Hume's Empiricism versus Kant's Critical Philosophy (in the Times of Artificial Intelligence and the Attention Economy)

The article exposes how one of the most fundamental oppositions in the history of modern philosophy – the opposition between Hume's empiricism and Kant's critical philosophy – is embedded in the current debate on the impact of artificial intelligence (in particular, the algorithmic selection of content) on human society. Hume's empiricism – with its deduction of subjectivity based on a process of habituation – corresponds to the functioning of recommending algorithms, while Kant's idea of autonomous subjectivity corresponds to the ideals underlying today's ethical attempts towards the regulation of artificial intelligence. According to such ethics, the use of empirical data can endanger humans; whereby our attention can be easily caught by sensationalist content and our autonomy replaced by the agency of machinic intelligence. However, as argued in the present article, such ethical positioning also reproduces the gap between the empirical reality and normative principles, which is why transcendental (Kantian) ethics should be supplemented with Hume's immanent and practical reasoning.

Keywords: *Habit, Recommendation Algorithms, Attention Economy, Artificial Intelligence, David Hume, Immanuel Kant*

Author Information
Jernej Kaluža, Social Communication Research Centre, Faculty of Social Sciences, University of Ljubljana, Slovenia; Jernej.Kaluza@fdv.uni-lj.si; https://orcid.org/0000-0003-1781-3713

1. Introduction

The "empirical turn" in the philosophy of technology, contrary to the opinion of some (see Cera 2020), may not imply the abandonment of all the great ontological questions from the history of philosophy. On the contrary, as a review of the vast literature that has followed the rise in the use of artificial intelligence in recent years reveals, it is possible to find deep philosophical problems embedded even in quite concrete technological innovations. An excellent illustrative example of a concrete new technology, the emergence of which has raised several major philosophical questions and ethical dilemmas, may be "AI-powered content recommendation engines" (see, among others, Sorbán 2021; Carlson 2018; Chun 2016; Kant 2020). Since the first studies and popular books on the subject, the algorithmic personalisation of content has sparked heated philosophical debates on various issues, particularly those related to the conceptions of privacy and publicity, subjective autonomy, and social power. Initially, the debate was mainly framed around issues of algorithmic surveillance, the right to privacy, and the potential misuse of the data obtained, which led to a re-actualisation of Foucault's thesis on the panoptic control of society. In parallel, the rise of algorithmisation and datafication has been accompanied by debates on cognitive biases, technological addiction, and political radicalisation (see for example, Williams 2018; Tufekci 2015).

However, the topicality and acuteness of the developments involving algorithms and the use of artificial intelligence have meant that attention has been repeatedly drawn to various empirical incidents that have arisen associated with their use (for example, fake news, hate speech, and political polarisation), while systematic and deeper attempts to philosophically ground the issues have been rarer. In this article, we aim to do just that: to highlight how one of the most fundamental oppositions in the history of modern philosophy, namely the opposition between Hume's empiricism and Kant's critical philosophy, is deeply embedded in today's debates on the impact of artificial intelligence (and in particular the algorithmic personalisation of content) on human society. As further explicated below, Hume's empiricism, with its deduction of subjectivity to the predictability of habit, i.e. to past experiences from which the patterns of future action can be predicted, can be seen to correspond to the big data-based operation of recommendation algorithms. On the other hand, Kant's idea of autonomous, rational, and free subjectivity corresponds to the ideals underlying today's ethical attempts to regulate and constrain the functioning of artificial intelligence.

Focusing on Hume's concept of habit seems appropriate for conceptualisation of the complex interaction between human agency and artificial intelligence. It also corresponds with the focus on other pertinent issues, such as the attention economy, cognitive biases, and mental health (which seem to be replacing the exclusive focus on surveillance as the main political problem related to recommending algorithms in the last few years). As Tanya Kant (2020, 9) argued: "Data tracking does not exist, in and of itself, simply to surveil or track users, but to anticipate them /.../, to 'know' some facet(s) of a user's identity in order to make 'personally' relevant some component of experience on their behalf" (see also Cheney-Lippold 2017). In this

text, I argue that the concept of habit, with all its associated contradictions, could provide a fundamental philosophical grounding, from the perspective of which one can comprehend the broader complexity of the problems associated with algorithmic content distribution (see also Kaluža 2022a; 2022b). Additionally, the concept of habit opens up questions related to the principles of ethical behaviour in times of the attention economy.

This article, based on a conceptual analysis of texts from the history of continental philosophy and contemporary publications in the scientific fields of media and communication studies, is structured as follows: The first part describes how the concept of habit is explicitly or implicitly situated in contemporary critical theory that addresses the problematic effects of algorithmic personalisation. The second part briefly situates the concept of habit within Hume's empiricism and shows how subjectivity can be influenced through habituation (as these ways are also used by contemporary, big data-based predictive analysis of human behaviour). The third part of the article explicates how Kant's response to Hume's philosophy, which seeks to establish a sphere of purity independent of habit, can be understood in a way analogous to the response to a certain part of the ethics debate today, which urges users to confront algorithms with the power of their free will. This kind of ethical positioning, as I argue in the concluding section of the paper, often reproduces the gap between empirical reality and normative principles, which is why I advocate that transcendental (Kantian) ethics should be supplemented with critical empiricism, as a form of immanent ethics derived from concrete empirical situations and which aims to change the state of affairs "from within".

2. The concept of habit in the debate on the algorithmic distribution of content

Algorithmic content selection challenges human agency and is increasingly replacing human curators of content. "AI-powered recommendation systems" have become an increasingly key focus of interest as AI has "recently reached the level of development that makes their functioning comparable to human thinking and allows them to perform tasks requiring (close to) human intelligence" (Sorbán 2021). Such assumptions, however, have not only remained at an abstract level, but have to a certain extent also materialised in a very concrete way in the modern world, as algorithmic content selection is now replacing some of the professions that are closest to the ideals of human curation and decision-making, such as journalists and editors (Carlson 2018; Vobič, Šikonja, and Kalin Golob 2019). Artificial intelligence "which is getting more and more independent from humans through machine learning methods", could, in this context, be considered "a major threat to the human race" as seen in "countless science fiction movies in which AI machines try to take control over mankind" (Hosseinpour 2020, 49–50).

However, the paradox is that this struggle between AI and humans is not caused by the disobedience of AI. On the contrary, it is fuelled by treating AI as a purely subordinate "tool or servant" (see Gyulai in Ujlaki 2021; Totschnig 2019). The superi-

ority of AI for content selection is conditioned precisely by the ability of technology to please the interests of users, while human agency is passivised (individual users) or excluded from the process altogether (professional human content curators, such as journalists). In this trickery, we can see the manifestation of "the tragedy of the master" (Coeckelbergh 2015), where, based on Hegel's dialectic between master and slave, all the work is left to the slave, while the master is completely passivised and becomes alienated from his or her own mastering agency. One could also argue that we are now faced with the realisation of Heidegger's reasoning that human thought – in its own calculating striving for domination – is becoming more and more subordinated to technology (Hosseinpour 2020).

However to date, less attention has been paid to the fact that this subjugation of humans to technology is today primarily conditioned by the habituation of the subject to technology made possible by artificial intelligence (machine learning, neural networks, etc.). Recommendation algorithms are a particularly good illustrative example in this context since they work by exploiting users' previous habits (recorded in the form of data on their history of online activity) to serve the users and recommend future activity. Such service also implies a passivisation of the user. It is this passivisation through repetitive activity, which leads to rigid and predetermined (one might even say programmed) behaviour, that has been one of the key features of the concept of habit that has been so discussed in the history of philosophy (see Malabou 2004; Chun 2016).

The concept of habit is also implicitly present in the debates on various cognitive biases, which are supposed to be encouraged by the algorithmic selection of content. One of the most well-known of these is the so-called confirmation bias, namely, the tendency to confirm one's past attitudes, beliefs, and values, which occurs mainly in the processes of searching for, filtering, and interpreting information. For example, Ariely (2008, xx) showed that there is a certain rigidity not only in the rational, but also in irrational thought, stating: "irrational behaviors of ours are neither random nor senseless. They are systematic, and since we repeat them again and again, predictable". This kind of thesis implies that there must be an instance – and habit seems to be an appropriate concept to name it – that makes the opposite of rational thought not chaotic and unpredictable, but predictably irrational. Habitual decision is the opposite of conscious, thoughtful, and rational decision: It follows certain patterns and can thus also be manipulated. This persuasive manipulation, which could be understood as a "hidden influence that attempts to interface with people's decision-making processes in order to steer them toward the manipulator's ends" does not "persuade people" rationally, but rather seeks to exploit their irrational "vulnerabilities", such as "hidden fears" (Lewandowsky et al. 2020, 23). Seaver (2018), in his analysis, went even further and claimed that the major tendency among today's developers of recommender systems is to "hook" people: He considers recommender systems as traps and compares their functioning with animal trapping based on anthropological theories.

The philosophical dichotomy between rationality and irrationality and the hierarchy between different levels of self-consciousness and self-possession are therefore implicitly ever-present in discussions on algorithmic personalisation. Even if

the creators of technologies "often justify their design decisions by saying they're 'giving users what they want', critical theory doubts that they really give users 'what they want to want'" (Williams 122). With such an argument, the tension between users' "higher" rational goals and users' "lower" irrational passions, which often go against the individual's interest, is confirmed. And it is this tension, which is indirectly connected with Kant's rejection of Hume's reduction of higher instances of mind to habitual functioning (see Deleuze 1967/2008), that is today most directly manifested in clickbait logic or in the logic of nudging, i.e. "influencing action by subtle prompts and signals" (Couldry and Mejias 2019, 140). It is not a coincidence, therefore, that behavioural economics is becoming a crucial part of the broader marketing profession. This development, though, has led to warnings that the potential "integration of neuroscience and marketing" could bypass "the speaking human subject" (Couldry and Mejias 2019, 141).

Kant's autonomy of the human mind has perhaps never been so severely challenged throughout history, since the struggle for attention requires the achievement of ends regardless of the means: "All these types of distractions undermine people's autonomy, by instilling habits and desires that are not voluntarily chosen. In other words, persuasive design steers users' attention towards irrational behaviours" (Voinea et al. 2020, 2354). This division between the higher (reason, rational thought) and lower (habit, instinct, desire) instances of the mind is in contemporary debate often expressed through the vocabulary of the "attention economy", according to which our attention is alienated from us through the use of psychological tricks that count on "our capacity to fool ourselves" (Galef 2021, 7).

The balance between a more goal-oriented reality principle on the one hand, and pleasure principle which aims for the immediate fulfilment of desire on the other hand, seems to be disturbed by our failure to fully distinguish rationality from desire and emotional from intellectual activity. However, this inability can not only influence our individual attention, but it can have crucial effects on broader society too. For example, it has radically changed the news industry by encouraging business models that are based on the monetisation of attention (see Myllylahti 2020; and Slaček Brlek 2018) and it has even led to a broader phenomenon related to the quantification of the public sphere (Splichal 2022). Data behaviourism, which is replacing "disinterested social knowledge", is also affecting the broader sphere of social sciences, in which "social knowledge becomes whatever works to enable private or public action to modulate others' behavior in their own interests" (Couldry and Mejias 2019, 128).

As a response to this development, today's critics of the algorithmically driven dominance of the instantaneous, the shocking, and the clickable, argue for the promotion of the subject's autonomy, morality, and rationality, but understood in a specifically Kantian way. By this we mean (as it will be further developed) Kant's fundamental philosophical orientation towards the establishment of a purified sphere that is separated and independent from anything empirical. On the other hand, the understanding and the design of artificial intelligence follows the logic of a distinctly Humean notion of the mind. Hume's empiricism, in which human subjectivity is formed solely through experience, corresponds with the basic idea

in data harvesting, machine learning, and pattern recognition techniques: The idea that habitual repetition in the past indicates a tendency towards repetition of the same also in the future.

3. Hume's concept of subjectivity and the big-data turn

It seems pertinent to ask: What can philosophical thought from almost 300 years ago and current trends in data science and artificial intelligence development possibly have in common? What is behind Wendy Chun's strange thesis that "Hume is the favored philosopher of Big Data analytics" (Chun 2016, 54)? And why did Hume's thoughts, as they took shape in his works *A Treatise of Human Nature* (1739/2000) and *An Enquiry concerning Human Understanding 1748/*2004), provoke similar moral and epistemological concerns that the rise of algorithms and big data are provoking today?

As a starting point from which answers to these questions will be developed, we could consider the radical nature of Hume's empiricism, which grounds "human nature entirely upon experience" (Hume 1739/2000, 407) and proposes that all our ideas derive from our impressions: "We can never think of anything which we have not seen without us, or felt in our minds" (ibid. 408). To put the same point in a different way: All that we know (or that we believe that we know) is derived from what we have encountered along our past path. "It follows, then, that all reasoning concerning cause and effect are founded on experience, and that all reasonings from experience are founded on the supposition, that the course of nature will continue uniformly the same" (ibid. 410).

That is to say, our current beliefs, expectations, and customs are what they are, therefore, solely because of our previous experiences. This is why Hume's empiricism reveals the fundamental "emptiness of the self" (Chun 2016, 6). As Deleuze (1977/1991, *x*) wrote in his book on Hume: "We are habits, nothing but the habit of saying 'I'". Subjectivity is thus formed through experience, and this implies that users with similar experiences – for example, those who have a similar search history, or who like similar things or live in a certain geographical area – can be compared. As many authors have mentioned also in the context of today's data turn, "your profile is only made meaningful and commodifiable to marketers in and alongside the context of other users' profiles" (Kant 2020, 34–35). It is this kind of "plasticity of subjectivity" (Malabou 2004), which implies that users can be changed by orienting their experience and manipulating their habits, that has guided the conception of many Silicon Valley products (see Eyal 2014).

Subjectivity as understood by Hume is thus predictable: We believe that something will repeat in the future because we are accustomed or habituated to the repetition of the same in the past. Custom or habit is, according to Hume (1739/2000, 411), "the guide of life". It "determines the mind, in all instances, to suppose the future conformable to the past" (ibid. 411). This same proposition presents the grounding of the algorithmisation and big data turn of today; whereby data representing past experiences functions as a basis for a recommendation of the same (or similar) con-

tent in the future. "Algorithmic personalization is premised on the idea that your future preference scan be inferred from your past interactions" (Kant 2020, 37) and that a user's previous online activity is the best predictor of what he or she wants to see next.

Hume's empiricism, in which human subjectivity is formed solely through experience, corresponds with the basic idea behind data harvesting, machine learning, and pattern recognition techniques. Hume's subjectivity is reduced to experience and online subjectivity of today is reduced to data (Cheney-Lippold 2017). This is the reason why data analytics are, according to Chun (2016, 57), usually related to habit: They focus on habitual actions (such as buying), they seek to change habits, and they "replace" causality with correlations between habits. This "process of commercial anticipation" was somehow neglected in the early problematisation of algorithmisation, which focused mostly on the issue of surveillance: the central aim of data harvesting is not just to watch an individual users, "but instead to act on [them], with, or against their experience of the web" (Kant 2020, 9).

4. Kant's answer to Hume's provocation: the autonomy and purity of subjectivity

It is exactly the predictability of subjectivity, which derives from its reductibility to previous experience, that was one of the main provocations of Hume's philosophical gesture as represented in the context of the history of philosophy. Kant's critical project could thus be understood as a conceptual counterpart to Hume's empiricism. As Kant admits, Hume interrupted his "dogmatic slumber" and gave completely different directions to his "researches in the field of speculative philosophy" (Kant 1783/2004, 10). However, to avoid Hume's sceptical conclusions, Kant argued that some concepts do not derive from experience, but from "pure understanding" (Ibid. 10). In general, Kant's "new science", for which Hume gave a "hint" (even if he latter "deposited his ship on the beach [of skepticism]") (ibid. 11–12), aims to re-establish the autonomy of human understanding, reason, and moral judgement.

Kant's project could therefore be understood as a continuation of Hume's empiricism (since it rejects old metaphysical notions of cause, God, and essence), but also as a negation of the empiricist understanding of human nature, which deprives human agency of autonomy and self-determination (see Deleuze 1967/2008). Our behaviours and beliefs namely seem in Hume' world to be defined by experience, by our encounters with the outside world, while Kant re-establishes the sphere of "purity" (the sphere of pure reason, pure judgement, pure understanding), which is defined exactly by its independence from anything empirical. As he wrote in the preface to the first edition of the *Critique of pure reason*, his project presents a critique "of the faculty of reason in general, in respect of all knowledge after which it may strive independently of all experience" (Kant 1781/1996, 8).

It is difficult to pin down exactly why Hume's thoughts were perceived as a scandal, since there is nothing particularly provocative or transgressive in his style, as there is, for example, with Nietzsche or Marx. Nevertheless, the reasons for this

can be found in a specific combination of Hume's implicit scepticism, determinism, and amoralism. Scepticism stems from Hume's rejection of the idea of a necessary connection between cause and effect (Hume 1748/2004, 42–43). This rejection is also manifested in today's data turn, which is based on the principle of correlation instead of on the principle of causality (see Splichal 2022; and Mayer-Schönberger and Cukier 2013). In Hume's world, we have no direct knowledge of the interconnectedness of phenomena. Similarly, one cannot speculate about the causes for patterns that are observed in big data, but can only draw correlations between them. "Unlike earlier modes of governance", David Chandler (2019, 29) argued, "digital governance does not seek to make causal claims" and it "increasingly focus[es] on the 'what is' of the world in its complex and plural emergence".

The consequence of such a functioning of social knowledge, which is "built from unstructured data, drawn directly from the traces left in the flow of everyday life" (Couldry and Mejias 2019, 125) is specific scepticism; whereby "the end product (the knowledge generated) is not explicable in terms of rules /.../ no one – not even the engineers who run the process – can account exactly for how that knowledge was generated" (Ibid.). Hume's abandonment of the necessary link between cause and effect is, besides scepticism, linked also with his associationism, according to which the main ability of the mind is to habitually associate ideas. The sceptical character of such reasoning is derived from its inability to point to a first cause, which could transcend experience. Even Hume's principle of habit as the central concept of his thoughts in this area should not be understood in such a fashion: "By employing that word, we pretend not to have given the ultimate reason of such a propensity. We only point out a principle of human nature /.../ Perhaps, we can push our enquiries no farther, or pretend to give the cause of this cause" (Hume 1748/2004, 32).

According to Hume, even in the case of the movements of one's own body, we have no direct perception of the cause for those movements (our own power or will): "The powers, by which bodies operate, are entirely unknown. We perceive only their sensible qualities" (Hume 1739/2000, 411). The perception of these qualities is based solely on our past experience. This experience is very close to the passivisation of users that is often attributed to algorithmic personalisation. Events come and go in Hume's world the same way content is recommended on YouTube: any subjective activity is, secondarily, transformed into data, which passivises users and deprives them of their own agency. Or, as Eli Pariser argued in his famous work on filter bubbles: "Algorithmic induction can lead to a kind of information determinism, in which our past clickstreams entirely decide our future. If we don't erase our web histories, in other words, we may be doomed to repeat them" (Pariser 2011, 75).

The power of habit, which in Hume's thought represents the only connection between the past and the future, manifests itself not only as a form of epistemological scepticism, according to which reason is dethroned at the same level as belief, but also as machinic determinism; whereby repetition in the past determines the expectations in the future. The guiding principle of human nature in Hume's thought thus corresponds with the core principle of the "intelligence" of machines (see Sterrett 2002). Even though some say that the term "machine learning" is misleading, because machine "does not have real cognition" and "only humans can learn" (see

Coeckelbergh 2020, 83), one can see that Hume's vision of the basic functioning (and the principle of learning) of the human mind is not presented as a fully conscious process. On the contrary, Hume may be the first modern philosopher who does not tie philosophical certainty and the possession of the truth to the concept of self-consciousness, as was the case with Descartes or Locke (see Balibar 2013). For Hume, subjectivity is never more or less (self)conscious, or more or less close to immediate and definite certainty: "everything what enters the mind is perception, it is impossible to be anything different for the feeling /.../ This is why we can be wrong even when we are most intimately conscious" (Hume 1739/2000, 127). Habit functions automatically, without the subject's conscious awareness and control – this functioning is perceived as a "black box" from the perspective of the human subject. Habits are like machine learning: they don't ask why, they simply aim to automate and economise processes and recognise patterns. They "learn" in the sense that they can serve the purpose of being applied to different similar problems in the future.

We can see how this specific combination of Hume's determinism and scepticism also leads to immoralism. From an empiricist perspective, both moral and immoral acts appear to be on the same level: they are all predictable and conditioned by experience. Hume distances his "experience-based science of human nature" from the natural sciences based on experiment (Norton 2000, I14–I17), but nevertheless he preserves the parallelism between both fields: experience replaces the methodological role of the experiment in predicting human behaviour. Our passions, for example, are essentially related to our experiences of pleasure and pain (Ibid. I46–I50), which implies that those sensation define even higher moral feelings: "pleasure is the 'very essence' of virtue, beauty, and wit" (Ibid. I53; see also Génova and Navarro 2018).

Not surprisingly, Kant rejects this kind of understanding of morality, which is dependent only on previous experiences. In Hume's theory, it is (according to Kant) impossible to locate an instance of a subject's autonomous activity, which is why a seemingly moral action is derived from previous experience (and in the strive for happiness) in exactly the same manner as an immoral one. To answer this problem, Kant's moral action is perceived as pure, and therefore as an action that is completely independent from experience and that is dictated only by pure reason. In his *Critique of the Practical Reason* (1788/2002, 34), Kant rejects "the principle of private happiness" and shows how "the law of the pure will – which is free – places the will in a sphere entirely different from the empirical one".

A moral act that follows the law of pure will could, therefore, be understood as a non-habitual act or as an act that resists habit. The ability to restrain, resist, and control habits and impulses has always been at the core of ethics, but Kant may be the first who systematically confronted moral acts (acts that follow a categorical imperative) and acts performed out of habit. A similar form is taken today by appeals on how individual users can limit the power of algorithms over their lives: They must count on their pure will and the power to resist impulses and stimuli, and on the power to break free from the Humean world of empirical determinism, and to autonomously direct their own attention and their own actions. To paraphrase a typical book on the problems of algorithmically induced biases: One needs a moti-

vation to get rid of the motivated reasoning, i.e. the "motivation to see things as they are (not as we wish them to be)" (Galef 2021, 7).

However, as Bergson (2014, 372–373) argued, Kant's categorical imperative with its form of a pure obligation ("you must because you must") itself takes a form that is not far away from the rigidity and repeatability of habit. It is also not completely clear if Kant's way of rejecting Hume's "machinic" understanding of human nature is entirely successful, since it establishes the ideal of morality, which remains completely untouched by any empirical instance. Paradoxically, such "cold" morality, even if (or exactly because) dictated by a pure mind, itself seems somehow robotic and rigid. For if all empirical aspects must be excluded from moral decision-making, then also all sympathy and compassion learned through life events, which are often understood as the foundations of human morality, would seem to be lost too.

5. The discrepancy between Kantian ethics and empirical practice

The question we must ask is whether this kind of "frontal attack" on empiricism (from the Kantian position of non-empirical purity) presents a fruitful approach for considering today's AI ethics, or whether the opposite is true, and – because of its excessive "purity" – such an approach could reproduce the gap between empirical reality and ethical ideals?

As Deleuze remarked, Hume's empiricism is a sort of science fiction in which one has an impression of a strange, foreign world, "seen by other creatures", but also "the presentiment that this world is already ours, and those creatures, ourselves" (Deleuze 2001, 35). This uncanny and strange impression may be the main reason why our common sense often joins the Kantian critique of Hume's empiricism. However, even if Hume's notion often seem like a strange mixture of scepticism, immoralism, and determinism, which is at odds with the common sense, it works in practice. Empiricism is empirically evident: it draws its strength from its capacity for verification and confirmation. "The result is a great conversion of theory to practice" (Ibid. 36), wrote Deleuze. Even if, according to Hume, it is not possible to predict what will happen tomorrow, it is possible to predict what people, according to the experience through which they have been formed, will believe about what will happen tomorrow. The concept of belief should therefore be understood as "a mental state with a certain dynamic role in the production of action" (Bell 2009, 37). It is exactly the ability to predict such beliefs that is today seen at the very core of the "habitual new media" and the machine learning algorithms (see Chun 2016; Kant 2020), which are, as it is empirically evident, actually quite good at predicting the behaviours, beliefs, and the needs of users.

In contrast, the existence of pure reason or pure will is not so empirically self-evident. This is, unfortunately, structurally a necessary result of the fundamental orientation of Kantian critical thought, which establishes a purified sphere that is separated and independent from anything empirical; whereby the more the critical thought is at odds with empiricism and the less power it has over the course of changing the world, the more it becomes invested in its own criticality. Everything it

criticizes is worth criticizing and – because critique is empirically ineffective – more critique is needed. As Krašovec (2018) wrote (in a slightly different context) the basic structural problem of humanistic ethical theory is that it exists institutionally and habitually separated from the concrete technical organisation of the world, that it is not involved in it, and that its basic premise (and at the same time also the reason for its existence) is precisely this separation.

In this context, it sometimes seems that the demands for AI ethics also represent a complete opposition to the very essence of the functioning of AI, which cannot be understood like any other technological tool that can be used and easily controlled, but must, by its own definition, demonstrate a certain autonomy of non-human action; that is, a capacity to learn and to reason. As Héder (2021, 120) argued, the current wave of artificial intelligence ethics guidelines "can be understood as desperate attempts to achieve social control over a technology that appears to be as autonomous as no other", indicating a misunderstanding of the very nature of artificial intelligence, which is inherently "resistant to such control".

Kantian and humanistic predispositions of contemporary AI ethics that emphasise the role of a "human-rights-driven approach" as "the key to building trustworthy AI" (Sorbán 2021) have undoubtedly contributed to the proper addressing of many of the problems posed by this and other such new technologies. However, they are also co-responsible for some of the shortcomings and failures in the regulation of the use of AI. As Vică, Voinea, and Uszkai (2021, 88) argued, "ethics [are] seldom 'pure', that is based solely on supreme or ultimate moral principles". Therefore, Kantian ethical predispositions may not always prove applicable to concrete situations. They can even function as an entrenchment of the status quo. "Cookie consent" notifications may be a good example in this context, since they are – in opposition with the initial intent to control the use of personal data – often "designed to maximise compliance" and treated "as one more persuasive interaction" (Williams 2018, 116).

Kantian foundations could be found, for example, in the approach of the European Union towards data-handling legislation, which it has based on the idea of informed consent to the collection of data. However, such an approach that counts on the user's ability to reject data collection sometimes does not consider that any kind of "data diet" is almost completely impossible in the era of "data-hungry" algorithms. The Kantian vision of the subject's independence from all that is empirical and/or from such data-gathering is nowadays difficult to ensure also because it significantly compromises the user experience. Sometimes the principles in ethical handbooks seem so high-flying that they are quite impossible to apply in reality, which is why some argue that the question of "enforceability" – how ethical principles can "become binding" – "should be taken into account more seriously" (Gyulai and Ujlaki 2021, 30). This divergence from empirical reality, which is also creating a gap between what is realistically possible and what is ethical, must be understood as a structurally necessary result of the excessive moral rigour arising from the categorical character of Kantian ethics.

A problem of the Kantian anti-empiricist position is also that it presupposes that there is something immoral and wrong with empirical data in itself, with the process of learning through experience, or with the habitual automation and simplifica-

tion of repetitive actions. On the contrary, I believe that it would be easy to imagine a situation in which such approaches could be used in a manner that could reduce compulsive and addictive behaviour patterns or human biases. Some such methods – for example, Twitter's warning when a user tries to share content that he or she has not yet read – are already in use. We therefore need to be more specific and empirical in our critique of datafication and algorithmisation. We also need to ask when, how, and why exactly the market logic of the attention economy came to dominate the Internet? Why is data being used mainly for the purposes of attracting users? Why are users – rather than rational citizens who are supposed to participate in public online interactions or as learning subjects who would like to enhance their cognition – being treated as consumers whose needs must be fulfilled?

6. Conclusion: towards the critical empiricism

Let us summarise: The Kantian confrontation with Hume's empiricism is today manifested as a confrontation between the ethical, normative, and regulatory approaches to technological novelties on the one hand, and the unprecedented rise of big data, algoritmisation, and the use of artificial intelligence on the other hand. According to (Kantian) normative ethics, the use of empirical data supposedly alienates humans from their authentic essence on several levels: our attention is stolen and given over to sensationalist content that appeals to the passions rather than to a reasonable contemplation (in a similar manner as Hume's theory devaluates reason at the same level as belief); our morality is threatened by empirical stimuli that seek to exploit our momentary desires (in the same way as Kant's morality is seduced by passions and instincts); our autonomy is replaced by the autonomous and automatic functioning of machinic intelligence (in the same way as there is no space for the concept of autonomous decision in Hume's empiricism). The "continuous tracking of human life" is, therefore, as many argue, "incompatible with the minimal integrity of the self that underlies autonomy and freedom in all their forms" (Couldry and Mejias 2019, xv).

Although in this article I acknowledge the analytical value of this reasoning, as well as the value of the (Kantian) ethics of artificial intelligence on which the regulation of new technologies is often based, I also note there are certain problematic patterns raised by this kind of frontal opposition to empiricism. Such critique is often too general and too abstract – it is non-productive to oppose data, numbers, and quantity in general or to idealise a romantic image of human nature that transcends material and empirical circumstances. This kind of logic is partly responsible for the gap between normative ethical principles and empirical reality. Even if much of the philosophical understanding of technology in the 20th century was based on such ethical principles – from Heidegger's rejection of cybernetics to Adorno's distrust of mass media – today's technologies based on machine learning and big data seem to require an upgrade of their ethical groundings.

This is why I argue in favour of critical empiricism – a theory that must meet two criteria: 1) it must be able identify the specific empirical actors and their interests

that ultimately reproduce data colonialism, communication inequalities, and digital exploitation in the online sphere; 2) it must be able to interfere in the empirical development of new technologies (and not only oppose this in its empirical totality).

Therefore, Hume's empiricism (similar to today's turn to big data and algorithmisation based on the concept of habit) does not have to be completely dismissed. We should not forget that Hume's sceptic thoughts always left open the principled possibility for different and new ideas to form and actions to occur in the future: perhaps the sun will not rise tomorrow, since it is merely a habit of the mind to conclude that tomorrow will be the same as yesterday. And it was exactly this openness to the future and to novelty that inspired Hume's interpreter Gilles Deleuze, who centred his philosophical *oeuvre* (and his practical ethics) around questions of difference, diversity, and novelty. "It is perhaps habit which manages to 'draw' something new from repetition" (DR, 9), he argued. Habit contracts all that is repetitive and identical, and in doing so it exposes the different, deviant, and heterogeneous. However, habit should not be understood only as that conservative power which reproduces the same, but also as a power that can produce novelty. As Malabou (2008, viii) argued, we should bring to light that habit in the history of philosophy does not represent only "addiction (machining repetition)", but also "grace (ease, facility, power)" – "it is one and the same force, one and the same principle".

Allow me a brief subjective opinion to conclude. Perhaps it is exactly the attempts of the attention economy to control our habits that have given rise to the rebellion against it in the form of a desire for a different, more diverse, and less monotonous internet of the future. Those desires are manifested, I believe, also in the formation of many underground internet movements and local communities of practice, such as those presented by the open-source initiatives. And maybe what is missing is not only the transcendental normative principles of (Kantian) ethics but also more support for practical and immanent attempts as part of the struggle for a free internet and for the democratic use of new technologies. There is no doubt that some of the problems with our habituation to new technologies should be solved in a Kantian manner – with a law, regulation, prohibition, or with withdrawal from the empirical data – however, we should also take into consideration the formative power of habit, which can help with the concrete application of ethical principles in empirical reality and produce something new through repetitive practical activities.

References

Ariely, Dan. *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins Publishers, 2008.

Balibar, Etienne (ed.). *Identity and Difference: John Locke and the Invention of Consciousness*. London: Verso, 2013.

Bell, Jeffrey. *Deleuze's Hume: Philosophy, Culture and the Scottish Enlightenment.* Edinburgh: Edinburg University Press, 2009.

Bergson, Henri. *Key Writings*. New York in London: Bloomsbury, 2014.

Carlson, Matt. "Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism." *New Media & Society* 20, no. 5 (2018): 1755–1772. https://doi.org/10.1177/1461444817706684

Cera, Agostino. "Beyond the Empirical Turn: Elements for an Ontology of Engineering." *Információs Társadalom* XX, no. 4 (2020): 74–89. https://dx.doi.org/10.22503/inftars.XX.2020.4.6

Cheney-Lippold, John. *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: NYU Press, 2017.

Chandler, David. "Digital Governance in the Anthropocene: The Rise of the Correlational Machine." In *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*, edited by David Chandler and Cristian Fuchs, 23–42. London: University of Westminster Press, 2019. https://doi.org/10.25969/mediarep/11926

Chun, Wendy Hui Kyong. *Updating to Remain the Same: Habitual New Media*. Cambridge and London: The MIT Press, 2016.

Coeckelbergh, Mark. *AI Ethics*. Cambridge, MA: The MIT Press, 2020.

Coeckelbergh, Mark. "The Tragedy of the Master: Automation, Vulnerability, and Distance." *Ethics and Information Technology* 17, (2015): 219–229. https://doi.org/10.1007/s10676-015-9377-6

Couldry, Nick, and Ulises A. Mejias. *The Cost of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, California: Stanford University Press, 2019.

Deleuze, Gilles. *Kant's Critical Philosophy*. London in New York: Continuum, 1967/2008.

Deleuze, Gilles. *Empiricism and Subjectivity*. New York: Columbia University Press, 1977/1991.

Deleuze, Gilles. *Pure Immanence: Essays on A Life*. New York: Zone Books, 2001.

Eyal, Nir. *Hooked: How to Build Habit-Forming Products*. Harmondsworth: Portfolio Books, 2014.

Galef, Julia. *The Scout Mindset: Why Some People See Things Clearly and Others Don't*. New York: Portfolio/Penguin, 2021.

Génova, Gonzalo, and Ignacio Quintanilla Navarro. "Are Human Beings Humean Robots?" *Journal of Experimental & Theoretical Artificial Intelligence* 30, no. 1 (2018) 177186. https://doi.org/10.1080/0952813X.2017.1409279

Gyulai, Attila, and Anna Ujlaki. "The political AI: A realist account of AI regulation." *Információs Társadalom* XXI, no. 2 (2021): 29–42. https://dx.doi.org/10.22503/inftars.XXI.2021.2.3

Héder, Mihály. "AI and the resurrection of Technological Determinism." *Információs Társadalom* XXI, no. 2 (2021): 119–130. https://dx.doi.org/10.22503/inftars.XXI.2021.2.8

Hosseinpour, Hesam. "Disobedience of AI: Threat or promise." *Információs Társadalom* XX, no. 4 (2020): 48–56. https://dx.doi.org/10.22503/inftars.XX.2020.4.4

Hume, David. *A Treatise of Human nature*. Oxford and New York: Oxford University Press, 1739/2000.

Hume, David. *Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 1748/2004.

Kaluža, Jernej. "Habitual Generation of Filter Bubbles: Why is Algorithmic Personalisation Problematic for the Democratic Public Sphere?" *Javnost - The Public* 29, no. 3 (2022a): 267–283.
https://doi.org/10.1080/13183222.2021.2003052

Kaluža, Jernej. "Far-reaching effects of the filter bubble, the most notorious metaphor in media studies." *AI & Society* (2022b).
https://doi.org/10.1007/s00146-022-01399-x

Kant, Immanuel. *Critique of Pure Reason (Unified Edition)*. Cambridge: Hackett Publishing Company, 1781/1996.

Kant, Immanuel. *Prolegomena to Any Future Metaphysics*. Cambridge: Cambridge University Press, 1783/2004.

Kant, Immanuel. *Critique of the Practical Reason*. Cambridge: Hackett Publishing Company, 1788/2002.

Kant, Tanya. *Making it Personal: Algorithmic Personalization, Identity, and Everyday Life*. New York: Oxford University Press, 2020.

Krašovec Primož. "Spremna Beseda." In *Molekularno rdeče: teorija za dobo antropocena, by* Mackenzie Wark, 311–320. Ljubljana: Sophia, 2018.

Lewandowsky, Stephen, Laura Smillie, David Garcia, Ralph Hertwig, Jim Weatherall, Stefanie Egidy, Ronald E. Robertson, Cailin O'Connor, Anastasia Kozyreva, Philipp Lorenz-Spreen, Yannic Blaschke, and Mark Leiser. *Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Luxembourg: Publications Office of the European Union, 2020.

Malabou, Catherine. *The Future of Hegel: Plasticity, Temporality and Dialectic*. London and New York: Routledge, 2004.

Malabou, Catherine. "Addiction and Grace: Preface to Félix Ravaisson's Of Habit." In *Of Habit,* edited by Ravaisson Félix, vii – xx. New York in London: Continuum, 2008.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.

Myllylahti, Meria. "Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms." *Digital Journalism* 8, no. 5 (2020): 567–575.
https://doi.org/10.1080/21670811.2019.1691926

Norton, David Fate. "Editor's Introduction." In *David Hume: A Treatise of Human nature,*
edited by David Fate Norton and Mary Norton. Oxford and New York: Oxford University Press, 2000.

Totschnig, Wolfhart. "The Problem of Superintelligence: Political, Not Technological." *AI & SOCIETY* 34, no. 4 (2019): 907–920.
https://doi.org/10.1007/s00146-017-0753-0

Tufekci, Zeynep. "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency." *Journal on Telecommunications and High Technology Law* 13 (2015): 203–217.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Random House, 2011.

Seaver, Nick. "Captivating algorithms: Recommender systems as traps." *Journal of Material Culture* 24, no. 4 (2019): 421–436,
https://doi.org/10.1177/1359183518820366

Slaček Brlek, Aleksader Sašo. "Audience Metrics as a Decision-Making Factor in Slovene Online News Organizations." In *Technologies of Labour and the Politics of Contradiction* (book series *Dynamics of Virtual Work*), edited by Bilić Paško, Primorac Jaka, and Bjarki Valtýsson, 211–232. Cham, Switzerland: Palgrave Macmillan, 2018.

Sorbán, Kinga. "Ethical and legal implications of using AI-powered recommendation systems in streaming services." *Információs Társadalom* XXI, no. 2 (2021): 63–82. https://dx.doi.org/10.22503/inftars.XXI.2021.2.5

Splichal, Slavko. *Datafication of Public Opinion and the Public Sphere: How Extraction Replaced Expression of Opinion*. London: Anthem Press, 2022.

Sterrett, Susan G. "Too many instincts: contrasting philosophical views on intelligence in humans and non-humans." *Journal of Experimental & Theoretical Artificial Intelligence* 14, no. 1 (2002): 39–60. https://doi.org/10.1080/09528130210153505

Vică, Constantin, Cristina Voinea, and Radu Uszkai. "The emperor is naked: Moral diplomacies and the ethics of AI." *Információs Társadalom* XXI, no. 2 (2021): 83–96. https://dx.doi.org/10.22503/inftars.XXI.2021.2.6

Vobič, Igor, Marko Robnik Šikonja, and Monika Kalin Golob. "Nazaj v prihodnost: avtomatizacija in preobrazba novinarske epistemologije [Back to the Future: Automation and the Transformation of Journalism Epistemology]." Javnost - The Public 26, Sup. 1 (2019): 41–61. https://doi.org/10.1080/13183222.2019.1696600

Voinea, Cristina, Constantin Vică, Emilian Mihailov, and Julian Savulescu. "The Internet as Cognitive Enhancement." *Science and Engineering Ethics* 26 (2020): 2345–2362. https://doi.org/10.1007/s11948-020-00210-8

Williams, James. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press, 2018.