HELEN KNOWLES

The Trial of Superdebthunterbot

Installation view (2019) at Kunstverein Hannover of *The Trial of Superdebthunterbot* (2016) by Helen Knowles; photograph by Raimund Zakowski.

Can an algorithm be found guilty in a court of law? What is the responsibility of a nonhuman entity? How might artistic practice contribute to the dialogue around this philosophical question? How might it contribute to the ethics of algorithmic governance? These were some of the questions I was exploring in May 2015, when I staged the first performance of *The Trial of Superdebthunterbot* in a small gallery in Oriel Sycharth, in Wrexham, Wales, for the *Collaborate!* exhibition, curated by Ivan Liotchev and Nicolas John Jones.

The artwork took the form of a scripted performance of a trial of an algorithm conducted by a judge, played by the actor Mark Frost and two real-life lawyers, Laurie Elks and Oana Labontu Radu. The lawyers wrote their own prosecution and defense speeches. In this first performance, the jury was comprised of the attendees to the opening night of the exhibition. There were minimal props, the main one being the algorithm, which was fallaciously housed in an exquisitely crafted, hand-built, see-through

computer made by the artist Daniel Dressel. The computer sat blinking sedately in the dock. I chose to dramatize the intangible algorithmic code, opting to manifest its physicality in the chips and wiring, graphics cards and LED lights that comprise computational hardware. The exterior casing of this computer was made from Perspex and enabled the jury (or the audience watching the performance) to peer in. However, since code is not physical but relies on the material infrastructure of engineered components, not just in the computer itself but across the networked world, it felt like an appropriate interplay between our imaginative understanding of these intangible entities and their existence in our material world. Palpating the line between fact and fiction, between our imagination and reality, was the axis around which *The Trial of Superdebthunterbot* was built.

In 2015, I read an article in the *London Review of Books* by Andrew McGettigan that illustrated the misconception around the actual monetary worth of the student loan book, a stock of student loans owed to the British government.[1] The article revealed how it had been oversold since most people were unable to pay back the debt. The fiction I concocted centered on an algorithm that is programmed by a debt-collecting company, Debt BB, to target loan defaulters—in this case vulnerable students who were defaulting on their student loans. Debt BB buys the student loan book for more than it is worth, on one condition: it can code an algorithm to use big data to scour the internet and place job advertisements on Facebook pages visited by student loan defaulters. The advertisements are for lucrative yet potentially dodgy employment such as sex work or medical trials. In *Superdebthunterbot*, we come to understand that the algorithm can learn on the job but that it has little moral capacity. Deaths result and the algorithm is tried for gross negligent manslaughter.

The artwork also took inspiration from Susan Schuppli's seminal essay, "Deadly Algorithms: Can legal codes hold software accountable for code that kills?," which asks what "existing legal frameworks are *capable* of attending to the emergence of these new algorithmic actors?"[2] In 2015, as details about Cambridge Analytica were emerging alongside a nascent awareness of the implications of big data, I began asking what kind of legal framework might exist for prosecuting a nonhuman entity. I did this by engaging in informal conversations with the lawyers, who collaborated with me to mold the story and make it feasible for a real trial. I then wrote the judge's speech, which was partially ripped from countless court transcripts available on YouTube, in order to have a framework for the judge to deliver an opening and closing speech. This was only after the lawyers had taken the floor with their closing arguments, which they had carefully crafted so they could deliver them with veracity. In essence, the trial is a shorthand of what might have been a much longer and drawn-out affair. This then enabled the jury to reach a verdict: guilty or not?

1. Andrew McGettigan, "Cash Today," *London Review of Books* 37, no. 5 (March 5, 2015), www.lrb.co.uk/the-paper/v37/n05/andrew-mcgettigan/cash-today.

2. Susan Schuppli, "Deadly Algorithms: Can legal codes hold software accountable for code that kills?," *Radical Philosophy* 187 (September/October 2014), www.radicalphilosophy.com/commentary/deadly-algorithms.

Installation view (2016) at Goldsmiths, University of London of *The Trial of Superdebthunterbot* (2016) by Helen Knowles.

In 2016, I staged the second performance of *The Trial of Superdebthunterbot* at Southwark Crown Court in London and shot the forty-five-minute film of the same title (which was released the same year). The jury was comprised of a variety of individuals from all walks of life whom I had invited to take part. In order to produce an interesting array of jury responses (the deliberations of the jury were never scripted), I actually staged the work in Southwark Crown Court twice, with two different juries. This enabled me to record two versions and opt for the one that seemed to ask the most interesting questions.

I premiered the film within an installation that was comprised of a modernist jury bench (made in collaboration with joiner Max Woodhall) complete with wireless headphones, to provide the audience with a sense of participating in the process of justice. Having visited many courtrooms at Southwark and the Old Bailey, I was driven to think about how a futuristic courtroom would look and feel. I chose the color blue to reference the digital era of the internet. If you ask Google, "What is the color of the internet?" it comes back in the search engine as "blue."[3] I chose to replace the terrestrial green leather seating with blue leatherette. In addition, I worked with artist Liza Brett to create a series of courtroom drawings. The work has been shown in many gallery settings since, including ZKM in Karlsruhe, Germany, as part of the exhibition *Open Codes* in 2017, and at the Illingworth Kerr Gallery at the Alberta University of the Arts in Calgary, Canada. It

3. See Dom Hennequin, "Why is Blue the Internet's Default Color?," January 9, 2018, www.envato.com/blog/blue-internet-default-color

has also been screened in law schools in the UK and at the Federal Ministry of Justice and Consumer Protection in Berlin, where I hope it has gone some way to impacting the practice of law and potentially the implementation of algorithms in decision-making.

Infrastructural technologies like algorithms are the bedrock of financial transactions and are now used universally in the quest to capture and learn from big data, predict risk and outcomes, and perform tasks. Given the opaque quality of algorithms, known for being "black boxes" where input and output are available but the actual processes that are used to arrive at a decision are notoriously murky, especially to a layperson, I was drawn to the context of the courtroom to open up the subject in the flesh. As Andrew King, a reader in medical image analysis at King's College London, states,

> all AI models have an input and output process, but it is not easily understandable to an everyday person . . . it is an opaque process. In reality, it is possible to inspect an AI black box model, but it is not so much about seeing as about understanding how complex decisions are reached.[4]

I wanted to make the culpability of an algorithm tangible, while expanding notions of transparency and opacity. Boris Groys's statement, "the visualization of the invisible is the most radical form of its profanation"[5] felt particularly appropriate in this case. My preoccupation with the way the immaterial meets forms of life, particularly in the social realm, was made manifest by my choice to use the state apparatus of the courtroom as a set to battle out questions on algorithmic culpability. To find ways to make subjects that are difficult to comprehend easier to understand, by facilitating the participation in, for instance, the mechanics of a courtroom trial and constructing a story and set that gives audiences the chance to work through difficult questions is an artistic strategy I have employed over many years. This tactic is also evident in my ongoing project *Birth Rites Collection*, which sheds light on the traditionally and universally taboo subject of the visual representation of birth by curating a collection of contemporary art on birth and exhibiting it in educational institutions.[6] To take art out of gallery spaces and into the world in which we live is challenging, maybe educational, and often revealing. It is also something I am committed to in my artistic practice.

The Trial of Superdebthunterbot also looked to explore Benjamin H. Bratton's concept of "planetary-scale computation," which is elusive, vast, and unfathomable. Bratton writes, taking

> different forms at different scales: energy grids and mineral sourcing; chthonic cloud infrastructure; urban software and public service privatization; massive universal addressing systems; interfaces drawn by the augmentation of the hand, of the eye, or dissolved into objects; users both overdetermined by self-quantification and exploded

---

4. Andrew King, "Trustworthiness in Healthcare AI," presentation at online event hosted by London Medical Imaging and AI Centre for Value Based Healthcare (November 18, 2021).

5. Boris Groys, *Art Power* (Cambridge, MA: MIT Press, 2008).

6. A collection of contemporary art on childbirth, www.birthritescollection.org.uk was founded in 2009 by Helen Knowles originally at the University of Salford (2009–17), then installed at King's College London (2017–21), and is currently hosted by the University of Kent (2021–present).

Film still from *The Trial of Superdebthunterbot* (2016) by Helen Knowles.

> by the arrival of legions of nonhuman users (sensors, cars, robots) . . . align, layer by
> layer, into something not unlike a vast (if also incomplete), pervasive (if also irregular)
> software and hardware *Stack*.[7]

Exposing the new sovereign territories of the internet and how these grind upon the old horizontal sovereign territories of land, sea, and mineral wealth became pertinent in the staging of the performance at Southwark. In the film, I looked to emphasize the old horizontal territory of the decrepit courtroom. Fraying carpets tacked together by tape, tired and threadbare, claim to insulate rubber vinyl floors that edge the peeling wooden laminate furniture and hoodwink us into believing they are doing a good job of muffling the aural proceedings. The material infrastructure of the courtroom is shown for what it is: falling apart. On a short research visit to the court at Southwark, I met a sarcastic and embittered clerk who told me that in the 1980s it was dubbed "a flagship court." The building's interior looked worn out and underfunded in the age of austerity.[8] I was keen to illustrate on film this clash between the physical spaces we inhabit and the ethereal "vertical"[9] territory of the internet, where algorithmic governance and data capture happens surreptitiously, with and without us humans.

The film was shot using three HD cameras, a camera drone, and several GoPros (even one worn as a head camera by a member of the jury), to emphasize the cutting-edge technology and surveillance themes behind the work. I attempted to convey Bratton's

---

7. Benjamin H. Bratton, "The Black Stack," *e-flux* 53 (March 2014), www.e-flux.com/journal/53/59883/the-black-stack.

8. After the 2008 economic crash, the conservative UK government was in the throes of eroding the welfare state and underfunding hospitals and civic spaces, highlighted most presciently during the recent pandemic as hospitals, courts of law, and the National Health Service duly suffered since they were the locus of these funding cuts.

9. Bratton, "The Black Stack."

"verticality" by using footage from these different cameras and stacking the imagery by editing between these different perspectives. Perhaps Bratton's emphasis on the autonomy of these interlocking machines, although incredibly poetic, is less based in reality than the idea of the assemblage between the human and machine. A technical assembly between a machine and a river was beautifully depicted by the philosopher Gilbert Simondon in his 1958 book *On the Mode of Existence of Technical Objects*, whereby the Guimbal turbine "integrates the river as both the driving force (of an engine wrapped by oil under high pressure) and the cooling agent"[10] because as the river flows faster, it drives the engine harder and equally cools it at the same pace. This appears as a somewhat tame and optimistic collaboration of nature and machine in comparison to the murderous outcome of *Superdebthunterbot*, which feeds off human desperation for money. When the trial of Adolf Eichmann, a Nazi war criminal, is held up as an analogy of distributed responsibility by a member of the jury in *Superdebthunterbot*, who asks if he "was a peg within a system [or whether] the system . . . ultimately determined his agenda," we become aware of the unimaginably destructive potential of these assemblages of human action and the (war) machine most famously described by Hannah Arendt as examples of the "banality of evil."[11] In *Superdebthunterbot*, the key issue battled out by the jury is the distribution of responsibility between machine and human. Does this distribution somehow water down the crime of murder? Although students have died as a result of the algorithm's intervention of placing job advertisements on Facebook pages, the charge of gross negligent manslaughter is the only available way to prosecute.

In *Unthought: The Power of the Cognitive Nonconscious* (2017), N. Katherine Hayles, a self-proclaimed feminist who embraces the corporeality of the world in which we live, clarifies her concept of assemblages (in opposition to the cleaner metaphor of the network), explaining that they "allow for contiguity in a fleshly sense, touching, incorporating, repelling, mutating. . . . interactions occurring across complex three-dimensional topologies [that] include information transactions across convoluted and involuted surfaces, with multiple volumetric entities interacting with many conspecifics simultaneously."[12] This messiness, codependency, and ultimate distribution of responsibility is a key theme that the jury members grapple with in the trial. One jury member argues, "We could say the same thing about a human being because that's actually a matter of, you know, kind of philosophy and religion to see whether or not the actions of our lives are predetermined by certain forces . . . outside of our control."

It remains to be seen whether "control" in this instance is achieved by employing cybernetic feedback loops. In her book *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (2020), Louise Amoore states these loops are either "contingent on a set of gradients and weighted probabilities," or their own truth logic, whose

10. Yuk Hui, "Machine and Ecology," *Angelaki* 25, no. 4 (August 2020): 54–66.

11. Hannah Arendt, *Eichmann In Jerusalem—A Report On The Banality Of Evil* (London: Penguin Classics, 1994).

12. N. Katherine Hayles, *Unthought: The Power of the Cognitive Nonconscious* (Chicago: University of Chicago Press, 2017), 118.

understanding of the world is seemingly at odds with "situated life," or both.[13] More specifically, who is deciding on the parameters of an algorithm, and what sociopolitical framework might influence the weighting of these parameters? Amoore petitions for the reinstatement of "doubt" when using such systems that could be read as hesitancy, humility, and constraint to inform a "cloud ethics." She illustrates the way algorithms make predictions from a risk calculus of "raw unlabeled data streams."[14] This nebulous process is embedded in the operations of algorithms and contrasts with the singular, actionable outputs spat out by machine learning. These variables range from an algorithm opting to act in one way despite potentially low-risk scores, to learning what a thing in the world is exclusively from attributes collected from image data. Imagine if we humans suddenly lost the myriad of senses that we use to make decisions and only employed vision. Efficient, cheap, and scalable decision-making at all costs may come with the problematic dismissal of complexity. Despite the inner variables of the recipe, the singular outcomes are taken as reliable, even fact. Eerily, this very human need to produce a discrete decision mirrors the sometimes unhelpful requirement of a trial to produce

Film still from *The Trial of Superdebthunterbot* (2016) by Helen Knowles.

13. Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Durham, NC: Duke University Press, 2020), 137.
14. Amoore, *Cloud Ethics*, 111.

a singular verdict, the binary guilty or not. The poignant plea by a member of the jury in the film, to use this as a

> policy opportunity here because ultimately what we see is, we can't sue the company that . . . was responsible for imbuing this algorithm with their agenda . . . What we have the opportunity to do is to actually use this algorithm as a certain vehicle for a just cause.

acknowledges the complexity of the problem and highlights a way to introduce standards for disciplining companies rather than simply stating whether the algorithm is guilty. When Hayles calls for "Critical Studies in Finance Capital"[15] my feeling is that theory may not be enough, so let us go further and embody these difficult decisions by using artistic practice in the process.

Of late, I have become interested in the work of the philosopher Yuk Hui. His books, *Art and Cosmotechnics* (2020), *The Question Concerning Technology in China: An Essay in Cosmotechnics* (2016), and *Recursivity and Contingency* (2019), critique the current hegemony of our capitalist paradigm, which has a direct influence on the tools we make and use, like algorithms, which are influenced by capitalist agendas like productivity, frictionless trade, and efficiency. This "ground," an idea put forward by Simondon that explains the individuation of technics from these forces, is now also becoming the ground from which they evolve, a cybernetic spiraling feedback loop. To be forever caught in this ground might to some seem dystopian, and yet somehow it is this eternal, darkly humorous dystopia whereby we can and can't give agency to the algorithm that is a strange and uncanny place in which to find ourselves. Yet Hui also explores the idea of acknowledging and making space for diverse ethical, moral, and cosmological forces to inform the development and use of technologies. An artwork like *The Trial of Superdebthunterbot* is part of that process. ■

HELEN KNOWLES is an artist and curator of the *Birth Rites Collection*.

15. Hayles, *Unthought,* 176.