xCoAx 2019 Conference on Computation, Communication, Aesthetics & X Milan, Italy

Catherine Griffiths
griffitc@usc.edu

University of Southern California

Approaches to visualise and critique algorithms for ethical scrutiny

Keywords
visualization
algorithms
ethics
machine learning
computer vision
surveillance
decision-making

This arts research project addresses the domain of obfuscation and ethics in algorithms, including computer vision and machine learning systems. The work presents a series of simulations as visual-critical arguments, proposed as methods to open the algorithmic black box to visualize and think through the meaning created by algorithmic structure and process deployed in ethically sensitive spaces. The project seeks to provide access to and elucidate the abstraction and obfuscation at the heart of algorithmic systems.

INTRODUCTION

A series of visual-critical simulations aims to bring arts knowledge to bear on the intersection of computation and ethics, using critical theory and creative coding to affirm computation as a site of the social and political. The paper presents two different approaches to visualising and critiquing computational ideas and algorithms to think through questions about aesthetics and social issues.

CELLULAR AUTOMATA, SURVEILLANCE, AND CREATIVE CODE

The first method presents an approach to reverse engineer a social issue, in this case, surveillance, back through a particular algorithm, or core computational concept, in this case, a cellular automaton. There is a line, conceptually and visually, to be drawn between the core computational logic of cellular automata, via image processing techniques, through computer vision algorithms, and into the gaze of a street surveillance camera. The approach seeks to make this argument visually, through a series of simulations. Contextually, this research begins by looking at open source algorithms and libraries and thinking through the social and political implications of them, addressing algorithms, not just as cultural artifacts but at the level of code syntax. Connecting to the practice of critical code studies, which looks at source code as also being a cultural text with the same potential for humanistic interpretation as other cultural texts, the work uses the OpenCV library.

There is an affinity between cellular automata and images through the computational grid system of cellular automata and the pixel array structure of digital images. A cellular automaton is a system of simple rules and states, operating on grids of cells, and from such seeming simplicity, complex behaviours emerge, leading to further-reaching possibilities. State is usually represented by black and white coloured cells, which are often interpreted as alive and dead, whilst a typical rule set might be: if a live cell has less than two live neighbours, then it dies (interpreted as isolation); if a live cell has more than three live neighbours, then it dies (interpreted as overcrowding); if a dead cell has three live neighbours, then it comes alive (interpreted as reproduction); otherwise a cell stays the same (interpreted as stasis). From such a seemingly simple computational system, far-reaching speculations have been developed in relation to artificial life and the computational universe.

A Langton's ant is a version of a cellular automaton in which only one cell in the grid changes at a time, so it functions as an autonomous agent. This agent was applied to a satellite image of a location in the Amazon known as the 'Meeting of the Waters', which is the confluence of two rivers, the darker coloured water of the Rio Negro and the sandy coloured water of the Amazon River. Due to each river's different water density, speed, and temperature, their waters do not mix for several kilometers and instead run alongside each other inside the same river channel, demarcated by their different colours. Several hundred Langton's ants were deployed across the structure of the image, using its data structure to compute across, generatively repatterning it, and transforming the landscape and the composition of the river. The choice of image works analogously, where one's understanding of the landscape is terraformed by the agents. This visualization is presented in two formats, one which foregrounds the algorithm's interpretation of the scene, as a simplified four-state grayscale image that the agents use to compute on to determine their state and change pathway. Another image foregrounds the human view, as the effects of the generative redesign of the landscape. The work is presented in this way to think through the difference between the simplified data and logic that the algorithm operates with, and the higher-level image that we see, and which might hold cultural or social meaning.

Continuing this mode of visual-critical argument to connect the logic of cellular automata computations to our social understanding of surveillance, the research engaged with image processing techniques, which are an important part of a computer vision library of algorithms. Images need to be heavily processed, broken down and simplified to be interpretable by an algorithm. Popular filters such as blur, sharpen, and edge detection are used and operate with similar logic to a cellular automaton. When background subtraction is applied to an image from a surveillance camera, the image is reduced in complexity to just two states and two rules, if a pixel's RGB value changes between video frames it is assigned white, and if it remains the same between frames it is assigned black. In this way, an algorithm reads motion in a video image, and the result is a rather sinister image of the surveillance camera's gaze, tracking people walking in urban space. The research works with a creative coding approach to creating a series of visualizations of the algorithm in action, first of all isolating motion in the image, and then printing only that motion. The work uses the image of a chameleon, because of the nature of the animal to conceal itself through stillness. By analogy, the chameleon reveals itself to the algorithm through movement and camouflages itself from the algorithm through stillness.

This arts research seeks to move forward from the tradition of data visualization, to experiment with ways of visualizing computational process or models, to open the black-box of algorithms that are used in socially contentious spaces and think through their inner workings by means of visual-critical arguments. From a computer science perspective, cellular automata systems are understood as expressions of foundational computational concepts including state machines and formal logic, they are also understood as neutral mathematical concepts, however, from an arts research perspective, the very foundations of computation and code can be questioned and contextualized within a social context.

MACHINE LEARNING, ETHICS, AND INTERACTION DESIGN

A second approach to building visual-critical arguments to address the ethics of algorithms has also been explored. Machine learning algorithms were investigated because of their emerging use in ethically sensitive spaces such as policing and welfare. The incidents of algorithms arriving at racist or sexist classifications or being used to determine who goes to prison and who receives leniency, have received important attention over the last few years. The ethical dilemmas that are arising from the use of machine learning algorithms include the likelihood of them generating mistakes and of augmenting biases hidden in data. The investigative journalism organization, ProPublica, investigated machine bias in the US justice system in 2016, pointing to how predictive systems can encode racial bias when used in criminal sentencing, and it was from there that this research began (Angwin, 2016).

However, initially on looking into and working with machine learning, another related phenomenon captivated the direction of the research: the new emergent type of computation that has come to the fore through the rise in machine learning practices, specifically deep learning. Through machine learning, computation has shifted from a system of pre-programmed rules that are executed iteratively, into a form of generative code in which an initial algorithm, written by a human, writes its own algorithm, from which humans are precluded from understanding its logic. This has been referred to as algorithms operating in the wild.

At face value, there is something fascinating and seductive about this new computational paradigm. However, it presents an ethical issue known as the interpretability problem, in which an increase in accuracy creates a simultaneous decrease in human readability. The sensitivity around the use of algorithms to make potentially life-altering decisions is exacerbated by “AI’s Unspoken Problem”, described by Will Knight as being that an algorithm cannot tell us why it made the decision it did, it can only present it’s predicted answer to a given question. The algorithms that are currently in use, do not have the quality of common sense or awareness of context incorporated into their models, and the need to ask an AI ‘why’ and receive an explanation is necessary for us to work in collaboration with them. (Knight, 2016) From an ethical vantage point, the idea that we deploy a system into an ethically sensitive space and cannot say how it works precisely, or how it arrives at a particular decision, requires a level of trust that has not been earned by such flawed systems. There is no right of appeal, of disputing the outcome of an algorithm, or asking why the algorithm arrived at a particular decision. When that decision is to recommend a person be jailed, or fired from a job, or refused a place in a state-funded drug rehabilitation program, it becomes a significant ethical problem.

From an interdisciplinary design research perspective, a framework to critically study algorithms needs to provide access to algorithms for observation, to promote literacy, enable reflection, and formulate a critical and ethical position in the discourse. An interactive visualization tool was developed to visualize a simple machine learning algorithm, a decision tree classifier, to think through some of these ideas and pose further questions. Classifiers were generated using the scikit-learn library and then rebuilt in Unity, a game engine, to drive an interactive visualization in real-time. way to temporarily isolate the meaning in data, to think about the meaning of structure and process in the algorithm instead. From a design perspective, a combination of tactics from interaction design, generative design, and to some extent critical code studies, have been employed. A decision tree classifier was used because it is one of the simplest types of machine learning that is already somewhat graphic, and whilst it is a form of machine learning, it should be noted that it is not deep learning, which is where some, but not all of the controversy lies.

The design tactics employed begin by mapping out the algorithm spatially, to look at its possibility space, at all of the various paths through the algorithm, and decisions that are made before arriving at a prediction. Then data is simulated through the algorithm, showing decisions being made in real time as the algorithm executes. The simulation of time is a tactic taken from some computer games, in which time can be scaled to see individual decisions being made at a slower, human scale of perception, through to a higher, emergent scale in which patterns of decisions can be seen forming. At this point, the visualization can point to mistakes in prediction, where the algorithm mis-classifies data. A user can also hover over each data point and reverse engineer the path it took through the algorithm, perhaps to see at which point it made a wrong decision and took a wrong path. The system also visualizes particular features of the data, through the physical proportions between the data points. The most popular and least popular pathways through the algorithm's network are also visualized. The prototype was built procedurally so that any classifier of the same type can be loaded and visualized, with the user interface supporting its structural self-organization, and aiding analysis.

In developing an interactive design tool such as this, the questions that come up include: to what extent visualization is an a-linguistic tool to re-engage with decision-making in prediction systems and provoke questions, where we are at risk of losing our connection to decision-making? Could visual tools be used by key workers in the field, who are expected to work with the results of these algorithms but so far are precluded for understanding their logic? To what extent interaction design, generative design, and critical code studies combine as an effective method to visualize an ethical position in algorithms? What does it mean to learn, in machine learning, and is the anthropomorphism of AI a productive analogy? The tool uses synthetic data, therefore artificially removing the social meaning from the data, and whilst the research is primarily motivated by that, it is in the hope to explore the concept of bias augmentation, which speculates that where there is a small bias in a dataset, this can become amplified through the iterative algorithmic process. Where most people today argue that bias is in the data, because the data is a reflection of bias in society, there is also speculation that the algorithm in its structure and process, can play its own role to augment bias. That is something to explore further, hence the focus on structure and process over data so far.

COMPUTATION IS POLITICAL

In his book *Ethical Programs*, James J. Brown Jr conceives of computational networks as swarms: entities without a face, without a front or center, dispersed, in constant communication, and able to attack from all directions. Unclear when it is an ally or adversary, Brown asks how we can conceive of the other that has no face, that resists representation and understanding, that is always here and yet we can't make sense of it. "How do we deal with an other that has no face, what is an ethics in the face of the swarm? (...) What rhetorical actions are possible?" (Brown, p.4) Brown proposes that new ethical programs need to be developed to address the ethics of existing inside the swarm, an ethics that likewise needs to be brought to bear on machine learning algorithms, which we should consider to be hostile, faceless, and permitting their own agency. Our understanding of trust, permission, and accountability has not been updated in line with an understanding of this new type of computation.

Steven Shaviro likewise critiques computation in the context of complex adaptive systems for masking the essential process of decision-making at its core. When decisions in systems are hidden, causality is dissolved, and accountability cannot be traced. For Shaviro, an uninformed approach to complexity and computation is often used as an alibi for an unaccountable subject that refuses participation and political engagement. Shaviro proposes to foreground "an aesthetic of decision, instead of our current metaphysics of emergence," (2009) to disentangle accountability in computational systems.

We can also look to the field of encryption, for further affirmation on why access, interpretability, and critique is an ethical argument for why we need to read the code. In Bruce Schneier's book *Secrets and Lies: Digital Security in a Networked World*, he presents the case for why it is important that encryption algorithms are in the public domain. In cryptography, when an algorithm's code is open source, it means that a significant number of people will have studied the algorithm and identified flaws, weaknesses, and possible hacks, which lead to continual updates and a stronger algorithm literally meaning that the more eyes that have looked at an encryption algorithm, the more secure it will be. Whereas, when algorithms are black boxes and protected by intellectual property, the code has not been scrutinized by many people, leaving it open to hidden flaws. (Schneier, 2000) This is an attitude we can take into the field of artificial intelligence, to argue for policy to ensure that in ethically contentious spaces, algorithms need to be available to as much public scrutiny as possible to ensure ethical robustness.

In his book *Virtual Migration*, A. Aneesh puts forward the concept of "algocracy - rule of the algorithm, or rule of code" (2006), which is described as a new kind of power that is created through the way that algorithms are embedded in software. It is a system which does not require monitoring through traditional surveillance systems, hierarchies, and forms of government, but instead governance and surveillance take place through the design of the algorithm and the ways it tacitly shapes behaviors and asserts authority, without public awareness. If we accept the role algorithms play, that their design can be a form of tacit governance, what would it mean to design an algorithm more comprehensively and consciously, with the scope of critical thinking that comes from the arts and humanities and social sciences? What if algorithms were not written solely by computer scientists, but designed by such an interdisciplinary team?

This research questions how aesthetic language and critical thinking from the visual arts can be brought to bear on algorithms and society? The research proposes to move toward a perspective that positions code as a political language. In the creative coding community, emerging from software studies, we are told that code is now a comprehensive language for creative and authorial expression. Cannot code also be a language of critique to probe its own social and political latencies?


Fig. 1. and 2.
Langton's Ant simulation (left);
Automata I, algorithm interpretation (right).


Fig. 3. and 4.
Automata I, human interpretation (left);
OpenCV library, background subtraction (right).


Fig. 5. and 6.
Automata II, detecting motion (left);
Automata II, drawing motion (right).

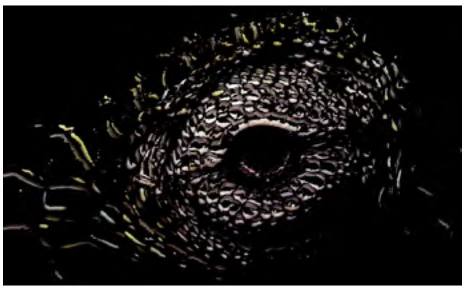
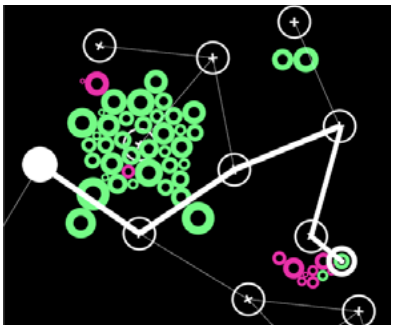



Fig. 7. and 8.
Automata II, detecting motion, close up (left);
Automata II, drawing motion, close up (right).



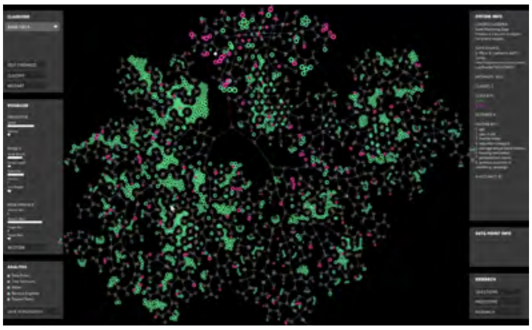


Fig. 9. and 10.
Interactive visualization tool for a simple machine learning algorithm (left); Detail of the machine learning visualization tool showing mistakes in the algorithm (right).

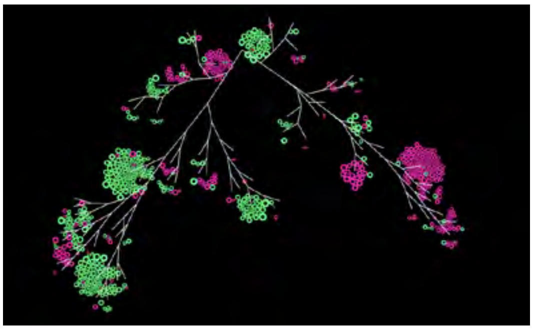


Fig. 11.
Detail of the machine learning visualization tool

REFERENCES

Aneesh, A.
Virtual Migration: The Programming of Globalization, Duke University Press, 2006

Angwin, Julia. et al.
Machine Bias. ProPublica, May 23rd 2016.

Brown Jr, James J.
Ethical Programs: Hospitality and t he Rhetorics of Software, University of Michigan Press, 2015

Knight, Will.
AI's Unspoken Problem, in MIT Technology Review, Vol 119, No. 5, 2016

Shaviro, Steven.
Against Self-Organization, blog post The Pinocchio Theory, May 26th, 2009
http://www.shaviro.com/Blog/?p=756

Schneier, Bruce.
Secrets and Lies: Digital Security in a Networked World, John Wiley & Sons, 2000