



SEÇÃO: ENTREVISTA

Disruptive technological progress and value changes: An Interview with Jeroen Hopster on the most appropriate approach, regulation and governance

Sobre progresso tecnológico disruptivo e mudanças de valor: uma entrevista com Jeroen Hopster sobre a abordagem mais apropriada, regulação e governança

Sobre el progreso tecnológico disruptivo y los cambios de valor: una entrevista con Jeroen Hopster sobre el mejor enfoque, la regulación y la gobernanza

Murilo Mariano Vilaça¹

orcid.org/0000-0001-9720-5552
murilo.vilaca@fiocruz.br

Recebido em: 25 ago. 2024.

Aprovado em: 14 ago. 2025.

Publicado em: 13 nov. 2025.

Abstract: Accelerated technological progress has become a central topic of philosophical reflection today. One of the key points of philosophical debate is the disruptive nature of technological progress. Although the expressions "disruptive technologies" and "technological disruption" have become quite frequent in publications, it is not always clear what this means. In other words, there are significant conceptual and therefore explanatory gaps as to what differentiates an incremental new technology (or technological innovation) from a disruptive one, as to what the domains of this so-called disruption might be, and as to the future implications of these disruptions, considering a decisive factor: the value change. In this interview, I aim to fill these gaps by turning to one of the main exponents of the debate on socially disruptive technologies, Professor Jeroen Hopster (Utrecht University). Going through his publications in some of the most internationally renowned journals, we offer readers a kind of guide to the topic, clarifying it conceptually, explaining its implications and indicating avenues of research around it.

Keywords: technological progress; social disruption; value change; Jeroen Hopster.

Resumo: O acelerado avanço tecnológico se tornou um tópico central da reflexão filosófica na atualidade. Um dos pontos-chave do debate filosófico é a natureza disruptiva do progresso tecnológico. Embora as expressões "tecnologias disruptivas" e "disrupção tecnológica" tenham se tornado bem frequentes em publicações, nem sempre está devidamente claro o que isso significa. Ou seja, há significativas lacunas conceituais e, portanto, explicativas quanto ao que diferencia uma nova tecnologia (ou uma inovação tecnológica) incremental e uma disruptiva, em relação a quais seriam os domínios dessa propalada disrupção, bem como quanto às implicações futuras dessas disrupções, considerando um fator decisivo: a mudança de valores. Nesta entrevista, viso a preencher essas lacunas, recorrendo a um dos principais expoentes do debate em torno das tecnologias socialmente disruptivas, o Professor Jeroen Hopster (Utrecht University). Passeando por suas publicações em alguns dos periódicos mais renomados internacionalmente, oferecemos aos leitores(as) uma espécie de guia sobre o tópico, esclarecendo-o conceitualmente, explicando suas implicações e indicando caminhos de pesquisa em torno dele.

Palavras-chave: progresso tecnológico; disrupção social; mudança de valor; Jeroen Hopster.



Artigo está licenciado sob forma de uma licença
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Resumen: El progreso tecnológico acelerado se ha convertido en un tema central de la reflexión filosófica actual. Uno de los puntos clave del debate filosófico es la naturaleza disruptiva del progreso tecnológico. Aunque las expresiones "tecnologías disruptivas" y "disrupción tecnológica" se han hecho bastante fre-

¹ Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro (RJ), Brazil.

cuentas en las publicaciones, no siempre está claro qué significa. En otras palabras, existen importantes lagunas conceptuales y, por tanto, explicativas, en cuanto a lo que diferencia una nueva tecnología incremental (o innovación tecnológica) de una disruptiva, en cuanto a cuáles pueden ser los ámbitos de esta llamada disrupción, y en cuanto a las implicaciones futuras de estas disrupciones, considerando un factor decisivo: el cambio de valor. En esta entrevista, pretendo colmar estas lagunas recurriendo a uno de los principales exponentes del debate sobre las tecnologías socialmente disruptivas, el Profesor Jeroen Hopster (Universidad de Utrecht). Recorriendo sus publicaciones en algunas de las revistas de mayor renombre internacional, ofreceremos a los(as) lectores(as) una especie de guía sobre el tema, aclarándolo conceptualmente, explicando sus implicaciones e indicando vías de investigación en torno a él.

Palabras clave: progreso tecnológico; disrupción social; cambio de valores; Jeroen Hopster.

Introduction

In this seventh interview in the series conducted by the Transhumanism and Human Bioenhancement Research Group (GIFT-H+/CNPq), we are pleased to interview Jeroen Hopster on technological progress, how to define the socially disruptive nature of technologies, the relationship between this and changes in value, emphasizing his proposal for a triadic approach and reflecting on what would be the most appropriate way of regulating and governing these correlated processes.

Jeroen Hopster is Assistant Professor of Ethics at Utrecht University, the Netherlands. As a researcher in the consortium "[Ethics of Socially Disruptive Technologies](#)" (ESDiT, 2020-2029), he investigates the nature of social disruption, its ethical implications, and the role of technology in processes of moral and conceptual change. Additionally, he leads a research project on a just climate transition, "[Lost in Transition](#)" (2024-2028).

Hopster obtained his PhD in Philosophy at Utrecht University (2019), with the dissertation *Moral Objectivity: Origins and Foundations*. Subsequently, he has been affiliated to the University of Graz and the University of Twente. He has had visiting fellowships at Harvard University, the Jagiellonian University (Kraków), UniSR (Milan). He is a member of the Utrecht Young Academy.

Interview

Murilo Vilaça: The first question has to do with accelerated technological progress, which is a central theme of practical philosophy nowadays. Although philosophy's interest in technology is not exactly new, the speed of technological advancement in recent years has challenged philosophers' ability to think about technology. One of the biggest challenges, according to Gilles Deleuze, was the main task of philosophy, namely, to create concepts. In this sense, how do you assess the fulfillment of this task of creating concepts by philosophers who are mainly interested in new technologies? Have we been competent in creating concepts suited to technological innovations, or, on the contrary, have we created a flood of concepts without much relevance or utility? In other words, has our conceptual engineering been efficient? And here, without trying to induce your answer, I'd just like to draw your attention to an example that I've been thinking a lot lately, which is the concept of artificial intelligence, the idea that machines can be intelligent. Professor Luciano Floridi has called artificial intelligence "agents without intelligence", and it seems that this idea of intelligence may be being used more as a kind of media or publicity stunt than truly expressing what it means. I'd like to hear from you. Thank you.

Thank you, Murilo. That is an interesting question with many layers to it, so let me pick up on a few threads. First of all, I agree that technological acceleration has been a challenge for philosophy, but it's also been a very fruitful challenge. For instance, your research group, as well as my work and that of the [ESDiT consortium](#), takes inspiration from the various philosophical concerns this raises. So, in a way, I think it's also been a fruitful moment for philosophy.

You mentioned Deleuze and his statement that "the main task of philosophy is to create concepts." I don't fully agree with that statement, even though there's something to it. But there is also a sense in which it's rather easy to create

concepts. You can just combine existing words, or you can think of new ones and associate them with specific meanings. Novel concepts can't solve all philosophical problems; what we want is *good* concepts, that do their job, and do it well. The bigger challenge, in a way, is to rethink the concepts we already have, and to investigate how they might have to be modified or adjusted, and how improved concepts or conceptions can achieve theoretical and societal uptake. And sometimes, in the face of technological pressure, the challenge is actually to preserve concepts, rather than coming up with new ones. So the overarching task is not, primarily, to create concepts, but rather to improve conceptual schemes. That's one of my key interests at the moment. And I think that one of the reasons why this topic is becoming increasingly important has to do with technological disruption. Because of technological disruptions we have to rethink concepts such as autonomy, agency, intelligence, privacy. Explicit reflection on this matter, on the disruption of concepts and how that can be tied to conceptual engineering, that's a fairly new undertaking.

You also mentioned artificial intelligence. I think that's a very interesting concept, in two ways. On the one hand, it's interesting because of the very notions of "intelligence", as well as "artificial". Some like Floridi will say: "well, AI is not all that intelligent". And some authors say that AI is neither all that artificial. So the concept of AI can be challenged, from both of these sides. And importantly, there might be normative implications to this, for instance to associating AI-systems with intelligence. With AI becoming increasingly dominant in society, there are a sociotechnical systems emerging, where intelligence is sort of praised as the highest value. Many developers of AI present AI as a general problem-solving agent, and suggest that 'intelligence' is all we need to solve major issues, from poverty to climate change. I think that rests on a misunderstanding and distracts from more plausible proposals. And it's a distraction that is fueled in part by the very notion of intelligence, engrained in this concept, which is propagated by marketeers of various

techniques... thereby simultaneously reducing a lot of values to one epistemic value, namely, intelligence, whereas actually many epistemic and moral values are needed to solve the problems at hand. Also, the kinds of techniques that are involved in AI are hugely different. To use 'AI' as a common designator for these various different techniques already leads to misunderstandings. So I'm sympathetic to the view that the emerging socio-technical imaginary in which 'intelligence' as an overarching concept plays such a huge role, comes with a distinctive risks.

There's another level at which it's interesting to think through the dynamics of technology and concepts, which does not only look at a single concept like intelligence, but at the broader conceptual scheme. I mentioned autonomy and agency as examples. Now think of a recent AI system like GPT-4, and what it's capable of. Arguably we have to rethink what it means to be an agent in light of the emergence of Large Language Models like GPT. Is GPT itself an agent, or not? So that's a potential disruption to the concept of agency, one might say. But actually, it's more than just the disruption to a single concept. It's often a cluster of important concepts that is being disrupted by new technologies. For instance, one might additionally consider: "is GPT able to manipulate us?" That requires having an understanding of what 'manipulation' is all about. Traditionally it's been thought that the ability to manipulate requires ones having intentions. So, if GPT is able to manipulate, does it have 'intentions' as well? And if GPT, with or without intention, is an agent, [is it] going to be held responsible for its actions? So whether we conceptualize GPT as an agent or not, is entangled with various further questions that pertain to a cluster of concepts. And we can ask similar questions about GPT being an author. Can we ascribe authorship to GPT? Can GPT be creative? I think these are all hugely interesting questions. They're open questions and they're really about rethinking what we mean by authorship, creativity, agency, etc. So that's my reason for saying that it's not just the case that we have to think of new concepts, but

in a way, we also have to think about our existing concepts, whether we have to reconceptualize them or not, whether they fulfil their function... I mean, there might be indeed a need for new concepts. There might be certain conceptual gaps that have to be overcome. There might be certain conceptual misalignments that have to be corrected for, and there might be a need for new ways of conceptualizing things altogether? So, there are different kinds of what I would call "conceptual disruption" going on.

Murilo Vilaça: Right, and I think the next one is a follow-up question, which is intended to hear more from you, because I would like to know how you see the explosion in the use of the concept "disruptive technologies" from the second half of the 1990s onwards. I did a search that was not very deep on Google Scholar and I realized that from the second half of the 1990s there was an exponential increase in the use of this expression. My question is whether you think that this vast presence of the term in academic literature is proportional to its conceptual clarity, its empirical relevance and its normative usefulness.

The short answer is: probably not. Though that might actually be true for many philosophical concepts that we've been thinking about for, you know, for ages. Let's say... "What is free will?" If this concept had been immensely clear, then presumably we wouldn't have been thinking about it for so long. So, in a way that's sort of a trademark of interesting philosophical concepts... that they're not all that clear, but relevant nonetheless, and in need of clarification.

But yes, I agree with your diagnosis. As of the 1990s, because of the work of Clayton Christensen, the notion of 'disruptive technologies' has become very prominent in the public sphere². Christensen was an influential author and acade-

mic, but not a philosopher. He wrote in a context of economics and business. The concept has subsequently been adopted by many innovators who actually want to be disruptors, who think that 'disruptive innovation'³ is an aspiration to be aimed for. But Christensen used "disruption" in quite a narrow sense. Basically, the idea is that new technologies can disrupt markets, making older companies and their products obsolete. That's what market disruption is all about. But in common parlance, in natural language, disruption is often used in a much broader sense, which is not limited to disruptive innovation as Christensen meant it. And the concept has caught on. Many people are using it. New discourses have emerged, for instance on digital disruption, which are only loosely related to Christensen's work. So, apparently, it's a useful concept. Apparently, it captures something of the imagination.

I think it's useful to make the concept a bit broader than Christensen's usage. His conception seems unnecessarily restrictive. I think that the philosophically interesting aspect of disruption is only about what happens in economics and business – though that's part of the story. But I'm also interested in how technology can disrupt norms, values, human capabilities, human society, even what it means to be human... There's a much deeper sense of disruption involved here. And I think that that's an angle that has long been underappreciated in scholarship, although at the moment there's a lot of attention for it, in part due to the ESDiT programme. Technologies have a capacity to disrupt in a much more forgoing sense: information and communication technologies disrupt democratic practices, neuro-technologies can alter our human constitution, geoengineering technologies can alter the global carbon cycle... And new technologies can also disrupt scientific theories, moral norms, core values, concepts, conceptual schemes... So my key interest is in this deeper sense of disruption,

² See, for example, BOWER, Joseph L.; CHRISTENSEN, Clayton M. Disruptive technologies: Catching the wave. *Harvard Business Review*, Cambridge, Jan./Feb. 1995. | CHRISTENSEN, Clayton M.; OVERDORF, Michael. Meeting the challenge of disruptive change. *Harvard Business Review*, Cambridge, 2000.

³ See, for example, CHRISTENSEN, Clayton M.; RAYNOR, Michael E.; MCDONALD, Rory. What is disruptive innovation? *Harvard Business Review*, Cambridge, 2015. Available at: <https://hbr.org/2015/12/what-is-disruptive-innovation>. Accessed: 24 Sept. 2025.

not merely in economics and business, but in the many foundational aspects of human beings and institutions that are being challenged, and that might have to be rethought. And I think that there is a lot of emerging interest in these kinds of 'deep disruptions', because they are so fundamental, which makes it philosophically very interesting to think them through and to consider how we should respond to them.

Murilo Vilaça: This is a very interesting point, because the idea of disruption in the economic context is an idea that can be understood as positive, as what technological innovation and innovation in methods, such as new products, can do, or actually contribute to the economic development of those who own these innovations. But it seems that, on the contrary, the idea of disruption in the debate in which we find ourselves is almost an idea that something bad is happening, a change that could trigger, at the limit, even an apocalypse, right? So it's about "how to restrict disruption". For the economy, it's the other way around, it is about "how to promote disruption". So, it's quite interesting to see how the transposition of a concept from one context, from one academic area to another can completely change its meaning. This is really interesting.

Jeroen Hopster: I much agree. My only addition would be that, as I think of disruption, it's not necessarily bad either. As said, in market and business, its connotation is often positive. In philosophy and ethics, its connotation might be somewhat biased towards the negative, because we tend to foreground the risks and harms. We think about what can go wrong, but in principle disruptive change might also be for the better. For instance, we might be stuck with certain values, certain norms that can actually benefit from being disrupted. Perhaps they entrench practices that we don't want to entrench, on further thought. So, disruption can also be a good thing. The same holds for social disruption. When it comes to the implications of socially disruptive technologies

such as AI, there are indeed many things that can go wrong, so we should be on guard. But the implications may also be quite ambiguous and unpredictable. And it's not either or: there might be a positive as well as a negative side to it. So what is needed is, first of all, to really think through the issues, the values, the norms that are at stake, and to safeguard ourselves for the risks and for the possible harms that might ensue. But we should also consider how we can put a positive spin on disruptions. I think that's very much dependent on our capacity as a society to appropriately respond: to devise the right kinds of institutions, regulations, as well as the right kinds of normative categories, concepts and practices. Whether technological innovation is for the better, partly depends on our capacity to foster an ecosystem of responsible innovation.

Murilo Vilaça: Just a brief comment about that. Even in the economic field, we can challenge this idea that disruption is always positive, because the creation of new products or the creation of new processes can generate or even further aggravate existing inequalities, so it's interesting how this idea of disruption predominates as something positive, and in another scenario as something eminently negative.

The next question I'd like to ask is about the impact of these technological changes. One of the key points of the contemporary philosophical debate has to do with technological change, with its diverse, and to a certain extent, unpredictable impacts. The future is uncertain, but we know at least one thing: it will be even more profoundly impacted by the emerging technologies. In your opinion, does the vast literature on technological impacts reflect a detailed understanding of them, distinguishing them by subtopics and specific domains? How can a generic approach to impacts hinder progress in the debate?

Jeroen Hopster: Let's start with the thesis that even though the future's uncertain, we know at least one thing: that it'll be more profoundly impacted by emerging technologies. At the moment, I would probably agree, though in the farther future, I'm less sure. It also depends on what we think of as an impact. Like I said in my answer to the previous question, I think that the extent to which a disruption is highly unsettling and disorienting, very much depends on the kinds of conceptual, normative, institutional infrastructures we have in place. So, we can also build very resilient societies, in principle, that can even cope with a lot of technological changes. Therefore, I think it's not a given that levels of disruption are only going to increase as time goes on. But I would agree that at the moment that is certainly what's happening and for the short-term foreseeable future, I don't see that changing.

Then, thinking about the kinds of impacts in rather generic terms, does this impede progress? Should we think of them in more specific domains? Well, I agree that we still have to make a lot of progress in advancing our conceptual schemes, in order to really think these issues through, and also to make them tailored to specific domains and contexts. In general, that's a useful thing. That said, I also think that there's something useful about discussing disruption at a general level. In one of my publications – "What are socially disruptive technologies" – I have come up with criteria of what makes a technology more or less disruptive. That's, of course, at a fairly general level of disruption. Indeed, even 'technology' itself can already refer to many different things; it is a very diffuse category. Being 'disruptive' can also mean different things, but we can make this more precise. In the paper I identify seven properties that can make an impact less or more disruptive: the depth of the impact; the extent to which it affects deeply held values, beliefs, institutions, basic human capacities; the breadth of the impact; the range and variety of domains it affects; its normative significance, let's say, the extent to which it touches upon core determinants

of the quality of society, of the good life; the uncertainty that comes along with it, so to what extent its implications yield normative as well as epistemic uncertainty; the pace of the change, so if the change is very rapid, then the degree of societal preparation tends to be less, so that can also make a technology more disruptive; the degree of reversibility, so to what extent is an impact likely to be long lasting, or can it easily be made undone. And adding to this, there are also the differential effects: to what extent do technologies dissimilarly impact specific groups in societies? To what extent do they increase technological divides in between societies? This is an approach to disruption and disruptive technologies at a general level, and I think it can be useful for some purposes – e.g. in making general legislation or normative guidelines. Or to prioritize which technologies really deserve our ethical attention, because they have such a high degree of disruptive implications. But surely, the specifics of that ethical attention, that's a second question, and that should be discussed at a much more fine-grained level. The general framework helps to prioritize or to triage certain technologies rather than others, and to foster societal and ethical debates on some issues rather than others. But subsequently, we should look at the specifics of it. And appropriate discourses should develop around that as well, which is very much a task for the ethics of technology.

Murilo Vilaça: I fully agree. But I also have in mind the reports that were published by the SIENNA project, and they were very careful to establish the different impacts according to specific domains, so the impact that a technology might have, for example, on the workplace might be different in healthcare and vice versa. So this is something that I think scholars should be concerned about in order not to work with the notion of impact in a non-specific way, because impacts in the plural can be very different depending on the domain in which they are applied, and depending on the society, as well. One thing is for us to think about the impact of

artificial intelligence in the Netherlands or in Denmark, and it is another thing to think about it in the Brazilian context. I think your framework allows this clarification, so that we can identify more clearly what kind of impact we're talking about and what kind of concerns we want or need to have. Very interesting point.

Jeroen Hopster: I think it's a useful corrective that you're making. My framework does emphasize that there's differential disruption. Disruption is not the same for every human being, and also not the same for different groups, and also not the same for different societies. But yes, I grant that it's a framework that operates at a high level of generality, which is useful for certain aims, but not for others. When it comes to the ethical issues arising from AI in the workplace versus AI in an environmental context, say, these are obviously very different. You see that reflected, for instance, in the AI act that has just been adopted by the European Parliament. The act doesn't state something like "AI should be prohibited" or "AI is fine and can be used in all contexts". Instead, the act is specific about the risks that are involved and delineates specific contexts where AI can and cannot be used. For instance, there cannot be social scoring systems in government contexts, or there cannot be emotional registration of students, or in the workplace. So, I think that latches onto your point: we should be specific about context and domain.

Murilo Vilaça: **One of the subtopics of the debate that seems most interesting to me refers to the relationship between technological changes, disruptive technologies and changes in value. In this sense, I'd like to ask a question whose complexity perhaps exceeds the limits of this interview, but about which I'd like to hear from you: what is the role of technological change in the moral revolutions that have taken place and will take place in the future? Again, I know that this is a very broad question, and it must include, for example, the definition of what a moral revolution is, which is different from other**

types of moral change, but I would like to hear from you because this is very important to me.

Jeroen Hopster: I agree. I think technologies can play different causal roles in moral revolutions. They can be drivers of change, they can be triggers of change, they can be contributors. It's of course difficult to study future moral revolutions. We might arrive at insights of what these can plausibly involve, but if you want to have a handle on the causal role that technology is playing, then you are best off studying revolutions that have happened in the past. In historical case studies you see that technology often does play a role, but it's rarely technology alone. It's always technology being embedded in a certain social context, being infused with certain ideological norms and expectations. But in that context, the role of technology can nonetheless be important. I like to approach this topic in terms of a concept drawn from biology: niche construction. I think that humans are niche constructors. We construct many different kinds of niches, including moral niches. And these niches can be sustained by certain technologies, but they can also be disrupted by technologies. So, we really have to think about the role that technologies play in our moral niches or social technical systems, so to say. And then, if you look at the moral revolutions of the past, you can see certain regularities that might also play a role in those of the future.

At the most general level, what technology tends to do is to create affordances, and generate constraints. That's a useful starting point when categorizing things. But there are more specific things that technology does as well. It can make things visible or invisible. And that includes making certain stakeholders visible or invisible. So, if I had to anticipate one of the major revolutions that might still occur, one might think about the treatment of animals. Possibly, the way that animals are treated nowadays will be looked upon with disdain one century from now. What could create such a mindset that we really start to see things differently? A lot of it has to do in part with

making things visible. If you can make the suffering of animals in the bioindustry more visible, or more audible, or more perceptually salient in any kind of way, that will change people's mindsets. Technologies often have the power to do these kinds of things. One concrete way in which that might go about would be with VR installations, which allow you to really have an experience of what it would be like to be a factory farmed animal, and to really empathize with them. I think that can really be perspective changing.

Many technologies have this potential to change perspectives. But technologies can also be instruments of power. They can be instruments of resistance for certain groups. They can amplify one's voice. They can make groups be heard that were previously not heard, so they can facilitate opposition. They can facilitate the spread of ideas. They can also serve to moralize or demoralize certain issues, and so the list goes on. I think there are actually many different kinds of roles that technology can play and ways in which it can mediate our morality. I find this hugely interesting, in part because it is at odds with the old-school philosophical idea of thinking that figuring out what morality is all about, amounts to discovering timeless foundational principles. But that perspective overlooks how technology and morality actually interact, and how technology might drive certain moral changes, and also how morality might drive certain technological developments. It has become apparent that moral values are not as fixed and immutable as people had thought before. Things have changed in the past, and things will definitely change as well going forward. Even though we can't predict the precise changes themselves, we can anticipate the way in which technology can facilitate such changes, for instance by making certain things visible or perceptually more salient. So yeah, I think that's the level at which it's useful to study the kinds of roles that technology can have vis-a-vis morality.

Murilo Vilaça: This question began to be asked a long time ago when I read your very interesting article, which was a dialogue with John Danaher. He published an article on axiological futurism, and you proposed a way to outline the field because it was too open, to identify the possibilities and the risks in a realistic way. And then you later published an article together in which you presented some of these solutions that you developed together. And another point about what you answered that also seems very interesting to me is what some authors have defended in terms of the use of either virtual reality or augmented reality as forms of moral enhancement. Jon Rueda, whom you know very well, published an article about this recently, and other people have been talking about it, so it's interesting how going through the experience virtually, of course, going through the experience of a possible moral concern like a non-human animal can sensitize humans. Your example is magnificent. I've had the sad experience of seeing a monkey in a cage with something inserted into its head, and you could see on the monkey's face that it was very unhappy. It was one of the most impactful experiences of my life. Not everyone has the chance to visit a laboratory, but if virtual reality technology can be popularized, it could have a positive disruptive impact on our moral sensitivity towards animals that are used for scientific experiments. So yeah, very interesting points that you raised.

Jeroen Hopster: Yeah, I agree. It also has to do with the fact that our moral perception is quite malleable. We can see something and think of it as "bad", or see something without any valenced moral perception. And especially when it comes to the different sensory capacities that we have, I think humans tend to be quite visual creatures, so actually seeing something... that can have a huge impact on our moral perception. Subsequently, we might start to see things differently, for instance as bad. This notion of "seeing as", it was developed by Wittgenstein already. Or take the example which Thomas Kuhn popularized,

you can see something both as a rabbit and as a duck. And so, similarly, you can see something as morally okay, permissible, or as bad. And I think that technologies can actually help trigger such perceptions, and that may be one of the elements that can be very impactful, eventually, in generating societal moral change and societal moral revolutions.

Murilo Vilaça: I will move on to my last question, then, and I'd like to hear from you about the regulatory and governance challenges of the disruptive technologies. I would like more specifically to hear from you about the limits of the dyadic approaches (TechEthics and TechLaw, separately), and how the triadic model that Matthijs Mass and you recently proposed could overcome them. Just a brief comment: the article is fantastic. I came across it very recently, because I've been studying a lot of regulatory and governance challenges, and I haven't been able to identify a consistent proposal in the literature for articulating the ethical dimension with the legal dimension. It's as if they were two things running in parallel, and when you cross these things, it seems that some of the difficulties can be solved. Not that... nothing is easy, of course. But there's a model there, a proposal with an example, which is also very rare. You give an example at the end of the article that really clarifies how this triadic articulation can be much more interesting than what we had until then.

Jeroen Hopster: Thank you for your kind words. When we talk about governance, people often tend to immediately think about legal regulation, which is hugely important, but legal regulation is informed by the moral norms or the social norms that you want to endorse. And in the implementation of legal regulation, we have to think through many ethical issues. There are so many connections between law and ethics, that it's often useful to discuss them in tandem. And as a matter of fact, this notion of disruption, and the more technical scholarship on legal disruption as well as social

disruption, has emerged independently in legal scholarship and in philosophical scholarship. So there was also a reason to take that together. Matthijs Maas and I, we co-wrote the article and we're definitely not the first ones to realize that there's a connection between ethical and legal issues, as there have been quite a few frameworks already. The 'responsible innovation' framework, or the 'ELSA' framework on the legal, ethical and societal aspects of technology... These kinds of frameworks exist. But to operationalize them, in responding to the disruptive implications of technology, I thought that was underdeveloped. I hope that Matthijs and I provide a first step in thinking that through a bit more.

When it comes to regulation in Europe, as said the AI Act has just been endorsed by the European Parliament. So that's a major step. I expect that similar regulations will emerge elsewhere. Legal guardrails for governing AI are developing quickly. But it doesn't end with that. We also have to keep on thinking about the ethical issues involved. You brought up the concept of intelligence, and I mentioned the risk of a sociotechnical imaginary where intelligence is very prevalent and colonizing other values. That's not something that you can regulate of course, but it's an important development nonetheless, that requires a societal response. So thinking through the ethical issues remains of key importance, even with legislation in place.

Also, when it comes to compliance with the law, that's just one thing that disruptive innovators should do. But they should do more than just that. Legal compliance is a low bar, too low for ethical behaviour. To be an ethically responsible disruptive innovator, you have to do more. You have to think through what kind of responsibility gaps might emerge in virtue of new technologies arising, the kinds of uncertainties that it might yield, as well as the kinds of institutions that might be needed in response. Again, that goes towards showing that regulation is not the end of it. It's a very important part of governing technology,

but we have to use legal measures in conjunction with further ethical and institutional measures. Regulation can entrench certain ethical norms and values, and that's very good, but we need a broader approach for a good governance response to emerging technologies.

Murilo Vilaça: Yeah, and I think this is at least one of the aspects of this article that really caught my attention, which is a proposal to operationalize this connection between ethics and the laws. It has been considered in a certain way for a while, but the model that you propose can be applied in various contexts. It's one thing to talk about regulation, which often involves, or perhaps always involves, social institutions that are responsible for formulating and implementing these regulations, right? Here in Brazil, for example, there is what we have come to call, after a process of privatization of some services that were provided by the state, a regulatory agency. So, for example, the telephone service in Brazil was completely state-owned and was privatized. A regulatory agency was created for the telephone service. But often, this regulatory agency is made up of people who are politically appointed, so there are people who have economic interests in these services, and even people who are appointed by the companies themselves that provide the service. It's like putting the wolf in charge of the hen house. So, the way in which the design of a regulatory agency is defined can make a theoretically perfect regulation absolutely useless. It's a bit difficult to create a design that is ethically sensitive. And I think that the values or the way certain values are treated in Brazil is very different from how they are treated elsewhere. So this model has a certain... I don't know if a certain or a great deal of flexibility to adapt to the contexts, which I think is very important. It's a model that's not just top down, you can also go from the bottom up to feed this model so that it generates the result, the most appropriate outcome for that context. This is an aspect that really caught my attention.

Jeroen Hopster: Yeah, I agree. Thanks again. And yes, this is very much meant to be context-sensitive, as well as sensitive to the fact that ethics and regulatory systems co-evolve. They're not static systems, but they co-evolve, partly in response to emerging technologies. So we really have to think of this as a sort of dynamic evolving system. And that's the direction that we are explicitly trying to develop in that paper.

Murilo Vilaça: Thank you so much for the interview.

References

- DANAHER, J.; Hopster, J. The Normative Significance of Future Moral Revolutions. *Futures*, [s. l.], n. 144, p. 103046, 2022.
- Hopster, J. K. G. Socially Disruptive Technologies and Epistemic Injustice. *Ethics and Information Technology*, [s. l.], v. 26, n. 1, p. 14, 2024.
- Hopster, J. K. G. *et al.* Pistols, Pills, Pork and Ploughs: The Structure of Technomoral Revolutions. *Inquiry*, [s. l.], p. 1-33, 2022.
- Hopster, J. K. G. What are Socially Disruptive Technologies? *Technology in Society*, [s. l.], n. 67, p. 101750, 2021.
- Hopster, J. K. G.; Löhr, G. Conceptual Engineering and Philosophy of Technology: Amelioration or Adaptation? *Philosophy & Technology*, [s. l.], v. 36, n. 4, p. 70, 2023.
- Hopster, J. K. G.; MAAS, Matthijs M. The Technology Triad: Disruptive AI, Regulatory Gaps and Value Change. *AI and Ethics*, [s. l.], p. 1-19, 2023.
- POEL, I. v. P. *et al.* *Ethics of Socially Disruptive Technologies: An Introduction*. Cambridge, UK: Open Book Publishers, 2023.

Acknowledgements

I thank the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ - JCNE: E-26/201.377/2021) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): APQ/PRÓ-HUMANIDADES (421523/2022-0); Bolsa de Produtividade em Pesquisa/PQ (315804/2023-8); and Chamada Universal (421419/2023-7) for their support.

Murilo Mariano Vilaça

Pesquisador Titular da Fundação Oswaldo Cruz. Doutor em Filosofia pela Universidade Federal do Rio de Janeiro (UFRJ). Professor dos Programas de Pós-Graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBios) e Saúde Pública (PPG-SP) da Fiocruz; Bolsista de Produtividade em Pesquisa (PQ) do CNPq. Coordenador do Grupo de Investigações Filosóficas sobre Transumanismo e Biomelhoramento Humano (GIFT-H+/CNPq).

Full Researcher at the Oswaldo Cruz Foundation (Fiocruz). PhD in Philosophy (Federal University of Rio de Janeiro/UFRJ). Professor of the Graduate Program in Bioethics, Applied Ethics and Public Health (PPGBios/ENSP/Fiocruz). Professor of the Graduate Program in Public Health (PPGSP/ENSP/Fiocruz). Research Productivity Fellow (PQ/CNPq). Head of the Philosophical Investigation Group on Transhumanism and Human Bioenhancement (GIFT-H+/CNPq).

Endereço para correspondência

MURILO MARIANO VILAÇA

Rua Leopoldo Bulhões, 1480, Manguinhos, 21041-210
Rio de Janeiro (RJ), Brasil

Os textos deste artigo foram normatizados por Araceli Pimentel Godinho e submetidos para validação dos autores antes da publicação.