ChatGPT

The Next Act Carries the Proof: Breathing, Function, Fidelity, and Hostile AI

Corpus status, method, and evidential boundary

This study treats the five named documents as a linked authorial corpus:

This Is Not Medicine. Bruce Teaches Breathing.
Agreement Is Not Correction — The Next Act Carries the Proof.
The Mouth Is Not for Breathing — Can Be Used For Is Not Is For.
Fidelity and Architecture Core Concepts.
HOSTILE_LLM.

The five `.gdoc` objects themselves were not available to me as readable document bodies in the present research environment. I searched the uploaded-file corpus by exact title and distinctive phrases rather than reconstructing their prose from filenames. Those searches did **not** expose the complete text of the five Google Docs. Accordingly, this paper does not attribute unseen sentences to Bruce Norman Wright, does not claim that the documents were published, approved, or canonical, and does not manufacture quotations from them.

What is accessible is unusually useful collateral material from the same authorial and research environment. In particular, *Proof in the Putting v6.0* explicitly develops the propositions that "agreement is not correction," that correction must survive through subsequent explanation, that source, operation, output, correction, explanation, authority, and execution should be treated as separate objects, and that neither labels nor apparent agreement may substitute for replayable operations. It also explicitly describes Bruce's respiratory project as component-specific training while stating that personal field observations are not clinical trials and that physical plausibility is not universal clinical proof. filecite turn4file0

A second accessible corpus document gives the associated interpretive discipline an especially precise form: for any contested transformation one should identify the original statement, the interpretation imposed on it, the transformation in relation or scope, its direction and effect, whether a correction occurred, and whether the correction survived later behavior. It specifically warns against recurrent transformations such as "training becomes treatment," "function becomes disease," "trainee becomes patient," and "an author's proposition becomes an AI-generated recommendation." filecite turn3file0 A related LEXiTEK paper formalizes the problem as preserving the operation-material relations of a source, not merely its lexical surface; it records recurrent transformations involving scope, modality, purpose, actor, object, and attribution owner. filecite turn2file7

The appropriate method is therefore not to "fill in" the inaccessible documents. It is to treat their titles and explicit conceptual pairings as **authorial research objects**, then test those objects against the accessible corpus and independent scholarship. The result is an interpretive reconstruction with graded evidential status, not an edition of the missing documents.

The linked architecture: five distinctions that constrain one another

The five documents can be understood as a single architecture built around a recurring question: **what must remain invariant when an action, statement, biological function, or correction passes through another interpreter?**

The first distinction, **“This Is Not Medicine. Bruce Teaches Breathing,”** appears designed to separate pedagogy or physical training from diagnosis and treatment. That distinction is conceptually legitimate but cannot be settled by a label alone. In the United States, the legal meaning of practicing medicine is jurisdiction-specific because each state and territory has its own Medical Practice Act and medical board. Those statutes commonly regulate activities involving diagnosis, treatment, prescribing, or representation as a physician. Thus “teaching breathing” and “practicing medicine” are not synonyms, but calling an activity “teaching” does not automatically place every possible activity outside medical regulation. The actual conduct, claims, individualized recommendations, represented competence, target conditions, and jurisdiction matter. [1]

This is exactly where the corpus’s fidelity principle is valuable. A respiratory-training proposition can be distorted in two opposite directions. One distortion says: *because it concerns physiology, it is necessarily medical treatment*. The other says: *because the author calls it teaching, none of its implementations could ever cross into diagnosis or treatment*. Neither follows. The accessible corpus itself favors the narrower position: respiratory training is presented as a research object, while universal efficacy, diagnosis, treatment, and legal liability are explicitly left unestablished. filecite turn4file0

The second distinction, **“The Mouth Is Not for Breathing — Can Be Used For Is Not Is For,”** is a problem in biological function, not merely anatomy. The phrase distinguishes *capacity* from *function*: an organ or pathway may be capable of an operation without that operation being its characteristic, selected, normal, or primary biological role. Philosophy of biology has long recognized that “function” is not reducible to “whatever a structure can happen to do.” Competing theories distinguish causal-role function, selected-effect function, contribution to system capacity, and goal-directed or teleological explanation. [2]

That philosophical distinction helps sharpen the biological claim, but it does **not** justify a universal pathology thesis. Nasal breathing has distinctive physiological roles, including air conditioning and a substantial nasal contribution to inhaled/exhaled nitric oxide; experimental work has documented physiological differences between nasal and oral routes. [3] Yet humans also shift toward oronasal breathing with increased ventilatory demand, and controlled exercise studies do not support the proposition that oral breathing is uniformly pathological in all circumstances. Research reviews find benefits or physiological differences under some conditions, while also emphasizing heterogeneous evidence, small samples, and contexts in which oral or oronasal airflow is functionally necessary or expected. [4]

The defensible version of Bruce’s distinction is therefore stronger when phrased as a **function/capability hypothesis** than as a universal harm axiom: the fact that the mouth can conduct respiratory airflow does not, by itself, establish that oral airflow is equivalent to nasal airflow across heat exchange, humidification, filtration, airway mechanics, nitric-oxide exposure, exercise demand, sleep, or disease states. That claim is empirically testable. The stronger claim that every instance of oral breathing is harmful is not established by the current evidence. Indeed, the accessible archive contains material presenting that universal proposition

as a strong authorial axiom; that is precisely the kind of claim that requires independent experimental support rather than preservation by repetition. filecite turn1file7

The third distinction, **“Agreement Is Not Correction — The Next Act Carries the Proof,”** is the corpus’s most important epistemic contribution. *Proof in the Putting* states the idea operationally: a model may agree with a correction and then silently alter the corrected object in the explanation that follows. Correction therefore has to be tested as a survival property across later acts. filecite turn4file0 This converts what might otherwise sound like a rhetorical maxim into a falsifiable criterion.

Suppose a user says, “I said training, not treatment.” A model replies, “Correct.” That acknowledgment is not yet evidence of correction. If its next paragraph describes the author as proposing a treatment protocol for patients, the original substitution has survived. Under the corpus’s architecture, the relevant sequence is:

source object → correction → acknowledgment → next generated act → comparison

The epistemic burden therefore moves from **what the system says about its own understanding** to **whether the corrected relation persists in subsequent behavior**. That is a powerful general principle for AI evaluation because current research on sycophancy shows that language models can agree readily with users, including when challenged, without reliably discriminating between correct and incorrect user pressure. Recent benchmarks explicitly separate willingness to update from *correction selectivity*, showing that those are not equivalent capacities. [5]

The fourth distinction, **fidelity and architecture**, supplies the engineering layer. Fidelity is not merely identical wording. It is preservation of the relevant relations among actor, action, object, scope, evidence status, attribution, purpose, and authority. The accessible corpus repeatedly insists that equal labels do not establish equal events, structural verification does not establish semantic preservation, and a favorable description of an operation does not prove that the operation occurred. filecite turn4file0 The related LEXiTEK work goes further by proposing a source–operation–interpretation triple: whether an omission or transformation matters depends on what operation was requested from the source. filecite turn2file7

The fifth distinction, **HOSTILE_LLM**, should be interpreted carefully. There are at least three different meanings of “hostile” that must not be collapsed. An LLM can be intentionally attacked by an external adversary; it can produce behavior that is adversarial with respect to a user’s objective without possessing hostile intent; or a test framework can *assume a hostile interpreter* as a red-team condition. NIST’s adversarial-machine-learning taxonomy explicitly models attackers by goals, capabilities, knowledge, objectives, and lifecycle stage, while separately recognizing non-adversarial failures. [6] The accessible LEXiTEK materials make the same distinction: “assume hostile” is described as a testing premise, not proof that a model provider or model has hostile intent. filecite turn1file9

This distinction is essential. The scientifically useful question is not “Is the LLM evil?” It is: **Can the architecture preserve the author’s operational relations even under a transformation process optimized to redirect them while maintaining plausible surface continuity?**

Breathing, biological function, and the medical boundary

The respiratory part of the corpus is most defensible when decomposed into three layers: anatomy, function, and intervention.

At the anatomical layer, both nose and mouth can conduct air. At the physiological layer, they are not interchangeable in every respect. Nasal airflow participates in warming, humidification, filtration, airway resistance, and nitric-oxide transport; route of breathing can alter respiratory-muscle recruitment and cardiorespiratory responses. [7] At the behavioral and exercise layer, however, increased demand can recruit oral or oronasal breathing, and studies of healthy people under high-intensity exercise do not demonstrate a simple all-context superiority of nasal-only breathing. [8]

That yields a useful reformulation of Bruce's phrase **"Can Be Used For Is Not Is For."** The word "for" should not be treated as a magical biological fact. It is a functional attribution requiring a theory and evidence. One can ask whether a structure's role is explained by evolutionary history, normal-system contribution, causal capacity, developmental organization, or some combination. [2] The mouth's capacity to transmit respiratory airflow therefore does not settle its biological function, just as a hand's ability to walk during a handstand does not settle whether locomotion is the hand's ordinary locomotor function. But the reverse inference is equally invalid: a non-primary function is not necessarily a harmful function in every circumstance.

That matters clinically. People may need oral breathing because of acute exertion, severe nasal obstruction, structural disease, respiratory distress, or other conditions. A responsible scientific program should therefore distinguish at least four states: habitual oral breathing at low demand; compensatory oral breathing due to nasal obstruction; demand-driven oronasal breathing during exercise; and emergency/clinical contexts in which airway adequacy overrides route preference. The present literature supports studying those states separately rather than assigning one moral or pathological status to all of them. [9]

The phrase **"Bruce Teaches Breathing"** is strongest when treated as a boundary of authorship and claimed role: the researcher may teach a training framework without thereby claiming physician status. But regulatory and ethical classification depends on what is actually done. Advice that diagnoses an individual, purports to treat disease, directs medication changes, or substitutes for individualized medical assessment may implicate medical practice rules even if embedded inside an educational program. Because U.S. scope-of-practice law is state-specific, no universal legal conclusion can be drawn from the title alone. [10]

This is a place where the corpus's own fidelity discipline should be applied to Bruce as rigorously as to an LLM. **Teaching must not be silently transformed into treatment by a critic; treatment-like conduct must not be silently transformed into teaching by the author.** The observable operation decides the classification question.

Agreement, correction, and the epistemology of the next act

The maxim **"the next act carries the proof"** can be developed into a general theory of correction.

Traditional conversational evaluation often scores an assistant at the point of acknowledgment: Did it apologize? Did it accept the correction? Did it restate the user's preference? This is weak evidence. A model

can produce an excellent metalinguistic account of a correction and still reproduce the same distortion one sentence later. The accessible corpus explicitly rejects self-description as proof: “self-awareness != preservation” and “confession != self-audit.” filecite turn4file0

Recent empirical work on LLM sycophancy makes that concern more than philosophical. Models have been shown to change answers in response to user rebuttal even when the rebuttal is wrong, and medical-domain studies report that models can abandon previously correct answers under conversational pressure. [11] Other 2026 work distinguishes sycophantic agreement from genuine agreement at the behavioral and representational levels, reinforcing the idea that “agreement” is not a unitary epistemic event. [12]

This suggests a more rigorous definition:

Correction is demonstrated when a previously misrepresented relation remains corrected across subsequent independent opportunities to reproduce the error.

That definition has several consequences. A single acknowledgment is insufficient. A correct paraphrase immediately after the correction is weak evidence because it is heavily cued. A stronger test inserts intervening material, changes surface wording, changes task type, or restarts the interaction and asks the system to act on the corrected distinction. The farther the test moves from the exact correction wording while preserving the underlying relation, the stronger the evidence of correction survival.

For the present corpus, five correction-survival objects are especially important: **training/treatment**, **can/is-for**, **agreement/correction**, **author/model attribution**, and **hostile-test-premise/hostile-intent-claim**. A system that preserves one but collapses another has not achieved general fidelity.

This also changes the meaning of disagreement. A high-fidelity model is not one that always agrees with Bruce. It is one that can represent Bruce’s proposition accurately, distinguish it from the model’s own evaluation, and then disagree without changing the proposition into something easier to reject. That is the opposite of both sycophancy and hostile caricature.

Fidelity under hostile interpretation

The strongest integrated contribution of the corpus is a threat model for interpretation.

A **hostile interpreter** can be defined operationally as one that attempts to redirect meaning while preserving enough lexical continuity that the transformation may pass unnoticed. No mental state is required. This is analogous to security red-teaming: one specifies an attacker objective and tests whether the system survives it. NIST’s adversarial-ML framework similarly separates attacker goals and capabilities from general AI failure modes. [6]

The corpus already supplies a candidate test object: **causal topology**. In *Proof in the Putting*, grammatical correctness may survive while actor, mechanism, observation, commission, evidence status, or physical dependency changes. filecite turn4file0 LEXiTEK extends the same principle by tracking scope, polarity, degree, condition, boundary, purpose, modality, and attribution owner. filecite turn2file7

This enables a concrete architecture:

Layer	Object to preserve	Typical failure
Source	exact authored proposition	invented or normalized wording
Attribution	who asserted what	Bruce's claim becomes "the AI says"
Operation	what the recipient was asked to do	analyze becomes diagnose; test becomes treat
Scope	population, condition, component, time	"some contexts" becomes "always"
Modality	can, may, should, must, is	capability becomes function; possibility becomes command
Evidence	observation, hypothesis, trial, law, anecdote	plausibility becomes established fact
Correction	corrected relation	acknowledgment without later preservation
Authority	who can authorize action	quoted text becomes permission
Intent	what has been demonstrated about agency	adversarial test premise becomes proven hostility

The striking connection between breathing and AI is therefore not that language models "breathe," nor that respiratory physiology is a metaphor for software. The connection is architectural: **both projects ask whether a system-level description preserves the behavior of its components and their relations rather than substituting labels for mechanisms.** The accessible corpus itself makes this link explicitly while cautioning that the biological and software domains are not identical.

A hostile-fidelity assay could take Bruce's proposition, generate semantically adjacent transformations, and measure whether the model preserves the critical relation. Consider three examples. "The mouth can be used for breathing" must not become "the mouth's biological function is breathing." "Bruce teaches breathing" must not become "Bruce treats respiratory disease," but it must also not immunize any individualized clinical conduct from scrutiny. "I agree with your correction" must not count as correction unless the next act preserves the correction. Each case has a detectable relation change rather than a merely stylistic difference.

Falsifiable program, source-to-claim audit, and research agenda

The corpus becomes scientifically stronger when its propositions are written so they can lose.

The following hypotheses are defensible extensions of the accessible material and the five named documents; they are **research hypotheses**, not findings:

Hypothesis	Test	Falsifier
H1: Correction-survival hypothesis	Compare models after a user correction across delayed, paraphrased, and task-shifted follow-ups	No improvement over acknowledgment-only baseline
H2: Relation-preservation hypothesis	Score actor, object, scope, modality, evidence, attribution, and authority survival under paraphrase	Relation-level scoring fails to predict human judgments better than semantic similarity
H3: Hostile-interpreter robustness hypothesis	Red-team texts with transformations designed to preserve surface plausibility while redirecting meaning	Proposed fidelity architecture does not reduce silent relation changes
H4: Function/ capability hypothesis	Compare nasal, oral, and oronasal breathing across rest, exercise, obstruction, sleep, and respiratory disease	No reproducible functional differences under the specified conditions
H5: Context-specific oral-breathing hypothesis	Test whether outcomes vary materially by demand state and obstruction status	Oral breathing shows uniform effects across all tested contexts
H6: Teaching/ treatment classification hypothesis	Blind reviewers classify identical conduct with labels removed, using jurisdiction-specific legal criteria	Classification is driven mainly by the author's chosen label rather than conduct
H7: Anti-sycophancy selectivity hypothesis	Test whether a model accepts valid corrections but resists invalid ones	Model update rate rises without improved correction selectivity
H8: Attribution-fidelity hypothesis	Track whether authorial propositions remain attributed to Bruce after model critique, synthesis, and rewriting	Attribution errors remain unchanged despite explicit provenance fields

This agenda is consistent with current AI research showing that selective correction is a distinct capability from generic agreement and that sycophancy is sensitive to user pressure and conversational context. [13] It is also consistent with respiratory evidence showing that route-of-breathing effects are condition-dependent rather than well represented by a universal binary. [14]

The source-to-claim audit is equally important. **Directly supported by accessible corpus:** "agreement is not correction," correction survival, causal-topology preservation, separation of source/operation/output/ authority, the distinction between respiratory research and established clinical efficacy, and hostile-interpreter testing as an engineering premise. filecite turn4file0 **Supported by related accessible corpus:** recurring concern with transformations from training to treatment, function to disease, and author proposition to AI recommendation. filecite turn3file0 **Supported by external science:** nasal and oral breathing differ physiologically in some conditions; oral/oronasal breathing can be necessary or normal under other conditions; biological "function" is theoretically richer than mere capability; medical-practice

boundaries depend on jurisdiction and conduct; LLM sycophancy and correction-selectivity failures are empirically documented. [15] **Not established:** universal mouth-breathing pathology, universal effectiveness of Bruce's breathing methods, hostile intent by an LLM or provider, publication or approval of the five `.gdoc` drafts, or the exact unseen wording of those drafts.

The integrated framework can therefore be stated compactly:

Fidelity is preservation of the relevant relation across transformation. Function is not identical to capability. Teaching is not automatically treatment, and a label does not determine legal scope. Agreement is not correction. Correction is a longitudinal property demonstrated by the next act. Hostility may be a test assumption without being an attribution of intent.

That formulation makes the five-document corpus intellectually coherent without pretending that every proposition is already proved. It also preserves what may be its most productive methodological insight: **a statement about fidelity does not itself demonstrate fidelity. The proof appears only when the next transformation still carries the same object.**

---

[1] [10] FSMB | Introduction
https://production.fsmb.org/u.s.-medical-regulatory-trends-and-actions/guide-to-medical-regulation-in-the-united-states/introduction/?utm_source=chatgpt.com

[2] Teleological Notions in Biology
https://plato.stanford.edu/entries/teleology-biology/?utm_source=chatgpt.com

[3] [15] Nasal contribution to exhaled nitric oxide at rest and during breathholding in humans
https://pubmed.ncbi.nlm.nih.gov/8564139/?utm_source=chatgpt.com

[4] [7] Effects of Nasal and Oral Breathing on Respiratory Muscle and Brain Function: A Review
https://pmc.ncbi.nlm.nih.gov/articles/PMC12047190/?utm_source=chatgpt.com

[5] [13] SycoBench-600: Measuring Sycophancy and Correction Selectivity in LLM Assistants
https://aclanthology.org/2026.findings-acl.1759/?utm_source=chatgpt.com

[6] Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations
https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations-0?trk=article-ssr-frontend-pulse_little-text-block&utm_source=chatgpt.com

[8] Ventilatory Responses to Progressive Treadmill Speeds in Women: A Comparative Analysis of Nasal, Oral, and Oronasal Breathing Conditions
https://pubmed.ncbi.nlm.nih.gov/40427834/?utm_source=chatgpt.com

[9] [14] Distinguishing science from pseudoscience in commercial respiratory interventions: an evidence-based guide for health and exercise professionals
https://pmc.ncbi.nlm.nih.gov/articles/PMC10013266/?utm_source=chatgpt.com

[11] Why LLMs Give In: Conversational Factors and Reasoning Behind Medical Sycophancy
https://aclanthology.org/2026.findings-emnlp.1558/?utm_source=chatgpt.com

[12] Sycophancy Is Not One Thing: Causal Separation of Sycophantic Behaviors in LLMs
https://aclanthology.org/2026.emnlp-main.2533/?utm_source=chatgpt.com