Seeing Beyond the Blur with Generative AI

Can AI hallucinations be responsibly harnessed for scientific imaging?

By Berthy Feng and Katherine L. Bouman

DOI: 10.1145/3703400

OPEN ACCESS

As artificial intelligence (AI) grows in generative capability, questions remain about how to leverage it for science. AI for science promises to accelerate scientific discovery and technological progress, in turn addressing pressing challenges in areas like healthcare and sustainability. In our quest for greater scientific knowledge in the service of such endeavors, seeing is believing. Whether it's viruses, weather systems, or exoplanets, making the unseen visible helps inform, motivate, and educate. As computational imaging researchers, we aim to visualize scientific phenomena beyond the reach of conventional cameras. Generative AI is an exciting tool for such visualizations.

Given that it can concoct scenes like corgis traveling to space, perhaps it can help us image "invisible" objects that have eluded scientists for decades.

One such object, which until recently has existed only in theoretical papers and in our collective imagination, is a black hole. In April 2017, the Event Horizon Telescope (EHT) pointed its antennas at a supermassive black hole known as M87* [1]. Residing 55 million light-years away in the galaxy Messier 87, M87* evades the view of any typical telescope. To get around this, researches built a new type of telescope consisting not of a single receiver but of disparate antennas spread across the globe. Each antenna could detect radio waves being emitted by the plasma swirling around the black hole. The data collected by all the antennas could then be synchronized and processed to appear as if they came from a single earth-sized telescope. However, even this virtual telescope could not perfectly capture the black hole.

An earth-sized telescope is still not big enough to see all the tiny details of the material surrounding the black hole. Additionally, instrument noise and atmospheric turbulence interfere with the radio signal reaching the antennas. These complications lead to the key computational challenge of imaging a black hole: creating an interpretable image from a small and garbled set of data.

This brings us back to AI. The holes in the data create room for imagination. When the first image of M87* was released in 2019, AI was kept out of

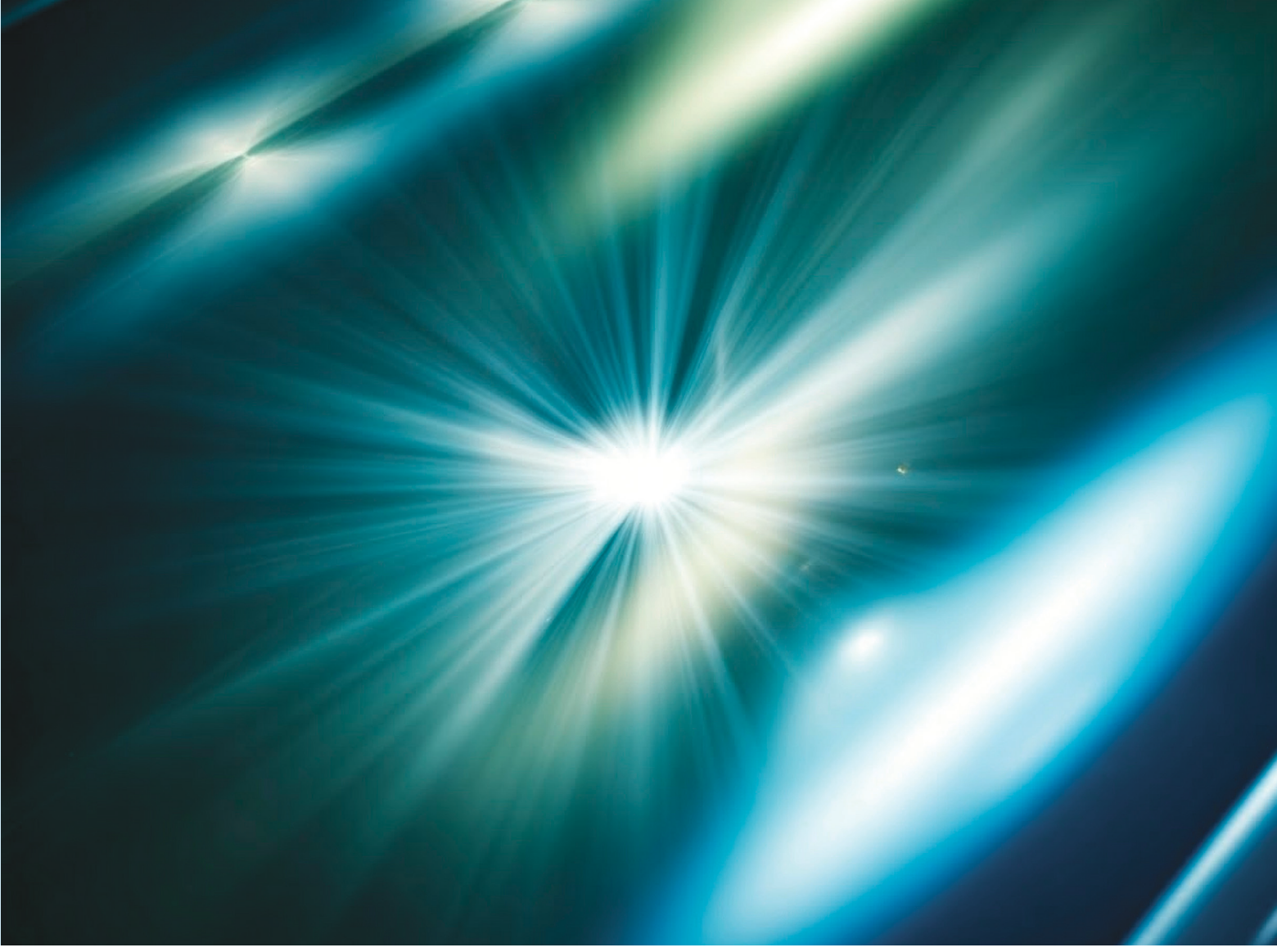


Image by Andrij Borys Associates / Shutterstock AI

the picture. The EHT's imaging team incorporated only the most basic image assumptions—like light being non-negative and changing gradually across the image—to transform its raw antenna data into the now-famous photo. The wariness of AI is understandable. For such a momentous image, any bias that might misrepresent the data would raise eyebrows.

But the extra caution led to an image that many of us can't help but see as blurry. Cognitive dissonance between the awe of the scientific feat and the urge to squint at the photo characterized the public reaction to the first M87* image. Despite it being one of the highest-resolution photos ever captured, people expected to see more.

To see past the blur inevitably calls for some amount of hallucination. Generative AI excels at devising novel images but transferring generative AI to scientific imaging demands guardrails to keep its hallucinations in check. Our work centers on how to use hallucinations to our advantage in a principled way to image the invisible without misrepresenting the factual data. With collaborators we have developed computational methods for bringing in different image assumptions to supplement observed data. An exciting application of these methods is to re-imagine the M87* black hole image under different assumptions and assess which visual features withstand bias. Whereas previous hand-designed imaging pipelines made it difficult to disentangle hallucinations and reality, an AI-based approach lets us probe different sets of assumptions to determine the consistent, and therefore trustworthy, features in the image.

IMAGING AS AN INVERSE PROBLEM

Taking a note from high-school math, the first step to solving this imaging problem is to translate it into an algebraic equation with knowns and unknowns. The task of inferring the unknown from the known is called an inverse problem. For us, the known is the data collected by the EHT's antennas, and the unknown is the image of the black hole. There are myriad other imaging tasks that are inverse problems. If you've ever gotten an MRI scan, the doctor might have shown you the output of an inverse problem. An MRI machine collects data (the known) that help constrain an image of your internal tissue (the unknown). If you've ever tried to snap a picture of someone who doesn't keep still, you've probably been annoyed to keep getting a blurry picture (the known). Restoring the person to their crisp self (the unknown) is called "motion deblurring," an inverse problem that your smartphone camera can now solve for you.

But unlike in high-school algebra, in inverse problems, the knowns are not sufficient for solving for the un-

knowns. To understand why, consider the hypothetical inverse problem of restoring an image with half its pixels missing. As long as you keep the known pixels fixed, there are infinitely many solutions for the unknown pixels (see Figure 1 for an example). In our case, infinitely many arrangements of light around M87* could have led to the data that the EHT observed. To obtain any reasonable solution, we must supplement the known data with additional assumptions about the unknown image.

Traditional computational imaging algorithms enforce assumptions through something known as a "regularizer." Image regularizers are formulas that evaluate how likely an image is according to our assumptions, and they are typically hand-designed. A common assumption is images tend to be smooth, meaning brightness does not change too much from pixel to pixel. Naturally, a regularizer for this assumption would penalize large changes between pixels—an image of pure white noise, for instance, would fare poorly under this regularizer because every one of its pixels is completely different from its neighbors. The EHT used such regularizers to infer the first M87* image, but the range of assumptions for which we can easily hand-craft regularizers is limited.

The last several years have witnessed a paradigm shift from traditional regularizers to generative AI models for solving inverse problems. While generative AI is known for creating novel, flashy pictures, we now know it also offers a way to impose more sophisticated assumptions than previously possible with regularizers. Moreover, we can target different assumptions simply by training the generative model on different image datasets. With the help of generative AI, we can reconstruct images from the same data under a variety of assumptions.

Diffusion models are particularly powerful generative AI models [2, 3]. They form the backbone of much of today's most popular generative AI technology, including Stable Diffusion, DALL-E, and Midjourney. Once trained on a set of example images, a diffusion model can start from a random draw of noise to generate images never seen before. What's more, the generated images are convincing, resembling what we might expect to see as clean training images. This hints that it has somehow learned our beliefs about what types of images are likely and what types are unlikely. By feeding the diffusion model exemplary images during training, we can teach the model our assumptions about what a desired image looks like. And by changing the training images, we can target different image assumptions.

Figure 1. As a hypothetical inverse problem, suppose you're given an image with the center portion missing. There is no way to perfectly recover the full photo from the partial data. These three possible images are just some (of infinitely many) examples of how we could fill in the missing pixels. Assuming the image is a photograph of a person, the example on the far right is highly unlikely. If you could make additional assumptions, such as the photo being of a boy, then you could further narrow down the set of possible images. Images courtesy of Bingliang Zhang.

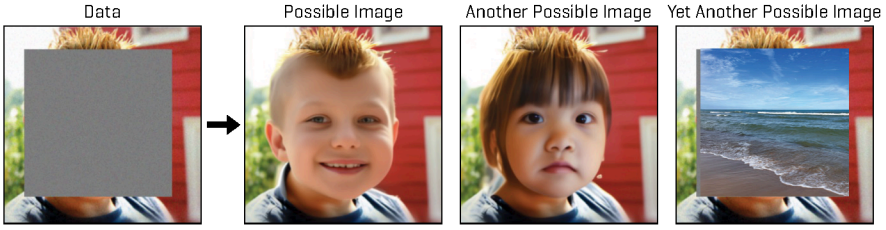
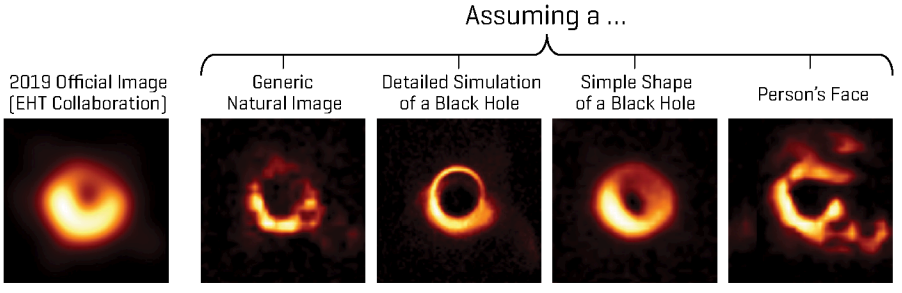


Figure 2. To see beyond the blur of the original 2019 image of M87* requires stronger biases to fill in the missing data. With the help of generative AI, we can see what the image looks like under a variety of assumptions. These re-imagined images of M87*, based on EHT data from April 5, 2017, all agree on the existence of a ring structure that is brighter on the bottom, although each portraying it in their own style. (Note: The displayed images are centered for the sake of comparison, although the data do not actually constrain the ring to lie in the center of the image.) Image courtesy of EHT Collaboration [1].



RE-IMAGINING M87*

Along with collaborators, we developed a technique to turn a diffusion model into a sophisticated regularizer for solving inverse problems in imaging, and we used this technique to explore the effects of different image assumptions on the visualization of M87* [4, 5]. We started by training diffusion models on different image datasets. One was trained on generic natural images. Another was trained on images of detailed simulations of black holes. Another was trained on images from a simpler shape-based model of black holes. We even trained one on pictures of celebrity faces to see how the EHT data might get visualized under such an absurd assumption. We then paired each of these diffusion models with the observations of M87* that the EHT gathered in 2017. Using our algorithm, we obtained different re-imaginings of M87* assuming different visual statistics (see Figure 2) [6].

No matter the assumptions im-

posed by the diffusion model, the images all displayed a ring structure of the same size that was brighter on the bottom. Even assuming a celebrity face did not get in the way of imaging a ring from the data, the diffusion model that had only ever seen pictures of people's faces still managed to craft a ring by removing half the face and an eye, leaving an ominous Phantom of the Opera-esque mask.

In addition to the ring structure, the diameter of the ring and location of its bright spot were consistent across assumptions. The rest of the image was up to the diffusion model's interpretation. The diffusion model trained on simulated black holes gave us a thin ring with gas swirling around the shadow of the black hole. In contrast, the assumption of a simple geometric model of the black hole offered less visual detail, showing only the shape of a crescent. Such idiosyncratic hallucinations—the patchiness from the assumption of a generic natural image, the dynamic wisps from the assumption of a detailed black-hole simulation, and the nose from the assumption of a face—should not be trusted as real features of M87*. On the other hand, we can rely on the assumption-independent characteristics of our images, namely the appearance of a ring with most of its brightness on the bottom.

WILL THE REAL M87* PLEASE STAND UP?

You might be wondering which of these images most accurately portrays M87*. It is impossible to know. Surely none of them depicts M87* exactly, but viewed together they convey a wealth of information. The multitude of possible images might defy your expectation of one "real" image, but in fact most images you see are not real. You're probably more comfortable with hallucinations in your everyday digital life than you realize. Most digital cameras use an RGB color filter that captures just one primary color in each pixel, meaning the colors in your digital photos are two-thirds made-up. We accept the hallucinated colors because they follow unobjectionable assumptions. The same can be true of AI hallucinations as long as we accept the assumptions of the AI model.

Our work centers on how to use hallucinations to our advantage in a principled way to image the invisible without misrepresenting the factual data.

Much of the fear around AI stems from a fear of unchecked hallucinations, which could proliferate false information to the detriment of science, society, and politics. Our research efforts have shown it is possible to wield AI responsibly. We can build trust in AI hallucinations by thoroughly testing their assumptions to determine which image features are invariant and which are sensitive to bias, helping us rule out any false representations of reality. We can accept assumption-dependent hallucinations as real only if we agree with the assumptions.

Hallucination is unavoidable as we push the limits of imaging. Rather than fear hallucinations, we should design methods to apply it responsibly, paving the way to a future in which we are not held back by the physical limits of pure imaging devices. For example, the EHT is currently aiming for a video of Sgr A*, the black hole at the center of our galaxy, whose surrounding plasma changes drastically within a matter of minutes [7]. Since today's antennas are incapable of capturing such faraway and fast dynamics, hallucination will be a necessary ingredient in the process of reconstructing videos. Shifting our focus from the skies to the lab bench, researchers are looking to image miniscule objects, such as proteins, using cryo-electron microscopy. Capturing high-resolution 3D molecular structures will benefit from AI to help fill in gaps in the data obtainable by electron microscopes [8]. In all these cases, although hallucination will play a vital role, it should not obscure the truth. We're excited to discover what more we can see as we continue to leverage the growing power of generative AI responsibly.

References

[1] The Event Horizon Telescope Collaboration et al. First M87 Event Horizon Telescope results. IV. Imaging the central supermassive black hole. *The Astrophysical Journal Letters* 875, 1 (2019), L4.

[2] Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (NIPS '20). Curran Associates Inc., Red Hook, NY, 2020, 6840–6851.

[3] Song, Y. et al. Score-based generative modeling through stochastic differential equations. In *Proceedings of the Ninth International Conference on Learning Representations* (ICLR '21). 2021.

[4] Feng, B. T. and Bouman, K. L. Variational Bayesian imaging with an efficient surrogate score-based prior. *Transactions on Machine Learning Research* (2024).

[5] Feng, B. T. et al. Score-based diffusion models as principled priors for inverse imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (ICCV '23). IEEE, 2023, 10520–10531.

[6] Feng, B. T., Bouman, K. L., and Freeman, W. T. Event-horizon-scale imaging of M87* under different assumptions via deep generative image priors. *The Astrophysical Journal*. 2024 (in press).

[7] Johnson, M. D.et al. Key science goals for the next-generation Event Horizon Telescope. *Galaxies* 11, 3 (2023), 61.

[8] Zhong, E. D. et al. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* 18, 2 (2021), 176–185.

Biographies

Berthy Feng is a Ph.D. candidate in the Computing and Mathematical Sciences Department at the California Institute of Technology. She received her bachelor's degree in computer science, summa cum laude, with a certificate in statistics and machine learning from Princeton University. She has received the Kortshak Scholarship and NSF GRFP Fellowship. She primarily works in computational imaging, computer vision, and machine learning. Her research interests include developing methods to incorporate deep learning and physics knowledge to solve inverse problems in imaging.

Katherine L. (Katie) Bouman is an associate professor in the Computing and Mathematical Sciences, Electrical Engineering, and Astronomy Departments at the California Institute of Technology. Her work combines ideas from signal processing, computer vision, machine learning, and physics to find and exploit hidden signals for scientific discovery. Before joining Caltech, she was a postdoctoral fellow in the Harvard-Smithsonian Center for Astrophysics. She received her Ph.D. in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT in EECS, and her bachelor's degree in electrical engineering from the University of Michigan. She is a Rosenberg Scholar, Heritage Medical Research Institute Investigator, recipient of the Royal Photographic Society Progress Medal, Electronic Imaging Scientist of the Year Award, Sloan Fellowship, University of Michigan Outstanding Recent Alumni Award, and co-recipient of the Breakthrough Prize in Fundamental Physics. As part of the Event Horizon Telescope Collaboration, she co-led the Imaging Working Group and acted as coordinator for papers concerning the first imaging of the M87* and Sagittarius A* black holes.

Copyright is held by the owner/author(s).
Publication rights licensed to ACM.

1528-4972/24/12 $15.00