



Article

Co-addictive human–machine configurations: Relating critical design and algorithm studies to medical-psychiatric research on “problematic Internet use”

new media & society

2024, Vol. 26(12) 7295–7313

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448231165916

journals.sagepub.com/home/nms**Paula Helm** 

University of Amsterdam, The Netherlands

Tobias Matzner 

Paderborn University, Germany

Abstract

Next to popular term Internet addiction, problematic Internet use (PIU) has established itself as an umbrella term for all types of repetitive impairing behaviors associated with new media technologies. Yet, debates about categorization, prevention, and treatment are nowhere near settled. When it comes to classification, medical-psychiatric research has so far retained authority. Here, PIU is examined primarily at the level of the individual user, and it is at this level that solutions are sought. Complementing this, research from critical algorithm studies and technology ethics emphasize the design of many applications as problematic, while cautioning against a determinist view of technology making people addicted. Based on new materialist conceptions of responsibility, the article argues for integrating the different perspectives into a relational understanding of *co-addictive human–machine configurations*. The goal is to capture the interactive character of PIU, and to achieve a well-calibrated distribution of responsibilities in avoiding destructive habits.

Keywords

AI-ethics, algorithmic accountability, critical algorithm studies, dark patterns, habit-forming design, Internet addiction, problematic Internet use, responsibility

Corresponding author:

Paula Helm, Media Studies Department, University of Amsterdam, Turfdraagsterpad 9, BGI, Office 1.04, Amsterdam, The Netherlands.

Email: p.m.helm@uva.nl

As new media technologies have become an integral part of our everyday lives, forms of problematic Internet use (PIU) have also become an issue of growing concern. While there is hardly anyone left in the psychological community who would deny the need for action in this area (World Health Organization [WHO], 2020), controversies heat up when it is no longer just about classifying domain-specific phenomena such as excessive video gaming or compulsive viewing of pornography. For more diffuse but widespread issues such as “binge watching,” compulsive social media use, or Internet-related attention and sleep disorders, opinions vary widely (Bickham, 2021; Reer et al., 2019).

When it comes to issues of official classification, it is primarily psychiatric, neurological, and pharmaceutical research and discourse that asserts authority. Here, problems related to Internet usage are usually thought of from the perspective of individual behavior and the cognitive and mental state of individual users.¹ In contrast, work from critical algorithm studies and technology ethics suggests that the design of many applications, as well as the functions they perform, may also be causal. The claim here is that new media technologies are purposefully designed to manipulate us, identify our vulnerabilities, lead us to build compulsive habits, and thus exploit our attention (Bhargava and Velasquez, 2021; Matzner, 2022; Susser et al., 2018; Williams, 2018). In this article we propose to integrate these different, complementary perspectives. We argue that while the role of design and technology should be given greater consideration in current PIU classification and prevention processes, one must be wary of technological determinism. This can quickly turn into a paralyzing fatalism or an exaggerated urge toward responsabilization when responding to a seemingly threatening and overpowering technology.

Starting from the sole criterion of problematizing harmful effects of new media on mental health and well-being, we identify five relevant disciplinary perspectives in this article. The perspectives do not only differ in their methodological and analytical approaches, but also in their foci, ranging from the psychology of the individual, over the technologies used to the socio-political contexts of usage. While we do not claim to be exhaustive, due to their variety, we assume that with these ideal-typically distinguished five perspectives, we can cover the most important dimensions that need to be considered when striving for a holistic approach. Moreover, it should be emphasized that we do not consider the perspectives to be mutually exclusive, but complementary. They include:

1. the *Medical-Psychiatric Model of Problematic Internet Use*. In this perspective the problem is viewed primarily from the user’s perspective. PIU is defined in terms of qualitative and quantitative patterns that describe how Internet use negatively affects social and psychological functioning. It is usually related to other psychological phenomena such as impulse control disorders or even addiction.
2. In contrast, research in the field of *Interface Design* focuses on the role that software and hardware design play in the causation of problematic habits. This perspective indicates that not only value judgments, but also bad intentions can be implemented into new media design. In consequence, rather than only the nature and extent of use, the design of many applications is considered problematic, such as when it is steered to create habits that tend to become compulsive or even addictive. Here, solutions are primarily sought in value-sensitive design approaches rather than medication or behavioral therapy, such as debated in psychiatric discourse.

3. Another stream of work belongs to the new field of *Critical Algorithms Studies*. Here, questions are raised concerning the impact of machine learning algorithms. Topics include biased training data or attempts at predicting character traits that facilitate manipulation. Issues of habit-forming design covered in Point 2 are discussed primarily from the view of Human–Computer–Interaction. The rise of digital platforms as powerful economic actors that integrate many applications pertinent to PIU and the rise of personalization based on machine learning make it additionally necessary to take these issues of distributed, networked, and market-driven socio-technical systems into account that operate “behind” the interfaces and seemingly innocuous utility functions.
4. In turn, *AI-Ethics* evaluates problems of loss of control regarding fundamental human values such as autonomy, which are analyzed as being threatened by manipulative design and technology. In this context, close references to works from Critical Algorithm and Design Studies can be found, but the focus is on the normative evaluation of the phenomena as well as the ethical justification of potential consequences, in addition to empirical analyses.
5. Finally, digital well-being movements are being addressed in *Studies of New Media & Digital Culture*. Related social trends such as digital detox react to phenomena of problematic use, proposing an alternative to medicalization and psychotherapy. Here, people are experimenting with ways in which individuals and communities can manage their Internet use more healthily in their everyday life’s.

In the following, we outline each of the five perspectives in more detail, followed by a discussion of their different sociopolitical implications. Thereafter, we present an approach to bringing the five perspectives together. Against this backdrop, we conclude by arguing for a multi-perspective responsibility regime that distributes responsibility among different actors. Such regime takes the increasing (manipulative) power and influence of digital platforms and ecosystems very seriously, while still acknowledging the diversity and autonomy of users.

Medical-psychiatric model of PIU

Along with other terms such as Internet addiction, social media dependency, and excessive usage, PIU has become a key expression in medical-psychiatric discourses dealing with the matter. According to the manifesto of the European Research Network into Problematic Internet Use (2018), PIU is defined “as an umbrella term for a range of repetitive, impairing behaviors related to Internet use” (Fineberg et al., 2018). Currently, there is much debate about how and in what form corresponding diagnoses might be integrated into the relevant medical diagnostic tools, particularly the Diagnostic and Statistical Manual of Mental Disorders (*DSM*) of the American Psychiatric Association (APA) and the International Classification of Diseases (ICD) maintained by the World Health Organization (WHO). However, discrepancies in classification and diagnosis have prevented inclusion to this date (Griffith et al., 2016).

A body of studies is emerging in psychiatric discourse that set out to examine the issue in both scale and detail. As a result of increased research activity, an initial consensus is

emerging on several key points. For example, the literature almost consistently highlights clear similarities between cases of PIU, addictive behaviors, and impulse control disorders (Spada, 2014). In addition, many argue that genetic and personality differences favor the development of PIU. Pharmacological and psychotherapeutic treatments are being tested in this context. Initial evidence appears to suggest that both psychotropic medications and cognitive behavioral therapy may have utility in the treatment of PIU (Bickham, 2021).

Beyond these initial movements toward consensus, profound discrepancies regarding fundamental issues related to PIU continue to prevent classification under the APA's DSM or, alternatively, the WHO's ICD (van Rooij and Kardefelt-Winther, 2017). For example, authors have cautioned that given the widespread proliferation of potentially compulsive patterns of behavior related to technology use, there is a danger of pathologizing society (Aarseth et al., 2016). A related debate focuses on the question to what extent PIU is a symptom that relates to deeper rooted psychological causes (Kardefelt-Winther, 2014). A peculiar problem of PIU is that it is often discussed regarding children and adolescents—leading to the question whether PIU as a parenting problem should be separated from possible forms of psychopathology (Griffiths et al., 2016; Livingstone and Helsper, 2008). This also raises the question whether PIU requires its own classification as a new problem, or whether it is to be seen as a variant of already existing categories. Foremost among the former is the suggestion that PIU is a new form of behavioral addictive disorder. This leads on to the question whether PIU should be labeled as a form of addiction—as it is now frequently the case in public discourse—or whether there are important differences to addictions—superficial similarities notwithstanding (Griffiths et al., 2016). For example, operationalizing central diagnostic elements such as “tolerance” or “withdrawal” is much more difficult even for gaming than for substance-related addictions, due to the breadth and complexity of both games and gaming behavior. Since gaming disorder (intending video games) was included in the newest version 11 of the ICD, as well as a condition requiring further study in the *DSM*, a related question is whether a broad category, such as Internet addiction, is an apt categorization, or if the phenomenon rather summarizes quite distinct issues that should be treated separately, for example, compulsive smartphone usage, binge watching, screen-time related sleeping disorders, and so on.

Such questions are familiar from the more general discussions of the similarities and differences between substance addiction and already more established behavioral disorders such as gambling or eating disorders (Holden, 2001). The current state of research can be summarized as suggesting that “behavioral addictions such as Internet addiction are similar to drug addiction, with the difference that in the former, the person is not addicted to a substance, but to the behavior or feeling evoked by the corresponding action” (Alavi et al., 2012).

However, with respect to drugs, the addictive features of the substance have been researched extensively. In contrast, psychological, and medical research posits “behavior or feeling” as alternative to “drug”; giving little consideration to the extent to which the technology used, with its particular design and features, is a critical element in triggering and maintaining potentially addictive behavior (with few very recent exceptions, Chen et al., 2022). In consequence, the notion of an Internet addiction, which is sometimes

discussed by analogy to food in eating disorders, sometimes to drugs in substance abuse, also should be reflected regarding the different implications and sociotechnical appropriateness of such comparisons (Sutton, 2017).

Given these controversial, but crucial questions with far-reaching medical and policy implications, calls for more extensive, epidemiological, and longitudinal studies have been raised. Many of the more recent studies consider social factors; in particular, issues such as isolation during the Covid-19 pandemic and its correlation with patterns of PIU are at the center of attention (Iqbal et al., 2022; Nguyen et al., 2020).

While this move from individual behavior to social issues is a step in the right direction, with this article, we aim to stimulate an even broader interdisciplinary approach. We believe that such an approach will contribute to a more comprehensive understanding of PIU, what is at stake, and how best to respond. Among other things, we assume that some of the hitherto controversial questions have remained unanswered precisely because they require, in addition to psychological and medical perspectives, an approach that is capable of adequately assessing both the role of the technologies involved and their complex embeddedness within socio-historically situated societies. Furthermore, we argue that an ethico-political evaluation should be considered alongside clinical and evidence-based psychiatric research when discussing politically relevant issues related to the classification, diagnosis, prevention, and treatment of PIU. We believe this is necessary to address the complexity of a phenomenon that is, by its nature, psychological, physiological, sociotechnical, and socio-political. To back this contention, in what follows, we will integrate the psychological-medical views with diverse scientific perspectives on the same complex, and finally relating them to each other.

Interface design

Turning to the existing research on PIU from the field of design studies, a not necessarily contradictory, but complementary perspective on the phenomenon emerges. Here PIU is seen less as a problem of individual users, their psychology, or their brains, and more as a problem that depends heavily on the design of certain technologies, their interfaces, their hardware, and their underlying functions.

This view strongly resonates with a public wave of criticism of digital platforms, which was initiated primarily by GAFAM (Google, Amazon, Facebook, Alphabet, Microsoft) whistleblowers such as Greg Hochmuth, Chamath Palihapitiya, Tristan Harris, James Williams, Sandy Parakilas, and Frances Haugen. These (former) employees of platform companies accuse them of distributing products designed to exploit their users' socio-emotional vulnerabilities for the benefit of corporate profits, with no concern for the potentially destructive consequences. They identify not only issues such as political and social polarization, surveillance, and fake news, but also users' mental and emotional health (Williams, 2018). The main targets of platform operators are users' attention and time because the longer users stay, the more data—the greatest asset of digital platforms—they produce (Zuboff, 2019). The problem is that triggering so-called compulsive behavior patterns is one of the most effective ways to achieve this goal.

There are several design mechanisms and features that have been created for the specific task of forming user behavior toward a specific direction. The products that exhibit

such design properties can be classified as what has been called persuasive (Fogg, 2002) or habit-forming technologies (Eyal, 2014), or “dark patterns” (Gray et al., 2018). They are designed to hook users to garner as much attention, attachment, and data as possible. Qualitative inquiry and ethical research have examined and analyzed the processes by which employees in large Silicon Valley tech companies have been progressively trained and professionalized to form a new guild of habit-forming design experts (Schüll, 2012; Williams, 2018).

To capture what is happening here in terms of human-computer interaction, and thus bridge the user-centered psychological and the techno-centered design perspectives, Natasha Dow Schüll has coined the term “addiction by design.” Based on the results of a long-term ethnographic study of slot machines and online gambling, she argues that gambling addiction disorder should be defined not only as an internalized cognitive state or individual mental disorder, but as the result of an asymmetrical interaction between humans and machines (Schüll, 2012). While the term “addiction by design” initially referred explicitly to gambling addiction, many studies from the field of media and communication studies now also apply it to the critical analysis of dating, social media, or booking platforms (Rosenblat and Stark, 2016; Turkle, 2016; Vaidhyanathan, 2018). In this context, Schüll’s work is sometimes interpreted and taken up in an overly techno-deterministic way.

In light of positions such as those of whistleblower Tristan Harris, one sometimes almost gets the impression that we are helplessly at the mercy of technology (Harris, 2020). Such views can have a counterproductive effect on our ability to act, especially when, instead of demystifying the industry they seek to criticize, they reproduce their imaginary of new media technology as a determinist instrument (Fogg and Hreha, 2010). This view is also echoed in some parts of the psychiatric-medical perspective, when technology is seen as something given and fixed that affects the human mind according to fundamental properties (such as cognitive biases). Schüll herself, contrary to deterministic readings of her work, repeatedly emphasizes that the phenomenon of people losing control over their usage can hardly be meaningfully understood without taking power asymmetries between users and providers into account. They point to the social and political dimensions inherent in the phenomenon.

While it is central to highlight that the behavioral effects of current digital platforms, social media, and other applications are no accident, but a strategic feature, the contextuality, malleability, and social embeddedness on both sides of the interaction (technology and user) should not be overlooked.

Given these considerations, we propose to speak of co-addictive human-machine configurations, thereby putting the focus on the *relationship* between users and their devices, both understood in a socially situated manner. We do this out of the conviction that it is important to precisely understand the interplay between users with certain dispositions, living under certain social circumstances, on one hand, and personalized data analytics and specific technology designs, on the other. In this way, it becomes possible to give greater consideration to the role of technology and its design than is the case in current psychiatric discourses; while remaining wary of technology determinism, which can quickly lead to sweeping generalizations and produce a crippling fatalism.

Critical algorithm studies

In addition to persuasive design, there is another technological factor that becomes relevant in the production of PIUs. Current media platforms involve artificial neural networks, and other recent machine learning technologies that are often referred to as “black boxes” (Pasquale, 2015). The use of these technologies complicates matters considerably, as not only can strategic design decisions create problematic impacts. Unanticipated machine learning dynamics, whose unfolding is often much more difficult to fathom, can also play a critical role here (Beer, 2017). Current regulatory regimes that rely on informed consent are criticized as inadequate when large platforms have the power to identify, create, and exploit cognitive dissonance and emotional vulnerabilities based on data analytics (Van Dijck et al., 2018).

Let us illustrate this point with an example. In an interview with Mark Zuckerberg’s speechwriter, Kate, she recalls how a gaming application called Zynga was developed. This gaming application involved an optimization algorithm that learned from real-time data fed back into the system as the users played. At some point, the application proved to be optimized to an extent that it performed the single utility function “increasing time spent on the device” so effectively that it was eventually deemed “too good” in the sense that it was considered too “addictive.” It threatened the platform’s reputation as a social platform:

With Zynga, you spent a lot of time on Facebook, and people got addicted to it. [. . .] that was a problem because it went too far. It was making tons of money without really putting anything back into the social media ecosystem that could then be used (Kulwin, 2018).

This is an example where it is not primarily a strategic, original design of the application, but a self-learning algorithm programmed to fulfill a seemingly innocent utility function that is considered responsible for the emergence of compulsive usage patterns. Prominent machine learning algorithms make it difficult even for the programmers themselves to reconstruct how exactly the highly addictive effects were created.

These are serious issues that need careful framing, however. Critical algorithm studies question the narrative of a seemingly innocent function as portrayed by Losse. It has been shown that the efficacy of algorithms emerges out of the combination of several factors, where programming is just one that has to be combined with the context of application, material conditions, data collection and preparation, and so on (Gillespie, 2014; Introna, 2011; Mann and Matzner, 2019).

Furthermore, critical algorithm studies have emphasized the relevance of unreflective or implicit factors that influence programming (Campolo et al., 2017; Rieder, 2020). In the case of the apparent self-critical take by Facebook’s engineers on Zynga, for example, the general view of a platform economy that subsumes friendship, news, games, research, and much more under the rubric of “content” that serves to create an audience for advertising is rarely questioned. More generally, factors like the demographic of a programming team or their location can import certain assumptions and biases into the program—often in a manner that is difficult to notice (Friedman and Nissenbaum, 1996). Finally, the changes in machine learning models are not entirely unpredictable, as they

are driven by a variety of socio-economic factors, such as adaption to new laws, dynamics of platformization, and changes in political regimes (Mackenzie, 2015). In particular, the emergence of platforms (Bucher and Helmond, 2017) entails a concentration of social networking, games, text messaging, and more while tracking their users over many devices (smartphones, laptops, smart watches, home assistants, etc.). This gives proprietary machine learning algorithms unprecedented amounts of data as input. At the same time, the impact of the algorithms is multiplied. This tight interdependence of platformization as a structural condition for the development of machine learning is rarely acknowledged by its practitioners, while it clearly orients their design and application (Mühlhoff, 2019). In summary, while it can well be the case that PIU is not the result of conscious design decisions, the emergence of factors that contribute to PIU via machine learning is no pure accident either. It can be traced to the complex interplay of the factors discussed in this section. This doesn't imply to reduce the responsibility of programming and design. In fact, it increases the responsibility by demanding a view toward a larger social context and socio-technical dynamics.

AI-ethics

Parallel to these empirically informed discourses, techno-ethical analyses discuss the destructive effects certain forms of technology design can have on personal as well as basic democratic values. For example, a major threat to personal autonomy is that it is not only extremely profitable for corporations to trigger compulsive patterns of use through habit-forming design. Given the increasingly precise profiling capabilities offered by comprehensive data analytics and the ubiquity of digital devices, the potential to implement manipulation much more easily, unobtrusively, and ubiquitously is also growing. According to the European Commission's (EU, 2021) proposal for the regulation of AI, so-called pattern-building technologies are to be classified as particularly risky. Compounding the technical systems and logics themselves is the enormous power of large platforms, resulting from the catalyzing effects of sheer numbers of users and the ensuing market advantages, which in turn attract even more users (Van Dijck et al., 2018). These effects make it virtually impossible for individual users to refrain from using certain platforms if they want to participate in social life and democratic deliberation (Helm and Seubert, 2020). If we further assume that these platforms have a manipulative, or even addictive effect on their users, at some point it becomes almost impossible to distinguish autonomous behavior from manipulatively influenced behavior. Moreover, notions of what counts as normal user behavior in terms of, for example, frequency and duration can creepily shift from 4 to 8 to 16 hours a day. Crises, such as the Covid-19 pandemic, can obviously exaggerate this process of shifting the norm (Nguyen et al., 2020). This is not to say that what is considered normal or pathogenic behavior is essentially contingent on cultural norms, but it is an argument for the situatedness of (self-) perception and judgment of one's own behavior and the behavior of others, which is arguably an important factor in terms of personal distress and social pressure (Helm, 2019; Mieczkowski et al., 2020).

This demarcation problem of where autonomous user behavior begins, and ends is further complicated by the fact that most scholars today would agree that autonomy does

not exist independently of social norms and structures anyway (Christman, 2014; Mackenzie and Stoljar, 2000; Oshana, 2006). Autonomy, that is, the ability to shape one's own life according to one's own choices (Rössler, 2021), must be understood as always already closely interwoven with the performative influence of sociopolitical norms on one's will, as well as the social conditions for its realization. Furthermore, human autonomy as we know it is itself dependent on technology (Chun, 2011; Haraway, 1985). Considering such a socio-technical understanding of autonomy, the question of online manipulation emerges less as an individual-ethical nor as a psychiatric or psychological one, but rather as an ethico-political question: which forms of autonomy are possible in our societies and for whom (Cheney-Lippold, 2011)? In which ways do we want social norms and cultural practices to be shaped by powerful platforms, and in what ways do we not?

To provide a possible answer to this question, Daniel Susser et al. (2018) have clarified what online manipulation is and analyzed why it is problematic. They define manipulative technologies as those that are "strategically designed to exert covert, exploitative, and targeted influences on users." An exemplary mechanism in this context is so-called automated "hypernudges." Based on an analysis of large data sets, emotional vulnerabilities are identified to target incentives that are meant to covertly induce users to behave in a certain way that may deviate from the user's original plan or will (Yeung, 2017). Two aspects are ethically crucial here: First, the fact that influence happens covertly, which prevents reflection, critical engagement, and thus autonomy. It is what distinguishes persuasion from manipulation. The second aspect is the exploitation of vulnerabilities that deserve protection. These two aspects enlarge the power asymmetry underlying this process, which has already been discussed in connection with Schüll's analyses of online gambling. Platform operators are today able to create very precise profiles about each individual user based on the automated analysis of data. These profiles enable platforms to get at users' vulnerabilities. The users themselves, then again, know very little about the platform operators' approaches.

Therefore, in addition to the already discussed factors it is additionally important to emphasize ethico-political issues such as the possibilities for autonomy or the asymmetric distribution of power. This applies not only to issues of a more precise and context-sensitive classification of PIU, but also to those of treatment and prevention, that is, what kind of measures are needed and where they need to be applied to effectively prevent various forms of PIU.

Studies of new media and digital culture

Cultural Studies has traditionally been concerned with how practices are shaped and how popular culture takes up, deals with, disguises, and influences social developments. Looking at how the issue of PIU is being taken up in popular new media, it is apparent that alongside the frightening visions of Internet addiction, digital detox is a forward-looking, less pathologizing or medicalizing, but very trendy trope. It can be seen as part of the larger digital wellness movement (Vanden Abeele, 2021). It has established itself as a fashionable expression, which is applied to very different things—ranging from cosmetic product lines to nostalgic back-to-nature camps. Despite this diversity,

however, there is now a definition that precisely characterizes and narrows down the concept. According to this, digital detox refers to “the periodic disconnection from digital media or a strategy to reduce engagement with media technology” (Syvertsen and Enli, 2019). Basically, digital detox stands for awareness and caution in usage. As such, it may involve many activities, far from being reduced to radical disconnection, such as during retreat. Instead, being an integral part of a digitized world, it can also involve activities such as just switching off the phone during meals, notifications during class, or certain apps during physical activity.

As a cultural practice, digital detox became internationally visible with the founding of the largest and probably best-known venue: Camp Grounded in California. Beginning with the first camp in 2012, the participants explicitly positioned themselves not only as a therapeutic but also as a social movement. According to Theodora Sutton’s (2017) analysis, the “detoxers” present wanted to express with this self-identification that they were interested not only in improving their personal health, productivity, and well-being but also in political issues related to the growing influence of large tech companies.

An interesting aspect of digital detox today, however, is the paradox that while it began as a media resistance movement, it is now increasingly being promoted and discussed through the very digital platforms that many detoxers originally sought to resist (Hesselberth, 2018). This can be interpreted not least as an expression of the enormous power of digital technologies that is very hard to confront for social movements. Besides, digital detox has been commercialized. Today, it is often promoted as a luxury product for successful tech people who can afford to take a break from their work and to refill themselves with new energy by taking a break while relaxing at remote places (Syvertsen, 2020). As a result, Digital Detox has lost both its original emancipatory character and its credibility as a social movement.

On a different level of analysis, digital detox has attracted criticism for its underlying ideologies (Fast, 2021). Various studies analyzed digital detox camps as being mobilized by the polarizing assumption of a conflict-ridden state of alienation and overload: The subject must detoxify itself to find its way back to a connection it once lost somewhere in the shallows of cyberspace. This idea finds expression in the popular slogan: disconnect to reconnect. This evokes romantic notions of a pristine nature and essentialist ideologies of a true pre-cultural self. These ideologies perpetuate a nature-culture and nature-technology dualism that has long been criticized (Haraway, 1985) for preventing people from making things better and fighting for their rights as digital citizens (Isin and Ruppert, 2015), that is, advocating for “better” culture and technology (Kaun and Treré, 2020). Instead, it polarizes and thus risks promoting radicalization on both sides: on the side of the techno-determinists and on the side of nature essentialists.

In a society where not using digital technologies is hardly an option, such a polarizing approach is not necessarily helpful in supporting people in preventing PIU. Instead, such a form of prevention, operating under self-optimization imperatives, may be more likely to build pressure and stress, which is obviously not good for mental health either. However, as an alternative to a neoliberal, self-responsible and market-oriented self-optimization model of digital detoxification, it is equally questionable whether digital detox as a mandatory component of state-funded treatment programs and even so-called “Internet Addiction Boot Camps,” which are now common in China and South Korea

(Koo et al., 2011), is something that Western societies should strive for. As a result, recent discussions have increasingly revolved around more nuanced versions that focus on disconnection activities as part of a broader range of digital well-being activities, such as turning off notifications, rather than radical detoxification (Vanden Abeele, 2021). More nuanced, gentle approaches, however, are also more focused on self-regulation in the case of normal patterns of use, but may do little in the case of actual, pathological symptomatology. They, moreover, move away from addressing the broader socio-political contexts, which was still the case in the early detox movements, and tend to focus one-sidedly on the options of the individual user.

Response-able use AND design

The different perspectives that we have presented correspond to very different regimes of responsibility, ranging from individualistic to technocratic, with far-reaching political and social consequences. In the following, we work toward an integrative perspective on responsibility. To that aim, it should first be made clear that responsibility begins with engagement through questioning. This questioning entails an already deeply political consideration of who is considered capable of providing an answer that either calls for action or can be meaningfully challenged. Karen Barad (2007) has developed the notion of “response-ability” around the very idea that not only is the assignment of responsibility political, but so too are the questions that precede that assignment. According to her theory of “agential realism,” response-ability—the ability to respond—is not primarily about the right answer. Rather, it is about inviting and enabling the response of an “other,” whether that “other” is a “user,” a “programmer,” a “designer,” a “data broker,” a “platform operator,” a “CEO,” a “machine,” an “algorithm,” a “political activist,” a “neurologist,” a “pharmacologist,” a “patient,” or a “psychotherapist.” In Barad’s relational “onto-epistemology,” these perspectives are not just individual parts that need to be summed up to get the full picture. According to Barad, each of the different perspective we reviewed amounts to a particular “cut” between an object of observation and its observers. This cut through a more primary relation or “intra-action”² brings certain phenomena to the fore—at the exclusion of others (Barad, 2007: 114). In consequence, the range of possible responses that are invited, as well as the types of responses that are excluded, is constrained, and conditioned by the questions asked, where each question implies a particular epistemic position or “cut.” This range, and the issues of answerability and exclusion it raises, differs significantly depending on which disciplinary perspective one considers.

Consequently, each perspective must include the reflection of positions or conceptualizations that are excluded by the present “cut.” This follows from a central insight by Barad. A full picture following traditional epistemics of disinterested objectivity is not possible, yet this does not mean that each position suffices. Since each position is the result of relations,

objectivity requires an accounting of the constitutive practices in the fullness of their materialities, including the enactment of boundaries and exclusions, the production of phenomena in their sedimenting historicity, and the ongoing reconfiguring of the space of possibilities for future enactments. (p. 391)

The question, then, is not only who is responsible according to certain criteria, but also who is engaged to be answerable, and what questions are asked that allow actors to answer or to be cloaked in disengaged silence. Responsibility, the ability to respond, consequently “entails an ongoing responsiveness to the entanglements of self and other, here and there, now and then” (p. 394).

In this article, we urge the development of such a response-ability based approach to PIU. Conceiving of the different perspectives here presented as cuts through a complex relationality provides a view that is interdisciplinary yet does not dissolve the important particularities of each discipline. In consequence, distributing responsibility means not to aim for a “grand unified theory.” Rather, it is necessary to ask for each field how it distributes response-ability and which other relations to the own objects of observation are necessary to make oneself, that is, the observer, more response-able.

Socio-political implications

In the first scenario of the medical-psychological model, technology designs are conceived as an externalized and thus fixed condition for peoples’ psychological or mental processes. Thus, neither engineers, nor designers, nor platform operators are asked about the causes or the possible effects of usage behavior. The goal here is, instead, to provide empirical evidence based on which to unambiguously define and classify compulsive patterns for a medial context. This evidence is collected regarding the user’s mental, cognitive, and emotional state as well as behavior. By establishing new classes of disorders based on such factors, it becomes possible to turn users into patients through appropriate diagnoses. Through this act, some of the user’s response-ability is lifted from them and shifted instead to therapists, doctors, psychiatrists. Such processes are essential to relieve users not only of the moral burden but also of the financial responsibility for treatment costs. This can be achieved by developing and providing medications, drop-in centers, and treatment programs that are part of standard health insurance plans, and whose use is thus not dependent on the economic wealth of users, as is currently the case with “luxurious” digital detox retreats.

However, the creation of such opportunities should also be accompanied by the recognition that PIU unfolds in the interaction between user and machine—in co-addictive human–machine configurations. Here, Barad’s observation that each perspective or cut simultaneously enables and excludes becomes relevant to questions of responsibility. In other words, treating PIU as a disease should not mean paying too little attention to sociotechnical factors and power issues.

Clearly, treatment cannot wait until society has adapted, design trends have changed, power dynamics have shifted, or a different technology is available. However, a thorough understanding of the issues and responsibilities should not be constrained by acute rationalities. Studies of affect and media have shown that many of the supposedly unconscious and reflexive responses of users, which arguably can be triggered by machines, are highly dependent on the relation of specific technologies with socio-biographical characteristics. This has been studied, for example, in relation to radicalization and the proliferation of hate online (Berlant, 2011). Wendy Chun has developed the concept of “habitual media,” which uses a socially situated notion of habitualization (Chun, 2016).

Such concepts of habitualization can help explain social as well as individual differences in the effects of algorithmic technologies (Matzner, 2019). Such findings should be interrogated regarding their potential to contribute to the elucidation of PIU.

Such a socio-technically situated approach is particularly important when considering questions of response-ability of the users themselves. If autonomy is a relational matter, as has been discussed above, it is important not to ask too much responsibility from persons whose social situation does not enable the required form of autonomy. If responsibility is the demand to respond to certain questions, autonomy can be conceived as the ability to respond in a manner that is appropriate to one's capabilities (Sen, 2004; Westlund, 2009). In consequence, both need to be carefully calibrated. Miscalibrations are discussed as responsabilization, an issue yet to be reviewed in more detail. In the worst case, miscalibration of responsibility and autonomy can lead to cases of victim blaming in media use. In this regard, rather than just focusing on lifting responsibility from patients to therapists, medial research could also play a bigger role in putting pressure on other potentially responsible actors, such as platforms or software engineers, by developing criteria for drawing clear lines between "licit" and "illicit" designs, for example.

The push for a more socio-economically situated perspective must not entail that PIU is merely a social problem that is mirrored in technology. It also derives from certain features of new media. This is where the perspective of Critical Algorithm Studies stands out. It focuses on the co-addictive and manipulative effects of certain technologies and thus emphasizes developers' responsibility. However, in the praxis of engineers and programmers, responsibility is often shifted away. This is done either by invoking a narrative according to which they initially pursued a utility function that seemed innocent enough to be ethically legitimate, which then happened to go wrong. Another strategy is to turn responsibility itself into an optimization problem to be "solved" by an algorithm.

This is where the design studies perspective excels, as it can quite clearly identify problematic design elements and thus provide a basis for pragmatic and relatively easy-to-operationalize recommendations for the health- and value-centered design of applications. This perspective can be extended to include mechanisms of dealing with the developments of learning algorithms. Being response-able here means to not just see them as more or less successful attainment of a clear aim—the utility function—but as more or less explicitly oriented by a wide range of factors, such as path dependencies, social situation of the programming context, general economic motivations of the platform, and so on. As soon as such aspects become part of the questions that programmers and platforms are asked to answer, the attempts at shifting responsibility away are losing appeal.

Current activities of social media platforms can already be seen as a particular form of addressing such questioning. Since the platforms know that discussions about the toxicity of their apps and services pose a risk to their marketing, they are addressing the problem head on by developing digital detox apps and digital well-being programs themselves. Recently, they even came up with an algorithm that collects data on when user behavior becomes pathogenic, and then generates a nudge stating something like: "You've been using this app for two hours now, would you like to take a short break?"

Yet, this is a kind of pseudo-responsibility because users still are manipulated. Neither algorithm nor nudge are there to encourage the user to respond; rather, they engage in behavior that is strategically calibrated according to the interests of other actors, which spills over just enough so that it does not yet cross the line into pathogenicity, but still maximally fills the entire space below.

Using an integrative perspective on the distribution of response-ability in co-addictive human-machine configurations, as proposed here, allows us to distinguish such enactments of a responsabilization as the opposite of inclusive response-ability. Here, companies invest in non-profit endeavors for marketing reasons, promoting values such as privacy, health, and consumer protection, but not for the sake of these values themselves, but for economic considerations. Building on this is a reversal of priorities, where values are subordinated to business logics, not the other way around (Shamir, 2008).

In stark contrast, in the original push for the Digital Detox movement, the actors took the problem into their own hands. This also makes Digital Detox interesting from an ethico-political perspective: Here, solutions are not developed and tested top-down, but bottom-up. This explains that contrary to the name, Digital Detox is not just about the distance from technology but an attempt in reversing power structures and fostering intra-actions that yield different configurations of agency. In the early Digital Detox movements, participants organized themselves, considered strategies for a sovereign, autonomous, healthy use of potentially manipulative media technologies and joined forces with people pursuing the same goal. The problem, however, is that this engagement was and still is dependent on the resources of the individual participants. A lack of institutional embeddedness means that Digital Detox has become a luxury phenomenon that depends on the resources of individuals and/or increasingly results in people trying to make money from it by offering courses and products privately and for a fee. This illustrates how response-ability is an ability that emerges in intra-action with other factors.

In its individualized form, Digital Detox is inevitably bound to become commodified, which is not problematic per se, but the commodification restricts its originally emancipatory and political goals to specific sections of the populace. Even more, digital detox can be seamlessly integrated into a pre-existing neoliberal regime of self-enhancement. What looks like responsibility here, then, is a necessity to respond—or rather perform—in a preconfigured way rather than fostering response-ability. Ultimately, it is this very unfolding of Digital Detox that shines a spotlight on the enormous importance of proceeding in matters of official classification of PIU and the corresponding establishment of government-aided treatment programs. Such mechanisms may in fact serve as safeguards against an overly simplistic incorporation of the problem of PIU into economy-driven logics of platforms. Acknowledging the seriousness of the medical problem might help to promote response-ability rather than responsabilization. This, in turn, is the basis for a plea to move the ongoing discussions in WHO and APA forward wisely, thoughtfully, and sustainably by integrating diverse interdisciplinary perspectives.

Conclusion: combining strengths, overcoming limitations


Five perspectives have been reviewed in this article, all problematizing detrimental mental health effects of new media, but from very different angles: the psychiatric-medical

one, perspectives that focus on the design of the technologies involved, the algorithms operating in the backend, ethical evaluations and perspectives studying cultural practices responding to PIU. Drawing these perspectives together, we sought to overcome a lack in integration, particularly between the medical-psychiatric one and the other ones. This is vital, as we identify this very isolation of perspectives as one of the main reasons for prevailing shortcomings regarding lack of or ineffective regulation, prevention, and treatment. This is because each discourse, while making important points, suffers from limitations. These can only be overcome by combining perspectives. We therefore advocate for an interdisciplinary, multi-perspective approach to the topic. Grasping the different forms of responsibility in each field needs a responsiveness to the complex entanglements of the phenomenon that other perspectives can provide better. Each of the reviewed perspectives does have its own logics and requirements that certainly cannot be easily overcome. Sometimes this sturdiness is indeed helpful, for example, in the possibility to curtail responsabilization from within the medical-psychiatric field. But even when it can serve such a function, it needs to be reconsidered in terms of its interconnectedness with other perspectives. In particular, the complex role of specific forms of design and new media as a socio-technical system must be considered when discussing new approaches to prevention, treatment, and regulation. As a first step toward such better integration, we propose to move away from talking about PIU, which suggests a focus on the individual, and instead use the term: co-addicted human-machine configurations. This term can contribute to a better understanding by highlighting the relational nature of the respective phenomena and promoting distributed responsibility regimes.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Paula Helm  <https://orcid.org/0000-0002-2719-9721>

Tobias Matzner  <https://orcid.org/0000-0002-1439-2418>

Notes

1. This applies to both the committees around the Statistical Manual for Mental Disorders of the American Psychiatric Association and the International Classification System of the World Health Organization. Both are decisive for how psychological illnesses are dealt with in Europe and the United States. Thus, it is a matter of social recognition, the establishment of specific treatment pathways, the assumption of treatment costs, and the establishment of institutional structures. In addition, official recognition can lay the foundation for further steps related to the regulation of disease triggers (à Role of Technology) and prevention (à Digital Literacy).
2. Physicist Barad (2007) defines intra-action “in contrast to the usual ‘interaction’, which presumes the prior existence of independent entities/relata.”

References

- Aarseth E, Bean M, Boonen H, et al. (2016) Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. *Journal of Behavioural Addictions* 6: 267–270.
- Alavi S, Ferdosi M, Jannatifard F, et al. (2012) Behavioral addiction versus substance addiction: correspondence of psychiatric and psychological views. *International Journal of Preventive Medicine* 3(4): 290–294.
- American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing.
- Barad K (2007) *Meeting the Universe Halfway*. Durham, NC: Duke University Press.
- Beer D (2017) The social power of algorithms. *Information, Communication & Society* 20(1): 1–13.
- Berlant L (2011) *Cruel Optimism*. Durham, NC: Duke University Press.
- Bhargava VR and Velasquez M (2021) Ethics of the attention economy: the problem of social media addiction. *Business Ethics Quarterly* 31(3): 321–359.
- Bickham DS (2021) Current research and viewpoints on internet addiction in adolescents. *Current Pediatrics Report* 9: 1–10.
- Bucher T and Helmond A (2017) The affordances of social media platforms. In: Burgess J, Marwick A and Poell T (eds) *The SAGE Handbook of Social Media*. London; New York: SAGE, pp. 223–253.
- Campolo A, Sanfilippo M, Whittaker M, et al. (2017) *AI Now 2017 Report*. AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf (accessed 5 September 2021).
- Chen C, Cohen O and Sundar S (2022) Differentiating problematic from habitual instagram use: a uses and grats 2.0 perspective. *Social Media + Society* 8(3): 20563051221116339.
- Cheney-Lippold J (2011) A new algorithmic identity: soft biopolitics and the modulation of control. *Theory, Culture & Society* 28(6): 164–181.
- Christman J (2014) Relational autonomy and the social dynamics of paternalism. *Ethical Theory and Moral Practice* 17(3): 369–382.
- Chun W (2011) Race and/as technology, or how to do things to race. In: Nakamura L and Chow-White P (eds) *Race after the Internet*. London: Routledge, pp. 38–60.
- Chun W (2016) *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press.
- Van Dijck J, Poell T, Waal M, et al. (2018) *The Platform Society. Public Values in a Connected World*. Oxford: Oxford University Press.
- European Commission (EU) (2021) Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206> (accessed 7 October 2021).
- Eyal N (2014) *Hooked. How to Build Habit-Forming Products*. London: Portfolio Penguin.
- Fast K (2021) The disconnection turn: three facets of disconnection work in post-digital capitalism. *Convergence* 27(6): 1615–1630.
- Fineberg N, Demetrovics Z, Stein DJ, et al. (2018) Manifesto for a European research network into problematic usage of the internet. *European Neuropsychopharmacology* 28(11): 1232–1246.
- Fogg BJ (2002) *Persuasive Technology: Using Computers to Change what we Think and do*. New York: Ubiquity.
- Fogg BJ and Hreha J (2010) Behavior wizard: a method for matching target behaviors with solutions. In: Ploug T, Hasle P and Oinas-Kukkonen H (eds) *International conference on persuasive technology*. Berlin; Heidelberg: Springer, pp. 117–131.

- Friedman B and Nissenbaum H (1996) Bias in computer systems. *ACM Transactions on Information Systems* 14(3): 330–347.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot K (eds) *Media Technologies*. Cambridge, MA: MIT Press, pp. 167–193.
- Gray CM, Kou Y, Battles B, et al. (2018) The dark (patterns) side of UX design. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. Available at: <https://dl.acm.org/doi/pdf/10.1145/3173574.3174108>
- Griffiths MD, van Rooij A, Kardefelt-Winther D, et al. (2016) Working towards an international consensus on criteria for assessing internet gaming disorder: a critical commentary on Petry et al. (2014). *Addiction* 111(1): 167–175.
- Haraway D (1985) Manifesto for cyborgs: science, technology, and socialist feminism in the 1980s. *Socialist Review* 80: 65–108.
- Harris T (2020) Addiction is a feature, not a bug, of social media platforms. *The New York Times*, 9 September. Available at: <https://www.nytimes.com/2020/09/09/movies/the-social-dilemma-review.html> (accessed 2 July 2021).
- Helm P (2019) Sobriety versus abstinence. How 12-stepper negotiate long-term recovery across groups. *Addiction Research & Theory* 271: 29–36.
- Helm P and Seubert S (2020) Normative paradoxes of privacy. Literacy and choice in platform societies. *Surveillance & Society* 18(2): 185–192.
- Hesselberth P (2018) Discourses on disconnectivity and the right to disconnect. *New Media & Society* 20(5): 1994–2010.
- Holden C (2001) “Behavioral” addictions: do they exist? *Science* 298: 5544.
- Introna L (2011) The enframing of code: agency, originality, and the plagiarist. *Theory, Culture & Society* 28(6): 113–141.
- Iqbal J, Asghar MZ, Ashraf MA, et al. (2022) Social media networking sites usage and depression among university students during the COVID-19 pandemic: the mediating roles of social anxiety and loneliness. *Social Media + Society* 8(3): 20563051221107633.
- Isin E and Ruppert E (2015) *Being Digital Citizens*. London: Rowman & Littlefield.
- Kardefelt-Winther D (2014) A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. *Computers in Human Behavior* 31: 351–354.
- Kaun A and Treré E (2020) Repression, resistance and lifestyle: charting (dis)connection and activism in times of accelerated capitalism. *Social Movement Studies*: 19(5–6): 697–715.
- Kulwin N (2018) A Former Zuckerberg Speechwriter on What He Doesn’t Understand. *Intelligencer*, 10. April 2018. <https://nymag.com/intelligencer/2018/04/kate-losse-former-zuckerberg-speechwriter-interview.html>.
- Leach A (2018) Iain M. Banks—human, posthuman and beyond human. *ELOPE: English Language Overseas Perspectives and Enquiries* 15(1): 69–81.
- Lemke T (2001) The birth of bio-politics”: Michel Foucault’s lecture at the College de France on neo-liberal governmentality. *Economy and Society* 30(2): 190–207.
- Livingstone S and Helsper E (2008) Parental mediation of children’s internet use. *Journal of Broadcasting & Electronic Media* 52(4): 581–599.
- Mackenzie A (2015) The production of prediction: what does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.
- Mackenzie C and Stoljar N (2000) *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford; New York: Oxford University Press.
- Mann M and Matzner T (2019) Challenging algorithmic profiling: the limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society* 6(2): 20539517198958.

- Matzner T (2019) The human is dead—long live the algorithm! Human-algorithmic ensembles and liberal subjectivity. *Theory, Culture & Society* 36(2): 123–144.
- Matzner T (2022) Algorithms as complementary abstractions. *New Media & Society*. Epub ahead of print 16 February. DOI: 10.1177/14614448221078604.
- Mieczkowski H, Lee AY and Hancock JT (2020) Priming effects of social media use scales on well-being outcomes: the influence of intensity and addiction scales on self-reported depression. *Social Media + Society* 6(4): 2056305120961784.
- Mühlhoff R (2019) Human-aided artificial intelligence: or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society* 22: 1868–1884.
- Nguyen MH, Gruber J, Fuchs J, et al. (2020) Changes in digital communication during the COVID-19 global pandemic: implications for digital inequality and future research. *Social Media + Society* 6(3): 2056305120948255.
- Oshana M (2006) *Personal Autonomy in Society*. London: Routledge.
- Pasquale F (2015) *The Black Box Society. The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
- Reer F, Tang WY and Quandt T (2019) Psychosocial well-being and social media engagement: the mediating roles of social comparison orientation and fear of missing out. *New Media & Society* 21(7): 1486–1505.
- Rieder B (2020) *Engines of Order. A Melancholy of Algorithmic Techniques*. Amsterdam: Amsterdam University Press.
- Rosenblat A and Stark L (2016) Algorithmic labor and information asymmetries: a case study of uber's drivers. *International Journal of Communication* 10: 3758–3784.
- Rössler B (2021) *Autonomy: An Essay on the Life Well Lived*. Cambridge, UK; Medford, MA: Polity.
- Schüll ND (2012) *Addiction by Design: Machine Gambling in Las Vegas*. Princeton, NJ: Princeton University Press.
- Sen A (2004) Capability and well-being. In: Nussbaum M and Sen A (eds) *The Quality of Life*. New York: Routledge, pp. 30–53.
- Shamir R (2008) The age of responsabilization: on market-embedded morality. *Economy and Society* 37(1): 1–19.
- Spada M (2014) An overview of problematic internet use. *Journal of Behavioral Addictions* 39(1): 3–6.
- Susser D, Roessler B and Nissenbaum H (2018) Online manipulation: hidden influences in a digital world. *Georgetown Law Technology Review* 4: 12–45.
- Sutton T (2017) Disconnect to reconnect: the food/technology metaphor in digital detoxing. *First Monday* 22(6). Available at: <https://firstmonday.org/ojs/index.php/fm/article/view/7561>
- Syvertsen T (2020) *Digital Detox. The Politics of Disconnection*. Bingley: Emerald.
- Syvertsen T and Enli G (2019) Digital detox: media resistance and the promise of authenticity. *Convergence: The International Journal of Research into New Media Technology* 26(5–6): 1269–1283.
- Turkle S (2016) *Reclaiming Conversation: The Power of Talk in the Digital Age*. New York: Penguin Books.
- Vaidhyanathan S (2018) *Anti-social Media. How Facebook Disconnects Us and Undermines Democracy*. Oxford: Oxford University Press.
- van Rooij A and Kardefelt-Winther D (2017) Lost in the chaos: flawed literature should not generate new disorders. *Journal of Behavioral Addictions* 6(2): 128–132.
- Vanden Abeele M (2021) Digital wellbeing as a dynamic construct. *Communication Theory* 31(4): 932–955.

- Westlund A (2009) Rethinking relational autonomy. *Hypatia* 24(4): 26–38.
- Williams J (2018) *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press.
- World Health Organization (WHO) (2020) Excessive internet use—developing policies and programs to address a growing problem. Available at: https://www.who.int/substance_abuse/excessive_internet_use/en/ (accessed 11 November 2020).
- Yeung K (2017) “Hypermudge.” Big data as a mode of regulation by design. *Information, Communication & Society* 20(1): 118–136.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. New York: Public Affairs.

Author biographies

Paula Helm is assistant professor in Ethics & Data at University of Amsterdam. Her work is situated at the intersection of STS, Media Studies, and Ethics. The larger goal and mission is to move AI-Ethics from the PR- to the Engineering- and Development-Level. To reach this goal, Paula is collaborating closely with partners from Computer Science. She holds a doctorate of philosophy (summa cum laude). Before starting her current position, Paula was a research associate at the MCTS at the Technical University Munich and the International Center for Ethics in Science at Tübingen University, where she was teaching in the RESET Elite MA Program as well as co-leading work-packages in several national and EU-projects concerned with the creation and ethical advancement of AI-Systems in various domains. Before that and while completing her PhD, she was a research associate at the University of Frankfurt in an interdisciplinary research project dealing with the “Structural Transformations of Privacy”. She was also a visiting scholar at the Center for Surveillance Studies at Queen’s University (invited by David Lyon) and at New York University, where she was involved in the Privacy Research Group (invited by Helen Nissenbaum). Her work has been published among others at *New Media & Society*, *Ethics & Information Technology*, *Addiction Research & Theory*, *Duke University Press*, *Surveillance & Society*, *Ephemera*, *ACM FAccT*, and *AAAI/ACM AIES*. Her research has further been featured in various media outlets including FAZ, Deutschlandfunk, Deutschlandradio Kultur and ZDF.

Tobias Matzner is professor for “Media, Algorithms, and Society” at the Department of Media Studies at Paderborn University in Germany. Tobias Matzner’s work combines theories of (digital) media and technologies with approaches from political philosophy, cultural studies, and social theory. Tobias Matzner’s research has been published in journals such as *Big Data & Society*, *AI & Society*, *Theory, Culture & Society*, *Internet Policy Review*, and *Surveillance & Society*. Tobias Matzner is board member of Big Data & Society and the Co-PI of the DFG-funded Transregio Research Focus Area “Constructing Explainability.” Recently, Tobias Matzner’s work has been focused on the entanglement of technology with concepts of subjectivity, affect, and autonomy—with a particular focus on algorithms and machine learning. Earlier (and related) research has been dealing with normative theories of privacy. In particular, Tobias Matzner is interested in ways of grounding the value of privacy beyond its traditional origins in liberal political thought. Tobias Matzner has studied philosophy and computer science in Karlsruhe, Rome, and Berlin and graduated with a “Diploma” (master’s degree equivalent) in computer science and a PhD in philosophy, both from the Karlsruhe Institute of Technology. Before joining Paderborn University, Tobias Matzner has been working at the International Centre for Ethics in the Sciences and Humanities in Tübingen and at the department of philosophy at the New School for Social Research in New York as a Feodor Lynen Fellow of the Humboldt-Foundation.