# Visualizing Algorithms: Mistakes, Bias, Interpretability

**Catherine Griffiths**
University of Southern California, School of Cinematic Arts
griffitc@usc.edu

## Abstract

This design research project addresses the domain of obfuscation and ethical bias at the heart of machine learning algorithms. By opening the algorithmic black box to visualize and think through the meaning created by algorithmic structure and process, this project seeks to provide access to and elucidate the complexity and obfuscation at the heart of artificial intelligence systems.

The questions being addressed include: Can tactics from the visual arts and digital humanities, including interaction design, generative design, and critical code studies, combine as an effective method to visualize ethical positions in algorithms, including bias, mistakes, and interpretability? How can visualization of algorithms be used as an a-linguistic tool to re-engage with decision-making in prediction systems, where humans are at risk of being precluded? When considering bias augmentation, what can be learnt by temporarily isolating the meaning in data, to focus on the effect that structure and process play in the generation of bias?

The work-in-progress prototype software visualizes a machine learning algorithm, a decision tree classifier. It simulates data flowing through the algorithm and predictions being made in real time. It is built procedurally as an interactive tool, so that any classifier of the same type can be loaded and visualized. The UI provides parameters to support the self-organization of the classifier structurally and to aid analysis. The loaded examples present different topologies of classifier based on machine learning data sets with different feature to class ratios. The prototype can currently visualize mistakes in prediction, where the algorithm misclassifies data. It can also reverse engineer each data point's path to visualize where in the algorithm an error was made. The most popular paths taken through the algorithm's complex network of decisions are also visualized.

The project is conceived using a conceptual approach to machine learning, to experiment with how aesthetics and design can be used as tactics for engagement with complexity. Tactics include: a move away from data visualization toward computational visualization to focus on real-time and even projected rule sets, rather than a retrospective and fixed approach to data. Adapted insights from programming games and animation are used to present both human-scale and emergent processing speeds, the flow of data through an algorithm, and how decisions are made in real-time.

The research is working toward the use of visual arts tactics as a means of "ethical debugging", in which complex terms, such as bias and interpretability can be presented visually, and algorithms can be engaged with aesthetically as socio-political systems. [1] As the research continues to develop, more speculative design avenues will be explored, alongside technical problems. The project so far has concentrated on developing a more robust visualization of a machine learning algorithm to engage and collaborate with computer scientists working in this field. As the research develops, the intention is to

develop further scenes of this application that navigate more strongly, even contentiously, back toward the visual arts, to explore the potential for “novel models of relationality and connectivity.” [2] An overarching question asks, how artistic knowledge can contribute to the issues of the day, generating new ideas, proposals, and methods, using aesthetics as the primary paradigm of knowledge generation, without solely assimilating with traditional scientific methods.

## References

1. Catherine Griffiths, “Visual Tactics Toward an Ethical Debugging,” *Digital Culture & Society: Rethinking AI*, 4, no. 1 (2018): 217.
2. Simon O’Sullivan, “Inquiry,” in *NJP Reader 1: Contributions to an Artistic Anthropology*, ed. Youngchul Lee and Henk Slager (Seoul: Nam June Paik Art Center, 2010), 52.

## Biography

Catherine Griffiths is a PhD candidate in Interdisciplinary Media Arts + Practice at the University of Southern California, School of Cinematic Arts. She researches at the intersection of visual art, computation and critical studies, focusing on the visualization of algorithms, in the context of machine learning and the ethics of algorithms debate. She has a bachelor's degree in Fine Art from the University of the Arts, London, and a master’s degree in Architecture from University College, London.