

Jun 25th, 9:00 AM

Prototyping tensions: How to talk to your colleagues about AI

Sonja Rattay
University of Copenhagen, Denmark

Mireia Yurrita
Delft University of Technology, The Netherlands

Ignacio Garnham
Aarhus University, Denmark

Jacob T. Browne
Philips Experience Design, The Netherlands

Follow this and additional works at: <https://dl.designresearchsociety.org/drs-conference-papers>



Part of the [Art and Design Commons](#)

Citation

Rattay, S., Yurrita, M., Garnham, I., and Browne, J.T. (2022) Prototyping tensions: How to talk to your colleagues about AI, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao*, 27 June - 3 July, Bilbao, Spain. <https://doi.org/10.21606/drs.2022.909>

This Research Paper is brought to you for free and open access by the DRS Conference Proceedings at DRS Digital Library. It has been accepted for inclusion in DRS Biennial Conference Series by an authorized administrator of DRS Digital Library. For more information, please contact dl@designresearchsociety.org.

Prototyping tensions: How to talk to your colleagues about AI

Sonja Rattay^a, Mireia Yurrita^b, Ignacio Garnham^c, Jacob T. Browne^{b,d,*}

^aUniversity of Copenhagen, Denmark

^bDelft University of Technology, The Netherlands

^cAarhus University, Denmark

^dPhilips Experience Design, The Netherlands

*corresponding author e-mail: j.t.browne@tudelft.nl

doi.org/10.21606/drs.2022.909

Abstract: Current imaginaries of AI are pulled in opposing directions. On the utopian side, AI is being projected as the better half of humankind, evening out the fallibility of human bias and our inability to “know it all”. On the dystopian side, AI will make humans either obsolete or turn humans into overly optimized cogs in a machine. The reality will most likely lie somewhere in between, in an uncomfortable grey zone of compromises, tradeoffs and negotiations of values and desired futures. In this conversation, we want to explore the tensions between contrasting imaginaries, which we present as four different positions towards how the future development of AI should be approached.

Keywords: artificial intelligence; prototyping; conversations

1. Context of the conversation

From developer-centric to critical studies perspectives, many different standpoints constellate the AI research space. A central challenge amongst AI research lies in rendering effective collaboration and conversation between these standpoints (Morley 2020). While the consequences of many AI behemoth companies hiring-and firing-AI researchers over these differences in standpoints is becoming increasingly visible (Schiffer 2021), there is a lack of attention in examining how academic research is being shaped through the diversity of research standpoints aiming to shape the AI conversation around their own priorities (Jobin et al., 2019)

For instance, different standpoints recognize and place focus on different unintended effects of AI. Pragmatic, constructive approaches focus on debiasing, while destructive, speculative approaches focus on how to challenge power structures and in some cases abolish AI altogether (Miceli 2022, Mitchell 2019, McQuillan 2019, Mohamed 2020). While the pragmatic v. speculative and constructive v. destructive axes are not comprehensive of all views, they serve as an exposition of the tensions within AI research.



This conversation aims to bring together these different standpoints in reconciling positions and finding common ground. Each convener will play as a representative of each of these standpoints, rendered as follows:

- **Speculative/Destructive.** AI is inherently harmful and should be abolished, together with the power structures which it upholds. To abolish something harmful means to rid the world of the conditions that allowed the harm to take place.
- **Speculative/Constructive.** Machine learning has enabled some truly innovative approaches to better address the complex demands of our current society. We need to think of different futures and idealistic visions from here and harness the opportunities it presents.
- **Pragmatic/Destructive.** To get rid of machine learning just because it can potentially be dangerous is not the solution. Instead, we need to be sensitive towards the risks this new technology poses and learn from the harms that have been done. Let's have more regulations and limit what we can do.
- **Pragmatic / Constructive.** We should focus on dealing with problematic aspects of the development and implementation of algorithmic systems and come up with strategies for dealing with those. Let's experiment with what we have and take small steps in the right direction.

DRS Participants will get a chance to voice their own standpoint, challenge other positions, and find common ground with other participants in a collaborative fashion. We hope that through this conversation we can engage a broader audience in how we, as design researchers, can bridge the conversational gaps in different positions on AI.

2. Organizing research questions

How can the structured use of physical space and public argumentation help academic researchers in negotiating standpoints and finding a productive common ground?

Are the destructive/constructive and speculative/pragmatic axes conducive to meaningful dialog that fosters collaboration?

3. What discussions took place

3.1 Structure of the session

The session is set up as a conversation between four positions- destructive, constructive, speculative and pragmatic-each represented by one of the contributors. We will utilize physical space by positioning each opinion in one corner of the room and creating a middle point where the axes intersect for a final roundtable discussion. The objective of the structure is to encourage participants to express their positionality through movement and proximity, exercising a modified version of the law of two feet (Owen 2008). The conversation is structured as follows:

- It begins with a short pitch from each contributor with the aim of providing scaffolding for participants to build on their positions. After this introduction, participants are asked to choose one of the positions by standing in the corresponding corner of the room.
- The resulting groups are then invited to discuss amongst themselves why each participant chose their position. During this stage, participants are asked to pay close attention to how others in their group describe their positionality, and to agree on a description of the position that represents the group.
- To continue, we initiate a round-based conversation where each group shares their position, after which participants across the axes are invited to change positions if persuaded to. As participants move between positions, the groups can adjust their arguments to adopt the new positionalities and to adapt to changes in how the other groups define their positions.
- When the movement of participants across axes has ceased, participants are then invited to take steps towards the middle of the room if they become at least partly convinced or agree to a point made by another group. At this stage people can still decide to shift position but the aim of this spiral-like movement is to find common ground within the different opinions and to move all four groups towards the middle.
- Lastly, participants engage in a round-based conversation to reflect on the process and share which conversations or actions led them to maintain or change positions.

With this set up, we hope to explore dynamics of compromise, and to prototype a conversation in which participants can explore ways of communicating and negotiating their positionality with other researchers. Ultimately, we want to explore how the validity and legitimacy of different concerns and approaches are being discussed, framed, and negotiated.

3.2 Session content

The session was performed with 11 participants with various backgrounds, mostly from design research. We started the session with the pitches for each position and then asked the participants to quickly note down a few notes, questions, and points for each pitch on post-its. Following the brainstorming, the participants joined the quadrant of the position they felt most comfortable with. For the warmup, the groups had a short discussion of the points and notes they came up with during the brainstorm among themselves. We then started mixing the groups – one or two participants from one corner started discussing their standpoints with one or two participants from the opposite corner and argue for why they assumed this position to be the most productive regarding engaging with AI.

After about 5 minutes, the group came together again and shared the discussed points. Participants who felt convinced by the arguments of a participant from another corner changed

their position within the quadrant accordingly. We repeated this exercise a second time with new pairings. Participants who had changed their quadrants one or even two times reflected together with the group upon which arguments contributed to their shift in opinion and how it affected their standpoint towards AI.

For the third round of paired argumentation, we manipulated the dynamic by asking each participant to take on the standpoint opposing to their current position and argue for that approach. For some participants, this led to arguing again from the position they started off from, while others had to now argue from a position they had argued against for the last two rounds. We wrapped up this dynamic part of the conversation with a short temperature reading of the group regarding how productive they considered each position to be.

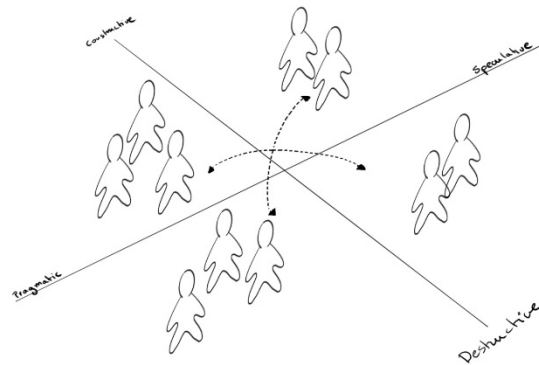


Figure 1. Layout of the space the participants traversed to negotiate their positions.

4. Outcomes, insights, and reflections

The reactions to the conversation were generally positive and the structure and organization led to a rich and engaged discussion across all parts of the conversation. We collected three main points of reflection about both how the structure facilitated a fruitful conversation and about which positions came to light based on it.

4.1 The format organizes different perspectives into productive tensions

While most participants initially leaned towards positioning themselves “somewhere in the middle”, the clear separation between the four positions pushed them towards thinking about which tendencies they could take to the most radical direction and still defend comfortably. While the positions are not exclusive and can co-exist, having to commit to one position exclusively for the sake of discussion creates tensions that in turn lead to more precise openings for reflection. Participants also observed that this way of structuring a conversation might work as productively for other technologies, as it forces a commitment and a clear articulation of why certain aspects of a position are deemed valuable and important, and others aren't.

4.2 *The arising tensions make participants reflect upon their positionality*

Participants reflected upon which aspects of their training, current roles and other positional factors contributed to their choosing of position. Some participants recognized patterns of arguments that they share with colleagues and were either able to take a different position than what they were used to or feel they must take in certain conditions of their research or work environment. For example, some participants noted that an engineering background might lend itself more towards a pragmatic outlook, while a humanist education might fit better with a destructive outlook. Similarly, participants noted that as a researcher, one should take a constructive outlook, as that would guide productive contributions to the field.

4.3 *Personality and general outlook on humanity might influence positionality more than the technical outlook*

One participant who changed their position after the first round noted that the reason to do so was connected more to the general way they want to see live, rather than an outlook on the technology, i.e., in this case they changed from speculative/destructive to speculative/constructive and argued that they want to be hopeful and have a positive impression on humanity rather than focus on the negatives. Similarly, other participants argued that they stuck to the pragmatic sides of the positions, simply because that is where they felt they could be active and contribute something concrete, rather than negotiate abstractions. In both cases, personal values contributed to the positionality of the participants and their take on productive positions more so than technological developments.

We want to conclude with the general observation that positionalities upon questions such as the ones we asked during this conversation are manifold and more complex than simply depending upon what is technically feasible or presented as “true”. Personal values, politics and power hierarchies strongly influence which perspectives might feel fitting at any given time for any participant. Positions such as the ones presented are not as fixed as we set them in this conversation, but fluid and shifting in relation to each other. However, acknowledging and clearly articulating the concrete directions is productive in that it allows to guide negotiations of value tensions.

Acknowledgements: We would like to thank all our participants for an engaging, fun session in Bilbao. This work is part of the DCODE project. The project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955990.

5. References

- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- McQuillan, D. (2019). Towards an anti-fascist AI. (Accessed February 10, 2022). <https://www.opendemocracy.net/en/digitaliberties/towards-anti-fascist-ai/> (Accessed February 10, 2022).
- Miceli, M., Posada, J., and Yang, T. (2022). Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power?. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 34 (January 2022), 14 pages. <https://doi.org/10.1145/3492853>

- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, Spitzer, E., Raji, I.D., and Gebru, T. 2019. Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229. DOI:<https://doi.org/10.1145/3287560.3287596>
- Mohamed, S., Png, M.T., and Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*. 33. 10.1007/s13347-020-00405-8.
- Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics* 26, 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Schiffer, Z. (2021). Timnit Gebru was Fired From Google - Then the Harassment Arrived. <https://www.theverge.com/22309962/timnit-gebru-google-harassment-campaign-jeff-dean> (Accessed February 10, 2022).
- Owen, H., (2008). A brief user's guide to open space technology. (Accessed February 14, 2022) <https://openspaceworld.org/wp2/hho/papers/brief-users-guide-open-space-technology/>

About the authors:

Sonja Rattay is a PhD fellow at University Copenhagen. Her research explores how we can engage ethics-in-action in new design practices for data-driven technology.

Mireia Yurrita is a PhD fellow at TU Delft. Her research focuses on developing design and engineering methodologies for shaping and deploying principled algorithmic systems.

Ignacio Garnham is a PhD fellow at Aarhus University. His research examines the capacity of algorithms to change patterns of social and cultural change from a design anthropology perspective.

Jacob T. Browne is a PhD fellow at Philips and TU Delft. His research is concerned with designing for appropriate trust in clinical AI.