Alois Joh. Buch

Rethinking Ethics in the Context of Informational Overload and Artificial Intelligences

Being keenly aware that philosophers and social scientists may not have detailed inside knowledge of the technology part of the issue at stake it is anyway advisable to leave that part to real experts – respectfully trusting that they can provide some important clarification and also will contribute to transdisciplinary communication about current state and future prospects of 'Artificial Intelligence' (AI) and of 'informational overload'. In this essay I also do not address some rather interesting aspects exceeding pure technological implications of the topic, like the simple but in some way important question what the term 'overload' may mean in the context of information, and particularly how this is being measured, to what it relates, and what therefore this term '*over*-load' would indicate. Being also aware that in view of the more general debate some basic issues still require thorough examination – not least in regard to the complex and probably mutual relation of 'disruptive innovation' and 'sustaining innovation' in

digitalisation[1], and even with respect to the precise meaning of so-called 'Ethics of Disruption'[2] – I would like instead, based on a phenomenological approach and with a specific social-ethical perspective, to present some challenging questions as well as a few considerations for further discussion.

1. … With Regard to the 'Personal' Level

While Australian researcher Toby Walsh concludes that Artificial Intelligences and humans will be on par by 2062[3] – '*Homo digitalis* will have won' – (thus, substituting us[4], or according to James Lovelock AI 'will even dominate' in the 'novacene'[5]), and German

1 Cf. Markus von Fuchs, Digitalisierung: Innovation und Disruption, in: Sonderpublikation der Handelsblatt Fachmedien (IT Special), https://www.skwschwarz.de/fileadmin/user_upload/Veroeffentlichungen_Dokumente/HBFM_05-2017_Digitalisierung_Innovation_und_Disruption_S4.pdf (January 24, 2020).

2 The term itself seems to be still unclear; respective issues are related to 'information ethics' – cf. Oliver Bendel, Art.: Disruptive Technologien, https://wirtschaftslexikon.gabler.de/definition/disruptive-technologien-54194/version-368845 (March 18, 2020).

3 Cf. Toby Walsh, *Das Jahr, in dem künstliche Intelligenz uns ebenbürtig sein wird* (München: Riva, 2019) (*2062. The world that AI made* [Carlton / Australia: La Trobe University Press, 2018]); it's about the year 2062 for which AI-experts believe AI being on par with humans in regard to intelligence, others expect that for 2220 (ibid., 35); Walsh dares to make his prediction although he is aware of warnings and serious uncertainties in this regard (cf. ibid., 46–50), and despite his critical attitude towards certain projections like so-called 'technological singularity' of AI as a radical turning point (50–71).

4 *Ibid.*, 12 f, 32 f, 70 f – yet, perhaps not simply replacing the homo sapiens (cf. 71), or wiping him out (cf. 78). – Toby Walsh takes a more favourable view on informational ‚overload', e.g. in the context of what he calls 'Global Co-Learning' (ibid, 18 ff), as he does regarding his vision of a 'fair, just, and beautiful' digital future (cf. ibid., 34); however, he raises a number serious ethical issues related to AI, like (end of) labour, robotic war etc.

5 Cf. Mark Siemons, "Können Intelligenzen uns noch retten? Der hundertjährige Gaia-Künstler James Lovelock preist das Anthropozän – und das Novozän, in dem KI uns beherrscht," in: *Frankfurter Allgemeine Zeitung* (FAZ), no. 4, January 26, 2020, 42; the comment refers to James Lovelock / Bryan Appleyard, *Novozän. Das kommende Zeitalter der Hyperintelligenz* {München: C. H. Beck, 2020) (German edition of: „, London: Allen Lane, 2019).

computer scientist Jana Koehler takes the opposite view – namely technology would 'not be equal to the human being'[6] –, another serious protagonist, UK-German brain researcher John Dylan Haynes, adopts a somewhat intermediate position: 'We tend to attribute too much intellectual potential to things'; 'We should get away from a perspective of *man against algorithm*.'[7] Noticing this variety, at first glance we may on a personal level – i.e. in regard to implications for humans as individual persons – perhaps not be overly concerned about developments around AI and AGI.[8] Pragmatically speaking, this could instead simply mean to deal with AI as we had dealt and are used to deal for quite some time already with technological developments, which as always would certainly have to include special attention to the ambivalence of opportunities and risks etc. that belongs *per se* to such developments**.**

However, at this first level, things seem to be quite complicated as well. Even when leaving aside the more metaphysical issue concerning the (legal or philosophical) status of AI compared to human persons,

6 Jana Koehler (translation: A. J. Buch), quoted by: Lisa Hegemann / Meike Laaff, "Welche Funktionen wollen wir auf Maschinen übertragen, welche nicht?", https://www.zeit.de/digital/2019-07/kuenstliche-intelligenz-algorithmus-dfki-jana-koehler (January 25, 2020).

7 Respective reference is made in: "Wie intelligent ist KI?," in: *Christ in der Gegenwart* (CiG) 71 (2019) 546; see also: Elisabeth Gräb-Schmidt / Christian P. Stritzelberger, "Ethische Herausforderungen durch autonome Systeme und Robotik im Bereich der Pflege," in: *Zeitschrift für medizinische Ethik* (ZfmE) 64 (2018) 357–372, esp. 376: With regard to 'robotics' and to AI, it is stressed that "performance of technologies tend to be largely overestimated." (Translation: A. J. Buch).

8 Artificial General Intelligence (AGI) means 'artificial super-intelligence', which Toby Walsh, 44 ff, expects as next step in development, despite a number of still existing problems. – Cf. also terms like 'seed AI' and 'augmented intelligence'.

it is not inconsiderable that AI experts themselves draw our attention to the question, 'which functions we would like to attribute to machines, and which ones not'.[9] Though according to observations of experts, we do not really know how realistic or unrealistic expectations regarding AI are, apparently *decision* becomes crucial here. Obviously this also carries an *ethical* dimension, in particular in respect of an alternative that may not really exist but can be framed as follows: Do we just decide working on Decision-Programming of complex learning or recursive enhancing AI systems – or are we willing to also hand over or to leave decision-making competencies to those systems?[10] More sharply stated, do we, on a personal level, feel comfortable with creating technically the space for significant decision making next to the one based on personal freedom? The underlying intention of mostly ethically dominated discourses about such questions, which according to some AI protagonists, in view of the already existing self-creative potential of AI, may turn out as just expressing a kind of naïve hope, seems to aim at ensuring *humans maintaining control* over sophisticated AI and associated areas like informational overload. No matter whether it proves to be illusory, this intention nonetheless includes the idea of encouraging AI technicians to combine conscientiously technical knowledge with responsibility. At least implicitly, this would require reference to profound ethical criteria. As a general orientation, in accordance with current philosophical concepts of humanity and humanitarian principles (like humanity, impartiality, non-discrimination), and also for reasons of broad acceptance, such criteria should be based on shared foundations of democratic, open and

[9] Lisa Hegemann / Meike Laaff. – (Translation: A. J. Buch).

[10] Depending on how Artificial Intelligences would have to be understood as 'entities' in a qualitative perspective, compared e.g. to humans, one would also face a terminological problem, like in regard to the proper term for respective creative operation, namely 'decision' *or* 'solution'? – As to 'recursive enhancement' see: Toby Walsh, 50–58.

pluralist societies, such as human rights, especially concerning dignity of every person and social justice, which altogether are inspired by history of philosophy and which are aligned with 'Integral Human Development' as well. We will return to this further in this essay.

In short: In awareness of the fundamental uncertainty of any technological upheaval, from an ethical point of view, *no basic objection* to AI would have to be raised within the kind of limited personal approach as chosen here – at least as long as the respective discourse would serve reaching a consensus about knowing what we do, about clarifying what we are willing to do, and about what we should do, and all this with a sound humane orientation in terms of exactly not endangering, but preserving dignified human existence. This kind of *critical yet open approach* to AI, which clearly transcends any simple personal level, can also be derived as a sort of lesson learnt from the multifaceted, ambivalent history of technology in general,[11] and in particular from recent discussions, e.g. about truth-claims in post-truth contexts,[12] about anti-social or socially damaging elements in social media,[13] and about in-human ingredients in human genetic research.[14]

[11] Cf. Alois Joh. Buch, "Technischer Fortschritt und Zukunft der Gesellschaft. Zu einem Grundproblem der Technik-Diskussion," in: *ZKTh* 109 (1987) 48–68.

[12] Cf. Alois Joh. Buch, 'Post-truth' – Challenging Academia to Re-think Truth?,' in: Viktor Poletko / Gregory Arblaster (ed.), *Responding to the Challenges of Post-Truth* (International Institute for Ethics and Contemporary Issues) (Lviv (Ukraine): Ukrainian Catholic University Press, 2019), 30–47.

[13] Cf. Alois Joh. Buch, "About connecting lines between Integral Human Development and Communication," pdf-publication of papers from the IIECI-conference on 'Friendship in the time of Facebook'. Lviv (February-March 2019), https://drive.google.com/file/d/1Au0LCfVpEJgd-n2K5DuSH-Oq1BYtL-gQb/view- June 2019, 5–7.

[14] Cf. e.g. Dietmar Mieth, "The Human Being and the Myth of Progress; or, The Possibilities and Limitations of Finite Freedom," in: Lisa Sowle Cahill, *Genetics, Theology, and Ethics. An Interdisciplinary Conversation* (New York: Crossroad, 2005), 53–73, esp. 61–73.

2. …From an 'Institutional' Point of View

What I call the 'institutional' point of view does not only complement the aforementioned one but in a way suggests turning the perspective: namely to consider digitalisation including AI at *first as a major factor of improving life conditions* and therefore as a means for further humanizing our world in its personal, social and institutional aspects. As part of the above described 'ambivalence' of all technique and technology, we have to notice that digital reality has already augmented, and potential prospects of the digital era will likely contribute to improvement of peoples' living conditions and of development opportunities of social groups, of society as a whole, even to the design of international relations, as well as to coping with global challenges. This means nothing less than another highly influential change in the framework conditions that essentially shape the living environment as a whole. Important examples for this 'institutional impact' of digitalisation as a structural factor can be found in the special field of AI in efficient and sustainable solutions, existing or expected, for trading and supply of goods and services, for (individual and public) transport and travel, for managing climate change, but also for the educational and cultural sector, and – last but not least – for medicine and top specialised medical treatment. In this respect, especially because of the life-serving effects that are inherent in it, special and primary attention must be paid to the right and ability of each person to participate.

However, exactly when recognizing the life-serving impact of digitalisation, we should point out some serious concerns in view of AI and some attached evolution. For illustration purposes only, let me just focus on *one selected topic from the health sector* which I consider rather enlightening, and which may serve to some extent also to outline the *ethical* impact of AI prospects and thus subsequently may foster specific attention. The health sector seems well suited

for exemplifying respective concerns, particularly since in this sector everybody is a potential stakeholder, and since discourses about digitalization in this field have existed for quite some time already – one need only think of issues like 'big data', the 'glass human being', mechanization of medicine, patient's anonymity and autonomy. In more detail, in this context I would like to refer to *geriatric and nursing care*; a bit more accentuated as it is about care for care-dependent people, for seriously diseased people, for terminally ill patients, and for the dying. As is well known, there is a long-lasting debate about improvement of sensitive nursing for these patients and of care in general – some key words of which are, for instance, highly individualized supply, holistic care, human attention, appropriate communication, spiritual care, and wide ranging aspects of palliative care as well. Moreover, this issue is especially relevant, given its urgency due to considerable shortages of doctors, qualified nurses, and of professional social carers – a problem that concerns hospitals, residential nursing homes as well as mobile care in quite a number of countries and societies and which results from a complex web of social, economic, and sometimes ethical and cultural backgrounds.[15]

Quite new opportunities and challenges in this area arise from recent developments in robotic technology in combination with AI,[16] particularly with regard to the upcoming use of various 'intelligent service robotics' in healthcare, which would include physical human-robot interaction.[17] Certainly, such new means may be considered

[15] Cf. Birgit Graf / Barbara Klein, "Robotik in Pflege und Krankenhaus – Einsatzfelder, Produkte und aktuelle Forschungsarbeiten," *ZfmE* 64 (2018) 327–343, esp. 328.

[16] Cf. Elisabeth Gräb-Schmidt / Christian P. Stritzelberger, 358 – the authors point at important convergences between robotics and AI; see also ibid. 363.

[17] For basic information about recent developments and current research see: Alexander Dietrich / Jörn Vogel et. al., "Feinfühlige, interaktive Roboter in

most helpful in view of efficient treatment, most notably in surgery, and of specific care e.g. in isolation wards, quarantine stations, and emergency care (partly combined with 'telepresence technologies').[18] In some environments, it can be considered useful as technical support serving patients' independence and autonomy, particularly for those suffering from specific diseases.[19] However, such means can be viewed quite differently in regard to 'regular' care and in regard to specific nursing for critically ill and dying people. Especially here, robotic technology, though according to experts *for technical reasons* currently still an exception,[20] may raise a number of questions. This applies particularly to care in the final phase of life, though the reduction of workload by help of intelligent technology may always be a strong argument, especially with respect to AI-induced autonomous robots with creative self-adaptation to new tasks in care environments that are facing important staffing problems.[21] Certainly, one could argue that robotic care is more than no care; yet, beyond this, some phenomena of the envisioned future of elderly care and nursing may create ethical issues. Just as one example for this can serve the vision of care in which professional nurses and carers would be replaced by care robots, and in which 'humanoid robots' more and more would mutate from technical means to 'technical colleagues', and moreover, in which they themselves become an active part or even an 'autonomous actor' of interaction with patients. Concerns

Krankenhaus und Pflege: Wo stehen wir und wohin geht die Reise?," in: *ZfmE* 64 (2018) 307–325.

18 Cf. Alexander Dietrich / Jörn Vogel et. al., 315; also: Toby Walsh, 42 f.

19 Cf. Birgit Graf / Barbara Klein, 332f. (with special reference to telepresence robotics, and pointing at potential impact particularly in paediatrics), also 336 f, 339.

20 Cf. Birgit Graf / Barbara Klein, 339 f; Alexander Dietrich / Jörn Vogel et al., 308.

21 Cf. Alexander Dietrich / Jörn Vogel et al., 316.

would certainly not decrease if economic reasons for establishing respective human-robot interaction were taken into consideration.[22]

In ethical terms, we have to notice that a certain intensity of robotic support of patients' autonomy may turn into a moral threat[23] – namely in contrast by endangering the freedom, autonomy, and hence dignity of patients. The issue at stake here is what our intentions in health care are, and what we consider as essential for geriatric care and nursing of patients that is appropriate to human life and to the 'principle of patient autonomy', especially in critical phases of life. Obviously, this requires moral options, choices, and again *decisions*. Such decisions would pertain to questions like: Would we personally, and would the community which is committed to basic values of humanity, in those extremely significant moments of human life like to see AI-governed assistance and care by robots – and if yes, to which extent? Could we think of and agree upon a defined setting of care that would take into consideration both improvement of care *and* compliance with fundamental ethical standards like respect, dignity, autonomy of humans – an ethically framed setting which consequently would imply institutionalizing care also in times of AI in such a way that robotics were *restricted to assist* nursing and care staff, while the latter in turn would be relieved and could focus more on social

22 Cf. Alexander Dietrich / Jörn Vogel et al., 322: "The lack of personnel in healthcare in our society is well known, it is now taken up from the perspective of robotics […]. Because of development leaps […] in recent years and because of progress in voice recognition, navigation, and artificial intelligence most likely robotic technology in healthcare will benefit too." (Translation: A. J. Buch). – Concerning the ethical issue see: Elisabeth Gräb-Schmidt / Christian P. Stritzelberger, 360f.

23 Another challenging ethical question, although not taken up in this article, would be if or to which extent robotics in combination with AI could be understood or would have to be seen as 'autonomous' entities in the ethical sense – cf. Elisabeth Gräb-Schmidt / Christian P. Stritzelberger, 363–366. Attached to this, also serious legal questions are arising – see: Eric Hilgendorf, "Recht und Ethik in der Pflegerobotik – ein Überblick," *ZfmE* 64 (2018) 373–385.

relationships, communication, attentiveness, and accompaniment?[24] Phenomena like socially destructive side effects of social media can be thought provoking when reflecting on so-called future 'social robotics'[25] or 'emotional robotics'[26] in healthcare. Also in practical terms this is not at all a trivial matter, since potential major upheavals in the human design of personal and social life come into view if we include at this point an additional question – namely (again only as an example of much broader and complex problems) whether it is ethically reasonable and desirable to create institutional frameworks of nursing care that tend to establish systems of robotic *virtual visits* that over time would relieve or even exclude people from *personal visits*, assistance and accompaniment of their sick and dying fellow humans.[27] It is remarkable at least, that in a comment on 'experiments with robotics in elderly care and nursing', the author claims to clarify 'what machines can do and for which tasks humans are still indispensable'. And this ideally, before we are in need of a 'Human's day' campaign for getting back into machine-dominated sectors.'[28]

What results from these considerations? In a nutshell, this exemplary case of AI-driven increasing use of robots suggests that within

[24] Cf. Birgit Graf / Barbara Klein, 340.

[25] Cf. Alexander Dietrich / Vogel, Jörn et al., 321; these 'social robotics' – the respective wording seems to be most enlightening – "would primarily serve interaction and communication and would imitate the shape of the human body in order to build trust in humans easier and faster" (Translation: A. J. Buch).

[26] Cf. Birgit Graf / Barbara Klein, 334 f. – See also Tobi Walsh, 101, who expects future machines being 'very likely' emotional ones.

[27] Cf. Birgit Graf / Barbara Klein, 332.

[28] "Männerjobs, Frauenjobs – Menschenjobs?," in: *CiG* 72 (2020), 63. (Perhaps we should "ask instead, which work machines can do – an where human beings still are indispensable. And this is the best thing to ask before we will be in need of a ‚Human's Day' [(…)], in order to regain an foothold in sectors dominated by machines." (Translation: A. J. Buch).)

and beyond healthcare, it may be desirable to recognize and to take up deeper-rooted dimensions of institutionalized digital framework conditions of life, getting to the core of the matter: Though the current debates reveal also fear-laden contexts, the main ethical issue emerging from rational reasoning concerns the *humane character of future care for human beings as persons*. Part of this touches on sound reflection about dealing with most vulnerable persons in general, or even more basic, on whether we would be willing and able to combine and closely link Artificial Intelligence and its technical products with the demanding (as well as essential) goal of real integral human development.

3. … From a Social and Political Perspective

It is no surprise that in different aspects 'decisions' are dominating future developments of digitalization and especially of AI. The importance of 'decisions' is even more evident when looking at our topic from a social and political perspective. As to the impact on social life and on community at large that emanates from AI, from growing informational resources which may be perceived as 'overload', and from accelerating digitalization processes and projects in general, it is not least about the *social and political implications of decision making* – or of avoiding decisions, or even of refusing to take decisions for whatever reason, be it a lack of interest, ignorance, or just lethargy.[29] In any case, the consequences of decision or non-decision in regard to socially and politically desired, undesired, or unacceptable actions and developments as well as in regard to respective

[29] See: Alois Joh. Buch, "Moral Particularism and Individualism – Challenging Reflection on Virtue Ethics," in: Volodymyr Turchynovskyy (ed.), *Ethics in the Global World: Reflections on Civic Virtues* (International Institute for Ethics and Contemporary Issues, ed. Volodymyr Turchynovskyy) (Lviv (Ukraine): Ukrainian Catholic University Press, 2013), 82–116.

commitments in these fields are and will be significant, and they even will become fundamental for shaping life and living together. Therefore agreement about deciding factors include diligent attention to issues important for humane development, like transparency of research, democratic practices, social participation, ethical business principles, defined critical limits – to name but a few.

Therefore discussion, clarification, and decision-making processes about basic options, about framework conditions and ethical standards require a public discourse, not just methodologically.[30] Given the nature and the scope of ethical questions to be addressed in this context, it is a reasonable assumption that this discourse does not pertain simply to a number of individual aspects. The actual challenge consists in exploring and thinking through basic elements that constitute an ethics covering moral principles and norms considered as essential and appropriate to handle accelerating digitalization, and especially AI, according to our human responsibility. Such a fundamental approach seems to make sense, as new problems of human judgement and action require also new ethical thinking, not least due to the specific innovative complexity and potential massive impact of respective technologies. Coping with risks and taking up opportunities arising from such technological developments correlate with the ethical task related thereto becomes particularly obvious if one thinks of the above-mentioned complicated relationship between spaces for decision-making in AI systems on the one hand, and on the other the specific ability and responsibility of human beings for making decisions and for taking action.

However, because of the characteristics of ethical demands at stake, this approach implies nothing less than *rethinking ethics as such*, at least to some extent. This would be in line with AI experts who claim: "Computer technologies require us to rethink our ethi-

[30] Cf. Elisabeth Gräb-Schmidt / Christian P. Stritzelberger, 359.

cal foundations."[31] Such a requirement must be perceived as different from other ethical views, e.g. the kind of stereotype saying that compared to technological progress, ethics always comes too late, or like the more radical pragmatic view according to which actually 'ethics has no chance against technological progress',[32] or even like a concept of 'digital humanism', which insists that still "technical progress is shaped by humans" and which therefore would simply "use digital technologies to expand" humans' abilities "rather than to limit them."[33] In comparison, the task at hand would be in a way modest

[31] Lisa Hegemann / Meike Laaff, ibid. (Translation: A. J. Buch) – See also: Toby Walsh, 35, 104, and in particular 204.

[32] Alain Veuve, "Ethik hat keine Chance: Jede Technologie wird früher oder später genutzt.", https://www.alainveuve.ch/ethik-hat-keine-chance-jede-technologie-wird-frueher-oder-spaeter-genutzt/ (January 24, 2020) – (Translation: A. J. Buch); ibid.: "Und darum haben unsere Wertevorstellungen und unsere momentane Ethik mittel- und langfristig keine Chance gegen den technologischen Fortschritt. Was wir mit neuen Technologien auf einer breiteren Zeitachse gewinnen, ist grösser als das, was wir als Opfer bringen müssten. Und darum werden wir über kurz und lang sämtliche verfügbare Technologie einsetzen." It is quite enlightening how the author, consistent with his basic argument, e.g. the voluntary self-commitment of genetic researchers in the famous 1975 Asilomar conference agreement calls, namely as 'inhibiting and controlling technological progress'.

[33] Julian Nida-Rümelin, "Digital Humanism", in: *Max Planck Research. The Science Magazine of the Max Planck Society* 2.2019, 10–15, 12. – https://www.mpg.de/13790224/W002_Viewpoint_010-015.pdf (March 12, 2020). Though claiming, that "Digital humanism counters both IT and Internet enthusiasts and the apocalyptics" (ibid.), one may consider this kind of 'digital humanism' a bit too optimistic in two aspects – namely taking it for granted that AI development would (just) be "shaped by humans" (ibid.) *and* stressing that "the further key goal of humanism" would (just) focus on "the formation of personality" (15). See also, Julian Nida-Rümelin / Nathalie Weidenfeld, *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*, 4th ed. (München: Piper, 2018); here the author opts for a 'middle way' between 'doom scenarios and hopes for salvation' (11) and underlines that "digital humanism does not transform man into machine as it does not interpret machines as humans. It does maintain peculiarity and abilities of the human being …" (10 f – translation: A.J. Buch).

yet ambitious: namely to notice seriously the ambivalence also of AI and to design ethical fundamentals for a sustainable human civilization in the digital age, with its rapidly changing circumstances and challenges. These ethical fundamentals should relate to enabling real personal and social responsibility in the broader context of respect for human dignity and human rights, which altogether are basic for integral human development. One should note that this endeavour would not aim at moralizing all sectors of life – something that critical observers identify as a means which some protagonists in politics prefer for fighting complexity, and what can be called 'excessive moralizing' or 'over-moralization'; on the contrary, instead of searching for detailed moral rules for everything possible, rethinking ethics would focus on clarification of *basic ethical components*, including viable guidelines and principles for action in the foreseeable AI future. One of these components, and not the least, is *responsibility* in its literal sense: 'to respond to'; this does not only mean looking back to realise responsibility for what has happened, it also implies looking ahead in order to take responsibility for future action. In view of the above, more detailed studies on this would have to include a thorough phenomenology of responsibility in its broadest meaning, and which thus would not exclude its spiritual dimensions – also taking critically up some enlightening insights from earlier phenomenological ethics concerning e.g. 'moral attribution', 'personal accountability', 'phenomenon of conscience', 'value awareness', and 'sense of guilt';[34] and, not least, it would even have to give some thought to an overall 'readiness for responsibility' as a kind of virtue,[35] particularly in times of AI.

[34] Cf. e.g. Nicolai Hartmann, *Ethik*, 4th ed., (Berlin: de Gruyter, 1962), esp. 132–136; 727–731.

[35] See: Alois Joh. Buch, Moral Particularism and Individualism – Challenging reflection on Virtue Ethics, esp. 107–111; id., "Bereitschaft zur Verantwortung. Reflexionen über eine christliche Grund-Tugend," in: Studia Teologiczno-Historyczne Slaska Opolskiego 28 (2008), (Opole / Polen: Uniwersytet Opolski, 2008), 125–139.

Clearly, this would entail efforts in answering some difficult questions. One of them would be what precisely 'dignity of the person' means in the digital age, and what follows from it in regard to AI; another would be what humane shaping of interaction and communication in society should look like, and how freedom of speech and press, transparency of information, and an ever-increasing flood of information ('overload') can ethically be weighed against each other; a third and somewhat broader question would be how 'integral human development' should be designed in an environment of growing AI.[36] Just to sharpen one point of principle in view of potential AI prospects: Will and should the shaping of the world, the shaping of social life, and more generally the decision making power remain in some way exclusive areas and responsibilities of humans – or have they been, are they about to, or will they, or even should they be taken over, albeit only partly, by 'artificial intelligences'? With reference to the history of ideas and specifically to the history of technology, what would this latter transformation imply for future self-awareness and self-conception of humans as persons and moral subjects? And, viewed from the opposite angle, which idea of the world, and which understanding of the human person, would underlie a concept of shaping the world basically by 'self-thinking robots'?[37] Characterizing all this just as "transition

[36] Cf. e.g. Julian Nida-Rümelin / Nathalie Weidenfeld, who claim their concept of 'digital humanism' being both 'technology-friendly and people friendly' (15).

[37] Toby Walsh considers precisely 'consciousness' as basic requirement for enabling potential 'conscious machines' to behave ethically (cf. 94), which consequently would imply serious ethical and juridical problems (cf. 97); similar considerations (in part referring to Isaac Asimov's 'laws') pertain to 'free will' of machines (102 ff) and to 'ethical guidelines' for robots (104 ff). – See also: Julian Nida-Rümelin / Nathalie Weidenfeld, who show themselves rather sceptical in this regard; certain expectations concerning 'strong artificial intelligence' they call 'digitalizing ideology' which eventually would turn out to be a kind of 'modern animism' (19), yet

from deterministic to probabilistic machines"[38] at best only indicates the problem. It is obvious that ethical questions in this context are closely linked to anthropological topics; consequently, the current radical change in digitalisation, not least around AI, should evoke renewed and sound anthropological reflection too. A focal issue of ethical as well as anthropological importance, not to be taken up merely by ethicists, would be whether or how far AI really serves and fosters human development – or whether it rather limits human development, so that part of its 'integral' characteristics would become less important or get lost.[39]

Finally, as to content orientation of such ethical and anthropological rethinking, the specific reference to the concept of 'integral human development' seems to be a promising option, as mentioned before. Just two remarks in this respect: First, some important insight precisely in terms of social and political ethics can be gained from reflection on 'Integral Human Development' as presented in Christian Social Thought and in the Social Teaching of Christian Churches; this particularly applies to 'Caritas in Veritate', the 2009 encyclical by Pope Benedict XVI[40] – a specifically relevant text concerning in-

combined with a totally 'deterministic understanding of the world' (cf. 56), altogether rooted in deeper going background of intellectual history.

38 Julian Nida-Rümelin, 12. – Toby Walsh says quite clearly that it is about AI activities that do not result from programming, and that develop their own creativity (cf. Toby Walsh, 27 f).

39 Cf. Toby Walsh, 204, who even envisages a new 'golden era of philosophy', in particular in regard to Computer-Ethics. – Julian Nida-Rümelin / Nathalie Weidenfeld argue, based on their differentiated approach, that those who opt for a categorical 'indistinguishability between man and computer' would mean 'questioning the foundations both of scientific practice and of human life-world' (28; translation A. J. Buch).

40 Benedict XVI, *Caritas in Veritate (CiV), Encyclical Letter to the bishops, priests and deacons, the men and women religious, the lay faithful and all people of good will on integral human development in charity and truth* (June 29, 2009), http://www.vatican.va/holy_father/benedict_xvi/encyclicals/documents/hf_ben-xvi_enc_20090629_

tegral human development as is already shown by its title, certainly in some aspects to be adjusted to and confronted with contexts of AI.[41] Second, rethinking ethics and its anthropological foundations in regard to the digital era may presumably benefit also from a closer look at what is being called 'ethics of relation', a phenomenologically based concept of ethical thought and moral argumentation from the recent past.[42] It stresses the wide-ranging 'relational' dimension of human existence, of responsible decision and action, and of a prosperous community at all levels, which thus may afford new views and some surprising rationale for contemplation of the way we would like to live together in a more and more digitalised world.

Concluding Remarks

Let me conclude these questions and considerations by pointing at three particularly thoughtful voices, which from different angles may inspire rethinking ethics and anthropology.

The first is a word from Isabella Guanzini, an Italian philosopher and theologian: She starts her critical as well as encouraging

caritas-in-veritate_en.html (17 January 2014). See also: Alois Joh. Buch, Universalism and Diversity. Reflecting on features of globalization – with reference to *Caritas in Veritate,* in: Volodymyr Turchynovskyy / Orysya Bila (ed.), *Ethics and Global Political Theory: the Encyclical Letter Caritas in Veritate and Critical Perspectives on Integral Human Development* (International Institute for Ethics and Contemporary Issues, ed. Volodymyr Turchynovskyy) (Lviv (Ukraine): Ukrainian Catholic University Press, 2016), 26–61.

41 Cf. Vatican Academy for Life, Rome Call for AI Ethics, February 28, 2020, http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_.pdf (18 March 2020); see esp. the chapter on 'Ethics,' and also the "desire […] to promote 'algor-ethics'".

42 Cf. Alois Joh. Buch, "Beziehungsethische Perspektiven der Theologischen Ethik," in: Chittilappilly, Paul-Chummar (Hrsg.) *Horizonte gegenwärtiger Ethik* (= FS Josef Schuster SJ) (Freiburg i. Br.: Herder, 2016), 309–321.

comment with the observation that 'we are over-flooded with information' within a 'digital cultural industry […] that leads to a loss of attentiveness and of fundamental relations […].'[43] 'This indeed does not guarantee humane circumstances […].'[44] Instead, 'in view of our shared vulnerability we all need tenderness', an appropriate 'language […], and a creative communication as human beings […] which brings us together […]'[45]; we are in need of 'new human communities', 'of social networks […] which provide a deeper and sustainable meaning than social media networks do.'[46]

The second is taken from Byung-Chil Han, a Korean-German philosopher, whose argument starts by a provocative thesis: 'The time, in which there was still *the other*, is over.' Han considers this time being replaced by what he calls 'terror of sameness',[47] emerging in a kind of 'formless mass'[48] which also results from 'informational overload';[49] in contrast, in the 'future there may be a new profession, which would be called listener. Being paid for, the listener gives the other a hearing. People go to the listener since there is hardly anyone else really listening. […] Listening means a specific activity. […] It is a sort of giving, a gift.'[50] 'Listening means something com-

43 Guanzini, Isabella, Zärtlichkeit. *Eine Philosophie der sanften Macht* (2017), (München: C. H. Beck, 2019), 63. (translation: A. J. Buch).

44 *Ibid.*, 109.

45 *Ibid.*, 111.

46 *Ibid.*, 124.

47 Han, Byung-Chul, *Die Austreibung des Anderen. Gesellschaft, Wahrnehmung und Kommunikation heute*, 3rd ed. (Frankfurt am Main: S. Fischer, 2018) (Erstausgabe 2016), 7 (Translation here and hereafter: A. J. Buch).

48 *Ibid.*, 9.

49 *Ibid.*,7.

50 *Ibid.*, 93.

pletely different from exchange of information. […] No community can ever develop without […] listening."[51]

The third voice can be heard from an article by an anonymous author in a leading German newspaper. He thinks of a more general vision of a new kind of 'progress with freedom' 'which would pave the way into a future, not yet thought up until now […] By no means an analogue future, that would of course be absurd. But can we imagine a world leaving a choice, at least temporarily, allowing to opt out, to log off for a while, in order to reflect – and this in an environment open for some more experiments in the field of life together, of designing spaces, and of arts? On what, and why, is it still worth to work if Artificial Intelligence in the digital world is assuming what we were being paid for over a long period of time?"[52]

Quoting these voices, one may ask whether one of them expresses a too-critical view that would be in line with well-known, usually a bit pessimistic concerns of philosophers. Which could mean, that the more optimistic, challenging and indeed critical as well as subtle question would be: "Can we truly think better without the digital net?"[53]

51 *Ibid.*, 98.

52 "Fragen an das neue Jahrzehnt. Gelegenheit Funkloch," in: *Frankfurter Allgemeine Zeitung* (FAZ), Nr. 301, December 29, 2019, 11 (translation: A. J. Buch); the German text reads as follows: "Es gälte, eine Idee von Fortschritt zu entwickeln, die in eine bisher noch nicht erdachte Zukunft weist […]. Keineswegs eine analoge Zukunft, das wäre absurd. Aber ist nicht eine Welt vorstellbar, in der es eine Wahl gibt, zeitweise? In der ein Ausklinken, ein Abschalten möglich ist, vorübergehend, um nachzudenken, und das in einer Umgebung, die offen ist auch für andere Experimente: des Zusammenlebens, der Gestaltung von Räumen, der Kunst? Woran lohnt es sich zu arbeiten, wenn Künstliche Intelligenzen in der digitalen Welt tun, wofür wir lange bezahlt wurden?"

53 Fragen an das neue Jahrzehnt, ibid: "Und denkt es sich wirklich besser ohne Netz?" (translation: A. J. Buch).

Of course, one must not answer this question, which in a way is a radical one. However, there are certainly sufficient arguments to give time and energy for reflection about anthropological foundations as well as about moral decision and action in a digital world with its artificial intelligences. This can be seen as an important basic step in assuming *responsibility* in its proper sense, with both ethical and practical intent, i.e. trying to *respond* to sincere questions and challenges precisely for the sake of indispensable *critical discernment* – namely carefully weighing up the relationship between technical feasibility and moral acceptance. According to experts, the question in regard to AI is not whether we can do it; the question is what exactly we *want* to do, what we *should* do, and *why*. What may sound like a simple question upon closer look will require major efforts in pointing out new opportunities to foster the ability and willingness to take responsibility, to promote attention for formation of conscience,[54] and thus to develop ethical thought and moral competence that could contribute to responsibly dealing with AI without, at least, contradiction to Integral Human Development.

[54] Cf. Alois Joh. Buch, "Vergewisserung des Gewissens. Zu Bedeutung und Deutung des sittlichen Urphänomens," in: J. Schmidt et. al. (ed.), *Mitdenken über Gott und den Menschen* (= FS Jörg Splett), (Münster: Lit, 2001,) 121–135; also id., "Gewissensentscheidung im Kontext von Prinzipienethik und Kasuistik," in: Bormann, Franz-Josef Wetzstein, Verena (ed.), *Gewissen. Dimensionen eines Grundbegriffs medizinischer Ethik* (= FS Eberhard Schockenhoff) (Berlin: de Gruyter, 2014), 283–309.

Peter McCormick

AI and Ethical Responsibility[1]

"a new spirit… a new heart" (*Ezek*. 36:26)

The intelligence of persons – the human capacity to know, to comprehend, to understand, and to judge – remain today unsatisfactorily explained.[2] One significant consequence is the still-widespread confusion between machine intelligence and human

[1] This text is a revised version of an invited paper presented in shorter form at the International Institute for Ethics and Contemporary Issues of the Ukrainian Catholic University's Second Annual International Conference Series on Integral Human Development in the Digital Age on the particular theme, "Informational Overload, Artificial Intelligence, and Responsibility," held at the Ukrainian Catholic University in Lviv from 26 to 28 February 2020. My thanks to Dean V. Turchynovskyy for his kind and generous invitation and to participants for their constructive comments and criticisms. Please note that more than the usual number of references are included for the interests of advanced students. pjmccormick@gmx.com.

[2] See for example *Science Advances* (14 February 2020) and D. Drenckham and J. Farago, "*L'IA, super-physicienne?*" *Le Monde: Science et Médecine,* 19 February 2020, p. 7. See also B. Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (Cambridge, MA: MIT Press, 2020), P. Bartolomeo, *La pensée*