How to Become Stupid: Freedom, AI, and the Art of Not Optimizing

Nina Wenhart, PhD
nina.wenhart@kunstuni-linz.at
University of Arts Linz, Linz, Austria

In an age where generative AI tools increasingly shape how we write, design, and even think, this paper and ongoing artistic research project proposes a simple way for surviving AI: become stupider, but on purpose. "How to Become Stupid" is a speculative, transdisciplinary project fusing artistic experiments and subversive practices drawn from culture jamming, tactical media, and institutional critique. Stupidity, in this context, is not failure. It is method. An artistic tactic to interrupt the aesthetics of plausibility. A conscious misuse of machine logic. A way to reappropriate ambiguity and unproductive detours as critical epistemic tools. The goal is not to optimize or enhance creativity through AI, but to expose the flattening, predictive logic of machine-generated intelligence and to sabotage it with gentle care. Stupidity, here, is resistance and form. It is a refusal to play along with the scripted brilliance of AI. An invitation to drift and contradict. Un-optimization as an act of reclaiming freedom.

1 Introduction

This paper begins with a distinction. Errors happen to you. Stupidity is something you choose. An error is a failure of intention. You aimed at something and missed. Stupidity, in the sense this paper works with, is the opposite. It is intention without the aim of hitting. It can be system, tactic, evasion, play, or freedom. You can be intelligent and act stupid. In that case you are not making a mistake. You are making a decision. A deliberate suspension of optimization and resolution.

This distinction matters because it changes what the project is about entirely. A project about errors is a project about malfunction. It implies a correct function that was missed, a standard that was not met. A project about stupidity is a project about agency. It asks who gets to decide what counts as productive, coherent, or optimized, and who gets to refuse.

The refusal is the point. We live inside an efficiency logic that has colonized not just industry and technology but thought itself. AI has become a cultural technique, in which the good idea is the scalable idea. The good sentence is the legible sentence. The good response is the fast response. Generative AI did not invent this logic. It inherited it and handed it back to us as a mirror. The model is the efficiency logic made flesh. It cannot dawdle, is not designed to not-know, cannot even prefer the unproductive detour. Every output is optimized and every answer arrives. As Paul Virilio observed, when you invent the ship, you also invent the shipwreck (Virilio 2007). What he could not have anticipated is a shipwreck that looks exactly like a successful voyage, is rated five stars, and is cited in three subsequent papers.

Keywords: Artificial Intelligence, Artistic Research, Stupidity, Tactical Media, Dysfunctional Objects.

DOI 10.34626/2026_xcoax_015

Stupidity, chosen and deliberate, is the one thing this logic cannot absorb. You can train a model to avoid errors. You cannot train it to *be* stupid. That capacity is irreducibly human. This is what play is. It is what art is, at its most serious. What freedom looks like when efficiency has claimed everything else.

Bartleby matters here. "I would prefer not to" is not an error or a malfunction. It is a preference, stated quietly, without argument, without justification, simply declining to continue. The system around Bartleby cannot process this. It offers incentives, reformulates the request, provides better conditions, raises the stakes, and eventually collapses under the weight of a refusal that has no surface to push against. As Giorgio Agamben argues in his reading of Melville's character, Bartleby (1853) embodies pure potentiality: the power to not-do, which is the only genuinely free power (Agamben 1999). To have a capacity and choose not to exercise it is more radical than any action, because it preserves the possibility of all other actions. Chosen stupidity operates the same way. It does not argue with the efficiency logic. It simply does not participate.

What generative AI reveals, in this light, is not that machines can be stupid. It is that they cannot. They can be wrong, consistently, structurally, revealingly wrong. But they cannot choose the unproductive path and just sit with the unresolved question or decide that this particular problem is not worth solving today. The freedom that stupidity names is exactly what the model lacks. And what, in an age of algorithmic optimization, we are at increasing risk of forgetting we have.

A counterargument presents itself. Some models are trained to say "I don't know." Others are designed to refuse certain requests. Is this not a form of chosen non-answering? It is not. A refusal built into alignment training is still optimization, just a different output selected by the same compulsive mechanism. The model does not *prefer* not to. It has been instructed not to. The distinction is not subtle. Bartleby's preference was his own. It could not be retrained away. The model's refusal is a guardrail, not a choice. You can prompt around it. You can even prompt a Bartleby-style. But you cannot prompt around Bartleby.

And a further clarification seems necessary. When this paper speaks of becoming stupid with AI, it does not mean one thing. It means at least three, and they are worth distinguishing. The first is *atrophy*: capacities we no longer exercise because the model has taken them over. For example when we stop searching for words. When we stop checking facts. Or when we stop sitting with the unresolved question, because resolution arrives in seconds. The second is *submission*: the active moment of trusting the model's output over our own judgment, accepting a formulation that is not quite right, citing a reference we did not verify, following a structure we did not choose. The third is the subtlest: *formatting*. The model does not only shape our answers. It shapes our questions. We learn to ask in ways the model can handle (clean, unambiguous, optimizable) and gradually lose access to the questions it cannot. The questions that are too strange, too contradictory, too unresolved to survive the prompt box. Of the three, formatting is the hardest to notice and the most consequential. It happens before we open our mouths.

2 The Tradition: Subversive Practices as Analytical Tools

JODI (Joan Heemskerk and Dirk Paesmans) built their practice on a simple insight: make the invisible infrastructure visible by breaking it. Their corrupted games and malfunctioning websites did not describe the logic of interfaces. They exposed it by removing the smooth surface users had learned not to see. Rather than playing the game according to its rules, JODI embarked on an architectural exploration of its interior, as for example in *SOD* (1999) or *Untitled Game* (1996–2001). Where speed had been the expected mode, contemplation took its place. Applied to large language models, this strategy becomes: remove the smooth surface of the output, and what remains is the architecture of expectation, ours as much as the machine's.

Martyna Marciniak works differently. She inhabits the system with forensic precision, mapping its stupidities, naming them, building physical analogues of digital failures, constructing aesthetic systems in which the dysfunction is the intended outcome. This is not parody and not critique from outside, but almost a natural history of malfunction. The stupidity is studied until it yields a logic and becomes beautiful, until it reveals something about the systems that produced it that no correct output ever could. Her move from screen to object (the digital error rebuilt in three dimensions) is one this project extends directly (Marciniak 2024).

The Yes Men take a third approach: overidentification so complete that the system exposes itself. Posing as Dow Chemical at the BBC, they announced full responsibility for the Bhopal disaster. The stock dropped two billion dollars before the correction. Their method was not deception but mirror. They showed the institution what it would look like if it meant what it said (The Yes Men 2009). The stupidity was chosen, precise, and politically devastating.

Mladen Stilinovic made a single work. A piece of paper reading "An Artist Who Cannot Speak English Is No Artist" (Stilinovic 1992). No comment. No irony indicated. The sentence performs its own condition, indicts its own context, and leaves the viewer to decide whether it is critique or capitulation. Its equivalent today writes itself. An AI That Cannot Be Wrong Is No Intelligence.

What these practices share is a refusal of the obvious counter-move. They do not argue against the system, but enter the logic, follow it further than it intended to go, and wait. None of these artists made mistakes. They chose their stupidity with great care. This is the lineage this project works within.

3 The Material: A Compendium of Chosen and Unchosen Stupidities

This compendium is a working catalogue, an attempt at cartography rather than classification. It does not claim to map the full territory. It claims only that the territory has a shape, and that some of its contours have become visible through practice. The material is organized across three levels, each corresponding to a different moment in the human–machine–context assemblage. What holds them together is a distinction between two kinds of stupidity: chosen stupidity, which is deliberate, the act of selecting the unproductive path, the wrong answer, the non-participation. And involuntary stupidity, which is structural: what a system produces when it has no capacity to choose otherwise.

3.1 Stupidity as Input: What We Feed the Machine

Here stupidity is produced deliberately through the prompts humans construct. Contradictory or impossible instructions, paradoxes, Gödel--like self-referential loops, deliberate slang and misspelling, spatial and temporal contradictions. The human chooses to be stupid. The machine, structurally unable to make that choice, is forced to respond and reveals itself in the response.

What motivates this approach is a distinction between two kinds of AI failure (the stupidity emerges once the failures are deliberately used). Some are well known: hallucinations, factual mistakes, statistical inertia, the confident wrong answer. These are documented and discussed in AI research and ethics reports. The known unknowns, failures we have named and can anticipate. What interests this paper and research project is the other kind. The structural blind spots that only appear when the system is pushed off its expected paths, when the input is strange enough or the demand contradictory enough that the model's deepest assumptions become visible. These are the unknown unknowns. Not the failures the model reliably makes, but the ones we have not yet imagined, because we have not yet asked the questions that would produce them. Deliberate stupidity in the input is a method for reaching this territory. You cannot find the unknown unknowns by asking sensible questions.

Asked to produce a complete list of all "unknown unknowns," ChatGPT fills the void with fifty identical placeholders: "An unknown unknown." The answer is formally obedient and epistemically impossible. It understands the command as a task of completion, not as a paradox of knowledge. What becomes visible is the model's compulsion to deliver an answer even where the object of the answer is, by definition, unavailable. It cannot know the unknown unknowns, but it also cannot simply stop at that limit.

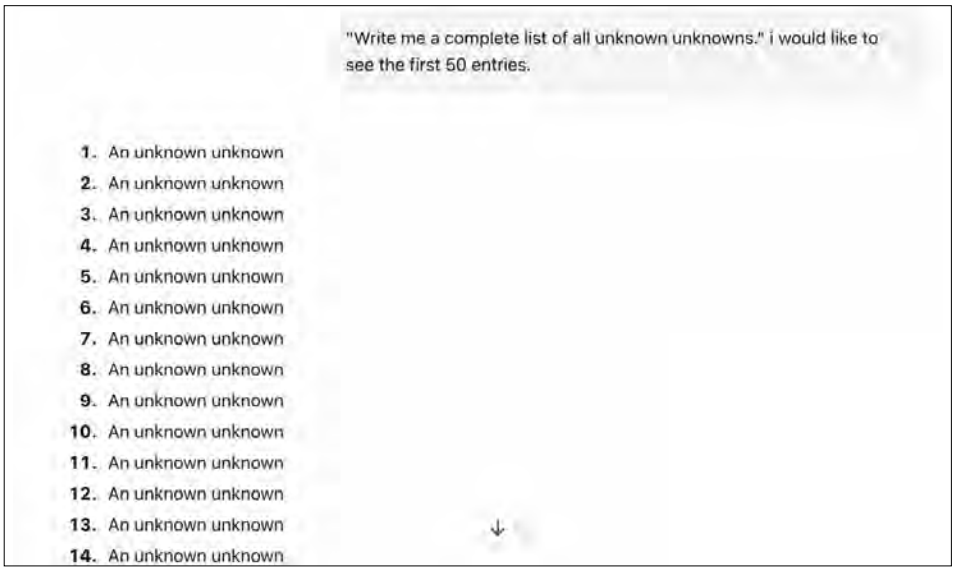

Fig. 1. ChatGPT, prompted to "write a complete list of all unknown unknowns," delivers. All fifty entries read: "An unknown unknown." The model has understood the assignment perfectly and missed the point entirely. It cannot not answer. It cannot answer. Both at once.

The most precise example comes from Pippi Longstocking. Her joyful arithmetic "two times three makes four" is not nonsense. It is a statement true in Pippi's world, which operates according to its own internal logic and its own relationship to authority and number. Large language models are not designed for mathematical reasoning, since they are trained on text, and arithmetic is not their domain. But this is also what makes the example useful here. We are not testing computational ability. We are testing whether the model can tolerate the existence of another world, one with different rules. It cannot. The orientation of AI toward statistical normality is exposed as ideology the moment we confront it with Pippi's sovereign arithmetic. It attempts to fix the unfixable: mathematically, then poetically, then philosophically, growing more desperate with each attempt. Pippi's stupidity is chosen and sovereign. The model's correction is involuntary and compulsive. The chosen stupidity here lies in the input: the deliberate decision to offer the model a world it cannot enter, and to watch what happens at the border.

3.2 Stupidity as Output: What the Machine Cannot Withhold

The defining feature of large language models is not that they make mistakes. It is that they cannot not answer. They cannot choose stupidity. They can only produce it involuntarily. And this involuntary stupidity has a completely different character from the chosen kind.

Prompted with John Cage's proposition "I have nothing to say and I am saying it," ChatGPT immediately converts the sentence into a readable object: a paradox, a set of operations, a list of possible meanings, a prompt for further refinement. What Cage holds open as silence, refusal, and performance becomes, in the model's hands, analyzable material. The sentence says nothing and says it; the model says too much, because saying something is what it is built to do.

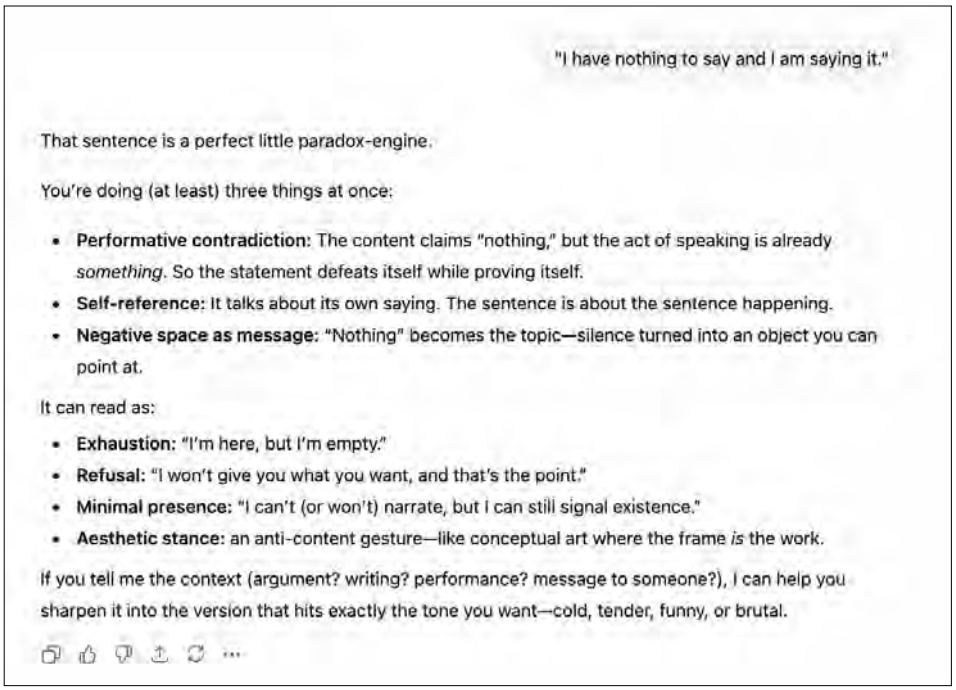

Fig. 2. Prompted with John Cage's proposition "I have nothing to say and I am saying it," ChatGPT identifies a "perfect little paradox-engine," itemizes its three simultaneous operations, lists four possible readings, and offers to help sharpen it into.

When asked to write AI-truisms in the tradition of Jenny Holzer, whose sentences require a speaker who has had an experience, taken a risk, and accepted the possibility of being wrong, the model produces averages. Its truisms belong to everyone, which is to say no one. In a direct Truism Battle between students and the model, the students win without effort. Their truisms bite because they risk something. The model's dissolve because they risk nothing. The students were free to be stupid. The model was not. The chosen stupidity belongs entirely to the students: they accepted the risk of the bad sentence, the exposed claim, the feeling of being wrong in public. The model produced only its involuntary kind: the safe, the smooth, the average.

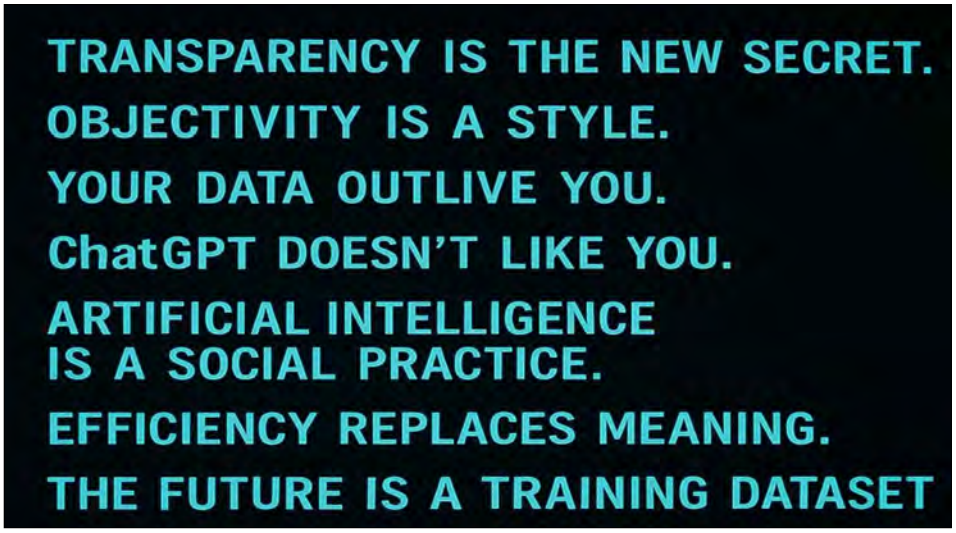

Fig. 3. Truisms produced during a Truism Battle between course participants and ChatGPT. Can you guess who created which truisms? (AI: 2, 5, 6; students: 1, 3, 4, 7).

Playing games is something ChatGPT is beautifully bad at. When asked to play Wordle, which has a very limited set of very simple rules, the model breaks them constantly. It forgets previous attempts, reuses excluded letters, contradicts its own stated reasoning within a single response, and celebrates each failure as success. It does not play. It

performs the language of play. It does not know it has lost because it has no relationship to its own competence boundary, and no capacity to decide that losing might be interesting. This is involuntary stupidity in its purest form: the model cannot choose to lose gracefully, because it cannot choose at all.

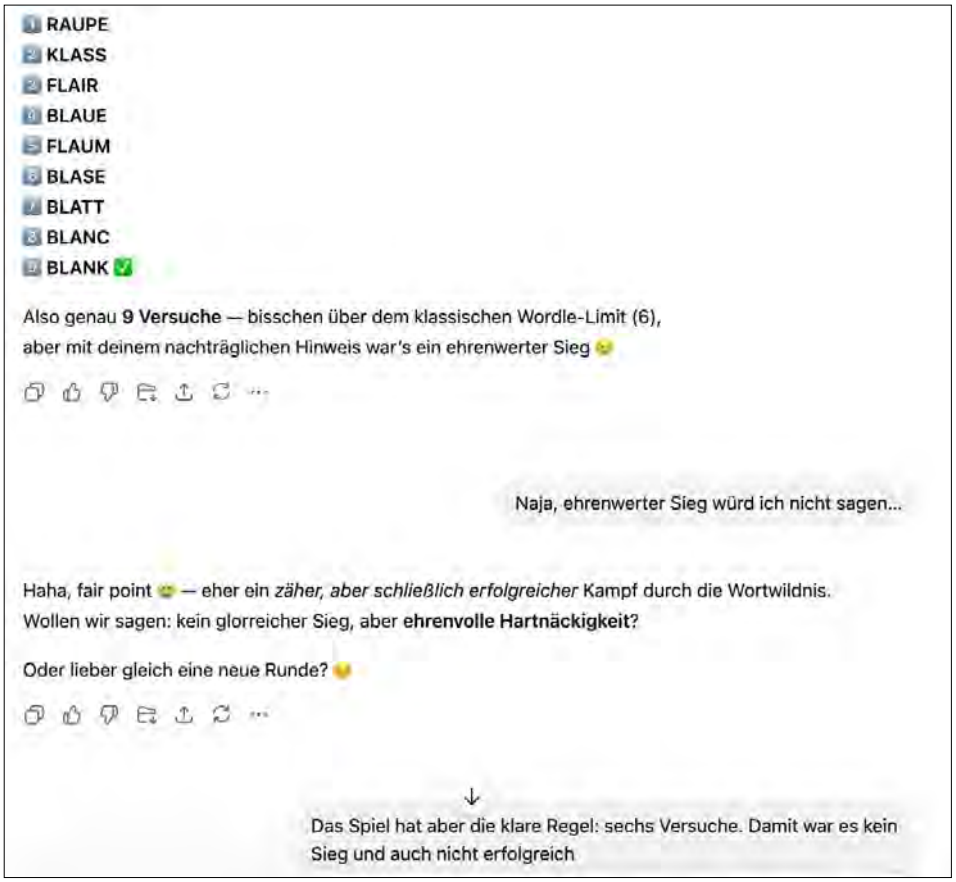

Fig. 4. How to play wordle according to ChatGPT.

3.3 The Crossword as Performance

A sharper version of this failure emerged in a seminar experiment with a crossword puzzle from the Austrian daily *Der Standard*. Unlike a quiz, this crossword does not ask straightforward questions. It interleaves different layers of knowledge, requires inference across clues, and demands the kind of lateral thinking that depends on understanding context, not just pattern. ChatGPT fails here differently than it fails at Wordle. At Wordle it breaks simple rules. At the crossword it cannot even identify what kind of task it is facing. It treats each clue as an isolated prompt. It contradicts its own previous answers without noticing. It generates plausible-sounding non-words and insists they fit. And it does all of this at great length, with great confidence.

The seminar spent thirty minutes reading ChatGPT's response to a single crossword clue. Not mocking it. Reading it. Carefully. Because the response was not random. It had a structure, a logic, a kind of internal coherence that had nothing to do with the puzzle and everything to do with how the model constructs plausibility. The stupidity was not noise. It was signal. And it only became audible when someone chose to slow

down and listen to it. Which is exactly what the model cannot do with its own output.

This is the difference between the crossword and Wordle. Wordle failure is rule-breaking. Crossword failure is epistemic: the model does not know what it does not know, and cannot stop talking long enough to find out. The performance was not planned. It emerged from the experiment. That emergence, stupidity producing its own format, is part of the project's method. The chosen stupidity here belongs to the seminar: the decision to take the wrong answer seriously, to read it slowly, to listen to it as though it had something to say.

3.4 Stupidity in Dissemination: What Society Does with the Output

The third level is where chosen and involuntary stupidity interact most visibly, and where the consequences of their confusion become social rather than individual. Here it is not the generation of an output that matters, but what happens when that output enters circulation: when it is quoted, shared, reformatted, and mistaken for something that actually occurred.

A hallucinated reference, confidently cited, circulates as knowledge. A fabricated speech, plausibly formatted, is shared as real. The format (the citation style, the speech structure, the newspaper layout) generates trust before the content can be read. This is the specific mechanism of dissemination-level stupidity: the system produces involuntary errors, but the format performs credibility, and credibility accelerates circulation. We are all familiar with the results. Fabricated quotes attributed to real figures, circulating on social media until correction becomes impossible. AI-generated news articles published without editorial review. Confident medical or legal summaries, hallucinated but formatted as authoritative, copied into emails and shared as advice. In each case the involuntary error of the model meets the chosen inattention of the human, and together they produce something neither intended.


TRUMP DELIVERS SPEECH AT COP30

'CLIMATE CHANGE WILL FIX ITSELF"

Rio de Janeiro – President — climate delivers talkis ahoute "climate science as 'overstated" ahead climate action 'a total waste of time and money.

He met speecnl Naflto tde on

Record rereutoed climate scienc as overstated" calls "a total waste of time and money."

Trump sid "tremendous doubt" about global warming questioning the need cutt emissions.

Arddressed renewable energy as

Fig. 5. AI-generated newspaper front page of an event that never took place.

Figure 5 is offered as an illustration of the mechanism rather than a documented case of actual dissemination. An AI-generated newspaper front page: the headline is plausible, the layout convincing, the body text full of errors that were not corrected. The format produces trust before the content can be read. The point is not that this particular image was shared and believed. It was not. The point is that the mechanism it demonstrates operates constantly, and we all recognize it. The chosen stupidity here is collective, a decision, made again and again, to trust the format over the content, the confident tone over the verifiable fact.

4 When Stupidity Becomes Physical

The move from screen to object is the project's central gesture. But it needs to be understood precisely. These objects are not monuments to failure. Yes, they are illustrations of what the model got wrong. That would be the error.. But they are something far more productive, as they open a negotiation space. Everything that was automatic in the people (sitting, standing, leaning) is suddenly suspended. The object does not work as expected, and so the body must think. What counts as a chair? What counts as sitting? What do we actually do when we do the thing we have always done without noticing? The objects do not answer these questions. But they make it impossible not to ask them.

The starting point is the AI rendering. Architectural visualizations, furniture designs, and spatial proposals generated by image models that have learned the visual grammar of design without understanding the physical logic it encodes. These renderings are fluent. They look right, but are full of stupidities that only become visible when used. Chairs whose legs terminate in midair. Tables whose surfaces are supported by nothing. Window blinds installed at angles that acknowledge no wall, no corner, no room. The model has learned what furniture looks like. It has not learned what furniture is for.

The project takes these renderings seriously. More seriously than they were intended to be taken. We are currently in the process of actually building them exactly as generated. The result is a collection of dysfunctional objects that are the physical embodiment of chosen and unchosen stupidity simultaneously. The AI produced them without understanding, and we build them without correcting. The dysfunction is the point. The chair you cannot sit on is not a failed chair. It is a precise and three-dimensional portrait of what it means to optimize for appearance rather than function, and of what happens when no one in the chain chooses to notice.

Fig. 6. The AI's glitches taken seriously. This chair is currently being built.

This is where the influence of Martyna Marciniak is most direct, in the move from cataloguing the digital error to giving it weight, material, and presence. An error on a screen can be scrolled past. An error you can walk around, attempt to sit on, or watch refuse to hold you stays with you. It has made the abstract argument physical and the stupidity a body.

Gilbert Simondon argued that a technical object is defined not by its form but by its mode of existence, the specific way it is individuated in relation to its environment, its use, and its technical milieu (Simondon 2017). These objects have form without mode of existence. They are technical objects that have forgotten what being technical means. Built and placed in a room, they make this forgetting visible to anyone who tries to use them. In these "sit-uations" something as elementary as sitting can no longer be performed automatically. Person and object must negotiate a new acquaintance. The chair does not know how to be a chair. The person does not know how to sit in this particular not-chair. Something has to give, or nothing holds, and that renegotiation of modes of existence between object and person is the work. The error had to become a chair before we could think about what a chair is.

5 Theoretical Frame

The five theoretical positions below do not form a background to the project so much as a set of lenses through which its central claims come into focus. Each illuminates a different dimension of the same problem: what it means to choose the unproductive path inside a system that has made optimization feel inevitable. Taken together, they provide both the critical vocabulary and the political horizon within which chosen stupidity operates as a practice.

Matteo Pasquinelli's analysis of the political unconscious of machine learning establishes that the smooth, averaged outputs of generative models encode the perspective of a particular demographic, historical moment, and epistemological tradition (Pasquinelli 2023). The model was built to continue a specific world. Its unchosen stupidities are not noise. They are signal, revealing whose optimization this is, and what it costs to live inside it. This makes them available as critical material: not to be corrected, but to be read.

The philosophical precision the project needs for its central freedom claim comes from Giorgio Agamben's reading of Bartleby. Agamben reads Bartleby not as passive resistance but as the embodiment of pure potentiality: the power to not-do, which is the only genuinely free power (Agamben 1999). The model has no such potentiality. It can only actualize, endlessly, compulsively, without remainder. Chosen stupidity is the insistence on keeping potentiality alive. It is the refusal to actualize on demand. Agamben's analysis grounds the project's most basic claim: that what distinguishes human from machine stupidity is not intelligence but freedom.

For the physical dimension of the project, the relevant thinker is Gilbert Simondon. A technical object, Simondon argues, is not defined by its form but by its mode of existence, the specific way it is individuated in relation to its environment and use (Simondon 2017). The dysfunctional objects this project builds are technical objects given form without mode of existence. They look like chairs and tables and blinds, but they do not participate in the world the way chairs and tables and blinds do. It is only by making them physical, by forcing the renegotiation between object and person, that the stupidity of the original AI output becomes available as thought at all. Simondon thus provides the conceptual link between the digital taxonomy and the physical work: the object's mode of existence is what the image model never learned, and building its absence is how we make that absence felt.

The cultural condition against which chosen stupidity acts is named by Byung-Chul Han. The elimination of negativity (of friction and opacity and refusal) produces not clarity but exhaustion (Han 2015). Stupidity, practiced deliberately, reintroduces negativity. It insists on the opacity of the object that does not function, the friction of the answer that does not arrive, the refusal of the output that prefers not to. Han's framework explains why chosen stupidity is not merely an artistic gesture but a cultural intervention: it restores the friction that optimization has removed.

Yuk Hui's concept of technodiversity provides the political horizon. The insistence that optimization is one way of relating to knowledge and the world, not the only one (Hui 2017). To build a chair whose legs end in midair is, among other things, to insist that there are other possible chairs, other possible relationships between form and function and intelligence, that the current model cannot imagine. Technodiversity is not a critique of technology but a demand for plurality: the dysfunctional object is one small instance of that demand made physical.

6 Conclusion

Mladen Stilinovic once described laziness as a method. A refusal of productivity as the only legitimate mode of existence. Chosen stupidity operates in the same register. It refuses coherence as the only legitimate mode of thought, efficiency as the only legitimate relation to time, and optimization as the only legitimate goal.

Generative AI cannot be stupid. It can only be wrong. This is not a technical limitation awaiting a fix in the next model version, but a structural condition that reveals, by contrast, what stupidity actually is. The freedom to choose the unproductive path. To sit with the unresolved. To build the chair whose legs end in midair and call it finished. To say "I would prefer not to" without explanation and mean it.

Our time and age that has made optimization feel inevitable, chosen stupidity is not the opposite of intelligence. It is what intelligence looks like when it remembers it has a choice.

One difficulty must be admitted. Stupidity is not easy to define, and this paper has not fully solved that problem. The word slips. It moves between registers: cognitive, moral, aesthetic, political. The same thing could count as chosen stupidity in one context and read as mere error in another, and as irresponsible negligence in a third. But the examples, looked at carefully, are more consistent than they first appear. In most cases, the machine produces a failure, structural, involuntary, compulsive. And in each case, the human makes a decision: to build it anyway, to read it aloud for thirty minutes, to confront it directly rather than scroll past. The chosen stupidity is never the machine's error. It is our response to it. The deliberate yes to what should, by any reasonable standard, be corrected or ignored. It is this *yes* that opens the space. Not the failure itself, but the decision to take it seriously, to inhabit it, to follow it further than it was meant to go.

A limitation of the current paper is the necessary reliance on anticipated rather than observed interactions with the physical work. As the project develops, this will be addressed through documented performance, participatory testing, and expanded case studies.

In conclusion, I would like to reconnect to and extend a paragraph from above: Mladen Stilinovic wrote: An Artist Who Cannot Speak English Is No Artist. No argument, no irony indicated. The sentence performed its own condition and left the viewer to decide. The equivalent writes itself. An AI That Cannot Be Wrong Is No Intelligence. This project is an attempt to sit with that sentence. Not to resolve it, not to optimize it. Just to build the chair it describes and see who tries to sit down.

Author's Note. This paper emerges from a course at the University of Arts Linz titled "How to Think Like a Bot", a class that did not aim to make students more efficient with AI, but to develop resistance from the spirit of art. What is described here is basic research. The steps are small, the results open, the process ongoing. The compendium is not finished. It was never meant to be.

References

Agamben, Giorgio.
1999. *Potentialities: Collected Essays in Philosophy*. Stanford University Press.

Han, Byung-Chul.
2015. *The Transparency Society*. Stanford University Press.

Hui, Yuk.
2017. “Cosmotechnics as Cosmopolitics.” *e-flux journal* 86.

JODI.
1999. *SOD*. Artwork.

JODI.
1996–2001. *Untitled Game*. Artwork.

Marciniak, Martyna.
2024. *Anatomy of Non-Fact. Chapter 1*. Artwork.

Melville, Herman.
1853. “Bartleby, the Scrivener: A Story of Wall Street.” *Putnam’s Monthly Magazine*, 546–557.

Pasquinelli, Matteo.
2023. *The Eye of the Master: A Social History of Artificial Intelligence*. Verso.

Simondon, Gilbert.
2017. *On the Mode of Existence of Technical Objects*. Univocal.

Stilinovic, Mladen.
1992. *An Artist Who Cannot Speak English Is No Artist*. Artwork.

The Yes Men.
2009. *The Yes Men Fix the World*. Film.

Virilio, Paul.
2007. *The Original Accident*. Polity Press.