The most promising aspect of the manuscript is not the descriptive discussion of *Ocarina of Time*, *The Wind Waker*, and *Breath of the Wild*, but rather the seven-dimensional framework introduced in Table 1. At present, however, these dimensions function primarily as narrative categories rather than scientific variables. The next stage of development should therefore focus on operationalization. The objective is not to prove the exact internal AI architecture used by Nintendo, since much of that information is proprietary or undocumented, but rather to develop defensible surrogate measures that quantify observable manifestations of those systems. By doing so, the framework could transition from a qualitative review into a reproducible comparative methodology applicable not only to Zelda titles but potentially to other game franchises as well.

For the dimension of NPC decision architecture, the manuscript currently distinguishes between finite-state systems, expanded scripted logic, and more adaptive systemic behavior. While the underlying architecture cannot be directly observed, the behavioral consequences of that architecture can be measured. One possible approach would be to expose NPCs to a standardized set of situations across all games and record the number of distinct context-dependent responses that occur. For example, an NPC might respond differently depending on the time of day, weather conditions, nearby threats, or player actions. A system exhibiting only one or two possible responses would suggest a relatively rigid decision structure, whereas a system producing a wider variety of context-sensitive behaviors would indicate greater decision-making flexibility. In this way, architectural sophistication would be inferred not from speculation about source code but from systematic observation of behavioral diversity.

The dimension of world simulation scope is particularly well suited to quantitative analysis. Rather than describing a world as "more immersive" or "more dynamic," the author could catalog the number of environmental variables that demonstrably influence NPC behavior. These variables might include day-night cycles, weather, temperature, lightning, fire, enemy presence, terrain conditions, and other world-state factors. The resulting score would represent the breadth of environmental systems integrated into NPC decision-making. Such a measure would be relatively objective and would directly address the paper's central argument that immersion increasingly emerges from interactions among interconnected systems rather than from scripted events alone. The resulting comparisons would provide concrete evidence for claims about the expanding role of environmental simulation throughout the franchise.

Movement and navigation can likewise be operationalized through direct observation. Rather than describing navigation as "fixed" or "adaptive," the author could systematically examine how NPCs respond when their intended paths become obstructed. The frequency with which NPCs successfully reroute around obstacles, adjust their paths, or recover from unexpected environmental changes could be quantified and compared across titles. Additional measures might include the number of alternative routes utilized to reach the same destination or the degree to which movement responds to dynamic world conditions. These metrics would allow navigation sophistication to be evaluated in a reproducible manner rather than through subjective gameplay impressions.

The immersion strategy dimension is currently the least well defined because it blends design philosophy with measurable behavior. To strengthen this component, the author could focus on the frequency of emergent interactions generated by the game world. Emergent interactions are events that arise from the interaction of multiple systems rather than from explicit scripting. Examples might include NPCs altering routines due to weather changes, reacting to nearby combat encounters, adjusting paths because of environmental hazards, or responding to wildlife behavior.

By recording the frequency of such unscripted interactions during standardized observation periods, the author could estimate the degree to which each game relies upon systemic rather than scripted mechanisms to create a sense of world liveliness.

The AI complexity emphasis dimension presents a different challenge because true algorithmic complexity cannot be measured without access to implementation details. Nevertheless, observable complexity can still be estimated by examining how many independent systems influence NPC behavior simultaneously. For instance, an NPC whose actions are governed only by a daily schedule and dialogue system would exhibit lower behavioral complexity than an NPC whose decisions are influenced by schedules, weather, threats, social interactions, pathfinding, physics systems, and environmental hazards. The number of interacting influences on behavior provides a reasonable surrogate for overall complexity and can be measured consistently across games without requiring access to proprietary technical documentation.

The player-NPC interaction model can be operationalized by examining the diversity and responsiveness of NPC reactions to player actions. A standardized testing protocol could be established in which players perform a predetermined set of actions, such as approaching NPCs, initiating dialogue, drawing weapons, creating disturbances, or interacting with nearby objects. The number of unique responses generated by NPCs, along with the proportion of player actions that elicit any response at all, would provide quantitative measures of interaction richness. Such metrics would directly address the paper's claims regarding increasing responsiveness and behavioral sophistication across the franchise.

Finally, the believability mechanism dimension may be the most important because it connects directly to the manuscript's broader claims about immersion. Unlike the previous dimensions, believability can be assessed through external observers. Independent raters could view short gameplay clips and evaluate whether NPC behavior appears appropriate given the circumstances. Alternatively, the author could construct a consistency-based metric by examining whether NPCs respond in ways that align with the established rules of the game world. For example, if weather affects players, believable NPCs should exhibit corresponding weather-related responses. If lightning is dangerous, NPCs should take precautions when exposed to it. The extent to which behaviors consistently align with environmental conditions would provide a measurable indication of perceived believability.

Perhaps the most compelling outcome of this process would be the development of an overarching **Environmental Coherence Index** derived from the seven dimensions. The central argument of the manuscript is not simply that AI has become more sophisticated, but that NPCs increasingly operate within a shared system of environmental rules. By combining measures of decision flexibility, world simulation integration, navigation adaptability, interaction richness, behavioral complexity, emergent activity, and believability, the author could construct a quantitative estimate of environmental coherence. Such an index would permit direct comparison among games and would transform the manuscript from a descriptive historical review into a framework-driven observational study. Importantly, this would still not prove that AI evolution causes immersion, since immersion itself would require independent measurement through player studies. However, it would provide much stronger evidence for the more defensible claim that later Zelda titles exhibit greater environmental coherence and behavioral integration than earlier entries. That claim alone would constitute a meaningful and potentially publishable contribution.

Example Tables:

Dimension	OOT	WW	BOTW
Decision Diversity	1.3	2.1	6.8
Simulation Scope	2	3	8
Navigation Adaptability	0.15	0.30	0.85
Believability Rating	2.8	3.6	4.7

Author's Dimension	Possible Surrogate Measure
NPC Decision Architecture	Number of distinct context-dependent behavior transitions observed
World Simulation Scope	Number of environmental variables affecting NPC behavior
Movement & Navigation	Percentage of obstacle encounters successfully rerouted
Immersion Strategy	Hardest one; proxy through density of unscripted interactions
AI Complexity Emphasis	Number of independent behavioral systems interacting simultaneously
Player-NPC Interaction	Number of unique NPC responses to different player actions
Believability Mechanism	Behavioral consistency score

Examples of methodology:

NPC Decision Architecture

The paper claims:

- Ocarina = FSM
- Wind Waker = richer scripted logic
- BOTW = adaptive systems

Rather than arguing architecture directly, measure:

Diversity responses observed stimulusDecision Diversity=same stimulusunique responses observed

Example:

Player approaches villager.

Ocarina:

- 1 response

BOTW:

- 5 responses depending on weather, time, location, threat level

Higher diversity is a reasonable surrogate for architectural sophistication.

World Simulation Scope

This is probably the easiest.

Count environmental variables influencing NPC behavior:

Variable	OOT	WW	BOTW
Time of Day	Yes	Yes	Yes
Weather	No	Minimal	Yes
Temperature	No	No	Yes
Physics	Minimal	Minimal	Yes
Enemy Presence	Limited	Limited	Yes

Then compute:

Scope Score of environmental variablesSimulation Scope Score=Number of environmental variabl es

Not perfect, but objective.

Movement & Navigation

Observe 100 obstacle encounters.

Measure:

Adaptability reroutes encountersNavigation Adaptability=obstacle encounters divided by successful reroutes

This is a classic AI performance metric.

Believability Mechanism

This one is surprisingly measurable.

Present observers with gameplay clips.

Ask:

> "Does this character behave as expected given the situation?"

Scale 1-5.

Average across raters.

This is essentially what the Warpefelt literature the author cites is trying to formalize.

AI Complexity Emphasis

Not actual algorithmic complexity (which Nintendo doesn't publish).

Instead:

Count interacting subsystems.

For example:

Ocarina NPC:

- schedule

- dialogue

Complexity score = 2

BOTW NPC:

- schedule
- weather
- danger
- pathfinding
- physics
- social interactions

Complexity score = 6

Again, a surrogate.

The Hardest One: Immersion Strategy

The author's "Immersion Strategy" dimension is the weakest scientifically because it mixes design philosophy with measurable behavior.

Replace it with:

Rate interactions of gameplay

Emergence Rate=unscripted interactions per hour of gameplay

Example:

Player witnesses:

- villager fleeing storm
- monster attacks village
- NPC changes route
- wildlife scares merchant

Those are emergent events.

More emergent events imply greater reliance on systemic immersion rather than scripted immersion.

The other very good reason for implementing and quantitating this framework is that it is **portable**. Another researcher apply this to:

- Skyrim
- Red Dead Redemption 2
- Witcher 3
- Elden Ring

and obtain meaningful results. This generalizability is valuable in scientific research.