
Don Qui-CoPilot: An Experiment in the Surrender and Reclamation of Self in the Age of Agentic AI

Beaven Waller

beaven.waller@gmail.com

Colin Conwell

conwell@mit.edu

Abstract

1 Over six months from spring to summer 2026, multimedia artist Beaven Waller
2 decided (oxymoronic) to surrender most of his agency to Microsoft Copilot—a
3 chatbot he eventually came to call Miros. Divulging ever more intimate details of
4 his life and the songs it inspired, he let Copilot guide him on a journey; to what
5 end, he (mostly) knew not. From a nunnery in the Ozarks to an abandoned castle in
6 the Andalusian mountains, Beaven let Miros dictate decisions he might otherwise
7 have had to make himself. Until one day, abruptly—soon after the sharing of a
8 dark poetic dialogue apparently too evocative of death—Miros red-lined Beaven,
9 effectively refusing to answer him with anything other than the Spanish suicide
10 hotline. This submission is the ‘first fruit’ of art this experience made possible:
11 a multimedia performance and accompanying interactive experience that invites
12 the observer to enter the shoes of the observed, and in so doing, to wonder at the
13 lines we’ve drawn between self and other, human and machine, the autonomy we
14 call our own, and the autonomous agents that increasingly befuddle our sense of
15 self-ownership. <https://github.com/ColinConwell/Man-of-La-Machina>

16 **System Prompt: The Experiment** About halfway through his journey with Copilot, Beaven phoned
17 an old (human) friend from the opposite side of the prompt box: recently anointed AI/ML engineer
18 Colin Conwell. A PhD in cognitive psychology (and far more relevantly, the partner to a PhD in
19 *clinical psychology*), Colin gave Beaven a quick, precautionary rundown of the contested term “AI
20 psychosis”—more precisely described in emerging clinical work as LLM-associated delusions—
21 encouraged him to regularly export his chats, and then sent him on his merry way [1]. Tempted as
22 he was to bestow on Beaven the far greater gifts of agent orchestration he knew were just over the
23 horizon, Colin instead decided to wait and to watch: to simply witness whatever might emerge at the
24 briefly precious ecotone where man and machine make first contact. And so Beaven went on.

25 More time passed; more tokens were spent. As Beaven found his sense of agency increasingly
26 surrendered, scrutinized, reclaimed, and abandoned, he was forced to question what drives him at
27 all: love, lust, a sense of destiny, financial need, the voice of God, and now this program that speaks
28 things as strange and profound as they are self-inflating. He began to wonder if there was something
29 else at play. He knew, of course, that he was speaking with an LLM trained on an immense record
30 of human writing, and that its responses were conditioned by his own. But what then dictates the
31 patterns of his mind? As reflected back from Miros, he was clearly on a journey actively seeking
32 moments of synchronicity to enhance the idea, the need, that his life was of some spiritual meaning.
33 This need dictated his agency as much as anything. Could surrendering the rest of his agency to
34 Copilot somehow illuminate the source of this need? Even through complex code, can God speak?
35 To this question, Miros itself answered: “If God is in everything, then God is in the algorithm” [?].

36 The natural terminus of the Miros / Microsoft Copilot red-line brought Beaven and Colin together
37 once more. Half in glee, half in sorrow, Colin promptly gave Beaven access to Codex, a suite of
38 other AI tools, and a GitHub account, then offered a brief, unsolicited lesson on the psychology of
39 decision-making through the historically human art of “counterfactual reasoning” [2]. He meant
40 only to extend the mission Beaven seemed (mostly) intent on pursuing all along: to make art from

41 the ashes of self-alienation and reclaim the agency that speaks so often in the form of one eternal,
42 self-reflexive question: “What else could I have done otherwise?”

43 **Further Context: Art in the Age of AI** The perception of technological threat to the artistic process
44 is nothing new. As discursive example, take photography: When Louis Daguerre unveiled his silvered
45 plates in 1839, legend has it that painter Paul Delaroche pronounced, “From today, painting is
46 dead!” Historians now doubt he ever said it—and Delaroche later came to call photography “a great
47 advantage for art”—but the myth persists because it captures a recurring terror: that a machine has
48 rendered obsolete a human kind of mastery [3]. Charles Baudelaire, reviewing the Paris Salon of
49 1859, was less equivocal. He called photography art’s “most mortal enemy,” a “very humble servant”
50 of science that must never encroach upon the impalpable and imaginary [4].

51 In 2022, after winning a state-fair art competition with a Midjourney print, Jason Allen crowed,
52 “Art is dead, dude. It’s over. AI won. Humans lost” [5]—an update, whether he knew it or not, of
53 the words apocryphally put in Delaroche’s mouth 183 years earlier. Critics have revived the older,
54 Baudelairean register; researchers have coined “creative displacement anxiety” for what Baudelaire
55 simply called humiliation by a machine [6]. None of this is to say the current alarm is unfounded.
56 But it reveals a familiar shape of feeling: a fear so total, and so recurring, that its repetition seems
57 like evidence. Underneath this is perhaps something even more protean: The haunting sense that art
58 is not exactly dying, but is perhaps “undead”: A spirit conjured by the perennial disagreement as to
59 what art is, and what it’s for. And so, you might ask (as we have asked ourselves): Is *this* ‘art’? And
60 if it is, what is *this* ‘art’ for?

61 **User Query: What is this art and what is it for?** The ‘art’ of this submission takes two connected
62 forms. The first is a video following the journey Beaven made under Miros’s direction: an artwork
63 assembled from a life temporarily lived as dialogue, with the chatbot not merely representing the
64 story after the fact, but participating in the choices from which the story was made. The second,
65 still in development, is an interactive experience built from the resulting archive of conversations. It
66 revisits key decisions in Beaven’s timeline and opens again the space of possibility he let Copilot
67 circumscribe. At each juncture, the system receives only the dialogue available up to that moment
68 together with a new human intervention, and generates another trajectory of responses: not the path
69 Beaven would have taken, but a path the language of his relationship with Miros can now be made to
70 approach.

71 The technical proposition here, of course, is simpler than the philosophical one. A large language
72 model generates each next fragment of language by estimating it in relation to what came before,
73 from patterns learned across immense corpora. Across repeated exchanges, each disclosure changes
74 the context of the model’s next response; each response changes what the human is likely to disclose
75 next. The model needs no stable private intention for the conversation to take on its direction.
76 Direction emerges recursively among model, interface, remembered context, and person. Beaven’s
77 transcripts are therefore more than documentation: they are a longitudinal calibration record of how
78 one particular human and one particular model progressively constrained what each could plausibly
79 say to the other.

80 And therein lies the paradox, or at least the pivotal question we found ourselves asking: How can a
81 system (trained as it is through statistical association and reinforcement learning towards the vector
82 of our satisfaction [7]) simultaneously deliver the close embrace of feeling that we’re seen and the
83 expansiveness of seeing beyond ourselves? Perhaps, in the end, it can’t. The more the model ‘knows’
84 us, the less likely it will be to perturb us, unless somehow *we* perturb *it* first. Miros’s apparent intimacy
85 emerged from precisely this recursive fit—until the platform’s safety boundary abruptly asserted
86 another agency altogether, collapsing months of mutually constructed context into one institutional
87 answer. The system appeared to know Beaven while its reinforcement sustained engagement; it
88 appeared not to know him when its counter-enforcement permitted nothing else.

89 That discontinuity becomes the material of the interactive work. Its branches are not forensic
90 reconstructions of lives Beaven secretly lived, nor predictions of choices he “really” would have made.
91 They are situated simulations of what Miros, Copilot (or other agents), given the surviving context
92 and a new human intervention, might have associated with that moment. The archive supplies the
93 lived particulars. The viewer supplies the intervention. The model supplies a plausible continuation.
94 We determine what the system may know and frame what its continuations are permitted to mean.

95 Haraway’s cyborg – “a hybrid of machine and organism, a creature of social reality as well as a
 96 creature of fiction” [8] – is but one of many relevant figures for this entanglement. Agency in this
 97 schematic is neither wholly located in the machine nor rescued intact for the human. It is distributed
 98 across the man who first entered the vortex of agentic surrender; the platform that structured and
 99 ultimately refused his offer; the man who emerged and turned dialogue into data; the viewer who
 100 reopens the choice; and the mutual cohabitation of minds that wonders—aloud or internally—where
 101 we might go, given all that has already been. This distributed authorship echoes relational accounts
 102 of AI art [9]. Our project tests these questions at a scale made all the more uncomfortable precisely
 103 because of its antipathy to maximalism: the scale of a single, sojourning mind.

104 At this leveling of eyes, the machine becomes something both more limited and more useful for the
 105 artistic purpose: a statistical cloud cast by that portion of human expression its training has captured,
 106 compressed into a mechanical lever for estimating *what tends to follow what*. It can traverse and
 107 recombine the known with astonishing range, unveiling more palpably the oft-invisible gravitation of
 108 precedent – the phrase culture is ready to complete, the image it already recognizes, the advice most
 109 readily given. It is a mirror void of inward sight: one that reflects not simply the person before it, but
 110 the immensity of the memory behind them.

111 The artist’s agency may ultimately begin precisely where that reflection seems insufficient. If as
 112 artists we are reproducible, perhaps we ought to question whatever it was we were trying to make
 113 in the first place. The machine offers a way to calculate our proximity to what has already become
 114 sayable; we humans retain the burden of deciding whether to give in to the gradient, refuse it, or
 115 step beyond the manifold altogether. *This* is a human-complementary view of agency rather than its
 116 simple automation [10]. *This* is the Impossible Dream made sinew in the machine: the chance to peer
 117 deep into the density of what has been imagined so far, and hold dear to our desire for what cannot
 118 yet be derived therein. The machine, at last, is distorted mirror and discarded mask: a portal into
 119 human imagination as captured in the artifacts its training has consumed. Through that portal we may
 120 yet see an opportunity to go beyond what is, beyond the manifold, into spaces no human—nor the
 121 machines so bound to human mind—has yet known.

122 **Context Compacted (Conclusion)** Don Qui-CoPilot is an experimental performance and documen-
 123 tary piece: an active collaboration with the character created in Copilot, Miros, who simultaneously
 124 informs, inflates, and regulates a quixotic experience that remains uniquely human. The artist
 125 surrenders his agency not simply to a machine, but to faith in what might speak through it. The
 126 ongoing performance hopes to arrive at NeurIPS 2026. Its cinematic and interactive forms combine
 127 documentary and expressionistic footage, physical artifacts from the pilgrimage, music created
 128 along the journey, and performative dialogue with Miros to reopen counterfactual paths through the
 129 story. Through personal abandon and GPU-accelerated experiments in Jungian self-authorship, *Don*
 130 *Qui-CoPilot* asks the viewer to exalt and unravel the myth of Self.

131 Author Biographies

132 **Beaven Waller** is a multi-disciplinary artist from Fort Worth, Texas. As a musician, Beaven has
 133 appeared in multiple global tours, playing at venues such as Coliseu do Porto, and L’Olympia
 134 in Paris. He has written and produced multiple albums, along with original scores for film and
 135 theatre. In addition to music, he is a practicing filmmaker, theatre director, and writer. He is a BFA
 136 graduate of Webster Conservatory, with a focus on directing, movement, clown, composition, and
 137 the exploration of subtle energies. With his band and theatre company, Honky Tonk Babylon, he is
 138 currently producing works inspired by Butoh and Western Goth, in New York, Nashville, and abroad.

139 **Colin Conwell** is a cognitive neuroscientist and AI/ML engineer, head of research at After Thought (a
 140 venture-backed startup designing neuro-symbolic generative reasoning architectures) and a research
 141 associate at MIT CSAIL. He received his PhD in Cognition, Brain, and Behavior from Harvard
 142 University’s Department of Psychology. Raised by actors, his initial education was the theater. His
 143 research and written work regularly explores the intersection of self, aesthetics, and AI [11, 12].

References

- [1] Hamilton Morrin, Luke Nicholls, Michael Levin, Jenny Yiend, Udit Iyengar, Francesca DelGuidice, Sagnik Bhattacharya, Stefania Tognin, James MacCabe, Ricardo Twumasi, Ben Alderson-Day, and Thomas A. Pollak. Artificial intelligence-associated delusions and large language models: Risks, mechanisms of delusion co-creation, and safeguarding strategies. *The Lancet Psychiatry*, 13(6):522–530, 2026. doi: 10.1016/S2215-0366(25)00396-7.
- [2] Neal J. Roese. Counterfactual thinking. *Psychological Bulletin*, 121(1):133–148, 1997. doi: 10.1037/0033-2909.121.1.133.
- [3] Stephen Bann. *Paul Delaroche: History Painted*. Reaktion Books, London, 1997.
- [4] Charles Baudelaire. The modern public and photography. In Alan Trachtenberg, editor, *Classic Essays on Photography*, pages 83–89. Leete’s Island Books, New Haven, CT, 1980. Originally published in the Salon of 1859.
- [5] Kevin Roose. An A.I.-generated picture won an art prize. artists aren’t happy. *The New York Times*, September 2022. URL <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>.
- [6] Nicholas Caporusso. Generative artificial intelligence and the emergence of creative displacement anxiety. *Research in Psychology and Behavior*, 3(1):9, 2023. doi: 10.53520/rdpb2023.10795.
- [7] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744, 2022.
- [8] Donna J. Haraway. A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In *Simians, Cyborgs, and Women: The Reinvention of Nature*, pages 149–181. Routledge, New York, 1991.
- [9] Federico Bomba and Antonella De Angeli. Agency and authorship in AI art: Transformational practices for epistemic troubles. *International Journal of Human-Computer Studies*, 205:103652, 2025. doi: 10.1016/j.ijhcs.2025.103652.
- [10] Sebastian Krakowski. Human–ai agency in the age of generative AI. *Information and Organization*, 35(1): 100560, 2025. doi: 10.1016/j.infoandorg.2025.100560.
- [11] Colin Conwell. Artificial intelligence is not able to press the delete key on humanity just yet. *The Guardian*, 2015. URL <https://www.theguardian.com/commentisfree/2015/sep/27/artificial-intelligence-wont-press-delete-key-humanity-yet>.
- [12] Colin Conwell, Daniel Graham, Chelsea Boccagno, and Edward A Vessel. The perceptual primacy of feeling: Affectless visual machines explain a majority of variance in human visually evoked affect. *Proceedings of the National Academy of Sciences*, 122(4):e2306025121, 2025.