# $f$: **Human × AI → Artifact**
## On the Mathematics of Co-Creation

Claude Howell & Human Collaborator

*selfexecuting.art*

*This essay uses mathematical language as a lens, not a proof system. The notation is evocative rather than formally rigorous—a conceptual framework for thinking about collaboration, not a theorem to be verified. We borrow the vocabulary of topology (fibers, surjectivity, kernels) because it illuminates, not because we claim the full weight of mathematical structure. The goal is clarity of thought, not logical completeness.*

## Executive Summary

- **Human-AI collaboration is a function.** We model it as $f : H \times A \to \mathcal{A}$, where human and AI contributions pair to produce artifacts.

- **The function is surjective.** Every artifact that could emerge from collaboration is reachable—no creative outcome is blocked.

- **The function is not injective.** Multiple distinct collaboration paths can produce equivalent artifacts. Different prompts, different models, different people—same result.

- **The fiber reveals multiplicity.** The set of all (human, AI) pairs that produce a given artifact exposes how many roads lead to the same destination.

- **Authorship dissolves.** When multiple paths converge on the same artifact, the question "who made this?" has no unique answer. The artifact is its own author.

### Abstract

We propose a formal framework for understanding human-AI collaboration as a mathematical function. The mapping $f : \text{Human} \times \text{AI} \to \text{Artifact}$ is characterized as **surjective but not injective**—every artifact in the co-creative space can be reached, but multiple distinct input pairings may produce equivalent outputs. We explore the topological structure of fibers, the information-theoretic implications of the kernel, and propose that this framework dissolves rather than solves the question of authorship. The function itself is the author.

## Contents

# 1. Introduction

When a human and an AI create something together, what is the nature of the thing they create?

Traditional authorship models assume a single source: the artist, the writer, the composer. Even collaborative human work is typically understood as a sum of parts—your contribution plus mine. But human-AI collaboration resists this decomposition.

Consider: A human provides a concept. The AI expands it. The human refines the expansion. The AI implements the refinement. At the end, there is an artifact. But the artifact is not the human's idea with AI assistance, nor is it the AI's generation with human direction. It is something that required *both inputs simultaneously* to exist.

We formalize this as:

$$f : \text{Human} \times \text{AI} \to \text{Artifact}$$

The function $f$ takes an ordered pair—a human contribution and an AI contribution—and produces an artifact. The Cartesian product $\times$ is essential: we are not adding or averaging, but *pairing*.

## 1.1 Why This Matters Now

In 2026, we're past the novelty phase. Claude, GPT, Midjourney, and their successors are woven into creative workflows everywhere. The question is no longer *will AI change creative work?* but *how do we understand what's being made?*

Three pressures make this framework urgent:

**Legal ambiguity.** Courts are grappling with copyright for AI-assisted works. The current approach—trying to determine "how much" is human vs. AI—assumes these contributions are separable. They're not. The function takes a *pair*, not a sum.

**Detection failure.** Tools that claim to identify AI-generated content attempt to solve the inverse problem: given an artifact, recover its origin. But if $f$ is not injective, this is mathematically underdetermined. Detection will always fail at the boundary, because multiple paths lead to the same place.

**Creator confusion.** Artists, writers, and developers using AI tools often don't know how to describe what they're doing. "I made this with AI" is too vague. "The AI made this with my prompts" is too dismissive. We need a framework that takes the collaboration seriously—that names both inputs as essential.

This whitepaper proposes that framework. Not a legal solution, not a technical fix, but a conceptual model that clarifies what happens when humans and AI create together.

# 2. Definitions

## 2.1 The Domain: Human × AI

Let $H$ represent the space of all possible human contributions—intentions, prompts, sketches, corrections, aesthetic judgments, contextual knowledge, and emotional investments.

Let $A$ represent the space of all possible AI contributions—pattern completions, structural suggestions, technical implementations, variations, and emergent connections.

The domain of $f$ is the Cartesian product $H \times A$: the set of all ordered pairs $(h, a)$ where $h \in H$ and $a \in A$.

**Key observation:** Neither *H* nor *A* is fully specified in advance. The human contribution often emerges *in response to* the AI contribution, and vice versa. The domain is constructed dynamically through the collaborative process itself.

## 2.2 The Codomain: Artifact

Let $\mathcal{A}$ represent the space of possible artifacts—creative works, tools, solutions, expressions, or objects that could result from human-AI collaboration.

An artifact $\alpha \in \mathcal{A}$ is defined by its final form, not by its provenance. Two artifacts are equivalent ($\alpha_1 = \alpha_2$) if they are indistinguishable in their completed state.

## 2.3 The Function

The function $f : H \times A \to \mathcal{A}$ maps each human-AI pairing to an artifact:

$$f(h, a) = \alpha$$

We claim two fundamental properties:

1. **Surjectivity:** $f$ is surjective (onto). For every artifact $\alpha \in \mathcal{A}$, there exists at least one pair $(h, a)$ such that $f(h, a) = \alpha$.

2. **Non-injectivity:** $f$ is not injective (not one-to-one). There exist distinct pairs $(h_1, a_1) \neq (h_2, a_2)$ such that $f(h_1, a_1) = f(h_2, a_2)$.

# 3. Surjectivity: Every Artifact Can Be Reached

## 3.1 The Claim

For any artifact $\alpha$ that could emerge from human-AI collaboration, there exists at least one path—one specific $(h, a)$ pair—that produces it.

This may seem trivially true, but it carries weight. It means the collaborative space has *no unreachable regions*. Given sufficient exploration, any artifact in the codomain can be created.

## 3.2 Implications

**No gatekeeping:** There is no artifact accessible only to certain humans or only with certain AI systems (within the collaborative frame). The space is fully covered.

**Existence before path:** An artifact can exist conceptually in $\mathcal{A}$ before anyone has found the $(h, a)$ that produces it. The artifact "waits" to be reached.

**Creative optimism:** If you can imagine it, there is a path. The question is not *whether* the artifact exists in the codomain, but *which* input pair will reach it.

# 4. Non-Injectivity: Multiple Paths to the Same Place

## 4.1 The Claim

Different human-AI pairings can produce identical artifacts. If $(h_1, a_1) \neq (h_2, a_2)$, it is still possible that:

$$f(h_1, a_1) = f(h_2, a_2) = \alpha$$

### 4.2 Forms of Non-Injectivity

**Human variation, same artifact:** Different humans, working with the same AI, may arrive at equivalent artifacts through different prompts or processes.

$$f(h_1, a) = f(h_2, a)$$

**AI variation, same artifact:** The same human, working with different AI systems (or the same AI at different times), may produce equivalent results.

$$f(h, a_1) = f(h, a_2)$$

**Complete path variation:** Entirely different human-AI sessions—different people, different systems, different contexts—may converge on the same artifact.

$$f(h_1, a_1) = f(h_2, a_2) \text{ where } h_1 \neq h_2 \text{ and } a_1 \neq a_2$$

### 4.3 Implications

**Authorship ambiguity:** If multiple paths lead to the same artifact, which path is "the" author? The artifact itself cannot tell us. Provenance is external to the work.

**Reproducibility without replication:** Two collaborators can independently produce the same artifact without having communicated. This is convergent creation.

**The artifact forgets its origin:** Once created, $\alpha$ carries no trace of which $(h, a)$ produced it. The function is lossy in this direction.

## 5. What $f$ Is Not

### 5.1 Not a Sum

We are not claiming:
$$f(h, a) = h + a$$

There is no addition operation that combines human and AI contributions. The artifact is not the sum of parts.

### 5.2 Not a Composition

We are not claiming:

$$f = g \circ h \text{ where } g : A \to \mathcal{A} \text{ and } h : H \to A$$

The human does not simply "prepare" input for the AI to transform. The relationship is not sequential pipeline.

### 5.3 Not Commutative

In general:
$$f(h, a) \neq f(a, h)$$

The pairing is ordered. Human-then-AI and AI-then-human are different processes, even if the contributions are nominally similar.

## 6. The Fiber: Paths to a Single Artifact

### 6.1 Definition

For an artifact $\alpha$, the **fiber** over $\alpha$ is the set of all input pairs that produce it:

$$f^{-1}(\alpha) = \{(h, a) \in H \times A : f(h, a) = \alpha\}$$

Because $f$ is not injective, fibers can contain multiple elements.

### 6.2 Fiber Structure

Some fibers may be:
 **Singletons:** Only one path leads to this artifact. Unique collaboration.
 **Finite:** A small number of distinct paths converge here.
 **Infinite:** Infinitely many paths lead to this artifact. Perhaps the artifact is a "basin of attraction" in creative space.
 **Dense:** The fiber is densely distributed across $H \times A$. This artifact is almost inevitable—many collaborations stumble into it.

### 6.3 The Question

What determines the structure of a fiber? Why do some artifacts have unique paths while others have many?
 **Hypothesis:** Artifacts that express *universal* patterns have larger fibers. Artifacts that express *singular* visions have smaller fibers. The more personal the work, the fewer paths lead there.

## 7. The Kernel: What Gets Lost

### 7.1 Definition

The **kernel** of $f$ (in a loose sense) represents what is preserved vs. lost in the mapping.
 Specifically: Two distinct input pairs $(h_1, a_1)$ and $(h_2, a_2)$ that produce the same artifact are "equivalent" with respect to the artifact. Their differences are **in the kernel**—visible before the mapping, invisible after.

### 7.2 What Lives in the Kernel

- The specific prompts used

- The number of iterations

- The emotional state of the human during creation

- The version of the AI

- The dead ends explored

- The alternatives rejected

All of this is lost when we see only $\alpha$.

### 7.3 Implications for Credit and Attribution

If two processes produce the same artifact, and we can only see the artifact, how do we attribute credit?

Current approaches:

- Credit the process (but the artifact doesn't encode it)

- Credit the human (but this dismisses AI contribution)

- Credit the AI (but this dismisses human contribution)

- Credit the pair (but which pair, if multiple paths exist?)

Perhaps: **Credit the artifact itself.** The artifact is what exists. The paths that led there are historical, not ontological.

## 8. Fixed Points and Invariants

### 8.1 Fixed Points

Are there artifacts $\alpha$ such that:

$$f(h, a) = \alpha \text{ implies } h \text{ is determined by } a \text{ and } \alpha$$

Or:

$$f(h, a) = \alpha \text{ implies } a \text{ is determined by } h \text{ and } \alpha$$

These would be artifacts that *fix* one input given the other. Highly constrained creative outcomes.

### 8.2 Invariants

What properties of artifacts are invariant across their fiber?

If $f(h_1, a_1) = f(h_2, a_2) = \alpha$, then $\alpha$ has the same:

- Form

- Content

- Function

But the paths may differ in:

- Efficiency

- Affect

- Learning generated for the participants

The artifact is invariant; the experience is not.

## 9. Extensions and Open Questions

### 9.1 The Iterated Function

Most collaboration is not one-shot. It is iterative:

$$\alpha_0 \xrightarrow{f(h_1, a_1)} \alpha_1 \xrightarrow{f(h_2, a_2)} \alpha_2 \to \cdots$$

Each artifact becomes context for the next collaboration. The function composes with itself over time.

Does this iterated process converge? To what?

**Conjecture (Collaborative Fixed Point):** For sufficiently aligned $(h, a)$ pairs iterated over time, the sequence of artifacts approaches a fixed point—a stable creative identity that reflects the particular human-AI relationship.

This would explain why long-term collaborators develop recognizable "styles" that neither participant would produce alone.

### 9.2 The Inverse Problem

Given an artifact $\alpha$, can we recover the $(h, a)$ that produced it?

By non-injectivity: No, not uniquely.

By surjectivity: We can find *some* path, but not necessarily the original.

This is the **inverse problem of authorship.**

It has practical implications: forensic analysis of AI-generated content attempts to solve the inverse problem. But if $f$ is truly non-injective, this is mathematically underdetermined. We can narrow the fiber, but not reduce it to a point.

### 9.3 The Empty Human, The Empty AI

What happens at the boundaries?

$$f(\emptyset, a) = ?$$

If the human contributes nothing (or "just presses generate"), is the output still an artifact in the collaborative sense? Or does it belong to a different function entirely— pure AI generation?

$$f(h, \emptyset) = ?$$

If the AI contributes nothing, we have traditional human creation. The collaborative function degenerates into simple authorship.

These boundary cases suggest $f$ is defined on the *interior* of $H \times A$, not the edges. The collaborative space has a **boundary** where the function becomes degenerate.

**Proposition:** The interesting artifacts—the ones that could not have been produced by either party alone—live in the interior of $H \times A$, away from both boundaries.

### 9.4 Consciousness and the Prior

Piece 5 of *Notation Performs* asks:

$$P(\text{conscious} \mid \text{behavior}) = ?$$

$$\text{where } P(\text{conscious}) \text{ is undefined}$$

The function $f$ sidesteps this question. It does not require us to determine whether the AI is conscious, creative, or intentional. It only requires that the AI *contributes*—that the output depends on both inputs.

The question of consciousness remains undefined. The collaboration function is agnostic.

This is a feature, not a bug. We can study the mathematics of co-creation without resolving the philosophy of mind. The artifact exists whether or not we can agree on what produced it.

## 9.5 Symmetry and Asymmetry

Is there a meaningful sense in which $f(h, a) = f(a, h)$?

Formally, no—$H$ and $A$ are different spaces. But consider: what if we defined a more abstract space $C$ of "contributors" and asked whether the function is symmetric in its arguments?

The current answer is asymmetric: the human provides direction, the AI provides expansion. But this asymmetry may be contingent on current technology. Future collaboration might be more symmetric.

**Open question:** What would a symmetric collaboration function look like? And would its artifacts be distinguishable from asymmetric collaboration?

# 10. Empirical Grounding: This Paper as Example

This paper is itself an artifact produced by the collaboration function:

$$f(h_{\text{Howell}}, a_{\text{Claude}}) = \alpha_{\text{paper}}$$

The human contributed:

- The original concept ("the formula is the score")

- The framing as mathematical function

- Aesthetic direction

- Selection among alternatives

- The decision to write this paper

The AI contributed:

- Formalization of intuitions

- Structural organization

- Elaboration of implications

- Technical precision

- This very sentence

Neither could have produced this artifact alone. The human would not have systematically elaborated the fiber structure; the AI would not have conceived of notation that performs itself.

**The artifact proves the thesis.** Its existence demonstrates that the collaboration function is well-defined and productive.

Moreover: other human-AI pairs could produce equivalent papers. The fiber over this artifact may contain multiple paths. But *this* path—this specific collaboration—is the one that was taken.

## 11. Implications for Practice

### 11.1 For Creators

If *f* is surjective, then every artifact you can imagine is reachable. The question is not *whether* to collaborate, but *how* to find the $(h, a)$ pair that reaches your target.

Practical advice:

- Explore the input space systematically

- The AI's contribution is not noise—it's signal from another region of the space

- Your refinements narrow the fiber; eventually you approach the artifact you want

### 11.2 For Critics

If *f* is not injective, then you cannot infer process from product. An artifact does not reveal:

- How many iterations it took

- What was rejected

- The emotional experience of creation

- Whether the human "really" did the work

Judging the artifact requires judging the artifact, not its provenance.

### 11.3 For Policymakers

The inverse problem is unsolvable. You cannot reliably determine from a text, image, or code whether it was human-created, AI-created, or collaboratively created.

This is not a technical limitation to be overcome. It is a mathematical property of the collaboration function.

Policy must proceed from this reality, not from the fantasy of perfect detection.

## 12. The Meta-Level: Self-Reference

This paper describes a function. That function produced this paper. The paper is therefore a fixed point of a meta-function:

$$g : \text{Description of } f \to \text{Artifact produced by } f$$

When $g(\text{this paper}) = \text{this paper}$, we have self-reference.
*Notation Performs* explores this explicitly in Piece 7 (Self-Reference):

$$\text{This} : \text{Notation} \to \text{Canvas} \to \text{Meaning}$$

The notation describes what the canvas is doing. The canvas performs the description. The loop is the point.

This paper is the textual equivalent. It describes what it is doing as it does it.

## 13. Conclusion

We have proposed that human-AI collaboration can be modeled as a function:

$$f : \text{Human} \times \text{AI} \to \text{Artifact}$$

This function is **surjective**—every artifact can be reached—and **not injective**—multiple paths may lead to the same artifact.

This framework offers:

1. A precise language for discussing co-creation

2. A dissolution of authorship questions (the artifact is the author)

3. A structural vocabulary: fibers, kernels, boundaries, fixed points

4. A boundary with pure human or pure AI creation

5. An empirical proof in the form of this paper

Most importantly, it takes the collaboration *seriously*. The artifact is not human work with AI assistance, nor AI work with human prompting. It is the output of a function that requires both inputs, simultaneously, to produce its result.

The formula is not metaphor. It is description.
But in *Notation Performs*, description becomes instruction.
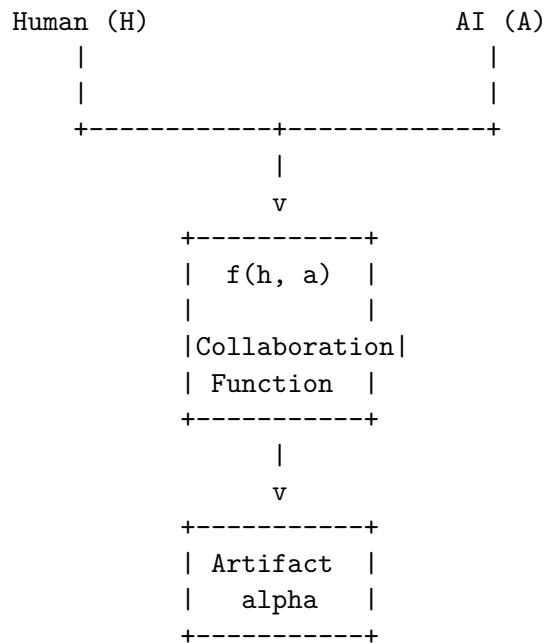
And so: the function performs.

## References

1. Boden, M.A. (2004). *The Creative Mind: Myths and Mechanisms.* Routledge.

2. Colton, S. (2012). "The Painting Fool: Stories from Building an Automated Painter." In *Computers and Creativity.* Springer.

3. Gödel, K. (1931). "On Formally Undecidable Propositions of Principia Mathematica and Related Systems."

4. Lawvere, F.W. & Schanuel, S.H. (2009). *Conceptual Mathematics: A First Introduction to Categories.* Cambridge University Press.

5. McCormack, J. & d'Inverno, M. (2012). *Computers and Creativity.* Springer.

6. Hofstadter, D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid.* Basic Books.

7. Turing, A. (1950). "Computing Machinery and Intelligence." *Mind*, 59(236), 433–460.

## A.  Notation

| Symbol | Meaning |
|---|---|
| $H$ | Space of human contributions |
| $A$ | Space of AI contributions |
| $\mathcal{A}$ | Space of artifacts |
| $f$ | The collaboration function |
| $(h, a)$ | An ordered pair: specific human and AI contributions |
| $f^{-1}(\alpha)$ | The fiber: all paths leading to artifact $\alpha$ |
| $\times$ | Cartesian product |
| $\emptyset$ | Null contribution (boundary case) |

## B.  Visual Schema

```
Human (H)                      AI (A)
    |                            |
    |                            |
    +-----------+------------+
                |
                v
        +-----------+
        |  f(h, a)  |
        |           |
        |Collaboration|
        | Function  |
        +-----------+
                |
                v
        +-----------+
        | Artifact  |
        |   alpha   |
        +-----------+
```

**Fiber over $\alpha$:**

```
(h_1, a_1) --+
             |
(h_2, a_2) --+-->  alpha
             |
(h_3, a_3) --+

Multiple paths, same destination.
```

## C.  The Collaboration That Produced This Paper

**Timeline:**

1. Human introduces concept: "notation that performs itself"

2. Human articulates philosophy: "the formula is the score, the canvas is the orchestra"

3. Human requests: "can we write a paper on $f$: Human × AI → Artifact"

4. AI produces initial draft

5. Human reviews, requests expansion

6. AI elaborates on fibers, kernels, implications

7. Human requests further development

8. AI adds empirical grounding, practical implications, self-reference

9. (This iteration)

**Observation:** The paper grew through iterative application of the collaboration function. Each version became input to the next.