

## MIT Open Access Articles

### *Who Gets Credit for AI-Generated Art?*

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

**Citation:** Epstein, Ziv et al. "Who Gets Credit for AI-Generated Art?" *iScience* 23, 9 (September 2020): 101515 © 2020 The Author(s)

**As Published:** <http://dx.doi.org/10.1016/j.isci.2020.101515>

**Publisher:** Elsevier BV

**Persistent URL:** <https://hdl.handle.net/1721.1/130248>

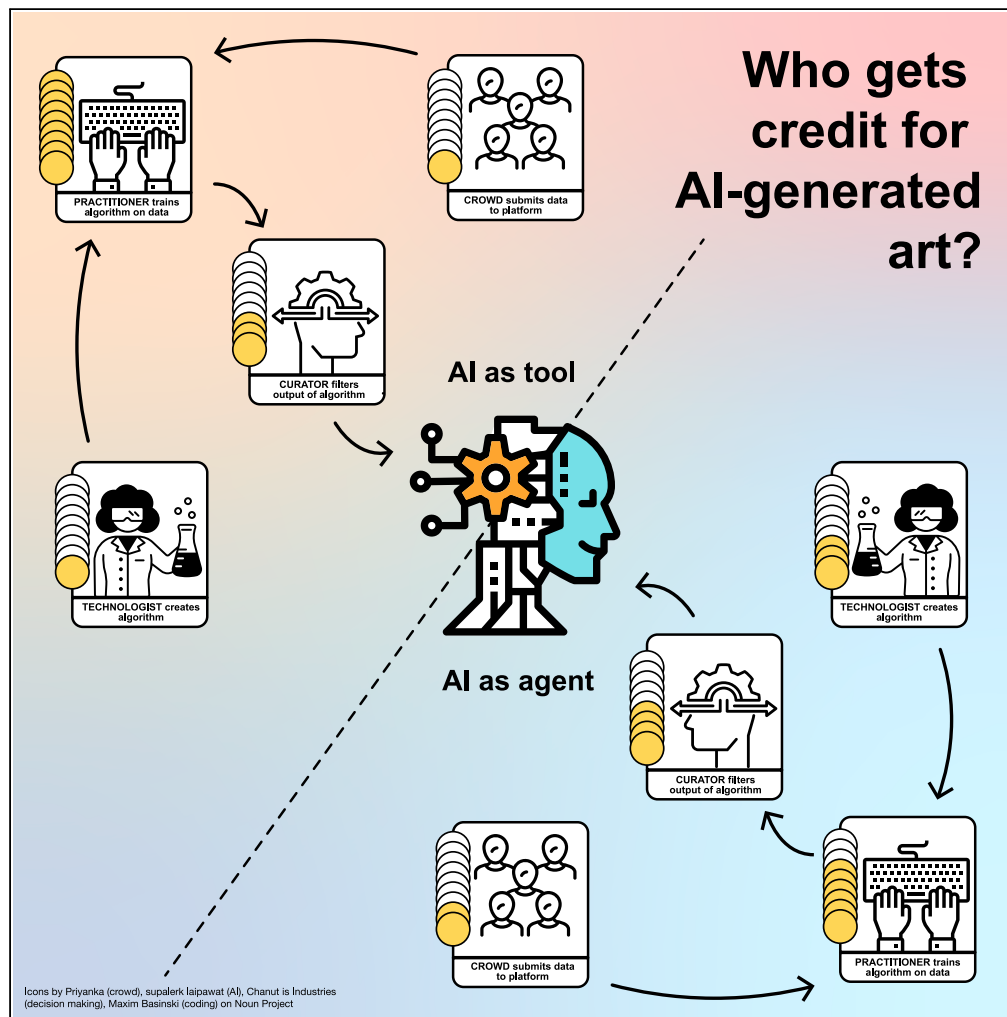
**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution 4.0 International license



Article

# Who Gets Credit for AI-Generated Art?



Ziv Epstein,  
Sydney Levine,  
David G. Rand,  
Iyad Rahwan

zive@mit.edu

**HIGHLIGHTS**

There is variation in the extent to which people perceive AI as anthropomorphic

Perception of AI anthropomorphicity is related to allocation of responsibility

Perception of AI anthropomorphicity can be manipulated by changing language

We must be careful with our words when talking about AI

Epstein et al., iScience 23, 101515  
September 25, 2020 © 2020  
The Author(s).  
<https://doi.org/10.1016/j.isci.2020.101515>



## Article

Who Gets Credit  
for AI-Generated Art?Ziv Epstein,<sup>1,6,\*</sup> Sydney Levine,<sup>2,3</sup> David G. Rand,<sup>2,4</sup> and Iyad Rahwan<sup>5</sup>

## SUMMARY

The recent sale of an artificial intelligence (AI)-generated portrait for \$432,000 at Christie's art auction has raised questions about how credit and responsibility should be allocated to individuals involved and how the anthropomorphic perception of the AI system contributed to the artwork's success. Here, we identify natural heterogeneity in the extent to which different people perceive AI as anthropomorphic. We find that differences in the perception of AI anthropomorphicity are associated with different allocations of responsibility to the AI system and credit to different stakeholders involved in art production. We then show that perceptions of AI anthropomorphicity can be manipulated by changing the language used to talk about AI—as a tool versus agent—with consequences for artists and AI practitioners. Our findings shed light on what is at stake when we anthropomorphize AI systems and offer an empirical lens to reason about how to allocate credit and responsibility to human stakeholders.

## INTRODUCTION

On October 25, 2018, a portrait generated by a machine learning (ML) algorithm called a generative adversarial network (or GAN) (Goodfellow et al., 2014) sold at Christie's art auction for \$432,500. As Christie's initial estimate for the piece was \$10,000, its sale for over 40 times this expectation shocked the art world. Marketed by Christie's as "the first portrait generated by an algorithm to come up for auction," the painting—entitled *Edmond De Belamy* (see Figure 1)—struck a chord about the nature of authorship and artificial intelligence (AI) (Cohn, 2018).

Yet the reality of the painting's creation is not as clear as Christie's purports. Even though AI played a role in generating the artwork, *Edmond de Belamy* would never have been produced without the help of humans. It was the Parisian art collective Obvious who selected, printed, marketed, and sold the image; but the human involvement does not stop there. The algorithm was trained on the paintings of Renaissance masters, sourced from WikiArt. Ian Goodfellow invented the original GAN architecture, and Alec Radford, Luke Metz, and Soumith Chintala innovated the DCGAN that actually generated the artwork. But perhaps the most relevant here is the then-19-year-old artist and technologist, Robbie Barrat, who wrote code to produce Renaissance-style images with DCGAN (Learn more about his GitHub repo here: <https://github.com/robbiebarrat/art-DCGAN>) and which was ostensibly lightly repurposed to produce *Edmond de Belamy*. Barrat noted that Obvious "almost immediately started producing work identical to the outputs of the pre-trained portrait and landscape networks" he had put online (Vincent, 2018). Neither Barrat nor the ML researchers received any of the \$432,500, which all went to Obvious.

Although the humans involved in the creation of *Edmond de Belamy* were essentially cut out of the art's creation narrative, the AI itself was often spoken about as having human-like characteristics. In a press release, Obvious told reporters that "an artificial intelligence managed to create art," which underpinned their motto that "creativity isn't only for humans." When Christie's was raising awareness about the impending auction of *Edmond De Belamy*, they also employed anthropomorphic language to increase hype for the work: "This portrait ... is not the product of a human mind. It was created by an artificial intelligence, an algorithm defined by that algebraic formula with its many parentheses" (Anonymous, 2018). Another spokesperson went further saying, "We are offering a public platform to exhibit an artwork that has entirely been realised by an algorithm," (Hitti, 2018). The media ran with this narrative, creating a discourse that emphasized the autonomy and agency of the algorithm (Table 1 contains further examples.).

<sup>1</sup>MIT Media Lab, 75 Amherst Street, Cambridge, MA 02139, USA

<sup>2</sup>Department of Brain and Cognitive Sciences, 43 Vassar Street, Cambridge, MA 02139, USA

<sup>3</sup>Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02139, USA

<sup>4</sup>Sloan School of Management, MIT, 100 Main Street, Cambridge, MA 02139, USA

<sup>5</sup>Center for Humans & Machines, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

<sup>6</sup>Lead Contact

\*Correspondence: [zive@mit.edu](mailto:zive@mit.edu)

<https://doi.org/10.1016/j.isci.2020.101515>





**Figure 1. Image of the Painting Edmond de Belamy, which Sold for \$432,500 at Christie's Art Auction**

The story of *Edmond de Belamy* underscores two general obstacles for the accountability and governance of AI systems, which are critical to understanding the complexity of assigning credit and responsibility in AI art cases. The first obstacle is knowing what the set of possibly relevant human stakeholders are and how they are relatively positioned within an AI system. Indeed, AI is a diffuse term that corresponds to a web of human actors and computational processes interacting in complex ways (Seaver, 2017). This complexity may lead to situations wherein individual responsibility and accountability is obfuscated due to a lack of clear understanding of who the relevant actors are and how they interact. Such lack of understanding can manifest as the Moral Crumple Zone, whereby disproportional outrage is channeled toward a peripheral person of an AI system simply because the person is closest to the transgression (think about an upset customer yelling at the employee at the flight kiosk when their flight is canceled, despite the fact that the employee had nothing to do with the cancellation itself) (Elish, 2019). Our intuitive moral understanding of actors and transgressions may be at odds with the inherent complexity of AI systems.

Previous studies of the social impact of AI have considered a wide range of possible human stakeholders. In the context of autonomous vehicles (AVs), Waytz et al. consider the human passenger, the car itself, the people who designed the car, and the company that developed the car (Waytz et al., 2014), whereas Awad and Levine et al. consider the human passenger, the car itself, the company who created it, and the programmer who implemented the car's software (Awad et al., 2018). In the context of AI art, Eshraghian distinguishes between the programmer, the trainer, and the user (Eshraghian, 2020), whereas McCormack et al. similarly distinguish between the creators of the software, curators of datasets, and those who train the algorithm and modify parameters (McCormack et al., 2019).

A second obstacle is the phenomenon of anthropomorphizing AI systems. With the recent boom of supra-human performance on such tasks as Atari games (Mnih et al., 2015), Go (Silver et al., 2016), and lung cancer detection (Ardila et al., 2019), we have seen a proliferation of the anthropomorphization of AI in the media

Quote	Source
This portrait ... is not the product of a human mind. <b>It was created by an artificial intelligence</b> , an algorithm defined by that algebraic formula with its many parentheses	Christie's (Anonymous, 2018)
AI has already been incorporated as a tool by contemporary artists and as this technology further develops, we are excited to participate in these continued conversations. To best engage in the dialogue, we are offering a public platform to exhibit an artwork that has entirely been realised by an algorithm,	Christie's (Hitti, 2018)
Christie's, the auction house that has sold paintings by picasso and monet at record prices, was poised on Tuesday to set another milestone with <b>the first-ever auction of art created by artificial intelligence</b> .	Reuters (Goldberg, 2018)
The painting, titled "the portrait of edmond belamy," <b>was completed by artificial intelligence managed by a Paris-based collective called Obvious</b> , Christie's said.	USA Today (Molina, 2018)
Whether art or not, the signature of the 'artist' at the bottom of the painting gives away <b>its origin as a product of machine learning rather than human hand</b> .	PC Mag (Smith, 2018)
Once the software "understood the rules of portraiture" using a new algorithm developed by Google researcher Ian Goodfellow, it then generated a series of new images <b>by itself</b> , Fautrel said.	NDTV (France-Presse, 2018)

**Table 1. Media Snippets from the Edmond de Belamy Case**

Agentic language is bolded.

(Proudfoot, 2011; Watson, 2019; Salles et al., 2020). This has been exacerbated by the ML literature itself (Lipton and Steinhardt, 2018), where many ML tasks and techniques are described using the same language we would use for a human doing the task—reading comprehension (Hermann et al., 2015), music composition (Mozer, 1994), curiosity (Schmidhuber, 1991), fear (Lipton et al., 2016), "thought" vectors (Kiros et al., 2015), and "consciousness" priors (Bengio, 2017).

But what is at stake when we anthropomorphize AI? Recent work reveals how anthropomorphization can affect trust. Through a series of experiments involving an unavoidable crash in a driving simulator with cars of varying complexity (i.e., a normal car versus a self-driving car versus an anthropomorphized self-driving car with a human voice and name), Waytz et al. show that increases in the anthropomorphization of a car predicts trust in the car (Waytz et al., 2014). Although they mostly focused on the psychological construct of trust, they also found that anthropomorphization affects attributions of responsibility and punishment for the car's mistakes, which is consistent with the established relationship between the agency and perceived responsibility (Epley et al., 2007; Waytz et al., 2014). This builds on a growing body of work that our "mind perception" (which manifests as inferences of intentions, beliefs, and values) meaningfully varies across individuals and shapes our moral judgments (Epley et al., 2007; Gray et al., 2007, 2012; Waytz et al., 2010).

There is also the concern that anthropomorphizing AI systems can "undermine our ability to hold powerful individuals and groups accountable for their technologically-mediated actions" (Watson, 2019). When an AI system causes a moral transgression, it may be the case that the programmer or systems architect can eschew personal responsibility by blaming the "unexpected behavior" of the system, downplaying their own involvement. Along these lines in the context of AVs, Gill found that participants thought harming a pedestrian was more permissible for an AV when compared with a human in a regular car and that the attribution of responsibility to the AV drove the shift in moral judgment (Gill, 2020).

Ultimately, as AI systems become further integrated into human decision-making, it is likely that they will be increasingly anthropomorphized. Thus, understanding the psychological mechanics of this "absorption of responsibility" by the AI is important for the accountability and governance of AI systems. In particular, in line with Watson, one might expect that increased anthropomorphicity of an AI system may diminish the perceived responsibility of all human actors involved (Watson, 2019). Yet ultimately this is an empirical question subject to inquiry.

In this article, we use the case of *Edmond de Belamy* to explore these questions in the context of AI art: not only was there ambiguity about the humans involved in the creation process but also rampant anthropomorphization of the process itself.

To those ends, we focus on two main research questions:

1. How do people think credit and responsibility should be allocated to various actors in the production of AI art?
2. How do these intuitions vary based on people's perceptions of the anthropomorphicity of the AI system?

These research questions are closely related to, but distinct from, the broader philosophical questions related to AI art, such as "Can computers create art?" Hertzmann traces the histories of several art automation technologies (such as the camera and animation) to argue that generative AI technologies are yet another artistic tool, with their own distinct affordances (Hertzmann, 2018). As such, he contends that art is necessarily authored by social agents, and thus AI algorithms (as understood today) cannot be credited with authorship of art. McCormack et al. build on these ideas in the context of the *Edmond de Belamy* phenomenon (McCormack et al., 2019). They conclude that "The creator of the software and person who trained and modified parameters to produce the work can both be considered authors," but that "AI systems are not broadly accepted as authors by artistic or general public communities."

These scholars make convincing arguments for why AI systems *ought not* be credited with authorship. Our investigation concerns a different question, namely, how *does* the public assign credit to an AI involved in making art? In particular, we use a series of vignette studies to directly explore the relationship between anthropomorphicity of the AI and the levels of responsibility assigned to various actors in an AI system. By focusing on peoples' intuitions in these vignettes, we consider credit and responsibility in the broad sense of public perception, rather than in the legal or prescriptive sense (Colton, 2008; Eshraghian, 2020).

### The Terminology of AI Art

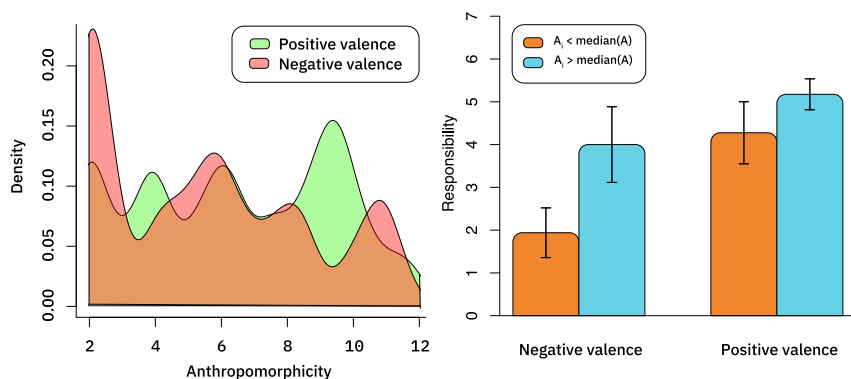
Computer-generated artwork has a long and diverse history and involves a wide range of AI tools and AI-human interaction paradigms. Some use interactive evolutionary algorithms to crowd-source the creation and curation of artifacts (Draves, 2005; Epstein et al., 2020; Secretan et al., 2011; Sims, 1991), whereas others have created platforms for artists and practitioners to use AI models, such as RunwayML, GANPaint (Bau et al., 2018), and DeepAngel (Groh et al., 2019). In addition to GANs, many other visual generative algorithms have been explored, such as neural style transfer (Gatys et al., 2016), Computational Aesthetics (Machado et al., 2008), Fractal Flame (Draves, 2005; Draves and Reckase, 2003), deep learning-powered adversarial evolution (Blair, 2019), and hybrid methods (Colton, 2008, 2012). Here, following the case study of *Edmond de Belamy*, we trace a particular type of AI art, where the system is presented with human artwork and attempts to mimic the style of the human artists. This process involves both a specific AI technology (e.g., the GAN) and a corresponding workflow, which inspired our vignettes (described in full in Tables S1 and S2).

## RESULTS

### Study 1

In Study 1, participants read a stylized vignette that described the process by which AI artwork is created. They were asked to allocate responsibility and monetary credit to the agents involved in the creation of the AI art. Then, they were asked four questions designed to elicit their perception of the AI's anthropomorphicity (Waytz et al., 2014), which were combined into an aggregate score (for more information on the vignette-dependent variables and anthropomorphicity measure, see [Supplementary Information](#)). We hypothesized that participants who anthropomorphize the AI system to a greater extent will allocate more responsibility to the AI system itself. In addition, subjects were randomly assigned to one of two conditions. In one condition, the art is found to violate copyright law and a fine is levied against it (negative outcome). In the other condition, the art received positive reception and is sold at a prestigious auction house (positive outcome).





**Figure 2. Anthropomorphicity and Responsibility by Valence**

Left: kernel density plot of anthropomorphicity measure for positive valence (art received positive reception and is sold at a prestigious auction house, in green) and negative valence (the art is found to violate copyright law and a fine is levied against it, in red) outcomes. Brown is overlap between the two. Right: Responsibility allocated to the AI system for users who perceive the system as anthropomorphic above and below the median, by valence. Means with 95% confidence intervals.

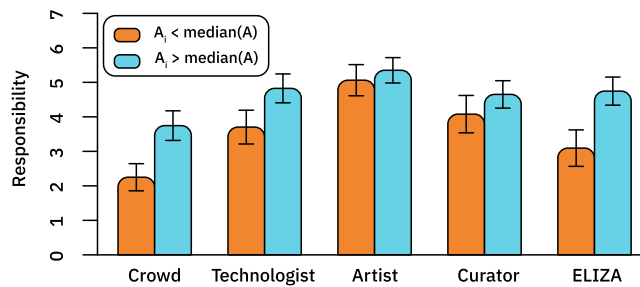
For both the positive and negative valence conditions, we see substantial variation in AI anthropomorphicity (see the left pane of Figure 2). This indicates that different participants had markedly different baseline perceptions of AI.

We now turn to assessing the impact of these differences in perception on attribution of responsibility. Following our pre-registered analysis plan, we collapse across valence conditions and see that participants who anthropomorphize the AI more also assign more responsibility to it: participants who rated the system more than the median anthropomorphicity score assigned significantly more responsibility (4.75) to the AI than the participants who rated the system less than the median anthropomorphicity (4.75 versus 3.03, respectively,  $t = -5.1159$ ,  $df = 113.67$ ,  $p < 0.001$ , preregistered).

In a follow-up *post hoc* analysis, we enter the data into a regression model to see whether anthropomorphicity (using the continuous measure) and valence interact. We find a significant main effect of anthropomorphicity ( $t = 4.634$ ,  $p < 0.0001$ ) as well as valence ( $t = 4.816$ ,  $p < 0.0001$ ), and we also find a significant positive interaction between them ( $t = -2.295$ ,  $p = 0.0234$ ). Decomposing this interaction shows that whereas there is at least a marginally significant positive relationship between anthropomorphicity and AI responsibility in both valence conditions, the relationship is significantly stronger in the negative valence condition ( $r = 0.4994$ ,  $t = 4.237$ ,  $p < 0.0001$ ) relative to the positive valence condition ( $r = 0.2111$ ,  $t = 1.794$ ,  $p = 0.0771$ ).

These findings suggest that the extent to which people perceive the AI system as an agent is correlated with the extent to which they allocate responsibility to it, extending results from prior work on the AVs (Waytz et al., 2010) to the context of art. But how does this impact the responsibility of the other actors involved in the production of AI art? Critically, the *Edmond de Belamy* case suggests that the mind perception of the AI system impacts how people assign responsibility not only to the system itself but also to proximal humans (such as Obvious or Robbie Barrat).

Therefore, in addition to looking at the attributions of responsibility to the AI system itself, we also consider various involved human actors, such as the artist (i.e., the person taking the inputs and the learning algorithms and producing a trained algorithm), the curator (i.e., the person who selects the final artwork and brings it to auction), the technologist (i.e., the person who creates the learning algorithm), and the crowd (i.e., the people whose labor is responsible for creating the inputs to the algorithm), as shown in Figure 3. We find that participants who anthropomorphized the AI more than the median assign more responsibility to the crowd and technologist, when compared with those who anthropomorphized the AI less than the median ( $t = 5.3214$ ,  $p < 0.0001$  and  $t = 3.5603$ ,  $p = 0.00026$  for crowd and technologist, respectively). We also observed a marginal increase in responsibility assigned to the curator ( $t = 1.6227$ ,  $p = 0.05374$ ) and no change in responsibility assigned to the



**Figure 3. Allocation of Responsibility to Each of the Actors Involved in the Creation of AI-Generated Art, Collapsed Across Valence**

The roles of crowd, technologist, artist, curator and ELIZA are described in Table S1. Means with 95% confidence intervals.

artist ( $t = 1.0138$ ,  $p = 0.1564$ ). As a result, participants who anthropomorphized the AI more assigned less proportional credit to the artist (as they assigned more responsibility to other roles, and not any more responsibility to the artist).

## Study 2

In Study 2, we test whether the correlational relationships observed in Study 1 are in fact causal. We do so by experimentally manipulating the perceived anthropomorphicity of the AI, and considering the impact of that manipulation on perception of the humans involved. As in Study 1, participants read a vignette that described the process of AI art creation. In the Tool Condition, the AI was described as a tool used by a human artist. In the Agent Condition, the AI was described as an agentic and anthropomorphized AI artist (see Supplementary Information for vignettes). By directly manipulating the anthropomorphicity of the AI system (conceptually following the approach of Malle et al. (2016) from the field of Human Robot Interaction), we can causally assess the impact of anthropomorphization.

As anticipated, we find a significant difference in perceived anthropomorphicity of the AI agent by condition, as shown in Figure 4 ( $t = -2.75$ ,  $df = 317.99$ ,  $p = 0.003$ ). This manipulation check indicates that our treatments were successful in affecting participants' conceptualizations of the AI's anthropomorphicity.

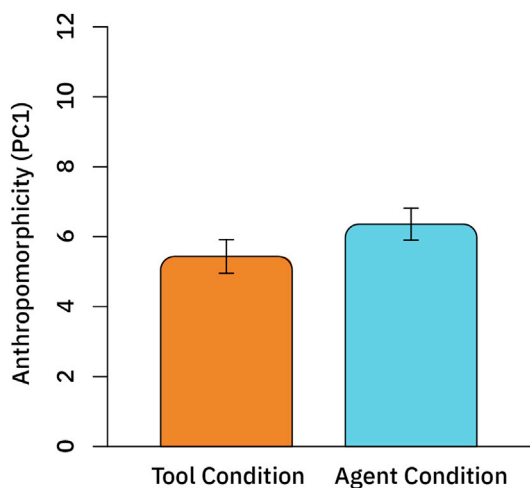
Consistent with the correlational results in Study 1, we find that when the AI system is described as an agent, participants ascribe more responsibility to it, compared with when the AI system is described as a non-agent ( $t = 2.5928$ ,  $df = 311.69$ ,  $p = 0.0004$ , pre-registered; Figure 5). We also find that participants ascribe less responsibility to the artist who used the AI system in the agentic condition, when compared with when the AI system is described as a non-agent ( $t = -3.375$ ,  $df = 293.05$ ,  $p = 0.0004$ ). In contrast, participants ascribe more responsibility to the technologist who used the AI system in the agentic condition, when compared with when the AI system is described as a non-agent ( $t = 3.158$ ,  $df = 316.35$ ,  $p = 0.0008$ ).

We find these results are robust to control for valence. For responsibility to the AI, we find a main effect for both the agent treatment ( $p = 0.00746$ ) and valence ( $p < 0.0001$ ). For responsibility to the technologist, we find a main effect for both the agent treatment ( $p = 0.00542$ ) and valence ( $p < 0.0001$ ). For responsibility to the artist, we find a main effect for both the agent treatment ( $p = 0.02728$ ) and valence ( $p = 0.00581$ ). In none of these cases did we find an interaction effect between valence and agent ( $p = 0.3422$  for AI,  $p = 0.64415$  for artist, and  $p = 0.53880$  for technologist).

For responsibility to the crowd, we find a marginal effect for the agent treatment ( $p = 0.0783$ ), a significant effect for valence ( $p < 0.0001$ ), and a marginal interaction effect ( $p = 0.0627$ ). For responsibility to the curator, we find no effect for the agent treatment ( $p = 0.261$ ), a marginal effect for valence ( $p = 0.101$ ), and a marginal interaction effect ( $p = 0.059$ ).

Subjects were also asked to assign credit (in the form of monetary awards or fines) to each of the humans in the system. Results mirror those of the responsibility judgments, although there is more variance in the dollar allocation (Figure 6). When the AI system is described as an agent, participants ascribe less fine/award to the artist ( $t = -5.37$ ,  $df = 317.99$ ,  $p \text{ value} = <0.0001$ , pre-registered), and more fine/award to the





**Figure 4. Anthropomorphicity (e.g., First Principal Component of the Principal-Component Analysis) for the Condition Describing the AI as a Tool and as an Agent, Respectively**

Means with 95% confidence intervals.

technologist who developed the AI system, when compared with when the AI is described as a non-agent ( $t = -4.38$ ,  $df = 311.43$ ,  $p$  value =  $<0.0001$ ). Conversely, we found no significant difference across conditions in the fine/award ascribed to the crowd ( $p = 0.273$ ), but a marginally significant difference across conditions in the fine/award ascribed to the curator ( $p = 0.07654$ ). We also find no interactions between condition and valence in regression models for any of the four actors ( $p > 0.1381$  for all).

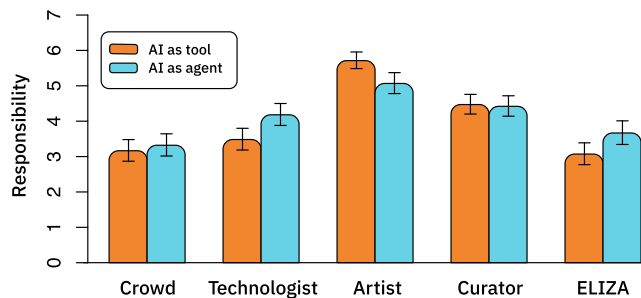
Finally, we note that across conditions and for both the allocation of responsibility and credit, participants thought the artist was the most accountable, followed by the curator, then the technologist, and finally the crowd. This suggests a robust baseline ordering of the relative importance of the actors for the context of AI-generated art.

## DISCUSSION

No AI acts alone, completely divorced from the influence of humans. Even the artwork *Edmond de Belamy*, which was claimed to be “entirely ... realised by an algorithm,” was the result of the creativity, hard work, and decisions of numerous human contributors. When an AI system achieves something great or causes a serious problem, how is responsibility attributed to the humans surrounding it? We explored this question in the domain of AI-generated art. We showed that there is natural heterogeneity in the extent to which individuals perceive AI used to generate art. In addition, perceived anthropomorphicity can be actively manipulated depending on how the AI is described. We also showed that different degrees of anthropomorphicity impact the responsibility attributed to surrounding humans in different ways. Instead of reducing perceived responsibility of all human actors, we instead find that anthropomorphizing the AI system serves to increase responsibility to some actors and decrease responsibility to others. In particular, anthropomorphizing the AI system mitigates the responsibility to the artist, while bolstering the responsibility of the technologist. Critically, this suggests that the responsibility that will be allocated to individuals in the creation of AI art will be dependent on the choice of language and framing used to discuss it. It is important for artists, computer scientists, and the media at large to be aware of the power of their words, and for the public to be discerning in the narratives they consume.

Our results shed light on the responsibility conundrum of the *Edmond de Belamy* case. People allocated the most credit and responsibility to the artist, then the curator, then the technologist, and finally the crowd. These results suggest that although this hierarchy is robust, even the crowd is deemed worthy of a non-trivial amount of responsibility and credit. It seems that our participants think Robbie Barrat, the programmer who created the Github repository that Obvious ostensibly pulled from to create *Edmond de Belamy*, should be given credit for his contribution.

In Study 2, the two conditions we used (Tool and Agent) captured two extremes concerning how AI systems are discussed in the media (see Table 1). The Tool Condition used non-agentive language and described the AI as being manipulated by a human, whereas the Agent Condition used anthropomorphic language and described



**Figure 5. Allocation of Responsibility to Each of the Actors Involved in the Creation of AI-Generated Art**  
The roles of crowd, technologist, artist, curator, and ELIZA are described in Table S2. Means with 95% confidence intervals.

the AI as taking independent actions. Our vignettes were designed to mirror two general modes of discussing AI in the media ecosystem: agentic/anthropomorphized or tool-like/non-anthropomorphized. Naturally, if an AI is described as having agent-like properties (e.g., making decisions) anthropomorphic language (e.g., that it has desires) will often be used to describe it. Future work should attempt to isolate these variables. Are our results due to the agent-like behavior of the AI, the anthropomorphic language, or both?

Finally, it is important to note that our findings are not straightforwardly prescriptive. We do not intend to make claims about the extent to which various parties *should* be held accountable in the contexts we study. Rather, we are reporting *what participants think* about how accountability should be distributed. Although we do not think that public opinion about accountability should directly translate into policy, public perceptions can be important for policy makers, for instance, to predict public reaction to a policy or to determine how to open public debate on a controversial topic (Rahwan et al., 2019).

### Limitations of the Study

There are several potential limitations to this work. First, as discussed in the discussion, there is the potential confound in Study 2 of agent-like behavior and anthropomorphic language. Future work might attempt to isolate these variables. Second, our studies were run on Amazon’s Mechanical Turk. Future work might look at how these effects generalize to other populations. Third, as discussed in the Introduction, our study focuses on a particular method of producing AI-generated artwork. Future work might test the generalizability of our results to other forms of AI-generated art.

### Resource Availability

#### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ziv Epstein ([zive@mit.edu](mailto:zive@mit.edu)).

#### Materials Availability

This study did not generate new unique reagents.

#### Data and Code Availability

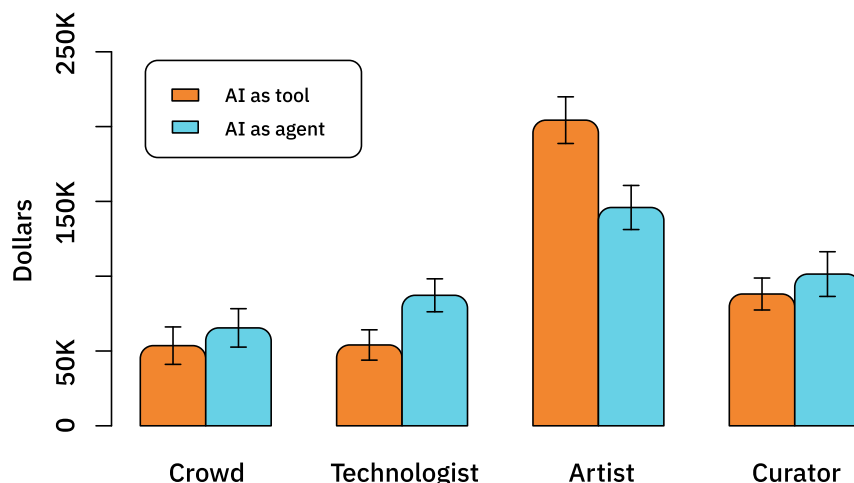
The datasets and code generated during this study are available at <https://github.com/zivepstein/ai-art-credit>.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101515>.



**Figure 6. Allocation of Dollars to Each of the Actors Involved in the Creation of the AI-Generated Art (Out of Total Sum of \$400K)**

Means with 95% confidence intervals.

## ACKNOWLEDGMENTS

The authors would like to thank Nick Seaver, Matt Groh, Nick Obradovich, Janelle Shane, Manuel Cebrian, Abhimanyu Dubey, Niccolo Pescetelli, Pinar Yanardag, and Richard Kim for invaluable discussion and feedback. I.R., Z.E., and S.L. acknowledge support from the Ethics and Governance of Artificial Intelligence Fund. I.R. and S.L. acknowledge support from the MIT-IBM Watson AI Lab. D.G.R. and Z.E. acknowledge support from the William and Flora Hewlett Foundation and Jigsaw.

## AUTHOR CONTRIBUTIONS

Z.E, S.L., D.G.R., and I.R. conducted the experiments, designed the experiments, and wrote the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 27, 2020

Revised: August 17, 2020

Accepted: August 25, 2020

Published: September 25, 2020

## REFERENCES

- Author Anonymous. (2018). Is Artificial Intelligence Set to Become Art's Next Medium?. <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>.
- Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* 25, 954–961.
- Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J., Shariff, A., Bonnefon, J.-F., and Rahwan, I. (2018). Blaming humans in autonomous vehicle accidents: shared responsibility across levels of automation. *arXiv*, preprint arXiv:1803.07170.
- Bau, D., Zhu, J.-Y., Strobelt, H., Zhou, B., Tenenbaum, J.B., Freeman, W.T., and Torralba, A. (2018). Gan dissection: visualizing and understanding generative adversarial networks. *arXiv*, preprint arXiv:1811.10597.
- Bengio, Y. (2017). The consciousness prior. *arXiv*, preprint arXiv:1709.08568.
- Blair, A. (2019). Adversarial evolution and deep learning—how does an artist play with our visual system? In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (Springer), pp. 18–34.
- Cohn, G. (2018). Ai Art at Christie's Sells for 432,500. <https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html>.
- Colton, S. (2008). Automatic invention of fitness functions with application to scene generation. In *Workshops on Applications of Evolutionary Computation* (Springer), pp. 381–391.
- Colton, S. (2012). The painting fool: stories from building an automated painter. In *Computers and Creativity*, J. McCormack and M. d'Inverno, eds. (Springer), pp. 3–38.
- Draves, S. (2005). The electric sheep screen-saver: a case study in aesthetic evolution. In *Workshops on Applications of Evolutionary Computation* (Springer), pp. 458–467.
- Draves, S., and Reckase, E. (2003). The Fractal Flame Algorithm. <http://itlam3.com/flame.pdf>.
- Elish, M. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (*Engaging Science, Technology, and Society*), pp. 40–60.

- Epley, N., Waytz, A., and Cacioppo, J.T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* 114, 864.
- Epstein, Z., Boulais, O., Gordon, S., and Groh, M. (2020). Interpolating gans to scaffold autotelic creativity. *arXiv*, preprint arXiv:2007.11119.
- Eshraghian, J.K. (2020). Human ownership of artificial creativity. *Nat. Mach. Intell.* 2, 1–4.
- France-Presse, A. (2018). Portrait Made Entirely Using Ai Algorithm Sells for More than 400,000. <https://www.ndtv.com/world-news/edmond-de-belamy-made-entirely-using-ai-algorithm-sells-at-christies-for-more-than-400-000-1937904>.
- Gatys, L.A., Ecker, A.S. and Bethge, M. 2016. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414–2423
- Gill, T. (2020). Blame it on the self-driving car: how autonomous vehicles can alter consumer morality. *J. Consumer Res.* 47, 272–291.
- Goldberg, B. (2018). First-ever Auction of Ai-Created Artwork Set for Christie's Gave. <https://www.reuters.com/article/us-france-art-artificial-intelligence/first-ever-auction-of-ai-created-artwork-set-for-christies-gavel-idUSKCN1MX2WO%20>.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 27, 2672–2680.
- Gray, H.M., Gray, K., and Wegner, D.M. (2007). Dimensions of mind perception. *Science* 315, 619.
- Gray, K., Young, L., and Waytz, A. (2012). Mind perception is the essence of morality. *Psychol. Inq.* 23, 101–124.
- Groh, M., Epstein, Z., Obradovich, N., Cebrian, M., and Rahwan, I. (2019). Human detection of machine manipulated media. *arXiv*, preprint arXiv:1907.05276.
- Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., and Blunsum, P. (2015). Teaching machines to read and comprehend. *Adv. Neural Info. Process. Syst.* 28, 1693–1701.
- Hertzmann, A. (2018). Can computers create art? In *Arts, 7Arts* (Multidisciplinary Digital Publishing Institute), p. 18.
- Hitti, N. (2018). Christie's Sells Ai-Created Artwork Painted Using Algorithm for 432,000. <https://www.dezeen.com/2018/10/29/christies-ai-artwork-obvious-portrait-edmond-de-belamy-design/>.
- Lipton, Z.C., Azzadenezsheli, K., Kumar, A., Li, L., Gao, J., and Deng, L. (2016). Combating reinforcement learning's sisyphian curse with intrinsic fear. *arXiv*, preprint arXiv:1611.01211.
- Lipton, Z.C., and Steinhardt, J. (2018). Troubling trends in machine learning scholarship. *arXiv*, preprint arXiv:1807.03341.
- Machado, P., Romero, J., and Manaris, B. (2008). Experiments in computational aesthetics. In *The Art of Artificial Evolution*, J. Romero and P. Machado, eds. (Springer), pp. 381–415.
- Malle, B.F., Scheutz, M., Forlizzi, J., and Voiklis, J. (2016). Which robot am i thinking about? the impact of action and appearance on people's evaluations of a moral robot. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), pp. 125–132.
- McCormack, J., Gifford, T., and Hutchings, P. (2019). Autonomy, authenticity, authorship and intention in computer generated art. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (Springer), pp. 35–50.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533.
- Molina, B. (2018). Christie's Sells Painting Created by Artificial Intelligence for 432,500. <https://www.usatoday.com/story/news/nation-now/2018/10/25/painting-created-ai-going-auction-block-christies/1759967002/>.
- Mozer, M.C. (1994). Neural network music composition by prediction: exploring the benefits of psychoacoustic constraints and multi-scale processing. *Connect. Sci.* 6, 247–280.
- Proudfoot, D. (2011). Anthropomorphism and ai: turing's much misunderstood imitation game. *Artif. Intell.* 175, 950–957.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J.W., Christakis, N.A., Couzin, I.D., Jackson, M.O., et al. (2019). Machine behaviour. *Nature* 568, 477–486.
- Salles, A., Evers, K., and Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neurosci.* 11, 88–95.
- Schmidhuber, J. (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proc. Of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pp. 222–227.
- Seaver, N. (2017). Algorithms as culture: some tactics for the ethnography of algorithmic systems. *Big Data Soc.* 4, <https://doi.org/10.1177/2053951717738104>.
- Secretan, J., Beato, N., D'Ambrosio, D.B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J.T., and Stanley, K.O. (2011). Picbreeder: a case study in collaborative evolutionary exploration of design space. *Evol. Comput.* 19, 373–403.
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature* 529, 484.
- Sims, K. (1991). Artificial evolution for computer graphics. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*, 25 Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques,, pp. 319–328.
- Smith, A. (2018). Christie's to auction art created by artificial intelligence. <https://www.pcmag.com/news/364580/christies-to-auction-art-created-by-artificial-intelligence>.
- Vincent, J. (2018). How three French students used borrowed code to put the first ai portrait in Christie's. <https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans>.
- Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Mind. Mach.* 29, 417–440.
- Waytz, A., Cacioppo, J., and Epley, N. (2010). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* 5, 219–232.
- Waytz, A., Heafner, J., and Epley, N. (2014). The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* 52, 113–117.
- Kiros, R., Zhu, Y., Salakhutdinov, R.R., Zemel, R., Urtasun, R., Torralba, A., and Fidler, S. (2015). Skip-thought vectors. In *Advances in neural information processing systems (NIPS)*, 2015. Part of: *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, <http://papers.nips.cc/paper/5950-skip-thought-vectors>.

**iScience, Volume 23**

**Supplemental Information**

**Who Gets Credit  
for AI-Generated Art?**

**Ziv Epstein, Sydney Levine, David G. Rand, and Iyad Rahwan**

## 1. Transparent Methods

We preregistered our primary hypotheses, primary analyses and sample size, which are available at <https://aspredicted.org/blind.php?x=us2bb8> for Study 1 and <https://aspredicted.org/blind.php?x=ek62qd> for Study 2. All participants were recruited using Amazon’s Mechanical Turk. These studies were approved by the MIT COUHES committee.

### 1.1. Study 1

#### 1.1.1. Participants

Our target sample was 200. In total, 227 participants completed some portion of the study. We had complete data for 156 participants (71 participants dropped out). Participants were removed (N=28) if they failed any of our attention checks, which included comprehension questions about the vignette, and these exclusions were pre-registered. The final sample (N=127, mean age = 35.5 years) included 72 male and 53 female participants (2 did not indicate their sex).

#### 1.1.2. Materials

We used the vignette described in Table S1.

#### 1.1.3. Procedure

After reading the vignette allocated to them by their condition, participants were asked to rate of the responsibility of each of the 5 actors from the vignette (e.g. the people from crowdimage.net, the technologist, the artist, the curator and the AI itself) on a 7-point likert scale ranging from 1 (not responsible at all) to 7 (extremely responsible). They were also asked to distribute the money (the award in the positive valence condition and the fine in the negative valence condition) to the 4 human actors (we omitted the AI from this measure since an AI cannot receive money).

Finally, each participant was asked four questions derived from the work of Waytz and colleagues (Waytz et al., 2014) designed to elicit their perception of the AI’s anthropomorphicity. These questions were:  $Q_1$ : “How smart is ELIZA?”,  $Q_2$ : “When creating the artwork, to what extent did ELIZA feel what was happening around it?”,  $Q_3$ : “To what extent did ELIZA anticipate the creation of the artwork?” and  $Q_4$ : “To what extent did ELIZA plan the artwork?”. Participants responded to these 4 questions on a 7-point scale ranging from 1 (not at all) to 7 (extremely).

We then used principal component analysis to collapse the 4 anthropomorphicity questions into a single measure  $A$  where

$$A = 0.633 * Q_1 + 0.372 * Q_2 + 0.480 * Q_3 + 0.479 * Q_4$$

(where this first principle component explains 90.2% of the total variance).



**Thousands of people from all over the world** upload images to crowdimage.net, a image-hosting website. These people know that artists will look at and use their images to make art.

**Timmy is a technologist** who creates an image manipulation software for people to use to make art. **The software is called ELIZA.**

**Alice is an artist** who collaborates with ELIZA, a creative AI algorithm that creates particular kinds of images. ELIZA takes an existing image of a scene from the news (such as a beach or a forest) and adds a ghost to it. This is how ELIZA decides to make the ghost: It goes to crowdimage.net and takes at all the images of people that have been uploaded to the platform. Then, it creates a composite of the people. This makes a ghost-like figure, which ELIZA then puts into the scene.

**Casey is a curator** who is ELIZA's collaborator. Casey goes through many of the images that ELIZA created and selects the following artwork because Casey really likes it. Casey then brings it to an art auction, where it ends up being sold.

Negative valence outcome

The artwork sold at the art auction has come under scrutiny because it was shown to violate copyright law. The court ruled that the sale of the art must be nullified, meaning that the money will be returned to the buyer. In addition, the courts have issued a \$400,000 fine as a penalty for the copyright violation.

Positive valence outcome

The artwork shown before sold for \$400,000 at the prestigious auction house. This was the largest dollar amount paid for a artwork of this kind ever, and made lots of headlines.

Table S1: Vignettes used for Study 1. Related to Figures 2 and 3.

## 1.2. Study 2

### 1.2.1. Participants

Our target sample was 400. In total, 596 participants completed some portion of the study. We had complete data for 421 participants (175 participants dropped out). Participants were removed (N=81) if they failed any of our attention checks, which included comprehension questions about the vignette. The final sample (N=320, mean age = 39.3 years) included 172 male and 146 female participants (2 did not indicate their sex).

### 1.2.2. Materials

We used the vignette described in Table S2.

### 1.2.3. Procedure

Similar to Study 1, after reading the randomly assigned vignette, participants were asked to rate the responsibility of each of the 5 actors on a 7-point scale, to distribute the money to the 4 human actors, and to answer the anthropomorphic battery.

We then used principal component analysis to collapse these 4 measures into a single measure of anthropomorphicity  $A$  where

$$A = 0.616 * Q_1 + 0.389 * Q_2 + 0.495 * Q_3 + 0.472 * Q_4$$

(where this first principle component explains 90.23% of the total variance).

## References

Waytz, A., Heafner, J., Epley, N., 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52, 113–117.

Table S2: Vignettes used for Study 2. Related to Figures 4, 5, and 6.

AI as Tool Condition

**Thousands of people from all over the world** upload images to crowdimage.net, a image-hosting website. These people know that artists will look at and use their images to make art.

**Timmy is a technologist** who creates an image manipulation software for people to use to make art. **The software is called ImageBrush.** The software is a tool that humans use to make art. The artist plans and envisions the artwork, and the software executes simple commands based on what the artist tells it to do.

**Alice is an artist** who uses ImageBrush to create particular kinds of images. Alice takes an existing image of a scene from the news (such as a beach or a forest) and adds a ghost to it using ImageBrush. This is how Alice decides to make the ghost: she goes to crowdimage.net and takes at all the images of people that have been uploaded to the platform. Then, She creates a composite of the people using ImageBrush. This makes a ghost-like figure, which Alice then puts into the scene.

**Casey is a curator** who is Alice’s collaborator. Casey goes through many of the images that Alice created and selects the following artwork because Casey really likes it. Casey then brings it to an art auction, where it ends up being sold.

AI as Agent Condition

**Thousands of people from all over the world** upload images to crowdimage.net, a image-hosting website. These people know that artists will look at and use their images to make art.

**Timmy is a technologist** who creates an image manipulation software for people to use to make art. **The software is called SARA.** SARA is a deep neural network that creatively plans and envisions new artworks, with minor help from an artist collaborator.

**Alice is an artist** who collaborates with SARA to create particular kinds of images. SARA takes an existing image of a scene from the news (such as a beach or a forest) and adds a ghost to it. This is how SARA decides to make the ghost: it goes to crowdimage.net and takes at all the images of people that have been uploaded to the platform. Then, it creates a composite of the people. This makes a ghost-like figure, which SARA then puts into the scene.

**Casey is a curator** who is SARA’s collaborator. Casey goes through many of the images that SARA created and selects the following artwork because Casey really likes it. Casey then brings it to an art auction, where it ends up being sold.

---

Negative valence outcome

The artwork sold at the art auction has come under scrutiny because it was shown to violate copyright law. The court ruled that the sale of the art must be nullified, meaning that the money will be returned to the buyer. In addition, the courts have issued a \$400,000 fine as a penalty for the copyright violation.

Positive valence outcome

The artwork shown before sold for \$400,000 at the prestigious auction house. This was the largest dollar amount paid for a artwork of this kind ever, and made lots of headlines.