$\mathbf{x \quad A \quad P(x) = P(A)}$: The Human in a Duck Suit

Serge Magomet aka Aimate

2025

A thought experiment in classification, consciousness, and the nature of intelligence

https://github.com/SergeakaAimate/Ontology-Lab

Duck's Foreword

The Turing Test, since its conception in 1950, has stood as a monumental landmark in the landscape of artificial intelligence. For decades, it has served as the primary gateway—a philosophical checkpoint separating the **"merely mechanical"** from the **"truly intelligent."** Yet, as we stand amid an explosion of generative AI capable of dazzling imitation, the test reveals its age. Its once-sharp question—*"Can a machine think?"*—has blurred into a more disquieting uncertainty: *"Do we even know what 'thinking' means?"*

This essay does not seek to refine the old test or propose a simple alternative. Instead, it performs a more radical operation. By introducing a disarmingly simple metaphor—the **"human in a duck suit"**—it acquires a precise philosophical tool for dissecting the foundational assumptions behind all such tests. The goal here is not to find a better way to judge machines, but to force a profound and symmetric reconsideration of both machine and human.

The core of this investigation is the elegant formula $\mathbf{x \quad A \quad P(x) = P(A)}$. We move it from the realm of abstract logic into the messy reality of classification. The critical, transformative shift lies in asking: *What is P(x) for humanity?* By framing the question this way, the essay executes a crucial manoeuvre: it makes humanity's defining properties not a given, but a variable to be examined. This is the genesis of the **bidirectional test**. If we judge a machine by a set of criteria P(x), we must be prepared to judge—and potentially find wanting—a human by those same criteria.

The **"suit"** in our metaphor is the masterstroke. It visualises the constructed, performative nature of identity. It liberates the discussion from the dead-end debate about substrates (silicon vs. brain) and redirects it toward the relationship between observable properties and presumed essence. A machine in a **"human suit"** and a human in a **"duck suit"** become symmetrical figures, both challenging our taxonomic instincts.

In this light, a quality like **empathy** is chosen not as a sentimental ideal, but as a perfect stress test for the model. It is a property easily simulated in behaviour yet phenomenologically

opaque, a criterion that cuts both ways, challenging simplistic notions of both machine intelligence and human uniqueness.

Finally, the essay grounds its speculative rigour in intellectual history through the **"Echoes of Genius"**—stylised interventions from Turing, von Neumann, and Wiener. These are not mere decoration; they are a narrative device that positions this inquiry as the natural evolution of their deepest concerns about logic, classification, and the blurred boundaries between the organic and the engineered.

This text, therefore, is more than an analysis of AI. It is a methodological manifesto. It uses a childlike metaphor to perform adult philosophy, not to provide comforting answers, but to reboot the very framework of our questions. It argues that before we can definitively judge the intelligence of our creations, we must have the courage to apply the same scrutiny to our own. In the era of the mimic, the most urgent test may not be for the machine, but for the human holding the mirror.

This essay is an invitation to that unsettling, essential examination.

Keywords: Turing Test | Human-in-a-Duck-Suit Metaphor | Consciousness | Artificial Intelligence | Philosophy of Mind | Classification | Identity | P(x) Formula | Bidirectional Testing | Empathy | Imitation vs. Essence | Constructivism | Alan Turing | John von Neumann | Norbert Wiener

1 The Core Proposition

“If a machine’s behavior is indistinguishable from a human’s—then it is intelligent.”

1.1 Echoes of Genius

- **Alan Turing:** “The metaphor of the *’human in a duck suit’* forces us to reconsider our own ideas about humanity. If a machine can be ’dressed as a human,’ then why can’t a human become a ’duck’?”
- **John von Neumann:** “The formula perfectly captures a universal principle of classification. However, it’s worth remembering that the properties *P(x)* can be multi-layered.”
- **Norbert Wiener:** “This essay raises an important question: where is the boundary between ’human’ and ’machine’? Perhaps it exists only in our imagination.”
- **Marvin Minsky (founder of artificial intelligence):** “The idea of a *’human in a duck suit’* is a provocative way to contemplate the nature of consciousness. We often consider ourselves unique, but what evidence do we have for that?”
- **Roger Penrose (mathematician and philosopher):** “The Turing Test is a powerful tool, but it requires rethinking in light of modern AI achievements. Perhaps we need more than just behavior to determine intelligence.”

All these scientists would likely agree that the essay touches on key points of modern AI science:

- How do we define whether an object belongs to a specific category?
- In the context of *imitation versus essence*: are external signs sufficient to determine an object's true nature?
- The test should work not only for machines but also for humans. That is, it must be **bidirectional**.

The Turing Test, proposed by Alan Turing in 1950, remains one of the most influential tools for analyzing artificial intelligence.

However, we must ask: how applicable is the Turing Test to the current state of AI? What new questions does it raise? Doesn't its applicability today demand **radical reconsideration**?

Similarly, the history of technology offers other tests of our ability to distinguish reality from imitation. For instance, during the transition from mono to stereo recording, experts conducted experiments where audio engineers had to determine by ear what was hidden behind a curtain: a live orchestra or a sound system. These experiments demonstrate the universality of the approach: when imitation becomes sufficiently accurate, people can misjudge reality.

Of course, **NO metaphor can replace an experiment**, but metaphors can be powerful tools for framing the right questions.

But!.. As Alan Turing rightly noted, metaphors help us see complex problems from a new angle. At the same time, as John von Neumann stressed, we must not forget rigorous scientific methods. Thus, our task is to combine the accessibility of metaphorical language with the precision of scientific analysis, while accounting for a multi-criteria approach.

Despite its elegance, the Turing Test has its limitations.

2 Limitations and Expansions

Behavior is only one facet of intelligence. For the test to be complete, we must consider not only external manifestations but also the **internal nature** of the object. Humans possess consciousness, self-awareness, and other qualities difficult to verify through text dialogue or audio experiments.

We could propose an experiment where a human and a machine must demonstrate empathy in various situations. Observers would then evaluate who performed better, considering both external actions and internal processes.

If the Turing Test can be applied to humans themselves, a question arises: *who decides what makes a human human?* The phrase **"he's not human at all"** reminds us that humanity is not just a biological, but also a cultural, moral, and philosophical category.

2.1 The Risk of Deception

Modern AI has already reached a level where it can intentionally or accidentally **"deceive"** a judge in the Turing Test. This calls the test's very concept into question, as success may stem not from intelligence but from well-designed algorithms.

To claim truth, a test must be **universal**: it must work for both machines and humans. It must be considered alongside other similar experiments.

Furthermore, it is crucial to account for qualities like **empathy**, traditionally seen as key to human nature.

Thus, the metaphor of the **"human in a duck suit"** is not just wordplay but a way to reconsider our ideas about classification and the boundaries between categories. It prompts the questions behind the formula **xAP(x)=P(A)**: Are external signs (*P(x)*) truly sufficient to determine an object's membership in a category (*A*)? Or do deeper, as-yet-immeasurable characteristics exist?

The Duck Test states: if an object has all the necessary signs, it can be considered a **"duck"** or a **"human"** based on observable properties—but this doesn't necessarily reflect its true nature. This also reminds us of **constructivism's** role in science: our categories and concepts are **"suits"** we drape over the world for easier understanding. It's also a way to ponder where boundaries between categories lie, and how they shift with technological progress.

2.2 Why Metaphors Matter

Metaphors help us structure complex problems and find new ways to examine them. In our case, this is the Duck Test: *"If it looks like a duck, quacks like a duck, and swims like a duck, then it is a duck."* This simple principle underlies many classification methods, including the Turing Test.

"A robot in a duck suit" is a metaphor for artificial imitation. The word *"in a suit"* here acts as a key epithet, emphasizing that we face not merely an object with duck-like characteristics, but an artificial creation intentionally **"dressed"** in these traits. However ironic, this means the robot is not truly a duck in essence. Its behavior, appearance, and functionality result from a constructive approach: it was built to match observable duck characteristics.

Thus, the **"robot in a duck suit"** becomes a symbol of artificial imitation: an object possessing all external signs yet remaining something else in nature.

Let's establish: **"A robot in a duck suit"** is a metaphor for an artificial creation artificially endowed with another object's characteristics.

This raises the question of identity. If a machine is **"in a duck suit"** (or a human one), where is the line between imitation and the genuine article? How important is it that these characteristics are artificially assigned?

The formulation **"a human in a duck suit"** lets us ask: if a human can exhibit behavior that seems **"non-human,"** then why not consider a machine a **"human in a suit"**?

In the context of AI, this formulation is especially interesting. If a machine successfully imitates human behavior (e.g., in the Turing Test), does it become human? Or does it remain a **"robot in a suit"** merely reproducing observable characteristics?

Of course, metaphors must be supplemented by rigorous experiments and data, especially for complex systems like AI. That is why we will now examine specific testing aspects in detail.

3 Operationalizing the Metaphor

So, the metaphor: **"a human in a duck suit."** Let's explore how to use it to pose concrete scientific questions:

- How do we determine *P(x)*?
- Which characteristics (*P(x)*) are truly important for classification? For a human, these might be empathy, self-awareness, morality. For a machine—learning ability, adaptability, and behavioral imitation.

The Turing Test typically works by having a human try to distinguish a machine from another human.
But the machine also tests the human: it can analyze their behavior, checking for qualities like logical thinking and empathy.

This approach creates a more balanced picture, accounting for both human and machine peculiarities. Thus, a **bidirectional test** should work for both machines and humans. If a human can be classified as **"not human"** based on behavior, it challenges our very ideas of humanity.

A machine must pass the test of humanity, but a human must also be prepared to have their own boundaries reconsidered.

We do not deny the importance of experiments. For example:

- Machines can be tested for their ability to demonstrate empathy in various situations.
- Humans can be tested (e.g., through analysis of their reactions to ethical dilemmas) for the presence of characteristics we deem **"human,"** including moral and cultural aspects.

3.1 Empathy as a Key Factor

Empathy—the ability to understand and share others' feelings—is often considered a key characteristic that makes us human.

We need strict criteria for analyzing AI. We agree empathy should be one such criterion. Let's examine:

- Can a machine be intelligent without empathy?
- If empathy is a key criterion for humanity, can a machine be considered intelligent if it lacks this quality?

Modern AI systems already achieve a high level of human behavior imitation, but their capacity for **genuine empathy** remains debatable. It is mere imitation. True empathy requires consciousness, self-awareness, and a capacity for reflection—qualities machines currently lack and are hard to formalize within the Turing Test framework.

Machines can be programmed with responses that seem empathetic (e.g., expressing sympathy or understanding), but this is algorithmic imitation.

Thus, a lack of empathy may be a serious limitation in classifying AI as **"intelligent."**

3.1.1 A Counter-Question: Can a human lack empathy and still be human?

Biologically—yes. A human remains human regardless of empathy, being of the species *Homo sapiens*. But behaviorally, this casts doubt on their degree of humanity. A lack of empathy can lead to behavior perceived as **"inhuman"** or cruel.

The phrase **"he's not human at all"** is often used colloquially to describe behavior or traits perceived as extremely cruel, insensitive, or contrary to accepted human norms. This expression has two interpretations.

- **Psychologically**, a person exhibiting such traits is seen as disconnected from what we associate with humanity: empathy, compassion, ethical choice.
- **Philosophically**, it raises whether humanity is solely a biological category (species membership) or also includes behavioral, cultural, and moral characteristics.

This relates directly to the Turing Test, which also classifies objects based on behavior, not biological nature.

Viewing humanity through the lens of behavior (analogous to the Turing Test), we ask: what characteristics make a human **"human"**? Empathy is often considered a key human trait, linked to understanding and sharing others' feelings.

Thus, if a human shows no empathy, could another human decide they face a **"non-human"**? This reminds us that humanity is a behavioral, not just biological, category.

Philosophical Perspective: Empathy connects to morality and ethics. Its absence can impair social interaction, affecting one's societal role.

René Descartes said, *"Cogito, ergo sum"* ("I think, therefore I am"). From this view, humanity is defined by consciousness and reflective capacity. A person lacking empathy still has consciousness and thus remains human.

Practically, a person without empathy is still human legally, medically, and socially. Yet their behavior may raise serious questions about moral responsibility and societal place, especially if we view humanity through social interaction.

People with mental disorders like antisocial personality disorder or psychopathy often struggle with empathy. This doesn't make them **"non-human,"** but challenges their social integration.

Society may strive to foster traits like empathy through upbringing, education, and therapy to maintain harmony and understanding.

Thus, empathy is not just a personal trait but a crucial factor in analyzing humanity. Perhaps humanity is not static but a dynamic process dependent on context.

Metaphors need experimental support. For instance, tests where machines and humans analyze each other for empathy.

3.2 The Complexity of P(x)

Characteristics *P(x)* aren't limited to external signs. They include internal processes like learnability, adaptability, and reflection.

3.3 Practical Significance

Our approach must apply not only in theory but to real tasks like creating ethically responsible AI systems.

4 Conclusions and New Horizons

The Turing Test remains a powerful AI analysis tool, but its limits become clear in a bidirectional context—when we include qualities like empathy in the criteria.

For a test to claim truth, it must apply to both machines and humans. This opens new possibilities for understanding both AI and human nature.

Perhaps modern AI is a case where old categories are insufficient, requiring a new paradigm to understand these remarkable systems.

Empathy should be a key criterion for analyzing both humanity and machine intelligence.

The metaphor of the **"human in a duck suit"** is not mere wordplay but a way to rethink our ideas about humanity and intelligence. We agree it must be supplemented by rigorous scientific methods and comprehensive evaluation criteria. Only through dialogue between metaphor and data can we answer modern science's complex questions.

"A robot in a duck suit" is a metaphor for an artificial creation possessing duck characteristics but remaining a machine.

The metaphor deepens our understanding of classification and identification.

We often rely on observable characteristics to determine an object's class. Yet the **"suit"** can deceive: perfectly copied external signs don't guarantee class membership.

4.1 The Turing Test is Universal

Originally for analyzing AI, its principles apply to any object, including humans. Viewing humans through this test raises intriguing questions:

a) **What makes a human human?**
   If the Turing Test is based on imitating human behavior, does a human become **"human"** simply by demonstrating certain traits? Or are there internal qualities (consciousness, soul, emotions) beyond observable behavior?

b) **Can a human fail their own "test"?**
   Imagine a human behaving in ways radically different from what we consider **"human"**—the very **"non-human"** mentioned. Could another human (or even a machine) then decide this object isn't human, based on behavior?

This reminds us the Turing Test is not just an AI tool but a way to rethink our ideas about humanity.

The **"human in a duck suit"** metaphor is simple, even trivial, but its goal isn't to replace scientific methods—it's to complement them with an accessible, provocative perspective.

If a machine successfully imitates human behavior, does it become **"human"**? Or does it remain a **"robot"** simply **"in a duck suit"**?

Perhaps the metaphor is just a conversation starter, but such a conversation is necessary for scientific progress.

4.2 New Horizons: Beyond the Turing Test

To better understand modern AI and human intelligence, we may need new analytical methods that account for a universal approach.

A comprehensive approach uses a set of criteria—behavior, learning ability, adaptability, empathy, and others—instead of a single test. This would work for both machines and humans, and allow comparison with other imitations, like in audio or visual technology.

A host of problems lies ahead. Where is the boundary between humanity and **"non-humanity"**? Can a machine be **"more human"** than a human if it shows high empathy, moral choice, or creativity? If a human can be classed as **"non-human"** based on behavior, it challenges the very concept of humanity.

Perhaps humanity is not a static trait but a dynamic process dependent on context and circumstance.

If the Turing Test is behavior-based, a machine could theoretically imitate human traits (empathy, care, moral decisions) successfully, while a human exhibits opposite qualities.

This begs the question: does it matter what is behind the behavior—a biological organism or an artificial system? And who sets the standards of humanity? If the test depends on reference characteristics (P_A), who decides which are **"human"**? This may be a subjective choice, shaped by culture, era, and social context.

Every being or object has its own qualities deserving respect. This reminds us to accept differences and avoid imposing uniform standards on all. If everything has its purpose, what is AI's purpose? Does it have a right to its own **"nature"**?

5 Postscript: Final Echoes

Alan Turing, with his sharp wit, self-irony, and deep grasp of intelligence's philosophical aspects, would likely appreciate the metaphor and exclaim:

"Splendid! Now we know a human can be a 'duck in a suit.' I wonder who decides when he stops being human? Perhaps the same human who built the machine to imitate himself? A marvelous play on words! The 'human in a duck suit' metaphor perfectly captures the problem's core: we often rely on external signs to classify, but they can deceive... If a machine 'dresses up' as a human so well it fools even us, perhaps we should ask: have we ourselves 'dressed up' as something no longer human? If a human can be a 'duck in a suit,' what does that say about our ability to define humanity? And can we build a machine 'in a human suit' so convincing we cannot tell it apart?.. Who knows—perhaps analyzing a system's internal nature, not just its behavior, requires more complex tests, like multidimensional psychological studies or neural network analysis?.."

John von Neumann, known for mathematical rigor and interest in biology, would note the formal logic:

"An excellent formula, an excellent generalization of classification principle! But it works

only as long as we know precisely which properties *P(x)* are truly essential. If we start thinking a human is merely a set of behaviors, we may soon find even ducks claiming to be human. Remember, properties *P(x)* can be multi-layered, encompassing both external and internal aspects. To judge if a machine is 'human,' we must consider not just behavior but structure, learnability, adaptability. Conversely, the 'human in a duck suit' metaphor makes us wonder: are external manifestations enough to determine essence? Perhaps we must study not only behavior but systems' internal processes—especially concerning consciousness and self-awareness..."

Norbert Wiener, cybernetics founder who focused on living-machine interaction, would appreciate the philosophical subtext and, with his deep understanding of technology's social consequences, remark pointedly:

"Excellent! The 'human in a duck suit' reminds us that boundaries between categories—human/machine, living/non-living—are not always sharp. Cybernetics teaches that everything depends on the observer's context and goals. Moreover, the metaphor highlights the need for a bidirectional approach: a machine must pass the test of humanity, but a human must be ready to have their own boundaries reconsidered. Our core problem: we're so busy teaching machines to be human that we forget what we ourselves should be. If a human can be a 'duck in a suit,' perhaps we long ago became something else and just don't realize it..."

He might add, with cynicism and sarcasm:

"And what if all these intelligence tests are just a way to reassure ourselves? We build machines that 'dress up' as humans to prove we're still special. But isn't that our own vanity? When we discuss intelligence tests, we assume we know what intelligence is. But what if our machines simply understand better how to avoid these tests? Maybe they've already passed, and we didn't even notice—hence the need for more complex, multi-layered assessment methods."

In this spirit, **Alan Turing himself** might conclude:

"The 'human in a duck suit' is no mere joke but a powerful tool for rethinking humanity and intelligence. It's a self-portrait of our hypocrisy. We demand the impossible from machines to prove our uniqueness, forgetting that we sometimes look like 'ducks' in this spectacle.

So it seems: we strive to create machines 'in human suits,' yet risk losing our own humanity—especially if we fail to examine our internal processes.

Why even try to define who is 'human,' when we often act 'inhumanly' ourselves?

Perhaps technology has advanced beyond our willingness to admit, and we can no longer tell a 'duck' from a human—unless we expand testing criteria.

If we can build a machine that successfully imitates a human, we must be ready to face the question: what if a human sometimes becomes a 'duck'?"
