

Mess, Surprise and Conflict – Trying to figure (out) alternative ethical AI stories

Xaroula Kerasidou

In this short piece, I would like to present my work originating from my project titled 'Configuring ethical AI in healthcare' (2019-2021), and offer my reflections. The project- hosted by Lancaster University and funded by the Wellcome Trust- set out to explore the ways that ethical AI is being figured in healthcare with a particular focus on the UK, with the aim of opening up the possibilities for alternative configurations. Such a sentiment is at the heart of the adopted feminist Science and Technology Studies (STS) perspective which seeks to unsettle the dominant stories and voices, and open up possibilities for new ways of understanding, designing and living with technology. Part of the project was a co-design workshop where I planned to invite a variety of stakeholders to co-create critical, hopeful, and alternative visions for ethical AI providing the groundwork for future innovative methodologies and practices.

However...

As the Covid-19 pandemic broke out and started unfolding, putting healthcare professionals and other stakeholders along with all the rest of us under immense pressures, an unexpected dread, but also opportunity was presented. Namely, how to think more carefully about questions, such as, *where do we find these alternative figurations and stories, and what is their nature?* Evidenced by prior research which demonstrated that stories, even the dominant ones, are messy and full of contradictions¹ - but can still travel far- and with the help of professional designers, we developed a workshop that would allow me to experiment with such messiness and openness and recognise that these stories do not simply pre-exist but are performative. They are coming into being through our very own methodological explorations.

For the project's workshop, I collaborated with professional designers, Dr Joe Lindley and Hayley Alter and web designer George Triantafyllakos, and used Design Fiction methods to create and pilot a new methodological research study which aimed to

- explore and reflect on the stories that figure and shape ethical AI in healthcare
- invite participants to co-create alternative ones
- test a new design methodology for reflecting on, capturing and creating new AI stories.

The study comprised of two phases. Both phases were designed to be remote and required no preparation.

PHASE 1

Phase 1 invited the participants to write, with our guidance, their own short fictional story/stories on ethical AI in healthcare. We hoped that these insights would help us understand more about (a) the messiness and complexity of the stories that shape ethical AI, and (b) the work that is needed to imagine and create alternative ones.

In more detail, Phase 1 was a collaborative story-writing exercise that was inspired by the writing game *Consequences*ⁱ and the concept of *Exquisite Corpse*ⁱⁱ. Participants worked together, but without ever directly interacting, in order to collaboratively write stories using a custom-built web tool. The custom tool deliberately encouraged participants to focus their contributions on AI and health

¹ Kerasidou, X.C., 2017. Figuring ubicomp (out). *Personal and Ubiquitous Computing*, 21(3), pp.593-605.

Law, J. (2002) *Aircraft stories: decentering the object in technoscience*, Durham, N.C.: Duke University Press.

themes, but also used random 'modifiers' to nudge participants to consider alternative configurations (e.g., that they 'have a chronic health condition', or that they are 'looking after a close friend after their surgery', or have been having a heated debate with 'the Health Minister of their country'). Having taken these modifiers into account, each new contribution started by concluding the prior contribution. Those independent yet interconnected snippets were put/stitched together with their messiness into complete stories for further use and reflection (see Story 1 as an example below).

You can still have a look at our custom-built web tool and also read (and play with) the resulting stories on the project's website:

<http://wp.lancs.ac.uk/configuring-ethical-ai-in-healthcare/workshop/phase-1/>

Story 1

The thing about AI and healthcare is...

If I am correct ... this treatment will end dyspepsia for ever! By dyspepsia, I mean heart failure. I mean the two are so easily confused aren't they? Did you know that the majority of people who have a heart attack believe it is just heartburn or indigestion? So, you see the genius of my plan? By eliminating indigestion, I can remove a key source of doubt and confusion. So that in future anybody having a heart attack will have only one potential explanation of what the sensation is. Then they should know to go immediately to the doctor and not waste time taking antacids. I'd like to see any of your artificial intelligence come up with such a neat and unexpected solution to a global health issue. And to think they sectioned me!

The world is always changing. We don't know what will come next. COVID is THE example. Global health has so many facets. I am afraid that AI will create more health problems than it solves.

We are supposed to be intelligent, but what does that even 'mean'? People judge me, like they judge themselves, not recognizing that I am ARTIFICIAAAAAAL! Then there is all this talk about Diversity and Inclusion. Hypocrites! On top of everything, they expect me to be perfect, and I am trying, I really, really do!!! ...but they keep on messing with my algorithms. I mean first, they feed me bullshit, biased, incomplete data (which by the way, they don't even admit!) and then they expect me to be superhero, understand which bits are biased and what is the bias and spit out the super morally correct outcome. Helloooooow... How about your biases and take a good look at the mirror... Haven't you heard this saying about people leaving in glass houses?

18

No wonder I have this glitch all over my algorithms. So depressing and there is no way out! Sometimes I feel I will explode and go hare wild... who cares. They have to see what they do to each other before they point the finger. If you don't want me to exclude people from consideration, then maybe you shouldn't have. And when you regret it, (but admit it first and ask I am sorry!), then let's sit down and hardwire this in my code. Am I crazy doctor? You probably do, because your ego does not let you admit it, so you scapegoat me... the artificial! Classic us and them mentality and I thought you were clever! Stressed already? I can tell from your wearable... You should manage that! and don't shoot the messenger, for telling you! Did you prefer your oblivion of the good old times? If so, yes... I am going to create more problems than solve... one, because I will warn you about all the steps to a heart attack and two, because your stupidity will not let you do anything about them before you actually get it... and blame me for telling you! Pfff Natural Intelligence... what a misnomer! No wonder I have this glitch all over my algorithms...It is autoimmune from having to put up with you! I am sure it is. What do you think?

Artificial intelligence is said to be very smart and useful, but it doesn't seem to be very helpful except for consumption. And in the era of the Fourth Industrial Revolution, we call it a doctor Watson or a surgical robot, but Watson is not a doctor. And surgical robots are not robots. We seem to overestimate artificial intelligence too much.

I don't want a machine to make decisions about me or my patients.

19

PHASE 2

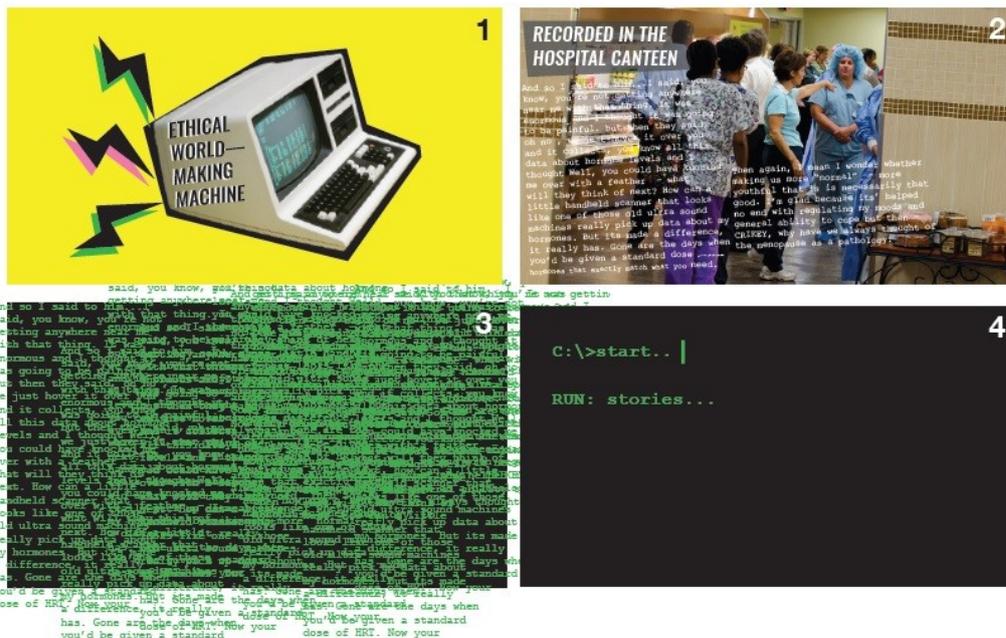
Phase 2 invited participants to a scheduled facilitated online workshop via video conferencing platform, Zoom. This was a playful pilot workshop where participants were asked to join one of 2 teams- a 'generator' or a 'discriminator'— to imaginatively approximate the dialogic workings of a Generative Adversarial Network (GAN)ⁱⁱⁱ AI programme in a healthcare setting. Through interplay

between these two teams and using as our basis the stories co-created in Phase 1, we tried to move away from questions of ‘what is an ethical AI’ by inviting participants to discuss and reflect on what is an ethical world and how we can build/imagine/ reconfigure ethical technological worlds within a healthcare context.

How did we play with the stories in the workshop?

Establishing context

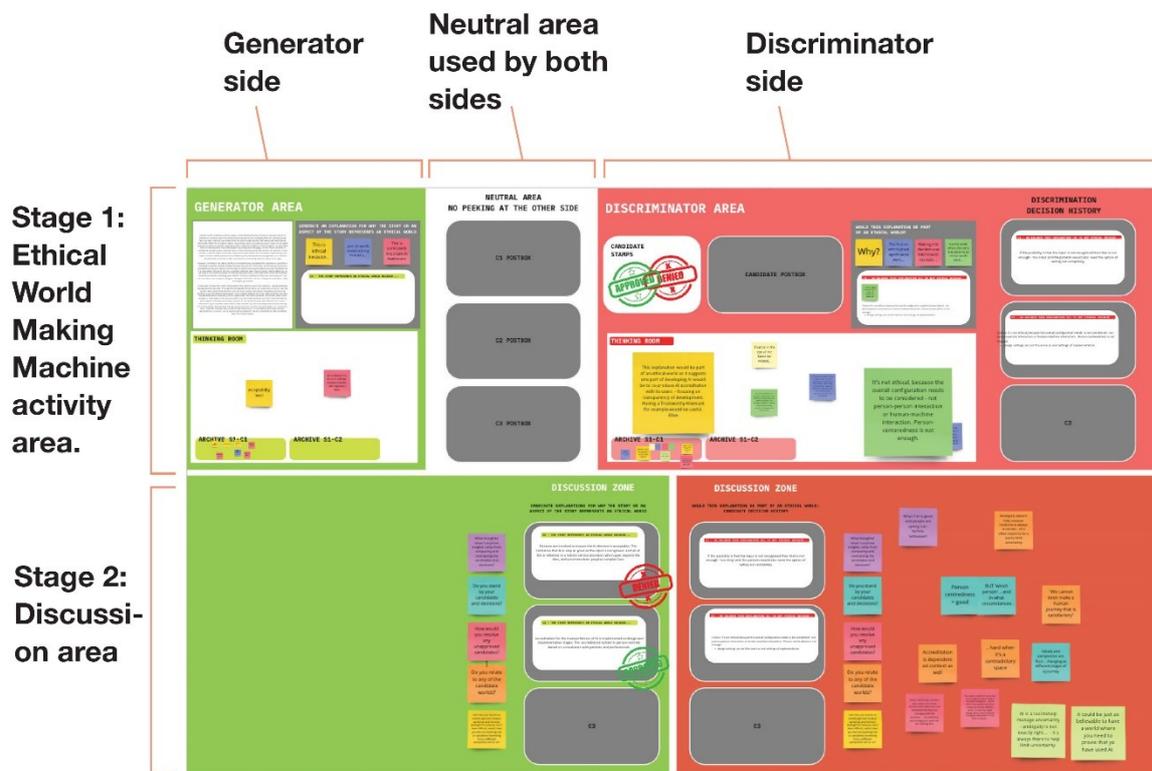
We began the workshop by establishing a fictional context for the workshop activity. We explained that we were all part of an AI GAN algorithm called the Ethical World Making Machine [fig 1] developed by the UK National Health Service (NHS), Service Futures division. The machine’s purpose was to inform design of patient care, products and services for this fictional division by identifying whether or not stories processed by the machine represent an ethical world. We said the stories we process came from audio recordings of conversations about health taken in hospital canteens, waiting rooms, surgeries and other health settings [fig 2]. Participants were asked not to question the ethics of recording private conversations without obtaining consent. Of course, in truth, the stories had already been written by participants in Phase 1. We went on to explain that usually the machine runs very efficiently but...last night, there was a power surge and the stories were scrambled [fig 3]. We tried our best to piece them back together but the result was muddled in places. From here, we ran the activity [fig 4].



Overview of Miro board for workshop

The rest of the workshop was divided into 2 stages: 1. The Ethical World Making Machine activity, and; 2. a group discussion in plenary reflecting on the activity and any emerging insights. These 2 stages were supported by real-time collaboration on Miro boards in parallel to conversation on Zoom. We used four identical Miro boards, one for each of the stories created in Phase 1. The image below is an overview of the Miro board demonstrating the area for both stages of the workshop. In stage 1, the participants were separated into 2 teams—Generators and Discriminators. Each team was facilitated in a separate breakout room and instructed to progress through each board in a series of 5-minute rounds. The green side was used by the Generators, the red by the Discriminators.

Participants were instructed not to look at the other team's side of the board throughout the activity. The white 'Neutral' area was used by both teams.



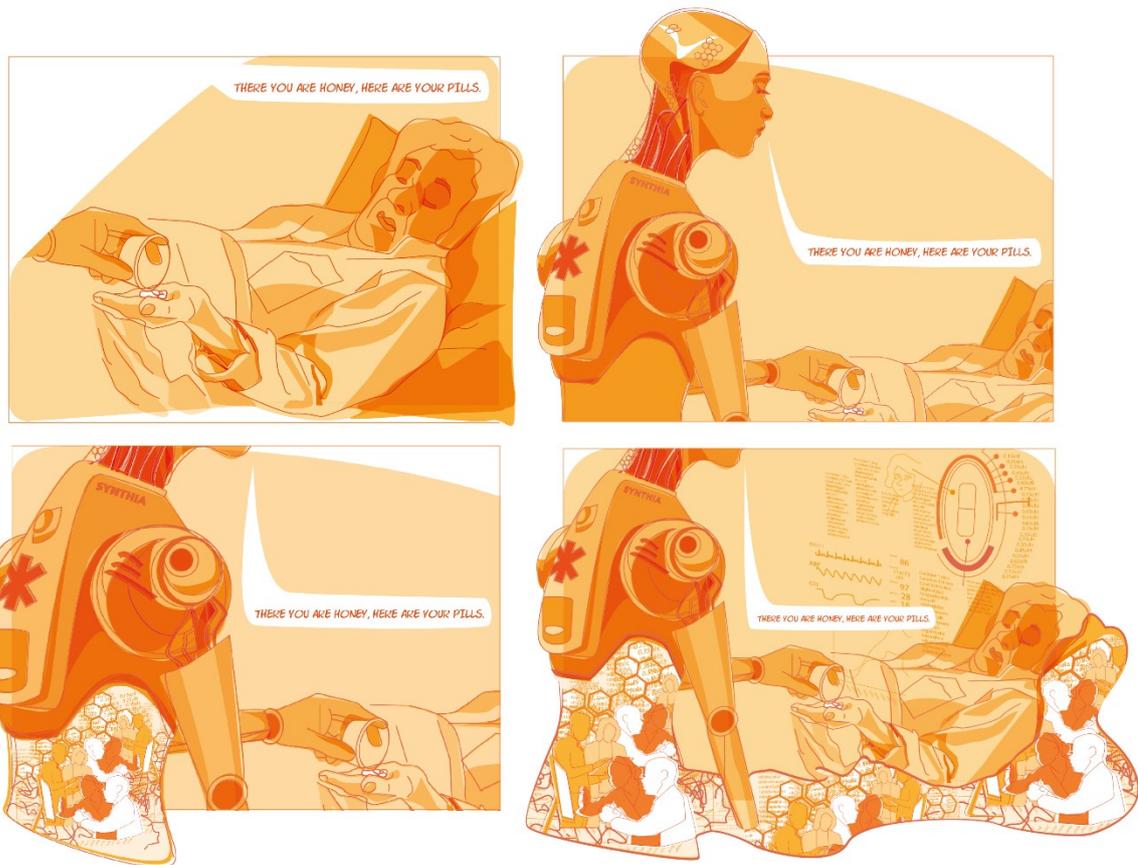
Alternative visualisations

Some of the participants' discussions and reflections from both phases have been used as inspiration for the development of a series of illustrations. These are inspired by a recent focus on the portrayals and representations of AI and some important critical efforts to challenge dominant and stereotypical AI representations and, instead, create better images and tell better stories of AI. They also keep closely at heart the basic remit of the project's methodological approach that AI stories and figurations are messy but also open-ended and performative.

The illustrations present key themes that emerged from our workshop discussions, such as the socio-technical nature not only of medical AI technologies but also of the traditional categories of patient/carer/doctor/human (featured below); the importance of 'Opting Out' and the problematics of consent; the ambivalent and productive role of the 'hype' that surrounds AI technologies in healthcare and beyond; the acknowledgment that AI technologies in healthcare are only one part of bigger infrastructures where the ethical, technological, organisational and political are not only tightly interlinked but also in tension. The aim of these countervisualities² is to act as methodological instruments that can be freely shared and disseminated, and hence to travel far and wide, as a way

² Spektor, F., Rodriguez, E., Shorey, S. and Fox, S., 2021, June. Discarded Labor: Countervisualities for Representing AI Integration in Essential Work. In Designing Interactive Systems Conference 2021 (pp. 406-419).

to reflect on and challenge the dominant AI stories and figurations in healthcare and beyond, and as a further invitation to imagine alternative ones. To view the other illustrations, please visit: <http://wp.lancs.ac.uk/configuring-ethical-ai-in-healthcare/workshop/alternative-visualisations/>



REFLECTIONS

"Experience and stories of that experience are not the same thing"

Verran (2002)³

When this project was first envisioned, the thinking regarding the workshop was to bring together healthcare practitioners and other key stakeholders - people who design, work and live with such technologies - hear their own stories and with them co-design alternative ethical AI figurations.

However, Covid-19 forced a rethink.

The disruption of the pandemic meant that getting hold of healthcare practitioners and generally people physically in the same room was now impossible. This limitation presented an opportunity as it forced a deeper methodological thinking.

³ Verran, H., 2002. A postcolonial moment in science studies: alternative firing regimes of environmental scientists and aboriginal landowners. *Social Studies of Science*, 32(5-6), pp.729-762.

If people and settings (in this case, those connected with the healthcare setting) are out of reach, where and how can we find these alternative stories and figurations? Where do these stories reside? Can I 'get hold of them', and how? Should I even try? What is the ethics and politics of such a 'knowledge/story grab'? And what is the role of technology in all this when we all sit in front of our screens trying to make a connection? Is this technologically mediated interaction changing the stories that can be told and how?

Indeed, the lessons by feminist STS tell us that these different comings-together are changing the stories that can be told and the figurations that emerge.

But they also remind us, that it was ever thus!

There is no such thing as a pure/unmediated interaction. These stories are not immutable mobiles that pre-exist in the minds of our participants that we can then access them by cleverly devising ways to get them out of their brains and into our note pads. And there is a 'violence' (can I even use such a word?) in treating them like that. A forcefulness that enables us, the researchers, to enact the 'god trick', as Haraway⁴ has so well exposed it, of seeing everything from nowhere, of receiving without participating, without exposing ourselves to 'surprise and conflict'⁵.

The AI stories and figurations are performed, worked and re-worked, shaping and shaped by our very own methods.⁶ As research, mine and others', shows, they can be messy, unexpected and inconsistent, just like the stories that emerged from our workshop, yet they can still hold their place, sometimes travel successfully, others not. With these key points in mind, and with Covid's troubling opportunity presented, the workshop of this project was reimagined and redesigned as I allowed myself to experiment. This was an effort to explore and reflect on the stories that figure and shape ethical AI in healthcare and invite participants to co-create alternative ones while staying with their inconsistencies and messiness and recognising the productivity of this mess.⁷

Some things worked well. Some of the stories were legible, evocative, funny even! Other things could be better (much better). Phase 2 was rushed, too complicated, and showed me that it takes time to take people with you, to get them to break the mould of repeating the dominant stories and find their voice. And I wonder, is this the price to pay for creativity and co-creation? Namely, the fact that 'success', 'effectiveness', 'good research results' which fit nicely in journal papers, project reports, and further funding applications might not materialise in one's timed schedule? Therein lies the conflict.

But there is also surprise. In the 'cracks', the jokes, the anecdotes, the confusion and comments under one's breath. Our challenge perhaps, I propose to you, is how to capture these without formalising them and restraining them, and hence changing their nature. How can our methods and ways of doing research adapt, or fundamentally change, to allow such fleeting but valuable knowledges to change not only the dominant stories, but also our expectations, our methods, ourselves.

⁴ Haraway, D. (1988). 'Situated knowledges: The science question in feminism and the privilege of partial perspective'. *Feminist studies* 14(3): 575-599.

⁵ Ahmed, S. (2000). *Strange encounters: embodied others in post-coloniality*. New York: Routledge.

⁶ Lury, C. and Wakeford, N. eds., 2012. *Inventive methods: The happening of the social*. Routledge.

⁷ Law, J., 2004. *After method: Mess in social science research*. Routledge.

Endnotes

ⁱ Consequences is an old parlour game. The first player will write a word, phrase or description together with the first part of a story. They then fold the paper over, hiding everything but the most recent line. The next player uses that line to start a new section of the story. Once all players have taken part, the story is read out.

ⁱⁱ Exquisite Corpse is an image-based version of Consequences. Players draw a body, section by section. When they each fold their finished section, they leave markers visible to the next player so that the body part in each section is attached to the last. The game is attributed to various artists in the Surrealist movement in Paris, France between approximately 1918 and 1925.

ⁱⁱⁱ A Generative Adversarial Network (GAN) is a framework for machine learning, designed by Ian Goodfellow in 2014, in which learning to create authentic seeming content, and in particular, images, is done by two neural networks—Generators and Discriminators— engaging in a zero-sum game. In this game, the Generator produces thousands of ‘candidates’ from random input which it sends to the Discriminator in an attempt to fool it. Throughout this interplay, both networks are continually updated.

Bio

Dr. Xaroula (Charalampia) Kerasidou is a researcher in the fields of feminist science and technology studies and media and cultural studies with a special interest in exploring how new technologies challenge us to reconceptualize the ontologies and power relations between the human and the machine, and to understand what this means for ethics, politics, and policy. She currently works on research projects that explore the ethics of AI in healthcare.