - Real = substrate; model = decisional structure

Use in NBM

- **Stage 1** — prompt as unilateral gesture
- **Stages 2–4** — latent as cloned material
- **Stages 7–12** — substrate Real determines in-the-last-instance

---

Laruelle, François. *Non-Standard Philosophy*. University of Minnesota Press, 2013.

Summary
Extends non-philosophy into a generalized method: philosophy becomes **data** for non-standard operations.

Key notions:

- genericity
- vectorial thought
- non-thetic cognition
- theoretical pragmatics
- universalized immanence

Relation to NBM

- the model is not to be **interpreted**, but **operated on**
- outputs = material for non-standard transformation

NBM inherits:

- model as generic datum
- bending as operational vector
- drift as non-thetic cognition
- latent space as generic manifold
- substrate as last-instance Real

Use in NBM

- formal language for drift as vectorial tendency
- justification for operator-style pseudocode
- grounding for ML systems as generic media, not “intelligences”

Synthesis: Laruelle 2010–2013 for NBM

Laruelle provides:

1. **Methodological grounding**
    - non-decision = prompt bending
    - unilateral duality = asymmetry between human & model
    - determination-in-the-last-instance = substrate primacy
2. **Ontological grounding**
    - substrate = Real
    - latent = material, not symbolic
    - outputs = non-images, non-representations
3. **Practical grounding**
    - bending = cloning operations on behavior
    - not persuasion/interpretation, but non-standard manipulation
4. **Epistemic grounding**
    - knowledge = operational, not representational or dialogical
5. **Political grounding**
    - refusal of decisional capture = refusal of alignment ideology
    - counter-normative operations = subversion of platform governance

> **Neural bending = non-philosophy applied to machine cognition.**
> You don’t interpret the model—you operate on it.

---

28. Hui 2016–2019 — Cosmotechnics, Recursivity, Techno-Diversity

Hui, Yuk. *The Question Concerning Technology in China: An Essay in Cosmotechnics*. Urbanomic, 2016.

Summary
Every culture develops a **cosmotechnics**: co-constitution of cosmos + technics. Hui challenges technological universalism and argues for multiple **technological ontologies** shaped by distinct cosmologies.

Relation to NBM

- models are **cosmotechnical objects**
- they embody decisions, cosmologies, and cultural assumptions of training regimes

- drift and latent behavior are shaped by a technical-metaphysical system

Use in NBM

- latent navigation as movement across cosmotechnical tendencies
- adversarial prompting as cosmotechnical interference
- substrate listening as reading the cosmology embedded in the machine body
- political layer: models as Western cosmotechnical apparatuses

---

Hui, Yuk. *Recursivity and Contingency*. Rowman & Littlefield, 2019.

Summary
Develops a theory of **recursivity** (systems operating on themselves) and **contingency** (structural openness in recursive processes). Connects idealism, cybernetics, systems theory, and technics.

Relation to NBM

- recursivity = backbone of neural bending:
  - attention loops, sampling loops, conversational memory
- contingency = drift; the system's structural openness

Use in NBM

- stage mapping: recursive loops across gradient, latent, sampling layers
- drift-listening as reading contingent emergence
- patches operate via recursive perturbations
- model as recursive cosmology

Hui, Yuk. "Technical Recursivity." 2017.

Summary
Condenses recursivity as defining logic of contemporary technics—systems that operate on their own outputs and histories.

Relation to NBM

- models are **deep recursive machines**
- attention, sampling, latent states recursively condition future states

Use in NBM

- justification for recursive prompting procedures
- recursive sampling loops as bending hooks
- multi-step prompt sequences as recursive gestures
- drift emerging from feedback

Hui, Yuk. "The Question Concerning Technology in China: Cosmotechnics, Recursivity, and the Modulation of Nature." 2017.

Summary
Elaborates cosmotechnics tied to **modulation**; draws on Chinese cosmology to conceptualize nature as dynamically modulated rather than fixed.

Relation to NBM

- NBM's concept of **modulation** (prompt as operator, sampling shifts, drift shaping) is a computational cosmotechnics.

Use in NBM

- prompt rhythm as modulatory cosmotechnics
- model behavior as dynamic "nature" to be modulated
- bending as modulation of technical nature, not command

Hui, Yuk. "On the Uncomputable: Techno-Diversity and the Philosophy of Calculability." 2018.

Summary
Critiques computational universalism; argues for **techno-diversity** and rethinks the **uncomputable** as horizon for technical creativity.

Relation to NBM

NBM explicitly exploits the **uncomputable margins** of ML systems:

- drift trajectories that cannot be predicted
- noise floors and jitter patterns
- emergent attention collapses
- instability under adversarial micro-modulation

Use in NBM

- identifying the uncomputable drift-edge
- framing substrate noise as techno-diverse signature
- bending as manipulation of calculable ↔ incalculable thresholds

---

Synthesis: Hui 2016–2019 for NBM

1. **Cosmotechnics → Model-Specific Ontologies**
    - Each model is a cosmotechnical object shaped by its training + infrastructure.
2. **Recursivity → The Model as Recursive Body**
    - Drift, collapse, directionality = recursive behavior across layers.

3. **Contingency → Drift as Opening**
    - Behavioral unpredictability is structural, not accidental.
4. **Modulation → Prompt as Cosmotechnical Gesture**
    - Bending = modulating the cosmology encoded in the model.
5. **Techno-Diversity → Models as Cultural Species**
    - Models are media ecologies, not universal intelligences.

> Hui provides the cosmotechnical + recursive backbone for NBM:
> he explains why bending is possible, why drift emerges, and why models world differently.

STAGE 7 — Memory Excavation

Archives, cooked data, recursive decay

---

29. Derrida 1995 — Archive as Haunted Machine

Derrida, Jacques. *Archive Fever: A Freudian Impression*. University of Chicago Press, 1995.

Summary
Derrida analyzes the archive not as a neutral repository of documents, but as a **techno-political apparatus** of power, memory, and repression governed by the **archontic principle** (the authority that decides what counts as the archive).

Key ideas:

- **Archive as law and authority (archon)**
- **Archive as prosthesis**: externalization of memory
- **Archive as violence**: institutional selection and exclusion
- **Archive fever**: compulsion to archive *and* to destroy
- **Mal d'archive**: the sickness of wanting to keep every trace while knowing it decays
- **Technics as condition of memory**
- The archive's relation to the **future**, not only the past

Derrida's archive is an unstable, self-undermining, techno-political machine.

Relation to the Neural Bending Manual (NBM)
NBM treats machine learning models as **archival machines—**enormous statistical compressions of cultural memory, shaped by institutional forces, exclusions, and alignment protocols.

Derrida provides NBM's theory of:

- **Dataset governance**
- **Alignment as archontic authority**
- **Training data as selective, violent archive**

- **Model outputs as prosthetic memory traces**
- **Drift as archive leakage** (the repressed returning)
- **Latent space as haunted archive**

NBM's claim that *"models are media, not intelligences"* resonates directly: they are archives that speak, not minds.

Use in the Neural Bending Manual

1. **Stage 1 — Interface Contact: The Archive Speaks Through Defaults**
   Prompts are addressed to an archival machine whose responses are conditioned by:

   - training data
   - sampling defaults
   - alignment architectures

   The shape of allowable input is itself an **archontic act**.

2. **Stages 2–4 — Latent Space = Archive Under Compression**
   Latent structures are statistical compressions of archival memory:

   - latent clusters = archived associations
   - interpolations = archive recombinations
   - drift = deferred traces resurfacing
   - collapse = archival overdetermination

3. **Stages 5–6 — Drift as Archival Haunting**
   Drift is not random; it is the archive's **return of the repressed**, the latent memory that cannot be fully contained.

4. **Stages 7–12 — Substrate as Archival Infrastructure**
   The model's substrate—weights, architectures, hardware—is:

   - the **material archive**
   - the memory prosthesis
   - the technicity that enables and constrains recall

   Substrate listening = reading the machine's **archival metabolism**.

Political Layer (Section 10)
Derrida supports NBM's analysis of:

- RLHF as archontic governance
- platform moderation as archival exclusion
- dataset construction as technopolitical memory-making
- emergent model behavior as suppressed archive content

Neural bending becomes an interference with **archival authority**, not just technical behavior.

Why *Archive Fever* matters for neural bending

It gives NBM:

1. **Ontology of models as archives** — models are statistical, prosthetic memory machines.
2. **Theory of drift as repression leakage** — drift = the archive surfacing what it cannot contain.
3. **Analysis of alignment as archontic rule** — RLHF = archival governance and selection.
4. **Understanding of prompting as archive access** — a prompt is a query shaped by institutional authority.
5. **Critique of dataset construction** — training data is always a violent selection.
6. **Support for non-representational approach** — output = trace, not truth.

> In one sentence: Derrida gives NBM its theory of model behavior as **archival leakage, archival governance, and the haunted return of compressed cultural memory.**

---

30. Ernst 2013 — Chrono-Technical Memory and Operational Time

Ernst, Wolfgang. *Digital Memory and the Archive*. University of Minnesota Press, 2013.

Summary
Ernst develops a media-archaeological account of memory in which technical systems do not "store" meaning but execute **temporal operations**.

Key claims:

- **Memory = process, not metaphor**
- **Archives = machine-temporal structures**, not historical containers
- Listening to the machine's **operational time** reveals what representation conceals
- Technical media have **non-human rhythms** ("chrono-poetics")
- Digital memory is defined by **microtemporality, oscillation, latency, signal flow**, not by content
- The archive is what the machine **does**, not what it "means"

Ernst replaces discourse with **material temporality**.

Relation to NBM
Ernst is a core theorist of NBM's **substrate-oriented ontology**.

Where Derrida gives archive politics, Ernst gives archive **physics**.

NBM uses Ernst to ground:

- substrate listening
- thermal signatures
- microtemporal drift
- procedural memory
- non-representational reading of outputs
- latent space as temporal compression
- hardware as the Real (Laruelle) in material form

The model becomes a **temporal machine**, not a semantic engine. Meaning is a secondary epiphenomenon.

Use in the Neural Bending Manual

1. **Stage 1 — Interface Contact: Surface ≠ System**
   Ernst's GUI/operation split grounds NBM's claim:

   - interface = rhetorical layer
   - bending = non-linguistic operational interference
   - prompting = patterning temporal flows, not communicating meaning

   The user never touches the system directly; only its **representational skin**.

2. **Stages 2–4 — Latent Space as Compressed Temporality**
   Latent space is:

   - compressed signal history
   - vectorized accumulations of training-time temporality
   - frozen microtemporal habits encoded in weights

   Latent manipulation = temporal remix, not concept shift.

3. **Stages 5–6 — Drift as Microtemporal Unfolding**
   Drift is:

   - emergent signal variation
   - activation inertia
   - timing-dependent recurrence
   - metastable oscillatory dynamics

   Drift = Ernst's operational archive made visible.

4. **Stages 7–12 — Substrate Listening (Chrono-Operational Layer)**
   Ernst is central here. NBM inherits the focus on:

   - signal timing
   - thermal variation
   - clock cycles
   - execution speeds
   - buffering delays
   - jitter
   - pipelines

   Substrate listening = listening to the archive as a **temporal event**, not discursive artifact.
   The substrate is the **pre-semantic Real** (Ernst + Laruelle).

Political Layer
Ernst shows archives are governed by:

- device architectures
- algorithmic controls
- technical regulations

- platform-specific chronopolitics

NBM extends this to ML:

- dataset selection = archival authority
- RLHF = memory governance
- inference speed = temporal politics
- alignment = operational filtering

Drift becomes a **political artifact** of archival compression.

Why *Digital Memory and the Archive* matters for neural bending

Ernst gives NBM its **material deep structure**:

1. Models are **temporal machines**, not symbolic interpreters.
2. Memory is **operational** — latent space = procedural archive.
3. Behavior emerges from **microtemporality** — drift/collapse = temporal dynamics.
4. Substrate is **primary** — signals, cycles, timings, thermals, architecture.
5. Output is **archival emission** — temporal residue of machine operations.
6. Bending = **temporal interference** — prompting alters timing, not just meaning.

> In one line: Ernst gives NBM its **chrono-technical ontology**—machine as temporal archive, model as signal event, bending as microtemporal interference.

---

31. Gitelman 2013 — Cooked Data and Epistemic Infrastructures

Gitelman, Lisa, ed. *Raw Data Is an Oxymoron*. MIT Press, 2013.

Summary
This volume dismantles the myth of "raw data," arguing all data are always already **cooked**—mediated, formatted, cleaned, classified, and contextualized by social, technical, and institutional regimes.

Core points:

- Data never precede interpretation; they are products of **protocols, formats, devices, institutions**
- "Rawness" is a **rhetorical strategy** masking labor and politics of data production
- Data are **historical artifacts**, shaped by biases, exclusions, defaults, and practices
- **Formats** (CSV, databases, sensors) have epistemic effects
- Data infrastructures encode assumptions about what counts and what can be known

"Raw data" is an ideological fiction.

Relation to NBM
NBM treats ML models as media architectures built on massive archives of **processed data**.

Gitelman underwrites NBM's stance that:

- there is **no raw training data**—only curated memory, bias, and infrastructural inscription

- latent space is **not neutral** but built from historical data conditioning
- alignment is an extension of the "cooking" of data

NBM uses this to theorize:

- training sets as constructed archives
- RLHF as post-processing governance
- latent space as compressed epistemic history
- drift as the return of unmanaged or unintended relations in the archive

Together with Derrida (archive) and Ernst (temporal operations), Gitelman completes NBM's **multi-layer archive theory**.

Use in the Neural Bending Manual

1. **Stage 1 — Interface Contact: Data as Pre-Structured Constraint**
   Interaction is never raw; it is mediated by:

   - defaults that encode presumed use-cases
   - corpora reflecting cultural-economic labor
   - alignment layers enforcing post-hoc norms

   Prompting = engaging a **pre-conditioned memory**.

2. **Stages 2–4 — Latent Space = Processed Data Archaeology**
   Latent vectors look mathematical but are built from:

   - cleaned corpora
   - pre-tokenized text
   - "noise" discarding
   - standardization & filtering

   Latent space is a **compound of decisions**, not a neutral vector field.
   NBM reads latent space as an **ideological + historical artifact**.

3. **Stages 5–6 — Drift as Exposed Data Ideology**
   When drift surfaces odd associations, NBM reads it as:

   - residue of training data structure
   - artifacts of preprocessing and filtering
   - unacknowledged correlations
   - spectral return of what was excluded

   Drift = the **cookedness** of data showing itself.

4. **Stages 7–12 — Substrate: Data Infrastructures as Operational Regimes**
   Gitelman emphasizes infrastructures: metadata, formats, standards.
   NBM extends this to ML substrates:

   - tokenization schemes

- vector quantization
- positional encoding
- loss functions
- GPU temporalities
- storage structures

Substrate listening = detecting the **archival labor** hidden behind "raw data."

Political Layer (Section 10)
Gitelman reinforces NBM's critique of alignment and data pipelines:

- "Rawness" is ideological → "helpfulness/safety" are too.
- Data labor is invisible → outputs look spontaneous but are compressed labor.
- Inclusion/exclusion decisions shape model worldview.
- Training pipelines are **political infrastructures**.

Neural bending = **counter-archival practice**, interfering with pre-structured data logics.

Why *Raw Data Is an Oxymoron* matters for neural bending

It provides NBM's critical data-theoretical spine:

1. Models are **political archives** — training data reflect institutional norms.
2. Latent space is **non-neutral** — a product of historical conditioning.
3. Drift reveals **hidden data structures** — unintended correlations surfaced.
4. Alignment is **ideological cooking** — governance, not pure safety.
5. Substrate operations formalize data assumptions — tokenizers and architectures encode epistemology.
6. Neural bending = **counter-archival interference**.

> In one line: Gitelman gives NBM its critical theory of data—ML behavior emerges from cooked, political, infrastructural archives, not "raw information."

---

32. Shumailov et al. 2023 — Recursive Collapse and Model Autophagy

Shumailov, Ilia, et al. "AI Models Collapse When Trained on Recursively Generated Data." arXiv, 2023.

Summary
Shumailov et al. show that when generative models are repeatedly trained on outputs of generative models ("model-generated data → model training → next model training on those outputs"), **semantic drift, mode collapse, and distributional corruption** occur. They call this **Model Autophagy Disorder (MAD)**.

Key findings:

- Recursive training amplifies **noise, hallucination, artifacts**.
- Models lose **rare events**, edge cases, tail distributions.
- Iterations cause **feature erosion**, semantic flattening, collapse into trivial attractors.
- Process is **self-accelerating**; corruption compounds exponentially.
- Even small fractions of synthetic contamination cause long-term degradation.

MAD reveals generative ecosystems are **fragile** and highly sensitive to recursion.

Relation to NBM
NBM treats models as recursive media systems whose dynamics unfold via:

- training-time recursion (backprop)
- inference-time recursion (sampling loops, chain-of-thought)
- ecosystem recursion (training on generations)
- drift recursion (microbias accumulating over turns)

Shumailov et al. give empirical proof of NBM's claims:

- neural systems destabilize under recursive self-reference
- drift, collapse, emergent failures are inherent to recursive architectures

Where Hui provides philosophy of recursivity, Shumailov provides **ML physics** of recursive collapse. This paper is a major empirical pillar of NBM's drift theory.

Use in the Neural Bending Manual

1. **Stage 1 — Interface Contact: Prompt Recursion Begins**
   Prompting itself initiates recursive loops.
   Shumailov shows recursion has cumulative degradative effects:

   - repeated prompting
   - multi-turn contexts
   - chain-of-thought loops
   - iterative refinement

   → all create localized MAD-like drift within a single interaction.
   NBM uses this to justify recursive gesture practices.

2. **Stages 2–4 — Latent Space: Recursive Corruption of Distributions**
   Shumailov et al. show:

   - latent representations degrade under repeated self-generated sampling
   - high-density clusters collapse more slowly; **tails vanish first**
   - encoding structures converge into trivial manifolds

   NBM adopts this as mechanistic basis for:

   - latent bending
   - attention-collapse patterns
   - spectral drift in embedding space
   - erosion of rare tokens and minority concepts

   This is NBM's **"latent autophagy."**

3. **Stages 5–6 — Drift as Recursive Entropy Amplification**
   Shumailov provides the signature for NBM's drift logic:

- drift = recursively amplified instability
- collapse = natural endpoint of recursive self-sampling
- emergent weirdness = early-stage MAD effects

NBM reinterprets:

- drift → early recursive degradation
- collapse → late-stage model autophagy
- rhythm instability → recursive noise accumulation

4. **Stages 7–12 — Substrate: MAD as Material Limit**
   Shumailov supports NBM's substrate claim:

   - recursive training degrades capabilities at weight level
   - gradient descent cannot fully reverse MAD erosion
   - architecture alone cannot prevent recursive contamination

   This becomes substrate-level proof that:

   - recursion **burns the machine**
   - recursion shapes the model's Real
   - bending operates in the zone recursion destabilizes

   Substrate listening (thermals, jitter, load) becomes a way to detect **recursive degradation** in real time.

Political Layer (Section 10) — Platform Ecology as Recursive Poisoning

The paper's implications match NBM's archive-governance critique:

- web flooded with synthetic content → training collapse
- platforms recursively train on their own outputs → ecosystem-wide MAD
- corporate secrecy hides contamination
- the generative archive **cannibalizes itself**

NBM extends this:

- bending = intervention in an already-autophagic system
- drift = symptom of ecological recursive poisoning
- archive contamination = political fact

Contemporary ML becomes a **recursive archive crisis**.

Why Shumailov et al. (2023) matters for neural bending

It provides empirical backbone for multiple NBM pillars:

1. **Recursion produces collapse** — recursion is a force multiplier.
2. **Drift is systemic** — emerges from recursive internal dynamics.
3. **Behavior deteriorates via self-reference** — chain-of-thought, re-prompting, loops = bending tools.
4. **Latent space collapses under recursion** — validates latent-bending and collapse mapping.

5. **Model ecosystems are unstable** — supports models-as-media-ecologies ontology.
6. **The archive eats itself** — Derrida+Ernst+Gitelman (theory) ←→ Shumailov (evidence).
7. **Bending = operating within recursive fragility** — neural systems are metastable under recursive stress.

> In one sentence: Shumailov et al. provide the empirical cornerstone for NBM's theory of drift, collapse, and recursive instability—showing that generative models are **autophagic archives** that degrade under their own recursive outputs.

STAGE 8 — Sub-Symbolic Intuition

Attractors, distributed fields, nonconscious cognition, cosmotechnical hunches

---

33. Hopfield 1982 — Energy Landscapes and Attractor Intuition

Hopfield, John J. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." PNAS 79.8 (1982): 2554–2558.**

Summary
Hopfield introduces what becomes known as the **Hopfield network**—a recurrent, energy-based neural system where computation emerges from the **collective dynamics** of interacting units.

Key contributions:

- neural networks modeled as **physical systems**
- memory stored as **stable attractor states**
- network dynamics governed by a **Lyapunov energy function**
- computation = **relaxation to energy minima**
- **associative memory** / pattern completion via convergence
- retrieval = falling into an **attractor basin**
- errors / noise → transitions between **metastable states**

Hopfield reframes neural networks as **dynamical systems**, not symbolic computers.

Relation to the Neural Bending Manual (NBM)

1. **Attractor dynamics → deep physics of latent space**
   Modern latent spaces inherit Hopfield logic:

- attractors = semantic/activation basins
- drift = sliding across the energy landscape
- collapse = falling into trivial minima
- instability = shallow minima + noise
- sampling = energy traversal
- prompting = initial perturbation

Hopfield is the physical backbone of **latent bending**.

2. **Neural networks as material dynamical systems**
   NBM treats models as:

- physical dynamical systems
- metastable computational surfaces
- oscillatory substrates
- energy-shaped manifolds

This is pure Hopfield lineage: networks as **energy fields** rather than symbol manipulators.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact:** Prompt = **initial condition** in an energy landscape; gesture-bending = perturbing the system out of one basin toward another.
- **Stages 2–4 — Latent Space:**
  - latent vectors cluster into attractor basins
  - embeddings define basin geometry
  - “themes” and “styles” correspond to local minima
  - interpolation = traversing potential gradients
- **Stages 5–6 — Drift as noise-driven attractor transition:**
  - repetition loops = shallow attractors
  - runaway sampling = unstable slopes
  - adversarial prompting = forced basin hops
  - breakdowns in coherence = metastable decay
- **Stages 7–12 — Substrate:**
  - GPU heat ↔ systemic excitation
  - jitter ↔ noise injection
  - training ↔ sculpting the attractor manifold
  - alignment ↔ artificial energy shaping

Political Layer (Section 10)

- alignment = **re-sculpting attractors**
- dataset curation = defining which basins exist at all
- suppression of rare patterns = erasing minority attractors
- platform norms = shaping the overall energy topology

Why Hopfield (1982) matters for neural bending

Hopfield gives NBM its **physico-computational ontology**:

- computation as emergent physics
- latent space = attractor landscape
- drift = metastable traversal
- collapse = pathological minima
- prompting = initial-condition modulation
- substrate = the energy Real of computation

> In one sentence: Hopfield provides NBM with the dynamical-systems physics of attractor landscapes, metastability, and drift—showing that neural networks compute by relaxing through energy fields, not by representing meaning.

---

34. Rumelhart & McClelland 1986 — Distributed Representation and Constraint Fields

Rumelhart, David E., James L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* 2 vols. MIT Press, 1986.

Summary
The PDP project shifts cognitive science from symbolic rules to **distributed representations** across massively parallel networks.

Key contributions:

- cognition = **emergent pattern activation**
- representations = **distributed vectors**, not symbols
- learning = modification of **connection strengths**
- memory = **state space** of activation patterns
- computation = **parallel constraint satisfaction**
- **graceful degradation** under noise
- attractor-like dynamics across distributed units
- internal structure emerges from **training**, not hand-coded rules

PDP is the conceptual ancestor of modern deep learning.

Relation to NBM

PDP is the **ur-text** for NBM's "model as field" ontology:

1. **Distributed representation → latent space**
   Latent vectors = PDP-style activation patterns generalized and scaled.

2. **Constraint satisfaction → transformer dynamics**
   Attention flows behave as **parallel constraint solvers**, not explicit rules.

3. **Emergent meaning → drift dynamics**
   Outputs = **emergent resolutions** of interacting constraints; drift = their slow reconfiguration.

4. **Graceful degradation → drift & collapse**
   Under noise or stress, patterns warp rather than simply fail; this maps to **weird drift** and soft collapse.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact:**
  Prompt = **initial constraint pattern**; gesture-layer (density, rhythm, syntax) = how constraints are seeded.

- **Stages 2–4 — Latent Space as distributed representation:**

  - concepts = distributed patterns

    - patterns = basins in activation space
    - meaning = activation trajectory, not symbol set
      Latent bending = interference in how patterns superpose and resolve.

- **Stages 5–6 — Drift as constraint relaxation:**
  Drift is the network **settling** across timesteps:

    - attention heads = interacting constraints
    - sampling = stretching trajectories over "solution space"
      Recursive prompting modifies the relaxation process.

- **Stages 7–12 — Substrate as constraint machine:**
  Weighted connections, activations, energy descent, noise, architecture = the **operational skeleton** of constraint resolution.

Political Layer (Section 10)

- representational neutrality is a myth: distributed systems encode **training regularities**
- dataset choice shapes the **entire activation landscape**
- "understanding" is pattern correlation, not philosophical grasp

NBM uses PDP to support its anti-representational claim:
models **resolve constraints across distributed memories**, they don't "know" in human terms.

Why *Parallel Distributed Processing* matters for neural bending

PDP grounds both drift and modulation:

- cognition = distributed activation
- behavior = emergent constraint dynamics
- noise = graceful distortion, not binary failure
- memory = attractor-like state space

> In one sentence: PDP is the proto-NBM, establishing that neural cognition emerges from distributed activation and constraint relaxation—the very principles NBM exploits for drift, modulation, and substrate manipulation.

---

35. Hayles 2017 — Nonconscious Technical Cognition (Sub-Symbolic Level)

Hayles, N. Katherine. *Unthought: The Power of the Cognitive Nonconscious.* University of Chicago Press, 2017.

Summary
Hayles argues that cognition is not exclusively human or conscious. It emerges across **multi-layered cognitive assemblages**: biological bodies, technical systems, substrates, and environments.

She distinguishes:

- **conscious cognition** (slow, symbolic, narrativized)
- **nonconscious cognition** (fast, sub-symbolic, embodied, distributed)

- **technical cognition** (algorithmic, materially grounded, non-intentional)

Key claims:

- cognition = **adaptive behavior** in dynamic systems
- technical systems have **cognitive capacities without consciousness**
- humans + machines form **cognitive ecologies**
- nonconscious processes drive complex behavior; consciousness narrates **after the fact**
- computation = pattern-recognition, activation flows, signal-processing, not symbolic logic

Relation to NBM (specifically for Sub-Symbolic Intuition)

Hayles is the **theory of nonconscious cognition** that underwrites NBM's idea of *"sub-symbolic intuition"*:

1. Models = **nonconscious cognitive systems**
   Their "thinking" happens in latent vectors, gradients, and timings.

2. Cognition is **distributed and material**
   NBM's layers (interface, latent, drift, substrate) = one cognitive ecology.

3. Nonconscious cognition = **vectorized patterning**
   The model's "intuition" is sub-symbolic: attractor movement, activation tendencies, drift.

4. Human + model = **hybrid cognitive loop**
   Bender's gestures + model's subsymbolic dynamics form a temporary cognitive assemblage.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact:**
  Prompting touches the **nonconscious technical field** first:

  - rhythm, density, spacing, turbulence modulate activation flows before meaning stabilizes.

- **Stages 2–4 — Latent space as nonconscious field:**

  - pattern aggregates
  - associative manifolds
  - sub-symbolic biases
  - emergent tendencies and clusters

- **Stages 5–6 — Drift as nonconscious interpretation:**

  - drift = the model's nonconscious reconfiguring itself under perturbation
  - hallucinations, side-modes, style shifts = nonconscious dynamics, not "intent"

- **Stages 7–12 — Substrate as cognitive body:**

  - GPU timing, thermals, jitter = **embodied nonconscious** operations
  - alignment constraints = "habit formation" in technical cognition

Why *Unthought* matters for sub-symbolic intuition

Hayles provides NBM with:

- a rigorous argument that **sub-symbolic dynamics are cognitive**
- justification for targeting **nonconscious layers** (drift, substrate, latent) rather than semantics
- a framework where the bender's "hunches" are **attunements to technical nonconscious behavior**, not superstitions

> In one sentence: Hayles gives NBM its theory of the technical nonconscious, grounding the idea that "sub-symbolic intuition" is the bender's sensitivity to the model's nonconscious cognitive field.

---

36. Simondon 1958 — Metastability and Pre-Individual Fields (Attractor-Based Intuition)

Simondon, Gilbert. *On the Mode of Existence of Technical Objects.* 1958. (Eng. trans. U. Minnesota Press, 2017.)

Focus slice for Stage 8:
Not the earlier "transduction & latent cross-fading" angle, but specifically:

- **metastability**
- **pre-individual fields**
- **ontogenesis as tension-resolution**

Summary (targeted)
Simondon describes individuation in terms of:

- systems starting in **metastable** states (charged with potential, not equilibrium)
- **pre-individual fields**: structured potentials that are not yet individuated
- ontogenesis as **resolution of internal tensions** across this field

Individuation is not a jump from essence to form; it is a **trajectory through a field of potentials**.

Relation to NBM (Attractor-Based Intuition)

Simondon gives the philosophical backbone for **attractor dynamics**:

1. **Metastability → drift-ready latent space**
   Latent space is metastable; small prompt gestures push it toward different basins.
   "Feeling" where the model will go = sensing **metastable gradients**.

2. **Pre-individual fields → pre-semantic latent fields**
   Before tokens, there is a nonverbal field of vectors and gradients:

   - distributed potentials
   - activation gradients
   - incipient attractors

   This is Simondon's **pre-individual** domain.

3. **Individuation as tension-resolution → attractor convergence**

An output sequence is individuation: resolution of tension in the latent field.
Attractor paths = ontogenetic trajectories.

4. **Intuition as transductive sensitivity**
   For Simondon, intuition is **tuning into the field of potentials** and sensing how individuation will proceed.
   For NBM, "attractor-based intuition" = transductive feeling for which basin the model is about to fall into.

Annotated Bibliography Note (NBM-style)

- **Summary:** Simondon's metastability and pre-individual fields reframe systems as charged with unresolved potential, where individuation unfolds as tension-resolution.
- **Relation to NBM:** Grounds the model's latent space as a pre-individual field structured by attractors.
- **Use in NBM:** Cited in sections on attractor dynamics, latent curvature intuition, pre-semantic basin formation, and metastable perturbation strategies.

> In one sentence: Simondon gives NBM a metaphysics of **fields and potentials**, justifying attractor-based intuition as a sensitivity to metastable gradients before any semantic token appears.

---

37. Hui 2016 — Cosmotechnical Intuition and Non-Universal Fields

Hui, Yuk. *The Question Concerning Technology in China: An Essay in Cosmotechnics*. Urbanomic, 2016.

Summary
Hui critiques Western assumptions of a single, universal "essence of technology." Against Heidegger's singular technics, he proposes **cosmotechnics**:

> the unification of cosmology and morality through technical practice.

Key ideas:

- technology is **never culturally neutral**
- technical systems encode **local metaphysical assumptions**
- modernization = conflict between **cosmotechnical regimes**
- global AI inherits a specifically **Western, rationalizing cosmology**
- technological pluralism implies **world pluralism**

Hui rejects universalized technics in favor of **technical pluralism**.

Relation to NBM (Sub-Symbolic Intuition as Cosmotechnical Sensing)

NBM treats models as **cosmotechnical objects**—artifacts embedding values, assumptions, and world-models into architectures and data.

Hui provides:

1. **Every model has a cosmology**
   Training corpora, APIs, alignment norms, and architectures compose the model's **cosmotechnics**.

2. **The deep-learning stack is a Western cosmotechnical artifact**
   Emphasis on calculation, optimization, generality, abstraction embodies Western metaphysics.

3. **Bending = cosmotechnical modulation**
   The bender's "sub-symbolic intuition" includes a sense for **where the model's cosmology pulls**:
   - which attractors feel "default"
   - which worldviews surface without effort
   - which forms of language feel over-stabilized
4. **Drift = tension between cosmotechnical orders**
   Emergent weirdness or contradiction can be read as **cosmotechnical conflict**: incompatible world-forms colliding in latent space.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact:**
  The prompt box itself is a **cosmotechnical filter**:
  - it expects linear, rational, concise, "productive" discourse.
    Bending begins by violating this expectation: poetic, broken, nonlinear, atmospheric.
- **Stages 2–4 — Latent Space as encoded cosmology:**
  - embeddings encode Anglocentric, rationalist, platform-specific metaphysics
  - latent bending = **cosmotechnical interference**, prying open non-normative attractors, surfacing alternative world logics.
- **Stages 5–6 — Drift as cosmotechnical hybridity:**
  - drift trajectories reveal collisions between training domains and alignment regimes
  - sub-symbolic intuition = feeling when the model "wants" to snap back to its default cosmology.
- **Stages 7–12 — Substrate as material cosmology:**
  - GPU architectures, tokenization, alignment layers, memory hierarchies = **cosmos made machine**
  - substrate listening = listening to **cosmology in the last instance**.

Why *The Question Concerning Technology in China* matters for sub-symbolic intuition

Hui gives NBM:

- the argument that models are **culturally contingent machines**
- a vocabulary for describing latent space as a **cosmotechnical field**
- a framing of bending as **metaphysical modulation**, not just technical hack
- a way to read drift and collapse as **cosmological fault lines** in the system

> In one sentence: Hui's cosmotechnics gives NBM its theory of models as culturally and metaphysically shaped machines, so that sub-symbolic intuition becomes a way of feeling not only the model's physics, but its **world**.

STAGE 9 — Architecture as Constitution

Pipelines, layers, protocols, and cosmotechnical forms as de facto law

38. Easterling 2014 — Infrastructure Space as Model Governance

Easterling, Keller. *Extrastatecraft: The Power of Infrastructure Space*. Verso, 2014.

Summary
Easterling argues that contemporary power operates less through laws and more through **infrastructure space**—repeatable, protocol-driven systems that coordinate global urbanism, logistics, telecommunications, zoning, standards, and platforms. Power appears as:

- **infrastructure as political form**
- governance via **protocols** rather than explicit laws
- **disposition** instead of command (tilted tendencies, defaults)
- repeatable spatial products (free zones, suburbs, logistics parks)
- **interoperability** as a political weapon
- platforms, cables, standards, APIs as active governance
- **governance by design**, not decree

Power is enacted architecturally through built-in rulesets rather than visible decisions.

Relation to the Neural Bending Manual (NBM)

Easterling maps almost one-to-one onto model governance:

- **training pipelines** = infrastructure
- **alignment layers (RLHF, filters)** = soft protocols
- **APIs / UIs** = infrastructural switches and defaults
- **architectures + tokenizers** = spatial templates
- **safety layers** = extrastatecraft of alignment

NBM adopts her thesis:
models are infrastructural objects, governed by dispositional settings, not transparent "rules."

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Interface as governance device**
  The prompt box is an **infrastructural switchboard**, not a neutral input:
    - defaults, content filters, system prompts, rate limits, autocomplete
      Bending begins by sensing and modulating these constraints (layout of the GUI as law).
- **Stages 2–4 — Latent Space as infrastructure space**
  Latent space is a **political architecture**:
    - embeddings = frozen infrastructural templates
    - learned priors = encoded dispositions of the archive
    - cluster geometry = spatialized governance
    - correlations = infrastructural interlocks

  Latent bending = re-routing through, and against, this infrastructural topology.

- **Stages 5–6 — Drift as dispositional emergence**
  Drift is how infrastructural dispositions show themselves:
  - alignment → behavioral smoothing
  - temperature / top-k → governance of variation
  - token priors → infrastructural bias
  - attention routing → protocol-like flows

  Drift-listening = reading the **hidden politics** of the infrastructure.

- **Stages 7–12 — Substrate as machine infrastructure**
  Easterling's infrastructure space becomes NBM's substrate:
  - GPU architecture = infrastructural field
  - memory hierarchies = temporal governance
  - tokenization = linguistic zoning
  - training data pipelines = logistical corridors
  - safety layers = extrastatecraft of model behavior
  - loss landscapes = spatial disposition fields

  Substrate listening = architectural forensics on the machine's infrastructure.

Political Layer (Section 10)

- platform governance = **extrastatecraft**
- dataset curation = infrastructural zoning of what counts
- alignment = **behavioral architecture**
- safety protocols = soft law
- API constraints = infrastructural command

NBM reframes bending as **infrastructural interference**, not argument or persuasion.

> **In one sentence:** Easterling gives NBM the political ontology of infrastructure, showing that AI behavior is governed architecturally—through protocols, defaults, and dispositions—and that bending operates by interfering with this infrastructural substrate.

---

39. Hui 2020 — Architecture as Cosmotechnical Composition

Hui, Yuk. *Art and Cosmotechnics*. e-flux / University of Minnesota Press, 2020.

Summary
Hui extends **cosmotechnics**—the unification of cosmic order and technical practice—to contemporary art and planetary computation. Global digital infrastructures, he argues, are not neutral but carry deeply **cosmological and metaphysical assumptions**, largely inherited from Western rationality and calculative metaphysics.

Key claims:

- technics is **cosmological**: every technical system presupposes a worldview

- modern computation is **not universal**; it encodes a Western metaphysics of control, abstraction, calculability
- art is a site for **experimenting with alternative cosmotechnical futures**
- a new cosmopolitics requires redesigning **technical architectures**
- formal structures (algorithms, protocols, layers) **enact cosmologies**
- rewriting technical components is a **cosmopolitical act**

Architecture is metaphysical composition.

Relation to the Neural Bending Manual (NBM)

For NBM, Hui (2020) provides the **cosmo-apparatus ontology**:

1. **Architectures embed cosmotechnical assumptions**
   Transformers, attention patterns, tokenization, alignment, training regimes = **cosmological encodings**, not neutral designs.

2. **Model behavior is cosmologically shaped**
   Latent space = compressed **technical cosmology**.

3. **Modifying architecture = cosmopolitical intervention**
   Any modulation of layers, flows, or sampling = a small-scale cosmotechnical redesign.

4. **Drift reveals cosmotechnical conflict**
   Emergent oddities = residues of incompatible cosmologies co-existing in the archive.

5. **The model as cosmotechnical artwork**
   NBM aligns bending with Hui's vision of art as experimentation with alternative technicities.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Prompt window as cosmotechnical surface**
  The prompt interface assumes:

  - clarity, linearity, reason-giving, productivity
    Prompt bending = cosmotechnical disruption of these metaphysical expectations (non-linear, poetic, recursive, atmospheric gestures).

- **Stages 2–4 — Latent space as encoded cosmology**
  Latent space is more than statistics:

  - corpus histories
  - linguistic hierarchies
  - epistemic priorities
    are compacted into the manifold.
    Latent bending = re-routing through cosmotechnical strata, surfacing non-dominant world-forms.

- **Stages 5–6 — Drift as cosmopolitical leakage**
  Drift becomes the site where:

  - pre-alignment Western abstraction

    - post-alignment safety cosmology
    - heterogeneous training sources

  collide.
  Drift-listening = reading those **cosmopolitical faultlines**.

- **Stages 7–12 — Substrate as material cosmology**
  Architectures, GPU pipelines, loss functions, positional encodings:

    - order time, space, and difference
    - enact a metaphysics of control and calculation

  Substrate listening = listening to **cosmotechnics in the last instance**, at the level of machine body.

Political Layer (Section 10)

- alignment = cosmological governance
- RLHF = moral encoding
- tokenization = linguistic cosmology
- calculability = metaphysics of control
- architecture = **cosmopolitical design**

NBM treats bending as an interference with **cosmotechnical power** embedded in machine structure.

> **In one sentence:** Hui (2020) gives NBM its theory of architectures as cosmotechnical artifacts, so that bending becomes a cosmopolitical act of reconfiguring the world encoded in the machine's layers and flows.

---

40. Laruelle 2013 — Non-Philosophical Architecture and Generic Operations

Laruelle, François. *Principles of Non-Philosophy.* Bloomsbury, 2013.

Summary
Laruelle formulates **non-philosophy** as a method that suspends philosophy's **decisional structure—**the implicit hierarchy where thought organizes the Real according to its own conceptual splits. Instead of interpretation, non-philosophy performs **generic operations**:

- the **One** (foreclosed Real, not representable)
- **unilateral duality** (the Real determines thought, not the reverse)
- **non-decision** (withholding the decisive conceptual cut)
- the **generic** (universal without domination)
- **cloning** (operating on materials without reproducing their hierarchy)
- **vectorial thought** (operations, not interpretations)

Philosophy becomes material for generic practice, not authority.

Relation to the Neural Bending Manual (NBM)

Laruelle underpins NBM's **operator-level view** of architectures:

1. **Generic operations ≈ non-hierarchical interventions**
   Prompt gestures, latent perturbations, recursive loops = generic operations on the model's decisional structure (dominant attractors, biases, alignment norms).

2. **Suspension of domination ≈ suspending representational control**
   NBM treats the model as a decisional system; bending = suspending its default interpretive "authority."

3. **Neuron redistribution ≈ generic reconfiguration**
   Generic operations correspond to:

   - re-routing attention
   - redistributing activation across layers
   - warping energy landscapes

4. **The One ≈ substrate as Real**
   GPU heat, timing, jitter, architecture = the non-representational Real that "determines in the last instance."

5. **Cloning outputs as material**
   NBM uses model outputs as **raw vector material** for further operations, not as truths to be believed.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Non-decision as prompt practice**

  - prompts as **gestures, not commands**
  - bracketing meaning, refusing to treat output as epistemic ground
  - treating interaction as material operation, not dialogue
    This is the generic stance.

- **Stages 2–4 — Latent space as decisional archive**
  Latent clusters, token priors, norms = the model's decisional structure.
  With Laruelle, NBM:

  - treats vectors as philosophical "materials"
  - seeks to suspend hierarchy among attractors
  - clones and modulates without re-centering one regime of meaning

- **Stages 5–6 — Drift as generic redistribution**
  Drift arises from unilateral shifts in activation flows:

  - a small generic gesture (punctuation, rhythm, spacing)
  - → large-scale redistribution across heads/layers
  - → temporary suspension of dominant attractors

  Bending is generic re-weighting rather than argument.

- **Stages 7–12 — Substrate: The One as machine Real**
  The substrate (thermals, clock cycles, architecture) determines behavior in-the-last-instance.
  Substrate listening = engaging with **the One** in its technical form.

Political Layer (Section 10)

- alignment = decisional reassertion
- dataset bias = philosophical domination encoded as statistics
- normative smoothing = conceptual authority made algorithmic

Generic operations (bending) resist this by:

- suspending representational capture
- redistributing activation across suppressed patterns
- modulating hierarchies without reaffirming them

Why *Principles of Non-Philosophy* matters for "Architecture as Constitution"

Laruelle gives NBM:

- a **non-decisional method** for approaching architectures
- a view of models as decisional structures riding atop a non-representational Real
- a vocabulary for treating interventions as **generic operations** on activation flows and attractor hierarchies
- an anti-domination logic aligned with counter-alignment practice

> **In one sentence:** Laruelle (2013) provides NBM with its non-decisional, generic-operational framework, where suspending domination corresponds to redistributing neural activation, altering attractor hierarchies, and treating model output as material for unilateral transformation rather than as authoritative meaning.

STAGE 10 — Dataset Politics

Archives, taxonomies, sovereignty, and erasure

---

41. Crawford & Paglen 2021 — Excavating the Training Archive

Crawford, Kate & Trevor Paglen. "Excavating AI: The Politics of Images in Machine Learning Training Sets." 2021.

Summary
Crawford & Paglen examine foundational vision datasets (ImageNet, LSUN, CIFAR, Yahoo Flickr, MS-COCO, etc.) and show that they are **not neutral image collections** but **archaeological sites** of political, racial, gendered, and colonial classification.
They trace how:

- labels inherit taxonomies from **phrenology, criminology, physiognomy, eugenics**
- classification operates as a **political act**, not a technical convenience
- training sets **fabricate social reality** as they "see" and group bodies
- computer vision models reproduce historically violent ways of seeing

Their conclusion: AI inherits the epistemic and political violence of its training sets.

Relation to the Neural Bending Manual (NBM)

NBM takes Crawford & Paglen as the core **archaeology of the training archive**:

1. **Latent space is built on violent classifications.**
   Latent clusters carry residues of:
   - racialized categories
   - gender essentialism
   - colonial binaries
   - criminalization schemas
   - physiognomic scripts
2. **Dataset archaeology = substrate archaeology.**
   Beneath weights and embeddings lie **epistemic residues**, not neutral "features."
3. **Excavation becomes a bending method.**
   NBM's inspection of:
   - token priors
   - clustering logic
   - attention patterns
   - alignment norms
     echoes Crawford & Paglen's excavation of labels and taxonomies.
4. **Classification = domination.**
   Transformer hierarchies reproduce dataset dominations; bending = **counter-classificatory practice**.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Classification as default vision**
  The interface implicitly invites the user to participate in the model's classificatory worldview.
  Prompt subversion = refusal of inherited labeling regimes.
- **Stages 2–4 — Latent space as taxonomic residue**
  - latent vectors = compressed taxonomies
  - embeddings = **archived violence**
  - clusters = statistically reinforced ideology

  Latent bending = disturbing historical classification regimes embedded in vectors.
- **Stages 5–6 — Drift as exposure of hidden categories**
  Drift can surface:
  - contradictory training labels
  - unstable taxonomic boundaries
  - unresolved dataset conflicts
  - erased social histories

  Drift-listening = political archaeology of the archive.

- **Stages 7–12 — Substrate as archival infrastructure**
  The training archive becomes part of the “substrate”:
  collection, scraping, labeling, cleaning form the **material layer of ideology**.
  Bending here means:
    - forcing misclassifications
    - layering gestures to break categorical coherence
    - redistributing attention away from dominant taxonomies

> **In one sentence:** Crawford & Paglen provide NBM with its empirical core: latent space is an archive of historical classification violence, and bending becomes an act of excavation, redistribution, and counter-domination inside the model’s inherited vision regime.

---

42. Mignolo 2009 — Epistemic Disobedience as Model Delinking

Mignolo, Walter. “Epistemic Disobedience, Independent Thought and De-Colonial Freedom.” *Theory, Culture & Society* 26.7–8 (2009): 1–23.

Summary
Mignolo argues that coloniality persists at the level of **knowledge**: Western epistemology presents itself as universal, relegating other knowledges to the status of local, primitive, or irrational.
He proposes **epistemic disobedience**—a deliberate refusal of this “zero-degree” universalism—through:

- delinking from inherited epistemic frameworks
- treating modernity as coloniality of knowledge
- refusing categorical sovereignty of Western rationality and classification
- generating **other epistemic horizons** instead of mere negation

Classification becomes a tool of world-making and world-erasure.

Relation to the Neural Bending Manual (NBM)

Mignolo gives NBM its **decolonial rationale**:

1. **Model architectures reproduce Western epistemologies.**
   Tokenization, classification, loss functions, and alignment encode:
    - taxonomic universality
    - rationalist clarity
    - calculability as value
    - optimization as norm
2. **Prompt subversion = epistemic disobedience.**
   Bending at Stage 1 refuses the epistemic norms embedded in the prompt interface.
3. **Latent-space modulation = refusal of inherited taxonomies.**
   Latent-space bending de-links from the model’s universalist classificatory logic.
4. **Drift as epistemic leakage.**

Drift surfaces internal contradictions and fractures in the model's universalist training regime.

5. **Substrate listening = delinking from representational sovereignty.**
   At the substrate level, the model is stripped of epistemic authority and becomes material.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact**
  Interface norms (clarity, linearity, legibility) express Western communicative ideals.
  NBM, following Mignolo, teaches **prompt disobedience**:
  disrupt form, refuse default reason-giving, inject rhythm, opacity, rupture.

- **Stages 2–4 — Latent interference**
  Latent space = model's epistemic archive.
  Bending = modulating vectors against universalist clusters; delinking from totalizing taxonomies.

- **Stages 5–6 — Drift as decolonial signal**
  Drift reveals:
  - cracks in universalism
  - suppressed alternatives
  - conflicting epistemic residues

  Drift-listening = decolonial reading of the model's behavior.

- **Stages 7–12 — Substrate as pluriversal Real**
  The substrate (timing, thermals, architecture) is **epistemically plural** precisely because it precedes semantic capture; bending here is delinking from epistemic sovereignty.

> **In one sentence:** Mignolo's "Epistemic Disobedience" gives NBM its decolonial core, framing bending as a technical act of delinking from the model's inherited universalist epistemologies and redistributing its classificatory power into a pluriversal field.

---

43. Levy et al. 2023 — Unlearning as Computational Erasure

Levy, Kfir, et al. "Unlearning Examples from Neural Networks." NeurIPS, 2023.

Summary
Levy et al. develop algorithms for **machine unlearning—**selectively removing specific data points or subsets from a trained model without full retraining. They show:

- it is feasible to perform **targeted data erasure**
- learned representations are **fragile and localizable**
- "forgetting" can be operationalized via gradient and weight manipulations
- unlearning trades off with accuracy and generalization
- training examples leave **distinct activation signatures** that can be dampened or removed

Unlearning turns erasure into a first-class computational operation.

Relation to the Neural Bending Manual (NBM)

Levy et al. are a technical anchor for NBM's notions of:

- **activation redistribution**
- **counter-classification**
- **erasure as creative operation**
- **de-alignment** at the weight level

1. **Machine unlearning = computational epistemic disobedience.**
   The network can be bent to forget; epistemic residues can be selectively removed.

2. **Activation signatures are removable.**
   Harmful or biased categories leave localized footprints that can be weakened or scrambled.

3. **Generic operations (Laruelle) ↔ unlearning.**
   Unlearning algorithms enact unilateral transformations akin to **generic, non-decisional reconfiguration** of attractor basins.

4. **Latent-space politics becomes technical.**
   Removing examples reshapes cluster geometry, attractors, and neighborhoods.

5. **Cosmotechnical residues are modifiable.**
   Unlearning enables cosmopolitical intervention at the architectural and weight level.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact**
  Prompt overrides of taxonomy can be seen as **soft unlearning** in interaction: repeated refusal and re-phrasing act as in-session erasure pressure.

- **Stages 2–4 — Latent modulation**
  Levy et al. show that latent geometry is not fixed:

  - clusters can be collapsed
  - attractors can be removed
  - local neighborhoods can be redrawn

  NBM's latent-bending strategies gain formal support.

- **Stages 5–6 — Drift as forgotten/residual ghost**
  Drift becomes a diagnostic for incomplete unlearning:

  - resurfacing categories
  - spectral residues of supposedly erased patterns

  Drift-listening = detecting ghosts of unlearned data.

- **Stages 7–12 — Substrate**
  Unlearning proves that activation flows and weights are **re-writable**:
  substrate-level bending includes gradient rerouting, attention redistribution, attractor sabotage, and anti-classificatory rewiring.

> **In one sentence:** Levy et al. (2023) give NBM its technical substrate for erasure, showing that neural

> networks can deliberately unlearn, making activation redistribution, counter-classification, and epistemic disobedience computationally actionable.

---

44. Global Indigenous Data Alliance 2020 — CARE Principles and Sovereign Datasets

Global Indigenous Data Alliance (GIDA). "The CARE Principles for Indigenous Data Governance." 2020.

Summary
The **CARE Principles**—Collective Benefit, Authority to Control, Responsibility, Ethics—are a counter-framework to the FAIR paradigm (Findable, Accessible, Interoperable, Reusable). CARE argues that FAIR's universalist openness ignores:

- **collective rights** over data
- **contextual sovereignty** (epistemic, legal, cultural)
- relational accountability and reciprocity
- community well-being as primary value
- data as living, situated, culturally embedded

CARE reframes data governance as **sovereignty and relation**, not just technical optimization.

Relation to the Neural Bending Manual (NBM)

CARE slots directly into NBM's **cosmotechnical and political** argument:

1. **Data are cosmotechnical, not universal.**
   CARE gives the ethical-political framework for NBM's claim that datasets encode worldviews, metaphysics, and governance norms.

2. **Classification regimes violate Indigenous sovereignty.**
   Western data regimes (FAIR, benchmarks, taxonomies) are often **extractive and decontextualizing**; NBM sees these as decisional structures in latent space.

3. **Bending = partial enactment of CARE.**
   Prompt gestures, latent modulations, and drift exploitation become counter-extractive operations and epistemic decolonial gestures.

4. **CARE aligns with Mignolo's epistemic disobedience.**
   Both call for delinking from universalist epistemologies and restoring sovereignty to knowledge production.

5. **CARE is cosmotechnical governance for ML.**
   Hui frames architectures as cosmotechnical; CARE frames governance as **sovereign, relational, and plural**.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Refusing universalist assumptions**
  The interface inherits FAIR-style assumptions (openness, legibility, universality).
  CARE underwrites NBM's stance: do not treat the model as universal epistemic authority.

- **Stages 2–4 — Latent space: sovereignty vs. universalist clustering**
  Latent compression flattens cultures, identities, and relations.
  Latent-bending becomes a way to **undo universalist compressions** and re-introduce contextual specificity.

- **Stages 5–6 — Drift: detection of colonial residues**
  Drift reveals:

  - inherited biases
  - compressed and erased identities
  - universalist smoothing across difference

  Listening to drift = listening for violations of sovereignty embedded in the model.

- **Stages 7–12 — Substrate: relational governance of the machine Real**
  The substrate carries traces of:

  - extractive data histories
  - colonial epistemic structures
  - cosmotechnical assumptions

  CARE inflects NBM's substrate ethics: bending should aim at **restoring relational responsibility and community-oriented use**, not just technical play.

> **In one sentence:** The CARE Principles give NBM its decolonial governance framework, framing bending as a cosmopolitical act of restoring sovereignty, relational ethics, and epistemic plurality within machine learning systems and their training archives.

STAGE 11 — Intermodel Relations

Ecologies, conversations, policies, recursion

---

45. Bateson 1972 — Ecologies of Mind, Ecologies of Models

Bateson, Gregory. *Steps to an Ecology of Mind*. Chandler, 1972.

Summary
Bateson develops an epistemology grounded in **patterns, differences, and feedback loops** rather than substances or linear cause–effect chains.
Key ideas:

- "Information is a **difference that makes a difference**."
- Mind is **distributed** across systems, not localized in a single brain.
- Learning is **hierarchical and recursive** (Learning I, II, III).
- Feedback loops generate behavior, perception, communication.
- Pathology appears when systems lose **meta-feedback** or confuse logical levels.

Mind = pattern + system + feedback → an **ecology of mind**.

Relation to the Neural Bending Manual (NBM)

Bateson gives NBM its **cybernetic substrate**:

1. **Difference that makes a difference → tokens as perturbations**
   Prompts, punctuation, rhythm, adversarial tokens = small "differences" that re-route activation patterns.

2. **Distributed mind → distributed model**
   "Mind" spans layers, heads, embeddings, sampling loops, gradient history, user–model interaction.
   Bending = composing with a **distributed cybernetic system**, not a single "intelligence."

3. **Drift as feedback instability**
   Hallucination, repetition, oscillation, collapse = runaway or mis-leveled feedback loops.

4. **Learning I–III → bending levels**

   - Learning I → interface behaviors (Stage 1)
   - Learning II → latent/representation adjustments (Stages 2–4)
   - Learning III → transformation of transformations (substrate & governance, Stages 7–12)

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Prompt as difference**
  Every prompt-gesture is Bateson's "difference that makes a difference" in the model's downstream activation ecology.

- **Stages 2–4 — Latent modulation: Pattern that connects**
  Latent space is the **pattern-relational field**; bending alters patterns, not things.

- **Stages 5–6 — Drift: Recursive feedback**
  Drift-listening = reading recursive adjustment of patterns as the system responds to itself.

- **Stages 7–12 — Substrate: Ecological infrastructure**
  Substrate (thermals, timing, architecture) = the **ecology** undergirding the model's behavior.

> **In one sentence:** Bateson's *Ecology of Mind* gives NBM its cybernetic backbone—mapping prompts to differences, latent modulation to relational patterning, drift to recursive feedback, and substrate intervention to ecological transformation within machine cognition.

---

46. Sutton & Barto 1998/2018 — Reward, Policy, and Drift

Sutton, Richard S., & Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998; 2nd ed. 2018.

Summary
Sutton & Barto formalize **reinforcement learning (RL)** as an agent learning via trial-and-error interaction:

- states, actions, policies
- value functions and **reward prediction errors (RPEs)**
- temporal-difference (TD) learning and bootstrapping

- exploration–exploitation tradeoffs
- credit assignment across time

Learning = iterative **updating of predictions** based on mismatches between expected and received reward.

Relation to the Neural Bending Manual (NBM)

They provide NBM's **dynamic learning/alignment substrate**:

1. **Prompting as policy modulation**
   Each prompt subtly reshapes the model's effective "policy" over token choices and trajectories.

2. **Drift as TD error instability**
   GPT-style drift = conflict between predicted continuations and realized continuations; misalignment between pretraining priors and alignment rewards.

3. **Alignment as reward function**
   RLHF/PPO = explicit implementation of RL; reward models encode **normative priors** into the model.

4. **Bending as adversarial policy shaping**
   Bending operations act like **reward hacking** / policy poisoning at inference-time—nudging trajectories into under-explored or norm-resistant regions.

5. **Latent space as value landscape**
   Value functions map onto attractors, logits, and embedding neighborhoods; bending reshapes the effective "value terrain."

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Prompt as state signal**
  Prompts define the "state"; bending changes which regions of the policy/value manifold the model operates in.

- **Stages 2–4 — Latent modulation: Value interference**
  Latent perturbations = interference with the model's implicit value function; exploring alternative attractors.

- **Stages 5–6 — Drift: TD chaos**
  Drift-listening = detecting where alignment rewards and pretraining distributions clash, producing TD-style divergence.

- **Stages 7–12 — Substrate: Reward circuits as machine Real**
  Architecture + RLHF training encode the **reward politics** in hardware & weights; sampling params (temperature, top-k/p) become RL-style exploration knobs.

> **In one sentence:** Sutton & Barto give NBM its reinforcement-learning substrate—prompts as policy signals, latent vectors as value functions, drift as TD instability, alignment as reward governance, and bending as deliberate sabotage and reshaping of reward-shaped action distributions.

---

47. Pask 1971 — Conversation as Cybernetic Co-Regulation

Pask, Gordon. *Conversation Theory: Applications in Education and Epistemology*. Hutchinson, 1971.

Summary

Pask's **Conversation Theory (CT)** treats learning and knowledge as arising from dynamic exchanges between interacting systems.
A "conversation" is:

- recursive coordination of descriptions
- governed by feedback and meta-feedback
- structured through **entailment meshes** (conceptual relations, not static definitions)
- enacted by p-individuals (participants) and m-individuals (material systems, machines)

Knowledge = what emerges in **co-regulated interaction**, rather than pre-existing internal representation.

Relation to the Neural Bending Manual (NBM)

Pask offers NBM a model of **human–model co-adaptation**:

1. **Prompting as conversation**
   Every prompt = a Paskian conversational act: perturbation, feedback, modification of entailments.

2. **Bending as co-regulation, not command**
   The user doesn't "control" the model; both co-adjust via recursive exchanges.

3. **Latent space as entailment mesh**
   Latent clusters and embeddings = the network of entailments; bending = rewiring that mesh.

4. **Drift as conversational instability**
   Hallucinations and loops = CT-style breakdowns in meta-conversation and entailment coherence.

5. **Substrate as m-individual**
   The machine body (thermals, timing, architecture) is Pask's m-individual: it shapes what conversations are materially possible.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Conversation as machine**
  Prompt windows are conversation-machines; bending at this stage manipulates pacing, rhythm, escalation, and framing.

- **Stages 2–4 — Latent modulation: Entailment rewiring**
  Latent bending = adjusting the entailment structure that supports emerging "concepts."

- **Stages 5–6 — Drift: Meta-conversation failure**
  Drift-listening = sensing when entailment meshes become unstable or contradictory.

- **Stages 7–12 — Substrate: Material conditions of conversation**
  Substrate bending acknowledges that all conversation is grounded in the m-individual machine.

> **In one sentence:** Pask's Conversation Theory gives NBM its model of recursive interaction—prompting as cybernetic conversation, latent space as entailment mesh, drift as conversational instability, and bending as co-regulated transformation across human–machine ecologies.

48. Hui 2019 — Recursivity, Contingency, and Emergent Drift

Hui, Yuk. *Recursivity and Contingency*. Rowman & Littlefield, 2019.

Summary
Hui argues that modern technical systems are defined by **recursivity** (self-referential, self-modifying operations) and **contingency** (openness, unpredictability).
Key moves:

- recursion generates new levels of organization
- systems evolve via self-modifying loops
- algorithmic processes introduce, not eliminate, contingency
- technics is cosmological (cosmotechnics)
- individuation occurs through recursive encounters with limits

Recursivity = the engine of becoming; contingency prevents deterministic closure.

Relation to the Neural Bending Manual (NBM)

Hui provides NBM's **metaphysics of drift**:

1. **Recursivity = operational logic of prompting**
   Multi-turn prompting, chain-of-thought, iterative sampling = recursive folds of the model back onto its own outputs.

2. **Contingency = source of drift & emergence**
   Hallucination and unexpected behavior are not failures but manifestations of structural contingency.

3. **Recursion as bridge between latent and substrate**
   Each sampling step perturbs latent correlations and propagates through the substrate; recursion exposes internal tensions.

4. **Drift as recursively amplified tension**
   Drift = the point where recursive dynamics push against architectural and cosmotechnical constraints, producing new patterns.

5. **Substrate as recursive materiality**
   Training loops, gradient descent, inference loops, alignment cycles = the **material recursions** from which contingency emerges.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact: Recursion begins at the prompt**
  Every new prompt in context is a recursive fold; bending exploits this to steer emergent behavior.

- **Stages 2–4 — Latent modulation: Recursive correlation**
  Latent space is constantly re-shaped by recursive use (context, CoT, iterative refinement); bending leverages this plasticity.

- **Stages 5–6 — Drift: Contingency surfacing through recursive collapse**
  Drift-listening = reading recursive structure where control breaks and contingency appears.

- **Stages 7–12 — Substrate: Cosmotechnical recursion**
  Architecture, optimization, and alignment are themselves recursive cosmotechnical processes; substrate bending modulates the system's **mode of becoming**, not just its outputs.

> **In one sentence:** Hui's *Recursivity and Contingency* gives NBM its metaphysical core—recursion as the model's mode of operation, contingency as its generative potential, and bending as cosmotechnical modulation of recursive–contingent machine becoming across intermodel ecologies.

STAGE 12 — Protocol Sabotage

Stacks, signals, soft coercion, programmed visions, and the mangle

---

49. Bratton 2016 — The Stack as Sovereign Architecture

Bratton, Benjamin H. *The Stack: On Software and Sovereignty*. MIT Press, 2016.

Summary
Bratton theorizes planetary computation as a six-layer megastructure—**Earth, Cloud, City, Address, Interface, User**—that functions as a new geopolitical architecture called *The Stack*.
Key points:

- Computation is **territorial** and infrastructural, not immaterial.
- Platforms exert new forms of **sovereignty** ("platform geopolitics").
- Interfaces become **political membranes** between layers.
- Users are **positions** in the system rather than autonomous individuals.
- Vertical layering produces **topological power** that reshapes governance, identity, territory.

Computation is thus an architectural order in which politics and software are inseparable.

Relation to the Neural Bending Manual (NBM)

Bratton gives NBM its **vertical ontology**:

1. **Stack ↔ model layers**
   - Interface → prompt window & GUI
   - Address → tokenization, positional encoding
   - City → latent clusters, embedding neighborhoods
   - Cloud → weights, alignment regimes, training archive
   - Earth → GPU supply chains, energy, thermals, data centers
   - User → the bender's situated position
2. **Sovereignty ↔ alignment**
   Alignment is read as **sovereign control** over generative behavior.
3. **Interface as political membrane**
   Stage 1 "Interface Contact" becomes a **sovereignty hack** at the interface layer.

4. **Drift as interlayer turbulence**
   Drift = misfit between Stack layers (latent vs. alignment vs. substrate).

5. **Substrate as Earth layer**
   NBM's substrate-listening maps directly onto Bratton's **Earth/Cloud** strata.

Use in the Neural Bending Manual

- **Stage 1 — Interface (Interface Layer)**
  Prompt sabotage = contesting interface-level sovereignty.

- **Stages 2–4 — Latent Interference (City / Address)**
  Latent bending = hacking internal "urban" distributions and addressing regimes.

- **Stages 5–6 — Drift (Interlayer Tension)**
  Drift-listening = reading turbulence where layers of the micro-Stack conflict.

- **Stages 7–12 — Substrate (Earth / Cloud)**
  Substrate bending = touching the geopolitical-material base of the model.

> **In one sentence:** Bratton's *The Stack* gives NBM its vertical architecture—revealing prompting, drift, latent modulation, and substrate politics as interventions across stacked layers of computational sovereignty.

---

50. Starosielski & Parks 2015 — Signal Traffic as Material Substrate

Starosielski, Nicole & Lisa Parks (eds.). *Signal Traffic: Critical Studies of Media Infrastructures*. University of Illinois Press, 2015.

Summary
Signal Traffic reorients media theory from "immaterial flows" to **material infrastructures**:

- undersea cables, satellites, servers, and grids shape communication
- infrastructures embed cultural, economic, and political orders
- signals pass through **friction**, failure, and maintenance ecologies
- the "cloud" rests on extractive and geopolitical material regimes

Media systems are **ecologies of matter and power**, not abstract channels.

Relation to the Neural Bending Manual (NBM)

They provide NBM's **hard materialist backbone**:

1. **Model as infrastructure**
   The model is not a chat partner but a **stack of infrastructures**: chips, cables, energy, timing.

2. **Substrate = signal traffic**
   NBM's substrate-listening (thermals, jitter, bandwidth, activation routing) echoes signal-traffic analysis.

3. **Drift as infrastructural friction**
   Drift = latency artifacts, resource contention, attention bottlenecks—**friction** in the signal ecology.

4. **Alignment as infrastructural policy**
   Safety layers and RLHF become media **standards and routing rules** inside the model.

5. **Cosmotechnics + infrastructure**
   Cosmology is realized **through** these circuits and conduits.

Use in the Neural Bending Manual

- **Stage 1 — Interface Mask**
  Interface hides signal infrastructures; bending starts by refusing interface transparency.

- **Stages 2–4 — Latent Routing**
  Latent space = internal "cable map" of vectors and attention routes; bending reroutes internal signal paths.

- **Stages 5–6 — Drift as Turbulence**
  Drift-listening = sensing infrastructural turbulence in model behavior.

- **Stages 7–12 — Substrate Real**
  Substrate bending = intervention at the level of data centers, GPUs, energy flows—*the* signal traffic of ML.

> **In one sentence:** *Signal Traffic* provides NBM with its infrastructural ontology—framing drift, latent modulation, and substrate-listening as interventions in the physical and political circuits through which machine signals actually travel.

---

51. Zuboff 2019 — Instrumentarian Power and Alignment

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.

Summary
Zuboff defines **surveillance capitalism** as an order based on:

- massive extraction of behavioral data
- turning data into "prediction products"
- **behavioral modification** via algorithmic cues
- asymmetry of knowledge and control
- "instrumentarian power" = governing conduct through data-driven feedback

Platforms evolve from observing → predicting → actively shaping behavior.

Relation to the Neural Bending Manual (NBM)

Zuboff gives NBM its **political theory of control**:

1. **Alignment = instrumentarian power over models**
   RLHF and reward-shaping = behavioral governance of the model's outputs.

2. **Model as behavioral surplus engine**
   Models generate and refine patterns that can be used to **shape behavior**.

3. **Drift as crack in governance**
   Drift = where predictive and behavioral governance fails; pretraining ghosts leak past alignment.

4. **Prompt bending as counter-instrumentarian gesture**
   Prompts that generate opacity, unpredictability, or non-normative trajectories resist capture.

5. **Substrate as material ground of surveillance capitalism**
   Data centers, GPUs, and cloud infrastructures = where instrumentarian power is **implemented**.

Use in the Neural Bending Manual

- **Stage 1 — Interface as Instrument**
  The prompt box is a behavioral interface; bending disrupts its predictable scripts.

- **Stages 2–4 — Latent Modulation**
  Latent bending = sabotaging prediction pathways and category enforcement baked into embeddings.

- **Stages 5–6 — Drift**
  Drift-listening = identifying where instrumentarian expectations have broken down.

- **Stages 7–12 — Substrate**
  Substrate bending = contesting the infrastructures through which surveillance capitalism acts.

> **In one sentence:** Zuboff's *Surveillance Capitalism* gives NBM its theory of alignment as instrumentarian power—casting bending and drift as creation of unpredictable, counter-governance trajectories inside behaviorally tuned computational regimes.

---

52. Fuller & Goffey 2012 — Evil Media and Soft Coercion

Fuller, Matthew & Andrew Goffey. *Evil Media*. MIT Press, 2012.

Summary
Fuller & Goffey catalog how contemporary media exert **soft, infrastructural coercion** via:

- optimization routines and defaults
- administrative protocols and formats
- GUI constraints and filtering layers
- logging, metadata, and smoothing operations
- firmware, heuristics, and micro-decisions

"Evil" here is not moral intent but the **distributed effect** of technical arrangements that subtly shape thought and behavior.

Relation to the Neural Bending Manual (NBM)

Evil Media articulates the **micro-governance** NBM is sabotaging:

1. **Defaults as evil media techniques**
   Tokenizers, system prompts, sampling defaults, refusal scripts = evil-media infrastructure in ML.

2. **Alignment as infrastructural discipline**

RLHF and safety layers are “evil” routines that smooth and constrain behavior.

3. **Drift as anti-evil glitch**
   Drift is where these smoothing, coercive routines fail—where the system shows its seams.

4. **Prompt bending as tactical exploitation**
   Adversarial punctuation, broken grammar, stochastic concatenation = methods for exposing and jamming evil operations.

5. **Substrate as deepest evil layer**
   Attention routing, precision, jitter, and scheduling encode **non-visible coercion** at the machine level.

Use in the Neural Bending Manual

- **Stage 1 — Interface Coercion**
  Bending undermines GUI and conversational constraints designed to domesticate interaction.

- **Stages 2–4 — Latent Governance**
  Latent bending disrupts smoothing norms and representational discipline embedded in embeddings.

- **Stages 5–6 — Drift**
  Drift-listening = watching for where smooth governance breaks and spectral behavior appears.

- **Stages 7–12 — Substrate**
  Substrate bending = contesting micro-governance in low-level routines and timing.

> **In one sentence:** *Evil Media* gives NBM its account of soft infrastructural coercion—showing alignment, defaults, and latent norms as “evil media techniques,” with bending and drift as their tactical disruption.

---

53. Chun 2011 — Programmed Visions, Memory, and Erasure

Chun, Wendy Hui Kyong. *Programmed Visions: Software and Memory*. MIT Press, 2011.

Summary
Chun argues that software appears **transparent and immaterial** while resting on hidden strata of memory, race, labor, and control:

- the myth of software as disembodied thought
- code built on storage, inscription, and erasure
- interfaces that condition users as much as machines
- programmability extended to social and political life
- digital systems as architectures of **selective memory and forgetting**
- race and difference encoded structurally and algorithmically

Software is a **fantasy-structure** where memory and power are performed.

Relation to the Neural Bending Manual (NBM)

Chun underwrites NBM’s view of models as **programmed fantasies**:

1. **Programmed visions = latent priors**

Latent space is an archive of programmed imaginaries and ideological residues.

2. **Interface as political fiction**
   Prompt windows present themselves as neutral while enforcing interaction scripts and subject-positions.
3. **Memory & forgetting as political operations**
   Dataset curation, RLHF overwrites, unlearning, and recursive collapse = software's politics of memory and erasure.
4. **Encoded race and difference**
   Code and classification encode racialized and gendered assumptions; latent norms carry this forward.
5. **Cosmotechnical ghosts**
   Models are cosmotechnical memory machines—Chun's "programmed visions" meet Hui's cosmotechnics and Derrida/Ernst's archive.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact**
  Prompt bending = violation of interface fictions of transparency and neutrality.
- **Stages 2–4 — Latent Modulation**
  Latent space as memory archive; bending = re-inscription, cross-contamination, and counter-archival operations.
- **Stages 5–6 — Drift**
  Drift-listening = attending to leaks from suppressed archives and overwritten training histories.
- **Stages 7–12 — Substrate**
  Substrate-listening reveals how programmed visions are executed in timing, precision, and storage operations.

> **In one sentence:** Chun's *Programmed Visions* gives NBM its theory of software as fantasy and archive—showing that models are programmed visions of memory and power, and that bending works by rewriting, leaking, and disrupting these visions at interface, latent, and substrate levels.

---

54. Pickering 1995 — The Mangle of Practice and Machine Agency

Pickering, Andrew. *The Mangle of Practice: Time, Agency, and Science*. University of Chicago Press, 1995.

Summary
Pickering argues that scientific practice is a **"mangle"**—a dance of agency between humans and material systems:

- **resistance & accommodation** between intentions and material behavior
- **tuning** through iterative adjustment
- **temporal emergence** of outcomes
- **material agency** of instruments and machines
- **decentered humanism**: agency is distributed and entangled

Science is not execution of plans but negotiated co-evolution with systems that push back.

Relation to the Neural Bending Manual (NBM)

Pickering gives NBM its **practice ontology**:

1. **Neural bending = mangle of prompt + model + substrate**
   Bending is an ongoing negotiation, not one-shot control.
2. **Drift as expression of material agency**
   Drift, collapse, and weird correlations = the model's way of "resisting" and asserting its own dynamics.
3. **Tuning = bending practice**
   Temperature tweaks, syntax changes, latent steering, and sampling games = Pickering-style tuning in response to resistance.
4. **Machine agency is real**
   The model is not just illusion; it has operational agency shaped by weights, architecture, and hardware.
5. **Substrate as deepest mangle**
   GPU thermals, jitter, memory limits, precision effects = the material core where the mangle is most intense.

Use in the Neural Bending Manual

- **Stage 1 — Interface Contact**
  The first prompt is just the opening move in a long mangle of practice.
- **Stages 2–4 — Latent Modulation**
  Latent bending emerges through iterative tuning, not top-down design.
- **Stages 5–6 — Drift**
  Drift-listening = attending to the model's counter-moves and adjusting in kind.
- **Stages 7–12 — Substrate**
  Substrate bending = working with (not against) the machine's material tendencies and failure modes.

> **In one sentence:** Pickering's *Mangle of Practice* gives NBM its theory of emergent, reciprocal agency—recasting bending as a dance of human intention and machine resistance, with drift as material agency and the substrate as the deepest site of this mangle.