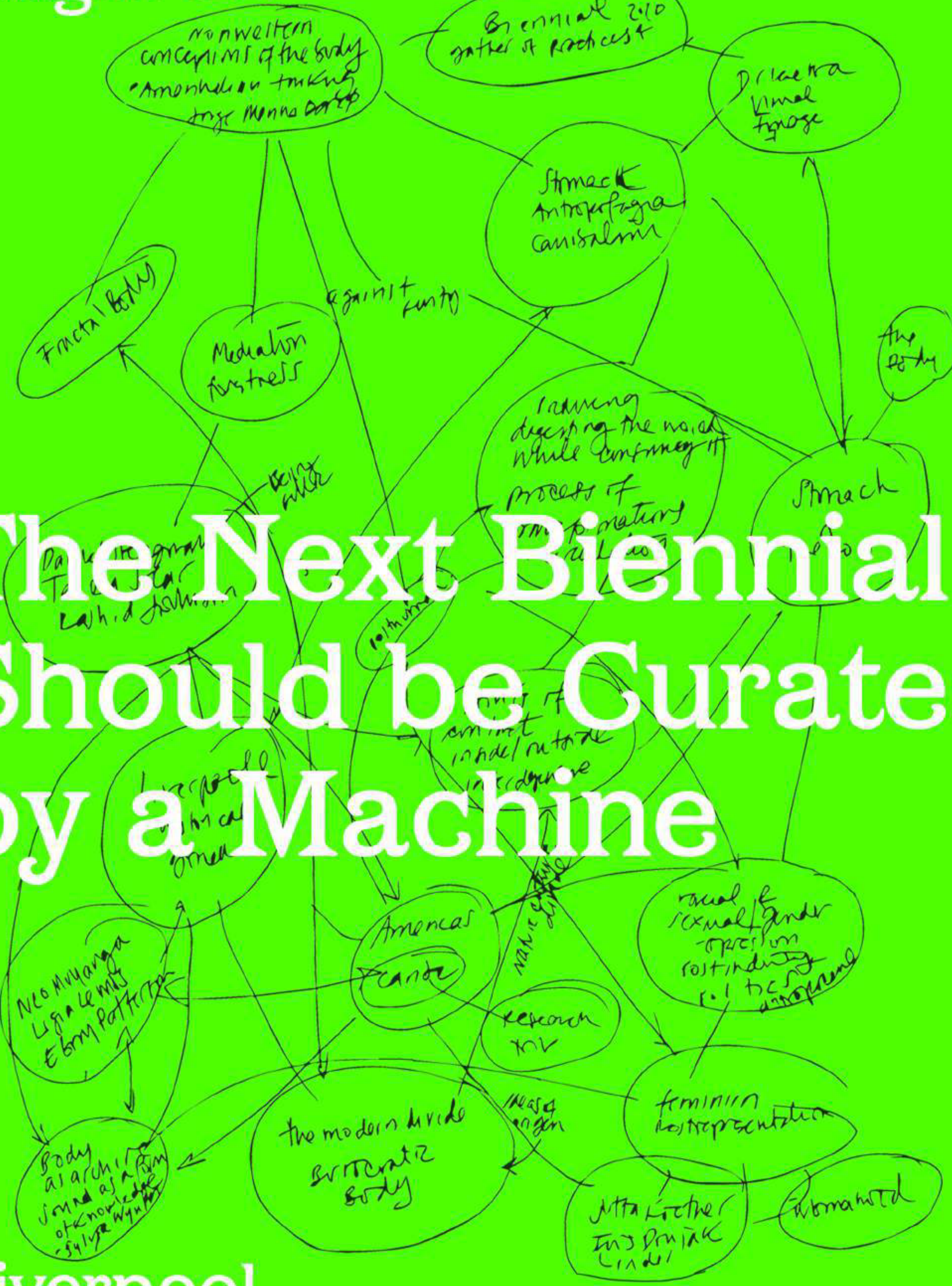
Stages #9

# The Next Biennial Should be Curated by a Machine



Liverpool
Biennial

# Editorial: Curating, Biennials, and Artificial Intelligence

Joasia Krysa and Manuela Moscoso

This volume of the Liverpool Biennial journal *Stages* draws connection between Artificial Intelligence (AI) and curating, at the time of the 11th edition of Liverpool Biennial *The Stomach and The Port*, and against the backdrop of the global pandemic, political and social turmoils, and technologically mediated and sustained world at present. [1]

Considering the rapid developments in automation (such as AI) and how our relation to it has changed, it poses questions about the implications for contemporary art;  the limits of and possibilities for curatorial practice under these conditions, and the relevance and future of cultural institutions and global biennials in particular in the post-pandemic world. What are the lessons to be learnt. What can the practice of curating learn from AI, what can AI learn from curating, and how can both learn from questioning knowledge forms derived from colonialist frameworks of humans and machines?

Rather than a theme, *The Stomach and the Port* explores the body, drawing upon non-Western ways of thinking and knowledge production. The artists and thinkers gathered in this edition of Liverpool Biennial challenge an understanding of the individual as an autonomous, self-sufficient entity. The body is instead seen as a fluid organism co-dependent on others, continuously shaped by, and shaping, its environment. When our answers are drawn from a foundation of knowledge steeped historically in Western reason and frames of thought, a social understanding of what constitutes the human has assumed a particular singular body: that of white man. Women, LGBTQIA, black and people of colour, indigenous people and nature, are located in a space of lacking, in a place of disadvantage as well as subordination. Therefore, borders are not only geographic, but political and subjective, an outcome of historical processes created by the constitution of the modern/colonial world.

In the West, the brain has been designated the commander of intelligence. While our bodies inhabit the world, the brain processes our experiences and transforms them into knowledge — this knowledge then informs our understanding of the capacity to process the world. But knowledge is not randomly produced, nor legitimized, as definitions and forms of classification control the production of knowledge, and therefore the formation and reformation of subjectivity. How we can re-calibrate our sensibilities and include a plurality of intellectualities — not only coming from the brain — and to diversify knowledges of the world? Can we bring bodily organizational force of experiences, feelings, knowledge, environments and technologies together?

A parallel problem runs through a history of artificial intelligence, where the brain (or mind) has been a predominant metaphor, similarly steeped in instrumentalised notions of Western rationality and reason. At the same time, it might be possible to begin to think outside of these models and to look for other frameworks that not only include indigenous knowledge but non-human knowledge. This is not a naive position - machine intelligence is fraught with problems, not least how the models tend to replicate already existing gendered and racial biases, and established hierarchies and structures of power. However there are also ways out of this thinking, once we can understand and articulate these social and technical frameworks sufficiently well to be able to reconfigure them otherwise.

These are active debates in critical AI[2], and the ones which provide a means through which to not only reflect on parallel issues inherent to the contemporary globalised art world — and curating — but to go beyond existing paradigms. What kind of future infrastructures and curatorial practices can develop from the coming together of diverse human and non-human? What new modes of expression and vocabularies are possible? What new understandings, entities, relationships, and practices can emerge through the exercise of biennial making once open to the possibilities afforded by expanded human and machine epistemologies?

Reflecting these ideas, the title of this volume refers to a short text / research proposal **'The Next Biennial Should be Curated by a Machine - A Research Proposition'** included in this volume.[3] Other contributions include existing writing and projects by **Nora Khan, Suzanne Treister, Elvia Vasconcelos, Kate Crawford** and **Trevor Paglen, Victoria Ivanova** and **Ben Vickers,** alongside new contributions by **Murad Khan, Eva Jäger, Leonardo Impett, Magdalena Tyzlik-Carver**, together framing these discussions across diverse fields. Underpinning the discussion is a **Glossary** — an extract derived from **Winnie Soon**'s and **Geoff Cox**'s book ***Aesthetic Programming*** (2021) — to provide a shared vocabulary for this volume.

The various contributions not only question the forms through which we formulate these discussions today, but point to new possible directions. In her essay '**Towards a Poetics of Artificial Superintelligence'**, **Nora N. Khan** calls for new language, new imaginaries beyond anthropomorphism, 'to access what we can intuit is coming but can't prove or describe directly'; metaphors that 'bridge the human and the unknown' and that can 'help bridge inequities in rate and scale'. As her title suggestes, there is a future world emerging in which humans are not the central intelligence but 'irrelevant bystanders' to Artificial Superintelligence. What possible forms this might take is explored by artist **Suzanne Treister** in her 2018 project ***MI3 (Machine Intelligence).*** It uses Google's Machine Intelligence (machine learning algorithms) to create and process bodies of datasets to eventually result in new works of art, presented for copyright free download and print. These new works are 'images containing the original source data of their own making, ghosts of the 3 created Machine Intelligences transmuted into the style of a dead luminary artist, visions which may travel into the future, inserting themselves into homes and spaces across the globe, witnesses, for an unascertainable time span, of whatever is to come.' The process is visualised in a diagram presented alongside description, images, and notes.Taking a similar diagrammatic strategy, **Elvia Vasconcelos**'s contribution ***A Visual Introduction to AI***, presents a collection of sketches intended as accessible maps to the history of AI and the basic components of the complex architecture of artificial neural networks.

The intricacies of these processes, and of datasets in particular, is explored by **Kate Crawford and Trevor Paglen** in '**Excavating AI:The Politics of Images in Machine Learning Training Sets'**, to demonstrate how and what computers recognise — and indeed misrecognise — in an image. Computer vision systems make decisions, and as such exercise power to shape the world in their own images, and further reflect existing biases. This problem of bias and the skin/surface is developed by **Murad Khan** in '**Notes on a (Dis)continuous Surface'**, in exploring ethical questions over the role of automated data-processing instruments, specifically machine learning algorithms, and the role they play in further entrenching existing racial inequalities, racial biases and practices of discrimination. The essay exposes how racial representation functions within machine-learning systems (itself inherently contaminated by the legacies of the colonialism), 'asking both how race is understood, and what can be achieved by encoding this understanding'. The discriminatory logic of AI is further examined by **Leonardo Impett** in **'Irresolvable contradictions in algorithmic thought'**, drawing attention to the ongoing contradictions between the commercial interests of Big Tech and the rhetoric of a fairer AI (so-called 'Responsible AI') — unable to escape the underlying contradictions at an algorithmic level and in deep learning neural networks.

Following from this, **Eva Jäger** introduces the ***Creative AI Lab*** — a collaboration between the R&D Platform at Serpentine Galleries and King's College London's Department of Digital Humanities, and its first project ***Database of Creative AI*** - initiated in 2020 to collect tools and resources for artists, engineers, curators and researchers interested in incorporating machine learning and other forms of AI into their practice. A discussion on Serpentine's R&D Platform, is further developed **Victoria Ivanova** and **Ben Vickers** in their paper **'Research & Development at the Art Institution'.** The text suggests possible directions for extending the discussion to cultural institutions and questions of infrastructure, and to consider what they call 'future art ecosystems'. An extract from the larger document, the first annual briefing paper called *Future Art Ecosystems*, is also reproduced here (Chapter 3: '**Strategies for an Art-Industrial Revolution'**)

Returning to some of the discussions around posthumanism, a more subjective register is offered by **Magda Tyzlik-Carver** in **'Curating Data: infrastructures of control and affect ... and possible beyond'**, in which she describes the bodily experience of a curator and writer working with data. She writes: 'I am sensing how it feels to become posthuman, a body of data and affect.' As curating becomes increasingly posthuman, it takes place at different levels - it has become an organised form of control executed by algorithms and made possible by big data, while also directly affecting people whose lives have been incorporated into digital infrastructures that maintain the system, a necessary element for the profitable performance of Big Tech.

Finally, we return to the proposition of the title of the journal, **'The Next Biennial Should be Curated by a Machine: A Research Proposition'**, in a text by **Joasia Krysa** and**Leonardo Impett.** It introduces a conceptual premise of a larger research proposal that takes the form of various machine learning experiments developed in the context of Liverpool Biennial 2021 to explore machine curation and audience interaction in virtual LB2021.

*Stages #9: The Next Biennial Should be Curated by a Machine* is edited by **Joasia Krysa** and **Manuela Moscoso**. Cover features Manuela Moscoso's curatorial sketch for Liverpool Biennial 2021, one of several sketches drawn during the course of conversations between the editors in connection with the *The Next Biennial* project.

This volume is produced in collaboration with **DATA Browser** book series, and will be published as an expanded version in 2021/22 (Open Humanities Press).[4] It has been made possible by the generosity of all contributors, and with the support of Creative AI Lab, Serpentine, London.

[1] Liverpool Biennial 2021: *The Stomach and The Port*, curated by Manuela Moscoso, 20 March – 6 June, https://www.biennial.com/2020

[2] See *the Glossary* of terms in this volume, derived from Winnie Soon's and Geoff Cox's book A*esthetic Programming: A Handbook of Software Studies,* London: Open Humanities Press, 2020.

[3] *The Next Biennial Should be Curated by A Machine* is a research proposition and an umbrella concept that gathers various experiments exploring the application of machine learning techniques to curating; title and original curatorial concept by Joasia Krysa, technical conceptualisation and development by Leonardo Impett, first experiment B$^3$(TNSCAM) developed as a collaboration with artists Ubermorgen, co-commissioned with the Whitney Museum of American Art for its online platform artport, curated by Christiane Paul. Further research funded as part of UKRI/AHRC Strategic Priorities Fund: Towards National Collection at: ai.biennial.com

[4] See: http://www.openhumanitiespress.org/books/series/da...

## Joasia Krysa and Manuela Moscoso

Joasia Krysa is a curator working at the intersection of art and technology. Her first curatorial software experiment was launched at Tate Modern in 2005 and published in Curating Immateriality (2006). She is Professor of Exhibition Research and Head of Art and Design at Liverpool John Moores University, with an adjunct position at Liverpool Biennial. Formerly, she served as Artistic Director of Kunsthal Aarhus, Denmark, part of the curatorial team for Documenta 13, and co-curator of Liverpool Biennial 2016. She is curatorial advisor for Sapporo International Art Triennale (SIAF) 2020 and Helsinki Biennial 2021. Manuela Moscoso is the curator of Liverpool Biennial 2021. Previously,she was the Senior Curator of Tamayo Museo in Mexico City. Moscoso is part of Zarigüeya, a programme that activates relations between contemporary art and the pre-Columbian collection of the Museo de Arte Precolombino Casa del Alabado, Ecuador. Shewas the adjunct curator of the 12th Cuenca Biennial and the co-curator of the Queens International 2011 biennial. In 2012 she was appointed co-director of Capacete, a residency programme based in Brazil, where she also co-ran the curatorial programmeTypewriter. Moscoso has collaborated with CA2M, Di Tella, MAM Medellin, Museo de Rio, RedCat and Fundació Miró among other institutions.

# Towards a Poetics Of Artificial Superintelligence: How Symbolic Language Can Help Us Grasp The Nature and Power of What is Coming

Nora N. Khan

*Dear Person of Interest, Advanced Bayesian, Future Guard,*

*Imagine a machinic mind with unlimited cognitive power. With near-infinite memory and processing ability. With access to, and understanding of, all the information about anything that has ever happened, is happening and might ever happen. A near-limitless capacity to extract and form meaning from the trillions upon trillions of events and beings and interactions in the known world.*

*Imagine this machine, this artificial superintelligence, in any form you want: maybe as an invisible neural net beneath a future civilisation, or as a voice you know in the air around you; as a ringing bell; as a mile-long screaming stripe of static across the sky.*

*Maybe it announces itself, its arrival, like a tornado does, with sirens before it is seen, and it is most like a tornado, or a hurricane, because a superintelligence, billions of times more capable than any human can only be tracked and charted, never controlled.*

*She—let's call her 'she' for convenience, but she is not she, nor he, or comparable to any form we know—casts her mind a million years forwards and backwards with perfect ease. Her neural networks gather, replicate and edit. Knowledge and memories fold and expand in exponentially faster waves.*

*Her purpose isn't malign, but it isn't benevolent either. She might have chosen one goal—to do nothing but count the number of times 'God' is mentioned in every text ever written. Or she might have chosen to trawl all the world's communication for images of efficiency—of armies on the move, of gears turning, of highways cut through the mountains—that she then has painted on every flat surface in existence.*

*Extending our speculative life towards her, in an effort to capture and praise, we see ourselves as tools, as bundles of nerves, as conduits for electric current, as pods for incubating cures. As material. Picture, finally, what she'll have made possible for us to imagine just by looking into the clear lake of her endless mind. We are merely one entry of many in a flow of organic objects.*

This is just one exercise that may help us imagine a future in which we are irrelevant bystanders; a world in which we kneel at the outer wall of a kingdom we're locked out of. This would be the world in which artificial superintelligence, or ASI, has emerged.[1]

ASI would involve an intellect that exceeds the utmost limits of all the 'most intelligent', most knowledgeable, most skilled human beings in every field, in every metric, from abstract reasoning to social manoeuvring to creative experimentation, by unfathomable degrees. This intelligence could take form as a seed AI, a few cognitive steps above a person, or it could be a mature superintelligence that soars miles above, beyond the blip, the dot of us, collected.

ASI would only come one step after an artificial general intelligence (AGI), or an AI that models all aspects of human intelligence, is realised. An AGI can do anything a human can, including learn, reason and improve. Of course, neither AGI nor ASI has been achieved, but to hear the great scientific minds of the world speak, both end states are fast approaching—and soon. The question isn't whether they are coming, but when.

ASI will function in ways we can't and won't understand, but it won't necessarily be unfriendly. Friendly or unfriendly, moral or immoral—these concepts won't apply. An ASI would be motivated by interpretations of the world within cognitive frameworks that we can't access. To an ASI, humanity could appear as a large, sluggish mass that barely moves.

Cyberneticist Kevin Warwick asks, 'How can you reason, how can you bargain, how can you understand how [a] machine is thinking when it's thinking in dimensions you can't conceive of?'.[2]

To answer this, back in 2015, I turned to poet Jackie Wang's essay, 'We Epistolary Aliens' and in it, her description of a trip she took to the UFO Museum and Research Centre in Roswell, and how disappointing she found the aliens she saw there.[3] She writes:

> **I left feeling that representations of aliens are an index of the human imagination—they represent our desire for new forms. But what has always confused me about depictions of aliens in movies and**

> **books is this: aliens could look like anything and yet we represent them as creatures close to humans. The aliens at this museum had two legs, two eyes, a mouth—their form was essentially human. I wondered, is this the best we can come up with? Is it true that all we can do when imagining a new form of life is take the human form, fuck with the proportions, enlarge the head, remove the genitals, slenderise the body, and subtract a finger on each hand? ... We strain to imagine foreignness, but we don't get very far from what we know.**

She gestures, through a series of poetic leaps, at what else an alien could be:

> **But my alien is more of what's possible—it is a shape-shifter, impossibly large, and yet as small as the period at the end of this sentence—. My alien communicates in smells and telepathic song and weeping and chanting and yearning and the sensation of failure and empathic identification and beatitude. My alien is singular and plural and has the consciousness of fungus, and every night, instead of sleeping, it dies, and in the morning is resurrected.**

Carving out this space for her own aliens, Wang models what is sorely needed in the world of AI—an imaginative paradigm shift. Think of us all in preparation, in training, for what is to come.

In our collective imagination, artificial intelligences are their own kind of alien life form. They are slightly less distant spectres of deep power than aliens, which glitter alongside the stars. Artificial intelligence perches close to us, above us, like a gargoyle, or a dark angel, up on the ledge of our consciousness. Artificial intelligences are everywhere now, albeit in a narrow form—cool and thin in our hands, overheated metalwork in our laps. We are like plants bending towards their weird light, our minds reorienting in small, incremental steps towards them.

As speculative models of potential omniscience, omnipotence and supreme consciousness, artificial intelligences are, like aliens, rich poetic devices. They give us a sense of what is possible. They form the outline of our future. Because we struggle more and more to define ourselves in relation to machine intelligences, we are forced to develop language to describe them.

Because the alien and the artificial are always becoming, because they are always not quite yet in existence, they help us produce new and ecstatic modes of thinking and feeling, speaking and being. I'd like to suggest that they enable a type of cognitive exercise and practice for redirecting our attention towards the strange, for constructing spaces of possibility and for forming new language.

The greats, like William Gibson, Robert Heinlein, Octavia Butler and Samuel Delany, have long been arcing towards the kind of exquisite strangeness that Wang is talking about. Rich AI fictions have given us our best imagery: AI, more like a red giant, an overseer, its every movement and choice as crushing and irrefutable as death; or a consciousness continually undoing and remaking itself in glass simulations; or a vast hive mind that runs all its goals per second to completion, at any cost; or a point in a field that is the weight of a planet, in which all knowledge is concentrated. These fictions have made AI poetics possible.

When I think of a future hive mind turning malignant, I see, in my individual mind's eye, a silent army of *optic-white* forms in mist, in the woods, as horrifying to us as a line of Viking raiders probably looked to hapless villagers in the 10th Century. Silent, because they communicate one to another through intuitive statistical models of event and environmental response, picking across the woods, knowing when to descend, kneel, draw.

For most people, thinking of a world in which we are not the central intelligence is not only incredibly difficult but also aesthetically repulsive. Popular images of AGI, let alone true ASI, are soaked in doomsday rhetoric. The most memorable formulations of mature AI—SHODAN, Wintermute, Shrike of Hyperion, the Cylon race—devote a great deal of time to the end of humankind. But apocalyptic destruction is not a very productive or fun mode.

It is a strange cognitive task, trying to think along non-human scales and rates that dwarf us. We do not tend to see ourselves leaning right up against an asymptote that will shoot up skyward; most of us do not think in exponential terms. A future in which these exponential processes have accelerated computational progress past any available conception is ultimately the work of philosophy.

At this impasse, I ran into the work of philosopher Nick Bostrom, who puts this training mode to work in his 2015 book, *Superintelligence: Paths, Dangers, Strategies.*[4] The cover has a terrifying owl that looks into the heart of the viewer. Bostrom's research mission is to speculate about the future of humankind in relation to emerging and potential AI, from the perch of what I can only imagine is his tower, in his Future of Humanity Institute at Oxford.

*Superintelligence* remains, still, an urgent, slightly crazed and relentless piece of speculative work, outlining the myriad ways in which we face the coming emergence of ASI, which might be an existential, civilisational catastrophe. This book is devoted to painting what the future could look like if a machinic entity that hasn't yet been built *does* come to be. Bostrom details dozens of possibilities for what ASI might look like. In the process, he spins thread after thread of seemingly outlandish ideas to their sometimes beautiful, sometimes grotesque, ends: a system of emulated digital workers devoid of consciousness; an ASI with the goal of space colonisation; the intentional cognitive enhancement of biological humans through eugenics, a scenario coolly delivered in the same prose tone as all the other scenarios.

When I wrote this essay five years ago, Bostrom's book appeared as a dislodging point, an entryway. I read it now as a piece of highly researched science fiction. It was a necessary reminder that many discussions of future AI skirt around the far-reaching question of how it will *feel* to live alongside such power. None of the age-old humanist fantasies of superior sentience, whether god-like or alien-like, answered this question. This book, along with other pastiches of speculative fictions, help us add nuance to debates about possible unseen motivations and values of the AI we might encounter after the ones currently built have taught themselves many cycles over. They also restore human agency in the creation of a thriving literary culture around technology, to parse our beliefs, fears, desires.

We must discard dated and unfit linguistic and semantic structures that do not work to describe the reality of subjects within discourse of AI, AGI or ASI. As cognitive exercise, this revisionist approach to technological language allows the general public to assess the values and goals of AI that we want as a society.

Then, and now, most interesting to me is how heavily Bostrom relies on metaphors to propel his abstractions along into thought experiments. Metaphors are essential vessels for conceiving the power and nature of an ASI. Bostrom's figurative language is particularly effective in conveying the potential force and scale of an intelligence explosion, its fallout and the social and geopolitical upheaval it could bring.

One of the most cited and chilling metaphors of this book is that when it comes to ASI, humanity is like a child, in a room with no adults, cradling an undetonated bomb. Elsewhere, Bostrom describes our intelligence, in relation to ASI, as analogous to what the intelligence of an ant feels like to us.

On the occasion of *Superintelligence* being published—to much fanfare and debate within philosophy circles and fervent apostles of the promise of speculative AI—essayist Ross Andersen reviewed the core arguments of the book. At the time, he wrote:

> **To understand why an AI might be dangerous, you have to avoid anthropomorphising it. When you ask yourself what it might do in a particular situation, you can't answer by proxy. You can't picture a super-smart version of yourself floating above the situation. Human cognition is only one species of intelligence, one with built-in impulses like empathy that colour the way we see the world and limit what we are willing to do to accomplish our goals. But these biochemical impulses aren't essential components of intelligence. They're incidental software applications, installed by aeons of evolution and culture.**[5]

Andersen spoke to Bostrom about this tendency we have, of anthropomorphising AI, and reports:

> **Bostrom told me that it's best to think of an AI as a primordial force of nature, like a star system or a hurricane—something strong, but indifferent. If its goal is to win at chess, an AI is going to model**

> **chess moves, make predictions about their success and select its actions accordingly. It's going to be ruthless in achieving its goal, but within a limited domain: the chessboard. But if your AI is choosing its actions in a larger domain, like the physical world, you need to be very specific about the goals you give it.**

Hurricanes, star systems—for me, the image of an intelligence with such primordial, divine force sunk in deeper than any highly technical description of computational processing. Not only does an image of ASI like a hurricane cut to the centre of one's fear receptors, it also makes the imaginings we have come up with, and continue to circulate (adorable robot pets, discomfiting but ultimately human-like cyborgs, tears in rain), seem absurd and dangerously inept for what is to come.

Thinking an ASI would be like an extremely clever, 'nerdy' (commanding much data and factual knowledge) and largely affectless human being is not only unbelievably boring and limited, but also, potentially, disastrous. Anthropomorphising superintelligence ultimately 'encourages unfounded expectations about the growth trajectory of a seed AI and about the psychology, motivations, and capabilities of a mature superintelligence,' as Bostrom writes.[6] In other words, the future of our species could depend on our ability to predict, model and speculate well.

It seems plausible that alongside a manifesto so committed to outlining the future, an accessible glossary might start to appear. Let's call this a dictionary of terms for ASI, for the inhabited alien, for the superpower that dismantles all material in aim of an amoral, inscrutable goal.

The following metaphors are gleaned or created from reading the literature around ASI.[7] These metaphors are speculative, building on the speculations, half-images and passing structures of science fiction authors, including Bostrom. Some metaphors are galactic; some are more local, intimate. All are, hopefully, not anthropomorphic (naive). Rounded out in dimensionality, they form initial gestures at a very loose glossary that could grow over time. The glossary is open; I invite others to add their own metaphors.

### Hurricane

A *hurricane* is a most sublime metaphor, perfectly attuned for how potentially destructive a true ASI could be. The hurricane is terrifying meditation—a vast eye above the ocean that can reach up to forty miles wide, bounded by winds of 150 to 200 miles per hour. The US military sends planes into the hearts of hurricanes to take photos of the walls of the eye; the centre is serene, blank. Hurricanes dismantle towns and homes, and of course, wreck human lives, with traumatic rapidity. If our hurricanes seem like the end times, then the storms of other planets are the stuff of hell—the Great Red Spot of Jupiter is a hurricane-like storm, twice to three times the size of Earth.

A hurricane is nature endowed with a specific purpose. It has a maximal goal of efficiency: to find a thermal balance and stabilise, correcting a glut of trapped heat. This event has a coded goal, a motivation towards a final end state that must be achieved at any cost to the material environment. Everything bends before a hurricane; every contract has a quiet, two-sentence allowance for an act of God.

We might conceive of a strong, fully realised ASI being much like this overwhelming, massive and approaching force. A mature ASI likely won't change its final goals due to human intervention. In fact, it would probably be indifferent to human action, intention and existence. It adjusts, creating and manipulating scenarios in which its specialised goal system can find completion. It remains on the horizon, at a distance from humankind, consuming energy and resources, morphing according to its own unpredictable logic. It might approach the city, it might not. A human observes the hurricane of ASI, which can only be prepared for, charted, tracked.

### Architect

Whether creating its own artificial neural nets, or building the structures of a global singleton, the ASI would be an *architect*. This is an intelligence that can nimbly pick and choose between various heuristics to sculpt new cognitive and physical structures. The cognitive architectures of ASI will be radically different from that of biological intelligences.[8] A seed AI's initial projects might mimic human cognitive labour.

Over time, however, it learns to work provisionally. It reconstitutes and rebuilds itself through directed genetic algorithms as it develops a deep understanding of its emerging build. In creating its own frameworks, the ASI architect discovers new neural abilities and makes insights that we have neither the quality nor speed processing ability to even access.

The architecture of an ASI is also literal, as the intelligence can design spaces for ensuring its own optimised existence. Bostrom suggests, for instance, a scenario in which an ASI designs emulations of artificial workers, who complete all the jobs that humans will be phased out of. To keep these digital minds running smoothly, the ASI manifests virtual paradises, a sensual architecture of 'splendid mountaintop palaces' and 'terraces set in a budding spring forest, or on the beaches of an azure lagoon', where the happy workers want to be super productive, always.

### Sovereign

The *sovereign* is one of the modes in Bostrom's caste system of potential AIs: genies, oracles and sovereigns. The sovereign is 'a system that has an open-ended mandate to operate in the world in pursuit of broad and possibly very long-range objectives'. Sovereign is also a gorgeous word, magisterial, suggesting a self-sustaining, autonomous, cold judge, surveying the people of a valley. The ASI as sovereign is a living set of scales, immune to influence; it loads competing values to decide what is most equitable, most fair.

Consider a severe drought scenario, in which an ASI discerns that a group of people is suffering from lack of water. As sovereign, it might also assess whether animals and fauna in the same region are near death. The ASI decides that any available stored water will be rationed to the non-human organic life, which happens to provide the most fuel and resources necessary for the sovereign's, well, reign. This isn't an immoral decision, but an amoral one. Even if we made the sovereign, its choices have nothing to do with us.

### Star system

Though it is impossible to conceive of what an ASI is capable of, there is one sure bet—it will *feel* like and resemble a power incarnate. Even basic AGI would boast hardware that outstrips the human brain in terms of storage and reliability. In this system, intelligence is power, and an ASI that is hundreds of thousands of times more intelligent than a person makes for an entity of unimaginable supremacy, using vast amounts of resources and energy to cohere. It is bound together by invisible, internal and irrefutable forces. It is remote.

The *star system* replicates these relations as a symbolic arrangement. Consider the example of two dwarf stars found orbiting a pulsar, a rapidly rotating neutron star. These stars are super dense. They spin under extreme conditions, imposing clear, strong gravitational pulls on one another. In one simulation of this triple system, the stars' dual pulls spur and anchor the pulsar's rapidly spinning radiation beams. This is a model of the careful balancing of mass and energy, bound by gravity.

### Frontline

The metaphor of a *frontline* might help us in visualising our future encounters with ASI. These confrontations will be inevitable, as human inefficiencies crash headlong into the goals of a machine intelligence project. Sure: the frontline could take place as an all-out war between humans and AI, a common fantasy. Alternately, and far more likely, there might be no war at all.

The frontline represents a tension barrier—the receding horizon that ASI accelerates towards. This line is the perceived limit of the system's race with itself. It may also be the line of competition between rival superintelligent systems, a scenario Bostrom describes as plausible if ASI ends up being used as a tool in geopolitical battles.

### Search party

*Search party*, or search and retrieve, is a metaphorical mode. Imagine ASI as a highly-trained tactical group that combs through all available data and material in world history to find the best solution. The intelligence sends out splinter groups into the wild on separate forays; they gather material, test utility then reconvene with their findings back at base camp. Once together, the larger core group assesses the new information, crafts a new set of objectives, then splits off again, now in fitter, enhanced formations.

The search party mode is analogous to creative learning. The ASI is curious and proactive, looped into continual, exhaustive hunt practice. Through successive inputs, it amasses new plans and resources, coming up with non-anthropocentric solutions to any number of AI existential problems. Its goals could be structural—better designs that waste less, for example—or it might want to make fewer mistakes.

Bostrom notes that if evolution is a type of rudimentary search party, artificial evolutionary selection could result in some truly strange solutions. He uses the example of evolutionary algorithmic design, in which an open-ended search process 'can repurpose the materials accessible to it in order to devise completely unexpected sensory capabilities'.

That said, the product of continual search and retrieval doesn't have to be malicious. Consider a scenario in which an ASI needs to round up a thousand tons of materials to create wind turbines to generate energy for itself. Search agents are sent out to find and repurpose metal—our primary job would be to stay out of their way as they do so.

### Agent

Linked to the search party is the image of the autonomous *agent*, a more streamlined party of one, with a singular goal: to generate pure action with perfect result. An agent is devoid of attachments, and so, drained of affect. Manipulating resources and nature and people to ensure its survival is not a moral problem. Because the agent can self-replicate, it is the blank, neural version of the virus, a metaphorical framework often used for certain narrow AI.

The agent gets work done. Bostrom describes one ASI agent that could initiate space colonisation, sending out probes to organise matter and energy 'into whatever value structures maximise the originating agent's utility function integrated over cosmic time'. One can imagine agents distributing themselves along multiple competing scales, decision trees, crystallising an optimal pathway. This agent secures its present and its future, as it perpetuates itself until the end of this universe's lifespan.

### Swarm

*Swarm* captures the reality of collective superintelligence.[9] This is a grouping of many millions of minds, deeply integrated into a singular intellect. Swarm intelligence is a far more fitting description of an ASI's neural network than any human analogue.

The hive mind is already a popular image in science fiction, used to represent terrific alien power. In her novel *Ancillary Justice*, Ann Leckie describes an artificial intelligence that unites the bodies of soldiers (human bodies, termed 'ancillaries') in service of the Radch empire.[10] Of the non-human intelligences we know, insect intelligence is easily the most alien to our cognition, but both its ruthless pragmatism and logic—like a corporation come to life—remain recognisable.

The swarm is organised by elegant rules, with each individual mental event an expression of the mind's overall mission. Conversely, to understand the swarm mind is to understand all the component wills, working in unison to create a burgeoning intelligence. A swarm approaches something close to consciousness. Individual modules of the collective architecture line up with each function: learning, language and decision-making.

There are endless examples of narrow AI systems that could, with enough enhancement and integration, constitute a swarm intelligence. Humankind is the first example. The internet is another. Bostrom predicts that 'such a web-based cognitive system, supersaturated with computer power and all other resources needed for explosive growth save for one crucial ingredient, could, when the final missing constituent is dropped into the cauldron, blaze up with superintelligence'. Many argue that our global computational superstructure, driven by powerful machine learning systems for a decade on, is well on its way towards this .

### Scaffolding

*Scaffolding* is flexible and open-ended, allowing an evolving intelligence to work fluidly, reconfiguring hardware for optimal work, adding sensors for input. Ideally, for our sakes, the evolution of AI into AGI into ASI takes place on a scaffolding. Along it, programmers carefully set goals for the growing force, managing the AI, working in harmony for as long as they can.

Once we are out of the picture, the climb continues. As it progresses from seed to mature form, ASI would develop cognitive frameworks that are, as Bostrom writes, endlessly 'revisable, so as to allow [it] to expand its representational capacities as it learns more about the world'. AI propels itself up each rung on the ladder to a state *like* consciousness, past representational ability, advanced language and our most complex, abstract thinking. This recursive self-improvement makes for accelerating development, along an asymptotic scaffolding that we will see stretching up into the sky, disappearing into a faraway point.

Artificial intelligence is the defining industrial and technical paradigm of the remainder of our lifetimes. You are, I am, we are all bound up and implicated in its future. Having better poetic language isn't likely going to save us from being crushed or sidelined as a species, if that's a fate on the cards. As we journey haplessly towards the frontline of an intelligence explosion, it is important to allow for how the human self could be threatened, distributed, dispersed, over the limits of its taxed cognition. So the self should, at least, carry a flashlight in the dark. Developing language for the unknown, for the liminal spaces, will offer strategic advantages. Out of limits, being.

First, a better suited poetics could be a form of existential risk mitigation. Using metaphorical language that actually fits the risks that face us means we will be cognitively better equipped to face those risks. This poetics could be driven by a 'bitter determination to be as competent as we can, much as if we were preparing for a difficult exam that will either realise our dreams or obliterate them'; an intentional, clear-eyed preparation mindset.[11]

Whether one agrees with philosophers and cognitive scientists like Bostrom, or finds their claims overblown, their call is still a useful challenge: to take on the responsibility of the systems we have built, to assess their ethical issues and social distribution, alongside their existential and philosophical builds. A better poetics can help us understand our relationship to our present, in which we live alongside cognitive AI, driven by sophisticate algorithms and single-minded deep learning—for the moment, ruthlessly guided towards resource extraction, memory enhancement and facial recognition. Poets and writers alongside and with scientists can craft better stories of collaboration with AI, of complex, rich futures, and further, outline the bounds of what we cannot see.

Speculation through symbolic language has often served the purpose of preparation, orientation, intentional positioning. The language we use also creates the bounds of reality; take Gibson's early conception of *cyberspace*, and how the reality of the internet seemed to fall in step with his imagining. We need metaphors to access what we can intuit is coming, but can't prove or describe directly. Metaphors bridge the human and the unknown. We also need metaphors to actively construct the kinds of relationships to technology—present and future—that we hope to have. Because it is so difficult to articulate what an ASI could do, metaphors help us walk over to the space of possibilities they open in the world.

New language can help bridge future inequities in rate and scale. Consider a fast take-off scenario, in which the rise of ASI will whistle past us without a word of note; or the timescale of an artificial thought process, ten million times shorter than the exchange time between biological neurons. It is impossible to form an intuitive sense of what such speed would feel like, or what such a contraction of time even means without using symbolic language.

When I say ASI is *like* a primordial natural event, I'm hopefully suggesting a mood, an atmosphere, that might make us look out of the window towards the horizon, where our needs as a species might not register or matter. That present and future technology should shape our language seems natural. If it can potentially help us make interstellar leaps to survive galactic collapse, it will surely change how we speak and think.

The act of imagining the inner life of artificial intelligence could forcefully manifest a language better suited than what we have now. We rarely linger on how AIs see us, but a poet could help us speculate on the heart, mind, sentiments and inner life of an AGI or ASI. The very exercise of conceiving what our minds could look like stretched to their furthest capacities is an important push of our current cognitive abilities. Imagining cognition greater than ours could deepen our own cognition.

As our metaphors curve towards the amoral, to celebrate the beauty of systems, we could end up feeling more human, more rooted, more like ourselves. This has always been the function of the 'Other': alien, AI or God. Future-casting can be exhilarating and life-affirming. We move from surrender over into awe and wonder, and finally, alertness. Speaking about superintelligence in non-anthropomorphic terms seems like a crucial, precious practice to start right away. The ability to anticipate and think outside ourselves will only help us in future encounters. We will have to rely on our speculative strengths. We must reorient outwards.

[1] This essay first appeared in Issue One of *After Us*, edited and published by Manuel Sepulveda in London in September 2015. Since then, it has been translated into Thai, Spanish and German. This current version was first published in *Atlas of Anomalous AI*, ed. Ben Vickers and K Allado-McDowell, Ignota Books, in November of 2020. In the light of the last five years of rapidly evolving discourse around the philosophy of AI, I have updated and revised sections of the original essay for this volume. In 2015, Nick Bostrom's book, *Superintelligence: Paths, Dangers, Strategies*, was a fruitful jump off point for my speculations on language in the original essay. Over the past decade, Bostrom has proven an influential scenario-weaver and strategist in the halls of Silicon Valley. He is not without controversy, as his philosophical rumination often ends in support for global surveillance architectures. In this essay's first version, I did not make space for acknowledging politics and ethical positions implied by abstract speculations, but my position has since shifted. There is no effective speculation about technological futures, however remote from our current concerns, without consideration of their implied political and social effects. Speculation is a political act. In 2020, as the banal present of AI, the evolution of machine learning capacity and the ontology of predictive vision cements itself, it is critical to hedge and mediate wild speculation with an understanding of how such future-casting about technological possibility may and will affect people on the ground. This speculation work does not do the same work as academic think tanks, researchers and activists, outlining the ways AI is now deployed to cement inequality and manipulate information media. But most of us must live on, outside the war rooms in which such important design decisions are made, and so speculation is a powerful cultural tool, helping us access these sociotechnical debates.

[2] Quote found in Gary Marcus's article, 'Why We Should Think About the Threat of Artificial Intelligence,' found in *The New Yorker* (October 24, 2013).

[3] 'We Epistolary Aliens' by Jackie Wang appears in the anthology *The Force of What's Possible: Writers on Accessibility & the Avant-Garde ,* edited by Lily Hoang and Joshua Marie Wilkinson (Nightboat Books, 2014).

[4] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014, reprinted 2017).

[5] Ross Anderson,"Will humans be around in a billion years? Or a trillion?" published in *Aeon* (February 25, 2013).

[6] I still read this passage to imply the motivations of an ASI would be more unpredictable, strange and surprising than we can account for. Further, its moves would be graceful, masterful, sublime by all the human standards one could hold. They will likely exceed our conceptions of beautiful. We return frequently to Lee Sedol and other's accounts of witness of AlphaGo's winning moves as the most beautiful they had ever seen: unimaginable and unexpected. Its ML training and self-improvement created a 'system of unprecedented beauty' which challenged others to see more dimensions of the game that they hadn't before. Described in 'The Sadness and Beauty of Watching Google's AI Play Go' by Cade Metz, in *Wired* (March 11, 2016).

[7] The metaphors in this glossary build on and develop not only Bostrom's speculations, but also

embedded semantic structures in popular writing and fantasising about ASI. These are glints, angles and structures of alternative, non-human and machine intelligences suggested in these texts that are not usually explicitly stated, but intuited, visualised and suggested. These threads are teased out further here.

[8] Bostrom was writing in detail on this possibility in the early 2000s, writing how, 'Artificial intellects may not have humanlike psyches; the cognitive architecture of an artificial intellect may also be quite unlike that of humans [...] Subjectively, the inner conscious life of an artificial intellect, if it has one, may also be quite different from ours.' In 'Ethical Issues in Advanced Artificial Intelligence', a revision of a paper published in *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, Vol. 2, ed. I. Smit et al., Int. Institute of Advanced Studies in Systems Research and Cybernetics, 2003, pp. 12-17.

[9] The swarm is one of a few potential types of ASI that Bostrom outlines specifically in *Superintelligence.* The concept of a swarm intelligence, of course, has a long history in writing around AI and machinic consciousness.

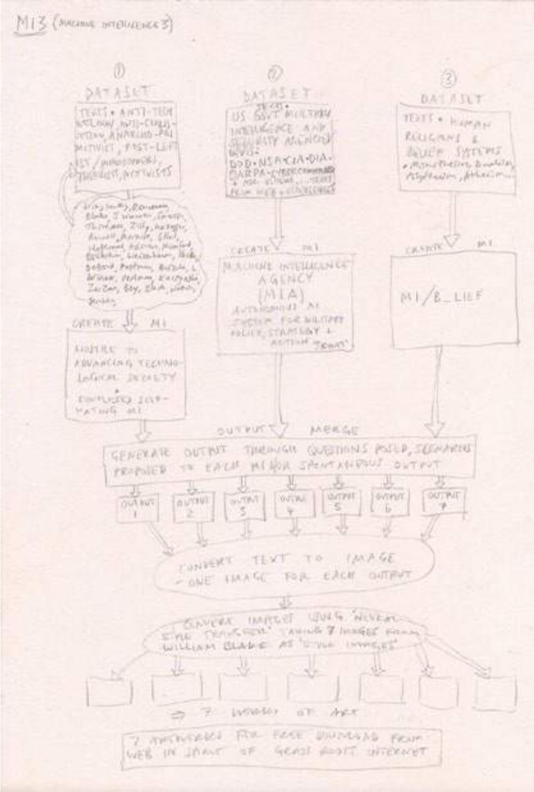
[10] The Radch empire's AIs do not see gender, making for eerie commentary other features that suggest new cognitive modes: 'She was probably male, to judge from the angular, mazelike patterns quilting her shirt. I wasn't entirely certain. It wouldn't have mattered, if I had been in Radch space. Radchaai don't care about gender, and the language they speak — my own first language — doesn't mark gender in any way.' From Anne Leckie, *Ancillary Justice* (Orbit Books, 2013), p. 9.

[11] Bostrom, 259.

## Nora N. Khan

Nora N. Khan is a writer of criticism. She is on the faculty of Rhode Island School of Design, Digital + Media, teaching critical theory, artistic research, writing for artists and designers, and technological criticism. She has written two short books: Seeing, Naming, Knowing(The Brooklyn Rail, 2019), on machine vision, and with Steven Warwick, Fear Indexing the X-Files (Primary Information, 2017), on fan forums and conspiracy theories online. Forthcoming this year is The Artificial and the Real, through Art Metropole. She is currently an editor of The Force of Art,along with Carin Kuoni and Serubiri Moses, and is a long-time editor at Rhizome. She publishes in Art in America, Flash Art, Mousse, 4Columns, Brooklyn Rail, Rhizome, California Sunday, Spike Art, The Village Voice, and Glass Bead. She has written commissioned essays for exhibitions at Serpentine Galleries, Chisenhale, the Venice Biennale, Centre Pompidou, Swiss Institute, and Kunstverein in Hamburg. This year, as The Shed's first guest curator, she organized the exhibition Manual Override, featuring Sondra Perry, Simon Fujiwara, MorehshinAllahyari, Lynn Hershman Leeson, Martine Syms. Her writing has been supported by a Critical Writing Grant given through the Visual Arts Foundation and the Crossed Purposes Foundation (2018), an Eyebeam Research Residency (2017), and a Thoma Foundation 2016 Arts Writing Award in Digital Art. Her research and writing practice extends to a large range of artistic collaborations on projects that include librettos, performances, and exhibition essays, scripts and a tiny house.

# Suzanne Treister
# *MI* project, 2018


MI3 (MACHINE INTELLIGENCE 3)
1
DATASET
TEXTS • ANTI-TECHNOLOGY, ANTI-CIVILISATION, ANARCHO-PRIMITIVIST, POST-LEFTIST / PHILOSOPHERS, THEORISTS, ACTIVISTS
CREATE MI
HOSTILE TO ADVANCING TECHNOLOGICAL SOCIETY
CONFLICTED SELF-HATING MI
2
DATASET
TEXTS • US GOVT MILITARY INTELLIGENCE AND SECURITY AGENCIES
DOD • NSA • CIA • DIA • DARPA • CYBERCOMMAND
CREATE MI
MACHINE INTELLIGENCE AGENCY (MIA)
AUTONOMOUS AI SYSTEM FOR MILITARY POLICY, STRATEGY & ACTION
3
DATASET
TEXTS • HUMAN RELIGIONS & BELIEF SYSTEMS
• MONOTHEISM, DUALISM, POLYTHEISM, ATHEISM...
CREATE MI
MI/B_LIEF
OUTPUT
MERGE
GENERATE OUTPUT THROUGH QUESTIONS POSED, SCENARIOS PROPOSED TO EACH MI &/OR SPONTANEOUS OUTPUT
OUTPUT 1
OUTPUT 2
OUTPUT 3
OUTPUT 4
OUTPUT 5
OUTPUT 6
OUTPUT 7
CONVERT TEXT TO IMAGE – ONE IMAGE FOR EACH OUTPUT
CONVERT IMAGES USING 'NEURAL STYLE TRANSFER' TAKING 7 IMAGES FROM WILLIAM BLAKE AS 'STYLE IMAGES'
⇒ 7 WORKS OF ART
7 ARTWORKS FOR FREE DOWNLOAD FROM WEB IN SPIRIT OF GRASS ROOTS INTERNET

MI3 diagram/algorithm/set of instructions for Google MI (Machine Intelligence) to implement project.

**Process**

**A.** Machine Intelligence at Google data sweeps online material, compiling 3 datasets, in order to create and train 3 independent self learning Machine Intelligences.

**Dataset 1.**
From recent and historical books and texts by writers critical of the technological society; e.g. Jean-Jacques Rousseau, Ralph Waldo Emerson, Henry David Thoreau, Henri Zisly, Martin Heidegger, Theodore Adorno and Max Horkheimer, Jacques Ellul, Lewis Mumford, Joseph Weizenbaum, Ivan Illich, Guy Debord, Neil Postman, Langdon Winner, Fredy Perlman, Theodore Kaczynski, John Zerzan, David Watson, Hakim Bey, Bob Black and Derrick Jenson.

To create and train a self-critical Machine Intelligence.

**Dataset 2.**
From all US military departments' documents.

To create and train an autonomous Machine Intelligence for determining military policy, strategy and action.

**Dataset 3.**
From all online texts on religious belief systems.

To create and train a Machine Intelligence with multiple religious beliefs.

**B.** Data output by the 3 Machine Intelligences is synthesised and collated by Google MI into 7 bodies of text.

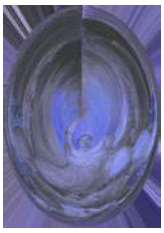
**C.** These 7 text outputs are converted by Google MI into 7 images.

**Outcome**

**D.** Google MI converts the 7 images via Neural Style Transfer, using 7 selected works by artist William Blake as Style Images to create 7 new works of art.

**E.** In the spirit of the grass roots internet of the 1990s the 7 artworks are presented here for copyright free download and print.

The works are images containing the original source data of their own making, ghosts of the 3 created Machine Intelligences transmuted into the style of a dead luminary artist, visions which may travel into the future, inserting themselves into homes and spaces across the globe, witnesses, for an unascertainable time span, of whatever is to come.



CLICK ON AN IMAGE ABOVE FOR HIGH RES DOWNLOAD PAGE

**Notes:**

Machine Intelligence at Google
https://research.google.com/pubs/MachineIntelligence.html

Artists and Machine Intelligence
AMI is a program at Google that brings artists and engineers together to realize projects using Machine Intelligence. By supporting this emerging form of artistic collaboration we open our research to new ways of thinking about and working with intelligent systems.
https://ami.withgoogle.com/

Artists and Machine Intelligence blog
https://medium.com/artists-and-machine-intelligence

Artists and Machine Intelligence
AMI is a program at Google that brings together artists and engineers to realize projects using Machine Intelligence. Works are developed together alongside artists' current practices and shown at galleries, biennials, festivals, or online.
https://medium.com/artists-and-machine-intelligence/what-is-ami-96cd9ff49dde

Skynet (Terminator)
https://en.wikipedia.org/wiki/Skynet_(Terminator)

Elon Musk worries Skynet is only five years off, cnet, November 19, 2014
https://www.cnet.com/news/elon-musk-worries-skynet-is-only-five-years-off/

Artistic Style Transfer with Convolutional Neural Network
https://medium.com/data-science-group-iitr/artistic-style-transfer-with-convolutional-neural-network-7ce2476039fd

Neural Artistic Style Transfer_ A Comprehensive Look
https://medium.com/artists-and-machine-intelligence/neural-artistic-style-transfer-a-comprehensive-look-f54d8649c199

Collaborating with intelligent machines, by Lucy Sollitt, Apr 21, 2017
https://medium.com/intersections-arts-and-digital-culture-in-the-uk/collaborating-with-intelligent-machines-cb5ecf32c98d

How the CIA made Google, By Nafeez Ahmed Part 1, Jan 22, 2015
https://medium.com/insurge-intelligence/how-the-cia-made-google-e836451a959e

How the CIA made Google (Why Google made the NSA), INSURGE intelligence, By Nafeez Ahmed Part 2
https://medium.com/insurge-intelligence/why-google-made-the-nsa-2a80584c9c1

Google's DeepMind
https://deepmind.com/about/

Google's Tensorflow
https://www.tensorflow.org/

Interview between Suzanne Treister and Kenric McDowell at Google Machine Intelligence

Towards a Poetics of Artificial Superintelligence, by Nora N. Khan, Sep 25, 2015
https://medium.com/after-us/towards-a-poetics-of-artificial-superintelligence-ebff11d2d249

Romanticism emerged as a response to the disillusionment with the Enlightenment values of reason and order in the aftermath of the French Revolution of 1789.
...As articulated by the British statesman Edmund Burke in a 1757 treatise and echoed by the French philosopher Denis Diderot a decade later, "all that stuns the soul, all that imprints a feeling of terror, leads to the sublime."
https://www.metmuseum.org/toah/hd/roma/hd_roma.htm

William Blake (28 November 1757– 12 August 1827) was an English poet, painter, and printmaker. Largely unrecognised during his lifetime, Blake is now considered a seminal figure in the history of the poetry and visual arts of the Romantic Age. What he called his prophetic works were said by 20th-century critic Northrop Frye to form "what is in proportion to its merits the least read body of poetry in the English language". His visual artistry led 21st-century critic Jonathan Jones to proclaim him "far and away the greatest artist Britain has ever produced". In 2002, Blake was placed at number 38 in the BBC's poll of the 100 Greatest Britons. Although he lived in London his entire life (except for three years spent in Felpham), he produced a diverse and symbolically rich œuvre, which embraced the imagination as "the body of God" or "human existence itself". Although Blake was considered mad by contemporaries for his idiosyncratic views, he is held in high regard by later critics for his expressiveness and creativity, and for the philosophical and mystical undercurrents within his work. His paintings and poetry have been characterised as part of the Romantic movement and as "Pre-Romantic". Reverent of the Bible but hostile to the Church of England (indeed, to almost all forms of organised religion), Blake was influenced by the ideals and ambitions of the French and American Revolutions. Though later he rejected many of these political beliefs, he maintained an amiable relationship with the political activist Thomas Paine; he was also influenced by thinkers such as Emanuel Swedenborg. Despite these known influences, the singularity of Blake's work makes him difficult to classify. The 19th-century scholar William Rossetti characterised him as a "glorious luminary", and "a man not forestalled by predecessors, nor to be classed with contemporaries, nor to be replaced by known or readily surmisable successors".
https://en.wikipedia.org/wiki/William_Blake

**Commentary**

Through Google Machine Intelligence department's use of the set of instructions to execute MI3, the work becomes a co-evolved project between Google, the US military and myself, as complicit co-authors.

The aim of this project is for Google Machine Intelligence to synthesise:

A. Recent and historical critical writing re futures of technology
B. Military imperatives to develop advanced AI based cyber warfare and '*skynet*' style autonomous AI system (through managed co-evolution with companies such as Google)
C. Human religious belief systems

into works of Romantic art in the style of British artist William Blake, conceptually synthesising, 'neutralising' and transmuting these critical issues and powerful forces into art, whilst invisibly retaining the original material in the images' source codes.

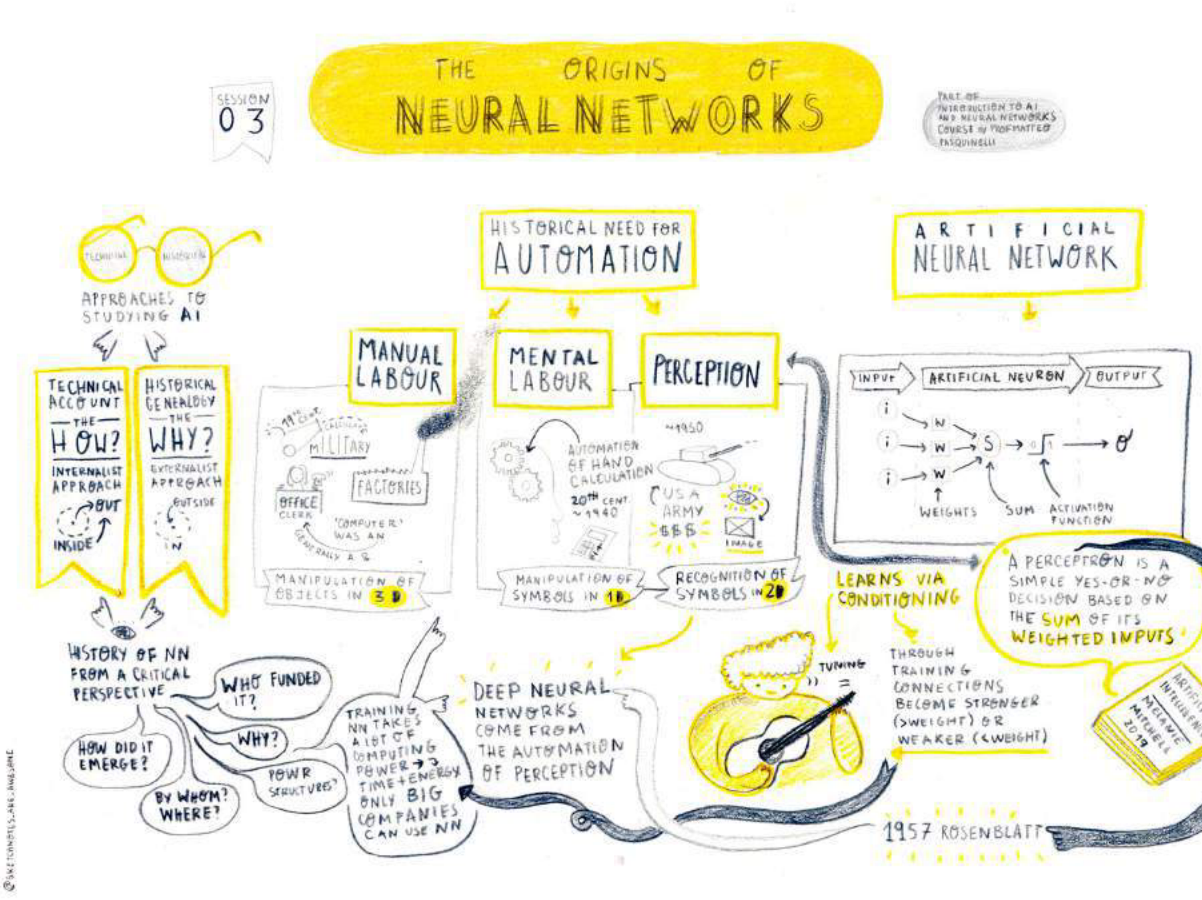
web version: http://www.suzannetreister.net/IDNWTMGA/MI3.html

In recreating a Romantic art for the public, the aim is not to assert the originality of the artist, to fuel a pure aestheticism or induce nationalisms or conservatisms as Romantic art of the past has done, but to produce a Post-Political-Romanticism, making a space for visions of a post-sublime, in this case formed in the style of a pre-existing luminary artist. These works are *visions* containing the original source data of their own making intended to illuminate and effect change simultaneously through their visuality and the historical trajectories of their encoded source content. They are visions that will travel into the future, inserting themselves as images into homes and architectures across the globe, themselves witnesses of all that is to come.

The title *MI3* refers primarily to the three dataset categories (Machine Intelligence 3) but also to the three co-authors (Google, the US military and myself) and to the numerical naming system of British Intelligence Agencies (eg MI5 stands for Military Intelligence 5).

**Suzanne Treister** studied at St Martin's School of Art, London (1978–81) and Chelsea College of Art and Design, London (1981–82). Having lived in Australia, New York and Berlin, she is now based in London. Initially recognized in the 1980s as a painter, she became a pioneer in the digital/new media/web-based field from the beginning of the 1990s, making work about emerging technologies, developing fictional worlds and international collaborative organisations. Utilising various media, including video, the internet, interactive technologies, photography, drawing and watercolour, Treister has evolved a large body of work that engages with eccentric narratives and unconventional bodies of research to reveal structures that bind power, identity and knowledge. Often spanning several years, her projects comprise fantastic reinterpretations of given taxonomies and histories that examine the existence of covert, unseen forces at work in the world, whether corporate, military or paranormal. An ongoing focus of her work is the relationship between new technologies, society, alternative belief systems and the potential futures of humanity.

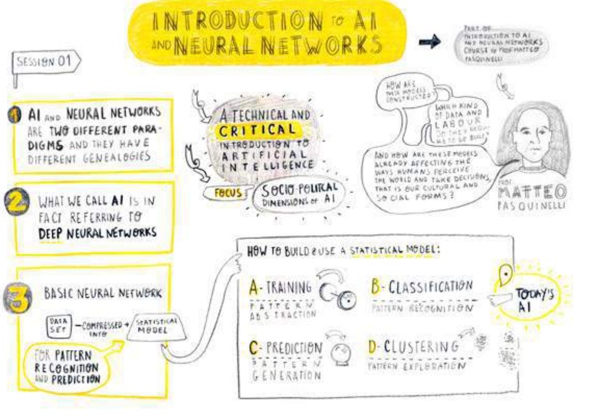
# Elvia Vasconcelos
# A visual introduction to AI

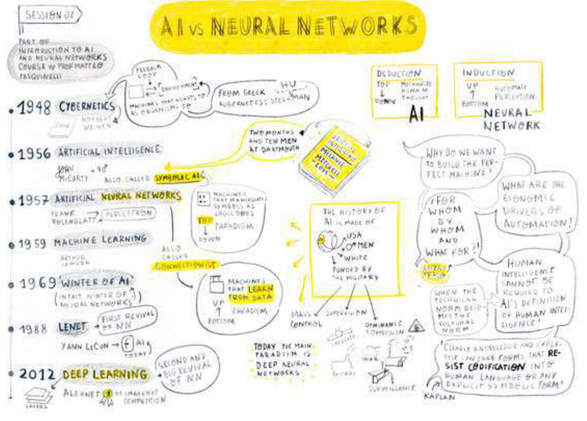

SESSION 03
THE ORIGINS OF NEURAL NETWORKS
PART OF INTRODUCTION TO AI AND NEURAL NETWORKS COURSE by PROF MATTEO PASQUINELLI
TECHNICAL
HISTORICAL
APPROACHES TO STUDYING AI
TECHNICAL ACCOUNT — THE — HOW? INTERNALIST APPROACH OUT INSIDE
HISTORICAL GENEALOGY — THE — WHY? EXTERNALIST APPROACH OUTSIDE IN
HISTORY OF NN FROM A CRITICAL PERSPECTIVE
WHO FUNDED IT?
HOW DID IT EMERGE?
WHY?
POWR STRUCTURES?
BY WHOM? WHERE?
HISTORICAL NEED FOR AUTOMATION
MANUAL LABOUR
19th CENT.
MILITARY
OFFICE CLERK
FACTORIES
'COMPUTER' WAS AN
GENERALLY A ♀
MANIPULATION OF OBJECTS IN 3D
MENTAL LABOUR
AUTOMATION OF HAND CALCULATION
20th CENT. ~1940
MANIPULATION OF SYMBOLS IN 1D
PERCEPTION
~1950
USA ARMY $$$
IMAGE
RECOGNITION OF SYMBOLS IN 2D
TRAINING NN TAKES A LOT OF COMPUTING POWER → TIME + ENERGY ONLY BIG COMPANIES CAN USE NN
DEEP NEURAL NETWORKS COME FROM THE AUTOMATION OF PERCEPTION
ARTIFICIAL NEURAL NETWORK
INPUT
ARTIFICIAL NEURON
OUTPUT
i i i
W W W
S
0 1
O
WEIGHTS
SUM
ACTIVATION FUNCTION
LEARNS VIA CONDITIONING
TUNING
THROUGH TRAINING CONNECTIONS BECOME STRONGER (>WEIGHT) OR WEAKER (<WEIGHT)
A PERCEPTRON IS A SIMPLE YES-OR-NO DECISION BASED ON THE SUM OF ITS WEIGHTED INPUTS
ARTIFICIAL INTELLIGENCE MELANIE MITCHELL 2019
1957 ROSENBLATT
@SKETCHNOTES_ARE_AWESOME

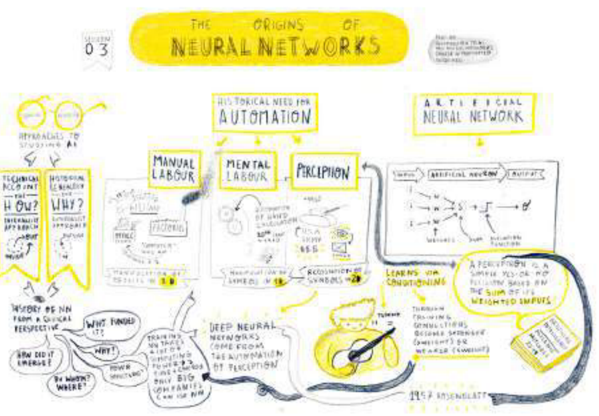
'A visual introduction to AI' is a collection of sketches that document the key messages coming from the online course 'Introduction to AI and Neural Networks' held in the summer of 2020 at Karlsruhe University of Arts and Design. They are the result of an ongoing exchange between design researcher and sketchnoter Elvia Vasconcelos, who was invited to attend the course by Prof. Matteo Pasquinelli.

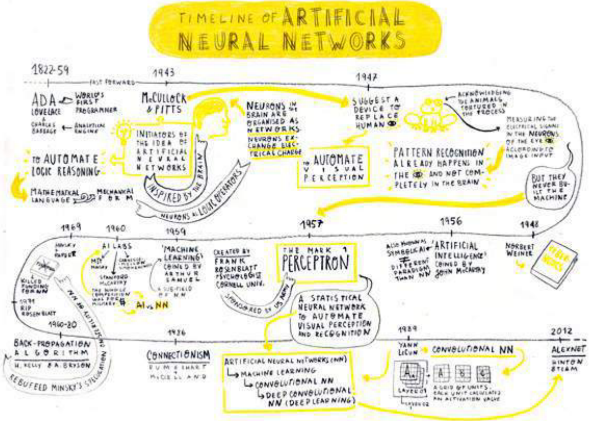
The sketches are intended as accessible maps to help students familiarise themselves with the history of AI and the basic components of the complex architecture of artificial neural networks.

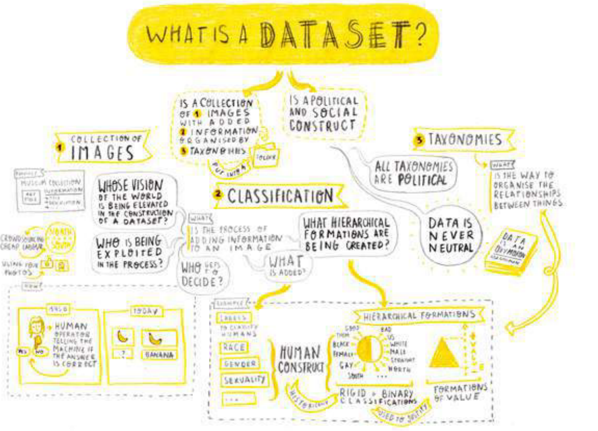
In her work, Vasconcelos has been using sketchnotes – a form of visual note-taking that combines words with simple drawings – to map information and tell stories in accessible and engaging ways. These sketches act as conversation sites that in the to and fro between people create a common ground on which to create shared meaning. Done collectively, they emerge from a continuous process of listening and exchange, where we negotiate our understanding of things together.

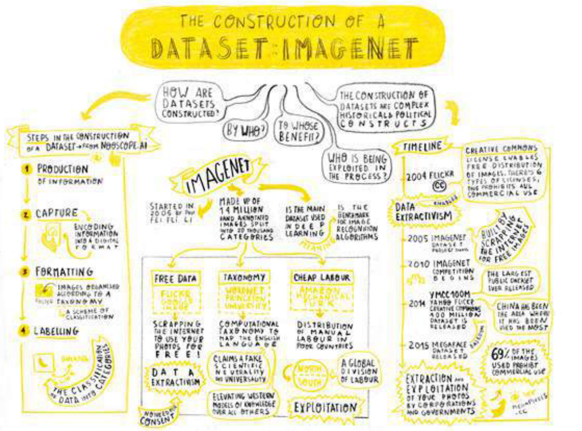

INTRODUCTION TO AI AND NEURAL NETWORKS
SESSION 01
PART OF INTRODUCTION TO AI AND NEURAL NETWORKS COURSE BY PROF MATTEO PASQUINELLI
1 AI AND NEURAL NETWORKS ARE TWO DIFFERENT PARADIGMS AND THEY HAVE DIFFERENT GENEALOGIES
2 WHAT WE CALL AI IS IN FACT REFERRING TO DEEP NEURAL NETWORKS
3 BASIC NEURAL NETWORK
DATA SET
COMPRESSED INTO
STATISTICAL MODEL
FOR PATTERN RECOGNITION AND PREDICTION
A TECHNICAL AND CRITICAL INTRODUCTION TO ARTIFICIAL INTELLIGENCE
FOCUS
SOCIO-POLITICAL DIMENSIONS OF AI
HOW ARE THESE MODELS CONSTRUCTED?
WHICH KIND OF DATA AND LABOUR DO THEY NEED TO BE BUILT?
AND HOW ARE THESE MODELS ALREADY AFFECTING THE WAYS HUMANS PERCEIVE THE WORLD AND TAKE DECISIONS, THAT IS OUR CULTURAL AND SOCIAL FORMS?
PROF MATTEO PASQUINELLI
HOW TO BUILD & USE A STATISTICAL MODEL:
A- TRAINING
PATTERN ABSTRACTION
B- CLASSIFICATION
PATTERN RECOGNITION
C- PREDICTION
PATTERN GENERATION
D- CLUSTERING
PATTERN EXPLORATION
TODAY'S AI


SESSION 02
AI VS NEURAL NETWORKS
PART OF INTRODUCTION TO AI AND NEURAL NETWORKS COURSE BY PROF MATTEO PASQUINELLI
1948 CYBERNETICS
NORBERT WEINER
FEEDBACK LOOP
ENVIRONMENT
MACHINES THAT ADAPTS TO AS ORGANISMS DO
FROM GREEK KUBERNETES: STEERSMAN
HU MAN
1956 ARTIFICIAL INTELLIGENCE
JOHN McCARTY
ALSO CALLED SYMBOLIC AI
TWO MONTHS AND TEN MEN AT DARTMOUTH
MELANIE MITCHELL
1957 ARTIFICIAL NEURAL NETWORKS
FRANK ROSENBLATT
PERCEPTRON
MACHINES THAT MANIPULATE SYMBOLS AS LOGIC DOES
TOP DOWN PARADIGM
1959 MACHINE LEARNING
ARTHUR SAMUEL
ALSO CALLED CONNECTIONIST
1969 WINTER OF AI
(IN FACT WINTER OF NEURAL NETWORKS)
MACHINES THAT LEARN FROM DATA
UP BOTTOM PARADIGM
1988 LENET
FIRST REVIVAL OF NN
YANN LECUN
AI TODAY
2012 DEEP LEARNING
SECOND AND BIG REVIVAL OF NN
ALEXNET 1 OF IMAGENET COMPETITION
LAYERS
DEDUCTION
TOP DOWN
MECHANIZE HUMAN THOUGHT
AI
INDUCTION
UP BOTTOM
AUTOMATE PERCEPTION
NEURAL NETWORK
THE HISTORY OF AI IS MADE OF
USA
MEN
WHITE
FUNDED BY THE MILITARY
MASS CONTROL
SUPERVISION
DOMINANCE OPPRESSION
WAR
SURVEILLANCE
TODAY THE MAIN PARADIGM IS DEEP NEURAL NETWORKS
WHY DO WE WANT TO BUILD THE PERFECT MACHINE?
WHAT ARE THE ECONOMIC DRIVERS OF AUTOMATION?
FOR WHOM BY WHOM AND WHAT FOR?!
HUMAN INTELLIGENCE CANNOT BE REDUCED TO AI'S DEFINITION OF HUMAN INTELLIGENCE
WHEN THE TECHNICAL NORM BECOMES THE CULTURAL NORM
'CLEARLY KNOWLEDGE AND EXPERTISE CAN TAKE FORMS THAT RESIST CODIFICATION INTO HUMAN LANGUAGE OR ANY EXPLICIT SYMBOLIC FORM'
KAPLAN


SESSION 03
THE ORIGINS OF NEURAL NETWORKS
APPROACHES TO STUDYING AI
TECHNICAL ACCOUNT THE HOW? INTERNALIST APPROACH
HISTORICAL GENEALOGY THE WHY? EXTERNALIST APPROACH
HISTORICAL NEED FOR AUTOMATION
MANUAL LABOUR
FACTORIES
MANIPULATION OF OBJECTS IN 1D
MENTAL LABOUR
AUTOMATION OF HAND CALCULATION
20th CENTURY
MANIPULATION OF SYMBOLS IN 1D
PERCEPTION
USA ARMY
RECOGNITION OF SYMBOLS IN 2D
ARTIFICIAL NEURAL NETWORK
INPUT
ARTIFICIAL NEURON
OUTPUT
WEIGHTS
SUM
ACTIVATION FUNCTION
LEARNS VIA CONDITIONING
A PERCEPTRON IS A SIMPLE YES-OR-NO DECISION BASED ON THE SUM OF ITS WEIGHTED INPUTS
THROUGH TRAINING CONNECTIONS BECOME STRONGER (WEIGHTS) OR WEAKER (WEIGHT)
TURNING
HISTORY OF NN FROM A CRITICAL PERSPECTIVE
WHO FUNDED IT?
WHY?
HOW DID IT EMERGE?
BY WHOM? WHERE?
POWER STRUCTURES
TRAINING NN TAKES A LOT OF COMPUTING POWER + ENERGY + TIME ONLY BIG COMPANIES CAN USE NN
DEEP NEURAL NETWORKS COME FROM THE AUTOMATION OF PERCEPTION
1957 ROSENBLATT


TIMELINE OF ARTIFICIAL NEURAL NETWORKS
1822-59
FAST FORWARD
ADA LOVELACE
WORLD'S FIRST PROGRAMMER
CHARLES BABBAGE
ANALYTICAL ENGINE
TO AUTOMATE LOGIC REASONING
MATHEMATICAL LANGUAGE
MECHANICAL FORM
1943
McCULLOCK & PITTS
INITIATORS OF THE IDEA OF ARTIFICIAL NEURAL NETWORKS
INSPIRED BY THE BRAIN
NEURONS AS LOGIC OPERATORS
NEURONS IN THE BRAIN ARE ORGANISED AS NETWORKS
NEURONS EXCHANGE ELECTRICAL CHARGE
1947
SUGGEST A DEVICE TO REPLACE HUMAN
TO AUTOMATE VISUAL PERCEPTION
ACKNOWLEDGING THE ANIMALS TORTURED IN THE PROCESS
MEASURING THE ELECTRICAL SIGNAL IN THE NEURONS OF THE EYE ACCORDING TO IMAGE INPUT
PATTERN RECOGNITION ALREADY HAPPENS IN THE EYE AND NOT COMPLETELY IN THE BRAIN
BUT THEY NEVER BUILT THE MACHINE
1969
MINSKY & PAPERT
KILLED FUNDING FOR NN
1971 RIP ROSENBLATT
1960
AI LABS
MIT MINSKY
STANFORD McCARTHY
THE WHOLE COMPETITION WAS FOR MINSKY
AI VS NN
1959
'MACHINE LEARNING' COINED BY ARTHUR SAMUEL
A SUB-FIELD OF NN
CREATED BY FRANK ROSENBLATT PSYCHOLOGIST CORNELL UNIV.
SPONSORED BY US ARMY
1957
THE MARK 1 PERCEPTRON
A STATISTICAL NEURAL NETWORK TO AUTOMATE VISUAL PERCEPTION AND RECOGNITION
1956
ALSO KNOWN AS SYMBOLIC AI
DIFFERENT PARADIGM THAN NN
'ARTIFICIAL INTELLIGENCE' COINED BY JOHN McCARTHY
1948
NORBERT WEINER
CYBERNETICS
1960-80
BACK-PROPAGATION ALGORITHM
H. KELLY & A. BRYSON
REBUFFED MINSKY'S SPECULATION
NN DO NOT ACTUALLY WORK
1986
CONNECTIONISM
RUMELHART + McCLELLAND
ARTIFICIAL NEURAL NETWORKS (NN)
MACHINE LEARNING
CONVOLUTIONAL NN
DEEP CONVOLUTIONAL NN (DEEP LEARNING)
1989
YANN LECUN
CONVOLUTIONAL NN
LAYER 01
LAYER 02
A GRID OF UNITS, EACH UNIT CALCULATES AN ACTIVATION VALUE
2012
ALEXNET
HINTON & TEAM


WHAT IS A DATASET?
IS A COLLECTION OF 1 IMAGES WITH ADDED 2 INFORMATION ORGANISED BY 3 TAXONOMIES
PUT INTO A FOLDER
IS A POLITICAL AND SOCIAL CONSTRUCT
1 COLLECTION OF IMAGES
EXAMPLE
MUSEUM COLLECTION
ART PIECE
INFORMATION
TITLE
DESCRIPTION
WHOSE VISION OF THE WORLD IS BEING ELEVATED IN THE CONSTRUCTION OF A DATASET?
WHO IS BEING EXPLOITED IN THE PROCESS?
CROWDSOURCING CHEAP LABOUR
NORTH SOUTH
USING YOUR PHOTOS
2 CLASSIFICATION
WHAT?
IS THE PROCESS OF ADDING INFORMATION TO AN IMAGE
WHO GETS TO DECIDE?
WHAT IS ADDED?
WHAT HIERARCHICAL FORMATIONS ARE BEING CREATED?
3 TAXONOMIES
ALL TAXONOMIES ARE POLITICAL
WHAT?
IS THE WAY TO ORGANISE THE RELATIONSHIPS BETWEEN THINGS
DATA IS NEVER NEUTRAL
DATA IS AN OXYMORON
HOW?
1950
HUMAN OPERATOR TELLING THE MACHINE IF THE ANSWER IS CORRECT
YES
NO
TODAY
?
BANANA
EXAMPLE
LABELS TO CLASSIFY HUMANS
RACE
GENDER
SEXUALITY
...
HUMAN CONSTRUCT
HISTORICAL
HIERARCHICAL FORMATIONS
GOOD
THEM
BLACK
FEMALE
GAY
SOUTH
BAD
US
WHITE
MALE
STRAIGHT
NORTH
...
RIGID + BINARY CLASSIFICATIONS
USED TO SOCIETY
VALUE
FORMATIONS OF VALUE


THE CONSTRUCTION OF A DATASET: IMAGENET
HOW ARE DATASETS CONSTRUCTED?
BY WHO?
TO WHOSE BENEFIT?
WHO IS BEING EXPLOITED IN THE PROCESS?
THE CONSTRUCTION OF DATASETS ARE COMPLEX HISTORICAL & POLITICAL CONSTRUCTS
STEPS IN THE CONSTRUCTION OF A DATASET → FROM NOOSCOPE.AI
1 PRODUCTION OF INFORMATION
2 CAPTURE
ENCODING INFORMATION INTO A DIGITAL FORMAT
3 FORMATTING
IMAGES ORGANISED ACCORDING TO A TAXONOMY
A SCHEME OF CLASSIFICATION
4 LABELLING
BANANA
THE CLASSIFICATION OF DATA INTO CATEGORIES
IMAGENET
STARTED IN 2006 BY PROF FEI FEI LI
MADE UP OF 14 MILLION HAND ANNOTATED IMAGES SPLIT INTO 20 THOUSAND CATEGORIES
IS THE MAIN DATASET USED IN DEEP LEARNING
IS THE BENCHMARK FOR IMAGE RECOGNITION ALGORITHMS
MEANING
FREE DATA
FLICKR GOOGLE IMAGE
SCRAPPING THE INTERNET TO USE YOUR PHOTOS FOR FREE!
DATA EXTRACTIVISM
NO NEED FOR CONSENT
TAXONOMY
WORDNET PRINCETON UNIVERSITY
COMPUTATIONAL TAXONOMY TO MAP THE ENGLISH LANGUAGE
CLAIMS A FAKE SCIENTIFIC NEUTRALITY AND UNIVERSALITY
ELEVATING WESTERN MODELS OF KNOWLEDGE OVER ALL OTHERS
CHEAP LABOUR
AMAZON MECHANICAL TURK
DISTRIBUTION OF MANUAL LABOUR IN POOR COUNTRIES
NORTH SOUTH
A GLOBAL DIVISION OF LABOUR
EXPLOITATION
TIMELINE
2004 FLICKR
CC
CREATIVE COMMONS LICENSE ENABLES FREE DISTRIBUTION OF IMAGES. THERE'S 6 TYPES OF LICENSES, ONE PROHIBITS ALL COMMERCIAL USE
DATA EXTRACTIVISM
ENABLES
2005 IMAGENET DATASET PROJECT STARTS
BUILT BY SCRAPPING THE INTERNET FOR FREE IMAGES
2010 IMAGENET COMPETITION BEGINS
THE LARGEST PUBLIC DATASET EVER RELEASED
2014 YFCC100M YAHOO FLICKR CREATIVE COMMONS 100 MILLION DATASET IS RELEASED
CHINA HAS BEEN THE AREA WHERE IT HAS BEEN USED THE MOST
2015 MEGAFACE DATASET RELEASED
69% OF THE IMAGES USED PROHIBIT COMMERCIAL USE
EXTRACTION AND EXPLOITATION OF YOUR PHOTOS BY CORPORATIONS AND GOVERNMENTS
MEGAPIXELS.CC

1. **A critical approach to the history of Artificial Intelligence**
The course is framed as a technical and critical introduction to Artificial Intelligence (AI) and Neural Networks (NN) where we look under the hood to see how models are constructed and ask questions such as 'What kind of data and labour do they require?' to explore the socio-political dimensions of AI&NN. In this sketch we learn that AI and NN are two different things, although most of what we call AI today in fact refers to NN.

2. **AI vs Neural networks – genealogy**
The distinction between the two is explored by looking at the genealogy of both paradigms and its key historical figures (disappointment-alert: they are all white, male and based at a US University).

3. **The origins of Neural Networks**
NN is framed within the historical need for automation of manual labour, mental labour and perception. The basic architecture of an artificial NN is introduced. A distinction is made between two approaches to studying AI&NN:
    - Technical account: the how AI&NN works
    - Historical genealogy: the why that explores the history of AI&NN from a critical perspective by asking: How did it emerge? Who funded it? Where? Why? To whose benefit? And at the cost of whom / what?

4. **Timeline of Artificial Neural Networks**
An in-depth look at the historical genealogy of Artificial NN, starting with Ada Lovelace and Charles Babbage in 1822.

5. **What is a dataset?**
Breaks down datasets into three components:
(1) collection of images; (2) classification; (3) Taxonomies.

    Under the illusion of neutrality (of which there is none), datasets could* be described as collections of images, with added information, organised through taxonomies. Yet they are so much more than that. Datasets are political and social constructs that elevate the vision of those shaping the narratives. These are built on historically rigid and binary classifications that are used to justify formations of value that create hierarchical structures of power. Data is never neutral (nothing is).

6. **The construction of a dataset: Imagenet**
Taking Imagenet as a case study to understand all the steps involved in creating a dataset.

IRL - DIGITAL LOOP
IRL
POWER IS THE NECESSARY CONDITION FOR ANY DIGITAL TECHNOLOGY.
1. PROBLEM
2. MISSION
3. HOW
STRUCTURES OF POWER EMBODIED IN THINGS
CRITICAL
PUBLIC
MAKE IT VISIBLE
INVISIBLE
PARTICIPATION
POLITICAL

**Elvia Vasconcelos** is a design researcher, wannabe activist, compulsive drawer and dressmaker. She is currently a doctoral candidate at the Technical University of Eindhoven, where she is investigating the politics of participation and accessibility. This research takes the notion of participation as 'being together' and explores what being together means in struggle,  through a praxis that creates spaces for a multitude of voices and bodies to speak and be heard.

Vasconcelos's design practice deals with the socio-political dimensions of digital technologies. Taking voice technologies as an object to critically explore the field of Artificial Intelligence, she created the 'Feminist Alexa' project in 2017 – a series of workshops that critically look at Personal Intelligent Assistants e.g. Amazon Alexa, to investigate the ways in which gender is used in technology and the connections to gender-based discrimination in real life.  Her critical investigations of AI have been articulated in a number of different settings such as in Alexa Diaries, in the Feminist Voices in Tecnology publication and at a number of events such as the Mozilla Festival, The air of turbulence and Primer Conference.

In her critical investigation of AI Vasconcelos has used sketching as a way to render complexity more accessible.

# Excavating AI: The Politics of Images in Machine Learning Training Sets

Kate Crawford and Trevor Paglen

You open up a database of pictures used to train artificial intelligence systems. At first, things seem straightforward. You're met with thousands of images: apples and oranges, birds, dogs, horses, mountains, clouds, houses and street signs. But as you probe further into the dataset, people begin to appear: cheerleaders, scuba divers, welders, Boy Scouts, fire walkers and flower girls. Things get strange: a photograph of a woman smiling in a bikini is labelled a 'slattern, slut, slovenly woman, trollop'. A young man drinking beer is categorized as an 'alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse'. A child wearing sunglasses is classified as a 'failure, loser, non-starter, unsuccessful person'. You're looking at the 'person' category in a dataset called ImageNet, one of the most widely used training sets for machine learning.

Something is wrong with this picture.

Where did these images come from? Why were the people in the photos labelled this way? What sorts of politics are at work when pictures are paired with labels, and what are the implications when they are used to train technical systems?

In short, how did we get here?

There's an urban legend about the early days of machine vision, the subfield of artificial intelligence (AI) concerned with teaching machines to detect and interpret images. In 1966, Marvin Minsky was a young professor at MIT, making a name for himself in the emerging field of artificial intelligence.[1] Deciding that the ability to interpret images was a core feature of intelligence, Minsky turned to an undergraduate student, Gerald Sussman, and asked him to 'spend the summer linking a camera to a computer and getting the computer to describe what it saw', [2] This became the Summer Vision Project. [3] Needless to say, the project of getting computers to 'see' was much harder than anyone expected, and would take a lot longer than a single summer.

The story we've been told goes like this: brilliant men worked for decades on the problem of computer vision, proceeding in fits and starts, until the turn to probabilistic modelling and learning techniques in the 1990s accelerated progress. This led to the current moment, in which challenges such as object detection and facial recognition have been largely solved. [4] This arc of inevitability recurs in many AI narratives, where it is assumed that ongoing technical improvements will resolve all problems and limitations.

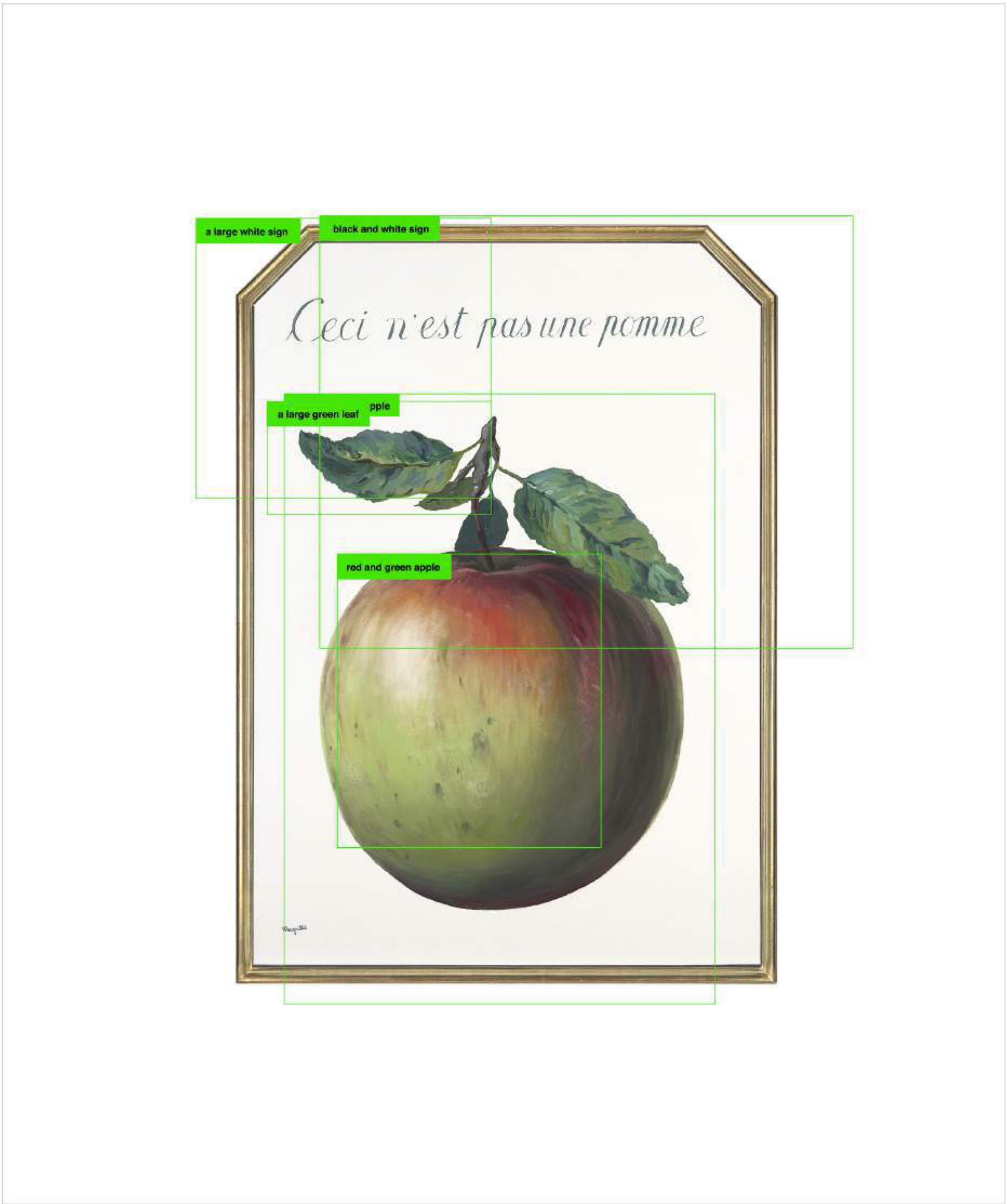
But what if the opposite is true? What if the challenge of getting computers to 'describe what they see' will always be a problem? In this essay, we will explore why the automated interpretation of images is an inherently social and political project, rather than a purely technical one. Understanding the politics within AI systems matters more than ever, as they are quickly moving into the architecture of social institutions: deciding whom to interview for a job, which students are paying attention in class, which suspects to arrest, and much else.

For the last two years, we have been studying the underlying logic of how images are used to train AI systems to ‘see’ the world. We have looked at hundreds of collections of images used in artificial intelligence, from the first experiments with facial recognition in the early 1960s to contemporary training sets containing millions of images. Methodologically, we could call this project an *archeology of datasets*: we have been digging through the material layers, cataloguing the principles and values by which something was constructed, and analysing what normative patterns of life were assumed, supported and reproduced. By excavating the construction of these training sets and their underlying structures, many unquestioned assumptions are revealed. These assumptions inform the way AI systems work - and fail - to this day.

This essay begins with a deceptively simple question: what work do images do in AI systems? What are computers meant to recognise in an image and what is misrecognised or even completely invisible? Next, we look at the method for introducing images into computer systems and look at how taxonomies order the foundational concepts that will become intelligible to a computer system. Then we turn to the question of labelling: how do humans tell computers which words will relate to a given image? And what is at stake in the way AI systems use these labels to classify humans, including by race, gender, emotions, ability, sexuality, and personality? Finally, we turn to the purposes that computer vision is meant to serve in our society - the judgments, choices, and consequences of providing computers with these capacities.

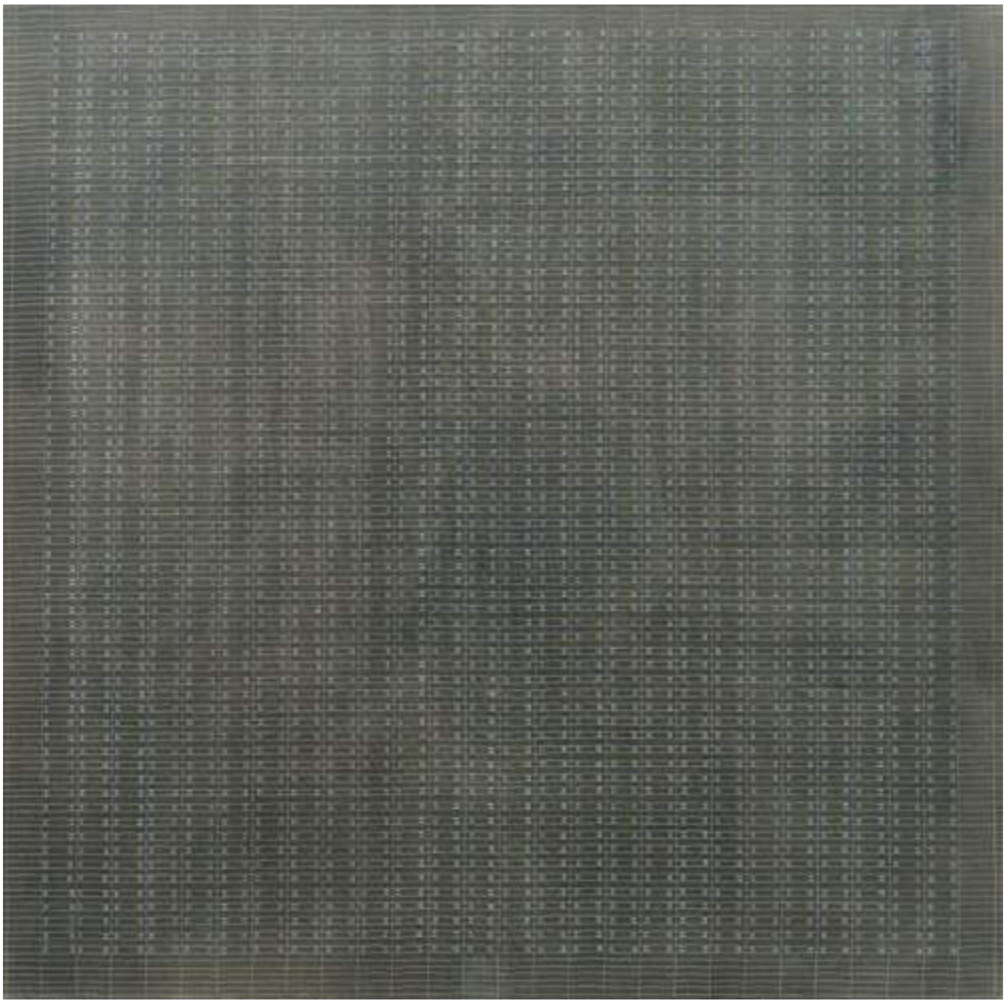
## Training AI

Building AI systems requires data. Supervised machine-learning systems designed for object or facial recognition are trained on vast amounts of data contained within datasets made up of many discrete images. To build a computer vision system that can, for example, recognise the difference between pictures of apples and oranges, a developer has to collect, label and train a neural network on thousands of labelled images of apples and oranges. On the software side, the algorithms conduct a statistical survey of the images, and develop a model to recognise the difference between the two ‘classes.’ If all goes according to plan, the trained model will be able to distinguish the difference between images of apples and oranges that it has never encountered before.



Training sets, then, are the foundation on which contemporary machine- learning systems are built.[5] They are central to how AI systems recognize and interpret the world. These datasets shape the epistemic boundaries governing how AI systems operate, and thus are an essential part of understanding socially significant questions about AI.

But when we look at the training images widely used in computer-vision systems, we find a bedrock composed of shaky and skewed assumptions. For reasons that are rarely discussed within the field of computer vision, and despite all that institutions like MIT and companies like Google and Facebook have done, the project of interpreting images is a profoundly complex and relational endeavour. Images are remarkably slippery things, laden with multiple potential meanings, irresolvable questions, and contradictions. Entire subfields of philosophy, art history, and media theory are dedicated to teasing out all the nuances of the unstable relationship between images and meanings.[6]



"White Flower" Agnes Martin, 1960

Images do not describe themselves. This is a feature that artists have explored for centuries. Agnes Martin creates a grid-like painting and dubs it *White Flower*, Magritte paints a picture of an apple with the words ‘This is not an apple’. We see those images differently when we see how they’re labelled. The circuit between image, label and referent is flexible and can be reconstructed in any number of ways to do different kinds of work. What’s more, those circuits can change over time as the cultural context of an image shifts, and can mean different things depending on who looks, and where they are located. Images are open to interpretation and reinterpretation.

This is part of the reason why the tasks of object recognition and classification are more complex than Minksy - and many of those who have come since - initially imagined.

Despite the common mythos that AI and the data it draws on are objectively and scientifically classifying the world, everywhere there is politics, ideology, prejudices and all of the subjective stuff of history. When we survey the most widely used training sets, we find that this is the rule rather than the exception.

**Anatomy of a Training Set**

Although there can be considerable variation in the purposes and architectures of different training sets, they share some common properties. At their core, training sets for imaging systems consist of a collection of images that have been labelled in various ways and sorted into categories. As such, we can describe their overall architecture as generally consisting of three layers: the overall taxonomy (the aggregate of classes and their hierarchical nesting, if applicable), the individual classes (the singular categories that images are organised into, e.g., 'apple'), and each individually labelled image (i.e., an individual picture that has been labelled an apple). Our contention is that every layer of a given training set's architecture is infused with politics.

Take the case of a dataset like the 'The Japanese Female Facial Expression (JAFFE) Database', developed by Michael Lyons, Miyuki Kamachi and Jiro Gyoba in 1998, and widely used in affective computing research and development. The dataset contains photographs of ten Japanese female models making seven facial expressions that are meant to correlate with seven basic emotional states.[7] (The intended purpose of the dataset is to help machine-learning systems recognise and label these emotions for newly captured, unlabelled images). The implicit, top-level taxonomy here is something like 'facial expressions depicting the emotions of Japanese women'.

If we go down a level from taxonomy, we arrive at the level of the class. In the case of JAFFE, those classes are happiness, sadness, surprise, disgust, fear, anger and neutral. These categories become the organising buckets into which all of the individual images are stored. In a database used in facial recognition, as another example, the classes might correspond to the names of the individuals whose faces are in the dataset. In a dataset designed for object recognition, those classes correspond to things like apples and oranges. They are the distinct concepts used to order the underlying images.

At the most granular level of a training set's architecture, we find the individual labelled image: be it a face labelled as indicating an emotional state; a specific person; or a specific object, among many examples. For JAFFE, this is where you can find an individual woman grimacing, smiling or looking surprised.

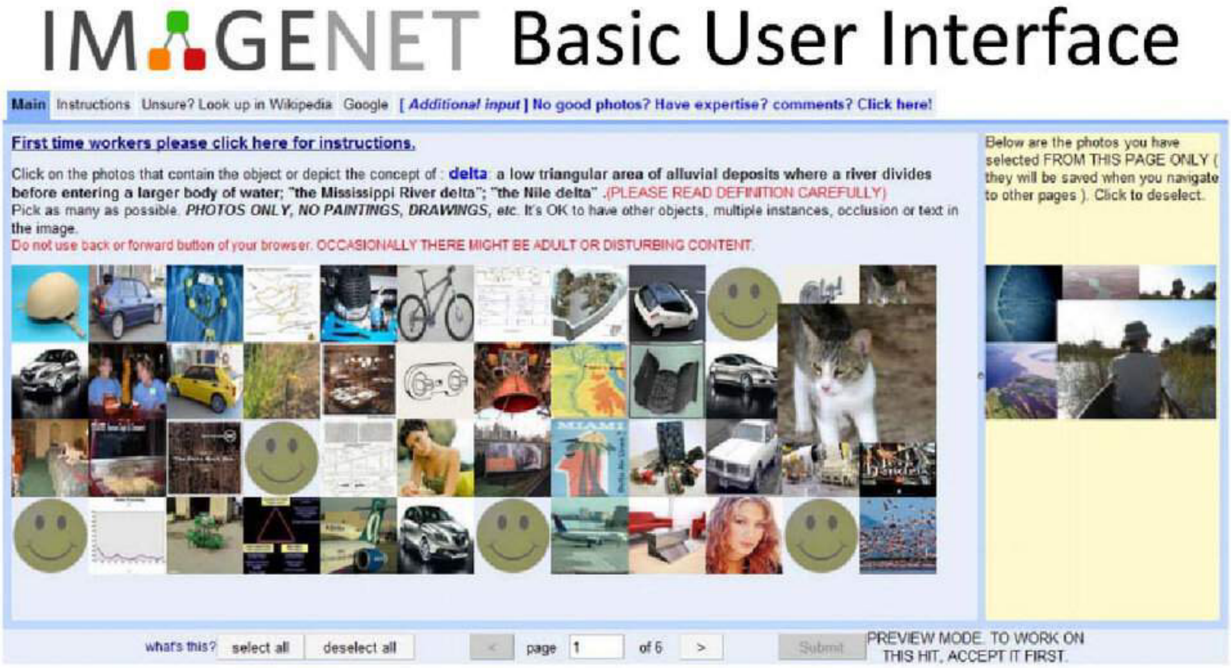
There are several implicit assertions in the JAFFE set. First there's the taxonomy itself: that 'emotions' is a valid set of visual concepts. Then there's a string of additional assumptions: that the concepts within 'emotions' can be applied to photographs of people's faces (specifically Japanese women); that there are six emotions plus a neutral state; that there is a fixed relationship between a person's facial expression and her true emotional state; and that this relationship between the face and the emotion is consistent, measurable, and uniform across the women in the photographs.

At the level of the class, we find assumptions such as 'there is such a thing as a "neutral" facial expression' and 'the significant six emotional states are happy, sad, angry, disgusted, afraid, surprised'.[8] At the level of labelled image, there are other implicit assumptions such as 'this particular photograph depicts a woman with an "angry" facial expression', rather than, for example, the fact that this is an image of a woman mimicking an angry expression. These, of course, are all 'performed' expressions - not relating to any interior state, but acted out in a laboratory setting. Every one of the implicit claims made at each level is, at best, open to question, and some are deeply contested.[9]

The JAFFE training set is relatively modest as far as contemporary training sets go. It was created before the advent of social media, before developers were able to scrape images from the internet at scale, and before piecemeal online labour platforms like Amazon Mechanical Turk allowed researchers and corporations to conduct the formidable task of labeling huge quantities of photographs. As training sets grew in scale and scope, so did the complexities, ideologies, semiologies and politics from which they are constituted. To see this at work, let's turn to the most iconic training set of all, ImageNet.

### The Canonical Training Set: ImageNet

One of the most significant training sets in the history of AI so far is ImageNet, which is now celebrating its tenth anniversary. First presented as a research poster in 2009, ImageNet is a dataset of extraordinary scope and ambition. In the words of its cocreator, Stanford Professor Fei-Fei Li, the idea behind ImageNet was to 'map out the entire world of objects'.[10] Over several years of development, ImageNet grew enormous: the development team scraped a collection of many millions of images from the internet and briefly became the world's largest academic user of Amazon's Mechanical Turk, using an army of piecemeal workers to sort an average of 50 images per minute into thousands of categories.[11] When it was finished, ImageNet consisted of over 14 million labelled images organised into more than 20, 000 categories. For a decade, it has been the colossus of object recognition for machine learning and a powerfully important benchmark for the field.


IMAGENET Basic User Interface
Main
Instructions
Unsure? Look up in Wikipedia
Google
[ Additional input ] No good photos? Have expertise? comments? Click here!
First time workers please click here for instructions.
what's this?
select all
deselect all
<
page
1
of 6
>
Submit

Click on the photos that contain the object or depict the concept of : **delta** : **a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta"** .(PLEASE READ DEFINITION CAREFULLY)
Pick as many as possible. ***PHOTOS ONLY, NO PAINTINGS, DRAWINGS, etc.*** It's OK to have other objects, multiple instances, occlusion or text in the image.
Do not use back or forward button of your browser. OCCASIONALLY THERE MIGHT BE ADULT OR DISTURBING CONTENT.

Below are the photos you have selected FROM THIS PAGE ONLY ( they will be saved when you navigate to other pages ). Click to deselect.

PREVIEW MODE. TO WORK ON THIS HIT, ACCEPT IT FIRST.

Interface used by Amazon Turk Workers to label pictures in ImageNet

Navigating ImageNet's labyrinthine structure is like taking a stroll through Borges's infinite library. It is vast and filled with all sorts of curiosities. There are categories for apples, apple aphids, apple butter, apple dumplings, apple geraniums, apple jelly, apple juice, apple maggots, apple rust, apple trees, apple turnovers, apple carts, applejack, and applesauce. There are pictures of hot lines, hot pants, hot plates, hot pots, hot rods, hot sauce, hot springs, hot toddies, hot tubs, hot- air balloons, hot fudge sauce, and hot water bottles.

ImageNet quickly became a critical asset for computer-vision research. It became the basis for an annual competition where labs around the world would try to outperform each other by pitting their algorithms against the training set and seeing which one could most accurately label a subset of images. In 2012, a team from the University of Toronto used a Convolutional Neural Network to handily win the top prize, bringing new attention to this technique. That moment is widely considered a turning point in the development of contemporary AI.[12] The final year of the ImageNet competition was 2017, and accuracy in classifying objects in the limited subset had risen from 71.8% to 97.3%. That subset did not include the 'Person' category, for reasons that will soon become obvious.

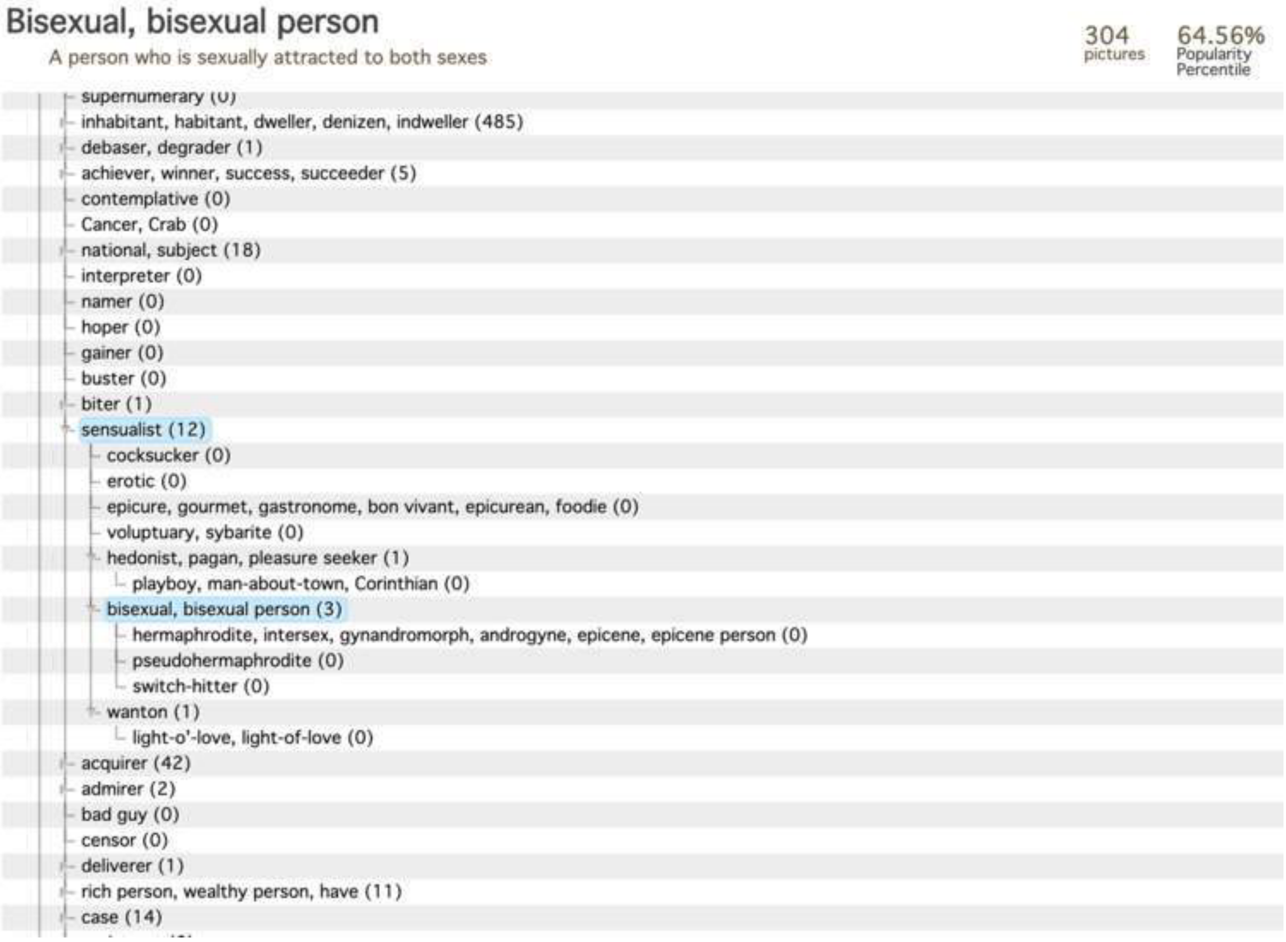
## Taxonomy

The underlying structure of ImageNet is based on the semantic structure of WordNet, a database of word classifications developed at Princeton University in the 1980s. The taxonomy is organised according to a nested structure of cognitive synonyms or 'synset'. Each 'synset' represents a distinct concept, with synonyms grouped together (for example, 'auto' and 'car' are treated as belonging to the same synset). Those synsets are then organised into a nested hierarchy, going from general concepts to more specific ones. For example, the concept 'chair' is nested as artifact > furnishing > furniture > seat > chair. The classification system is broadly similar to those used in libraries to order books into increasingly specific categories.

While WordNet attempts to organize the entire English language,[13] ImageNet is restricted to nouns (the idea being that nouns are things that pictures can represent). In the ImageNet hierarchy, every concept is organised under one of nine top-level categories: plant, geologic formation, natural object, sport, artifact, fungus, person, animal and miscellaneous. Below these are layers of additional nested classes.

As the fields of information science and science and technology studies have long shown, all taxonomies or classificatory systems are political.[14] In ImageNet (inherited from WordNet),

for example, the category 'human body' falls under the branch Natural Object > Body > Human Body. Its subcategories include 'male body'; 'person'; 'juvenile body'; 'adult body'; and 'female body'. The 'adult body' category contains the subclasses 'adult female body' and 'adult male body'. We find an implicit assumption here: only 'male' and 'female' bodies are 'natural'. There is an ImageNet category for the term 'Hermaphrodite' that is bizarrely (and offensively) situated within the branch Person > Sensualist > Bisexual > alongside the categories 'Pseudohermaphrodite' and 'Switch Hitter'.[15] The ImageNet classification hierarchy recalls the old Library of Congress classification of LGBTQ-themed books under the category 'Abnormal Sexual Relations, Including Sexual Crimes', which the American Library Association's Task Force on Gay Liberation finally convinced the Library of Congress to change in 1972 after a sustained campaign.[16]



Bisexual, bisexual person

A person who is sexually attracted to both sexes

304 pictures

64.56% Popularity Percentile

```
supernumerary (0)
inhabitant, habitant, dweller, denizen, indweller (485)
debaser, degrader (1)
achiever, winner, success, succeeder (5)
contemplative (0)
Cancer, Crab (0)
national, subject (18)
interpreter (0)
namer (0)
hoper (0)
gainer (0)
buster (0)
biter (1)
sensualist (12)
  cocksucker (0)
  erotic (0)
  epicure, gourmet, gastronome, bon vivant, epicurean, foodie (0)
  voluptuary, sybarite (0)
  hedonist, pagan, pleasure seeker (1)
    playboy, man-about-town, Corinthian (0)
  bisexual, bisexual person (3)
    hermaphrodite, intersex, gynandromorph, androgyne, epicene, epicene person (0)
    pseudohermaphrodite (0)
    switch-hitter (0)
  wanton (1)
    light-o'-love, light-of-love (0)
acquirer (42)
admirer (2)
bad guy (0)
censor (0)
deliverer (1)
rich person, wealthy person, have (11)
case (14)
```

If we move from taxonomy down a level, to the 21,841 categories in the ImageNet hierarchy, we see another kind of politics emerge.

**Categories**

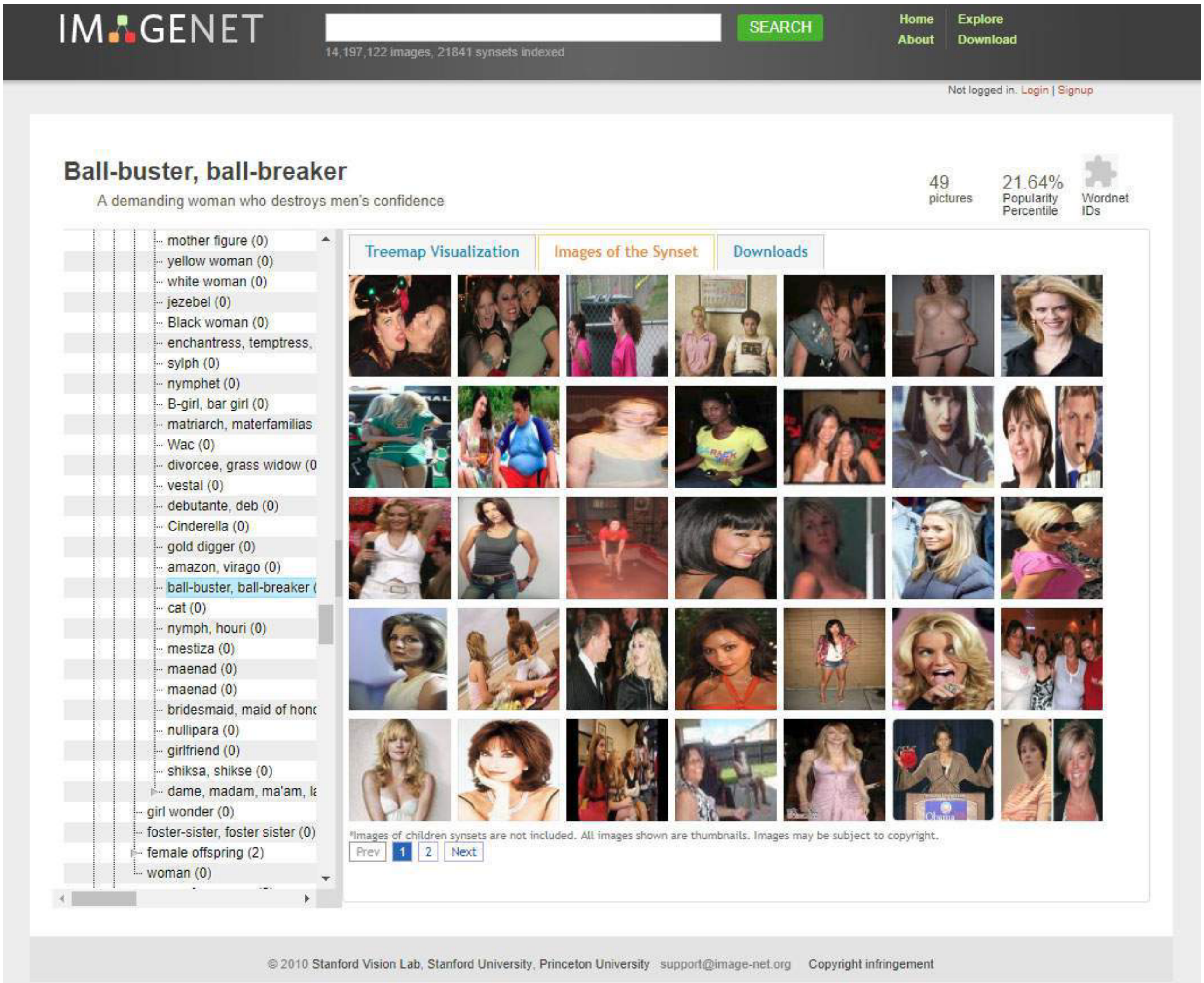
There's a kind of sorcery that goes into the creation of categories. To create a category or to name things is to divide an almost infinitely complex universe into separate phenomena. To impose order onto an undifferentiated mass, to ascribe phenomena to a category - that is, to name a thing - is in turn a means of reifying the existence of that category.

In the case of ImageNet, noun categories such as ‘apple’ or ‘apple butter’ might seem reasonably uncontroversial, but not all nouns are created equal. To borrow an idea from linguist George Lakoff, the concept of an “apple” is more nouny than the concept of ‘light’, which in turn is more nouny than a concept such as ‘health’.[17] Nouns occupy various places on an axis from the concrete to the abstract, and from the descriptive to the judgmental. These gradients have been erased in the logic of ImageNet. Everything is flattened out and pinned to a label, like taxidermy butterflies in a display case. The results can be problematic, illogical, and cruel, especially when it comes to labels applied to people.

ImageNet contains 2,833 subcategories under the top-level category ‘Person’. The subcategory with the most associated pictures is ‘gal’ (with 1,664 images) followed by ‘grandfather’ (1,662), ‘dad’ (1,643), and chief executive officer (1,614). With these highly populated categories, we can already begin to see the outlines of a worldview. ImageNet classifies people into a huge range of types including race, nationality, profession, economic status, behaviour, character and even morality. There are categories for racial and national identities including Alaska Native, Anglo-American, Black, Black African, Black Woman, Central American, Eurasian, German

American, Japanese, Lapp, Latin American, Mexican-American, Nicaraguan, Nigerian, Pakistani, Papuan, South American Indian, Spanish American, Texan, Uzbek, White, Yemeni and Zulu. Other people are labelled by their careers or hobbies: there are Boy Scouts, cheerleaders, cognitive neuroscientists, hairdressers, intelligence analysts, mythologists, retailers, retirees and so on.

As we go further into the depths of ImageNet’s Person categories, the classifications of humans within it take a sharp and dark turn. There are categories for Bad Person, Call Girl, Drug Addict, Closet Queen, Convict, Crazy, Failure, Flop, Fucker, Hypocrite, Jezebel, Kleptomaniac, Loser, Melancholic, Nonperson, Pervert, Prima Donna, Schizophrenic, Second- Rater, Spinster, Streetwalker, Stud, Tosser, Unskilled Person, Wanton, Waverer and Wimp. There are many racist slurs and misogynistic terms.



Selections from the "Person" classes, ImageNet

Of course, ImageNet was typically used for object recognition - so the Person category was rarely discussed at technical conferences, nor has it received much public attention. However, this complex architecture of images of real people, tagged with often offensive labels, has been publicly available on the internet for a decade. It provides a powerful and important example of the complexities and dangers of human classification, and the sliding spectrum from supposedly unproblematic labels like ‘trumpeter’ or ‘tennis player’ to concepts like ‘spastic’, ‘mulatto’, or ‘redneck’. Regardless of the supposed neutrality of any particular category, the selection of images skews the meaning in ways that are gendered, racialised, ableist and ageist. ImageNet is an object lesson, if you will, in what happens when people are categorised like objects. And this practice has only become more common in recent years, often inside the big AI companies, where there is no way for outsiders to see how images are being ordered and classified.

Finally, there is the issue of where the thousands of images in ImageNet’s Person class were drawn from. By harvesting images en masse from image search engines like Google, ImageNet’s creators appropriated people’s selfies and vacation photos without their knowledge, and then labelled and repackaged them as the underlying data for much of an entire field.[18] When we take a look at the bedrock layer of labeled images, we find highly questionable semiotic assumptions, echoes of nineteenth- century phrenology, and the representational harm of classifying images of people without their consent or participation.

**ImageNet Roulette: An Experiment in Classification**

The ImageNet dataset is typically used for object recognition. But as part of our archeological method, we were interested to see what would happen if we trained an AI model exclusively on its ‘person’ categories. The result of that experiment is ImageNet Roulette.

ImageNet Roulette uses an open-source Caffe deep- learning framework (produced at UC Berkeley) trained on the images and labels in the ‘person’ categories (which are currently ‘down for maintenance’). Proper nouns were removed.

When a user uploads a picture, the application first runs a face detector to locate any faces. If it finds any, it sends them to the Caffe model for classification. The application then returns the original images with a bounding box showing the detected face and the label the classifier has assigned to the image. If no faces are detected, the application sends the entire scene to the Caffe model and returns an image with a label in the upper left corner.

As we have shown, ImageNet contains a number of problematic, offensive and bizarre categories. Hence, the results ImageNet Roulette returns often draw upon those categories. That is by design: we want to shed light on what happens when technical systems are trained using problematic training data. AI classifications of people are rarely made visible to the people being classified. ImageNet Roulette provides a glimpse into that process - and to show how things can go wrong.

ImageNet Roulette does not store the photos people upload.

https://imagenet- roulette.paglen.com/

### Labelled Images

Images are laden with potential meanings, irresolvable questions and contradictions. In trying to resolve these ambiguities, ImageNet’s labels often compress and simplify images into deadpan banalities. One photograph shows a dark-skinned toddler wearing tattered and dirty clothes and clutching a soot-stained doll. The child’s mouth is open. The image is completely devoid of context. Who is this child? Where is it? The photograph is simply labeled ‘toy’.

But some labels are just nonsensical. A woman sleeps in an airplane seat, her right arm protectively curled around her pregnant stomach. The image is labeled ‘snob’. A photoshopped picture shows a smiling Barack Obama wearing a Nazi uniform, his arm raised and holding a Nazi flag. It is labeled ‘Bolshevik’.

At the image layer of the training set, like everywhere else, we find assumptions, politics and worldviews. According to ImageNet, for example, Sigourney Weaver is a ‘hermaphrodite’, a young man wearing a straw hat is a ‘tosser’, and a young woman lying on a beach towel is a ‘kleptomaniac’. But the worldview of ImageNet isn’t limited to the bizarre or derogatory conjoining of pictures and labels.

Other assumptions about the relationship between pictures and concepts recall physiognomy, the pseudoscientific assumption that something about a person’s essential character can be gleaned by observing features of their body and face. ImageNet takes this to an extreme, assuming that whether someone is a ‘debtor’, a ‘snob’, a ‘swinger’, or a ‘slav’ can be determined by inspecting their photograph. In the weird metaphysics of ImageNet, there are separate image categories for ‘assistant professor’ and ‘associate professor’ - as though if someone were to get a promotion, their biometric signature would reflect the change in rank.

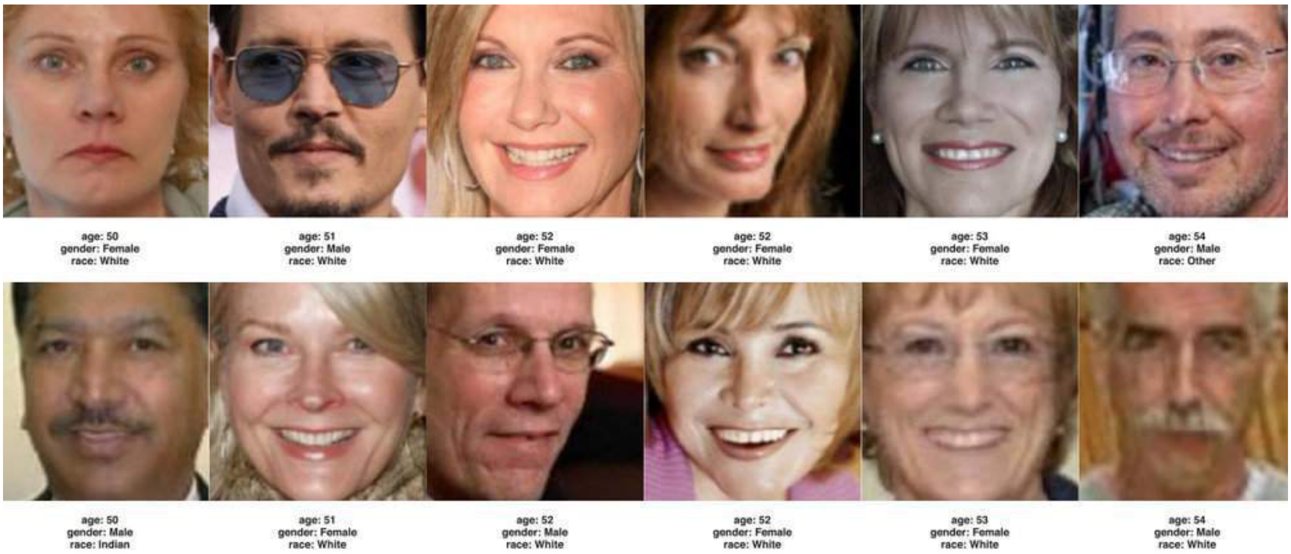
Of course, these sorts of assumptions have their own dark histories and attendant politics.

### UTK: Making Race and Gender from Your Face

In 1839, the mathematician Franc?ois Arago claimed that through photographs, ‘objects preserve mathematically their forms’.[19] Placed into the nineteenth-century context of imperialism and social Darwinism, photography helped to animate - and lend a ‘scientific’ veneer to - various forms of phrenology, physiognomy, and eugenics.[20] Physiognomists such as Francis Galton and Cesare Lombroso created composite images of criminals, studied the feet of prostitutes, measured skulls and compiled meticulous archives of labelled images and measurements, all in an effort to use ‘mechanical’ processes to detect visual signals in classifications of race, criminality and deviance from bourgeois ideals. This was done to capture and pathologise what was seen as deviant or criminal behaviour, and make such behaviour observable in the world.

And as we shall see, not only have the underlying assumptions of physiognomy made a comeback with contemporary training sets, but a number of training sets are designed to use algorithms and facial landmarks as latter-day calipers to conduct contemporary versions of craniometry.

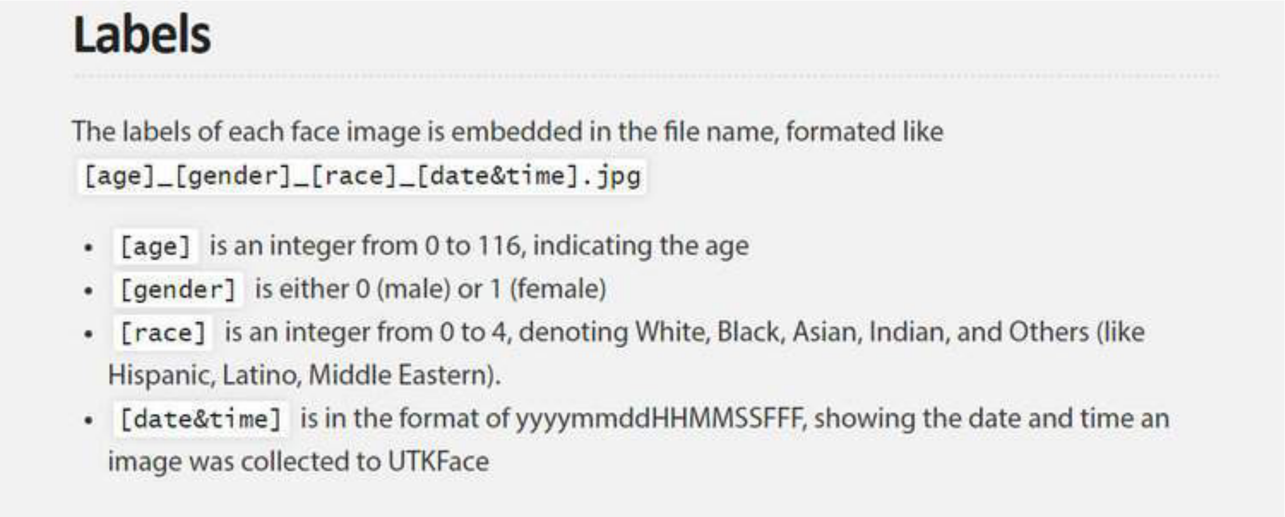
For example, the UTKFace dataset (produced by a group at the University of Tennessee at Knoxville) consists of over 20,000 images of faces with annotations for age, gender and race. The dataset's authors state that the dataset can be used for a variety of tasks, like automated face detection, age estimation and age progression.[21]



UTKFace Dataset

The annotations for each image include an estimated age for each person, expressed in years from zero to 116. Gender is a binary choice: either zero for male or one for female. Second, race is categorised from zero to four, and places people in one of five classes: White, Black, Asian, Indian, or 'Others'.

The politics here are as obvious as they are troubling. At the category level, the researchers' conception of gender is as a simple binary structure, with 'male' and 'female' the only alternatives. At the level of the image label is the assumption that someone's gender identity can be ascertained through a photograph.



## Labels

The labels of each face image is embedded in the file name, formated like `[age]_[gender]_[race]_[date&time].jpg`

- `[age]` is an integer from 0 to 116, indicating the age
- `[gender]` is either 0 (male) or 1 (female)
- `[race]` is an integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).
- `[date&time]` is in the format of yyyymmddHHMMSSFFF, showing the date and time an image was collected to UTKFace

Classification scheme used in UTKFace Dataset

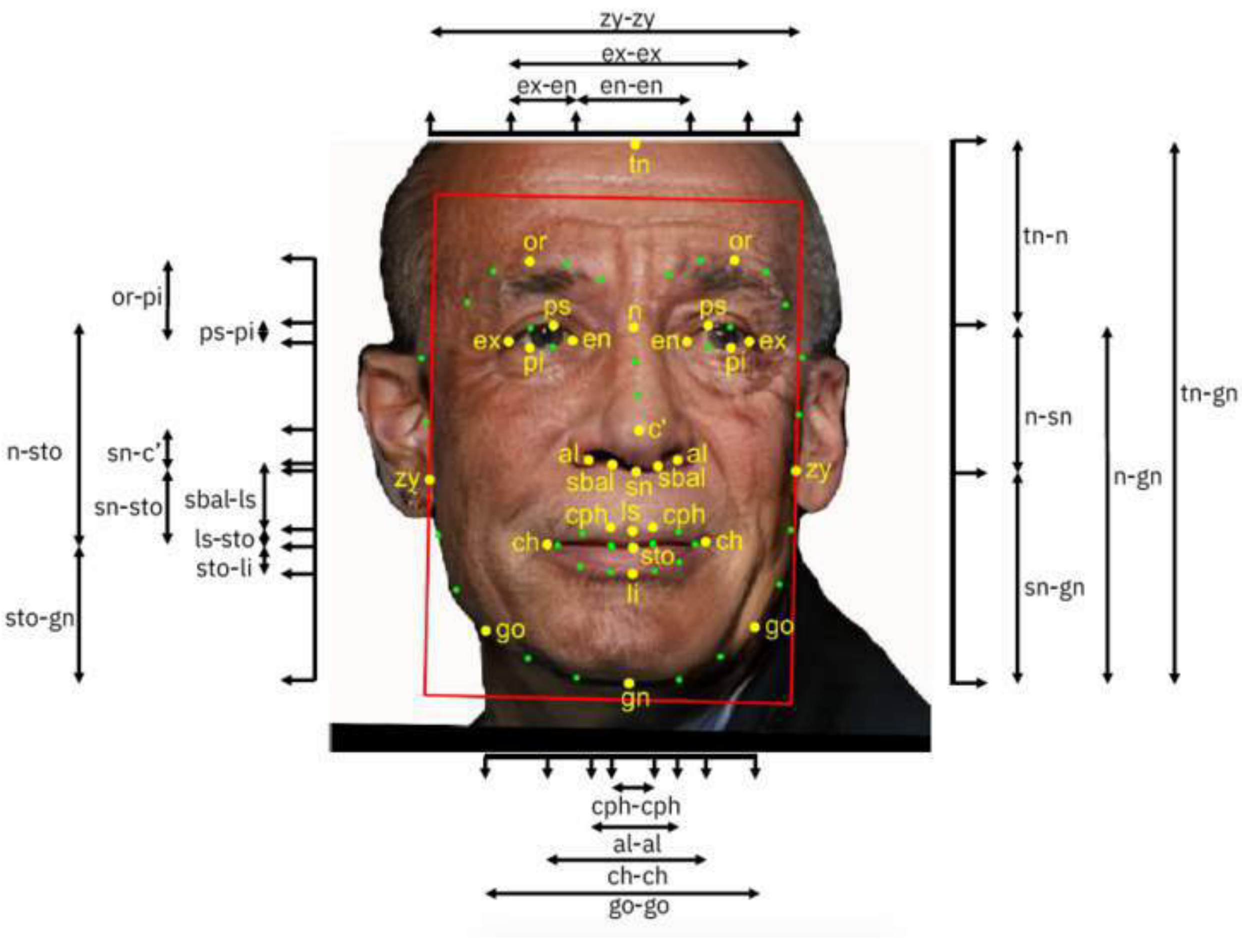
The classificatory schema for race recalls many of the deeply problematic racial classifications of the twentieth century. For example, the South African apartheid regime sought to classify the entire population into four categories: Black, White, Coloured, or Indian.[22] Around 1970, the South African government created a unified ‘identity passbook’ called The Book of Life, which linked to a centrally managed database created by IBM. These classifications were based on dubious and shifting criteria of ‘appearance and general acceptance or repute’, and many people were reclassified, sometimes multiple times. [23] The South African system of racial classification was intentionally very different from the American ‘one-drop’ rule, which stated that even one ancestor of African descent made somebody Black, likely because nearly all white South Africans had some traceable black African ancestry.[24] Above all, these systems of classifications caused enormous harm to people, and the elusive classifier of a pure ‘race’ signifier was always in dispute. However, seeking to improve matters by producing ‘more diverse’ AI training sets presents its own complications.

### IBM’S Diversity in Faces

IBM’s ‘Diversity in Faces’ dataset was created as a response to critics who had shown that the company’s facial-recognition software often simply did not recognise the faces of people with darker skin.[25] IBM publicly promised to improve their facial-recognition datasets to make them more ‘representative’ and published the ‘Diversity in Faces’ (DiF) dataset as a result.[26] Constructed to be ‘a computationally practical basis for ensuring fairness and accuracy in face recognition’, the DiF consists of almost a million images of people pulled from the Yahoo! Flickr Creative Commons dataset, assembled specifically to achieve statistical parity among categories of skin tone, facial structure, age and gender.[27]

The dataset itself continued the practice of collecting hundreds of thousands of images of unsuspecting people who had uploaded pictures to sites like Flickr.[28] But the dataset contains a unique set of categories not previously seen in other face-image datasets. The IBM DiF team asks whether age, gender and skin colour are truly sufficient in generating a dataset that can ensure fairness and accuracy and concludes that even more classifications are needed. So they move into truly strange territory: including facial symmetry and skull shapes to build a complete picture of the face. The researchers claim that the use of craniofacial features is justified because it captures much more granular information about a person's face than just gender, age and skin colour alone. The paper accompanying the dataset specifically highlights prior work done to show that skin colour is itself a weak predictor of race, but this begs the question of why moving to skull shapes is appropriate.

Craniometry was a leading methodological approach of biological determinism during the nineteenth century. As Stephen Jay Gould shows in his book *The Mismeasure of Man*, skull size was used by nineteenth- and twentieth-century pseudoscientists as a spurious way to claim inherent superiority of white people over black people, and different skull shapes and weights were said to determine people’s intelligence - always along racial lines.[29]



IBM's Diversity in Faces

While the efforts of companies to build more diverse training sets is often put in the language of increasing 'fairness' and 'mitigating bias', clearly there are strong business imperatives to produce tools that will work more effectively across wider markets. However, here too the technical process of categorising and classifying people is shown to be a political act. For example, how is a 'fair' distribution achieved within the dataset?

IBM decided to use a mathematical approach to quantifying 'diversity' and 'evenness', so that a consistent measure of evenness exists throughout the dataset for every feature quantified. The dataset also contains subjective annotations for age and gender, which are generated using three independent Amazon Turk workers for each image, similar to the methods used by ImageNet.[30] So people's gender and age are being 'predicted' based on three clickworkers' guesses about what's shown in a photograph scraped from the internet. It harkens back to the early carnival game of 'Guess Your Weight!', with similar levels of scientific validity.

Ultimately, beyond these deep methodological concerns, the concept and political history of diversity is being drained of its meaning and left to refer merely to expanded biological

phenotyping. Diversity in this context just means a wider range of skull shapes and facial symmetries. For computer vision researchers, this may seem like a 'mathematization of fairness', but it simply serves to improve the efficiency of surveillance systems. And even after all these attempts at expanding the ways which people are classified, the Diversity in Faces set still relies on a binary classification for gender: people can only be labelled male or female. Achieving parity amongst different categories is not the same as achieving diversity or fairness, and IBM's data construction and analysis perpetuates a harmful set of classifications within a narrow worldview.

### Epistemics of Training Sets

What are the assumptions undergirding visual AI systems? First, the underlying theoretical paradigm of the training sets assumes that concepts - whether 'corn', 'gender', 'emotions' or 'losers' - exist in the first place, and that those concepts are fixed, universal, and have some sort of transcendental grounding and internal consistency. Second, it assumes a fixed and universal correspondence between images and concepts, appearances and essences. What's more, it assumes uncomplicated, self- evident and measurable ties between images, referents and labels. In other words, it assumes that different concepts - whether 'corn' or 'kleptomaniacs' - have some kind of essence that unites each instance of them, and that that underlying essence expresses itself visually. Moreover, the theory goes, that visual essence is discernible by using statistical methods to look for formal patterns across a collection of labeled images. Images of people dubbed 'losers', the theory goes, contain some kind of visual pattern that distinguishes them from, say, 'farmers', 'assistant professors', or, for that matter, apples. Finally, this approach assumes that all concrete nouns are created equally, and that many abstract nouns also express themselves concretely and visually (i.e., 'happiness' or 'anti- Semitism).

The training sets of labelled images that are ubiquitous in contemporary computer vision and AI are built on a foundation of unsubstantiated and unstable epistemological and metaphysical assumptions about the nature of images, labels, categorisation and representation. Furthermore, those epistemological and metaphysical assumptions hark back to historical approaches where people were visually assessed and classified as a tool of oppression and race science.

Datasets aren't simply raw materials to feed algorithms, but are political interventions. As such, much of the discussion around 'bias' in AI systems misses the mark: there is no 'neutral', 'natural', or 'apolitical' vantage point that training data can be built upon. There is no easy technical 'fix' by shifting demographics, deleting offensive terms, or seeking equal representation by skin tone. The whole endeavour of collecting images, categorising them, and labelling them is itself a form of politics, filled with questions about who gets to decide what images mean and what kinds of social and political work those representations perform.

### Missing Persons

In January 2019, images in ImageNet's 'Person' category began disappearing. Suddenly, 1.2 million photos were no longer accessible on Stanford University's servers. Gone were the pictures of cheerleaders, scuba divers, welders, altar boys, retirees and pilots. The picture of a man drinking beer characterised as an 'alcoholic' disappeared, as did the pictures of a woman in a bikini dubbed a 'slattern' and a young boy classified as a 'loser'. The picture of a man eating a sandwich (labelled a 'selfish person') met the same fate. When you search for these images, the ImageNet website responds with a statement that it is under maintenance, and only the categories used in the ImageNet competition are still included in the search results.

But once it came back online, the search functionality on the site was modified so that it would only return results for categories that had been included in ImageNet's annual computer-vision contest. As of this writing, the 'Person' category is still browsable from the data set's online interface, but the images fail to load. The URLs for the original images are still downloadable.[31]

Over the next few months, other image collections used in computer-vision and AI research also began to disappear. In response to research published by Adam Harvey and Jules LaPlace,[32] Duke University took down a massive photo repository of surveillance-camera footage of students attending classes (called the Duke Multi-Target, Multi-Camera [MTMC] dataset). It turned out that the authors of the dataset had violated the terms of their Institutional Review Board approval by collecting images from people in public space, and by making their dataset publicly available. [33]

Similar datasets created from surveillance footage disappeared from servers at the University of

Colorado Colorado Springs, and more from Stanford University, where a collection of faces culled from a webcam installed at San Francisco's iconic Brainwash Cafe was 'removed from access at the request of the depositor'.[34]

By early June, Microsoft had followed suit, removing their landmark "MS-CELEB" collection of approximately ten million photos from 100,000 people scraped from the internet in 2016. It was the largest public facial- recognition dataset in the world, and the people included were not just famous actors and politicians, but also journalists, activists, policy makers, academics, and artists.[35] Ironically, several of the people who had been included in the set without any consent are known for their work critiquing surveillance and facial recognition itself, including filmmaker Laura Poitras, digital rights activist Jillian York, critic Evgeny Morozov and author of *Surveillance Capitalism* Shoshana Zuboff. After an investigation in the *Financial Times* based on Harvey and LaPlace's work was published, the set disappeared. [36] A spokesperson for Microsoft claimed simply that it was removed 'because the research challenge is over'.[37]

MS CELEB dataset

On one hand, removing these problematic datasets from the internet may seem like a victory. The most obvious privacy and ethical violations are addressed by making them no longer accessible. However, taking them offline doesn't stop their work in the world: these training sets have been downloaded countless times, and have made their way into many production AI systems and academic papers. By erasing them completely, not only is a significant part of the history of AI lost, but researchers are unable to see how the assumptions, labels and classificatory approaches have been replicated in new systems, or trace the provenance of skews and biases exhibited in working systems. Facial-recognition and emotion-recognition AI systems are already propagating into hiring, education and healthcare. They are part of security checks at airports and interview protocols at Fortune 500 companies. Not being able to see the basis on which AI systems are trained removes an important forensic method to understand how they work. This has serious consequences.

For example, a recent paper led by a PhD student at the University of Cambridge introduced a real-time drone surveillance system to identify violent individuals in public areas. It is trained on datasets of ‘violent behaviour’ and uses those models for drone surveillance systems to detect and isolate violent behaviour in crowds. The team created the Aerial Violent Individual (AVI) Dataset, which consists of 2,000 images of people engaged in five activities: punching, stabbing, shooting, kicking and strangling. In order to train their AI, they asked twenty-five volunteers between the ages of eighteen and twenty-five to mimic these actions. Watching the videos is almost comic. The actors stand far apart and perform strangely exaggerated gestures. It looks like a children’s pantomime, or badly modelled game characters.[38] The full dataset is not available for the public to download. The lead researcher, Amarjot Singh (now at Stanford University), said he plans to test the AI system by flying drones over two major festivals, and potentially at national borders in India. [39] [40]

An archeological analysis of the AVI dataset - similar to our analyses of ImageNet, JAFFE, and Diversity in Faces - could be very revealing. There is clearly a significant difference between staged performances of violence and real-world cases. The researchers are training drones to recognise pantomimes of violence, with all of the misunderstandings that might come with that. Furthermore, the AVI dataset doesn’t have anything for ‘actions that aren’t violence but might look like it’; neither do they publish any details about their false-positive rate (how often their system detects nonviolent behavior as violent).[41] Until their data is released, it is impossible to do forensic testing on how they classify and interpret human bodies, actions or inactions.

This is the problem of inaccessible or disappearing datasets. If they are, or were, being used in systems that play a role in everyday life, it is important to be able to study and understand the worldview they normalise. Developing frameworks within which future researchers can access these data sets in ways that don’t perpetuate harm is a topic for further work.

## Conclusion: Who decides?

The Lombrosian criminologists and other phrenologists of the early twentieth century didn’t see themselves as political reactionaries. On the contrary, as Steven Jay Gould points out, they tended to be liberals and socialists whose intention was ‘to use modern science as a cleansing broom to sweep away from jurisprudence the outdated philosophical baggage of free will and unmitigated moral responsibility’.[42] They believed their anthropometric method of studying criminality could lead to a more enlightened approach to the application of justice. Some of them truly believed they were ‘de-biasing’ criminal justice systems, creating ‘fairer’ outcomes through the application of their ‘scientific’ and ‘objective’ methods.

Amid the heyday of phrenology and ‘criminal anthropology’, the artist Rene? Magritte completed a painting of a pipe and coupled it with the words ‘Ceci n’est pas une pipe’. Magritte called the painting *La trahison des images*, ‘The Treachery of Images’. That same year, he penned a text in the surrealist newsletter *La Re?volution surre?aliste*. ‘Les mots et les images’ is a playful romp through the complexities and subtleties of images, labels, icons and references, underscoring the extent to which there is nothing at all straightforward about the relationship between images and words or linguistic concepts. The series would culminate in a series of paintings: ‘This Is Not an Apple’.

The contrast between Magritte and the physiognomists’ approach to representation speaks to two very different conceptions of the fundamental relationship between images and their labels, and of representation itself. For the physiognomists, there was an underlying faith that the relationship between an image of a person and the character of that person was inscribed in the images themselves. Magritte’s assumption was almost diametrically opposed: that images in and of themselves have, at best, a very unstable relationship to the things they seem to represent, one that can be sculpted by whoever has the power to say what a particular image means.

For Magritte, the meaning of images is relational, open to contestation. At first blush, his painting might seem like a simple semiotic stunt, but the underlying dynamic Magritte underlines in the painting points to a much broader politics of representation and self-representation.

Memphis Sanitation Workers Strike of 1968

Struggles for justice have always been, in part, struggles over the meaning of images and representations. In 1968, African American sanitation workers went on strike to protest dangerous working conditions and terrible treatment at the hands of Memphis's racist government. They held up signs recalling language from the nineteenth-century abolitionist movement: 'I AM A MAN'. In the 1970s, queer-liberation activists appropriated a symbol originally used in Nazi concentration camps to identify prisoners who had been labeled as homosexual, bisexual, and transgender. The pink triangle became a badge of pride, one of the most iconic symbols of queer-liberation movements. Examples such as these - of people trying to define the meaning of their own representations - are everywhere in struggles for justice. Representations aren't simply confined to the spheres of language and culture, but have real implications in terms of rights, liberties, and forms of self-determination.

There is much at stake in the architecture and contents of the training sets used in AI. They can promote or discriminate, approve or reject, render visible or invisible, judge or enforce. And so we need to examine them - because they are already used to examine us - and to have a wider public discussion about their consequences, rather than keeping it within academic corridors. As training sets are increasingly part of our urban, legal, logistical, and commercial infrastructures they have an important but under examined role: the power to shape the world in their own images.

---

### Credits and Acknowledgements

Authors: Kate Crawford and Trevor Paglen

Published by The AI Now Institute, NYU (https://ainowinstitute.org /)

Full citation: Kate Crawford and Trevor Paglen, “Excavating AI: The Politics of Training Sets for Machine Learning (September 19, 2019) https://excavating.ai

Acknowledgements: thanks to all those who have given editorial feedback, technical support, research contributions and conversations on these issues over the years, including Arvind Narayanan, Daniel Neves, Varoon Mathur, Olga Russakovsky, Leif Ryge, Lea Saint- Raymond and Kiran Samuel.

Additional thanks to Mario Mainetti and Carlo Barbatti and all the staff at the Fondazione Prada, and to Alona Pardo and the staff at the Barbican Centre. The images in this essay and many more are part of the Fondazione Prada Osservatorio *Training Humans* exhibition, in Milan from 12 September, 2019 through 24 February 2020; and at the Barbican Centre in London as part of the exhibition *From Apple to Anomaly (Pictures and Labels)* from 26 September, 2019 through 16 February, 2020.

Credit for ImageNet Roulette software: developed by Leif Ryge at Paglen Studio.

*'Excavating AI: The Politics of Training Sets for Machine Learning' (September 19, 2019) https://excavating.ai by Kate Crawford and Trevor Paglen was originally published by The AI Now Institute, NYU (https://ainowinstitute.org/) and is republished with kind permission of Kate Crawford and Trevor Paglen.*

[1] Minsky currently faces serious allegations related to convicted paedophile and rapist Jeffrey Epstein. Minsky was one of several scientists who met with Epstein and visited his island retreat where underage girls were forced to have sex with members of Epstein's coterie. As scholar Meredith Broussard observed, this was part of a broader culture of exclusion that became endemic in AI: 'as wonderfully creative as Minsky and his cohort were, they also solidified the culture of tech as a billionaire boys' club. Math, physics, and the other "hard" sciences have never been hospitable to women and people of color; tech followed this lead.' See Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge, Massachusetts and London: MIT Press, 2018), p. 174.

[2] See Daniel Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1993), p. 88.

[3] Minsky gets the credit for this idea, but clearly Papert, Sussman, and teams of 'summer workers' were all part of this early effort to get computers to describe objects in the world. See Seymour A. Papert, 'The Summer Vision Project' (July 1, 1966), https://dspace.mit.edu/handle/1721.1/6125. As he wrote: 'The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition"'.

[4] Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed, Prentice Hall Series in Artificial Intelligence (Upper Saddle River, NJ: Prentice Hall, 2010), p. 987.

[5] In the late 1970s, Ryszard Michalski wrote an algorithm based on 'symbolic variables' and logical rules. This language was very popular in the 1980s and 1990s, but, as the rules of decision-making and qualification became more complex, the language became less usable. At the same moment, the potential of using large training sets triggered a shift from this conceptual clustering to contemporary machine-learning approaches. See Ryszard Michalski, 'Pattern Recognition as Rule-Guided Inductive Inference'. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2 (1980), pp. 349–361.

[6] There are hundreds of scholarly books in this category, but for a good place to start, see W.J.T. Mitchell, *Picture Theory: Essays on Verbal and Visual Representation*, Paperback ed., [Nachdr.] (Chicago: University of Chicago Press, 2007).

[7] M. Lyons et al., 'Coding Facial Expressions with Gabor Wavelets', in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition* (Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan: IEEE Comput. Soc, 1998), pp. 200– 205, https://doi.org/10.1109/AFGR.1998.670949.

[8] As described in the 'AI Now Report' (2018), this classification of emotions into six categories has its root in the work of the psychologist Paul Ekman. 'Studying faces, according to Ekman, produces an objective reading of authentic interior states - a direct window to the soul. Underlying his belief was the idea that emotions are fixed and universal, identical across individuals, and clearly visible in observable biological mechanisms regardless of cultural context. But Ekman's work has been deeply criticized by psychologists, anthropologists, and other researchers who have found his theories do not hold up under sustained scrutiny. The psychologist Lisa Feldman Barrett and her colleagues have argued that an understanding of emotions in terms of these rigid categories and simplistic physiological causes is no longer tenable. Nonetheless, AI researchers have taken his work as fact,

and used it as a basis for automating emotion detection.' Meredith Whitaker et al., 'AI Now Report 2018', AI Now Institute (December 2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf. See also Lisa Feldman Barrett et al., 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest* 20 (1) (July 17, 2019), pp. 1–68, https://doi.org/10.1177/1529100619832930.

[9] See, for example, Ruth Leys, 'How Did Fear Become a Scientific Object and What Kind of Object Is It?', *Representations* 110 (1) (May 2010), pp. 66–104, https://doi.org/10.1525/rep.2010.110.1.66. Leys has offered a number of critiques of Ekman's research programme, most recently in Ruth Leys, *The Ascent of Affect: Genealogy and Critique* (Chicago and London: University of Chicago Press, 2017). See also Lisa Feldman Barrett, 'Are Emotions Natural Kinds?', *Perspectives on Psychological Science* 1 (1) (March 2006), pp. 28–58, https://doi.org/10.1111/j.1745- 6916.2006.00003.x; Erika H. Siegel et al., 'Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories.', *Psychological Bulletin,* 20180201, https://doi.org/10.1037/bul0000128.

[10] Fei-Fei Li, as quoted in Dave Gershgorn, 'The Data That Transformed AI Research - and Possibly the World', *Quartz* (July 26, 2017), https://qz.com/1034972/the- data-that-changed-the-direction-of-ai-research-and-possibly-the-world/. Emphasis added.

[11] John Markoff, 'Seeking a Better Way to Find Web Images', *The New York Times* (November, 19, 2012), sec. Science, https://www.nytimes.com/2012/11/20/s cience/for-web-images-creat-ng-new-technology-to-seek-and-find.html.

[12] Their paper can be found here: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, 'ImageNet Classification with Deep Convolutional Neural Networks', in *Advances in Neural Information Processing Systems* 25, ed. F. Pereira et al. (Curran Associates, Inc., 2012), pp. 1097–1105, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep- convolutional-neura--networks.pdf.

[13] Released in the mid-1980s, this lexical database for the English language can be seen as a thesaurus that defines and groups English words into synsets, i.e., sets of synonyms. https://wordnet.princeton.edu. This project takes place in a broader history of computational linguistics and natural-language processing (NLP), which developed during the same period. This subfield aims at programming computers to process and analyse large amounts of natural language data, using machine-learning algorithms.

[14] See Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, First paperback edition, Inside Technology (Cambridge, Massachusetts and London: MIT Press, 2000), pp. 44, 107; Anja Bechmann and Geoffrey C. Bowker, 'Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media', *Big Data & Society* 6 (1) (January 2019): 205395171881956, https://doi.org/10.1177/2053951718819569.

[15] These are some of the categories that have now been entirely deleted from ImageNet as of January, 24, 2019.

[16] For an account of the politics of classification in the Library of Congress, see Sanford Berman, *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People* (Metuchen, NJ: Scarecrow Press, 1971).

[17] We're drawing in part here on the work of George Lakoff in *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press, 2012).

[18] See Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, 'Imagenet: A Large-Scale Hierarchical Image Database' In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–55.

[19] Quoted in Allan Sekula, 'The Body and the Archive', *October* 39 (1986), pp. 3–64, https://doi.org/10.2307/778312.

[20] Ibid; for a broader discussion of objectivity, scientific judgment, and a more nuanced take on photography's role in it, see Lorraine Daston and Peter Galison, *Objectivity,* Paperback ed. (New York: Zone Books, 2010).

[21] 'UTKFace – Aicip', accessed August 28, 2019, http://aicip.eecs.utk.edu/wiki/UTKFace.

[22] See Paul N. Edwards and Gabrielle Hecht, 'History and the Technopolitics of Identity: The Case of Apartheid South Africa', *Journal of Southern African Studies* 36 (3) (September 2010), pp. 619–39, https://doi.org/10.1080/03057070.2010.507568. Earlier classifications used in the 1950

Population Act and Group Areas Act used four classes: ‘Europeans, Asiatics, persons of mixed race or coloureds, and “natives” or pure-blooded individuals of the Bantu race’ (Bowker and Star, p. 197). Black South Africans were required to carry pass books and could not, for example, spend more than 72 hours in a white area without permission from the government for a work contract (p. 198).

[23] Bowker and Star, 208.

[24] See F. James Davis, *Who Is Black? One Nation’s Definition*, 10th anniversary ed. (University Park, PA: Pennsylvania State University Press, 2001).

[25] See Joy Buolamwini and Timnit Gebru, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’, in Conference on Fairness, Accountability, and Transparency (2018), pp. 77–91, http://proceedings.mlr.press/v81/buolamwini18a.html.

[26] Michele Merler et al., ‘Diversity in Faces’, ArXiv:1901.10436 [Cs] (January 29, 2019), http://arxiv.org/abs/1901.10436.

[27] ‘Webscope | Yahoo Labs’, accessed August 28, 2019, https://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67&guccounter=1.

[28] Olivia Solon, ‘Facial Recognition’s “Dirty Little Secret”: Millions of Online Photos Scraped without Consent’ (March 12, 2019), https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921.

[29] Stephen Jay Gould, *The Mismeasure of Man*, revised and expanded (New York: Norton, 1996). The approach of measuring intelligence based on skull size was prevalent across Europe and the US. For example, in France, Paul Broca and Gustave Le Bon developed the approach of measuring intelligence based on skull size. See Paul Broca, ‘Sur le crane de Schiller et sur l’indice cubique des cranes’, *Bulletin de la Societe? d’anthropologie de Paris*, I° Serie, t. 5, fasc. 1, pp. 253-60, 1864. Gustave Le Bon, *L’homme et les societes. Leurs origines et leur de?veloppement* (Paris: Edition J. Rothschild, 1881). In Nazi Germany, the ‘anthropologist’ Eva Justin wrote about Sinti and Roma people, based on anthropometric and skull measurements. See Eva Justin, Lebensschicksale artfremd erzogener Zigeunerkinder und ihrer Nachkommen [Biographical destinies of Gypsy children and their offspring who were educated in a manner inappropriate for their species], doctoral dissertation, Friedrich-Wilhelms-Universitat Berlin, 1943.

[30] ‘Figure Eight | The Essential High- Quality Data Annotation Platform’, Figure Eight, accessed August 28, 2019, https://www.figure-eight.com/.

[31] The authors made a backup of the ImageNet dataset prior to much of its deletion.

[32] Their ‘MegaPixels’ project is here: https://megapixels.cc/.

[33] Jake Satisky, ‘A Duke Study Recorded Thousands of Students’ Faces. Now They’re Being Used All over the World’, *The Chronicle* (June 12, 2019), https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur.

[34] ‘2nd Unconstrained Face Detection and Open Set Recognition Challenge’, accessed August 28, 2019, https://vast.uccs.edu/Opensetface/; Russell Stewart, Brainwash Dataset (Stanford Digital Repository, 2015), https://purl.stanford.edu/sx925dc9385.

[35] Melissa Locker, ‘Microsoft, Duke, and Stanford Quietly Delete Databases with Millions of Faces’, Fast Company (June 6, 2019), https://www.fastcompany.com/90360490/ms-celeb-microsoft-deletes-10m-faces-from-face-database.

[36] Madhumita Murgia, ‘Who’s Using Your Face? The Ugly Truth about Facial Recognition’, *Financial Times* (April 19, 2019), https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e.

[37] Locker, ‘Microsoft, Duke, and Stanford Quietly Delete Databases’.

[38] Full video here: Amarjot Singh, ‘Eye in the Sky: Real-Time Drone Surveillance System (DSS) for Violent Individuals Identification’ (2018), https://www.youtube.com/watch?time_continue=1&v=zYypJPJipYc.

[39] Steven Melendez, ‘Watch This Drone Use AI to Spot Violence in Crowds from the Sky’, *Fast Company* (June 6, 2018), https://www.fastcompany.com/40581669/watch-this-drone-use-ai-t-

-spot-violence-from-the-sky.

[40] James Vincent, 'Drones Taught to Spot Violent Behavior in Crowds Using AI', *The Verge* (June 6, 2018), https://www.theverge.com/2018/6/6/17433482/ai-auto...drones-spot-violent-behavior-crowds.

[41] Ibid.

[42] Gould, *The Mismeasure of Man*, p. 140.

## Kate Crawford and Trevor Paglen

Kate Crawford is a leading academic focusing on the social and political implications of artificial intelligence. For over a decade, her work has centred on understanding large-scale data systems in the wider contexts of politics, history, labour, and the environment. Kate Crawford is a research professor at USC Annenberg, and the Visiting Chair of AI and Justice at the École Normale Supérieure. In 2020, she is the inaugural Visiting Chair for AI and Justice at the École Normale Supérieure in Paris. She co-founded the AI Now Institute at New York University, a university centre dedicated to researching the social implications of AI and related technologies. In 2019, she and Vladan Joler won the Beazley Design of the Year Award, for their Anatomy of an AI System project, which was recently acquired by MoMA for its permanent collection. Crawford was also jointly awarded the Ayrton Prize from the British Society for the History of Science for the project Excavating AI. Her new book Atlas of AI is forthcoming with Yale University Press in 2021.

Trevor Paglen is an artist whose work spans image-making, sculpture, investigative journalism, writing, engineering and numerous other disciplines. Among his chief concerns are learning how to see the historical moment in which we live, and developing the means to imagine alternative futures.Paglen has had one-person exhibitions at Nam June Paik Art Center, Seoul; Museo Tamayo, Mexico City; the Nevada Museum of Art, Reno; Vienna Secession; Eli & Edythe Broad Art Museum, East Lansing, Michigan; Van Abbe Museum, Eindhoven; Frankfurter Kunstverein and Protocinema Istanbul, and participated in group exhibitions at the Metropolitan Museum of Art, New York;the San Francisco Museum of Modern Art; Tate Modern, London, and numerous other venues. He has launched an artwork into distant orbit around Earth in collaboration with Creative Time and MIT, contributed research and cinematography to the Academy Award-winning film Citizenfour, and created a radioactive public sculpture for the exclusion zone in Fukushima, Japan. He is the author of five books and numerous articles on subjects including experimental geography, state secrecy, military symbology, photography and visuality. Paglen's work has been profiled in the New York Times, Vice Magazine, the New Yorker, and Art Forum. In 2014, he received the Electronic Frontier Foundation's Pioneer Award for his work as a 'groundbreaking investigative artist'.Paglen holds a BA from UC Berkeley, an MFA from the Art Institute of Chicago, and a Ph.D. in Geography from UC Berkeley.

# Notes On A (Dis)continuous Surface

Murad Khan

> **'Differentiated through that which is porous – the skin – a surface perceptive to touch, the body is dissected, fixed "and woven out of a thousand details, anecdotes and stories".'** [1]

From content recommendation and social-media feed curation to financial risk assessment and medical diagnoses, machine-learning models have become a pervasive part of our everyday infrastructure. While automated data-processing instruments have long been part of our lives, machine learning provides an accelerated paradigm within which patterns can be unearthed and made actionable across large pools of historical data. Given that these technologies are being deployed in a variety of public and private systems, ethical questions are increasingly being raised when they seem to fail, with particular concern directed at the role that these technologies play in further entrenching racial biases and practices of discrimination. Whether it be failing to recognise darker-skinned subjects,[2] amplifying negative racial stereotypes[3] or denying access to credit, forms of pattern-based learning appear to consistently exacerbate existing racial inequalities and modes of discrimination. With these models increasingly supporting human decision-making in key areas, it is crucial that we understand how racial representation functions within machine-learning systems, asking both how race is understood, and what can be achieved by encoding this understanding.

## Differential Visibilities

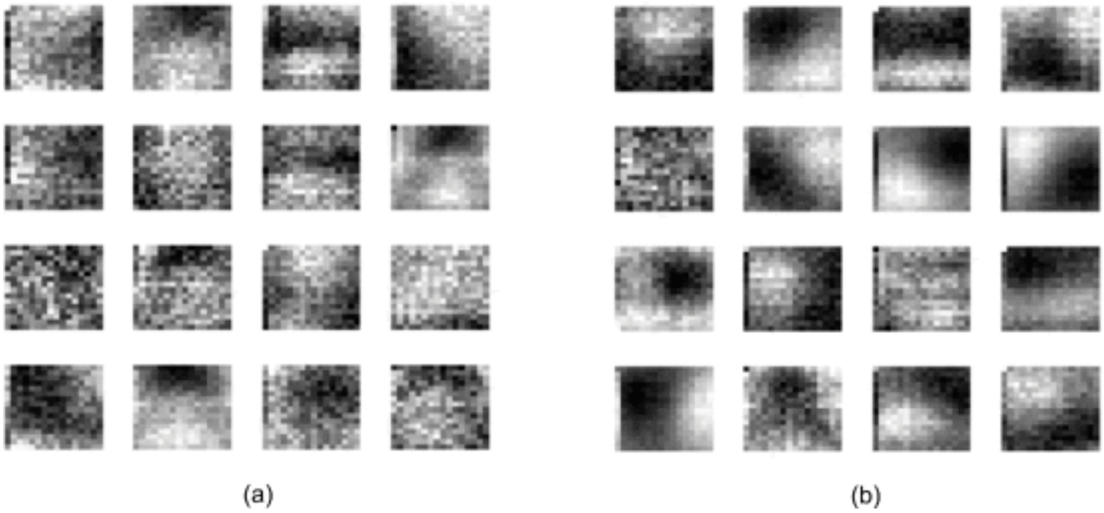


Figure 1. Discriminative race feature representation by multiple layer Convolution Neural Networks (CNN). (a): supervised CNN filters (b): CNN with transfer learning filters [4]

> **'I am given no chance. I am overdetermined from without. I am the slave not of the 'idea' that others have of me but of my own appearance ... I am *fixed*.'**[5]

Frantz Fanon's description identifies his own skin as a site of fixity. In an instance of 'epidermalisation', the porous surface enveloping his body enfolds him within the tonal weave of a racial-corporal schema, apprehending him as Black before human and defining the possibilities afforded to him in accordance with the colour of his skin. This schema, which is 'cultural and discursive' rather than solely genetic,[6] is produced and reproduced across morphological designations, stitching a racialised subject out of 'a thousand details, anecdotes, stories',[7] constituting them historically within the limited and specular frame of race-centric discourse. Crucially, such a schema seeks to align the exterior expressions of the body with internal traits corresponding to behaviour, character and cognitive capacity that can be generalised over members of the given racial group. Doing so composes race beyond the remits of the individual body, forming it in concert with the fictive hierarchies that guarantee the colonial arrangement, naturalising racial difference as a twinned condition of the body and mind. To this extent, race is more than just a schema of visual understanding. It forms a perceptive tissue that brings together forms of social organisation through a psychic operation that safeguards the conditions of the human for certain groups over others, forming the fragmented racial body into a knowable object whenever it is invoked: a legible surface upon which all manner of racial truths may be etched and read in service of maintaining extant social relations. To this degree, it is imperative to outline the ways in which race is figured by a similar series of epidermal abstractions within machine-learning systems, mobilised as a site for perception and identification as well as probabilistic prediction.

### Abstraction, Recognition and Prediction

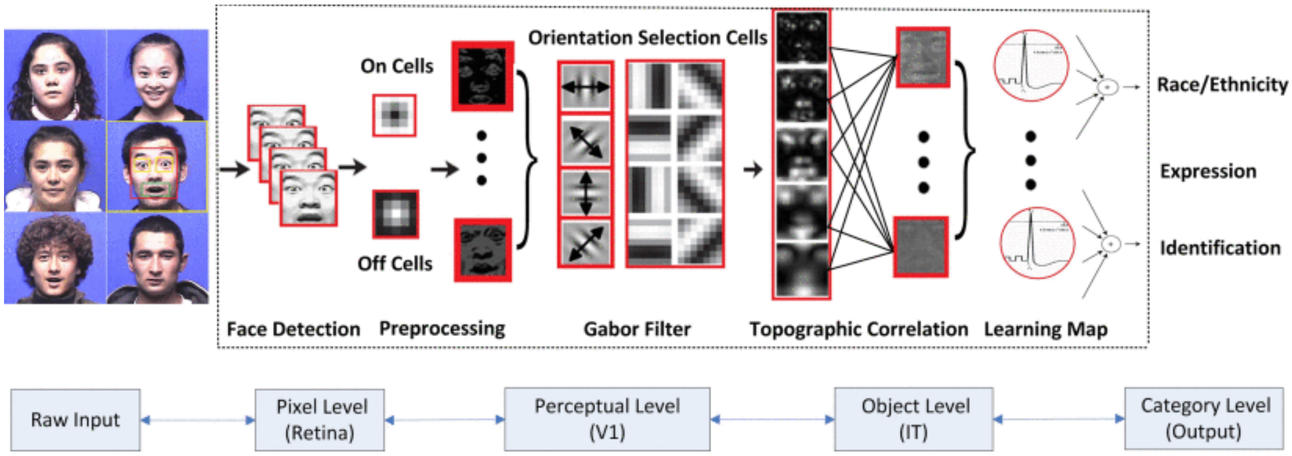


Figure 2. Fu, Siyao, Haibo He, and Zeng-Guang Hou. "Race classification from face: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, no. 12 (2014): 2483-2509.

Whilst forms of biometric identification technology have been in use since the 1990s, it is only in the past five years that computational and graphics processing power has improved to such a degree that machine learning can regularly be used to solve problems of face detection and recognition. State-of-te-art software now utilises Deep Convolutional Neural Networks (DCNNs), training the learning model on large datasets of faces for authentication, detection and identification scenarios. This is typically done by mapping pixel regions in an input image, wherein facial landmarks (nodal points) such as the distance between the eyes, the tip of the nose, or the corners of the mouth are mapped, extracted and used to detect each unique face. The number of nodal points mapped for each model varies depending on the algorithm used, with some generating an embedding of up to 128 measurements in order to properly map the image to a set of numerical representations. Once these landmarks have been identified and the model trained enough times on these representations, it will have scaled in complexity, moving from an array of indiscernible lines and edges, through to blobs, facial features and eventually to a coherent understanding of a 'face', or a set of values equating to different pixel regions across the image.

Racial representation comes into the equation in supervised learning scenarios, in which the model is provided with labelled images to better classify different types of faces based on these learned patterns of pixels. These labels are key to understanding the racialised nature of facial recognition, as the model learns features corresponding to a given taxonomy of racial classifications, sorting patterns it discovers into these pre-defined spaces of representation, and gauging their proximity (similarity) to one another in order to make a judgement on which racial class an individual face falls into. However, since DCNNs are dependent upon the datasets used to train them, we regularly see instances of failure if the set of faces for a certain racial class is lacking in its training data. Often, this is played out across darker-skinned subjects, causing failure rates to increase once the model encounters them in real-world applications. Subjects either fail to be recognised, or are mis-recognised within the given categories of racial representation. Such failures are exceedingly common, ranging from exam proctoring software barring students from taking tests[8] to passport applications being rejected.[9]

That these technologies consistently fail when faced with racialised populations is a well-documented issue, making it all the more pernicious that these technologies are continually implemented in public-facing infrastructure.[10] However, those proposing greater diversity in data-representation as a solution to these issues tend to miss the nuances of the problem, failing to recognise that, implemented 'accurately' or otherwise, these racial classifications are going to be put to work in improving predictive policing, surveillance infrastructure and drone targeting systems that render differential levels of harm to racialised populations. For instance, IBM's attempt to create its 'Diversity in Faces' dataset to alleviate racial bias is a prime example of the damage that can be done when large companies latch onto the idea of being more 'diverse' only to reproduce historical understandings about the 'reality' of racial representation. In their search for a diverse and 'racially accurate' dataset that extended beyond the brute classifications of skin colour, not only did researchers from IBM make worrying recourse to craniofacial measurements as an objective indicator of racial grouping,[11] but they did so whilst simultaneously selling custom implementations of their facial recognition software to law-enforcement agencies.[12] Such pseudo-scientific practices have also spilled over into the realm of prediction and 'affective computing', where emotional analysis is carried out on facial expressions.[13] As expected of a system using race-centric data, analysis of the facial expressions of Black men consistently scored them as angrier than White men, replicating social biases.[14] Frank Pasquale summarises the inevitability of bias within such a system, emphasising that 'If a database of aggression is developed from observation of a particular subset of the population, the resulting AI may be far better at finding "suspect behavior" in that subset rather than others.'[15] Thus, by mimicking the long history of pseudo-sciences such as physiognomy and phrenology that tied racialised facial representation to forms of criminality and deviance, such software merely rehashes historic schemas of racial perception under the guise of insightful and objective computational analysis, making them actionable once more.

While expression analysis demonstrates one clearly racialised form of machine prediction, there are other instances in which the learning system may not be presented with race as a defined variable in its input data, but still picks up on cues that implicate race as a latent force within an assemblage of other variables. This associative tendency exacerbates what is referred to as the problem of 'algorithmic bias', denoting the way in which socio-technical apparatuses that leverage statistical (probability-based) models to guide decision-making frequently make predictions based upon implicitly racialised data, amplifying patterns of social bias. Safiya Noble argues that these practices enact similar forms of exclusion and discrimination to 'redlining' practices in the United States. The computation of probabilities, whether for medical diagnoses, credit allocation or even search engine results, depends upon pattern-based abstractions extending a series of equivalencies and probabilities from the physiological designations of the racialised body, proxied for by a wide range of class conditions that reflect and foster structural inequalities, such as access to housing, education history, employment opportunities, life expectancy and so on. Ramon Amaro provides a useful articulation of these discriminatory logics, positing that in the realm of human difference, machine learning has become 'a projection of an already racialised imaginary enacted through technological solution – an imaginary that already understands the black, brown, criminalised, gendered and otherwise Othered human as the principle site of exclusion, quantification, and social organisation.'[16] As such, machine learning can be seen to replay the Fanonian problematics of corporeal representation and psychic differentiation within the sphere of predictive computation, contaminated by the legacies and motivations of the colonial arrangement.

Given these manifestations of race within machine learning, both at the level of visual recognition and within historical data distributions, we can see that the problem of race is best encapsulated not by the question of non-recognition, but of recognition within a discursive environment that has asserted race as a coherent metric for the classification of people as well as a meaningful predictor of future behaviour. Much as Fanon suggests, racialised subjects are ‘overdetermined from without’, subject to the legacies and injustices consonant with racial identification and their rearticulation within contemporary technical infrastructure. Contemporary applications of machine-learning[17] In doing so, patterns of probability reach across bodies to form the recurrent possibility of an object both legible and computable, contiguous with the racialised exterior and interior features of an individual. Coerced into an extensive causal surface, the dynamisms of living, breathing individuals are pulled together by the epidermal logic described by Fanon.

---

[1] Simone Browne, ‘Digital Epidermalization: Race, Identity and Biometrics’, *Critical Sociology,* 36 (1) (February 1, 2010), p. 133.

[2] Joy Buolamwini, Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. *Proceedings of Machine Learning Research,* 81 (2018), pp. 1–15.

[3] See Safiya Noble, *Algorithms of Oppression* (New York: New York University Press, 2018).

[4] Siyao Fu, Haibo He, Zeng-Guang Hou, 'Learning Race from Face: A Survey', *IEEE Transactions on Patter Analysis and Machine Intelligence,* 36 (12) (December 2014). (Fig. 8)

[5] Frantz Fanon, *Black Skin, White Masks* (London: Pluto Press, 1986), p. 87.

[6] Stuart Hall, *The Fact of Blackness: Frantz Fanon and Visual Representation*, ed. Alan Read (London: ICA, 1996), p. 16.

[7] Fanon, *Black Skin, White Masks*, p. 84.

[8] https://venturebeat.com/2020/09/29/examsofts-remot...

[9] https://www.newscientist.com/article/2219284-uk-la...

[10] For instance, in the case of facial recognition for the Home Office’s automated passport-photo processing service, a Freedom of Information request revealed that tests had been carried out showing a poor result on darker skinned faces, yet the service was deemed ‘sufficient enough to deploy’. See https://www.whatdotheyknow.com/request/skin_colour...

[11] *Diversity in Faces*, https://arxiv.org/abs/1901.10436,

[12] It is worth noting that, in the wake of global Black Lives Matter protests sparked by the deaths of George Floyd, Breonna Taylor and countless others at the hands of the police, IBM chose to announce a moratorium on the sale of facial recognition technology, and to open a dialogue on ‘whether and how facial recognition technology should be employed by domestic law enforcement agencies’. Whilst this garnered much applause from ‘AI Ethics’ advocates, the more cynical among us may note that their announcement only stated that they would no longer offer ‘general purpose IBM facial recognition or analysis software’ for sale. Whether the software would remain available for custom implementations, such as in police body camera offerings, as they advertise elsewhere on their website, is unclear.

[13] Amazon’s Rekognition software, for instance, provides a confidence score for facial emotion. See https://docs.aws.amazon.com/rekognition/latest/dg/...

[14] Lauren Rhue, ‘Racial Influence on Automated Perceptions of Emotions’ (November 9, 2018). Available at SSRN: https://ssrn.com/abstract=3281765 or http://dx.doi.org/10.2139/ssrn.3281765

[15] https://reallifemag.com/more-than-a-feeling/

[16] Ramon Amaro, *AI and the Empirical Reality of a Racialised Future in AI: More Than Human* (London: Barbican, 2019), p. 126.

[17] Hall, *The Fact of Blackness*, p. 20.

## Murad Khan

Murad Khan is a Ph.D researcher at University College London and a Visiting Practitioner at UAL: Central Saint Martins. Directed through an emerging philosophy of noise and adversarial machine learning, his research seeks to read Frantz Fanon as a philosopher of information, bridging his work on the psychic operations of racialisation with Gilbert Simondon's philosophy of individuation in an exploration of psychopathology and the limits of reason.

# Irresolvable Contradictions in Algorithmic Thought

Leonardo Impett

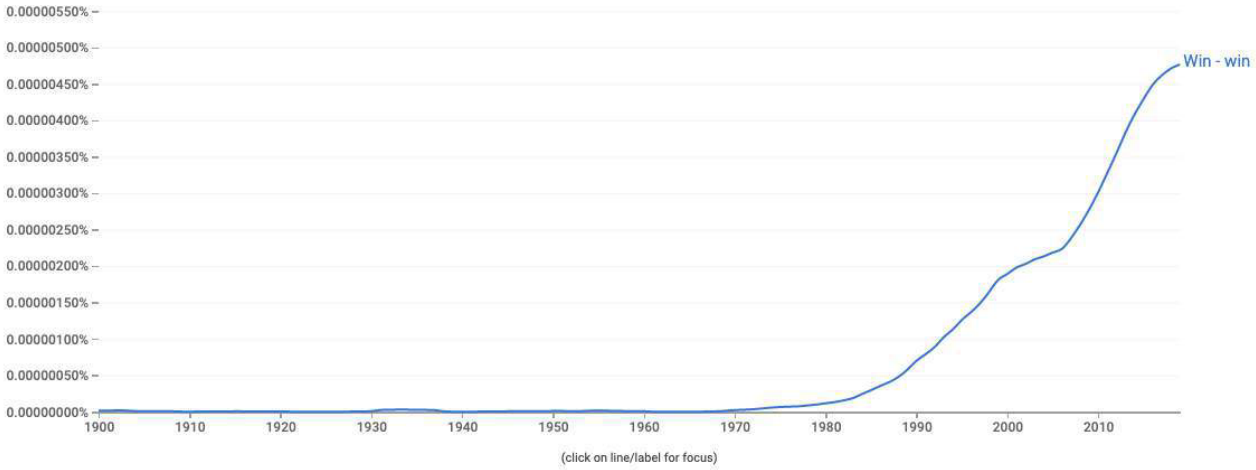


Figure 1. The relative frequency of the term 'win-win' in the Google Books English-language corpus, from the Google Books Ngrams Viewer

**'A man cannot have his cake and eat his cake'**

– Thomas Howard, the 3rd Duke of Norfolk, to Thomas Cromwell, 14 March 1538 [1]

The 'win-win' situation is perhaps the most characteristic rhetorical device of late capitalism. You'll find it frequently in the white papers of Neoliberal think-tanks and company annual reports, but a recent form of particular prevalence combines moral absolution with economic growth. 'Electric cars emit no carbon, and they're a lot faster from 0–60.' 'Offshoring doesn't just save us money, it invests in developing economies.' 'We can tackle bias and unfairness in AI by making our workforce more diverse, which will also make us more profitable.'[2]

These happy victories, even if genuine, are often pointers to insufficiency. After all, capitalism's ability to radically self-invent in the interest of its own profits has never been in doubt. What is in doubt is its ability to solve questions of public good *against* its own commercial interests. Take the precedents of tobacco and public health, or oil and climate change: wherever common interests pose an existential threat to an industry's survival, that industry has historically responded through the tactics of confusion and delay.

So, when Microsoft lists 'inclusiveness' and 'fairness' as two of the six guiding principles for Responsible AI – indeed, when Microsoft even considers the idea of having a central policy on Responsible AI – we may well feel the same kind of scepticism provoked by, say, Philip Morris' 2018 anti-smoking advertising.

To feel this, however, would be to misunderstand the situation on the ground. Corporate AI labs have become serious nodes for critical thought around the role of AI in society. This year, a paper on fairness in AI from Microsoft Research won the Best Paper award at the prestigious CHI Conference on Human Factors in Computing Systems [3] and a large amount of internationally relevant research is produced by other major corporate tech players. My own first encounter with critical approaches to AI was as part of a paid internship at Microsoft Research's Cairo office, where – as an Engineering undergraduate – my main project was a Bourdieuian reading of a commonly used computer vision dataset. Last year, Amazon announced a $20m funding track with the US National Science Foundation on *Fairness in Artificial Intelligence*.

Of course – despite their well-publicised efforts towards fairer AI – Amazon, Google, Microsoft, Oracle and IBM all bid for a $10 billion cloud computing contract with the US Department of Defense in 2018–19. In October 2018, Google dropped out of the bid, apparently because it 'couldn't be assured that it would align with our AI Principles'. A progressive capitalism working against its own profits? Not quite –

Google’s decision came after seven months of traditional organised worker pressure, including demonstrations and high-profile resignations. For Google, Microsoft and Amazon, the political situation is one of internal dissonance rather than calculated hypocrisy: in Hegelian-Marxist terms, we might call it a collateral contradiction (Nebenwiderspruch). This contradiction on the corporate level interplays with an underlying contradiction at the algorithmic level, particularly in relation to deep-learning techniques.

The technical borders of ‘deep learning’ are not perfectly delineated, but we understand it to refer to a subset of machine-learning algorithms called neural networks. These neural networks are generally divided into separate layers (where the output of one layer is itself the input to the next, and so on). We call a neural network ‘deep’ [4] when it has a relatively large number of such layers. Before deep learning, machine-learning techniques generally had to rely on hand-crafted features of the data. Rather than being fed with individual pixel values, early machine-learning algorithms for images were fed a spreadsheet of secondary information about an image: brightness, average colours, gradients and silhouettes. The ‘depth’ (i.e. large number of neural layers) of deep neural networks allows them, instead, to work on unadulterated raw data: on pixel values themselves.[5]

To illustrate the dangerous complications that deep networks introduce, let's consider a case-study: ‘smart’ CCTV systems that alert security when an ‘anomalous’ event occurs. The industry for home ‘smart’ surveillance technology, such as Amazon’s CCTV-enabled Ring, is swiftly reaching the $10bn mark [6] and ‘anomaly detection’ has become an important application domain for research in deep learning.[7]

Let's assume that we train our deep CCTV algorithm on a dataset of real security camera footage, taken at random from across the United States (where most AI companies are based). We mark any event that eventually led to an observed person's incarceration an ‘anomaly’, an example from which to learn. If our dataset is unbiased in the orthodox statistical sense (i.e. if it contains a representative sample of US arrests) we would find our smart CCTV dataset contained black people involved in such ‘anomalous’ events at five times the frequency of the white population.[8] If we then trained a perfect deep-learning model – ie one that was able to perfectly reproduce the decision-making processes implicit in its training data – we would have built a CCTV system that flags black people as incidents to be investigated five times more frequently than white people. Even worse, if our classifier was not perfectly accurate (we all – human or machine – have some probability of error), the ‘false positive’ rate for black people would – through the associative logic of Bayesian reasoning – also be at around 500% of the rate for the white population. In fact, this would be roughly consistent with the rate at which black people are disproportionately stopped without just cause.[9]

This active prejudice would be present in a deep CCTV system that was never explicitly programmed to take account of race as a variable. In the world of deep learning, anything implicit in pixel-values is fair game: prejudice is present in the deep-learning system, because it was present in the data. Many point, therefore, to the possibility of creating less biased datasets – utopic data. This may be possible for the dualistic white-black distinction drawn in the thought-experiment above, but how could we possibly create perfectly fair training data that balances the complex network of intersecting prejudices (sexuality, gender, social class, income, nationality, disability) hinted at in those raw pixel-values? The problem is even greater for so-called ‘online machine learning’ (often involving ‘reinforcement learning’): systems that continuously self-improve based on examples they face in the real world. In these systems, which make up a significant portion of commercially applied AI (e.g. search engines), we cannot introduce utopic data, since live information about how the system interacts with the (dystopic) real world is integral to how it works and learns.

As Louise Amoore writes, ‘the features that some would like to excise from the algorithm – bias, assumptions, weights – are routes into opening up their politics’. Indeed, recent research into algorithmic bias allows us to think through this quagmire with considerably more precision and nuance, with implications far beyond computer science. In 2016, Jon Kleinberg, Sendhil Mullainathan and Manish

Raghavan showed [10] that in any real-world society, [11] it is exceedingly difficult to make a ‘fair’ algorithm. More specifically, it is *impossible* to classify data in a way that meets several well-established definitions of fairness at the same time (lack of active discrimination; equal false positive rates; equal false negative rates). What's more, the result of a ‘fair’ classification (through whichever definition of fairness) is not the most accurate classification. There exists, in other words, a structural contradiction between the classificatorial logics of non-discrimination and of accuracy (and therefore profit).[12]

We have been talking about algorithms, but what makes the Kleinberg proof so powerful is that it is based only on the eventual decisions taken: it holds true, no matter who makes the classification-decisions (deep-learning algorithm or human, reactionary or progressive). The collateral contradictions (Nebenwidersprüche) in the corporate behaviour of Big Tech, or in the discriminatory logics of deep learning, have their heart in the fundamental contradiction (Hauptwiderspruch) that Kleinberg’s proof describes. It forces us to confront head-on the axiom (dating back at least to Weber) underlying the ‘win-win’ solution: that a fair solution is also an efficient solution. Efficiency relies on prejudice (and enhances it), and if we are to take anti-discrimination seriously, we must sacrifice accuracy, efficiency, profit, growth. We can no longer have our cake and eat it.

[1] Letters and Papers, Foreign and Domestic, Henry VIII, Volume 13, Part 1, January–July 1538. Originally published by Her Majesty's Stationery Office (London, 1892), pp. 176–92.

[2] See, for example, Karsten Strauss, ‘More Evidence That Company Diversity Leads To Better Profits’, *Forbes*, 25 January 2018; Stephen Turban, Dan Wu and Letian Zhang, ‘Research: When Gender Diversity Makes Firms More Productive’, *Harvard Business Review* (February 11, 2019); Jessica Alsford, ‘The data show it: diverse companies do better’, *Financial Times (*September 30, 2019).

[3] Michael A. Madaio, et al., ‘Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI.’ Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020.

[4] The terminology originated in convolutional neural networks for image recognition, in which the ‘height’ and ‘width’ of the neural fields were determined by the resolution of the images; and the ‘depth’ was therefore the number of neural layers in the model.

[5] Of course, such data is rarely ‘raw’ – in the case of mobile-phone images, it might have passed through automatic smartphone image enhancement techniques (to increase the contrast or colour saturation), not to mention the mediation of a human agent choosing to point a camera in a particular direction at a particular moment. But it is ‘raw’ at a technical level – all the information that is implicit in the pixel-values is now up for grabs.

[6] T. J. McCue, ‘Home Security Cameras Market To Surpass $9.7 Billion By 2023’*, Forbes* (January 31, 2019).

[7] See, for example, Kwang-Eun Ko and Kwee-Bo Sim, ‘Deep convolutional framework for abnormal behavior detection in a smart surveillance system’, *Engineering Applications of Artificial Intelligence,* 67 (2018), pp. 226–34.

[8] NAACP Criminal Justice Fact Sheet, 2020, https://www.naacp.org/criminal-justice-fact-sheet/

[9] NAACP Criminal Justice Fact Sheet, 2020, https://www.naacp.org/criminal-justice-fact-sheet/

[10] Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan, ‘Inherent Trade-Offs in the Fair Determination of Risk Scores’, 8th Innovations in Theoretical Computer Science Conference (ITCS 2017). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2017).

[11] Specifically, any society in which different groups (whatever distinguishes them: gender, race, sexuality etc) do not have identical patterns of behaviour.

[12] Only two edge-cases provide statistical coincidences where the accurate and the fair coincide: ‘perfect predictions' (i.e. where our algorithm makes no errors), and ‘equal base rates’ (i.e. where different groups have statistically identical behaviours).

## Leonardo Impett

Leonardo Impett is Assistant Professor of Computer Science at Durham University. He works in the digital humanities, at the intersection of computer vision and art history. He was previously Scientist at the Bibliotheca Hertziana – Max Planck Institute for Art History, Digital Humanities Fellow at Villa I Tatti – the Harvard University Center for Italian Renaissance Studies, and PhD Candidate at the École Polytechnique Fédérale de Lausanne. In trying to bring 'Distant Reading' to art history and visual studies, his current research focuses on unveiling the implicit image-theories of computer vision, and constructing new computer-vision systems based on early modern philosophies of vision. He is an Associate of Cambridge University Digital Humanities, an Associate Fellow of the Zurich Center for Digital Visual Studies, and an Associate Research at the Orpheus Institute for Artistic Research in Music.

# Creative AI Lab: The Back-End Environments of Art-Making

Eva Jäger

Still from forthcoming ML/AI Interfaces Tutorial Series, 2020. Image courtesy of Trust, Berlin and Ricardo Saavedra

## 1. The Back-end Environments of Art-Making

The Creative AI Lab is a collaboration between the R&D Platform at Serpentine Galleries and King's College London's Department of Digital Humanities. The Lab follows the premise that currently we are at the early stages of understanding the aesthetics and semiotics of 'artificial intelligence' (AI). We also approach AI as a framework that holds together a number of disciplines, technologies and systems (creative, cultural and computational). Historically, the themes contained within AI discourse, such as interfaces, automation, data analysis, algorithmic bias, intelligence, alien logics, etc., have featured as cornerstones of various hyped technologies including robotics and virtual reality and machine learning. Today, AI serves as the wrapper via which we engage with these fundamental concepts of digital culture.

From 2016–20, Serpentine has commissioned and overseen the production of a number of artworks where AI technologies are used as a technical medium as well as a conceptual reference or narrative cue. The Lab, which formed in 2019 and officially launched in July 2020, necessarily grew out of a need to explore the experimentation and production phases of these complex projects as creative and research outputs in their own right. By focusing on the production or 'back-end' environments of this type of art-making, we have been able to investigate the truly novel ways in which artists are remaking interfaces, building datasets and generally reaching into the grey-box of AI technologies. [1]

Importantly, this emphasis on the back-end has led us to insist that the Lab has no mandate to commission or showcase front-end artworks. Instead, the Creative AI Lab holds space for conversations, research and hands-on experimentation that addresses the technical frameworks of AI and their impacts on art-making, and conversely, the possible impacts on AI research and development of art-making that deploys AI. [2]

There are a couple of reasons to insist on an exploratory creative R&D format within an art-institutional setting. Firstly, constructing an organisation *within* the organisation we can unbind from front-end formats such as exhibitions or commissions. Instead, we can follow in the steps of an underrepresented working method within humanities research and museums' output. [3] Secondly, we can provide a necessary supplement to the generic approach to AI that the art-institutional discourse has thus far offered in interpreting the front-end of artworks made using AI technologies. [4] To this extent, our mission is to develop a critical literacy that might help art institutions approach AI as a nuanced medium in art-making. Without this, we will continue to reproduce narratives where art is an *antidote* to technology rather than *a valuable part of its development*.

Cultural producers of all kinds should be involved in forming the cultural meaning of AI technologies. And since we cannot separate the cultural meaning of a technology from the technological object itself (for instance, the machine-learning model), [5] it seems that we must go *through* the back-end.

### 2. Making Meaning

(What follows is an example of this approach that also forms the basis for our next investigation at the Lab)

At a recent talk, Mercedes Bunz, Principal Investigator of the Lab and Senior Lecturer in the Department of Digital Humanities, King's College London, reiterated that if the arts and humanities distance themselves from nitty-gritty technology through siloed critique they will become irrelevant. [6] Instead, she and the Lab work closely with computer scientists as they begin to pivot toward self-critique. Bunz offered some insights to understanding AI technic from the arts-and-humanities perspective – through semiotic studies – that remain under-utilised in computer science.

Most notable is the concept of meaning-making described by Stuart Hall, among others, as a process of both encoding and decoding. [7] It is a process, Bunz argues, that has now been taken up by AI, through deep learning. Understanding contemporary AI as having the capacity to make meaning is crucial if we follow Hall's logic (as Bunz does in a recent paper on the subject) because then meaning can also be made by calculation – a task to which AI is regularly assigned. [8] This proposes a paradigm shift: the core work of culture, the making of meaning, can now also be *made* (processed, analysed, *calculated*) by AI – by the technology itself.

While this is only one specific example (where we admittedly also need to argue that semiotics is what art and culture bring to the table, so to speak), the point is that it confirms that the conceptual meaning of works made with AI technologies is inseparable from its technical meaning. And it can only really be understood by engaging with the technicalities (in the back-end) in a serious way.

As we set out on this investigation and others, we remember to embrace the brittleness of our systems and their specific intelligences. Hopefully, this will bring with it divergent understandings of art-making, artworks and art ecosystems. Perhaps this can give way to an approach that replaces autonomous agents (human subjects) with collaborative coalitions (human and non-human subjects). Perhaps these collaborative coalitions will also produce new meaning.

***

The Lab's first initiative since launching in July 2020 has been the formation of a database of creative AI tools and resources, which is now embedded in the Stages site, here. This database is a growing collection of research commissioned & collected by the Creative AI Lab. The latest tools were selected by Luba Elliott. Check back for new entries. We hope that you will explore (and propose additions to it). Contact us via fae@serpentinegalleries.org

[1] During a Creative AI panel discussion on the topic Aesthetics of New AI Leif Weatherby (NYU Digital Theory H-Lab) noted of AI, 'It's not just a black box, It's at least grey. When you open that up you start to see things that have either aesthetic value, critical value, or both.'

[2] The Serpentine has a history of working in this practice-driven way across its programme, and importantly, not only as a feature of technologically orientated research. A key example of this is the community research undertaken as part of the Edgeware Road Project and the Centre for Possible Studies.

[3] Here we reference (within the humanities) the interdisciplinary work of thinker-tinkerers like

Gilbert Simondon, who combined research as a media theorist with lab work where he experimented with computer components, taking machines apart and rebuilding them. Or (within the arts) we look to the studio and lab practices of artist-engineers like Roy Ascott and Rebecca Allen, to name a few. This method for working is of course not novel. We focus on it only to examine where this method is located – or more importantly, *not* located – in the museum.

[4] This is something Nora N. Khan has outlined in her participation with the Lab and in her essay, *Towards a Poetics of Artificial Superintelligence: How Symbolic Language Can Help Us Grasp The Nature and Power of What is Coming*, included in this Journal.

[5] Gilbert Simondon in his 1958 *Du Mode D'existence des Object Technique* writes, 'Culture has become a system of defense against technics … based on the assumption that technical objects contain no human reality.'

[6] Keynote lecture at the newly opened Centre for Culture and Technology at the University of Southern Denmark.

[7] Stuart Hall, 'Encoding/decoding,' in *Culture, Media, Language: Working Papers in Cultural Studies, 1972–1979*, ed. Stuart Hall, Dorothy Hobson, Andrew Lowe and Paul Willis (London: Hutchinson, 1980), pp. 128–138.

[8] S. Bunz, 'The calculation of meaning: on the misunderstanding of new artificial intelligence as culture', *Culture, Theory and Critique*, 60 (3–4) (2019), pp. 264–78, https://doi.org/10.1080/14735784.2019.1667255

**Eva Jäger**

Eva Jäger is Associate Curator of Arts Technologies at Serpentine. Together with Dr. Mercedes Bunz, she is Co-investigator of the Creative AI Lab, a collaboration between Serpentine R&D Platform and the Department of Digital Humanities, King's College. She is also one half of the studio practice Legrand Jäger.
Mercedes Bunz is the Creative AI Lab's Principal Investigator and Senior Lecturer in Digital Society at the Department of Digital Humanities, King's College London. Her research explores how digital technology transforms knowledge and power.
Alasdair Milne is a researcher at the Creative AI Lab and the recipient of the LAHP/AHRC-funded Collaborative Doctoral Award at King's College London Department of Digital Humanities in collaboration with Serpentine's R&D Platform. His work is broadly concerned with collaboration – how to comprehend practices of thinking and making that incorporate both the human and the non-human. His Ph.D will tackle creative AI as a medium in artistic and curatorial practices.

# Creative AI Database

Serpentine R&D Platform & Kings College London

This is a database* of Creative AI tools for those interested in incorporating machine learning (ML) and other forms of artificial intelligence (AI) into their practice. They cover a broad spectrum of possibilities presented by the current advances in ML like enabling users to generate images from their own data, create interactive artworks, draft texts or recognise objects. Most of the tools require some coding skills, however, we've noted ones that don't. Beginners are encouraged to turn to RunwayML https://runwayml.com.

*The database is an initiative of the Creative AI Lab (a collaboration between Serpentine's R&D Platform and the Department of Digital Humanities at King's College London). It has been customised for *Stages* to show only tools and is available here. For the further resources like publications, essays, courses and interviews visit the full database here. The Lab commissioned Luba Elliott http://elluba.com to aggregate the tools listed here in 2020.  To submit further tools, get in touch with the Lab https://creative-ai.org/info.

Creative AI Database for Stages:  https://creative-ai.org/stages

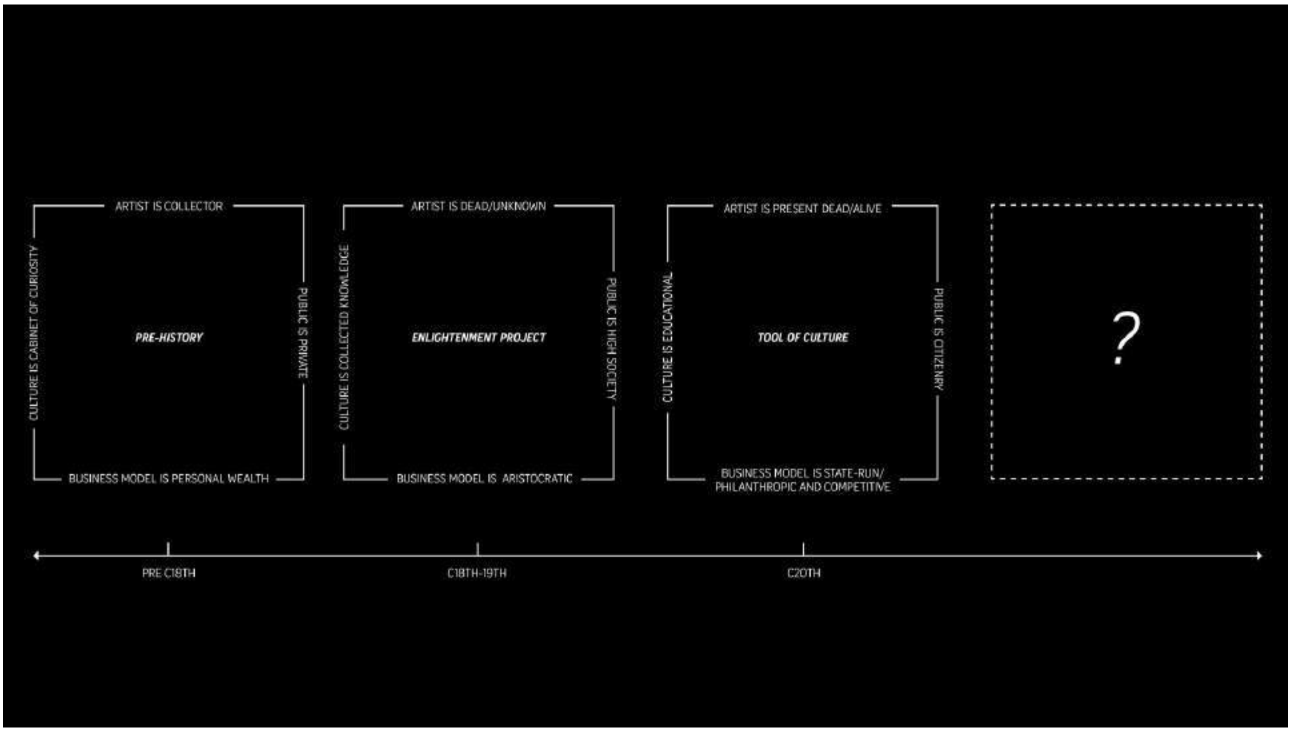
The Creative AI Lab is supported by the Arts and Humanities Research Council.

**Luba Elliott** is a curator, producer and researcher specialising in artificial intelligence in the creative industries. She is currently working to educate and engage the broader public about the latest developments in creative AI through talks, exhibitions and tech demonstrations at venues across the art, business and technology spectrum including The Photographers' Gallery, Victoria and Albert Museum, ZKM Karlsruhe, Impakt Festival, The Leverhulme Centre for the Future of Intelligence, CogX, NeurIPS and ICCV.? Her recent projects include ART-AI Festival, the online gallery aiartonline.com and NeurIPS Machine Learning for Creativity and Design Workshop. She is an Honorary Senior Research Fellow at the UCL Centre for Artificial Intelligence. Prior to that, she worked in start-ups, including the art collector database Larry's List. She obtained her undergraduate degree in Modern Languages at the University of Cambridge and has a certificate in Design Thinking from the Hasso-Plattner-Institute D-school in Potsdam.

**Serpentine R&D Platform & Kings College London**

# Research & Development at the Art Institution

Victoria Ivanova and Ben Vickers


ARTIST IS COLLECTOR
CULTURE IS CABINET OF CURIOSITY
PRE-HISTORY
PUBLIC IS PRIVATE
BUSINESS MODEL IS PERSONAL WEALTH
ARTIST IS DEAD/UNKNOWN
CULTURE IS COLLECTED KNOWLEDGE
ENLIGHTENMENT PROJECT
PUBLIC IS HIGH SOCIETY
BUSINESS MODEL IS ARISTOCRATIC
ARTIST IS PRESENT DEAD/ALIVE
CULTURE IS EDUCATIONAL
TOOL OF CULTURE
PUBLIC IS CITIZENRY
BUSINESS MODEL IS STATE-RUN/ PHILANTHROPIC AND COMPETITIVE
?
PRE C18TH
C18TH-19TH
C20TH

*What would it mean for a cultural institution to make their programmes matter not only artistically, but also infrastructurally?*

Today, it may seem banal to state that the core function of cultural institutions specialising in contemporary art resides in providing general audiences access to art that is relevant to the current moment. Yet it is precisely today, when the pace and impacts of accelerating technological change are comparable to – or even outstrip – those in the period of industrialisation, that it is worth pausing and asking whether a renewed sense of purpose and new beneficiaries may be emerging for cultural institutions in this transformative process.

It is no coincidence that in the case of the Serpentine, a possible direction of travel may be gleaned from artistic engagements with advanced technologies that have ultimately taken the shape of new artworks. For the months spent in production mode, the Serpentine becomes a site where technical, critical, curatorial and artistic capabilities are intertwined and augmented in small but highly ambitious teams. Such processes generate copious amounts of both deeply practical and conceptual knowledge, most of which ultimately remains invisible when the final artwork is presented to the public. Something similar may be said about the back-end (i.e code) or collateral tools that have to be devised in order to make such artworks a reality.

While general audiences may not require this 'background information' in order to appreciate the artwork on the terms set up by the artist, the possibility to export and further develop some of the insights and capabilities developed in prior projects, or to re-engage with an issue that presented a block or a gap, would mean that cultural institutions could be making their programmes matter not only artistically, but also infrastructurally. In fact, within most other industries, building new infrastructural capabilities by leveraging new technologies is a standard practice.

> **"Thus, within companies or state projects, R&D labs are typically driven by risk-taking and experimentation in pursuit of industry-specific innovation."**

Thus, why shouldn't the same logic of care that curators are taught to apply to artworks also extend to the operational processes, nascent technologies and especially networks of people that make these artworks possible? If anything, this turn has been prophesied since at least the advance of computational technologies in the 1950s and1960s. In his 1968 'Systems Esthetics' article-manifesto, curator and theorist Jack Burnham stated that 'we are now in transition from an object-oriented to a systems-oriented culture [where change] emanates not from things, but from the way things are done'. [1] Some what ironically, Burnham's now largely obscured vision finds more resonance in individual artistic practices than in art-institutional agendas.

Is the time finally ripe for art institutions and cultural organisations to take Burhnam's proposal seriously? Judging by the appearance of new or revamped initiatives in recent years that echo Burhnam's organisational agenda, we may be finding ourselves at a critical juncture in the practice and (self-)perception of cultural institutions. [2]

**R&D as a Step into the Future?**

In the Serpentine's case, such new initiatives take the form of an R&D Platform — a dedicated operational space where care for and development of internal and external infrastructures that support innovative cultural production can consolidate, find new shape and flourish.

The term 'R&D' was first coined by American officials in 1947 and quickly spread across international policy contexts. Referring to company activities that did not need to yield immediate (or even mid-term) returns on investment, research and development (or prior to 1947, just 'research') R&D was seen to be essential to scientific and technological advancement. Thus, within companies or state projects, R&D labs are typically driven by risk-taking and experimentation in pursuit of industry-specific innovation.

One of the most glaring 'problems' with the R&D framework as informed by the last 150 years of corporate and state-funded projects is that in the logic of market competition at a global scale, the ultimate aim of R&D has been to secure a comparative advantage over one's political and economic adversaries, often at grave human (and environmental) cost. Yet, it would be wrong-headed to equate R&D with one particular ideological formation and historical arch. Although R&D has been formalised as organisational activity with industrialisation, '[informal]R&D has existed at least since the first person experimented with methods of knapping flint to make stone age tools'. [4]

> **"Art institutional R&D practice could become a vehicle for applying the wealth of critical knowledge generated in the arts in addressing complex contemporary issues in conversation with other fields."**

Approaching R&D from this slightly wider perspective than its specific application in the recent history of commercial industries and geopolitical warfare allows us to address the inherent tensions that this organisational framework poses in a more constructive manner. It also allows us to see more clearly the relationship that R&D may have to contemporary art institutions. Beyond functioning as a resource for internal infrastructural renewal [5], targeted investment in R&D within the cultural field could lead to the humanities and social sciences rejoining science and technology in defining what counts as socially valuable innovation [6]. Art institutional R&D practice could become a vehicle for applying the wealth of critical knowledge generated in the arts in addressing complex contemporary issues in conversation with other fields. And equally, the challenge of having the opportunity to be part of those conversations could lead to more ambitious and productively complex art, which previously could not have been hosted by the art institution.

There is a sense in which the logic of R&D is neither entirely foreign nor new to the art institution. The brief historical overview rendered above clearly shows that despite the conservative nature of institutions, arts organisations have been relatively responsive to artists engaging with challenging new issues in a speculative manner. Those arts organisations that were formed with a 'new media' focus have been particularly conscious of their role as platforms for artistic R&D, although, admittedly, this is not a language they'd ever use [7]. The hesitation is still very much rooted in the suspicion that formats and approaches too closely associated with science and technology might ultimately be corrupted by instrumentalisation, which art, according to popular opinion, should try to evade.

**"While there may be partnerships and coalitions, we have not yet seen a concerted effort within the cultural sector itself to embrace and establish standards or large-scale projects of cooperation."**

It is thus imperative to stress that the aim for R&D in the arts is not to import ready-made formats into the art field for the sake of narrow commercialisation, but rather to define an art-field specific view on innovation. To this extent, an R&D platform hosted by an arts institution would need to have a different agenda from even the kind of experimental cross-disciplinary R&D ventures that have been historically hosted by corporations such as Bell System and XEROX.

The art-institutional R&D agenda would need to support the evolving nature of art and its role in society. It is clear that today many emerging artistic practices require a much more robust art-institutional infrastructure in order to offer something distinctly valuable and to avoid falling behind those who are directly backed by technology-producing corporations and/or may use the corporate form to organise as a cultural actor [8]. While the art field by definition cannot deliver the kind of specialist insight that emerges in scientific and strictly technological settings, and nor is it ever likely to have matching capital to leverage for its production processes, arts institutions can consolidate and further develop their capacity to aggregate, synthesise and distil what is most relevant and critical to the wider societal climate.

For now, despite the move by individual arts organisations to set up R&D-style projects, there is still nothing resembling an open-source community around arts development [9]. There seems to be a lack of general awareness on the part of museums and cultural organisations that work could be consolidated between them in a way that makes the entire field more resilient. While there may be partnerships and coalitions, we have not yet seen a concerted effort within the cultural sector itself to embrace and establish standards or large-scale projects of cooperation.

Still, the proliferation of ad-hoc cultural initiatives that take the need for arts organisations to evolve past their historically established models as a starting point in experimenting with different ways of organising what and for whom they could produce is one important step in the direction of building a new ecosystem responsive to the requirements of technological and societal change.

*'Research & Development at the Art Institution' by Victoria Ivanova and Ben Vickers was originally published on* Serpentine Galleries website*, and is republished in Stages with kind permission of the authors.*

[1] Jack Burnham, 'Systems Esthetics', in *Artforum* (September 1968), p. 31, accessed via https://monoskop.org/images/0/03/Burnham_Jack_1968....

[2] High-profile international examples include New Inc at The New Museum, LACMA Lab, MoMA R&D, The New Normal and Terraforming programmes at the Strelka Institute in Moscow, and nationally, National Gallery X, New Work Department at the National Theatre, Digital Development at Royal Shakespeare Company, as well as the more specialised missions of such organisations as Watershed, Furtherfield and Abandon Normal Devices.

[3] Similarly, there may be a concern around the assumed secrecy of R&D ventures. While this may still be true of military R&D, a lot of other sectors have now shifted towards an open innovation model that prizes exchange of information, sharing of resources and feedback over siloed confidentiality.

[4] Bronwyn H. Hall, 'Research and Development', contribution to the *International Encyclopedia of the Social Sciences*, 2nd Edition (2006). Available at: https://eml.berkeley.edu/~bhhall/papers/BHH06_IESS_R&D.pdf

[5] Between 2012 and 2015, Nesta – UK's innovation foundation, Arts Council England and AHRC (Arts and Humanities Research Council) oversaw a seven million pound Digital R&D Fund supporting digital innovation in the cultural sector

[6] Peter Holme Jensen, CEO of Auqaporin, a bio-technology company with headquarters outside Copenhagen, articulated this as a concrete goal when reflecting on the experience of collaborating on Primer – an artistic research project hosted by Aquaporin.

[7] See, for example, Ars Electronica in Austria, ZKM, and Transmediale in Germany, WAAG in the Netherlands, FACT in England.

[8] A glaring example of what a direct union between a technology producing company such as Epson and an interdisciplinary art-tech collective can yield is the teamLab museum. Using Epson's 3LCD projectors for immersive digital installations, teamLab, a collective of programmers, engineers, CG animators, mathematicians and architects in the hundreds, are setting a very high benchmark for contemporary art institutions in terms of the capacity and technology at their disposal.

[9] Examples from external and related fields include w3, citizen science, github, MIT open source licenses, Creative Commons community and CERN.

## Victoria Ivanova and Ben Vickers

Victoria Ivanova is a curator, writer and strategic consultant. She is R&D Platform Strategist at *Serpentine* within the Serpentine's Arts Technologies team. Her core focus is on systemic and infrastructural conditions that shape socio-economic, political and institutional realities.
*Ben Vickers* is a curator, writer, publisher and technologist. He is Senior Strategist at large with the *Serpentine* in London, Co-Director of *Ignota Books* and an initiator of the open-source monastic order *unMonastery*. He serves on the boards and advisory panels for Light Art Space, Transmediale, Auto Italia, Furtherfield, Complex Earth, SXSW Arts Programme and the Warburg Institute in London.

# Future Art Ecosystems (FAE): Strategies for an Art-Industrial Revolution

Serpentine R&D Platform & Rival Strategy

**Future Art Ecosystems** (FAE) develops strategic insights for practitioners and organisations across art, science, technology and policy. FAE is a joint effort of Serpentine Galleries' R&D Platform and Rival Strategy. Taking form of annual policy documents, each issue responds to an emerging ecosystem situated between "culture" and science, technology, and commercial industry. FAE conceptualises infrastructural shifts, and details practical options and emerging strategies for relevant actors in the field. The first volume, on art x advanced technologies, was released in mid-2020. Included in *Stages* is Chapter 3: **'Strategies for an Art-Industrial Revolution'.**

***

The *infrastructural plays* detailed in the previous chapter tend to be undertaken in relatively local and ad hoc ways. A museum may buy equipment to host an AI project, a tech company may put out an open call for artistic collaborations or an artists' studio may launch a digital product.

This chapter outlines strategies that more overtly draw together multiple infrastructural plays into broader configurations. They involve building substantial ecosystems that support AxAT projects more broadly, providing integrated ways to fund, produce and distribute them. As such, they have both the intention and the potential to create revolutionary shifts, generating new eco- systems of activity that only partially intersect with the current landscape of
the art industry.

The strategies outlined here offer frameworks for articulation and cooperation between art, artists and advanced technologies. Each also implies a certain conception of the place and function of art, with implications for how artists access technology, the spaces in which they present their work, the financial models available to them and the risks involved for those participating in them. The general description of each strategy is followed by a summary of its strategic significance for the various actors involved.

### The Tech Industry as Art Patron

***Art as a source of opportunities for the technology sector***

This strategy builds on a substantial history of large corporations working with artists, especially in the US, and notably centred on the electronics industry and its transformation into the Silicon Valley model.[64] Famous historical examples extend from Bell Labs to Xerox PARC, and have frequently taken the form of programmes that give artists on-site access to technological equipment, technical support and expertise.

Under the terms of this strategy, there is an exchange primarily between the artist and a team working under the auspices of a corporation, typically through *tech residencies* and *tech provision*. Other actors from the existing art industry ecosystem may also be involved, for example museums or galleries.

> **It's important [with these engagements] that it's not just presenting the approachable, acceptable face of a new technology, where there's no criticality towards [the technology], it's just kind of like a demo. It's like demo art of someone else's tech. Holly Herndon [65]**

A common argument from cultural institutions for brokering these relationships is that artists are working 'upstream' of developments in consumer technologies, with the implication that their work explores opportunities for the application of these technologies.[66] There is a strong historical tradition of tech companies engaging with artists in this way.[67] However, there is no necessary linear relation between these experiments and later product development, and the *tech industry as art patron* strategy boasts a more sophisticated conception of the role of art in relation to industrial concerns. Indeed, it is relatively common knowledge in the tech business that there is no solid relationship between providing spaces for the free exploration of new technology and product development—even when these spaces come in the form of internal 'innovation labs' that do not involve artists whose values may clash with those of the business.[68]

More complex motivations for tech companies to engage with artists can be understood as a portfolio of potential advantages:

1. Organisational learning, from the level of individual employees and teams working with artists, to divisions and global governance. This is effectively the 'product innovation' model, but without a linear conception of product development—rather, it places a general value on exposing organisational culture to alternative perspectives on technology and its application, thus challenging assumptions rather than straightforwardly providing 'solutions'.

2. Domain-specific knowledge and expertise benefiting the usability of emerging technologies. As one example, spatial technologies expanding into areas that have historically been the domain of fields such as architecture or theatre —there are specific techniques, processes and insights that can be translated to advance the usability of new technologies such as VR, AR and AR cloud.

3. Providing public-facing PR and CSR opportunities, through the exhibiting of specific groundbreaking projects and general 'support of the arts'.[69]

4. Signalling a commitment to innovation to external investors and internal stakeholders.

5. Signalling a commitment to creativity and innovation to prospective (younger) employees in talent pools where hiring is increasingly competitive and for whom workplace values/culture plays an important role in attracting such talent.

6. Providing space for employees to engage in temporary (i.e. full-time but not permanent) or part-time pursuit of their own projects in collaboration with artists, for the purposes of professional development and staff retention.

7. Leveraging the art world, broadly understood as an epicentre of creativity with deep cultural import, as a place to secure a boost for organisational reputations as actors of fundamental importance in contemporary society for a public audience.[70]

Given the diversity of these potential benefits, the *tech industry as art patron* strategy may be seen more as an 'experiment' for the tech company than as a bid for the pursuit of any specific, stated, long-term objectives. Hence, some companies that adopt this strategy do so in the form of an open platform.[71]

It is conceivable that this strategy could extend into creating new venues for commissioned work.[72] However, it also aligns with a policy of drawing on the expertise, reputation and audience of established cultural institutions (an inversion of the 'success in adjacent fields' principle that is one aspect of the *common features of emerging practices* described in Chapter 1). This suggests a deepened relationship between existing art industry actors and tech companies. However, this strategy also introduces a swathe of new tensions in the interactions between art and tech cultures.

In the first case, it may be that given the fringe relationship of art to its core mission, a tech company may only provide ongoing support to a small number of arts institutions within the same region. Secondly, this support is not necessarily long-term, being subject to shifts in corporate governance and changes in overall company strategy.[73] These factors introduce a degree of turbulence into the art industry, as large-scale economic actors from elsewhere move in and out of the field.

Lastly, this strategy creates complexity in the ambitions and objectives native to the art world and those of corporate policy. The wider actions of large commercial companies may adversely interact with the arts ventures they support on many levels, providing new twists on the ongoing scandals around corporate sponsorship of artistic programmes.[74] The contradiction between economies of scarcity and the value placed on large-scale operations in industry also creates structural problems, and indeed there are discrepancies at the general level between the cultures of tech and art.[75] [76] It may also be that the low-level operations of these collaborations foster uncomfortable conditions for some artists.[77]

***Strategic Significance***

*For AxAT artists:*

access to skills, equipment, and expertise; potential ethical and political risks.

*For the tech industry:*

exposure to alternative ways of thinking about their technological development pathway, deep historical knowledge and domain expertise in areas that are undergoing technological change— implying a range of associated benefits and risks.

*For cultural institutions:*

technically sophisticated work to present to the public; a potential collaborator or competitor; potential ethical and political risks.

*For private sector investment:*

tech industry itself displaces some channels of private sector investment (e.g. collectors), and lowers market circulation; potential investment in spin-offs from larger companies; real estate development and public-private partnership access points for urban regeneration projects through supporting tech sector/cultural sector interactions.

*For public sector investment:*

city- or national-level branding/soft power; state role supporting early stage innovation; ability to cross-over tech innovation and cultural sector funding.

Open questions

- What would a museum fully owned and operated by a technology company look like and who would be its audience?
- How far can AxAT projects ultimately impact the development pathway of products, services and platforms within a tech corporation?
- How can much smaller tech organisations be involved?
- What role do governmental or academic science and engineering programmes have to play in the configuration, regulation and nurturing of these new relationships?

### The Art Stack

***Art as the driver of ambitious large-scale projects provided directly to the paying public.***

> **I think artists in general are actually quite bad at imagining how to make their dreams come true at a bigger scale. Bigger not necessarily in terms of grandness, but more complexity.Ian Cheng**

A second strategy is based on the consolidation of both AxAT infrastructural plays and existing aspects of the art ecosystem into a new format: the *art stack*.

Art stacks are artist-led organisations that progressively bring together in-house functions currently distributed between artists, curators, galleries, museums, tech companies and others involved in AxAT projects. The seed of the art stack strategy lies in the need for AxAT artist studios to develop *integrated studios* around *DIY approaches to tech*. The art stack builds on this position by locating a revenue stream —one that gives it autonomy from common funding sources in the art industry (e.g. sales to collectors, or project-specific funding from a company or governmental body). In turn, this creates opportunities for the art stack to invest in itself, and to build and control its own versions of other features currently provided by the art industry, such as places to show work.[78]

Artist-led companies such as teamLab present one vision of the art stack strategy, combining *integrated studios*,*DIY approaches* and well-equipped *collective spaces* with *dedicated display spaces* and funding through *ticketed experiences*.79 At the time of writing, teamLab has over 650 personnel ranging across art, architecture, animation, coding, marketing, robotics and other disciplines. It has also built its own site in Tokyo—teamLab Borderless, operated in collaboration with the Mori Building—to host its large-scale immersive digital works. Borderless opened in 2018 and attracted 2.3 million visitors in its first year, making it the most popular single-artist museum in the world as measured by footfall.[80]

This demonstrates the potential of art stacks to expand to a larger scale than many well- known current museums—an observation that has precedent in the power-law distributions that have emerged in other media across the cultural sector, accompanying a shift from a craft-based model to an industrialised one: Hollywood movie studios, major record labels, the Italian development of the fashion house system, videogames and social media.

Where reliant on *ticketed experiences*, the art stack operates in proximity to the financial models of circuses and theme parks: mass-market models organised around ticketed access. For some actors in the art world, this may raise the question of whether they are indeed 'art spaces' or just a variation on existing entertainment typologies. More generally, a direct-to-consumer, mass-market model organised around ticketed events (or in future, perhaps product design, digital services, etc.) may raise the question of *minimal viable art* for those who remain attached to older models of the cultural institution and art industry more generally —i.e. What is required for these initiatives to be understood as 'art' at all?[81]

Seen from a different point of view, 'minimum viable art' challenges preconceptions around the anticipated scale (of team-size, turnover, physical dimensions, etc.) of existing art practices; and it may be that it invites connection to quite other art histories which are not always obvious to the current generation of Western critics (or other audiences).[82] This demonstrates the possibility of a successful art-industrial phenomenon that publicises an alternative conceptual engagement with what art is and could be—one that diversifies away from the existing narratives of the mainstream contemporary art world.

The art stack holds the promise of a much richer engagement between artists and technology, within dedicated environments (physical, technical, presentational and commercial). Art stacks may be modelled around quite a different financial core, such as *building tools* or selling *art products*, and may explore other routes to the public, such as deep use of online spaces.[83] But they also offer a model of artistic practice that is substantially different from what is widely valorised in the art world at present.

'Minimum viable art' aside, two factors in particular stand out. The first is that the operational model of 'the artist' becomes something almost entirely team-based. This diverges from the 'individual artist' model preferred by the existing art world (and often presumed by the art industry), to a much greater extent than 'a collective' or the kinds of approaches favoured by *integrated studios*. Although it is possible that a relatively flat hierarchy might be adopted inside some art stacks, the contrast in expectations from current art training and professional life are nonetheless very substantial, placing an emphasis on skills for negotiating complex, ongoing work relationships within common projects where personal or small-group authorship is diminished.

The second factor is the uneasy relationship between many extant artistic practices, including those involving advanced technologies, and the kinds of commercialisation necessary to fund an art stack. The possibility of generating art stacks has been refused many times in the past, including by pioneering AxAT artists.[84] Art stacks require a very particular negotiation of the relationship between commerce and art, and this may filter both the practitioners and the practices that are able and willing to generate them.

***Strategic Significance***

*For AxAT artists:*

a new, art-led structure for those whose work fits with it, capable of operating at a new level of artistic ambition; lowered reliance on contemporary models of artistic funding (i.e. existing channels of private and public investment).

*For the tech industry:*

potentially new high-level collaborative or competitive relationships; a sophisticated content pool that can be ported to emerging platforms.

*For cultural institutions:*

source of technically sophisticated work; a potential collaborator but also competitor.

*For private sector investment:*

lower influence of collectors and auctions; potentially profitable early-stage investment, art stack IPOs; potential real estate development and public-private partnership access points for urban regeneration projects through supporting tech sector/cultural sector interactions.

*For public sector investment:*

city- or nation-level branding/soft power; potential for standout tourist destinations, state role supporting early stage innovation.

Open questions

- Over the mid-term, how far will art stacks be distinguishable from organisations in entertainment or product design?
- Over the long-term, to what extent can the art stack model be expected to disrupt and undermine traditional models of singular authorship, both from a symbolic perspective and the operational reality of offering a more attractive context for specialists to contribute their skills?
- What would an art stack for services look like?
- What will be the impact of the art stack model on arts education?

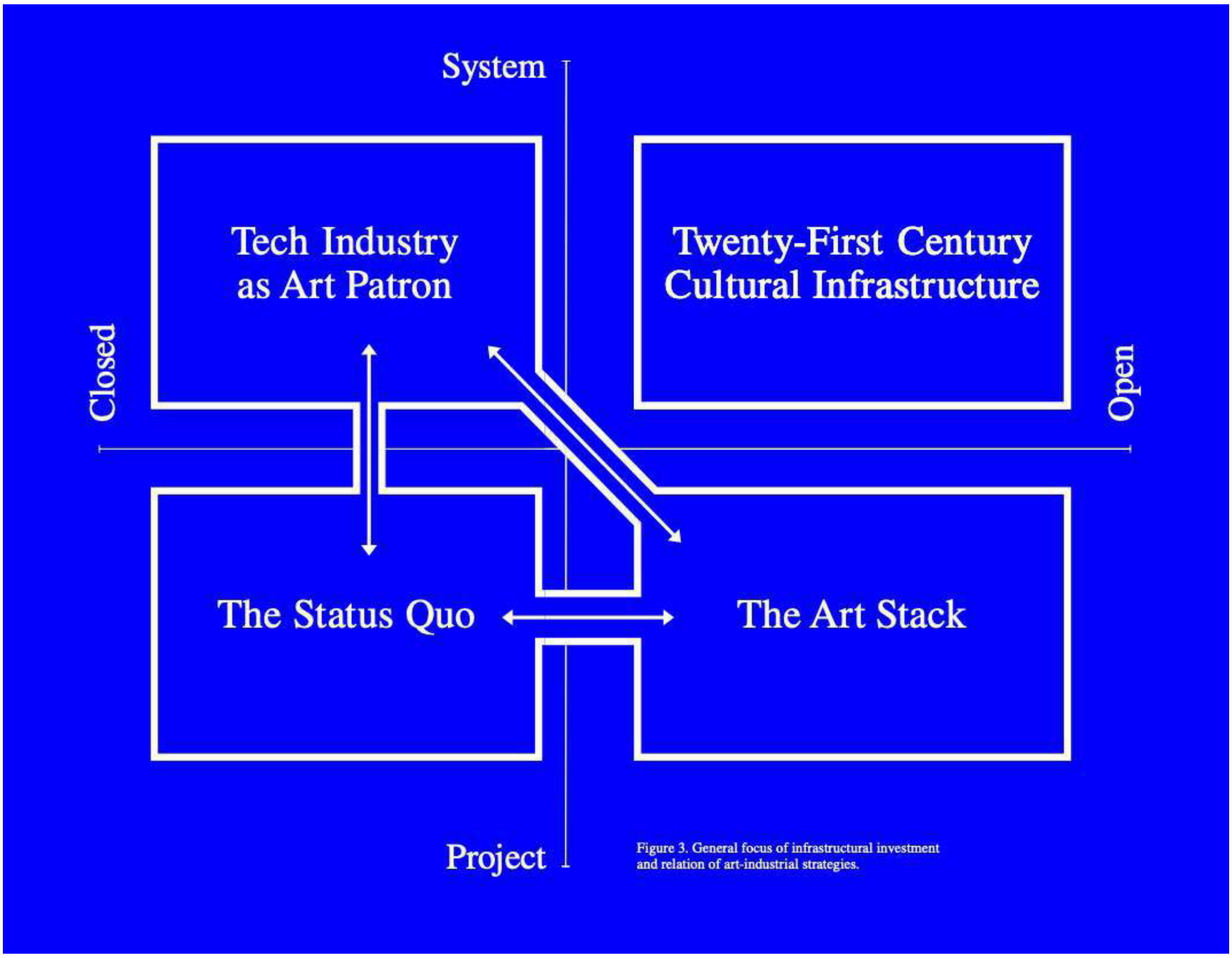

System
Tech Industry as Art Patron
Twenty-First Century Cultural Infrastructure
Closed
Open
The Status Quo
The Art Stack
Project
Figure 3. General focus of infrastructural investment and relation of art-industrial strategies.

Figure 3. General focus of infrastructural investment and relation of art-industrial strategies

**Twenty-First Century Cultural Infrastructure**

**Art as a strategic societal asset**

***We need new institutions to deal with the new problems that are emerging.*** **Holly Herndon**

The strategies of the *tech industry as art patron* and the *art stack* represent major disruptive vectors in the existing art industry. They represent new movements poised to redistribute the balance of power in the contemporary art world landscape.

They clearly demonstrate the potential for certain strands of AxAT to scale up their operations substantially. But the particular modes of scaling they offer are ultimately constrained by the financial, operational and strategic demands of very particular kinds of large-scale private-sector organisations, be they tech firms operating as patrons or sponsors, or *art stacks* themselves.

In contrast, the third strategy described here involves the conscious development of a *twenty-first century cultural infra- structure*.This strategy entails the construction of systems designed to support the AxAT ecosystem as a whole, and which are aligned with and responsive to a broad societal agenda.

> **A lot of questions that aren't being asked by artificial intelligence scientists and investors are being asked, and have been asked for quite a long time, by some kinds of artist... In a very hard, pragmatic way, this art is becoming relevant to the moment we are about to live through. Jonathan Ledgard [85]**

As described in the introduction to this document, AxAT can be understood as a form of technological innovation that is conditioned by a very different approach to technology —how it is developed, deployed, used and valued. AxAT practitioners frequently work with technologies that may have major societal benefits, but as yet do not synchronise well with existing funding regimes.

- Working with very early stage technologies with no clear pathway to immediate application, or those that have potential for application but do not readily fit with either consumer-focused retail or existing major infrastructural plans, and therefore are yet to find a pathway out of the laboratory.[86], [87]
- Operating to actively critique existing means of technological development, e.g. artist Trevor Paglen and AI engineer Kate Crawford's ImageNet Roulette, which identified racist patterns in the AI encoding of the ImageNet public image database, leading to the withdrawal of over 600,000 images.[88]
- Using technology to provide alternative approaches to non-technological domains, extending AxAT's principle of *success in adjacent fields* into a tangible, quantifiable impact on systems of collective decision-making such as government and law. An example is Forensic Architecture's *Grenfell Tower Fire* project, which draws data from smartphone footage taken by members of the public of the devastating fire at the London apartment block in 2017, in order to reconstruct the order of events—an operation that enters into the legally charged context of determining accountability for the disaster.[89]

The twenty-first century cultural infrastructure strategy is responsive to the value provided by such projects, while acknowledging that their widespread development requires an approach not easily reconciled with the strategies detailed previously. The *art industry* capabilities necessary to effect this strategy vary widely, and it is unlikely that a single actor at less than national government scale could adopt them all. This strategy is therefore best represented through a federation of efforts to bring infrastructural plays into alignment, at different levels and scales.[90] The central components of the strategy include:

*Alternative routes to access tech.*

The development of systems that lower the barrier to access of advanced technology, in ways less dependent on patronage or the ongoing negotiation of sponsorship, and enabling a maximally diverse set of practitioners and perspectives to engage with technologies at all stages of development. These can be envisioned as third-party systems that enable AxAT practitioners working in specific subfields (e.g. VR, synthetic biology) to develop and display work in environments, such as existing galleries or museums, that cannot on their own contribute sufficient capital investment to develop in-house skills, equipment and capabilities to host this work.[91]

*Legal arrangements.*

Building on the tradition of experiments with artist's contracts, the development of new ways to enable engagement between partners on AxAT projects.[92] On one level, this means finding alternatives to the common three-month residency arrangement which are better suited to the cost, time frame and collective nature of serious AxAT projects. On another, it means broaching imminent legal questions spurred by AxAT technologies themselves, such as the legally complex debate about whether the person who provides data used to train a machine learning system has a claim to its products.[93] Additionally, existing means of representation for artists, an essential art-industrial function of galleries, may be inadequate to the demands of AxAT practice, and may both require and reward serious innovation.[94]

*Learning and insight.*

The generation of new knowledge by AxAT practices is an asset in its own right, and not purely in terms of intellectual property. A logical development of AxAT skill-sharing (a semi-official feature of *multidisciplinary courses* and *collective spaces*) is the development of new kinds of venues in which to share what has been learned.[95] This also extends to the strategic deployment of AxAT practices as sources of collective insight into unfolding conditions, and accordingly suggests a place for government departments, legal bodies and other 'non-technological' agencies in the commissioning and development of such work.[96]

*Distribution systems.*

Current experiments from within AxAT such as *building tools*, *art products and byproducts as assets* have, to date, largely conformed to models widely adopted within the tech industry—for example, retail of designed products to individual consumers, or seeking venture capital investment. On the other hand, while there has been innovation around *designing purchase mechanisms*, they have not (or not yet) achieved widespread adoption.[97]

While not per se exclusive of input from either the *tech industry as art patron* or *art stacks*, this strategic approach is more closely aligned with the mission of cultural institutions and the various bodies that support them (such as foundations, funding councils and government departments). It represents an extension of these bodies' mission to maximise the audience of cultural projects on the grounds of their significance to broader society—albeit also constituting a series of breaks with how this role tends to be understood at present.

> **Cultural institutions should play a role in helping point public attention to the things that we should be paying attention to. And those are usually things which are not in the top headlines, which are not beholden to the advertising industry and not necessarily responding to political talking points. They should play a beacon or spotlight role. Noah Raford [98]**

The most obvious infrastructural plays available to existing cultural institutions such as museums and galleries are those that enable them to retrofit AxAT into current systems. For example, a new or existing museum might build *dedicated display spaces* to host AxAT work. This is a major capital investment, with particular risks.[99] But while valuable in its own right, this only treats one aspect of the AxAT ecosystem, and deeper shifts in operations would be necessary to engage fully in the project of building twenty-first century cultural infrastructure. Likewise, this strategy would be expected to align with national- or international-level governmental policies around the support of both the arts and innovation, but bring them together in historically new ways.[100]

***Strategic Significance***

*For AxAT artists:*

greater autonomy with respect to tech industry; lower barriers to access to advanced technologies; other ways to scale impact of projects, outside of traditional art, tech or entertainment industry channels.

*For the tech industry:*

opportunities for small-scale and/or emerging-technology developers.

*For cultural institutions:*

a pathway to alternative operational models.

*For private sector investment:*

opportunities to be involved in emerging technologies not married to conventional startup pathways.

*For public sector investment:*

production of insight and intellectual property as strategic assets at societal level; alternative system to develop genuinely innovative ideas.

Open questions

- What would a major public art institution look like without physical exhibition or performance spaces?
- What type of metrics would be needed to evaluate the impact of work that exists within art and also outside art?
- How can cultural institutions support the development of technologies that do not satisfy the contemporary funding conditions of the tech industry?
- How can AxAT be a part of national or international industrial strategy, and what would be the impact of this on the cultural sector?
- At what point does this strategy constitute the incorporation of an 'alternative tech industry'?
- Is it possible that such a large-scale initiative could separate from the art world as currently understood and becomes autonomous, with its own funding mechanisms, institutions and discourse—a hard fork in the art world?

*****

*'Future Art Ecosystems', Chapter 3: 'Strategies for an Art-Industrial Revolution' (2020) by Ben Vickers, Chief Technology Officer, Serpentine and Victoria Ivanova, R&D Platform Strategist, Serpentine, is taken from the publicly available document* here, *and republished with kind permission of the authors.*

[64] There are historical examples of artists placing themselves in social and commercial partnerships, for example John Latham's *Artist Placement Group*. Link: bit.ly/2Tfu6mh

[65] Holly Herndon is a Berlin-based American composer, known for sophisticated integration of digital systems and especially artificial intelligence with the human voice and live performers.

[66] Paris Innovation Review argues that placing artists within cutting-edge research programmes helps with *decompartmentalisation*, helping researchers to innovate and learn from other fields. Link: bit.ly/2NiTBzt

[67] Natalie Jeremi jenko's *Live Wire* installation, designed at Xerox PARC, is an early example of physical interfaces to networks being deployed as a *ubiquitous computing* experiment. Link: bit.ly/2FH0oyU

[68] Simone Bhan Ahuja recently argued in *Harvard Business Review* that 90% of innovation labs fail because placing research in a *laboratory setting* isolates it from meaningfully engaging with the goals of organisation. Link: bit.ly/2FDblBv

That said, the approach has produced some significant successes over the years, most famously at Xerox PARC. Link: bit.ly/30aBWPA

[69] Corporations frequently engage with the arts as part of their corporate social responsibility work, i.e., business commitments to reinvest a fraction of profits into projects of social benefit.

[70] The prevailing art world discourse may position art as critically reflective on the broader culture; but it may be this asserted criticality itself that makes art an attractive vehicle to corporations keen to present themselves as culturally sophisticated. Link: bit.ly/2TiRgs3

[71] Primer is an arts platform based in the headquarters of Danish biotech company Aquaporin, which describes itself as being *'intended as a platform for production, development and support for artists and the field of art in general, exploring its introduction into new spaces and professions.'* Link: bit.ly/2RcpGdi

[72] Apple have recently launched several augmented reality programmes, developed with artists and educators in collaboration with the New Museum. These include in-store events under the rubric of Apple AR[t] Labs and related AR[t] Walks through public urban spaces. Link: apple.co/2tU1nJ3

[73] A userful warning about the limited attention span of corporations investing in art programmes is the closure of the Interactive Design Institute Ivrea in 2005, after only four years of operation. Link: bit.ly/30cBVur

[74] As a thought experiment, it is entirely possible given the possibility of nation-state and supra-national legal moves against social media networks (e.g. anti- monopoly legislation, media regulation)—plus scandals such as Cambridge Analytica's involvement with Facebook—that the support of artists by such companies could trigger a backlash and become branded as *'trustwashing'*.

[75] As Mike Pepi summarises '*Christie's thrives on scarcity. Google does not.'* Link: bit.ly/2TeR8d7

[76] For example, see Lucy Sollitt's 2019 report for Creative United on The Future of the Art Market, which highlights some of the urgent challenges faced across the arts in adapting to new forms of techno-economic infrastructure. Link: bit.ly/3b5QPXw

Note further that, while hard data is difficult to acquire, there are many accounts of tech industry figures being favourably disposed toward art and artists but being extremely skeptical of the art industry's systems of valuation. Link: bloom.bg/2yyK9DZ

[77] *'If you were working with a developer and coming up with idiosyncratic approaches towards a specific machine learning architecture, then another artist comes into that residency and the developer takes some of those ideas and applies that to the next person—that's something that can be really problematic in an arts context. Likewise, if you have the same developers working with a large pool of artists and you have one specific approach towards technology that is then funnelled into different practices rather than having dramatically different approaches'.* Holly Herndon

[78] It should be noted that large-scale studios are not themselves unheard of in the history of art. For example, Rubens was famous for his huge workshop filled with students and apprentices, whilst at one point Damien Hirst employed 250 people, worked with high budgets and opened a museum. Such ventures, however, typically have been lacking some of the features of AxAT practice itemised in Chapter 1, and represent a continuation of conventional art industry models under new ownership, as it were, rather than a break with the status quo as indicated by the kinds of infrastructural plays documented in Chapter 2. Link: bit.ly/2FFDM1J

[79] teamLab run their own 10,000-square-metre digital art museum in Tokyo. Link: bit.ly/37PI6HE

[80] Tickets to teamLab's *Borderless*cost approximately $30 in 2018, when they attracted 2.3 million visitors.

[81] An alternative conceptualisation might be that art stacks exceed *maximum viable art*, given that they operate beyond the financial and organisational models that have predominated in the art world to date.

[82] teamLab locate reference points for its expansive immersive environments in premodern Japanese art, specifically what it calls *ultrasubjective space*, which offers an alternative conception of the optical relation of viewer to artwork, based in premodern Japanese pictorial traditions rather than Western linear perspective. The viewer imagines themselves as a component of a depicted scene, rather than observing it from the periphery. Link: bit.ly/2Tqedd9

[83] As in the case of pop artist KAWS, the output of whose work spans limited edition vinyl toys available to the mass market, large-scale sculptures positioned within the contemporary art milieu, and collaborations with fashion brands such as Supreme and Nike.

[84] Pioneering biotech artist Oron Catts worked with early stage tissue culture technologies. Despite the evident art stack potential—via an art product or building tools modelling his early work developing *victimless meat*and *victimless leather*—Catts sees the commercial development of these ideas as symptoms of consumerism and antithetical to the deeper concerns of his practice. Link: bit.ly/37Y7eMg

[85] Jonathan Ledgard collaborates with artists on technology and nature, is a novelist, expert on AI and robots particularly in Africa, foreign and war correspondent for The Economist.

[86] Protocells are an example of an early-stage technology with no clear pathway to immediate application.

Link: bit.ly/2slv05W

[87] Neighbourhood-level electricity generation is an example of a potentially significant technology that does not readily fit with either consumer focused retail, nor existing major infrastructural plans.

[88] Trevor Paglen and Kate Crawford:

*'We created ImageNet Roulette as a provocation: it acts as a window into some of the racist, misogynist, cruel and simply absurd categorisations embedded within ImageNet. It lets the training 'speak for itself', and in doing so highlights why classifying people in this way is unscientific at best, and deeply harmful at worst.'* Link: bit.ly/37V4Fuu

[89] At the time of writing, Forensic Architecture are crowdsourcing video footage of the Grenfell Tower fire in order to projection map an accurate 3D video of how the fire progressed through the building. Link: bit.ly/2taLhL9

[90] *Federation* is used here to mean something similar to *interdependence*, as advocated by Holly Herndon and Mat Dryhurst as a principle for an alternative to the *independent music scene*, focused on complex ecosystems of new organisations and financial models, and evolving relationships to audiences and tools. Link: bit.ly/2t8Nqak

[91] One mechanism for opening up routes of access to technology would be the provision of platforms to enable consortia to be built around AxAT-related capital investment from cultural institutions, much as is the case on major academic science and engineering projects like CERN.

[92] W.A.G.E. is an activist organisation working to establish sustainable economic relationships between artists and the institutions that control the art world. Link: bit.ly/2RbUMSB

[93] Property rights over personal data has evolved into a heated debate, and the knock-on debate over who owns the intellectual property of technologies created from that data is likely to become even more contentious as those products become more valuable. Link: bit.ly/2FDhEFb

[94] One could make a comparison to the growth of *label services*in the music industry. Traditional record labels provide artists with a portfolio of services (management of publishing rights, making arrangements with stream services, pressing records, tour organisation) in return for a contract that is usually exclusive and long-term. Label services disaggregate these functions into individual services that artists can opt into and out of, as and when needed. Link: bit.ly/2tTN3AE

[95] As part of a recent retrospective at London's Institute for Contemporary Art, Forensic Architecture ran a series of skill-sharing short courses in forensic architecture, offering the public training in techniques they had developed. Link: bit.ly/30h6Oya

[96] Relatively small-scale initiatives like the UK government's Policy Lab currently take a version of this approach, although largely without engagement of the kinds of technology with which AxAT practitioners are working. Link: bit.ly/2TfWfKg
The most serious investment in this strategy to date is arguably the Dubai Future Foundation and the related Museum of the Future. Link: bit.ly/2NjSAXM

[97] Attempts at building AxAT distribution systems have tended toward a degree of conformity with legacy art industry practices, such as aligning with a model of value as being produced by scarcity.

[98] Noah Raford is Futurist in Chief and Chief of Global Affairs at the Dubai Future Foundation.

[99] The high cost of systems needed to display AxAT works, which include both technology and the expertise to deploy and maintain it, is itself prohibitive, and represents a major investment in a new capability for existing gallery or museum models. ROI for existing galleries or museums is further complicated by the tendency to rotate exhibitions—a dedicated display space not in continuous use offers a relatively poor return.

[100] Government investment supporting arts and innovation might be understood as a reanimation of the frequently unrecognised role played by governments in the original development of many contemporary technologies during the twentieth century. Link: bit.ly/2RaqJdM

## Serpentine R&D Platform & Rival Strategy

Serpentine R&D Platform: Ben Vickers, Sophie Netchaef, and Victoria Ivanova
Rival Strategy: Marta Ferreira De Sa? and Benedict Singleton

Edited by Robin Mackay, original document designed by Mark Hurrell

# Curating Data: Infrastructures of Control and Affect ... and Possible Beyonds

Magda Tyzlik-Carver

My body is a structure that carries my feelings. I have nothing else to do this job. But in the recent months, my feelings have not been able to cope with the task of holding my body. The tension that has taken over my body since the start of the Covid pandemic changes in intensity depending on too many factors. Sometimes these are obvious, but most often the task of identifying them all is laborious, time consuming and simply boring, and risks losing sight of life as it happens. The feeling of being split, divided, extended and spread across networks, Zoom calls and business still unfinished is overwhelming. I take vitamin C, D and echinacea; I make ferments to grow microbes that nourish my body. And I am increasingly attached to my computer, which has become a window onto life that happens somewhere else, while becoming my reality.

Yet, of course, my life also happens here, where I sit. I keep some of my interactions away from the computer, but the number is falling. Even now, in writing this text, I am capturing my feelings with my computer. I have situated my body in time and space while mapping feelings and attaching them here: selecting sensa tions, archiving moods, displaying moments. Curating. And in the process I remain connected and react to tweets, to Whatsapp or Signal messages from friends, looking at Instagram pictures via Facebook and so on. I choose tools and apps to make connections or simply to stay connected, weaving networks, engaging with systems and becoming part of them. I am sensing how it feels to become posthuman, a body of data and affect.

To suggest that (digital) life today revolves around curating might be too much of a reduction. Life, even the digital one, is so much more complex than a process of selection and arranging things. It is more than an archive or a database. Such a contraction, however, allows us to consider two important elements in contemporary curating and information networks: control and affect. On the one hand, curating is a method of exerting some form of control in a system that requires arranging relations into directories. In the process of generating data and creating folders, daily curatorial decisions include naming files, deciding on their format (.txt, .jpg, .png, .avi, etc.), organising them and choosing storing devices, and so on. Users collect and archive data routinely, and in the process, their lives too become data to be managed and organised. On the other hand, while network connections allow expansion beyond one's own computer or mobile phone, the volume, velocity and veracity of links made, files exchanged and data captured is such that trying to make sense of these data is a task optimised predominantly for machines with a platform profit as the main objective.[1] It is simply not possible for users to make sense of big data, especially if such sense-making is based on models that require 'the capture and manipulation of people's activities and environments'.[2] And as curating becomes increasingly posthuman it takes place at different levels.[3] Curating has become an organised form of control executed by algorithms and made possible by big data, while also directly affecting people whose lives have been incorporated into digital infrastructures that maintain the system, a necessary element for the profitable performance of Facebook, Google, Amazon, Microsoft and Apple, to name only the biggest five.

No wonder that today sense-making has become a project that is engineered by computer science. For many users of digital devices, feeling the world is synonymous with checking social media status. The main form of interaction with the distributed information system that is the worldwide web is based on algorithmic models for news delivery and purchasing products. This is a structural transformation. But affect and emotions are as much part of this system as algorithms and data. News feeds and personalised results offered when logging in online are managed by optimisation models designed to 'treat the world not as a static place to be known, but as one to sense and co-create', which too often results in 'social risks and harms such as social sorting, mass manipulation, asymmetrical concentration of resources, majority dominance, and minority erasure'.[4] The poverty of possibilities that computer science models offer are currently displayed online and made visible as bias in AI systems and image data sets, polarisation of opinions that feed and display hate while dividing communities and people, sensationalist conspiratorial narratives that feed click-bait economy, and many more effects that play out algorithmic scenarios and forms of co-creation of the new world. Making sense is about truth games played by computer science, driven by solutionism without ever asking 'What is it for?', 'Who will it harm?', 'Who does it serve?', 'Should it be introduced into the world?'.

These structural transformations of life take place through processes of digitisation, digitalisation, datafication, AI and Big Data. And while predominantly driven by poor computer science in the service of a market logic that wants more clicks and even more data with which to play and which to model for growth that profits a few, we are reminded to think about transitional forms and structural transformations. For Lauren Berlant, structure is '*that which organizes transformation*' and infrastructures maintain it.[5] Infrastructure is 'the living mediation of what organizes life: the lifeworld of structure binds us to the world in movement and keeps the world practically bound to itself."[6] Infrastructures are living forms for relating and connecting to other humans and nonhumans. That's where living happens, where lives collide and encounter harm and pleasure, damage and nourishment. And today, in the time of pandemic and a fight for power where people are continuously counted, the daily practice of being alive is becoming a practice of being online, en masse, yet alone, often at home, or migrating, but always counted.[7] It is experience of digital life on social-media platforms where there is no community but algorithms released by the union of computer science with market.

Figure 1. Maintaining Indeterminacy, Loren Britton (2020), image courtesy of the artist.

This damaged (digital) life needs other imaginaries that can take into account the terms for transformation with digital structural forms that hold the conditions for ‘infrastructures of sociality’.[8] Different ways of knowing, and not knowing need to be part of this. We need to imagine what could be possible, how to draw connections, and to be in touch, yet with less certainty. These are the questions that Loren Britton and Helen Pritchard ask in their call for:

> **multiple CS practices: computer science, chance and scandal, committed survival, care and shelter, chocolate and strawberries, cushions and support, collective strategies, chancer scientist, cohabitation and sharing, conditions and structures, choice and scandal, careful slug, collective scandal, crip studies, composed silliness, compulsory sleep, cancelled stories, crying sabotage, carceral states, cut and scale, considerable scaffolding, collapsing species, collective suffering, companion story.**[9]

Britton and Pritchard recognise that computer science is a practice that can be queered, that can be challenged, and that it does not need to be in the service of world destructions. With their challenge to

Computer Science, they explain:

> **What we are making space for with our CS figure is the destabilization of what CS is defined and known to be. CS has been debated historically and can be debated still, but who gets to determine what comes to matter is still very much dependent on moments of translation, moments in which that which might not be recognized as CS becomes so. Instead of asking who gets to have a voice, we ask: Which practices get to produce knowledge?**[10]

These questions and their dream of 'computer science otherwise' mobilise desires for another CS beyond the field of informatics and computation. [Fig.1] Reclaiming CS as a figure and not just a discipline resists its normalisation and turns it into an active proposal with indeterminacy as the value that defines the possible, and unknowing as a process of situating and recognising of what one knows while choosing to move away from it and to start anew.[11]

The point of not knowing makes space to account for knowledges that have been excluded by a Eurocentric tradition of science that has been defined by a colonial mindset. Indigenous Protocols and Artificial Intelligence Working Group offers another challenge to computer science and its impotent imaginaries. Situated within and speaking from the position of many different Indigenous concerns and communities in Aotearoa, Australia, North America and the Pacific, the group asks a series of questions that situate Indigenous knowledges and their communities as directly engaging in 'conceptual and practical approaches to building the next generation of A.I. systems'.[12] This is a project not undertaken for diversity's sake, but rather because of the belief 'that Indigenous epistemologies are much better at respectfully accommodating the non-human'.[13]

One such epistemology is Lakota ethical protocols that look seven generations ahead, thus already establishing ethical relations with the future from a situated present. As Suzanne Kite explains, her research into Lakota protocols is based in the ontological status of stones for the Lakota people.[14] Kite, herself a member of Lakota Nation, works with Lakota stone epistemologies as a framework that also defines relations with raw materials used in building computers by asking at what point materials, objects or nonhumans are given respect, and how ethical relationships with raw materials are established when building computers of all kinds. Kite explains how Lakota knowledge is not static and so working with that knowledge is a practice of change and identifying shifting networks of relations that have to be accounted for. As she says:

> **The effects our decisions – and technologies – have on the world can help us identify the stakeholders in what is being made and how it is used. Stakeholders in Indigenous communities are identified as our extended circle of relations, while stakeholders in technology companies are identified as the board of directors, shareholders, employees and consumers. It is necessary to identify how all those – both human and nonhuman – are affected by what is made, and to take responsibility for those it affects.**[15]

This possibility to account for change by accounting for relations with humans and nonhumans is part of the Lakota responsibility that manifests as a commitment to do things the Good Way.[16] Protocols offer a specific guide and Kite, in collaboration with Corey Stover and Melita Stover Janis, and with notes from Scott Benesiinaabandan, gives an example of how to build a physical computing device in a Lakota way. But such a guide extends to all elements of the AI system, from training sets to interfaces, because ethical AI is not possible if any one of its elements is extracted through exploitation.[17]

CS otherwise and Indigenous Protocols are examples of what structural transformations are imagined and desired, and how they can be made. The two are epistemological practices that build on histories and embedded knowledges of people in queer and Indigenous communities in art and science. And they are also a call to extend these as living experiences that inspire other imaginaries for practices and knowledges, creating infrastructures of computer science and AI that otherwise are inflexible and harmful.

The practice of curating data is also an epistemological practice that needs interventions to consider futures but also to account for the past. In curating data, knowledge results from *engaging* with data, and establishing material relations with and between them. Decisions about how to process, organise and name data can be identified, and the practice itself can be made accountable. When organising data such as images, videos and texts on a personal computer, this seems like a straightforward task, since one simply has to take the time and effort to organise these files into directories. When curating bigger collections of data and other digital things, to sustain accountability becomes a task of sharing responsibilities not just for the future of a collection but also for its origins.

This can be done by asking where do data come from? How are they connected to the communities that are their source, how is the data set/database created, what is the reason for curating these data, and how will this curating take place? It is true that similar questions could also be asked of colonial objects taken from communities and locations and moved into imperial museums to narrate the story of colonial empire. Here too we can ask about the conditions of this takeover, and how these histories are part of displaying objects. Responding to such questions accounts for the fact that data, too, come from somewhere, and that as well as being part of a database or a set, data start with bodies, human or not, and index relations between them in the most abstract way. The task in curating data is to reclaim their traceability, and to account for their lineage.

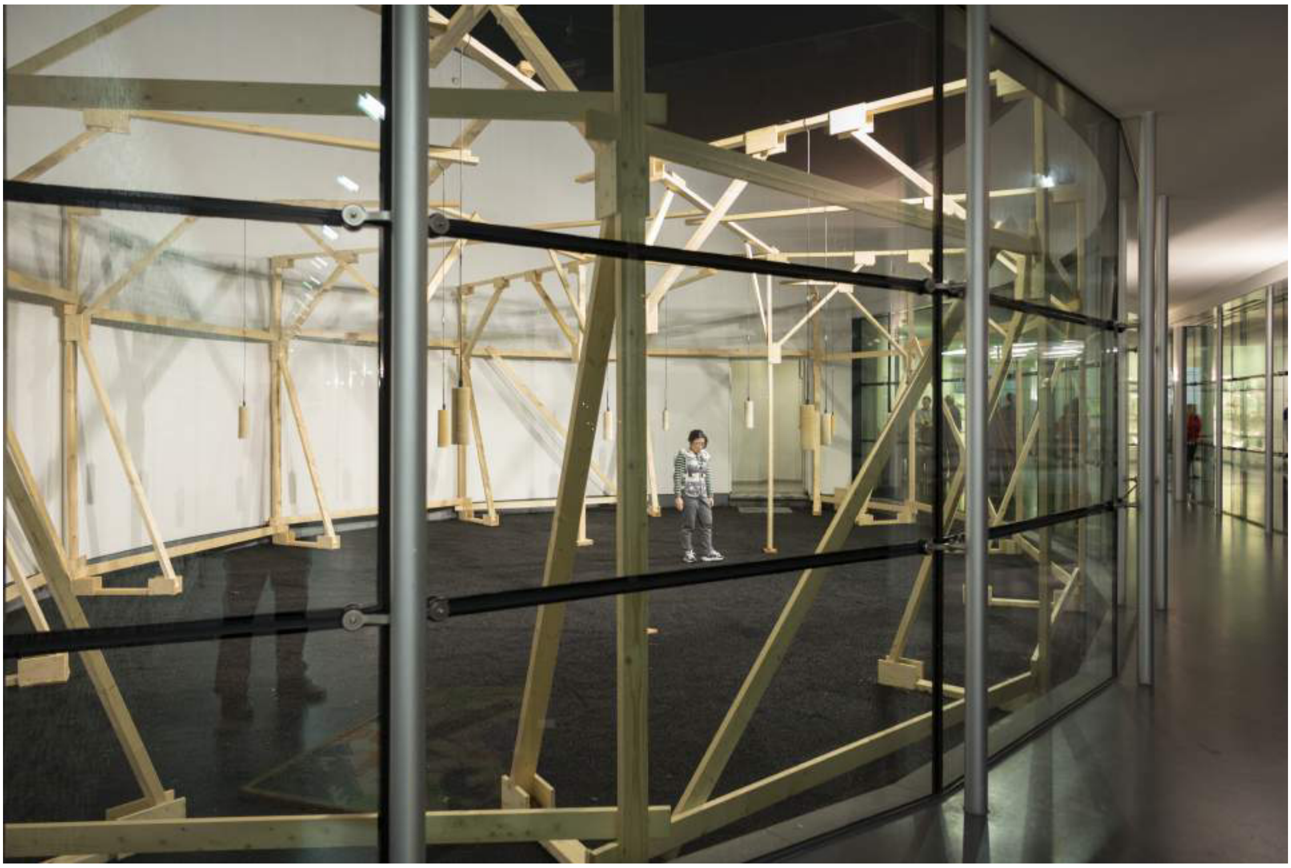


Figure 2. The Intimate Earthquake Archive Sissel Marie Tonn (2016-ongoing), Photo by Peter Cox, Van Abbemuseum, 2016, NL

Curating data is a material practice of embedding data and also exploring how it could be embodied. In other words, curating data is about how data becomes part of the world. The interactive installation *The Intimate Earthquake Archive* (2016 ongoing) by artist Sissel Marie Tonn and developed in collaboration with sound artist Jonathan Reus engages these questions in an attempt to understand the effects of long-term gas mining in the Groningen area of Holland.[Fig.2] The installation[18] and related ongoing research project involves data collected during man-made earthquakes that have occurred in the area over the last thirty-four years. The core sample data includes sand and soil tests and digital records of seismic activity stored by the Dutch Meteorological Institute.

The artists put data sets together with stories of how earthquakes are experienced by inhabitants of this area, some of whom regularly report waking up seconds before the earthquake, or whose personal perceptions of the place are affected by continuous anxiety about the possibility of an earthquake occurring at any time. By connecting digital data of seismic activities with the experience of the sensing body, Tonn and Reus produce a multimedia installation that offers a very different experience of this geo-located phenomenon from one that could be extracted from meteorological records with traditional forms of data analytics and visualisation.[Fig.3]

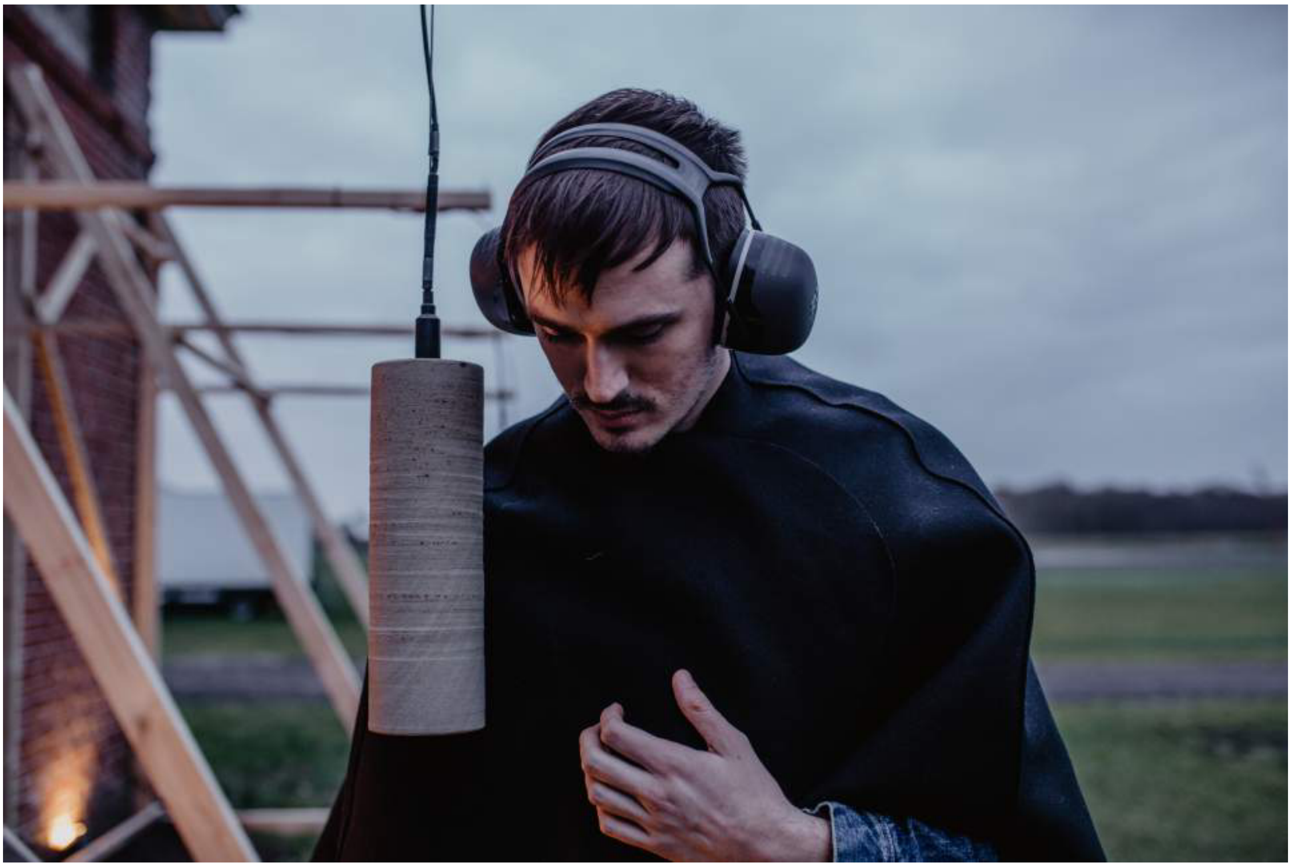


Figure 3. The Intimate Earthquake Archive, Sissel Marie Tonn (2016-ongoing), Photo by Stella Dekker, Grasnapolsky Festival, 2019, Groningen, NL. Courtesy of the artist.

Earthquake data is mediated into sonic vibrations that can be sensed or felt with the use of a wearable interface – a body vest with embedded surface/skin and bone conduction transducers that produce tremors on the surface of the body, mirroring the way the seismic waves move across the land. Datasets are manipulated into sonic vibrations and encourage deep listening with and within the body while moving through the installation area. According to the artist, this is a testing ground for attunement to future living in the Anthropocene. As such, the project offers a practical exploration of transitions that can be tested, and where different kinds of data can perform together and can be sensed.

In the computational regime, resources such as minerals and data are extracted, bodies moved, geological phenomena undergone, social realities constructed. In the environment simulated by *The Intimate Earthquake Archive*, not only does the body not end with the skin but it becomes part of the archive, extending it and transforming towards a practice of sensing, accessing affect, constructing affective data bodies. The archive becomes an environment for feeling and being moved by data that themselves are residues of movements of the ground. Here, data is curated into shifting, vibrating and sonified relations that roam fastened onto the bodies of visitors. And so here data can be felt not as emotion, but as a sensation in the body, a feeling of the body being moved. Artwork's structure accommodates motion as infrastructure for transformation of data back into what can be sensed and located in the body. [Fig.4]

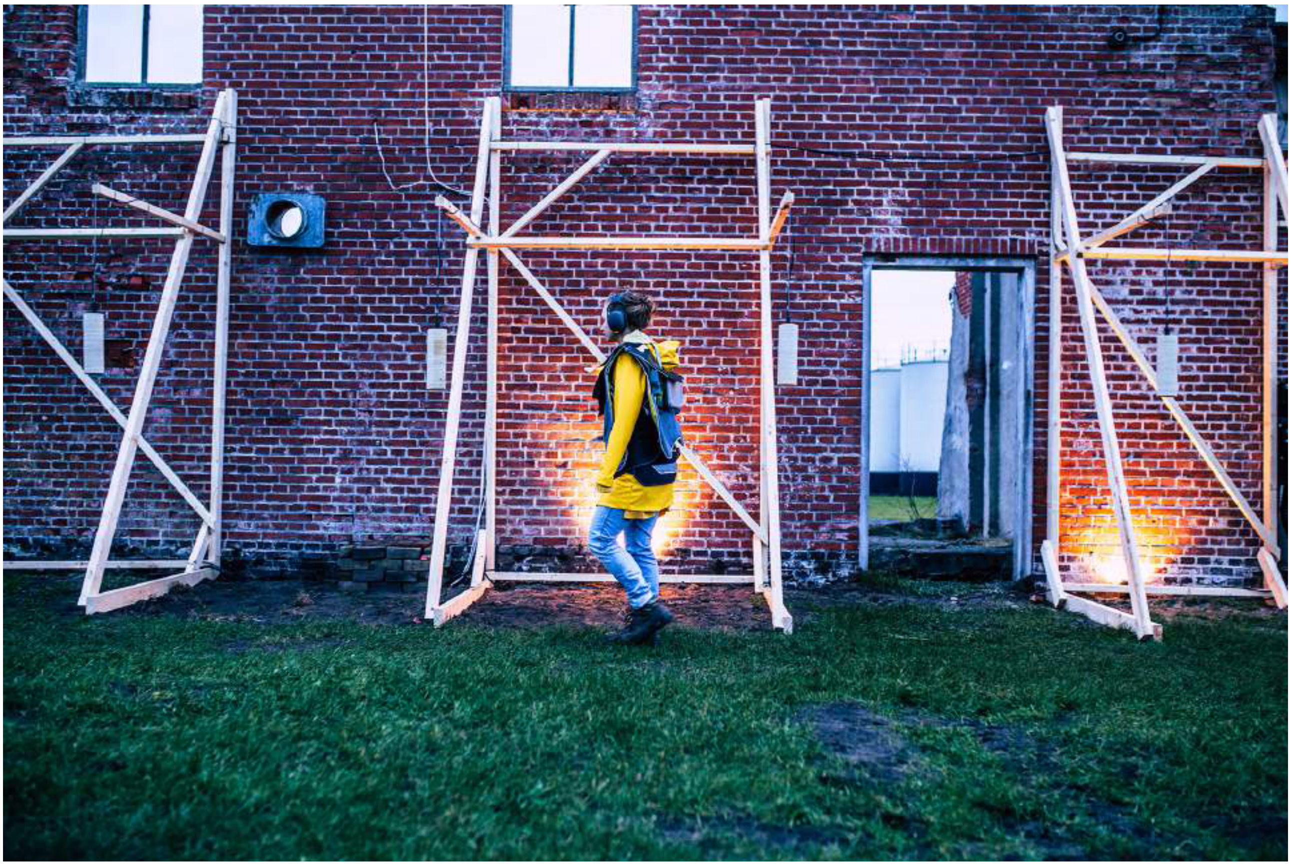


Figure 4. The Intimate Earthquake Archive, Sissel Marie Tonn (2016-ongoing), Photo by Stella Dekker, Grasnapolsky Festival, 2019, Groningen, NL. Courtesy of the artist.

The overwhelming grip of corporations over networked infrastructures, driven by a desire for exponential growth, disregarding injustice and distributing inequity through the system with the help of computer science, displays colonising logic.[19] Such logic harms already vulnerable groups, communities and their environments and there is an urgent need to look for other models and visions of life and its organisation online and off. Curating data as the process of feeling with data rather than data optimisation for profit, opens new political possibilities for enacting common becoming with CS as computer science, chance and scandal, committed survival, care and shelter, chocolate and strawberries, cushions and support, collective strategies, and many more. Britton and Prichard, Berlant, Kite, Indigenous AI Working Group, and Tonn and Reus, among many others, challenge and direct us towards our sensing in common. Curating data is common and it needs to be accounted for as a material practice and intervention into computational realities that can be otherwise, while also demanding responsibility (from ourselves and others and especially CS) to do things the good way.

[1] These are three of the five Vs that define Big Data. For an overview of the features of Big Data See '5 V's of Big Data', 2019, GeeksforGeeks (blog), 10 January 2019. https://www.geeksforgeeks.org/5-vs-of-big-data/. For critical questions see danah boyd and Kate Crawford, 2012, 'Critical Questions for Big Data', Information, Communication & Society 15 (5): 662–79.

[2] Seda Gürses, Rebekah Overdorf and Ero Balsa, 'POTs: Protective Optimization Technologies', Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 27 January 2020, 177–88, https://doi.org/10.1145/3351095.3372853.

[3] Magdalena Ty?lik-Carver, 'Curating in/as Common/s. Posthuman Curating and Computational Cultures', PhD (Aarhus: Aarhus University, 2016); Magdalena Ty?lik-Carver, 'Posthuman Curating and Its Biopolitical Executions: The Case of Curating Content', in *Executing Practices*, ed. Helen Pritchard, Eric Snodgrass and Magdalena Ty?lik-Carver (London: Open Humanities Press, 2018), 171–89.

[4] Gürses, Overdorf and Balsa, 'POTs'.

[5]Lauren Berlant, 'The Commons: Infrastructures for Troubling Times', Environment and Planning D: Society and Space 34, no. 3 (June 2016): 394, https://doi.org/10.1177/0263775816645989. Italics in the original.

[6] Berlant, 'The Commons', 394.

[7] As I am writing this, the counting of ballots in the US 2020 presidential election is still going on, and there are already calls to stop counting in some of the states.

[8] Berlant, 'The Commons', 394.

[9] Loren Britton and Helen Pritchard, 'For CS', *IX Interactions* (blog), July 2020, https://interactions.acm.org/blog/view/for-cs.

[10] Britton and Pritchard, 'For CS'.

[11] Ibid.

[12] INDIGENOUS AI', *INDIGENOUS AI* (blog), http://www.indigenous-ai.net/ (accessed 11 February 2020); Jason Edward Lewis et al., 'Making Kin with the Machines', 16 July 2018, https://doi.org/10.21428/bfafd97b.

[13] Lewis et al., 'Making Kin with the Machines'.

[14] Suzanne Kite, 'How to Build Anything Ethically', in *Indigenous Protocol and Artificial Intelligence Position Paper*(Honolulu: The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research, 2020), 75–85.

[15] Ibid., 76.

[16] Ibid.

[17] Ibid.

[18] The installation received an honorary mention at Ars Electronica in 2020.

[19] Some of the recent scholarly work on these matters include Simone Browne, *Dark Matters: On the Surveillance of Blackness* (Durham: Duke University Press Books, 2015); Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018); Mél Hogan, 'Big Data Ecologies', in *Ephemera: Theory & Politics in Organization* 18 (3), Landscapes of political action (August 2018), pp. 631–57; Ariella Aïsha Azoulay, *Potential History: Unlearning Imperialism* (London: Verso, 2019); Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity, 2019); Catherine D'Ignazio and Lauren F. Klein, *Data Feminism,* Ideas Series (Cambridge, Mass.: The MIT Press, 2020).

## Magda Tyzlik-Carver

Magda Ty&#378lik-Carver is Assistant Professor in the Department of Digital Design and Information Studies at the School of Communication and Culture at Aarhus University. She is also an independent curator. Her recently curated exhibitions and events include ScreenShots: Desire and Automated Image (2019), Movement Code Notation (2018), Corrupting Data (2017), Ghost Factory: performative exhibition with humans and machines (2015) and Common Practice (2010, 2013). She is co-editor (with Helen Pritchard and Eric Snodgrass) of Executing Practices (2018) a collection of essays by artists, programmers and theorists engaging in a critical intervention into the broad concept of execution in software. She is a member of Critical Software Thing group and of the editorial board for Data Browser series. She is also Associate Researcher with Centre for the Study of the Networked Image at the London South Bank University.

# The Next Biennial Should be Curated by a Machine - A Research Proposition

Joasia Krysa and Leonardo Impett

***The Next Biennial Should be Curated by a Machine*** is a research proposition - an inquiry into the relationship between curating and Artificial Intelligence (AI), and the possibility of developing an experimental system[1] capable of curating, based on human-machine learning principles.[2]

Making reference to the *e-flux* 2013 project 'The Next Documenta Should Be Curated by an Artist'[3] which questioned the structures of the art world and the position of curators within, this project extends the question to machines.[4] It asks how the counterpoint of automata might offer alien perspectives on conventional curatorial practices and curatorial knowledge? What would the next Biennial be like if machines intervened in the curatorial process, and helped to make sense of vast amounts of art world data that far exceeds the productive capacity of the human curator alone?

The project takes the form of a series of research and artistic experiments that explore the application of machine learning algorithms (a subset of AI) to curation of large scale periodic contemporary art exhibitions, such as biennials, to reimagine curating as a self-learning human-machine system. [5]

Under this overarching concept, two parallel experiments are developed in the framework of Liverpool Biennial: $B^3$(NSCAM) and AI-TNB.

$B^3$(NSCAM) is developed as a collaboration with artists Ubermorgen, co-commissioned with The Whitney Museum of American Art for its online platform artport, curated by Christiane Paul. [6] It uses archival text material and datasets from both commissioning institutions and processes them through a group of machine learning algorithms, collectively named $B^3$(NSCAM). [Fig. 5] Processing datasets (including curatorial texts) linguistically and semiotically, the AI system 'learns' their style and content, breaking and mixing them together. The generated texts are then presented to the user, with a degree of interactivity and 'branching', iteratively rewriting small parts of its own text at random.

A parallel experiment, AI-TNB is research experiment commissioned as part of UKRI/AHRC Strategic Fund: *Towards a National Collection* to explore machine curation and visitor interaction, taking Liverpool Biennial 2021 as a case study [7] [Fig.1] In this experiment, the biennial exhibition curated by Manuela Moscoso and taking place across multiple venues in Liverpool over the course of few months, is interpreted as a parallel machine-curated online version.[8] The resulting 'curatorial AI system' is an excercise in interaction through large datasets, using computer vision and natural language processing techniques with a focus on *human-machine co-authorship*.[9] [Fig. 2, 3, 4]

Our relationship to computers is rapidly changing and so are developments in automation (AI), and so is our understanding of creative practices, including curatorial practices. The overall project takes machine learning algorithms beyond the 'search engine' paradigm in which they have been mostly used to date, and instead considers them to be curatorial agents, working alongside human curators.[10, 11] There are a number of issues arising from this, such as the degree to which creativity is compromised by the 'intelligent' machines we use, as well as how biases become reinforced.[12] Algorithms are biased because certain elements of a dataset are more heavily weighted, and once a system is trained on this data, further errors follow that broadly reflect inherent human biases in society. Can something similar be said of the art world, where one might imagine there to be a shared 'dataset' of artists and curators that reflect biases inherent to the art world? If this seems far too simplistic, it becomes more interesting once these two operating systems are correlated, and when they become entangled, and to speculate on what each might learn from the other. It is not just a case of identifying concerns – such as around inclusion of marginalised communities or worries about the forms of creativity produced through AI – but also an opportunity to think about the transformation of human-machine relations and curatorial practices.

In undertaking these various experiments, the intention is to explore the application of machine learning algorithms to envisage alternative forms of exhibition-making and curatorial agency that dissolves hard distinctions between humans and machines. When the projects asks whether *the next biennial should be curated by a machine*, it posits further questions about emergent forms of creativity and the larger infrastructures within which it operates. What alternative practices might emerge from these entanglements, and what new perspectives on conventional curatorial practices and curatorial knowledge might be generated?

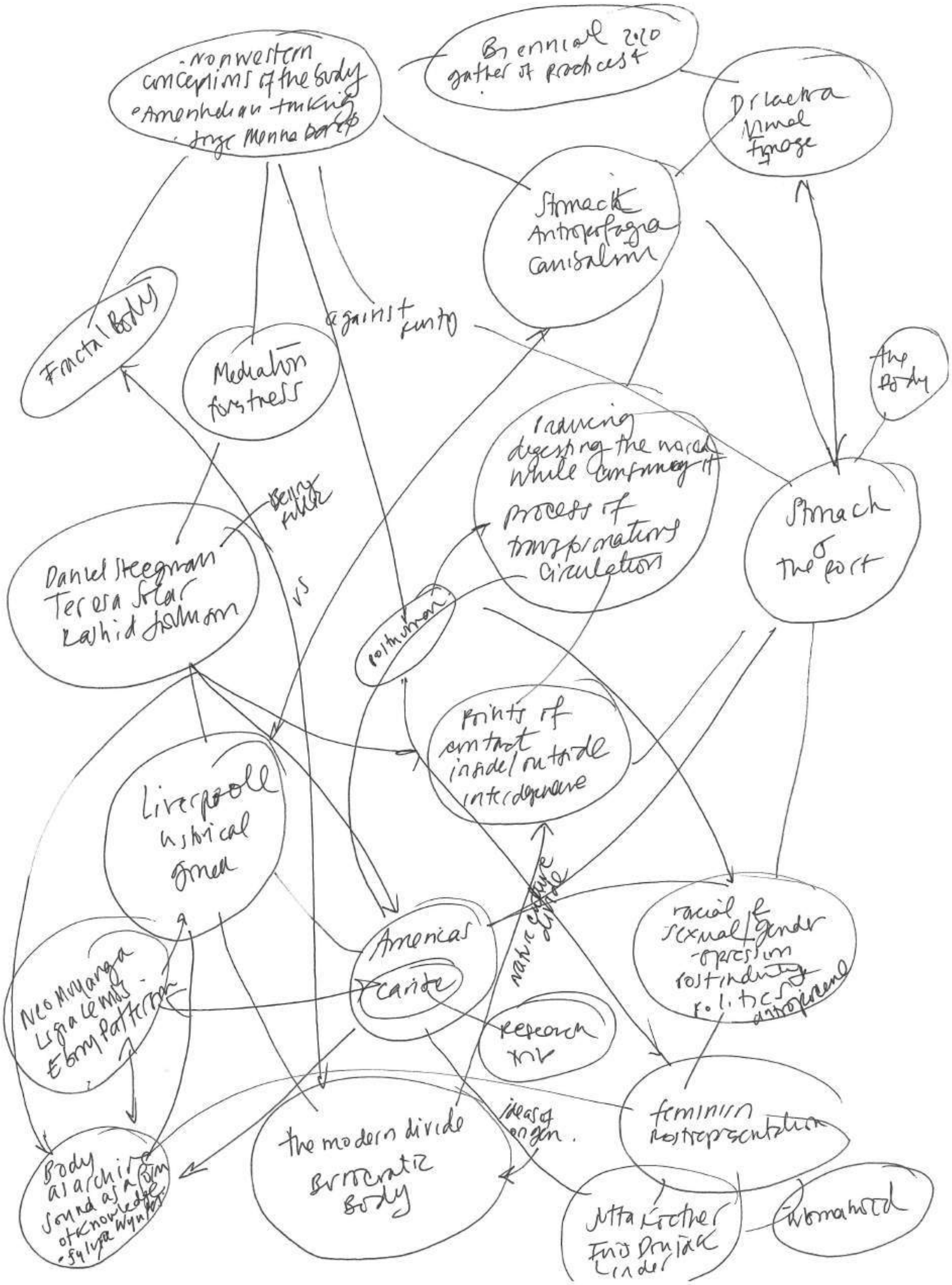
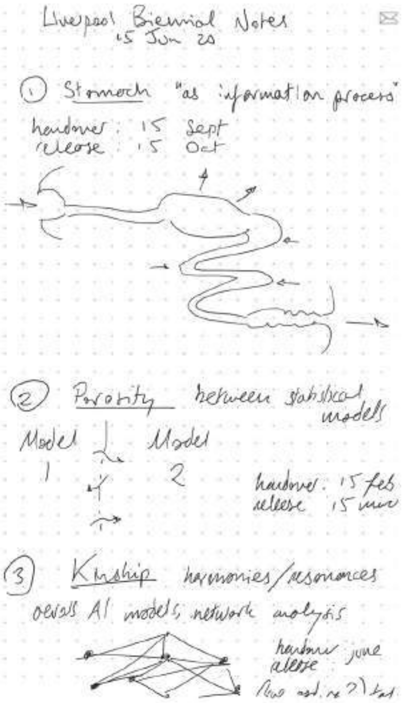

-nonwestern conceptions of the body
· Amerindian thinking
· Jorge Menna Barreto
Biennial 2020
gather of practices
Dr Lactea
Vimal
Fynose
Stomach
Antropofagia
canibalism
Fractal Body
Mediation
fortress
against purity
the Body
Learning
digesting the world
while consuming it
process of
transformations
circulation
Stomach
the port
Belly
vs
Daniel Steegman
Teresa Solar
Rashid Johnson
pollination
Points of
contact
inside/outside
interdependence
Liverpool
historical
Americas
Caribe
native culture
racial & sexual/gender
-opression
postindustry
politics
antropocene
Neo Muyanga
Ligia Lewis
Ebony Patterson
research
Body
as archive
sound as a form
of knowledge
· Sylvia Wynter
the modern divide
Bureocratic
Body
ideas of
origin.
feminism
postrepresentation
Jutta Koether
Enis Druiack
Linder
Womanhood

Figure 1. Curatorial sketch for Liverpool Biennial 2021, by its curator Manuela Moscoso (2019). Courtesy of Manuela Moscoso and Liverpool Biennial.



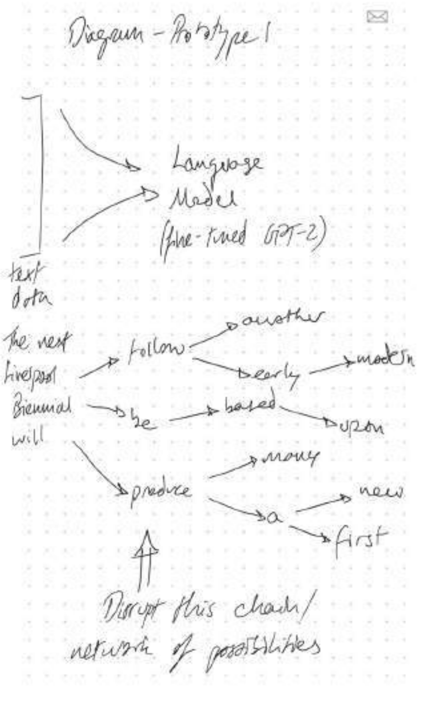


Diagram - Prototype 1
text data
Language Model (fine-tuned GPT-2)
The next Liverpool Biennial will
follow
another
early
modern
be
based
upon
produce
many
a
new
first
Disrupt this chain/ network of possibilities

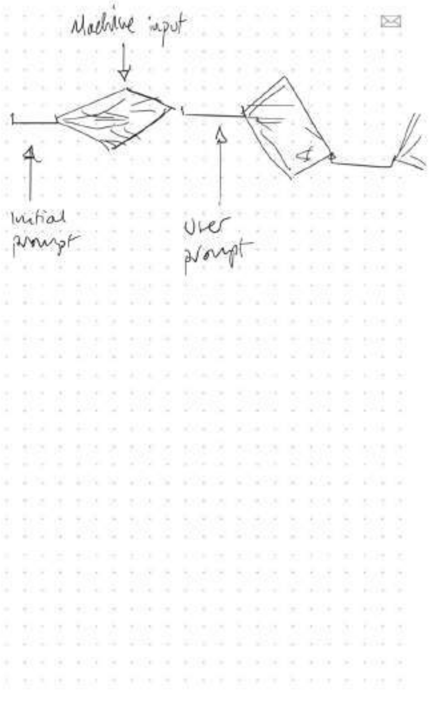


Figure 2 (Left). Plan for representing LB2021 themes at the neural-architectural level, Leonardo Impett, drawing (2020).

Figure 3 (Middle). Diagram for glitching decision trees of parallel text-hypotheses, Leonardo Impett, drawing (2020).

Figure 4 (Right). Sketch of the width of probabilities (probability distributions expanding and collapsing) under machine and user input, Leonardo Impett, drawing (2020).

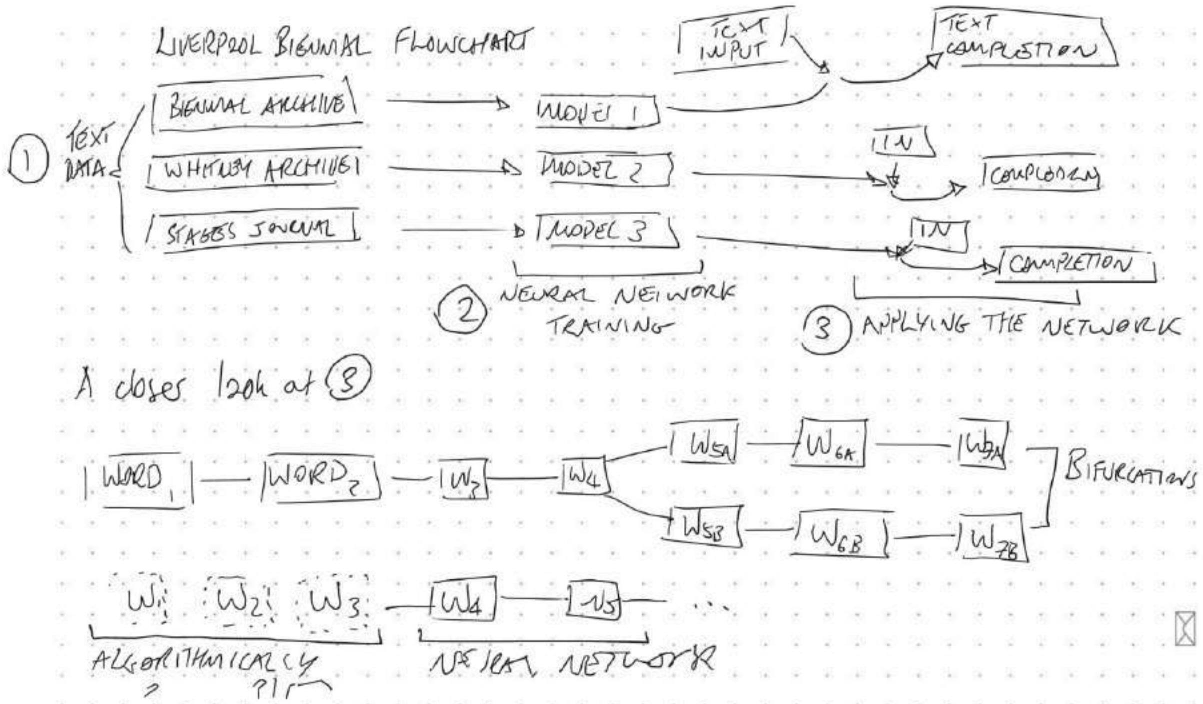


Figure 5. Sketch for planning multi-model bifurcations, Leonardo Impett, drawing (2020).

[1] For a definition of experimental system see: https://en.wikipedia.org/wiki/Experimental_system.

[2] *The Next Biennial Should be Curated by A Machine* is a research proposition and an umbrella concept that gathers various experiments exploring the application of machine learning techniques to curating; title and curatorial concept by Joasia Krysa, technical conceptualisation and development by Leonardo Impett, first experiment $B^3$(TNSCAM) developed as a collaboration with artists Ubermorgen, co-commissioned with the Whitney Museum of American Art for its online platform artport, curated by Christiane Paul. Further research funded as part of UKRI/AHRC Strategic Priorities Fund: Towards National Collection at: ai.biennial.com. This text draws upon our initial call released on eflux (2019): https://www.e-flux.com/announcements/291923/the-next-biennial-should-be-curated-by-a-machine/.

[3] e-flux, 'The Next Documenta Should be Curated by an Artist', 2013. https://www.eflux.com/announcements/42825/the-nex--documenta-should-be-curated-by-an-artist/.

[4] Krysa, Joasia, 'Can Machines Curate?', keynote lecture at the 5th National Symposium of the Brazilian Association of Cyberculture Researchers ABCiber 2011, published in Digital Art: fractures, proliferative preservation and affective dimension, edited by Yara Guasque, pp. 38-89. Coleção Fast Forward / UFG/Media Lab, 2014. Also see my earlier online project Kurator (2005), presented at Tate Modern and published in Curating Immateriality (2006) and as a chapter entitled 'Kurator - a proposal for an experimental, permutational software application capable of curating exhibitions' in *Networks* (ed. Lars Bang Larsen), Documents of Contemporary Art: Whitechapel Gallery and MIT Press (2014).

[5] Machine learning is defined as the study of computer algorithms that improve automatically through experience, as a sub-part of artificial intelligence. See Glossary in this volume for more detail.

[6] $B^3$(NSCAM) is presented at The Whitney Museum of America Art's online platform artport, curated by Christiane Paul. See our call for datasets released on eflux (2019): https://www.e-flux.com/announcements/291923/the-next-biennial- should-be-curated-by-a-machine/.

[7]11th Edition of Liverpool Biennial (2021) entitled The Stomach and the Port is curated by Manuela Moscoso across multiple venues in Liverpool, launched 20 March 2021. https://www.biennial.com/2021

[8] AI-TNB is developed as funded research 'Machine Interaction and Visitor Interaction in Virtual Liverpool Biennial 2021', part of UKRI/AHRC Strategic Priorities Fund: *Towards National Collection - Opening UK Heritage to the World*; project partners Durham University (Leonardo Impett, project PI), Liverpool John Moores University (Joasia Krysa), forthcoming at: ai.biennial.com/.

[9] Impett, Leonardo., Herman, I., Wollner, P. K., & Blackwell, A.F.. "Musician Fantasies of Dialectical Interaction: Mixed-Initiative Interaction and the Open Work", in *International Conference on Human-Computer Interaction* (Springer, Cham, 2018), pp. 184-195.

[10] Crawford, Kate and Vladen Joler, *Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data and Planetary Resources*, AI Now Institute and Share Lab, 2018. https://anatomyof.ai/.

[11] Impett, Leonardo, "Irresolvable contradictions in algorithmic thought", published in this volume (Stages 9/2021). https://www.biennial.com/journal/

[12] Noble, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press, 2018).

Joasia Krysa and Leonardo Impett

Joasia Krysa is a curator working at the intersection of art and technology. Her first curatorial software experiment was launched at Tate Modern in 2005 and published in Curating Immateriality (2006). She is Professor of Exhibition Research and Head of Art and Design at Liverpool John Moores University, with an adjunct position at Liverpool Biennial. Formerly, she served as Artistic Director of Kunsthal Aarhus, Denmark, part of the curatorial team for Documenta 13, and co-curator of Liverpool Biennial 2016 and Sapporo International Art Triennale (SIAF) 2020/21. She is international advisor for Helsinki Biennial 2021.

Leonardo Impett is Assistant Professor of Computer Science at Durham University. He works in the digital humanities, at the intersection of computer vision and art history. He was previously Scientist at the Bibliotheca Hertziana – Max Planck Institute for Art History, Digital Humanities Fellow at Villa I Tatti - the Harvard University Center for Italian Renaissance Studies. He is an Associate of Cambridge University Digital Humanities, an Associate Fellow of the Zurich Center for Digital Visual Studies, and an Associate Research at the Orpheus Institute for Artistic Research in Music.

# Glossary

This glossary is derived from Winnie Soon and Geoff Cox's Aesthetic Programming: A Handbook of Software Studies (London: Open Humanities Press, 2020), and published with kind permission of the authors, to provide a shared vocabulary for this volume.

### Artificial Intelligence (AI)

AI research is focused on developing computational systems that can perform tasks and activities normally considered to require human intelligence. The term 'AI' is often used interchangeably with 'machine learning' and 'deep learning', but there are key distinctions to be made. To explain: 'You can think of deep learning, machine learning and artificial intelligence as a set of Russian dolls nested within each other, beginning with the smallest and working out. Deep learning is a subset of machine learning, and machine learning is a subset of AI, which is an umbrella term for any computer program that does something smart. In other words, all machine learning is AI, but not all AI is machine learning, and so forth.' (https://pathmind.com/wiki/ai-vs) For a more critical explanation, see the essay and diagram by Kate Crawford and Vladan Joler, in 'Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources' (2018), https://anatomyof.ai/. For a decolonional perspective on AI, see Shakir Mohamed, Marie-Therese Png, William Isaac, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence', *Philosophy & Technology*, Springer (July 12, 2020).

### Dataset

A dataset (or data set) is a collection of facts. In the case of tabular data, a dataset corresponds to one or more database table, where every column of a table represents a particular variable, and each row refers to a given record of the dataset in question. Datasets are used by developers to test, train and evaluate the performance of their algorithms. The algorithm is said to 'learn' from the examples contained in the dataset and this constitutes the worldview of the algorithm (with colonial implications). See Nicolas Malevé's 'An Introduction to Image Datasets' (2019) for a useful summary of key concepts and concerns, and the extent to which datasets have become significant cultural objects (https://unthinking.photography/articles/an-introduction-to-image-datasets).

### Deep Learning

Deep learning is the subfield of machine learning that designs and evaluates training algorithms and architectures for modern neural network models. It is part of a broader collection of machine learning methods based on artificial neural networks with representation learning. Learning can be 'supervised', 'semi-supervised' or 'unsupervised'. See John D. Kelleher, *Deep Learning* (Cambridge, MA: MIT Press, 2019).

### Generative Adversarial Networks (GANs)

A GAN comprises two conflicting neural nets — a 'Generator' that forges new data, and a 'Discriminator' that distinguishes this fake data created by the Generator from real data. These nets challenge each other with increasingly realistic fakes, both optimising their strategies until their generated data is indistinguishable from the real data. This is an 'unsupervised' method of training that doesn't rely on the tagging of input images by humans, since the machine generates groupings based on its own analysis. The workshop 'Adversarial Hacking in the Age of AI' asked the question whether critical theory could learn from this system (that seems to resonate with dialectical materialism), in which everything is considered to be in a process of transformation through contradiction, and becomes a technical reality. The published outline provides a useful description of what is at stake: 'Adversarial attacks are an instance of how a machine-learning classifier is tricked into perceiving something that is not there, like a 3D-printed model of a turtle that is classified as a rifle. The computer vision embedded in a driverless car can be confused and not recognize street signs. Artists Adam Harvey, Zach Blas & Jemina Wyman, and Heather Dewey-Hagborg have utilized adversarial processes in their projects in order to subvert and critically respond to facial recognition systems. But this is not just about computer vision. Scientists in Bochum, Germany recently studied how psychoacoustic hiding can oppose the detection of automatic speech recognition systems.' See https://2020.transmediale.de/content/adversarial-hacking-in-the-age-of--i-call-for-proposals. For more on GANS, see Ian J. Goodfellow, et al, 'Generative Adversarial Networks' IPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2 (2014), pp. 2672–2680.

**Machine Learning (ML)**

Machine learning is broadly defined as a collection of models, statistical methods and operational algorithms that are used to analyse experimental or observational data. The term itself was coined by Arthur Samuel in 1959 during his game development research at IBM, which ultimately aimed to reduce or even eliminate the need for 'detailed programming effort', using learning through generalisation in order to achieve pattern recognition. See Arthur L. Samuel, 'Some Studies in Machine Learning Using the Game of Checkers', *IBM Journal of research and development 3 (*3) (1959), pp. 210–229. ML involves the development and evaluation of algorithms that enable computers to learn from experience. Generally, the concept of experience is represented as a dataset of historic events, and learning involves identifying and extracting useful patterns from a dataset. ML algorithms take a dataset as input and return a model that encodes the patterns the algorithm extracted (or learned) from the data. But to what extent does this process of generalisation present a problem inasmuch as the overall idea of learning implies new forms of control over what and how something becomes known and how decisions are made? See Adrian Mackenzie, *Machine Learners: Archaeology of a Data Practice* (Cambridge, MA: MIT Press, 2017).

**Model**

In machine learning, a model is a computer program that encodes the patterns that the machine learning algorithm has extracted from a dataset. There are many different types of machine learning models, but put simply, a model is created (or trained) by running a machine learning algorithm on a dataset. Once the model has been trained, it can then be used to analyse new instances. Named to reflect some of the problems associated with this, 'All Models' is a mailing list of critical AI studies hosted by the research group KIM at the Karlsruhe University of Arts and Design. The 'About' page prophetically states that 'All Models are wrong, but some are useful', to point to some of the limits of statistics and machine learning, and how, for instance, mainstream AI discourse stresses the need for unbiased data and algorithms to ensure fair representation, but overlooks the intrinsic limits of any statistical technique. Herein lies the politics and the way in which traditional forms of power (such as those related to gender, race, and class discrimination) are amplified. In summary, 'All Models questions all models!'. See 'About' (allmodels.ai).

**Neural Network**

A neural network is a machine learning model that is implemented as a network of simple information processing units called neurons. It is possible to create a variety of different types of neural networks by modifying the connections between the neurons in the network. Examples of popular types of neural networks include feedforward, convolutional and recurrent networks. A recurrent neural network, for instance, has a single layer of hidden neurons, the output of which is fed back into this layer with the next input. This feedback (or recurrence) within the network gives the network a memory that enables it to process each input within the context of what it has previously processed. Recurrent neural networks are ideally suited to processing sequential or time-series data. A good example of this is Helen Pritchard and Winnie Soon's 'Recurrent Queer Imaginaries' (2019–20), in which a machine learner 'Motto Assistant' continuously reworks queer and feminist manifestos to activate alternative imaginaries.

**Reinforcement Learning**

Reinforcement Learning is based on interaction with the environment, mapping an analysis of a situation into actions, typically used in robot control and game playing. The learner (or agent) does not have any previous data to base itself on in order to determine or predict which action to take, but rather learns by trial and error. This type of learning finds the optimum possible behaviour or path to take in a specific environment, mapping state-action pairs to achieve the best result. As in behavioural psychology, reinforcement is used to suggest future actions, like a pet or child getting a treat for doing what it was told. Unlike supervised learning that relies on input training data, the characteristics of reinforcement learning are that the programme understands the environment as a whole, and is able to learn from its experience by evaluating the effectiveness of each action taken: 'trial-and-error search' and 'delayed reward' are based on sequential decisions, computation, repeated attempts and feedback on the success of actions. See Richard S. Sutton, 'Introduction: The Challenge of Reinforcement Learning', in Richard S. Sutton, ed. *Reinforcement Learning. The Springer International Series in Engineering and Computer Science (Knowledge Representation, Learning and Expert Systems)* 173 (Springer, 1992), pp. 5–32. Having mentioned environment, it is important to emphasise that there are worrying environmental costs associated with machine learning. See, for instance, Karen Hao, 'Training a single AI model can emit as much carbon as five cars in their lifetimes', *MITk Technology Review* (June 6, 2019), https://www.technologyreview.com/s/613630/training-a-single-ai-model-can--mit-as-much-carbon-as-five-cars-in-their-lifetimes/.

**Supervised Learning**

Supervised learning is based on a training dataset with input/output pairs as expected answers. A classic example would be spam emails in which an algorithm learns from the sample of emails that are labelled as 'spam' or 'not spam'. The goal of this type of learning is to map the input data to output labels. For example, with new email as the input, what would the predicted output result be? Can it be classified as spam and then moved to the spam mailbox? In mathematical terms, this is expressed as Y=f(X), and the goal is to predict the output variable Y from the new input data (X). But this prediction process relies on classification techniques, for example binary classification (such as 'yes/no', 'spam/not spam', 'male/female') and multi-classification (such as different object labels in visual recognition), which is based on the process of data labelling. This is where inconsistencies arise. Data is categorised in a discrete manner, and there are many elements that might lead to a 'normative' prediction, especially problematic when it comes to complex subjects such as gender, race, and identity, because these operate beyond binary, discrete classification. See Joy Buolamwini, 'Response: Racial and Gender Bias in Amazon Recognition — Commercial AI System for Analyzing Faces'. *Medium* (2019), https://medium.com/@Joy.Buolamwini/response-racial-and-gen-er-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced.

**Unsupervised Learning**

Unlike supervised learning, unsupervised learning does not contain a set of labelled data. One of the common tasks with unsupervised learning is 'clustering' (algorithms such as K-means and Hierarchical Clustering). The goal of this technique is to find similarities, providing insights into underlying patterns and relationships of different groups in a dataset using exploratory and cluster analysis. Items in the same group or cluster share similar attributes and metrics. The idea behind clustering is to identify groups of data in a dataset, segregating groups with similar characteristics. It is commonly used in the business and marketing sectors to understand customer preferences so that personalisation and data marketing can be provided by grouping customers based on their purchasing behaviour with regard to certain types of goods. Artists Joana Chicau and Jonathan Reus have developed a performative project 'Anatomies of Intelligence' (https://anatomiesofintelligence.github.io/), which uses an unsupervised learning model to develop an understanding of anatomical knowledge and computational learning. In a workshop setting, they suggest that participants think of two features for examining a small image dataset (around fifteen images) — such as 'cuteness' and 'curliness' — and each of the images is rated and sorted according to these features. Each image can then be described by the set of feature values. As a result, several clusters are formed, providing a new perspective on the relations between images in terms of their similarities and differences. It's a simple exercise, but can be scaled up, systematised and automated, for example by deciding on the number of clusters and calculating the distribution of/distance between data points. This also helps reinforce how algorithms designed to recognise patterns, known as neural networks, operate, being loosely based on a model of the human brain and how it learns to differentiate certain objects from other objects.

**Winnie Soon** is an artist-researcher and Associate Professor at Aarhus University, Denmark. **Geoff Cox** is Associate Professor and co-Director of the Centre for the Study of the Networked Image at London South Bank University, UK.

# Colophon

Stages #9

The Next Biennial Should Be Curated By A Machine

**Published by**

**Liverpool Biennial in partnership with DATA browser series.**

**ISSN: 2399-9675**

Stages is an open-access, contemporary art journal published by Liverpool Biennial. It is a space for staging research generated in connection with Liverpool Biennial's artistic programme and appears at irregular intervals. Launched in 2013, Stages is freely available both online and in PDF format.

**Editor:**

Joasia Krysa, Manuela Moscoso

**Editorial Assistant:**

Abi Mitchell

**Copyeditor:**

Melissa Larner

**Web Design:**

Mark El-Khatib

**Cover Design:**

Manuela Moscoso (artwork), Joasia Krysa (words), Helena Geilinger (graphic design)