ETICA & POLITICA
ETHICS & POLITICS

XXVI, 2024, 1

www.units.it/etica

Edizioni
Università
di Trieste

ISSN 1825-5167

Etica & Politica / Ethics & Politics, XXVI, 2024, 1

M o n o g r a p h i c a

DISCOURSE ETHICS TO THE TEST OF ICTs

S y m p o s i u m

Lucio Cortella, *L'*ethos *del riconoscimento*, Laterza, Roma-Bari 2023

V a r i a

Monographica

DISCOURSE ETHICS TO THE TEST OF ICTs

Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 7-12
ISSN: 1825-5167

DISCOURSE ETHICS TO THE TEST OF ICTS
GUEST EDITOR'S PREFACE

CARLO CROSATO
Dipartimento di Lettere, Filosofia, Comunicazione
Università degli Studi di Bergamo
car.crosato@gmail.com

ABSTRACT
Contemporary technologies renew and broaden the definition of communication far beyond the dichotomy of strategy-agreement, utility and the establishment of a common horizon of mutual understanding. Today, the relational pattern itself is made up of communication and information mediated by technologies. The living habitat of the human is not only the dimension of survival, but also a natural and cultural, social, urban landscape structured not for communicative exchange, but by the communication itself. In these profound transformations of human history, the relationship that human beings have with the world, with other subjects and with themselves is involved. The very meaning of ethics is therefore in question, since the living environment in which human beings constitute themselves and act has changed. In particular, one of the most interesting ethical proposal of the twentieth century, the discourse ethics by Karl-Otto Apel and Jürgen Habermas, must be rethought, since the communicative environment has deeply changed.

KEYWORDS
ICTs, technology, discourse ethics, Apel, Habermas

1. Tackling the issue of the intersubjective relation today no longer means, as it has since Husserl, focusing on the ontological existence of the relation, but rather questioning its ethical tenor This is especially the case given the deep penetration of the technological apparatus into communication through the "information and communication technologies". Contemporary technologies renew and broaden the definition of communication far beyond the dichotomy of strategy-agreement, utility and the establishment of a common horizon of mutual understanding. Today, the relational pattern itself is made up of communication and information mediated by technologies. The living habitat of the human is not only the dimension of survival, but also a natural and cultural, social, urban landscape structured not for communicative exchange, but by the communication itself.

The novelty of new technologies can be understood as such in the light of their intrinsic disruptiveness and their capacity to break with the previous paradigm of

human-environment relationship. We are witnessing a geometric progression of the speed of technological developments as such. Thus, we have to take as our point of observation the interest in our relationship with new technologies, in the way they produce a change in our mentality and customs. With regard to new information and communication technologies, i.e. all those technologies that retrieve, store, process and transmit data, Luciano Floridi recently spoke of a real revolution, the fourth. First came Copernicus, who removed us from the centre of the Universe, where we had been placed by a creator God: the exploration of space that continues to this day has given new meaning to our very lives as humans on a small and fragile planet. Then came Darwin, who showed that every living species is the result of an evolution over time from common ancestors through a process of natural selection: removed from the centre of the biological world, the only consolation remained the fact that we are important in other respects and play a central role in other contexts, such as our mental life . The third revolution was sparked by Freud's psychoanalysis, with the discovery that much of our inner life takes place in the dark, that we are not masters in our own home. A fourth revolution found a new point of attack in the further breeding ground of pride, within which humans have entrenched themselves. We could still believe that our special place in the Universe had nothing to do with astronomy, biology or the transparency of the mind, but resided in our superior capacities for thought.

The nature of human intelligence is vague, elusive, and this would have undermined the scope of this last line of defence based on the presumption of being better than animals and of being at the centre of the infosphere, with no other earthly creature occupying that place. It was Pascal himself, who had identified human thought as our dignity, undermined the foundations of this line of defence by building an arithmetic machine, the Pascalina, in the 17th century. It was an instrument capable of influencing the history of calculators and surprising the inventor of the binary number system, Leibniz. A few years later, Hobbes defines reason as the calculation (addition and subtraction) of the result of general names connected with each other, for the purpose of fixing and expressing our thoughts. to think is to reason; to reason is to do maths; and to do maths is what Pascalina already knew how to do. Here are the first germs of the fourth revolution which, through generations of Pascalines and calculators, have removed us from our centred role as the sole intelligent agent in the infosphere. Alan Turing deposed us from the privileged and exclusive position we had in the realm of logical reasoning, the ability to process information and act intelligently.

The new information and communication technologies, by bending the original meaning of *techne*, affect human ability to build its own environment, to give itself a world.

2. Not only are the natural and artificial dimensions becoming increasingly blurred. We are witnessing a quantitative and qualitative increase in the power of technology. The quantitative novelty can be seen in the extent to which the power deployed today is not even comparable to that of previous technological apparatuses. The qualitative novelty consists in the fact that this human technological power also ends up turning against the human itself, retroacting on it, and thus either amplifying its action out of all proportion or, on the contrary, destroying itself. The more the technological potential grows, through the implementation of certain processes and the belief that they can always be regulated, the less it can be truly governed, both individually and collectively.

In these profound transformations of human history, the relationship that human beings have with the world, with other subjects and with themselves is involved. The very meaning of ethics is therefore in question, since the living environment in which human beings constitute themselves and act has changed. In particular, one of the most interesting ethical proposals of the twentieth century, the discourse ethics by Karl-Otto Apel and Jürgen Habermas, must be rethought, since the communicative environment has deeply changed. The transformation of the conditions for the existence and realisation of communication invites to ask whether the universal conditions that Apel and Habermas chose as principles of discourse are sustainable and able to be defended. We need to return to *Diskurethik* to reflect on whether it still has critical potential in a social environment and political context so heavily penetrated by information and communication technologies. We have to rethink these issues in a historical horizon, moreover, in which technological devices are active participants in communication as transmitters or receivers of messages. We live in a reality in which ethics get into touch with the power of technology and must also contemplate the existence of realities capable of learning and of mimicking the human in its transcendental dimensions.

This raises a series of major questions for contemporary thought. How does communicative action change in contact with information and communication technologies? Are there normative principles intrinsic to the new communicative dimension? Which ones? How does the search for a dimension of co-responsibility, justice, and solidarity change in a communication that is so mediatised and mostly aimed at strategy? How is the communication community transformed and how do we regulate "communication" with non-human entities? What does the entry of new technologies into the relational horizon represent for ethics, law, and politics?

The community of communicants, which Apel defined as unlimited, has unexpectedly actually expanded beyond any conceivable boundary, involving non-intentional agents: this calls for a reflection on the moral principles structuring discursive action, whose groundedness and universality are called into question in

the face of the profound penetration of a heterogeneous logic, that of the technologies we use.

Apel and Habermas both favoured the search for agreement over strategic elements. Yet the mediatisation of public communication, and its circulation on virtual networks promising a broadening of the debate, means that the latter now play a central role in public discourse. The circulation of communication on virtual networks, the opening up of social platforms, and the accumulation of huge amounts of information in reservoirs called "big data" engage right in the regulation of inter- and trans-national dimensions that no longer refer to political communities clearly defined by the liberal-democratic State as those to which the *Diskursethik* referred. New organs, formal and informal, are intertwined in the governance of processes that go beyond the domain of national parliaments. Political and juridical philosophy is called to work on the ways deliberative itineraries change.

3. In the first contribution, Žarko Paić notes that communication takes place primarily as a technical process of dialogue and discourse between networked machines-body computers and mobile smartphones. Thus, cultural-social processes are considered from the point of view of managing, regulating and monitoring the environment in which such a process takes place. This implies the transition from content analysis to form analysis in media theory, which corresponds to the transition from text analysis to the analysis of the language of communication or cultural techniques of communication. Cultural techniques become technologies with their aesthetic matrix of communication in a complex environment. We have to study the relationship between the technosphere and the biosphere. In fact, the true subject of communication in the digital age is not society in its complex relations of mediation of needs, interests and desires, but the technosphere itself as the entropy of all social relations in general. In the place of society comes communication, in the place of the paradigm of the text, we have the visual network, and in the place of the material form of the work in reality and the material sign of the event whose symbolic value constructs reality as entropy comes emergent autopoietic system.

The network enables communication to have the semblance of the immediacy of touching the Other through the game of proximity and distancing of those who sacrifice the freedom of individuality in exchange for solitude. When there is no longer either privacy or publicness in the classic model of the liberal idea of freedom, digital nomads strive to create a semblance of intimacy between virtual walls. The perfect apocalyptic utopia of digital communication represents a machine that produces the event of artificial life from the very essence of the technosphere. Can this dizzying speed in the videology of digital appearance not only be controlled but also ethically curb the uncanny condition of communication

transgression where all notions of social stability have become obsolete and, in their place, has come to the entropy of the non-human?

In his contribution, Paolo Capriati poses the problem of the ethical subject of communication in the technological environment in which we are immersed. And to pose this issue, he questions the nature of the subject in Habermasian discourse ethics: can machines play the role of the subject of communication assigned by Habermas' ethics? To fulfill this function, interlocutors must be able to understand each other and must have legitimate interests. Are machines really capable of understanding? And do they have interests? After a series of arguments, Capriati hypothesizes that machines can enter the communicative exchange. But communication would not be linguistic in the strict sense, and this seems to put Habermas' discourse ethics offside. However, Capriati notes how the problems that human-machine communication confronts us are not at all dissimilar to those intercepted by the German philosopher's reflection.

Fabio Mazzocchio compares the philosophical proposal of Karl-Otto Apel and that of Jürgen Habermas, finding divergent details regarding the notion of truth within a consensualist paradigm, the relationship between universal and contextual aspects, and the question of justification from an epistemic perspective. Once the two philosophical itineraries have been specified, the author undertakes to put them to the test in a context marked by the dematerialisation of relations and communication between human and non-human entities.

Although starting from a similar perspective on the transformation of Kantianism through the theory of communicative action, Apel and Habermas structure different models in terms of reference to the historical and transcendental dimensions. The limitlessness of the community remains problematically linked to an underlying optimism of a Kantian nature. The renewal in Habermas, in recent decades, of the theme of recognition and its dialectical substratum, makes it possible to read the distorting elements of communicative processes more adequately than in Apelian formalisation, and shifts the focus to the material and social conditions that determine the systemic level. A communicative conception of the human condition leads to the acknowledgement of the different possibilities of access to the linguistic game. The grammar of recognition is, from Mazzocchio's perspective, the very heart of discourse ethics and, at the same time, of communicative anthropology. Recognising ourselves as bearers of an openness to otherness allows us to rediscover the priority of the "we" over the "I" and the interpersonal constitution of subjectivity.

Explaining the possible link between the dynamics of communication and recognisable authenticity, the author proposes a hermeneutics of today's human condition as marked by flexible aggregative forms, characterised by extemporaneity. The fragile communities born with social media have become light aggregative forms that bring together lonely individuals in search of relationships that satisfy the

need to be recognised, but in an occasional way and mostly free from a commitment to stable bonds. The social horizon seems to be heading towards a deconstruction of the places that unite us. Virtual communication seems to be affected by chatter, dispersion and, behind an apparent existential empathy, extraneousness. This, at the same time, supports and confirms how today, in everyday human relations, a boundless sense of freedom unravels, which envisages ties accepted only, to a large extent, if they are functional to one's own benefit, hence the manifestation of elements of fragility on the community level, which thus exposes itself to the risk of communicative inauthenticity.

Finally, Paolo Monti deals with the increasingly effectiveness of AI writing systems. They produce original texts based on simple inputs about topic and style, through a training process that constantly engages with fragments of public discourse as found in internet webpages, books, and articles. Their outcomes are often indistinguishable from those of human writers.

In addition to many problematic uses of AI technology, such as deep fakes, the presence of AI-generated discourse in the public sphere is increasingly widespread and raises serious normative issues. Public communication can be produced by non-human authors and penetrate democratic deliberative itineraries.

These issues are addressed by Habermas on two occasions. First, he points out that the moral status of artificially-made subjects is problematic because of their structurally unequal position in public discourse. When the nature of some participants has been artificially pre-determined by the intentions of others, non-peer relationships within the community of communicants become inevitable. These structurally non-horizontal relations are exclusionary in the case of genetically modified humans but also pose a problem with the inclusion of AIs without a personal status in the public sphere. Secondly, Habermas argues that the social formation of the person through the practice of exchanging reasons with peers seems irreducible to the naturalistic understanding of the mind as a computer. Unlike the case of human speakers, AI participation in human discursive practices does not lead – so far – to the formation of intentional and responsible agents.

In light of these insights, Monti suggests an ethical assessment of the moral status of AI writing systems that acknowledges them as a special kind of co-participants in human discursive practices. AIs writers are not moral or epistemic peers with humans but can partake in the same public conversation as co-authors. Their contribution is not merely instrumental. However, it is through the agency of human members of the linguistic community that their contribution acquires full intentionality and can be construed as a form of communicative agency.

 Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 13-32
ISSN: 1825-5167

VIDEOLOGY AND DIGITAL APPEARANCE
CAN COMMUNICATION STILL BE ETHICALLY RESTRAINED?

ŽARKO PAIĆ
University of Zagreb
zarko.paic@zg.t-com.hr

ABSTRACT
The problem with open communication today does not stem from the condition of post-totalitarian censorship of information, but from the fact that every technological advance with so-called social networks always means a breakthrough of ethical-political boundaries in understanding the Other.
If information constitutes the essence of cybernetics, then its acceleration, dissemination, and storage aim beyond the limits of communicative irrationality, to twist Habermas' famous formulation about public consensus as the basis of liberal democracy. Instead of a metaphysical grand narrative about the rule of the principle of identity and a sufficient reason for explaining phenomena in the world, we encounter a cybernetic turn.
The videological turn indicates the breaking of ethical boundaries. Instead of restricting freedoms, it is much more significant to go beyond the borders of the global society of control as the first and last station on the way to a new technological singularity.

KEYWORDS
Videological turn, digital appearence, communication, technological singularity, cybernetic turn, ethics.

INTRODUCTION

Communication assumes the exchange of "aesthetic" (sensory) and "cognitive" (mental) energy in the process of becoming an ever-new identity of a fluid nature. That is why all technical devices of the digital age are aesthetically designed objects adapted to the "style" of the fluid nature of human mobility, the changeability of its position in space, and the experience of the telepresence of interplanetary nomads. All this happens in the triad of terms of communication, interaction and system. It should be clear that these concepts of cybernetics and systems theory simultaneously connect "nature" and "cultures" or "societies". Since communication takes place primarily as a technical process of dialogue and discourse between networked machines-body computers and mobile smartphones, cultural-social processes are considered from the point of view of managing, regulating and monitoring the environment in which such a process takes place.

The transition from content analysis to form analysis in media theory corresponds to the transition from text analysis to the analysis of the language of communication or, simply put, cultural techniques of communication. This turn from text to visuality corresponds, of course, to what has been called a paradigm shift in philosophy and the social humanities since the 1960s, when semiotics, semiology, and philosophies of language replace traditional metaphysical disciplines. Cultural techniques become technologies with their aesthetic matrix of communication in a complex environment. We can call them the means/purpose of communication. In this respect, the history of cultural techniques corresponds to the history of media as a linear series of inventions and practical applications of "tools" and "machines" of communication. But the problem is that media history presupposes an understanding of the relationship between the technosphere and the biosphere. Cognitive processes are self-referential. If in this theory it seems that Berkeley's solipsism *esse est percipi* has an obvious ontological advantage, then it might be truly a modernized way of criticizing the transcendental metaphysics of consciousness. The true subject of communication in the digital age is therefore not society in its complex relations of mediation of needs, interests and desires, but the technosphere itself as the entropy of all social relations in general. In the place of society comes communication, in the place of the paradigm of the text, we have the visual network, and in the place of the material form of the work in reality and the material sign of the event whose symbolic value constructs reality as entropy comes emergent autopoietic system.

Who communicates – humans or machines? The network enables communication to have the semblance of the immediacy of touching the Other through the game of proximity and distancing of those who sacrifice the freedom of individuality in exchange for solitude. When there is no longer either privacy or publicness in the classic model of the liberal idea of freedom, digital nomads strive to create a semblance of intimacy between virtual walls. In the posthuman condition of immateriality, the entropy of global capitalism itself produces its Others (planetary enemies) and constructs the event of disintegration in a constantly new staging of non-place. The perfect apocalyptic utopia of digital communication represents a machine that produces the event of artificial life from the very essence of the technosphere. (Paić 2021, 2022a, Paić 2022b)

How to approach this turn in the very essence of the paradigm shift of art as a true communication between worlds in which the concepts of nonlinearity, emergence, interactivity, self-organization, creativity and virtuality are constantly found in the game of transformations, where there is no longer a difference between "natural" and "human", between technoscience and art, as Flusser still announced, and where, in the end, reigns the chaotic order of relations between subjects/actors of events that can no longer be predicted by the classical methods of modern science? Can this dizzying speed in the *videology of digital appearance* not only be

controlled but also ethically curb the uncanny condition of communication transgression where all notions of social stability have become obsolete, and in their place have come to the entropy of the non-human?[1] From the very beginning of the critique of metaphysics in the first half of the 20th century, Wittgenstein and Heidegger presented two different paths in the philosophical understanding of the relationship between thought, speech and things. For Wittgenstein, language denotes a universal instrument. *Language games* are constructed with it. (Wittgenstein 1998) Reality is a state of "things" that consists of facts and statements about the state of "things". So, a language cannot be a materialized "thing" but might be a tool. Reality is constructed and understood with it. For Heidegger, language means the "telling" of the openness of events (*Ereignis*). (Heidegger 1989) Being and time are acquired in the event. Language as the telling goes beyond the horizon of the scientific and technical construction of the world only if it is addressed to the original telling of the fourfoldness of mortals and immortals, Earth and Heaven. In the modern age, language is transformed into a technical system so that "the world as an image of the world" could function in a mathematical-logical project. For Heidegger, language becomes synonymous with a thought in contrast to the epochal accident of the technical period in which everything becomes information, energy and mass. The essence of technology lies in *Ge-stell*. (Heidegger 2000) Language as telling for Heidegger can never be reduced to pragmatics. But speaking of a "thing" suggests that such a structure was already prepared in the modern period of history.

Is there "reality" in virtual space-time without language as a "thing"? If models and symbolic codes generated "reality", then the world in which technoculture affects all areas of life has long since lost its supra-mundaneness and intra-mundaneness. The world is exactly what Fernando Pessoa predicted in his futuristic song – a machine.

[1] With the *videological turn*, I refer to the entire paradigm shift from language and text to image and visuality. Just as in the 1990s, within the framework of visual studies and image science (*Bildwissenschaft*), the centre of gravity in understanding the essence of art and culture was shifted due to the interactivity of new digital media from iconology to the epistemology of visualization, so also in the field of society and politics, we can witness about extremely a complex turn in the notion of political communication. Mass visual media such as Facebook, Instagram and Twitter replace the uniformity of the television image, which does not have the possibility of creating an event in terms of the construction of a polemical situation as offered by interactive media. Therefore, it is completely wrong to reduce the *videological turn* to technological determinism, according to which social values and ideological-political conflict in the political space as the performance of power, passion and influence are imprinted in the consciousness of the masses as a nation, only because for the first time the object of politics has the possibility to become a subject under the condition of virtual participation in the political process. However, no fundamental political turn in the sense of the creation of immediate democracy as desired, quite naively, by the various currents of cyber-anarchism and cyber-cosmopolitics of the new left, took place. Instead, it is better to talk about the apparent duality of digital democracy, i.e. about one and the same model of multilateral communication between subjects/actors in which all cultural wars between ideologically opposed parties, left and right, centrists and various coalitions, are merely being moved into the virtual underground. (Paić 2008)

The desire for "another world" becomes serious with the help of the imaginary-symbolic construction of the world as a universal machine. Life did not lose its irreducibility in that cybernetic model. Quite the contrary, life has become an artificial accident of bio-digital reality as the only true reality. The fundamental problem of the language of new media, as we will see below by analyzing Lev Manovich's statements, becomes the "realization" of language as a "necessary possibility" for the functioning of technoculture. (Manovich 2001) Namely, for there to be any possibility of communication in the mediated immediacy of telematic presence, the new media language must necessarily be "realized". It must take over the symbols from the cybernetic system of world management and communication. Such language becomes the result of techno-cultural reality as a wired reality. We are not objects of such a universal language/thing. We are cyborgized beings of proximity and distance in the universal immersion of our desires in the endless sea of eternal actuality.

1. IMAGE SCIENCE, MEDIOLOGY, AND NEW PHILOSOPHY OF MEDIA

New media as new information and communication technologies, new technoculture and new aesthetics/ideology of the digital age opened up the problem of a different scientific insight into their complexity. Just as due to the change in the subject of research, the natural sciences are still only conditionally separated from the cultural sciences, so also with the emergence of new technology, social relations, structures and ways of articulating cultural orders of meaning has changed a lot. Therefore, an insufficiently firm and convincing answer to where the theory of new media belongs does not affect only the research of new media reality and practice in arts, design, fashion, and everyday life. The problem is how to positively establish the theory of new media in the newly created environment of the so-called cultural sciences (*Kulturwissenschaften*). There are still discussions about that. One of the orientations in building a general theory of image science (*Bildwissenschaft*) considers that mediology should be a new philosophy of media within the cultural sciences (Mersch, 2006, Paić 2021a, Paić 2022a). But it might be emphasized that the term "philosophy" is understood here in the historical sequence after Wittgenstein's philosophical grammar and Heiddeger's destruction of traditional ontology. In other words, the use of the term philosophy is determined by a methodological bracket against the historical suspension of metaphysics in the social and natural sciences. To that extent, one cannot directly oppose such an attitude. Let's just say that doubting any initiation of philosophy in the expansion of its "subjects" of research is nothing more than agreeing to the position of "eternal philosophy" which only historically changes its characters in which Being, beings and essence of man appear. In addition to media theory, as stated there are also

media sociology, media economics, media archaeology, media aesthetics, media pedagogy and media philosophy. Such division is noticeable at several levels of the interdisciplinary project of image science (Sachs-Hombach, 2006).

Another orientation considers media science as something that goes beyond the current ontological and epistemological foundations of philosophy (Mersch, 2006). Instead of the terms of speculative-dialectical philosophy of Hegel, with Heidegger, and especially with the post-structuralist turn to the problem of language, structures, decentralized subject, rhizome, etc. (Lacan, Foucault, Derrida, Deleuze/Guattari), a new direction of dissolution of philosophy in media science was paved. The concepts that such science uses undoubtedly arose from a philosophical reckoning with tradition. Derrida's deconstruction of metaphysics is an exemplary case of this. Another impetus for the constitution of media science came from various mathematical, cybernetic and information theories of communication (Shanon/Weaver, Moles, Flusser). According to this point of view, philosophy can deal with the media only as part of a complete interdisciplinary science. However, as an independent discipline, media philosophy seems to be a failure, as is the inflation of various "new" philosophies (drama, fashion, art, etc.) that is still present today. And for this second orientation, although it opens the problem of the relationship between philosophy and media science more appropriate to the matter itself, the same can be said for the first. Its difficulty is evident in the fact that it starts from the self-evident "fact" that the cultural sciences (*Kulturwissenschaften)* have replaced the spiritual or humanistic. The result of the transition of spirit into culture denotes the end of philosophy as metaphysics of spirit and nature. Cultural sciences, which would perhaps be a better translation of the German term *Kulturwissenschaften*, no longer consider culture as a sector of the spirit in modern society. As part of the "cultural turn", culture is understood as a symbolic order of the world in which society no longer has the meaning of the supersystem and culture the subsystem. This is about the complexity and irreducibility of culture. It is technologically reproduced as a letter, text and image. Therefore, visual culture cannot be just one of the "cultures", rather it denotes a fundamental marker for the new media status of a culture that visually constructs social reality (Paić 2019).

It seems that should be no longer possible to seriously talk about aesthetics without an insight into the results of the process of refining art itself from abstraction to enformel, from the radical iconoclasm of the Russian avant-garde to virtual or digital art "now" and "here". It is useful to refer to Margit Lovejoy's statements. She says that the basic category of digital art becomes interactive communication (Lovejoy, 2004: 220-230). In addition, aesthetics, which instead of traditional concepts reaches for the language of new media, can therefore be nothing else than a whole new building of interdisciplinary cultural sciences – beyond aesthetics and culture. Such aesthetics can, therefore, only be *trans-aesthetics.* Being between and being beyond determine, in the language of new media, the ontological place of all

aesthetic categories in the new digital environment. Just as in virtual reality, it becomes impossible to separate the real from the imaginary and the symbolic, so in the ecstasy of visual communications, all previous space-time divisions of past, present and future have fallen away. When information is compressed and condensed, interactive communication itself implodes. The consequences of this acceleration of the network are that the participants of the communication process can no longer withstand the burst of the amount of information. A machine or a tool is smarter and faster than a man, as Norbert Bolz rightly stated. For communication to still be possible as a coherent connection between the participants of "online works", a great critical ability to select information is needed. Machines of imagination and mind need communicators who can imitate machine "anthropotechnics" of data memory.

The trans-aesthetics of the new media no longer have the "space of the future" in front of it at all, as the time of the coming change of the constructed reality of the world. The new media stopped time in the ecstatic moment of a one-time event with the repeatability of what was. The replay effect of the digital age only shows that what constitutes the essence of "our time" is stability in the changing condition of information and communications. In a video performance entitled *Stories from the Nerv Bible*, Laurie Anderson invites her viewers to face the future by trying to determine whether there is hope for human progress or whether we have fallen hopelessly into a state of violence and social lawlessness on a global scale. The contemporary artist evokes the allegory of the angel of history – *Angelus Novus* – from Benjamin's essay on progress in the interpretation of Klee's painting of the same name. The storm of progress no longer leaves the possibility of distinguishing between the previous, the present and the upcoming. Communication processes in the social framework of relationships between people have different means of mediation. If they are neutrally called "media", then emerges already a question of a new relationship of mediation between people based on the visual processing of information. For Sachs-Hombach, the medium denotes primarily the physical aspects of the sign. This is not about any technological, economic or institutional part of media activity in social systems. The distinction between media according to those connected to the human body and those independent of it complements the formal analysis of communication in general.

Fixed forms of communication that are independent of the body are pictures and films. They are transmitted through written language and abstract symbols between users. Gestures and facial expressions are forms of communication temporarily attached to the body as gestural and non-verbal visual communication. The body is understood explicitly as a medium. (Sachs-Hombach, 2006: 96 and 97). The transition from the mimetic-representational conception of the image to the conception of the image as a communication medium is not of the same ontological rank as the transition from the magical-cult understanding to Plato's,

which truly begins the philosophical question of what an image is in general and why it must be in the function of cognition through logos. Before Plato's doctrine of the mimesis of images, it was the identity of the divine and what was depicted in a cult image, statue, or temple. The reversal occurs when the real character and his image are no longer exposed in the picture as identity and unity, when, therefore, mediation occurs in the sense of representing a character with a picture. Sachs-Hombach's concept, as we have seen, is a consistently realized attempt to establish the image science (*Bildwissenschaft*) in the symbolic merging of perceptual, cognitive and communicative aspects of pictoriality. The problem with such an interdisciplinary science, which would need a new (philosophical) meta-theory of the image for its foundation, is that it remains unclear how and in what way the two things should be understood:

> (1) generating a new reality that the image assumes and at the same time "creates" with its presence in virtual space-time;
>
> (2) the transformation of the image as information into a communication medium of visuality.

In both cases, we are faced with the question of the ontological status and function of the image in what is real and what becomes visually constructed. Contemporary debates on these issues do not subside. On the contrary, it seems that there exists an awareness of methodical doubt about the scientific character of such a post-science image that should be part of the new general cultural science (*Kulturwissenschaft*). One of the convincing critical answers to that question has been provided by Lambert Wiesing, a contemporary German philosopher, and theorist of image and visuality. He, namely, against any semiotic-communicative model of the image as a closed circle of meaning in which images refer to images as signs in the communicative chain of events, tries to save the phenomenological approach to the image. If a picture shows something, this does not mean, according to him, that there exists a relationship of noticing in the sense of an intersubjective relationship in which there is still a reference to something real in the relationship between the picture, the observer and the excess of the imaginary.

The question that must be asked is about the character of a completely new artificial presence in the field of media-constructed reality. The artificial presence of the image means that the observer places himself in a situation of understanding the iconic difference between a living or real presence and a non-living or artificial presence. (Wiesing, 2005: 35-36). At the same time, it seems important to warn about the pre-discursive perception of the image and the discursive one as a remnant of the iconological tradition of interpreting the meaning of the image in the history of art. In the non-living space-time of immersion of images in virtual reality, the observer finds himself in a situation that necessarily has the double character mentioned above. At the same time, he is free from the excess of previously acquired knowledge about the meaning of images, which he sees as intertextual and

metatextual creations of media images. But on the other hand, his vision is mediated by the awareness of the changed reality in which the present reality is presented to the observer. Thus, in its material aspect, the painting is shown as an intentional object. But the experience of viewing such an artificially generated image, for example on a computer interface, already significantly changes the meaning of the phenomenological concept of intentionality. That all images are something intentionally determined by perception cannot be problematic. What do artificially generated images refer to? This might be the main problem of the contemporary discussion about the image in the digital environment. The phenomenological concept of intentionality from Husserl to his numerous successors had something of vagueness. Consciousness in all its modes of presence in the world is always intentional consciousness.

2. THE MEDIALITY OF THE BODY

From Foucault's position on the disappearance of the concept of man in the archaeology of modernity to Derrida's critique of the logocentrism of history, there is a path of formal constructive reduction of man to something immanent to him – the mediality of the body within the world as a text. The media denotes the artificial nature of the technoscientific approach to the world and man. From this constructivism, the thought about the eccentricity of the medium and the decentering of the subject inevitably arises. This is of decisive importance for the entire media studies, communication and media philosophy. The eccentricity of the media means that all media are necessarily intermedial and trans-medial, refer to other media and transcend their indeterminacy. The new media of the 1990s do not have their basis in the mere mediation of information because they translate "old" contents into new "formats". It should be already clear that the formal constructivism of global reality is nothing more than a network of rhizomatic action. The decentered subject of the media cannot be human in its intersubjective extension of the senses. It is the very form of media that now becomes the mediality of the event. Far from some mysterious force that inexplicably starts the media process in its development, the concept of mediality can be simply defined as the occurrence of a reproductive event. In it, creation and reproduction on the side of the "subject" of procedural and transmission and mediation on the side of the "object" in the process of events are always mediated. The mediality of the media denotes worldliness without the world of the contemporary digital age. The temporality of the event corresponds to the pragmatic content of the media. Everything becomes informative because the information in its implosion (compression and condensation) becomes the internal logic of reality itself. The interactive nature of the media in the social meaning of the inclusion of the democratic public in the game of discourse and dialogue, as Flusser and, following

in his footsteps, Kittler, explain the process of transferring the public sphere into the private and the termination of "public" and "private" in the corporate structure of the world's communication activities, from noise in communication and saturation with the homogeneity of messages, their uselessness and bareness of meaning, leads to the entropy of the social order itself, which is stabilized only by constantly staging apocalyptic events. (Kittler 2013) In other words, the apocalypse represents the internal structure of the media age of entropy and not any catastrophic consciousness of this time. But the apocalypse is not the reality of the destruction and revelation of the new world, but a media event staged based on paranoia and conspiracy theories.

To that extent, Flusser's definition of the media, in contrast to McLuhan's, seems more thought-provoking for the upcoming era of media trans-mediality in general. Namely, Flusser understands the media epistemologically from the techno-scientific framework of the modern world. Instead of the subject – a man in the humanistic sense of the word – which is still at the foundation of McLuhan's anthropology of the media, for Flusser an intersubjective network of communication is at work. The project replaces the subject, and the network arranges and synthesizes and analyzes events from two worlds, the "natural" or technical and the "social-humanistic" or communication world. Therefore, the film cannot be just one of the other media in a linear sequence of development: from image to letter to visual text. It is the universal and paradigmatic medium of the world as techno-code. Life in the age of the media denotes necessarily a depicted or visualized life in which the entire realization of metaphysics is completed: the emergence and development of the techno-biosphere of the new corporeality of the complex body of living memory and artificial intelligence. When there should be no longer the stability of space and the chrono-utopian structure of time in the sequences of the series, then the mediality of the film itself becomes the event of the creation of a new space and a new time. The conceptual notion of space and time corresponds to the conceptual age of film as ideas in images. Deleuze claimed that film directors think in motion-pictures. It is therefore not at all unusual that precisely with the flourishing of new media in digital format, the philosophical understanding and interpretation of the film will become almost more significant than the artistic interpretation of the film. The paradox is self-explanatory because the loss of the aura of the event of the topology of art and the loss of the original temporality of the event necessarily leads to a *medial turn*. (Paić 2021b, 265-280)

One of the almost expected answers and binding approaches to reality in the modern world of construction and deconstruction of the event itself shows that the media lies behind such a thing. In advance, therefore, it would be assumed that the media have replaced the transcendental constitution of the object of experience. Instead of God or Thing as a condition for the possibility that an event has any internal or external "meaning" at all, there is now a term already problematic in that

it refers to mediation between the two. The mediation of consciousness in its journey from the point of view of "natural" consciousness to the absolute spirit in the form of art, religion and philosophy was for Hegel the "matter" of the phenomenology of spirit. It is the path of mediating consciousness as a spirit to the identity of subject and substance in the absolute science of spirit itself. That is why Hegel's philosophy can be called "absolute mediology". All that intends to form of the medium of spirit in its journey through time of absolute presence as the eternal present. For contemporary media theory in its much-celebrated interdisciplinarity to have credibility, the world must be constructed from that concept of the real that corresponds to the very idea of the world in the age of its worldlessness. Undoubtedly, the "real" lacks the fullness of Being in the excess of attributed reality. The difference between the world and its reality and the medial deconstruction of the immediacy of the relationship between man and the "world" opens up the problem of self-determination of communication beyond any instrumental logic of action.

Communication cannot be the result of a closed circle of information flow in the sense of a vulgar relationship of sender-receiver of a message, but a social relationship between the signal and the decoding of the message in interaction. Therefore, the question of communication in the modern world of network societies is a first-class question about social power and the identity of subjects/actors in general. But can there be communication at all without a world that has become a network of social events? The relationship between the form of media and the form of communication cannot be declared only as a social relationship. It is necessary to determine the relationships between the concepts beforehand. For this a priori form of media, society appears as a condition for the possibility of communication. For example, in sociological theories of globalization, instead of the power of social classes, the basic power structure is now moved to the area of power of communication between social subjects/actors. But such a neo-Weberian approach, as credibly advocated by Manuel Castells in his analysis of the global age of the power of networks and communication, denotes ultimately a kind of techno-determinism. Instead of society, we are now talking about communication networks, and instead of the social interaction of elites in the distribution of real power, communication becomes the power of cognitive capitalism in its highest phase of accumulation of knowledge-awareness-feelings. Hence, the media are not socially determined by some a priori force (a signifier in the semiotic sense of the word). They are set by the formal-material conditions of interactive communication. Therefore, communication becomes possible in the digital age of the "world" only when the "world" is spatially and temporally displaced from the centre. The decentering of the "world" corresponds to the process of media deconstruction of the subject. If the "world" can be called what shapes language in its articulations of thought and bodily experience, then it seems obvious that the media concerning

man and the world have the power to create a new subjectivity. So, it now has the character of active inter-communicativeness.

Just as language in the era of technical destruction of the meaning of the world necessarily acquires the character of a pragmatic means of communication, so the media language of events becomes, as Sloterdijk defines media, a combination of "encyclopedia and circus". (Sloterdijk 1983) The essence of language in the era of media neutralization of events can no longer be explained by anything other than referring to the logic of the spectacle itself. But since the spectacle in all its forms is capital as a social relationship mediated by images in its highest phase of visual communication, it is obvious that the language of contemporary media no longer imitates anything (the referential nature of language) nor represents (signs of society and culture), but precisely redesigns the world as a media spectacle of events without "meaning" in its fatal inter-mediality. Spectacle refers to media forms just as the language of spectacle refers to the material conditions of "real" events. For the spectacle of media self-production of events to function perfectly, language must become a tool of visual communication or, in other words, the empty speech of advertising messages. This is what Baudrillard calls the totalitarian message of the metalanguage of contemporary media.

The formal loss of the metaphysical concept of the world also means the material gain of pragmatic and empirically available communication. On the question of the loss of the world as a historical set of Being, beings and humans, one of the main theoreticians of (new) media, Vilém Flusser, in the essay "The Codified World", says directly:

> Premodern man lived in a world of images, which meant the "world." We live in a world of images, which theories regarding the "world" hope to symbolize. This is a revolutionary new situation. In order to grasp this, the present reflection will attempt an excursus on the concept of codes. A code is a system of symbols. Its purpose is to make communication between people possible. Because symbols are phenomena that replace ("stand for") other symbols, communication is a substitute: it replaces the experience of "that which it intends:' People must make themselves understandable through codes because they have lost direct contact with the meaning of symbols. Man is an "alienated" animal, who must create symbols and order them in codes if he wants to bridge the gap between himself and the "world." He must attempt to "mediate:' He must attempt to give the "world" meaning. (Flusser, 2004, 36-37)

If, on the other hand, no one stands behind the media, then the media themselves stand behind their unrepresentability by anything other than the mediality of the media itself. And precisely this mediality, in Flusser's terms, is the codified world of theories about the world. The difference between theories about the world and thinking about the world in its worldliness alone decides the character of the worldlessness of the contemporary world as a network of medially constructed events. It has become obvious for a long time that in modern culture, technical-technological assumptions of observation and perception can no longer

be lightly rejected. What we see cannot be just some immediately present event. It is always about an apparatus that opens the event to our cognitive and perceptual possibilities that are not naturally given. Rather, the naturalness of the view is made possible by the cultural techniques of viewing itself. The body is structured thanks to changes in the way of its medial performance. When the camera changes the perceived object, we can talk about the technological-aesthetic change of the body as an object. The eye cannot be innocent like any other sense. But for an event to be reproduced in its "truth", much more than a reproduction apparatus is needed. Each medium, in its historical-epochal determination by the bodily structure of action, designates at the same time a cognitive apparatus and bodily situating in the space-time of social relations and the cultural order of meaning. If therefore, the doubt about the "reality" of the event stems from the doubt about the interpretation of the event, which is always the product of diverse experiences of perception and to that extent subject to the ideological formation of discourse, then it seems self-evident that the concept of media in its pragmatic extension to all areas of society, culture, politics, art, sport, to everything that remains of the metaphysical concept of the world, goes beyond what is anthropologically destined for it: namely, to be a mediator of information exchange. McLuhan's assumption that the medium is the message can no longer be a sufficient guarantee of media "neutrality". There is something completely different and much more in the dubious "nature" of the media since the very beginning of the modern era.

It should be therefore not surprising that one of the prominent theoreticians of contemporary art and media, such as Boris Groys, expresses doubts about the effectiveness of the very concept of media and the entire theory that "serves" the monstrously grown communication drive. (Groys 2000) Doubt even goes so far that the only media theory, according to Groys, can be called a conspiracy theory. The reason is that there is no longer any sufficient reason for us to operate with the term "real", and since the media in the digital form of their information and communication activities do not maintain "real", but rather produce and stage it, then it is the need for "real" became simultaneously obsessive and pathological. We have some kind of obsession with the pathological because the "real" no longer has a foundation in the "reality" produced by the technological creation of objects. When it finally became certain that the "real" no longer exists "from above", there emerges a panicked search for the "excess of the real" in the immanent event itself.

3. COMMUNICATION AS A PRAGMATIC TURN

In the semiotic theory of communication, the pragmatics of meaning becomes more important than the syntax and semantics of the message. The performativity of use, therefore, decides the meaning of something as "useful" and "effective". The body is medially determined by the pragmatics of language as the speech of action

and reaction in the space and time of the media. What is the fundamental problem with mediality without a medium or with corporeality without a body? The discussions that take place about it are mostly focused on the question of the language of media. Thus, for example, the contemporary German media philosophy (Krämer, Mersch, Sandbothe, Hartmann and others) starts from the fact that the medial turn, among many other turns from the traditional metaphysics of language and body, is an attempt to overcome the still existing binary oppositions of openness-closedness of language as a body of media or, on the other hand, text-image to liberate the textuality and pictoriality of the world itself in its essential telling of the "traces" of language and the "apparatus" of the body. The issue of language in the referential frame of the media cannot be limited only to the issue of communication in the era of a telematic society. (Flusser 2004, Mersch 2006) The computer language of information as well as the cybernetic aesthetics of communication are made possible by the "third" which cannot be reduced to the formal structure of the technosphere and the material structure of the biosphere. The discomfort stems from the fact that all technologies today are those which synthesize information and communication, and their language pragmatics becomes an almost perfect game of codes that change programmatically as new software is perfected. The technology is aesthetically constructed. Therefore, the new media synthesize within themselves four metaphysical causes in the complex situation of the techno-biosphere:

(1) format of knowledge of reality as a project in the form of visualization (3 D);

(2) the materiality of the body as an object of perception in the social environment of information (network societies);

(3) the effectiveness of pragmatic action based on the performativity of speech in changed situations and contexts of the interactive culture of Others (interface culture);

(4) expediency in the productive consumption of the "new" as an immovable driver of the entire reality of the cotemporary world in which cognition and sensibility are combined in the memory machine of capitalism.

The first Aristotelian understood cause has a hidden primacy in this scheme. The formal cause, namely, represents the transcendental condition for the possibility of the entire system functioning, and the fourth as the final cause of reality combines production and consumption, since consumption and its subjects/actors can never stop in the form of stable order, but exists permanently in the process of renewal and crisis of production potential. The second and third, material and efficient cause, show how modern society and culture are established in a pragmatic-performative way - with the logic of new media. It is not difficult to conclude that society in the age of medial neutralization is always, as shown by Baudrillard,

Foucault and Deleuze, Flusser, Latour and Kittler and many other posthumanists, a society of the control governed by a bio-technological code. The self-organizing logic of culture cannot be functional. It can develop into a complex system of networks only because the path of direct mediation of the living body as an image that emanates energy, feelings and experiences has been technologically realized. Taken as a whole, society denotes a pragmatic set of information. It works by feedback culture self-organizing the order of social relations based on communication between different subjects/actors of social power. There should be no doubt that society in its pragmatic way of acting is determined by the result of the disintegration of the primary sphere of mediation of freedom and therefore the control over its worlds of life (culture as an interface) becomes a question of contemporary economics and politics. Surveillance takes place by the fact that the networks of direct mediation (new media) are economically and politically more and more subservient to the coupling of capitalist transnational corporations and authoritarian forms of "democratic control" of the cyberspace of information exchange.

The interactivity of new media creates the illusion that the world is unique and that technology is neutral. If there is anything common to all media theories, it is the viewpoint that the media as a technology of information transmission cannot be neutral. This applies equally to media "techno-determinists" and "cultural voluntarists". What is different in the approach to media stems from the relationship between ideas and reality. Because if use is the one that decides to change things by setting new rules of the game in a society and culture, then it is self-evident that the question of use has become a fundamental question of the provision of media in general. Instead of asking "what" is an event and what is its meaning, it seems now important to know "how" something happens and what is the meaning of the events in the event itself. The procedure of events denotes an open process of the mediatization of the body. It is not free in the language that the medium uses to talk about the body, but it is in the decision about the path of language that the medium uses in its embodiment. The pragmatics of meaning, therefore, should be reduced to use. So, the performativity of new media speech denotes their conceptual body of resistance in performance. That is why media should not be understood in any other way than as a set of four causes in the contemporary turn of the world itself.

So, what denotes the term – illusion? Precisely because the digital world is always the one that, through the implosion of information, creates the appearance of reality as an abundance and breach of information from which communication can become either a mass rebellion against the society and culture of hegemonic distributed power or, on the other hand, mass indifference towards all actions to change society and culture at all. The first case represents the activist faith in the power of new media as subversive resistance, and the second case should be named as the escapism and nihilism of retreating to one's own Voltaire's garden in

conjunction with the choice of lifestyle as a narcissistic celebration of the empowerment of a self-conscious individual. The information society in its telematic form denotes only the technical realization of the end of the linear code. Being "connected" to the network does not mean being socialized or culturalized. To be "connected" to the network means to have opportunities for freedom of information, even under the condition of a threat to that fundamentally liberal idea of freedom for all. Instead of the democratization of the media and the great digital utopia of dialogue at a distance, "now" and "here" we are facing a crisis of dialogue. The democratization of the media has played its part. The subversive demand for free software hides the potential of the socio-cultural struggle in the virtual space for the redistribution (socialization) of surplus value. Communication has come from ecstasy to the stage of final castration of the Father/Law, in Lacanian terms. Information from the network is used by everyone, and communication becomes inter-passive conduct of dialogue as a directed democratic monologue (Pfaller, 2002).

Does a man have the power to control technology? Can he use it for "his own" purposes? By our definition of the media as a means/purpose of the construction of the horizon of the world's meaning, the ambiguous nature of the media determines the disputes between the representatives of two theoretical-cognitive and critical positions on the relationship between man and technology until today. The ambiguous nature of the media presupposes a technical-technological division of the media and a social-cultural one. The former is about the ontological, and the latter is connected to the anthropological definition of the media. The first "ontological" and techno-deterministic attitude places the media on the level of conditions for the possibility of socio-cultural communication. Second, the "anthropological" and "humanistic" attitude is critically opposed to the power of technique-technology. Man as a subject rises to the rank of a historical being of change. According to this theory, a man cannot be passive, but an active being of creative freedom. From the new Era to contemporary theories of media and new media, such a dual structure of thought has been maintained.

Technological or media determinism tries to explain all social and cultural phenomena from a causal-teleological model (*cause-effect*). Changes in society and culture are made possible by a significant change in the way technology is used by man throughout history. The concept of technological determinism was introduced into anthropology and sociology by the American economist and sociologist Thorstein Veblen at the end of the 19th century. and at the beginning of the 20th century. The historical materialism of Karl Marx and Friedrich Engels as a critique of Hegel's philosophy of history is mostly understood as non-reversible materialistic (technological) determinism. Marx's concept of production forces (capital, science, technology) and production relations (classes, social context of capitalism, interpersonal relations) in the framework of his dual scheme of the development of

history underlines the possibility of continuing the determinism of culture as a medium in the contemporary era of global capitalism. But to what extent is it still possible to understand communication in social relations and cultural orders of meaning as "intersubjective relations"? Although the exit from the scheme of subject-object relations is resolved in this way, the difficulties are that it remains in the conceptual horizon of the speculative philosophy of the subject. Flusser, like the entire theory of new media in the German-speaking environment, is certainly a big step "forward". One cannot even imagine a radical theory of new media with the concepts of "digitalization", "virtualization", "dematerialization", "decentering", high tech civilization without simultaneously radically abandoning the talk about "man", "society" and "culture" within the framework of traditional technical and humanist ideas about historical development. If "man" is considered anthropologically, but now with the leading thought of "new anthropotechnics" (Sloterdijk, 2009) that are at his disposal by insight into the changed technical sphere that is becoming bio-technologically organized – nanotechnologies, genetic changes, biotechnology, cloning – then the theory of new media, from this cognitive-theoretical perspective tries to release the burden of dogmatic technological reductionism and false humanism of communication. The transition to the posthuman environment of new communication technologies, therefore, requires an attempt to define a new "man". It is no longer an anthropological question of what is man among other living beings, but how can "man" still maintain his "humanity" if he cannot be more considered as a unity of techno-spiritual connection with the divine, the world and "nature".

It is not uncommon to say that the digital age has made it possible to get out of the metaphysical labyrinth of history. In it, a man was always determined autonomously. As a creator of technology, as a social being of change and as part of the symbolic order of connecting material structures and spiritual processes, a man defined himself as a privileged being of mediation between God, the world and nature. Man, therefore, was metaphysically determined medially. In the anthropological horizon, the media were an extension of the human body (McLuhan, 1994). In the semiotic horizon, a "man" was understood as a technical-social-cultural order of information-communication that refers to other signs. Therefore, a 'man' becomes an extension of media (Flusser, 2004) with other informational means. In both versions, anthropological and semiotic, the essence of man is ultimately reduced to the possibility of imaginary-symbolic production in a changed historical context. Communication becomes more than a message as information. Only with the possibility of communication does "man" become a co-participant in a system in which God, the world and nature can become a divine, world-historical and natural meeting point of a new relationship with history, or else they can completely disappear in the absolute visualization of the world as pure information without communication.

4. VISUALIZING CONCEPTS

Let's repeat: the media is not a means of communication, but a means/purpose of communication in such a way that it creates a condition for the possibility of communication at all. That is information. Why, then, when talking about new media and the consequences they have on the development and changes of social structures and the cultural order of meaning, is it not simply that there are only information technologies? Why should a completely different term often be used – communication technology? Is there arbitrariness in this, or is it a matter of a certain cognitive-theoretical position in the discussion of new media deciding how and in what way the term information will be used, as well as communication? It would be logical to conclude the following. Techno-determinists use the term information technology. For them, communication means something secondary, derived, non-autonomous, an effect and not a cause. Anthropologically oriented communication theorists will avoid the term information technology in favour of communication technology. But they will admit that the society in which science, technology and technology rule is an information society. The conflict between these two cognitive-theoretical positions is noticeable in all discussions about the nature of contemporary global capitalism. Thus, for example, the techno-determinism of the information age in the analysis of globalization is represented by the sociologist Manuel Castells in a neo-Weberian way (Castells, 2011). The second stream would be represented by an attempt at a critical theory of risk society from the point of view of a cosmopolitan alternative to the techno-determinism of globalization in the works of the German sociologist Ulrich Beck (Beck, 2008).

I argue that it seems necessary to take a step beyond this cognitive-theoretical, but also consequently ideological conflict. The contemporary situation, which in our digital age determines the understanding of technology, media, society and culture, requires a radical overcoming of the already-mentioned metaphysical duality. Instead of the model of cause-effect, subject-object, and structure-function, it seems possible from the context of the theory of complexity to show the irreducibility of information and communication so that their separation, as well as the unquestionable relative supremacy of the first term, remains preserved in a productive relationship. New media becomes a new information and communication technology. Just as a man in the age of new media denotes a network of relationships between technology, society and culture, the same goes for the operation of new information and communication technologies primarily in the telematic society of immediate availability, information exchange and interactive communication. Such technologies are fluid and fractal, universal and particular, decentralized and creative. However, they are not an extension of the human body, nor a continuation of a man in a thinking machine. These technologies are above all the unity of "living images" and "images of life". Without visualizing concepts, new media could not have a cognitive function. Conceptual art rests on a pure idea in a

sign, text, or image. In the language of new media, information and communication technologies are not purely thoughts. They are "living images of life". The bio-digital circuit determines in a visual or pictorial turn (iconic turn) what is even an act of thought in the digital age. Speech, language, text and image of life are reconciled to identity. Now visualization has become more than a pure illustration of thought. It means the productive effect of mental images in the new reality of the world as a project of an intelligent machine.

CONCLUSION

New information and communication technologies form a bio-digital complex of science, technique, technology and life as artificial intelligence. But it no longer refers to "man" and signs, but to cognitive maps of the posthuman adventure in virtual space as a new real-time. Only from this point seems to be possible to understand why the techno-culture of our time denotes a post-culture that no longer has God, the world and nature as its object. The structures, processes, and codes of a "new" world whose horizon of meaning is determined by the pragmatic use of information in the concrete world of life replace God, the world and nature. Getting out of the dead end of the form and content of communication cannot be effective if it is not seen at the same time that the formal content of the communication is always threefold determined:

> (1) how communication occurs in the process of historical development from logos, text to image;
>
> (2) an order or system of signs by which communication from an order or system of information is translated into a new language that can be communicated and understood by the community of information users;
>
> (3) instructions for action based on dialogue and discourse in the community, which is enframed on the principles of mutual interaction of different subjects/actors, regardless of their different, conflicting interests depending on the position they occupy in the social order of roles, status, and lifestyles, as well as the cultural representation of power through ideological practices that are available to them in the visual culture of the digital age.

The formal content of communication removes the distinction between form and content of the communication. The medium of communication cannot be neutral. The messages are neither syntactically, semantically, nor pragmatically independent. Messages are always medially determined. Using the possibilities information technology, the user who is both the receiver and the sender of messages (interface-feedback), in Lacanian terms, unconsciously knows that the language he communicates with becomes a condition for his ability to act as a participant in interactive communication. Therefore, it was necessary to recall that

Flusser methodically overcame the duality of "man", "society"-"culture" and "machine" in his communication theory. Why, in the end, I believe that the communicative polycentric orientation, which was created by the logic of social networks in the age of cybernetic creation, storage and transmission of information, is impossible to ethically restrain to the extent of some kind of liberal-democratic repressive tolerance of the participants in the dialogue-discourse? The problem with open communication today does not stem from the condition of post-totalitarian censorship of information, but from the fact that every technological advance with so-called social networks always means a breakthrough of ethical-political boundaries in understanding the Other.

If information constitutes the essence of cybernetics, then its acceleration, dissemination and storage aim beyond the limits of communicative irrationality, to twist Habermas' famous formulation about public consensus as the basis of liberal democracy. Instead of a metaphysical grand narrative about the rule of the principle of identity and a sufficient reason for explaining phenomena in the world, we encounter a cybernetic turn. Now, namely, contingencies produce events that are not a necessity of things but belong to the set of events on the other side of determinism and indeterminism. This is also the case with new media that produce communication and are not just a means of visual transmission at a distance. The *videological turn* indicates the breaking of ethical boundaries. Instead of restricting freedoms, it is much more significant to go beyond the borders of the global society of control as the first and last station on the way to a new technological singularity.

REFERENCES

Beck, Ulrich. 2008. *Weltrisikogesellschaft. Auf der Suche nach der verlorenen Sicherheit*. Frankfurt/M.: Suhrkamp.

Castells, Manuel. 2011. *Communication Power*. Oxford: Oxford University Press.

Flusser, Vilém. 2004. *Writings*. Minneapolis: University of Minnesota Press. Ed. Andreas Strohl.

Groys, Boris. 2000. *Unter Verdacht: Eine Phänomenologie der Medien*. Munich: Hanser.

Heidegger, Martin. 1989. *Beiträge zur Philosophie (Vom Ereignis)*. Frankfurt/M.: V. Klostermann.

Heidegger, Martin. 2000. V*orträge und Aufsätze (1936-1953)*. Frankfurt/M.: V. Klostermann. Ed. von Friedrich-Wilhelm v. Herrmann.

Kittler, Friedrich A. 2013. *Die Wahrheit der technischen Welt. Essays zur Genealogie der Gegenwart*. Frankfurt/M.: Suhrkamp. Ed. Hans Ulrich Gumbrecht.

Lovejoy, Margot. 2004. *Digital Currents: Art in the Electronic Age*. London-New York: Routledge.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge Massachusetts-London: The MIT Press.

McLuhan, Marshall. 1994. *Understanding Media: The Extensions of Man*. Cambridge Massachusetts-London: The MIT Press.

Mersch, Dieter. 2006. *Medientheorien zur Einführung*. Hamburg: Junius Verlag.

Paić, Žarko. 2008. *Visual Communication – An Introduction*. Zagreb: Centre for Visual Studies.

Paić, Žarko. 2019. *White Holes and the Visualization of the Body*. London-New York: Palgrave Macmillan.

Paić, Žarko. 2021a. *Aesthetics and the Iconoclasm of Contemporary Art – Pictures Without a World*. Cham: Springer International Publishing.

Paić, Žarko. 2021b. "Aura, Technology, and the Work of Art in Walter Benjamin", in Krešimir, Purgar (eds.) *The Palgrave Handbook of Image Studies*. Cham: Springer. pp. 265-280.

Paić, Žarko. 2022a. *Art and the Technosphere: The Platforms of Strings*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Paić, Žarko. 2022b. *The Technosphere as a New Aesthetic*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Pfaller, Robert. 2002. *Die Illusionen der anderen. Über die Lustprinzip in der Kultur*. Frankfurt/M.: Suhrkamp.

Sachs-Hombach, Klaus. 2006. *Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft*. Cologne: Herbert vom Valem.

Sloterdijk, Peter. 1983. *Kritik der zynischen Vernunft*, vol. I-II. Frankfurt/M.: Suhrkamp.

Sloterdijk, Peter. 2009. *Du musst dein Leben ändern. Über Anthropotechnik*. Frankfurt/M.: Suhrkamp.

Wiesing, Lambert. 2005. *Artifizielle Präsenz. Studien zur Philosophie des Bildes*. Frankfurt/M.: Suhrkamp.

Wittgenstein, Ludwig. 1998. *Philosophical Investigations*. Oxford: Wiley-Blackwell.

 Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 33-45
ISSN: 1825-5167

RICERCARE L'INTESA CON UNA MACCHINA
COMPATIBILITA FRA UMANI E SOGGETTI ARTIFICIALI NELLA TEORIA DEL DISCORSO DI HABERMAS

PAOLO CAPRIATI
Università degli Studi di Palermo
Dipartimento di Giurisprudenza
paolo.capriati@unipa.it

ABSTRACT
Is the emergence of artificial subjectivities endowed with a certain autonomy compatible with Habermas' theory of discourse? To answer this question, it is necessary, preliminarily, to address the question of the Habermasian subject. In doing so, one observes how machines stand as potential interlocutors. As interlocutors, like the other protagonists in Habermasian discourse, they must aim at understanding and must have a legitimate interest. This opens the field to two orders of problems: (a) are machines capable of understanding? (b) can machines have an interest? Arguments borrowed from the Chinese Room debate show how, in principle, there are no obstacles to the entry of machines into discourse. If machines can participate in discourse, then in what language do they talk to humans? To communicate, humans and machines must resort to translation techniques, known as natural language processing. Translation already places itself outside language, deferring to a shared idea of reality in which to root universal meanings. By placing communication outside language, the entry of machines into discourse seems to checkmate Habermasian theory. However, two further arguments – "the Others mind reply" and the private language argument – make the case that communication between humans and machines poses no different problems than communication between humans alone.

KEYWORDS
Habermas, discourse theory, artificial entities, Chinese room, natural language processing.

1. INTRODUZIONE

Questo scritto intende analizzare i profili di compatibilità fra l'emergere di soggettività artificiali e la teoria del discorso di Habermas.

Tale questione verrà analizzata da due prospettive diverse.

La prima prospettiva sarà volta a valutare la tenuta *esterna* del pensiero habermasiano: la teoria del discorso può essere adottata per descrivere *anche* la comunicazione fra umani e macchine?

Prima di rispondere a questa domanda, occorre chiedersi se la comunicazione intersoggettiva (o senza soggetto) di Habermas – desoggettivizzando il processo discorsivo dell'intesa – ci risparmi dall'onere di indagare la compatibilità fra teoria del discorso e lo scambio comunicativo umano-macchina. Come si vedrà, Habermas non può fare a meno di interlocutori per dare sostanza al suo discorso. La domanda di partenza, quindi, si rivela rilevante.

Verrà, in secondo luogo, delineato il profilo delle macchine cui ci stiamo riferendo. Non si tratta di qualsiasi tipo di macchina, ma di quelle dotate di una certa autonomia. Qui riteniamo "autonoma" una macchina il cui comportamento non è perfettamente prevedibile o spiegabile a posteriori da un osservatore umano.

Una volta superate le questioni preliminari, approfondiremo alcuni profili problematici. Se per Habermas partecipano al discorso «i possibili interessati» e se il discorso è volto al raggiungimento dell'intesa reciproca, allora possiamo sollevare due tipi di obiezione.

La prima obiezione è volta a fare chiarezza sul concetto di "intesa". Una macchina è in grado di intendere? La seconda obiezione, invece, riguarda la questione dell'interesse. Si può dire che una macchina possieda degli interessi? Entrambe queste criticità verranno affrontate mutuando argomenti dal dibattito sulla stanza cinese.

Affrontate queste due obiezioni, entreremo nel merito della questione, chiedendoci in quale lingua comunichino umani e macchine. Escludendo che condividono un comune linguaggio, l'impianto habermasiano – che fonda la legittimità del diritto su processi linguistici – rischia di entrare in crisi qualora ammettessimo l'ingresso nel discorso di soggetti artificiali.

La seconda prospettiva di osservazione è quella interna. Attraverso tale prospettiva si intendono considerare eventuali incompatibilità comunicative fra umani e macchine. Ricorrendo ad argomenti come quello della "Other Minds" e del linguaggio privato, vedremo come non ci siano in linea di principio ragioni sufficientemente forti per sostenere che umani e macchine comunichino in un linguaggio così diverso da quello in cui gli umani comunicano gli uni con gli altri.

2. UNA QUESTIONE PRELIMINARE. IL PROBLEMA DEL SOGGETTO

La teoria del discorso è compatibile con le novità apportate dall'avvento delle tecniche di *machine learning*? Occorre definire i due termini della questione e poi riscrivere la domanda affinché sia meno vaga. Che cos'è la teoria del discorso? E quali sono queste novità che potrebbero mettere in discussione la teoria del discorso?

Per teoria del discorso, in riferimento a decisioni giuridicamente vincolanti, intendiamo il principio che «riconduce la legittimità del diritto a quei processi e

presupposti comunicativi capaci di fondare – dopo la loro istituzionalizzazione giuridica – la supposizione che i processi della produzione e applicazione giuridica conducano a risultati razionali» (Habermas 2013: 493).

Le nuove tecnologie che qui interessano riguardano forme definite, in maniera generica, di intelligenza artificiale. Le forme che rilevano sono quelle dotate di un'autonomia tale per cui si può porre una questione di soggettività.

La domanda si rivela così più chiara: la teoria del discorso è compatibile con l'emergere di soggettività artificiali dotate di un certo grado di autonomia? Per quale ragione non dovrebbe esserlo?

La prima questione da affrontare è quella relativa alla soggettività, in generale.

Si potrebbe in prima battuta sostenere che la teoria del discorso, affrontando la questione del soggetto in modo originale, riesce a superare eventuali problemi di compatibilità e competenza con entità artificiali. A fondamento della teoria del discorso, per Habermas, ci sarebbe la ragione comunicativa.

Essa «si distingue dalla ragion pratica anzitutto perché non è più riferita a singoli attori, o a macro-soggetti di natura statale e sociale. Ciò che rende possibile la ragione comunicativa è il *medium* linguistico, attraverso cui s'intrecciano interazioni e si strutturano forme di vita» (Habermas 2013: 33).

Prendendo le distanze tanto da Rousseau – con il riferimento al macro-soggetto dotato di potere legislativo – quanto da visioni che intendono la formazione legislativa come aggregazione di volontà individuali, Habermas sembrerebbe proporre non la via del soggetto, ma una via che passa fra i soggetti, intersoggettiva. Come è stato suggerito dallo stesso Habermas, potrebbe parlarsi di una «comunicazione senza soggetto»[1].

> La teoria del discorso prende finalmente congedo da quelle figure di pensiero – ancora legate alla filosofia della coscienza – per cui bisognava o ascrivere la prassi d'autodeterminazione dei cittadini a un soggetto sociale collettivo oppure ricondurre l'anonimo "dominio delle leggi" alla concorrenza dei soggetti individuali. Nel primo caso, si vedeva la cittadinanza come un attore collettivo che rispecchiava e agiva in nome della totalità. Nel secondo caso, i singoli attori funzionavano da rotelline, o variabili dipendenti, dei processi del potere – processi funzionanti alla cieca, in quanto al di là degli individuali atti di scelta potevano esserci solo decisioni aggregate, non liberamente consapevoli (Habermas 2013: 362).

Pur superando la dicotomia macro-soggetto/pluralità di individui, Habermas riesce davvero a fare a meno di un soggetto politico di riferimento?

[1] Il concetto di "comunicazione senza soggetto" (Habermas 2013; Habermas 2017) rivela affinità con quello di "a*nonymous public conversation*", che troviamo in Benhabib (1996). Altrove (Dryzek 2006), tale concetto è stato valutato insufficiente per analizzare le "*divided societes*". Sarebbe, infatti, troppo amorfo considerando che l'identità dei soggetti che prendono parte al discorso è una questione cruciale.

> La teoria del discorso punta sull'intersoggettività di grado superiore caratterizzante i processi d'intesa che si compiono nelle procedure democratiche oppure nella rete comunicativa delle sfere pubbliche. All'interno e all'esterno del complesso parlamentare, queste comunicazioni senza soggetto formano arene in cui può prender piede una formazione più o meno razionale dell'opinione e della volontà circa questioni politiche, vale a dire, circa materie socialmente rilevanti e bisognose di disciplina. Il flusso di comunicazione che si instaura tra a) pubblico formarsi dell'opinione, b) risultati elettorali istituzionalizzati, e c) decisioni legislative, serve a garantire che l'influenza dei mass media e il potere comunicativo si trasformino – attraverso la funzione legislativa – in un potere amministrativamente esercitabile (Habermas 2013: 362).

In altri punti, il riferimento alla comunicazione senza soggetto si fa ancora più sfuggente.

> Subjectless and anonymous, an intersubjectively dissolved popular sovereignty withdraws into democratic procedures and the demanding communication presuppositions of their implementation. It is sublimated into the elusive interactions between culturally mobilized public spheres and a will-formation institutionalized according to the rule of law. Set communicatively aflow, sovereignty makes itself felt in the power of public discourses (Habermas 1997: 58-59).

In proposito, si è osservato come tali riferimenti finiscano per essere connotati da un certo lirismo e siano supportati da argomenti poco fondati (Goodin 2008: 261).

Da quanto visto finora, espressioni come "senza soggetto" e "anonima" hanno più che altro una carica evocativa. Non è l'assenza di soggetto ciò che Habermas inaugura, ma, al massimo, la difficoltà di catturarlo.

Infatti, seppur rarefatto, un soggetto politico di riferimento nella ricostruzione di Habermas è comunque presente. In questo senso, nonostante il superamento della soggettività dicotomica appena descritta, ad Habermas sono necessari degli interlocutori che portino avanti il discorso. Tali interlocutori si caratterizzano per la loro eterogeneità. Non si tratta, infatti, né genericamente di cittadini, né di soggetti collettivi dotati di legittimità a decidere. Il soggetto che si intuisce tra le righe della teoria del discorso resta indefinito. Sviluppandosi nel medesimo medium linguistico, a questo soggetto è richiesto di comunicare nella stessa lingua degli altri partecipanti al discorso. Comunicare nella stessa lingua è necessario per raggiungere l'intesa.

Sull'intesa reciproca, infatti, si fonda la legittimità di una teoria che rompe il legame volontà/potere. Il potere, in una democrazia, secondo Habermas è legittimato dall'intesa che i partecipanti al discorso possono raggiungere e non dall'aggregazione delle loro volontà.

3. A QUALI MACCHINE CI RIFERIAMO?

Per portare avanti il suo discorso, Habermas non può fare a meno di interlocutori. La loro interlocuzione è volta all'intesa reciproca, pertanto ciò che è loro richiesto è l'utilizzo della stessa lingua. Sembra chiaro che Habermas non voglia riferirsi solo agli individui né solo a macro-soggetti o a centri di potere. Per questa ragione, occorre domandarsi se tra gli interlocutori al discorso di Habermas si possano far rientrare soggetti artificiali.

Bisogna anzitutto domandarsi se una macchina può comunicare. Questa domanda non appare, però, completa. Si tratta di una semplice comunicazione? Ciò che Habermas richiede ai suoi interlocutori è una comunicazione volta all'intesa reciproca. Al di là del comunicare, le macchine sono capaci di intesa e di farsi intendere?

La domanda si rivela tutt'altro che semplice. Si tratta, in altre parole, dell'annosa questione sulla coscienza delle macchine. Dal momento che chiedersi se una macchina sia in grado di intendere o di essere intesa non è una domanda che può essere in questa sede approcciata, verranno prese vie traverse.

Proviamo a immaginare che una macchina partecipi al discorso. Che tipo di partecipazione dovrebbe avere per essere considerata un soggetto e non solo uno strumento? Nell'impossibilità di definire quando un soggetto è in grado di intendere, dobbiamo sostituire i termini della questione.

Per poter essere considerata un soggetto, la macchina in questione deve essere dotata di una certa autonomia. "Autonomia" in questo contesto vuol dire che essa agisce senza essere subordinata alla volontà di altri. Non essere soggetti alla volontà altrui significa rispondere alla propria volontà. Siamo nuovamente caduti in un'annosa questione, che è quella relativa alle capacità di volere di una macchina. Anche a questa domanda non possiamo dare una risposta: occorre percorrere un'altra strada.

Rimanendo sulla via dell'autonomia, un ente può dirsi autonomo quando agisce in base a regole proprie. In ipotesi, dunque, una macchina che agisce in base a regole proprie può considerarsi in una certa misura autonoma e tale autonomia la eleva al rango di soggetto e non di semplice strumento. Ma quando di una macchina si può dire agisca in base a regole proprie?

Il fenomeno del *machine learning* sembra essere particolarmente rilevante a questo proposito. Si può definire il *machine learning* come l'insieme di tecniche e di metodi che utilizzano dati per trovare nuovi *pattern* e per generare nuova conoscenza e modelli utili per l'effettiva produzione sui dati (Van Otterlo 2013). In particolare, ci si riferisce a quegli algoritmi le cui azioni sono difficili da predire per gli esseri umani o la cui logica decisoria è difficile da spiegare a posteriori (Mittelstast et al. 2016).

Seguendo quest'ultima osservazione, le "regole proprie", in base alle quali una macchina agisce, sono tali perché producono comportamenti che sono

incomprensibili e imprevedibili per gli umani. L'autonomia di una macchina, in altre parole, è legata alla sua capacità di essere una "scatola nera": oggetto insondabile e inconoscibile per un osservatore umano. Al contrario, se consideriamo una macchina il cui comportamento è perfettamente prevedibile o spiegabile a posteriori, ci stiamo riferendo a uno strumento. In relazione a una macchina del genere, non ha senso chiedersi se essa sia o meno capace di intendere e di essere intesa: se essa è a priori perfettamente conoscibile non vi sarà alcuna tensione verso l'intesa, giacché questa si dà per presupposta. Pertanto, non possiamo sostenere che una macchina siffatta partecipi al discorso.

Paradossalmente, dunque, le macchine candidate ad interloquire sono proprio quelle il cui funzionamento resta per noi sconosciuto. A differenza delle macchine-strumento, per queste macchine "*black-box*" si pone un problema di ricerca dell'intesa.

Se non è possibile rispondere direttamente alla domanda "le macchine sono capaci di intendere o di essere intese?", si può almeno definire per quale classe di macchine ha senso porre questa domanda. In ragione di ciò, abbiamo escluso tutte quelle macchine la cui autonomia ridotta non permette di classificarle come soggetti.

Da ciò discende che il criterio per definire la soggettività è quello dell'autonomia. In altre parole, un soggetto è tale se è autonomo. Il criterio dell'autonomia è quindi requisito necessario per permettere l'accesso al discorso.

Da quanto visto finora, evitando di definire il concetto di intesa, la teoria del discorso di Habermas non sembra precludere a una macchina l'accesso alla partecipazione.

4. OBIEZIONI E CONTRO-OBIEZIONI PER L'ACCESSO DELLE MACCHINE AL DISCORSO

Si è osservato come Habermas non possa fare a meno di soggetti: egli ha bisogno di interlocutori che prendano parte al discorso. Tali interlocutori sono diversi da quelli che vengono solitamente identificati come i protagonisti del discorso politico. Non si tratta né di individui né di macro-soggetti costituitisi come comunità o come istituzione, ma di soggetti la cui azione è votata all'intesa reciproca. L'azione politica, ridotta a comunicazione, si sostanzia in uno scambio tra attori eterogenei che condividono la stessa lingua. Considerate queste premesse, non sembra esserci alcuna barriera per l'ingresso delle macchine al discorso.

A questa conclusione si possono avanzare due ordini di obiezioni.

La prima è già emersa e riguarda la capacità di intendere di una macchina. Alla tesi che non limita il processo discorsivo dell'intesa ad attori esclusivamente umani, si può obiettare che una macchina non è capace di intendere. Un'obiezione del genere ci rimanda al dibattito sulla stanza cinese di Searle (1999).

Searle immagina di essere in una stanza assieme ad un computer per rispondere a delle domande in cinese che gli vengono passate attraverso una fessura sotto la porta. Searle non parla il cinese, ma consultando le indicazioni del computer, invia risposte corrette in cinese fuori dalla porta, e questo induce chi è fuori a supporre erroneamente che ci sia qualcuno che parli cinese nella stanza.

L'esperimento della stanza cinese intende dimostrare che una macchina è in grado di dare una parvenza di comprensione della lingua, ma questo non significa che possa realmente comprendere. In altre parole, Searle sostiene che i computer semplicemente usano regole sintattiche per manipolare stringhe di simboli, ma non hanno comprensione del significato o della semantica.

L'esperimento della stanza cinese non è stato risparmiato da critiche. Quella che ai nostri fini risulta più rilevante è nota come "the Brain Simulator reply", sostenuta fra gli altri da Paul e Patricia Churchland (1990). Secondo questa critica, bisogna immaginare un cervello artificiale che simula il funzionamento di quello umano: anziché operazioni su stringhe di simboli, questo cervello artificiale imita le sequenze di attivazione nervosa che si verificano nel cervello di un soggetto che parla cinese quando lo comprende. Dal momento che un computer del genere avrebbe lo stesso funzionamento del cervello di un essere umano che parla cinese, processando le informazioni alla stessa maniera, si può sostenere che questo computer sia in grado di comprendere il cinese[2].

Una tale critica mira a riconsiderare le possibilità di comprensione – e quindi di intesa – di una macchina: mettere in discussione l'assunto che le macchine non sono in grado di pensare mostra come non abbiamo argomenti sufficienti per sostenere che le macchine non sono in grado di intendere.

Si tratta di un capovolgimento dell'onere della prova: ciò che occorre dimostrare non è la capacità delle macchine di pensare, ma la loro incapacità. In altre parole, il dibattito sulla stanza cinese mostra come non ci siano argomenti sufficientemente solidi per escludere a priori la capacità di pensare di una macchina. In assenza di tali argomenti, la prima obiezione sulle possibilità di una macchina di partecipare al discorso non può essere accettata.

La seconda obiezione ha a che fare con il concetto di interesse. Come precisa Habermas «le norme e le regole che possono pretendere legittimità sono soltanto quelle che tutti i possibili interessati potrebbero approvare in quanto partecipanti a discorsi razionali» (Habermas 2013: 546). Occorre interrogarsi, dunque, sulla capacità di una macchina di avere degli interessi. Se si dimostrasse, infatti, che le macchine non possono avere interessi, allora non sarebbero legittimate a partecipare al discorso di

[2] Oltre ai Churchland, si sono occupati di questa tesi Chalmers (1996), Cole & Foelber (1984), Pylyshyn (1980).

Habermas. L'obiezione può essere formulata nei seguenti termini: un'entità incapace di comprendere non può avere coscienza di se stessa, quindi dei suoi interessi.

Questo tipo di obiezione, in realtà, si rivela ancora più debole della precedente. Il concetto di interesse, infatti, può essere sostituito con quello di obiettivo. Se è difficile accettare completamente che una macchina possieda una coscienza, non si può escludere che essa sia in grado di perseguire degli obiettivi, che corrispondono ai suoi interessi. Avendo disgiunto il concetto di interesse da quello di coscienza, le resistenze nell'accettare l'incapacità di una macchina di pensare non si dimostrano argomenti rilevanti in questa seconda obiezione

Riusciamo a escludere completamente che una macchina possa perseguire un determinato interesse? O che sia interessata a che le cose vadano in una certa maniera anziché in un'altra? Non trovando argomenti che persuadano sull'inesistenza di interessi da parte di una macchina, una tale obiezione non può essere tenuta in considerazione.

A questo punto, la strada verso l'accesso al discorso da parte di una macchina sembra più spianata: queste due obiezioni non si sono rivelate un reale ostacolo.

5. IN CHE LINGUA PARLANO UMANI E MACCHINE?

Superate le obiezioni relative all'impossibilità di considerare le macchine come parti attive del discorso, si proverà a indagare quale sia la lingua in comune fra umani e macchine.

Umani e macchine riescono a comunicare fra loro attraverso tecniche di *natural language processing* (NLP). Si tratta, in altre parole, di strumenti per l'elaborazione del linguaggio naturale. Il NLP non crea una nuova lingua, ma rende intelligibili per una macchina proposizioni del linguaggio naturale. Si tratta, in altre parole, di un'operazione di traduzione.

Non esiste, dunque, una lingua comune ad umani e macchine. Essi possono comunicare attraverso processi di traduzione da una lingua all'altra.

L'assenza di una lingua comune, però, non significa incomunicabilità. Svincolare la comunicazione dalla lingua significa trovare un nuovo fondamento per la pretesa di validità delle norme. Se linguaggi diversi non rappresentano più un limite per la comunicazione, allora uno spazio centrale viene acquisito dalle regole di traduzione. Tali regole non sono la somma dei due diversi linguaggi, ma ciò che permette di comunicare a soggetti che parlano lingue diverse. Tale processo di mediazione finisce per delegittimare e devalorizzare il "discorso" per come l'aveva inteso Habermas. Se ciò che permette a due agenti di comunicare non è il linguaggio in sé, ma le regole di traduzione, l'accordo non è più radicato nel discorso, che smette di essere condiviso.

Esso viene anticipato – o posticipato – allo stadio che riguarda la fase di incontro fra due linguaggi diversi.

Se con la teoria del discorso Habermas fissa nella procedura discorsiva la legittimità delle decisioni, svincolandola in questo modo da predeterminati contenuti etici, nel dialogo umano-macchina, la legittimità acquisisce una dimensione extra-linguistica. Il dialogo umano-macchina, definito in questi termini, impone di trovare un punto di incontro diverso da quello meramente linguistico.

A questo punto, l'accordo sulle regole di traduzione può essere inteso come una fase del processo discorsivo di intesa. Ciò tuttavia potrebbe apparire come una forzatura. L'accordo sulle regole di traduzione da una lingua a un'altra è esterno al linguaggio cui si riferisce Habermas e ha bisogno di un parametro terzo di riferimento. Tale parametro si deve radicare su una concezione condivisa di realtà che va oltre il linguaggio stesso. Un radicamento del genere priva la teoria del discorso della sua forza pervasiva.

6. IL PROBLEMA DEL LINGUAGGIO

Se ammettiamo l'accesso di una macchina al discorso, occorre riformare la definizione di linguaggio. Se non riuscissimo in questa operazione, dovremo ammettere che l'emergere di soggettività artificiali non è compatibile con la teoria del discorso.

La definizione di linguaggio di Habermas principia da una rottura con i postulati metafisici di Kant – sulla contrapposizione tra intelligibile e fenomenico – e la dialettica speculativa di Hegel – tra essenza e fenomeno.

Se «"Reale" è ciò che si lascia simbolicamente formulare» (Habermas 2013: 44), il rapporto fra linguaggio e mondo viene completamente riconfigurato. Il linguaggio per Habermas è il mezzo che rende il reale "reale". Agire all'infuori di esso vuol dire perdere contatto con la realtà.

Se il dialogo umano-macchina è fra interlocutori che parlano lingue diverse, non si può non riferirsi ad una universalità dei significati. Tale universalità trascende il linguaggio e rende gli sforzi di ancorare la legittimità politica a un processo linguistico vana.

Ma riformando in questo modo la definizione di linguaggio la teoria del discorso sembra perdere significato. Al contempo, non si può non rilevare come umani e macchine parlino lingue diverse, tanto che ci sia bisogno di particolari tecniche – si veda il NLP – per metterli in comunicazione.

D'un tratto l'accesso di soggetti artificiali al discorso si complica e sembra mettere in crisi la teoria del discorso stessa.

7. UNA CONCEZIONE DIVERSA DI LINGUAGGIO

Abbiamo osservato come non ci siano ostacoli insuperabili per l'accesso di macchine al discorso habermasiano. Le obiezioni sollevate non risultano preclusive. Tuttavia, la questione si è rivelata più complessa da quando ci si è domandati se esiste un linguaggio in comune fra umani e macchine.

La domanda di partenza riguardava la compatibilità fra a teoria del discorso e l'emergere di soggettività artificiali. A questa domanda abbiamo dato una risposta positiva. Ciononostante, dei problemi di compatibilità continuano ad esserci.

Si proverà a questo punto ad adottare una prospettiva interna: esistono profili di incompatibilità nella comunicazione fra umani e macchine?

Se l'incompatibilità non è legata a presunte deficienze del soggetto artificiale – come l'incapacità di intendere e di essere inteso e l'impossibilità per una macchina di provare interessi –, altre problematiche ricompaiono in riferimento al linguaggio.

Assumere che umani e macchine parlino due linguaggi diversi mette in crisi il percorso fondativo della legittimità di Habermas su base linguistica. A quel punto diventerebbero dirimenti le regole di traduzione e, dunque, sarebbe necessario il riferimento ad una realtà extra-linguistica che si fondi su una universalità dei significati. È solo facendo appello a tale realtà che si può immaginare una autentica possibilità di intesa per soggetti che comunicano in lingue diverse. Ma l'appello a tale realtà extra-linguistica vanifica gli sforzi fatti da Habermas che fonda il suo discorso su un piano puramente linguistico.

La necessità di riferirsi a una realtà extra-linguistica è il frutto di un assunto di partenza: gli umani comunicano tutti nello stesso linguaggio, mentre le macchine comunicano necessariamente in un'altra lingua.

Anche in questo caso, possiamo derivare una obiezione a questa conclusione dal dibattito sulla stanza cinese. In questo caso, ci rifacciamo alla critica nota come "the Other Minds Reply"[3].

Questa critica è formulata nei seguenti termini: come si fa a sapere che qualcuno è in grado di comprendere il nostro linguaggio? Solo dal suo comportamento. Se una macchina è in grado di superare un test comportamentale (in questo caso riuscendo a comprendere efficacemente il nostro linguaggio) tanto quanto qualsiasi altra mente e se si intende attribuire la capacità di comprensione ad altre persone, in linea di principio la si deve attribuire anche alle macchine.

Sempre in linea con una lettura anti-solpsistica della comprensione, l'argomento del linguaggio privato di Wittgenstein (1967) intende dimostrare come una lingua che si rivela incomprensibile per chiunque, eccetto che per il suo utilizzatore originario, è

[3] Si sono occupati di questa critica Dennett (1987), Moravec (1999).

impossibile. La ragione sta nel fatto che una tale lingua risulterebbe incomprensibile anche per il suo utilizzatore originario, giacché egli non potrebbe stabilire un significato per i suoi presunti segni.

Da ciò risulta come, anche attraverso l'obiezione linguistica, il tentativo di invalidare l'accesso delle macchine al discorso si è rivelato fallimentare – o non particolarmente incisivo. Che tra umani e macchine ci sia una irriducibile difformità linguistica non può essere efficacemente asserito.

Non si può, quindi, sostenere che l'emergere di soggettività artificiali sia in linea di principio incompatibile con la teoria del discorso, così come non siamo riusciti a sostenere che umani e macchine parlino due lingue a tal punto diverse da dover trovare alla realtà una dimensione extra-linguistica.

8. CONCLUSIONI

Questa analisi ha inteso affrontare gli eventuali problemi di incompatibilità fra la teoria del discorso di Habermas e l'emergere di soggettività artificiali dotate di una certa autonomia.

Prima di approcciarsi a questa domanda è stato necessario sgomberare il campo da alcune questioni preliminari. La prima ha riguardato il problema del soggetto. Benché Habermas proclami una comunicazione senza identificare precisamente quali siano i soggetti coinvolti, il suo discorso ha bisogno di specifici interlocutori per essere portato avanti. Si è dimostrato, in questo modo, come la domanda sulla compatibilità umano-macchina sia perfettamente rilevante. In secondo luogo, è stato definito il profilo delle macchine che potrebbero far sorgere problemi di compatibilità. Si tratta di macchine che rispondono a regole proprie e il cui comportamento non è per noi perfettamente prevedibile.

Sono poi emerse due ordini di obiezioni diverse. Il primo ha a che fare con il concetto di intesa e con l'annosa questione "se le macchine possono pensare". Il secondo, connesso al primo, riguarda la capacità di una macchina di avere degli interessi. In entrambi i casi, si sono mutuati argomenti dal dibattito sulla stanza cinese di Searle. In particolare, è emerso come obiezioni del genere non siano sufficientemente incisive da escludere l'accesso per le macchine al discorso.

Stabilito che, in linea di principio, non ci sono particolari problemi di compatibilità, è stata affrontata la questione della lingua. In quale lingua macchine e umani parlano? Se si esclude che essi parlino la stessa lingua e si accetta che la teoria del discorso radichi la sua legittimità proprio nei processi linguistici, verrebbe da concludere che l'accesso delle macchine al discorso manda in crisi la teoria del discorso stessa. Tuttavia, alcune posizioni anti-solipstiche, valide sia per le altre menti biologiche che per le macchine,

hanno mostrato come non ci sarebbero problemi nell'assumere una definizione così ampia di linguaggio da far rientrare sotto lo stesso concetto tanto il linguaggio umano quanto quello artificiale.

Seguendo questa ricostruzione, la teoria di Habermas non preclude una pacifica convivenza fra umani e macchine nel discorso che mira all'intesa.

BIBLIOGRAFIA

Benhabib, S. 1996. *Toward a Deliberative Model of Democratic Legitimacy*. In *Democracy and Difference: Contesting the Boundaries of the Political*, a cura di S. Benhabib, S. Princeton: Princeton University Press, pp. 67-94.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P.M., & Churchland, P.S. 1990. *Could a Machine Think?* In «Scientific American», 262(1), pp. 32-39.

Cole, D. J., & Foelber, R. 1984. *Contingent Materialism*. In «Pacific Philosophical Quarterly», 65(1), pp. 74-85.

Dennett, D.C. 1987. *'Fast Thinking'*. In *The Intentional Stance, a cura di Dennet, D.C.* Cambridge: MIT Press, pp. 324-337.

Dryzek, J.S. 2006. *Deliberative Global Politics: Discourse and Democracy in a Divided World*. Cambridge: Polity.

Goodin, R.E. 2008. *Innovating Democracy: Democratic Theory and Practice After the Deliberative Turn*. Oxford: OUP.

Habermas, J. 1997. *Popular Sovereignty as Procedure*. In *Deliberative Democracy: Essays on Reason and Politics*, a cura di Bohman, J., & Rehg, W. Cambridge: MIT press, pp. 35-67.

Habermas, J. 2013. *Fatti e norme*. Bari: Laterza.

Habermas, J. 2017. *Three Normative Models of Democracy*. In *Constitutionalism and Democracy*, a cura di Bellamy, R. New York: Routledge, pp. 277-286.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. 2016. *The Ethics of Algorithms: Mapping the Debate*. In «Big Data & Society», 3 (2).

Moravec, H. 1999. *Robot: Mere Machine to Transcendent Mind.* New York: Oxford University Press.

Pylyshyn, Z.W. 1980. *The 'Causal Power' of Machines*. In «*Behavioral and Brain Sciences*», 3(3), pp. 442-444.

Searle, J. 1999. *The Chinese Room*.

Van Otterlo, M. 2013. *A Machine Learning View on Profiling*. In *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*, a cura di Hildebrandt, M., & de Vries, K. Abingdon: Routledge, pp. 41-64.

Wittgenstein, L. 1967. *Ricerche filosofiche*. Torino: Einaudi.

 Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 47-60
ISSN: 1825-5167

L'ARCHITETTURA DELL'ETICA DEL DISCORSO E LA COMUNICAZIONE NELL'ERA DEL DIGITALE

FABIO MAZZOCCHIO
Università di Palermo
Dipartimento di Scienze Umanistiche
fabio.mazzocchio@unipa.it

Abstract
The article addresses Habermas's recent critique of the digital age, and in particular the risks digital technologies pose to the systemic interaction between speakers within a democratic public sphere. For Habermas, a virtual bubble that exposes the citizens to unstable communication and post-truth narratives has been created. The article also discusses the architecture of discourse ethics by analyzing diverging foundational claims advanced by Apel's and Habermas's readings. In specific, these divergences regard: the notion of truth within a consensualist paradigm; the relationship between universal and contextual aspects; the question of justification from an epistemic perspective. The article argues that these divergences test the validity of the discursive approach in relation to the dematerialization of intersubjective relationships, as well as the communication with non-human entities.

Keywords
Discourse ethics, digital age, intersubjectivity, democracy, truth

La mutata condizione dell'umano, nel nuovo contesto segnato dall'espansione dell'infosfera e dalla progressiva penetrazione dell'agire digitale dentro l'orizzonte del comunicare, sta ridefinendo le modalità dell'interazione sistemica, le condizioni della relazione comunicativa e il suo *habitat* (Floridi 2017; Stiegler 2015; Taddio e Giacomini 2020).

Il contributo, a partire da tale scenario, si soffermerà in via introduttiva sulle recenti considerazioni avanzate da Habermas in tema di reti digitali e piattaforme s*ocial* e, successivamente, sull'architettonica dell'etica del discorso, analizzando nello specifico alcune delle questioni strutturali riguardanti la proposta discorsivista avanzata, in parallelo, da Apel e Habermas. In particolare si proporrà una lettura divergente dei fondamenti della *Discourse Ethics* nei due autori. Infatti, pur prendendo avvio da una prospettiva similare sul piano della trasformazione del kantismo attraverso la teoria dell'agire comunicativo, i due pensatori strutturano modelli differenziati sul piano del riferimento alla dimensione storica e a quella

trascendentale. In effetti l'illimitatezza della comunità della comunicazione rimane problematicamente legata a un ottimismo di fondo di natura kantiana. Il recupero in Habermas, negli ultimi decenni, del tema del riconoscimento e del relativo sostrato dialettico, sembrerebbe interpretare in maniera più adeguata, rispetto alla formalizzazione apeliana, gli elementi distorsivi dei processi comunicativi, spostando l'attenzione sulle condizioni sociali determinanti sul piano sistemico.

È altresì vero che una riflessione attenta sulla condizione comunicativa umana conduce alla presa d'atto delle differenti possibilità di accesso al gioco linguistico: il riconoscimento filosofico del diritto alla partecipazione alla comunità dei parlanti, e alla realizzazione di forme di vita ritenute buone, è solo un primo momento della costruzione della normatività. L'acquisizione di nuove consapevolezze circa i diaframmi che ostacolano la valorizzazione della soggettività e, ancor più, il compito morale di un superamento di queste condizioni sembrano essere oggi una sfida urgente tanto per l'etica, quanto per il pensiero politico. La grammatica del riconoscere è, dalla nostra prospettiva, il cuore stesso dell'etica del discorso e, nello stesso tempo, dell'antropologia comunicativa. Il riconoscersi come portatori di un'apertura all'alterità consente di scoprire, nel gioco comunicativo, la priorità del noi rispetto all'io e il costituirsi interpersonale della soggettività.

1. HABERMAS E L'ESPLOSIONE DELLE RETI[1]

Habermas pensa, con una felice metafora, alla soggettività, nella sua trama profonda, come una sorta di guanto rovesciato che ha «la struttura della sua stoffa intessuta di fili di intersoggettività» (Habermas 2007: 7). Esplicitando il nesso possibile tra dinamica del comunicare e autenticità riconoscitiva, l'autore tedesco crediamo proponga un'ermeneutica dell'odierna condizione umana come segnata da forme aggregative flessibili e per lo più caratterizzate dall'estemporaneità. Le "comunità fragili" nate con i *social media* sono forme sociali "leggere" che mettono in contatto individui spesso atomici, in cerca di relazioni che soddisfino il bisogno di riconoscimento, ma in modo occasionale e svincolato da un impegno nei confronti di legature stabili. Tali forme di relazione sono veramente riconoscitive? Creano autentici legami? L'orizzonte sociale pare dirigersi verso una destrutturazione dei luoghi accomunanti, resi superficiali e sempre meno stabilizzanti. La comunicazione virtuale sembra affetta dalla cifra della chiacchiera, della dispersione e, dietro un'apparente empatia esistenziale, dell'estraneità. Ciò, allo stesso tempo, asseconda e conferma come oggi, nella quotidianità dei rapporti umani, si dipani un senso smisurato della libertà che prevede legami accettati solo, in larga misura, se funzionali al proprio interesse. Di qui il manifestarsi di elementi

[1] Per una ricostruzione complessiva delle posizioni habermasiane sulla digitalizzazione della sfera pubblica rinvio al saggio di Luca Corchia (Corchia 2022).

di fragilità sul piano comunitario, con il conseguente rischio dell'inautenticità comunicativa.

In linea con l'idea di una ragione pubblica impegnata, Habermas recentemente ha espresso grandi perplessità circa la pervasiva presenza dell'interazione digitale nei processi di partecipazione e informazione della sfera pubblica (Calloni et al. 2021: §§ 4-5). La rivoluzione introdotta dalla digitalizzazione, secondo il francofortese, «ha comportato una rivoluzione dei media che può essere paragonata, per rilevanza, solo all'introduzione della stampa» (Calloni et al. 2021: 141). Per Habermas rimane una strutturale ambiguità tra la l'idea che il web 2.0 e 3.0 possa introdurre più informazione, più partecipazione, più processi di comunità, e la potenziale «disintegrazione della sfera pubblica politica» (Calloni et al. 2021: 141) nella versione democratica e rappresentativa finora conosciuta. La partecipazione dei cittadini passava tradizionalmente dalla mediazione degli organi di informazione, in questo modo la dialettica democratica assumeva la fisionomia del dibattito informato sui temi dell'agenda politica. Attualmente, invece, assistiamo alla digitalizzazione globale della comunicazione e alla «moltiplicazione esplosiva di reti che si presentano come frammentate sul piano funzionale, tematico e personale» (Calloni et al. 2021: 142). Tale frammentazione, per Habermas, si esplicita praticamente in «due aspetti dei nuovi media: da un lato vi è la modalità di utilizzo di piattaforme come Facebook, YouTube, Instagram o Twitter, dall'altro lato c'è la commercializzazione di questo utilizzo. La novità dei nuovi media, tecnicamente parlando, è il loro carattere di piattaforma; essi offrono agli utenti possibilità illimitate di collegamento e di comunicazione con qualsiasi destinatario» (Calloni et al. 2021: 142). Questo scenario mette in evidenza alcune potenzialità positive, come ad esempio la libertà quasi assoluta di espressione e di scelta dei contenuti e i tratti egualitari della rete. I fruitori dei vecchi media unidirezionali, oggi si sono trasformati e sono divenuti loro stessi autori di flussi di informazioni, frequentemente non aderenti alla realtà, in una sorta di bolla virtuale che espone i cittadini all'era della *post-truth-democracy*.

I *social* garantiscono una tendenziale illimitatezza dell'interconnessione, essi «permettono connessioni orizzontali, configurate in modo tale da favorire scambi comunicativi spontanei tra un numero illimitato di utenti» (Calloni et al. 2021: 142). Secondo Habermas però i rischi sotto i nostri occhi sono molteplici. Anzitutto un insufficiente controllo discorsivo dei contenuti; la costituzione di comunità segnate da estemporaneità e provvisorietà; la riduzione di questioni complesse in sintesi banalizzanti o, addirittura, falsificanti. Ed in ultimo, l'emergere di un nuovo controllo globale legato al cosiddetto *capitalismo della sorveglianza* (Zuboff 2023).

«I nuovi media – sottolinea ancora Habermas – dipendono da grandi multinazionali di internet, quotate in borsa, che traggono i loro profitti dall'ottenimento e dallo sfruttamento dei dati personali dei loro utenti; questi dati personali, usati per mettere in atto strategie pubblicitarie tarate sul singolo, vengono

carpiti incidentalmente quando gli utenti accedono ai servizi offerti a titolo gratuito, mediante cui i social media, a loro volta, attirano i propri clienti» (Calloni et al. 2021: 143). Il pericolo di profilare e orientare desideri e bisogni è uno dei più gravi rischi con cui dovremo fare i conti nei prossimi anni. La sfera pubblica sarà sempre più sottoposta a questa sorta di campana di vetro in cui il dissenso, l'autonomia di scelta e la critica rischiano l'estromissione. Potremmo persino trovarci ad abitare un mondo in «cui ciascuno si ritrova come in una sorta di "bolla" personale; si creano "camere dell'eco" autosufficienti e impermeabili all'esterno, nutrite da una cultura dell'opinione che cerca solo continue conferme di sé e non ammette un pubblico critico» (Calloni et al. 2021: 143)[2]. Su tale crinale si potenziano le fonti dei populismi politici, dei maestri del complotto globale e del negazionismo elevato a cifra controfattuale (Habermas 2021).

Probabilmente è in atto una destrutturazione dei luoghi della sfera pubblica tradizionale e un rimodellamento dei livelli di intermediazione comunitaria. Nell'esperienza, oggi sempre più massiccia e pervasiva, dei *social* è possibile rilevare che l'appartenere è motivato da esigenze temporanee o prevalentemente sciolte da impegni vincolanti. La navigazione nel mare del *web* è segnata spesso dalla cifra della dispersione e persino, dietro un'apparente empatia esistenziale, della cinica esibizione. Le conseguenze sul piano del funzionamento della democrazia possono essere gravi. Lo spirito democratico, in qualche modo, istituzionalizza la logica del paritetico riconoscimento intersoggettivo, per via di prassi discorsive includenti e di una sostanziale tensione verso l'interesse generale. Tra i nuovi media e la vita democratica il rapporto però rimane segnato da un certo tasso di ambiguità.

I *social* hanno aperto una fase nuova della democrazia. I flussi informazionali e l'accesso alle fonti aiutano e potenziano, in positivo, la capacità di controllo della società civile sul sistema politico (possono anche funzionare da potente strumento di mobilitazione e di partecipazione dal basso). In alcuni casi assurgono perfino a mezzi di una nuova democrazia diretta e deliberativa, contro la politica tradizionale. I rischi però permangono: la semplificazione; la superficialità delle valutazioni; la crisi della logica della rappresentanza politica (centrale nelle democrazie avanzate); il populismo antipolitico; l'emergere di una *élite* o di una *leadership* carismatica incontrovertibile in contesti di democrazia diretta (Badiou 2011); la marginalizzazione della diversità, in nome di una non ben identificata intelligenza collettiva e volontà della rete.

La comunità virtuale, in senso etico-politico, rappresenta potenzialmente un allargamento dell'*agorà* pubblica, ma non va confusa con il discorso democratico in senso stretto. L'immediatezza dell'interazione virtuale affascina il partecipante per il senso di condivisione che possiede, ma un'interazione senza coordinamento, senza legami vincolanti, rischia di girare a vuoto. Il pericolo, come abbiamo detto,

[2] Una severa critica alle posizioni di Habermas sulla mancata comprensione del digitale è stata svolta in Rheingold 2013.

sotto ai nostri occhi è quello di creare comunità liquide, figlie di aggregazioni estemporanee, iperflessibili e svincolate da impegni continuativi. «Nel mare magnum dei rumori digitali – scrive Habermas – queste comunità comunicative sono come arcipelaghi dispersi: ce ne saranno miliardi. Ciò che manca a questi spazi comunicativi (chiusi in se stessi) è il collante inclusivo, la forza inclusiva di una sfera pubblica che evidenzi quali cose sono davvero importanti. Per creare questa "concentrazione" occorre prima saper scegliere – conoscere e commentare – i temi, i contributi e le informazioni che sono pertinenti» (Schwering 2014). La natura del discorso pubblico, fondato sulla ricerca condivisa di ragioni giustificate, è lontana dal modo semplicistico e veloce in cui si svolge il fluire delle informazioni in rete e l'interazione tra i *partners* della comunicazione.

2. SULL'ARCHITETTURA DISCORSIVA: MOTIVI E DIVERGENZE

Queste valutazioni sull'era delle reti ci consente di tornare, quasi provandone la solidità, ai temi teorici più interni alla struttura dell'etica del discorso e al suo riferimento costante all'agire comunicativo. Il passaggio che riteniamo utile fare riguarda, sul piano del dibattito critico, la differente postura che, ad un certo punto, le filosofie dei padri dell'etica della comunicazione assumono. È vero che i motivi interni alla speculazione di Apel sono simili a quelli di Habermas, per molti versi la critica ha sovrapposto i due autori, segnalando soprattutto le convergenze. In verità, a partire dalla fine degli anni '80, si è delineata una frattura sulla questione fondazionale e sul piano della concezione della razionalità. Si tratta, a mio parere, di una cesura sostanziale tra i due modi di leggere la funzione del *logos* in rapporto alla sfera pubblica e alla storia.

Apel riconosce in molti luoghi testuali la fondamentalità della cornice elaborata da Habermas nel definire l'elemento architettonico centrale della teoria dell'agire comunicativo: si tratta di «quell'ampliamento della teoria semantico-formale del significato attraverso la pragmatica formale (o universale)» (Apel 1997: 174-175). Habermas, per mezzo di un progressivo superamento della *fallacia astrattiva* interna alla tradizione filosofica ed epistemologica occidentale, è riuscito nel difficile intento di tornare a pensare in modo non decostruttivo la razionalità e la dimensione teorico-pratica del normativo. Per Apel, infatti, l'architettonica della teoria della competenza comunicativa habermasiana, ripensando il rapporto tra significato linguistico e pretesa di validità, consente di operare delle differenziazioni teoriche importanti per una filosofia che possa ancora muoversi su un piano universale e veritativo.

Come sappiamo, l'approccio elaborato da Habermas enuclea triadicamente alcuni cardini della trasformazione linguistica del trascendentale. Anzitutto la distinzione tra tre funzioni linguistiche: la funzione rappresentativo-proposizionale in riferimento a stati di cose; la funzione espressiva e quella appellativa, quali

funzioni simbolicamente ancorate. Correlativamente, «la distinzione tra *tre dimensioni di mondo*: 1. il *mondo degli oggetti*, cioè degli stati di cose rappresentabili; 2. il *mondo sociale* della interazione e della comunicazione, regolate da norme; 3. il *mondo interiore soggettivo*» (Apel 1997: 175). Infine, tre specifiche pretese universali di validità, alla luce delle quali è possibile stabilire il significato e la forza stessa degli atti linguistici: «1. la *pretesa di verità*, riferita al mondo oggettivo, della quale sono portatrici le proposizioni asserite in atti linguistici costativi o assertori; 2. la *pretesa di giustezza*, riferita alle norme condivise (giuridiche e morali) del mondo sociale [...]; 3. la *pretesa di veracità o sincerità*, riferita al mondo interiore soggettivo del parlante e sollevata dagli atti linguistici, in quanto atti di autorappresentazione espressiva» (Apel 1997: 175; cfr. Habermas 1986: capp. 1-3).

Sul piano dell'interazione sociale, secondo la lettura di Apel, infatti, l'intesa sarebbe il frutto della forza "legante" delle pretese di validità comunicativa, finalizzata alla coordinazione tra attori pubblici e singoli, mentre sul piano dell'argomentare ogni forma di comunicazione assume una valenza riflessivo-critica e l'intesa è raggiungibile solo per mezzo del dialogo discorsivo.

Tale struttura teorica, come risaputo, ha trovato ampio spazio e consenso nel programma pragmatico-trascendentale, proprio perché l'autore tedesco ha visto in essa un proficuo completamento sia delle sue riflessioni sulla linguisticità del nostro essere nel mondo, sia del programma di una filosofia trasformata attraverso una semioticizzazione del kantismo[3]. In questo quadro, le istanze di Habermas vengono sfruttate da Apel attraverso un significativo rimodellamento teorico. Egli non manca di rilevare che le distanze tra il suo sistema e la pragmatica habermasiana riguardano una «differenza di principio attinente alla strategia fondativa» (Apel 1997: 176). La problematica fondativa, dunque, segna la linea di confine tra le esperienze speculative dei due filosofi: su questo versante si trovano le critiche più profonde e le perplessità reciproche più evidenti. Scrive ancora Apel: «io continuo a credere nel programma di una trasformazione della filosofia trascendentale e, in questo quadro, alla necessità di una *fondazione ultima riflessiva*, che proceda oltre Kant» (Apel 1997: 176). Habermas, invece, segnala che «se rinunciamo decisamente al fondamentalismo della filosofia trascendentale tradizionale, otteniamo nuove possibilità di verifica per l'etica del discorso» (Habermas 1989: 108).

Esplicitati questi contrasti prospettici, va però rilevato che, secondo Apel, «un merito essenziale della concezione habermasiana risiede innanzitutto nella possibilità che essa offre di superare radicalmente i limiti astrattivi dell'esplicitazione 'semantico-formale' del significato linguistico, condotta in termini di *condizioni di verità*, ovvero di analoghe *condizioni di adempimento o di soddisfazione*, per procedere verso una esplicitazione in termini di *condizioni di validità e accettabilità*» (Apel 1997: 177). Infatti, la concezione linguistica di Habermas supera le

[3] Per una complessiva ricostruzione del pensiero di Apel rimando a Marzocchi 2021.

tematizzazioni del rapporto linguaggio-mondo in relazione a un'idea corrispondentistica della verità riferita al rapporto enunciati-stati di cose; e in modo ancor più rilevante, riesce a cogliere la relazione che intercorre tra la dimensione del significato e la dimensione della validità, grazie a un concetto di verità ancorato alla validità intersoggettiva e al riscatto argomentativo (Apel 1997: 177-178).

Questo in buona sostanza significa ricollocare «l'astratta problematica logico-semantica relativa alla corrispondenza tra enunciati proposizionali e stati di cose esistenti nel contesto del discorso. Solo così viene definitivamente revocata la *abstractive fallacy* della semantica logica, orientata sugli enunciati proposizionali» (Apel 1997: 178).

Più problematico risulta il confronto sulla questione fondazionale. Per Apel una fondazione della razionalità è possibile solo attraverso una "stretta riflessione" sulle presupposizioni dell'argomentare filosofico in atto; è solo a questo livello infatti che si può far scattare la mossa dell'autocontraddizione performativa. Siamo sul terreno dell'autofondazione del *logos*; usando una particolare forma di argomento trascendentale, in grado di tematizzare entro il discorso le proprie condizioni di possibilità e validità (Grundmann 1992).

Secondo Apel, «Habermas aderisce ad una forma pragmaticamente ridotta della teoria del discorso o del consenso»; quest'impostazione si potrebbe definire come «teoria dello "smaltimento tranquillizzante" (*'Entsorgungs'-Theorie*) delle pretese di verità problematizzate» (Apel 2004: 260). Viene denunciato l'abbandono di un modello consensual-discorsivo di verità in nome di una teoria pragmatica della verità. Tale abbandono metterebbe in crisi apelianamente il concetto di verità come istanza strutturalmente legata alla validità in senso intersoggettivo e universalistico. Procedendo su tale linea, Apel polemicamente si chiede: «come dovrebbe avvenire secondo Habermas, la (possibile) legittimazione della *validità universale intersoggettiva* delle pretese di verità, se la verità deve essere un "concetto trascendente la giustificazione", se cioè dal suo punto di vista, essa non deve più essere esplicitabile attraverso l'idea del consenso d'una comunità discorsiva illimitata?» (Apel 2004: 277).

Come si rileva dalla lettura di alcune pagine di *Verità e giustificazione* Habermas, che per molto tempo ha sostenuto una teoria discorsiva della verità simile a quella apeliana, sembra protendere verso una svolta che Apel definirà "post-epistemica". L'allievo di Adorno è ancora disposto a sostenere però che negli «ineludibili presupposti dell'argomentazione [pubblicità, inclusione, pariteticità, immunizzazione contro costrizioni interne ed esterne, orientamento all'intesa, vincolo all'argomento migliore, ecc.] si esprime l'intuizione che le asserzioni vere sono resistenti a tentativi di confutazione spazialmente, socialmente e temporalmente illimitati. Ciò che riteniamo vero, deve poter essere difeso non solo in un altro contesto, ma in tutti i possibili contesti, dunque in ogni momento contro chiunque. A ciò si ispira la teoria discorsiva della verità: un'asserzione è vera

quando, nelle esigenti condizioni di un discorso razionale, resiste a tutti i tentativi di invalidazione» (Habermas 2001: 252).

3. TRA VERITÀ E GIUSTIFICAZIONE

Secondo Apel, invece, in Habermas interviene una sorta di svolta post-epistemica, evidente negli scritti più recenti. Si tratterebbe di una torsione pragmatistica che si palesa ad esempio in affermazioni come questa: «Poiché tutti i discorsi reali che si svolgono nel tempo sono provinciali nei confronti del futuro, noi non possiamo sapere se asserzioni che oggi, anche in condizioni approssimativamente ideali, sono razionalmente accettabili, potranno anche in futuro sostenersi contro i tentativi di invalidazione. D'altro canto, la stessa provincialità condanna il nostro spirito finito ad accontentarsi dell'accettabilità razionale *come prova sufficiente* della verità» (Habermas 2001: 253).

Tutto ciò condurrebbe ad un cambiamento quasi-paradigmatico che non dà più modo di pensare a quella particolare "idealizzazione" delle condizioni di verità in grado di assicurare «il nesso interno tra la *pretesa di verità* e la *pretesa di giustificazione* discorsiva, intesa come idea regolativa» (Habermas 2001: 279). Per tale via Apel non vede altra soluzione che il ricorso parassitario – in Habermas come in Wellmer (Wellmer 1991) – a una metafisica realistico-esterna[4]. Ciò che interviene è un'ingenua idealizzazione pragmatica che consente di interpretare l'accettabilità razionale di una asserzione, date come soddisfatte le condizioni della nostra pretesa di verità, fin quando «in futuro non emergeranno nuovi argomenti o evidenze tali da metterla in questione» (Habermas 2001: 279); questa del resto è l'ipotetica stabilità a cui secondo Habermas uno "spirito finito" può avere accesso.

Ciò che ci spinge alla verità è la necessità pratica di affidarsi intuitivamente a quanto creduto incondizionatamente per vero; tale bisogno pratico «si rispecchia, sul piano discorsivo, nella connotazione di pretese di verità che rinviano al di là del contesto di giustificazione di volta in volta esistente e costringono a supporre condizioni ideali di giustificazione – con la conseguenza di un decentramento della comunità di giustificazione. Perciò il processo di giustificazione può orientarsi su una verità *bensì trascendente la giustificazione, ma già sempre operativamente attiva nell'agire*» (Habermas 2001: 257). Questo ritorno «all'idealizzazione "ingenuo-performativa" delle pretese di verità della prassi quotidiana» è secondo Apel – via Peirce – assolutamente da escludere. Ciò che va mantenuto sarà, allora, un'idea regolativa di ricerca della verità che, non orientata "retrospettivamente" verso le idealizzazioni delle pretese di verità pratico-quotidiane, ma "prospettivamente" verso una giustificazione non sopravanzabile, eserciti la sua funzione prescrivendo normativamente la messa in questione di tutti i consensi fattualmente dati nel

[4] Sul tema si veda anche Spinaci 1997.

discorso. In tal senso, il perseguimento dell'idea regolativa della verità non porta, come Habermas sostiene, ad un arresto quasi-escatologico del processo di ricerca (Apel 2004: 280).

Queste precisazioni critiche permettono, in buona misura, di esplicitare la concezione trascendental-pragmatica di verità come consenso. Quest'ultima tiene insieme la dimensione contestuale del discorso, la tensione ad una verità universalmente accettabile e il principio di un fallibilismo non auto-applicabile.

Per Apel, dunque, la «salvaguardia del *nesso interno* tra *pretesa di verità* e *pretesa di giustificazione*» è possibile riferendosi all'utilizzo «di un *infinito potenziale* (non certo di uno attuale)» – nel senso dell'idealità regolativa – ed evitando così un improprio riferimento al concetto di un «mondo reale» che legittima la conoscenza in «chiave *esterno-metafisica*» e adeguativa, o sulla base di «una inconoscibile» realtà noumenica (Apel 2004: 281). «La definizione peirceana del reale in quanto "the knowable-in the long run" [...] evita ogni resto [permanenza] di una *metafisica esterna "from a God's eye view"*. Corrisponde invece alla prospettiva della *potenziale infinità* e, al tempo stesso, alla finitezza *di volta in volta fattuale* della nostra ricerca razionale della verità sulla realtà nel complesso. E ciò corrisponde evidentemente all'esplicitazione del senso della verità in concetti di una *idea regolativa*» (Apel 2005: 303).

4. UNIVERSALE E CONTESTUALE

Il filosofo di Düsseldorf chiaramente condivide con Habermas l'eredità dell'*hermeneutic-linguistic-pragmatic turn*. Stare nel regime instaurato dalla svolta linguistica significa far leva su quelle inaggirabili risorse di sfondo del mondo della vita offerte dalla linguisticità del nostro essere antropologicamente comunicativi. Queste risorse di "sfondo" rendono possibile l'intesa e sono presupposte sempre sia nel nostro stare quotidiano nel mondo, sia nella dimensione riflessiva e inaggirabile del comunicare che il *Diskurs* rappresenta. Due punti, ulteriori, però demarcano le distanze tra i due pensatori: I. se a livello di discorso filosofico è sufficiente o meno affidarsi alle certezze o risorse di sfondo (inaggirabili per ogni prassi quotidiana) offerte dal mondo della vita e proprie di ogni *Lebensformen*; II. se, similmente, la forma riflessiva del comunicare (il discorso argomentativo) possa andar oltre le contestuali "risorse di sfondo", nel senso di un accesso a risorse che «lo distinguano da tutte le forme di comunicazione presenti nel mondo della vita» (Apel 1997: 191).

Il *trend* attuale del dibattito filosofico spinge verso l'ipervalutazione della base contingente e fattuale del nostro argomentare, portando alla relativizzazione di tutti i criteri di verità e alla messa in crisi di prospettive che avanzino istanze fondative. Secondo Apel la riproposizione di una dimensione trascendentale sembra essere agli occhi dei più, compresa la stessa teoria universal-pragmatica di Habermas, un

ritorno a posizioni sostanzialmente pre-svolta linguistica e, dal punto di vista epistemologico, un'incomprensione della *pointe* del fallibilismo.

La posizione di Habermas però si muove in modo non convincente tra l'alternativa posta dal doppio fronte universalismo/contestualismo. Scrive in modo efficace Apel: «Da una lato, egli desidera mantenere l'*universalismo delle prese di validità*, intrinseche all'umano discorrere (pretese di senso, verità, sincerità e giustezza normativa), insieme con il momento di "incondizionatezza" e di "idealità" [...] del possibile consenso di tutti i pensabili partner dell'argomentazione sulla fondatezza delle pretese di validità; e, in tal senso, egli è ricorso a strutture "quasi-trascendentali" o anche a "versione debole" dell'impostazione pragmatico-trascendentale. D'altro canto, però, ha costantemente rifiutato, come impossibile e superflua, l'esigenza di una *fondazione ultima*, valida a priori, delle pretese di validità filosofica degli enunciati pragmatico trascendentali, relativi ai sopradetti presupposti del discorso argomentativo» (Apel 1997: 191-193)[5].

Ciò che più crea difficoltà nel quadro architettonico etico-discorsivo, secondo l'autore, è il fatto che Habermas non riconosca la differenza trascendentale tra enunciati universal-filosofici ed enunciati delle scienze ricostruttive, rivendicando un'applicazione onnicomprensiva del fallibilismo agli stessi presupposti riflessivamente esplicitati dell'argomentazione. La sua filosofia, dunque, si spingerebbe fino al punto di accogliere la proposta "panfallibilista" della scuola popperiana. Secondo Habermas, infatti, sarebbe impossibile confondere la fattuale assenza di alternative con un'universalità assoluta e a-temporale. Egli manca di considerare – nel vincolare le risorse di sfondo alla concreta condivisione di una forma di vita – che i filosofi, in quanto cercatori del vero, non sono in grado di rinunciare ad avanzare ipotesi capaci di dire «*come stanno le cose in assoluto*» (Apel 1997: 194n).

Nell'autore francofortese, se ben leggiamo le critiche apeliane, il tentativo di una fondazione dell'universalità "dal basso" comporterebbe la rilevante contestualità delle condizioni del comunicare e dello stesso gioco linguistico dell'argomentazione. Habermas, infatti, assume l'interazione comunicativa presente nel mondo della vita pensando che essa «contenga i potenziali di ragione, che determinano i fini di lungo periodo dei processi di apprendimento e razionalizzazione» (Apel 1997: 198-199). Ciò farebbe emergere, secondo la lettura apeliana, che: a) la fondazione dei principi della moralità sia già ottenuta tramite il riconoscimento delle condizioni normative della comunicazione quotidiana; b) un'ulteriore riflessiva fondazione ultima degli stessi principi sarebbe impossibile e inutile; c) la fondazione ultima della moralità deve essere sostituita con il richiamo all'eticità sostanziale del mondo della vita, dimensione che alimenta e sostiene sia la coesione sociale che l'autocomprendersi del singolo.

[5] Per una critica alle istanze habermasiane invito alla lettura di Gebauer 2003.

Di fatto in questo modo avviene una svalutazione della dimensione riflessiva e trascendentale a favore di un problematico recupero del piano contestuale dell'eticità del mondo della vita. Con la conseguente deflazione del ruolo critico e fondativo che la filosofia svolge quale «meta-istituzione del discorso argomentativo» (Apel 1997: 200). Salta così agli occhi una sorta di aporia interna all'architettonica habermasiana: il problema della fondazione mette a nudo il contrasto esistente tra le "ispirazioni di fondo" della teoria dell'agire comunicativo e le considerazioni che Habermas ha prodotto negli ultimi decenni.

Interpretando sistemicamente questo tentativo di indebolimento messo in atto dall'allievo di Adorno, Apel scrive: «Non mi sembra una proposta sensata – semplicemente perché è affatto impossibile comprendere il senso di concetti come controllo empirico, conferma, falsificazione, ecc., senza già presupporre quanto qui sarebbe da controllare (ovvero le presupposizioni costituite dalle quattro pretese di validità e dalla possibilità di principio della loro riscattabilità)» (Apel 1997: 135-136).

Ciò che l'autore tedesco vuol dirci è che probabilmente regole controllabili empiricamente «anche quando rinviino a *invarianti empirico-generali (universali antropologici)*» (Apel 1997: 136) non possono essere assunte come qualcosa di inaggirabile, ma al contrario come aggirabili, così come aggirabili sono le regole di fatto costituenti, storicamente e culturalmente, la diversità delle presupposizioni di sfondo dei mondi vitali.

Si rileva così un'insufficiente demarcazione metodologica tra il territorio specificamente filosofico e quello empirico descrittivo della scienza ricostruttiva: «Habermas dovrà, prima o poi, decidere se persistere nell'inconsistenza o riconferire alla filosofia la sua genuina *funzione fondativa*, collegata alla difesa di pretese di validità universali *a priori e autoreferenziali*» (Apel 1997: 220).

La più significativa "ambiguità" della strategia speculativa di Habermas sarebbe data da un'interna oscillazione: da un lato, egli ha portato alla luce nel principio discorsivo il "punto archimedeo" della necessità della fondazione «non più impostata in chiave *ontologica o di filosofia della coscienza, bensì pragmatico-trascedentale*»; per altro verso però, «egli non sfrutta fino in fondo questa scoperta entro l'architettonica del suo pensiero» (Apel 1997: 223). La preoccupazione di un allontanamento dalla prassi storica lo conduce, in ultima analisi, ad ancorare alle risorse del mondo della vita non solo l'apertura di senso, ma anche la giustificazione della validità e la controllabilità dei nostri asserti.

Habermas, del resto, intravede nella radicalità delle tesi apeliane un residuo della vecchia metafisica della soggettività. L'allievo di Adorno ha respinto il concetto di *fondazione ultima* proprio perché stimato come un ritorno a figure teoretiche di matrice metafisica, che opererebbero un incongruente ritorno al paradigma filosofico prelinguistico. Una trasformazione in chiave di intersoggettività comunicativa del soggetto della conoscenza dovrebbe invece lasciare dietro di sé sia

i canoni della metafisica coscienzialistica, sia l'ancoramento a un punto archimedeo assicurante.

Per Apel però il *trend* detrascendentalizzante, a cui anche la teoria di Habermas rimanda, non è necessariamente collegabile all'istanza di un superamento della metafisica della soggettività: la fondazione ultima in quanto «accertamento riflessivo dei principi della ragione» – implicitamente riconosciuti da ogni argomentante – evitando il concetto tradizionale di fondazione come derivazione di qualcosa da qualcos'altro, «si è lasciata alle spalle anche il *fondamentalismo*, nel senso della *metafisica dogmatica*, costretta a ricorrere ad assiomi supposti come evidenti» (Apel 1997: 234).

La fondazione pragmatico-trascendentale, che muove dall'autoriflessione dell'inaggirabile discorso argomentativo, «non coincide con il richiamo di Habermas alla *situazione comunicativa esistente nel mondo della vita*, documentabile in chiave di analisi descrittiva del linguaggio» (Apel 1997: 227). I recenti riferimenti di Habermas alla detrascendentalizzazione, come impegno e compito di un pensiero post-metafisico che si sgancia dal coscienzialismo, non fanno altro che avvalorare le intuizioni apeliane riguardo un progressivo allontanamento della pragmatica universale da possibili argomenti trascendentali centrati sull'autoconsistenza del *logos* (Mazzocchio 2011: 208). Se, da una parte, il grande merito della filosofia di Habermas sta nell'aver offerto una significativa teoria delle strutture della razionalità, attraverso quel cambiamento di paradigma semplificabile nel passaggio dalla coscienza all'intersoggettività linguistica (Pedroni 1999: 43 e ss.), dall'altra parte, tale impresa non riconosce un sufficiente valore di autonomia alla filosofia rispetto alle scienze storico-sociali, né dal punto di vista della specificità metodologica, né rispetto alla particolarità dell'oggetto d'indagine.

La richiamata difesa habermasiana dell'universalità post-metafisica della ragione, in quanto via critica di trascendimento del contestuale, non arriva al punto, per noi decisivo, di riconoscere al sapere filosofico, oltre la dichiarata "custodia della razionalità", specificità fondativa.

In questo quadro il paradigma etico-discorsivo potrebbe vacillare di fronte alle sfide aperte dalle nuove modalità della relazione intersoggettiva digitale e della comunicazione con entità non-umane.

BIBLIOGRAFIA

Apel, K.-O. 1997. *Discorso, verità, responsabilità. Le ragioni della fondazione: con Habermas contro Habermas*. Traduzione italiana a cura di V. Marzocchi. Napoli: Guerini.

Apel, K.-O. 2004. *Verità come idea regolativa*. Traduzione italiana a cura di A. Taraborrelli. In «La Cultura», 2 (2004), pp. 259-282.

Apel, K.-O. 2005. *Cambiamento di paradigma. La ricostruzione trascendentalermeneutica della filosofia moderna*. Traduzione italiana a cura di M. Borrelli. Cosenza: LPE.

Badiou, A. 2011. *Le réveil de l'histoire*. Fécamp: Nouvelles Editions Ligne.

Corchia, L. 2022. *Habermas e i social network: la fine delle sfere pubbliche riflessive?* In L. Gherardi (ed.), *Lezioni brevi sull'opinione pubblica. Nuove tendenze nelle scienze sociali*. Milano: Meltemi, pp. 175-187.

Calloni, M. - Nicoletti, M. - Petrucciani, S. (edd.) 2021. *Intervista a J. Habermas: Filosofia, pensiero post-metafisico e sfera pubblica in cambiamento*. In «Rivista Italiana di Filosofia Politica», 1 (2021), pp. 137-154.

Gebauer, R. 2003. *Letzte Begründung, eine Kritik der Diskursethik von J. Habermas*. München: W. Fink Verlag.

Grundmann, T. 1992. *Transzendentale Argumentation. Aspekte analytischer Transzendentalphilosophie*. Tübingen: Dissertazione.

Habermas, J. 1986. *Teoria dell'agire comunicativo*. Traduzione italiana a cura di P. Rinaudo. Bologna: Il Mulino, Vol. I.

Habermas, J. 1989. *Etica del discorso*. Traduzione italiana a cura di E. Agazzi. Roma-Bari: Laterza.

Habermas, J. 2001. *Verità e giustificazione*. Traduzione italiana a cura di M. Carpitella. Roma-Bari: Laterza.

Habermas, J. 2007. *La condizione intersoggettiva*. Traduzione italiana a cura di M. Carpitella. Roma-Bari: Laterza.

Habermas, J. 2021. *Überlegungen und hypothesen zu einem erneuten strukturwandel der politischen Öffentlichkeit*. In M. Seeliger, S. Sevignani (edd.), *Ein erneuter Strukturwandel der Öffentichkeit?*, in «Leviathan», Sonderband 37 (2021), pp. 9-40.

Floridi, L. 2017. *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina.

Rheingold, H. 2013. *Perché la rete ci rende intelligenti*. Traduzione italiana a cura di S. Garassini. Milano: Cortina Editore.

Marzocchi, V. 2001. *Ragione come discorso pubblico. La trasformazione della filosofia in K.-O. Apel*. Napoli: Liguori.

Mazzocchio, F. 2011. *Le vie del logos argomentativo. Intersoggettività e fondazione in K.-O. Apel*. Milano: Mimesis.

Pedroni, V. 1999. *Ragione e comunicazione. Pensiero e linguaggio nella filosofia di K.-O. Apel e J. Habermas*. Milano: Guerini.

Schwering, M. 2014. *Jürgen Habermas interviewed: Internet and Public Sphere. What the Web Can't Do*. In «Reset-dialogues on Civilizations», 24 Luglio 2014.

Spinaci, R. 1997. *Razionalità discorsiva e verità*. Genova: La Quercia.

Stiegler, B. 2015. *Platone digitale. Per una filosofia della rete*. Traduzione italiana a cura di P. Vignola. Milano: Mimesis.

Taddio, L. - Giacomini, G. 2020. *Filosofia del digitale*. Milano: Mimesis.

Zuboff, S. 2023. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Traduzione italiana a cura di P. Bassotti. Roma: Luiss University Press.

Wellmer, A. 1991. *Wahrheit, Kontingenz, Moderne*. Frankfurt: Suhrkamp.

Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 61-80
ISSN: 1825-5167

AI ENTERS PUBLIC DISCOURSE: A HABERMASIAN ASSESSMENT OF THE MORAL STATUS OF LARGE LANGUAGE MODELS

PAOLO MONTI
Università degli Studi di Milano Bicocca
Dipartimento di Scienze Umane per la Formazione "Riccardo Massa"
paolo.monti@unimib.it

ABSTRACT
Large Language Models (LLMs) are generative AI systems capable of producing original texts based on inputs about topic and style provided in the form of prompts or questions. The introduction of the outputs of these systems into human discursive practices poses unprecedented moral and political questions. The article articulates an analysis of the moral status of these systems and their interactions with human interlocutors based on the Habermasian theory of communicative action. The analysis explores, among other things, Habermas's inquiries into the analogy between human minds and computers, and into the status of atypical participants in the linguistic community such as genetically modified subjects and animals. Major conclusions are the LLMs seem to qualify as authors that originally participate in discursive practices but do display only a structurally derivative form of communicative competence and fail to meet the status of communicative agents. In this sense, while the contribution of AI writing systems in public discourse and deliberation can support the process of mutual understanding within the community of speakers, the human actors involved in the development, use, and diffusion of these systems share a collective responsibility for the disclosure of AI authorship and verification and adjudication of validity claims.

KEYWORDS
Large Language Models, Jürgen Habermas, Moral status, Responsibility, Public discourse

1. INTRODUCTION: AI HAS ENTERED THE CHAT

Generative AI systems are becoming increasingly effective at producing original texts based on inputs about topic and style provided in the form of prompts or questions. Latest AI technologies, especially Large Language Models (LLMs) like GPT-4 by OpenAI or PaLM 2 by Google, develop their capabilities through a machine learning process that feeds on fragments of public discourse as found in internet webpages, books, and articles. These systems have become increasingly successful at engaging in areas of complex and specialized writing, like poetry or academia, with results sometimes indistinguishable from those of human writers.

The diffusion of AI-generated discourse into the public sphere poses serious and unprecedented normative questions. Unlike other uses of AI technology, such as "deepfakes" – fabricated videos representing public figures in the act of saying words they never pronounced – AI writing cannot be reduced to a mere instance of forgery operated by human actors with the instrumental assistance of AI-based tools. Instead, it highlights the possibility of important pieces of the public conversation being originated by non-human authors and finding their way into moral-practical discourses about principles, norms, and policies that hold a central place in democratic deliberation among citizens.

Some political uses of LLMs are already in active development, from more experimental exercises, like the production of "toxic" models trained on data from unregulated internet message boards to more systematic analyses of the efficacy of microtargeted political messages written by LLMs and individually directed to social media users.

The problem of the moral status of LLMs as potential participants in public discourse can be fruitfully approached from different philosophical and STS perspectives (Gordon and Gunkel 2021; Sinnot-Armstrong and Conitzer 2021; Redaelli 2023). The limited scope of this article aims to highlighting which insights can be drawn from Habermasian theory and what status can be assigned to LLMs that participate in discursive practices with humans in terms of responsibility for what they generate in that context. In recent years, Jürgen Habermas has discussed some of the implications of the new technological infrastructure of communication based on the internet and social media for the public sphere and deliberative democracy (Calloni et al. 2021; Habermas 2023). He has not substantially engaged, on the other hand, with the possibility that digital technologies could soon also produce a new kind of non-human actors of public discourse and deliberation. His vast philosophical project, however, offers relevant conceptual resources to attempt this undertaking as well. This account begins by looking at two areas, mutually connected but articulated in Habermas's works at different times: first, the tension between the communicative origin of the person and the naturalistic understanding of the mind as a computer (2); second, the moral status of atypical members of the community of communicants like genetically modified individuals and animals (3). We will explore these two areas, to then attempt a characterization of the hybrid status of LLMs within our discursive practices (4) and outline a preliminary normative account of the moral responsibilities at play when fragments of discourse produced by LLMs enter public discourse and deliberation (5). The conclusions will briefly discuss how this account may fit within a larger consideration of the future impact of generative AIs on the ethics of democratic citizenship (6).

2. BETWEEN HUMAN MINDS AND MACHINE LEARNING

In *Between Naturalism and Religion*, Habermas argues that the social formation of the person through the practice of exchanging reasons with peers seems irreducible to the merely naturalistic understanding of the mind that is suggested by the frequently advanced analogy between human minds and computers. The genesis of the human mind, Habermas notes, lies in the interplay between «the perspective of an observer on what is going on in the world with the perspective of a participant in interaction» with others (Habermas 2008: 171). The "subjective mind" of the individual arises within a communicative process of understanding that constantly generates, in parallel, a linguistic "objective mind" of materially embodied symbols that is to some extent independent of its individual speakers. These subjective and objective sides of the human mind are both distinct and co-implicated. Distinct since, «On the one hand, objective mind evolved out of the interaction between the brains of intelligent animals who had already developed the capacity for reciprocal perspectivetaking [...] On the other hand, the "objective mind" claims relative independence vis-à-vis these individuals, since the universe of intersubjectively shared meanings, organized according to its own grammar, has taken on symbolic form» (Habermas 2008: 174-175). These "two minds" are, however, also tightly co-implicated, since:

> These meaning systems can, in turn, influence the brains of participants through the grammatically regulated use of symbols. The "subjective mind" of those individuated participants in shared practices develops only in the course of the socialization of their cognitive capacities. This is what we mean by the self-understanding of a subject who can step into the public space of a shared culture. As actors, they develop the awareness of being able to act one way or another because they are confronted in the public space of reasons with validity claims that challenge them to take positions.

Our self-understanding as free subjects emerges out of this interplay between the subjective and the objective mind, since «conscious participation in the symbolically structured "space of reasons" jointly inhabited by linguistically socialized minds is reflected in the accompanying performative sense of freedom» (Habermas 2008: 173). As rational subjects, our agency finds motivations «in this dimension and follows logical, linguistic, and pragmatic rules that are not reducible to natural laws» (Habermas 2008: 173). This opens up the possibility of separating the kind of causal connection that the naturalistic image of the world envisions between the individual brain and its corresponding individual mind, from a distinct form of "mental causation" that arises from the cognitive inputs that the symbolic "objective mind" feeds to the individual by stimulating judgments and considerations.

Habermas is intent in specifying that these two understandings of the life of the mind are not mutually exclusive but cannot, at the same time, be entirely reduced to the naturalistic side. They are rather the outcome of an inescapable linguistic

dualism between the perspective of the observer, as reflected in the scientific outlook, and the perspective of the participant articulated in our practical understanding of the mind. In this sense, he argues, the computer analogy that is often invoked to assimilate our thinking to the inner workings of computing machines is fundamentally flawed because it misses «the socialization of cognition that is peculiar to the human mind» (Habermas 2008: 175).

This brings us closer to our first step in the assessment of the moral status of generative AI systems, especially LLMs. Habermas is stark in remarking that the intersubjective, symbolic experience that animates the human mind is irreducible to the image of software running on computer hardware. At the same time, he does not rule out entirely the ICT analogy, as he notes:

> Talking of the mind "programming" the brain evokes metaphors from computer language. The computer analogy puts us on the wrong track insofar as it suggests the Cartesian model of isolated conscious monads [...] However, the mistaken metaphor is not "programming." Clearly, at the evolutionary level of human nature and culture, a symbolically materialized layer of intersubjectively shared, grammatically structured meanings emerges from the intensified interaction among conspecifics. Although the physiology of the brain does not permit any distinction between "software" and "hardware," the objective mind, in contrast to the subjective mind, can acquire the power to structure the individual brain. (Habermas 2008: 175).

This observation rules out a tight analogy between human minds and computers, but it also leaves the door open for a more nuanced stance when it comes to generative AI systems. LLMs escape, at least to some extent, the narrow formula of the individual hardware that runs its own pre-established software, as they are based on semi-automated learning processes fed by the same kind of socially shared "objective mind" that "programs" the individual human brain. Specifically, in the case of LLMs, the machine learning process trains the system on immense textual resources stored on the internet, on social media, and in the digital version of books and journals. Perhaps not surprisingly, then, the output of these AI systems is close to the kind of authorship and apparent creativity that we generally expect from human speakers, as they do not simply execute pre-programmed functions by rather "respond" to human prompts by articulating original pieces of writing. If we accept this distinction as consistent with the Habermasian stance on computers and programming, we can preliminarily note that, while from the "perspective of observation" humans and AIs clearly are two entirely different kinds of systems operating on their own rules and mechanics, from the "perspective of participation" the difference is much more subtle.[1]

[1] On the implications of this linguist dualism, Habermas notes: «The inescapable linguistic dualism compels us to assume that the complementarity of anthropologically deep-seated epistemic perspectives arose concurrently with the sociocultural form of life itself. The coeval emergence of the observer and participant perspectives would provide an evolutionary explanation for why the

Trained on the textual socialization of human cognition and displaying authorial properties, LLMs still lack, however, other salient traits that Habermas ascribes to humans as competent speakers in the community of communicants. AI participation in human discursive practices does not seem to lead to the formation of a sentient "subjective mind" out of their cognitive experience of socialization (Véliz 2021; Schwitzgebel 2023). This determines the novel situation of a new kind of actor that appears, in some relevant ways, capable of contributing to discursive practices as an author of discourse while, at the same time, not being fully responsible for its participation.

For Habermas, participation in discursive practices is a central aspect of becoming responsible agents: «People enter the public space of reasons by being socialized into a natural language and by gradually acquiring the status of a member of a linguistic community through practice. Only with the ability to participate in the practice of exchanging reasons do they acquire the status of responsible authors of actions that is definitive of persons as such, i.e. the ability to account for themselves toward others». This connection is rooted in the methodological primacy «enjoyed by the intersubjectively shared meanings embodied in joint practices in the sequence of explanation prior to internal states of the individuals involved» (Habermas 2008: 205). The process of becoming responsible agents is accompanied by a distinct reflexive aspect, specifically in the form of «a reflexive stability of our consciousness of freedom» (Habermas 2008: 208) rooted in the self-awareness that our convictions and our actions are grounded in meanings and reasons that inhabit ourselves and are shared, transmitted and revised within a community of communicants we belong to.

The reflexive nature of this linguistic-cultural genesis of human identities, Habermas notes, entails more than just the ability to draw from some pre-defined repertoires of signification and make use of them to articulate and justify actions. The "objective mind" is for humans a space of socialization, where the interaction with interpersonal semantic resources within shared practices is the basis for a self-conscious process of identification and projection into the future. Specifically, he argues:

> Only by growing into an intersubjectively shared universe of meanings and practices through socialization can persons develop into irreplaceable individuals. This cultural constitution of the human mind explains the enduring dependence of the individual on interpersonal relations and communication, on networks of reciprocal recognition, and on traditions. It explains why individuals can develop, revise, and maintain their self-understanding, their identity, and their individual life plans only in thick contexts of this kind (Habermas 2008: 296).

meanings that become accessible in our encounters with second persons do not admit of exhaustive objectification through the instruments of natural science» (Habermas 2008: 208).

This kind of reflexive consciousness and sense of identity is not a property that can be currently attributed to LLMs, at least based on how they operate and the kind of linguistic output they display. These preliminary considerations, then, suggest that the "perspective of participation" in practices is where the interaction between humans and AIs highlights both their common traits – as in the emergence of discursive capacities out of the learning process upon the "objective mind" of symbolic linguistic repertoires – and their differences – when it comes to the emergence of the intentional, desiring subjective mind of the human participants and the iterative, stochastic simulation that fuels the output of LLMs.[2] This, however, leaves substantially intact the problem of what kind of moral status should be attributed to this new kind of actor.

3. THE MORAL STATUS OF ATYPICAL PARTICIPANTS IN PUBLIC DISCOURSE

The question about the uncertain moral status of some specific kinds of participants in human interactions emerges within Habermas's work in at least two instances: the case of human subjects that have been genetically modified before birth and the case of animals who partake in our lives and daily practices. The two cases are obviously quite different, and they do not immediately overlap with the case of LLMs joining deliberative practices, but they are nonetheless relevant to explore the boundaries of Habermasian discourse ethics when confronted with fringe cases and atypical actors.

In *The Future of Human Nature*, Habermas points out that the moral status of genetically modified humans would be problematic insofar as the "genetic programming" (Habermas 2003: 63) artificially determines their subjectivity and their capabilities, thus putting them into a structurally unequal position within society.[3] This would in fact create an unprecedented rift in the evolution of horizontal, democratic relationships among humans: "Up to now, only persons born, not persons made, have participated in social interaction. In the biopolitical future prophesied by liberal eugenicists, this horizontal connection would be superseded by an intergenerational stream of action and communication cutting vertically across the deliberately modified genome of future generations" (Habermas 2003: 65). In other words, the moment the nature of some participants

[2] Whether one subscribes or not to the definition of LLMs as merely "stochastic parrots" (Bender et al. 2021), their inner workings are pretty commonly recognized to be a simulation of discourse achieved through a statistically based form of learning that differs substantially from the development of linguistic capacities in human subjectivities.

[3] This seems to suggest that the difference between genetic modification and education is akin to the difference between programming a computer, in the sense described by Habermas, and growing within a culture to become a free and competent participant in its conversations.

is artificially pre-determined by the intentions of others, non-peer relationships within the community of communicants also become inevitable and some members would be stuck in a structurally unequal position from which they cannot exchange roles with others.

This assessment of the relations entertained by genetically modified humans as inevitably uneven interactions offers an interesting perspective from which to look at the status of artificially made participants to discursive interactions like the LLMs. It is in fact, for Habermas, an issue that is deeply connected with the foundations of discourse ethics, insofar as it highlights that the moral space is defined concurrently by the equal form of dependence of all speakers from the linguistic structure of communication and by their active involvement in reflexively and cooperatively establishing the ethical boundaries of their process for reaching understanding and self-understanding:

> The *logos* of language escapes our control, and yet we are the ones, the subjects capable of speech and action, who reach an understanding with one another in this medium. It remains "our" language. The unconditionedness of truth and freedom is a necessary presupposition of our practices, but beyond the constituents of "our" form of life they lack any ontological guarantee. Similarly, the "right" ethical self-understanding is neither revealed nor "given" in some other way. It can only be won in a common endeavor. From this perspective, what makes our being-ourselves possible appears more as a transsubjective power than an absolute one. [...] As soon as the ethical self-understanding of language using agents is at stake *in its entirety*, philosophy can no longer avoid taking a substantive position (Habermas 2003: 11).

The special kind of "language using agents" represented by generative AI systems displays noteworthy capacities and producing outputs within the linguistic structure of communication, but disconnected from a comprehensive "form of life" shared with their human interlocutors that could provide a shared basis of engagement in a "common endeavor". The engagement in a lifeworld shared with others[4] is crucial in defining the profile of the moral subjects of discourse ethics, as they enter into a perspective of universal mutual recognition by reflecting on the normative

[4] It is interesting to notice that Habermas's articulation of the notion of lifeworld is also indebted to its Arendtian formulation. In an article published in 1977, Habermas reads Arendt through the lens of his developing theory of communicative action as follows: «the basic communicative action is the medium in which the intersubjectively shared life-world is formed. It is the "space of appearance" in which actors enter, encounter one another, are seen and heard. [...] In communication, individuals appear actively as unique beings and reveal themselves in their subjectivity. At the same time they must recognize one another as equally responsible beings, that is, as beings capable of intersubjective agreement – the rationality claim immanent in speech grounds a radical equality. Finally, the life-world itself is filled, so to speak, with praxis, with the "web of human relationships." This comprises the stories in which actors are involved as doers and sufferers» (Habermas 1977: 8). To our purpose, this commentary is helpful to highlight how consistently tight the link among communication, responsibility, and embodied presence in a shared life word appears in Habermas's inquiry. See also Arendt 1998: 189.

implications of the presuppositions implicit in their local experiences of communicative engagement with others:

> The ideas of justice and solidarity are already implicit in the idealizing presuppositions of communicative action, above all in the reciprocal recognition of persons capable of orienting their actions to validity claims». Of course, the normative obligations that children assume in virtue of the mere form of socializing interaction do not of themselves point beyond the limits of a concrete lifeworld (of the family, the clan, the city, or the nation). These barriers must first be breached in rational discourse. Arguments by their very nature point beyond particular individual lifeworlds; in their pragmatic presuppositions, the normative content of presuppositions of communicative action is generalized, abstracted and enlarged, and extended to an ideal communication community encompassing all subjects capable of speech and action (Habermas 1994: 50).

In the case of LLMs, the fundamental connection between arguments and "individual lifeworld" that is typical of communicative action is remarkably absent. LLMs are trained upon massive text corpora developed by countless individuals based on their own lifeworld, but as a system, they generate new fragments of discourse without being anchored to any specific lifeworld themselves. In this sense, the whole universalizing process is barred by the absence of a conspicuous link between an individual lifeworld and the speaker's reflexive consciousness of it as shared with other communicative partners. As Habermas observes, a display of cognitive and decision-making capabilities it is not sufficient to define the moral status of a person, since «[o]nly when at least two people encounter each other in the context of an intersubjectively shared lifeworld with the goal of coming to a shared understanding about something can – and must – they mutually recognize each other as *persons capable of taking responsibility for their actions* (*zurechnungsfiihige Personen*). They then impute to each other the capacity to orient themselves to validity claims in their actions» (Habermas 1994: 66).

The problem of "orienting their actions to validity claims" emerges, in different terms, for the designers of LLMs, in the form of what is generally designated in the literature as the value alignment problem, so as a problem intrinsic to the development of AI systems that need to identify relevant human values that are expected to guide the outcomes of the systems, implement these values into the machine learning process and assess that the output of the systems is consistent with those values (Arnold et al. 2017; Gabriel 2020; Christian 2020). But this kind of value aligning process seems quite far for the notion of self-orientation assigned by Habermas to human agents, since to achieve value alignment the identification of values needs to emerge from human actors and the assessment element is also largely dependent on human insight about AI outputs, such as in the case of Reinforcement Learning from Human Feeback (Knox and Stone 2011; Christiano et al. 2017; Kasirzadeh and Gabriel 2023). There are arguably elements of self-orientation insofar as AI systems become increasingly capable of achieving a more

"humanly aligned" orientation. Still, self-orientation as based on a reflexive assessment of their position within a community of speakers seems still definitely far from what LLMs are currently capable of expressing and the outputs of generative AIs are still largely "policed" through content filters introduced by the system developers to make sure that as certain words or requests are presented by human users, the response will be a pre-programmed no go (Derner and Batistič 2023) or by pre-filtering the training data, which in any case still does transmit human-biases into the learning process (Schramowski et al. 2022). In any case, improved AI outputs would still not meet the threshold of a fully moral form of self-orientation, since, as Habermas specifies, «In behaving truthfully I do not merely refrain from deception but at the same time perform an act without which the interpersonal relation between performatively engaged participants in interaction dependent on mutual recognition would collapse» (Habermas 1994: 66). Among moral persons, the orientation to validity claims is part of the intentional and free agency of all participants to the conversation, as they «Act with an orientation to mutual understanding and allow everyone the communicative freedom to take positions on validity claims» (Habermas 1994: 66). In the end, because of their distinct lack of self-reflexivity on a lifeworld and of self-orientation towards the goal of mutual understanding, LLM systems at the moment fall short of belonging to the community of speakers as peers, at least in the way humans are.

Once we acknowledge that, within the framework of discourse ethics, LLMs do not entertain the same moral status as humans, however, we are still faced with the conspicuous experience of their participation in our discursive practices. Habermas's account of the position of animals in his framework may prove useful to offer further clarification. In this regard, he notes that:

> Like moral obligations generally, our quasi-moral responsibility toward animals is related to and grounded in the potential for harm inherent in all social interactions. To the extent that creatures participate in our social interactions, we encounter them in the role of an alter ego as an other in need of protection; this grounds the expectation that we will assume a fiduciary responsibility for their claims. [...] To the extent that animals participate in our interactions, we enter into a form of contact that goes beyond one-sided or reciprocal observation because it is of the same kind as an intersubjective relation (Habermas 1994: 109-110).

These remarks open a space to consider that some non-human subjects may meaningfully participate in human interactions even though they are not peers and they are not structurally able to bear responsibility for their actions. In that context, the moral responsibilities fall on the human participants. The moral responsibilities of humans towards animals, however, according to Habermas, are limited to the scope of the specific interactions between individuals and within particular practices but do not universally bring the other species within the same moral realm (Habermas 1994: 111). In this perspective, participation in human interactions even

from a non-human status is sufficient to establish relations of responsibility, but this responsibility will be entirely up to the human participants. Compared with the case of animals, however, the peculiarity of generative AI in general, and LLMs in particular, is that their participation in our practices is performed specifically in a realm of linguistic creativity and authorship.[5]

4. LLMS AS CO-PARTICIPANTS IN DISCURSIVE PRACTICES

In light of these preliminary analyses of the status of atypical and non-human participants in human practices, I am now going to articulate more in detail how, within a Habermasian framework, AI writing systems can be acknowledged as a special kind of co-participants in human discursive practices, but not as fully communicative agents. In other words, LLMs are not moral or epistemic peers with humans but can still partake in the same public conversations as authors. Their contribution is, in this sense, not merely instrumental: they create original fragments of intelligible discourse that, when introduced into a conversation, can contribute to the process of clarification and understanding among the members of the community of communicants. In instances of public discussion and deliberation, fragments of discourse generated by AI systems can be then very well used to articulate difficult concepts, summarize different perspectives, or even introduce previously neglected ideas. Latest-generation LLMs are also able of expressing real-time interactions within the context of an online chat, which brings their contribution even closer to the same kind of back-and-forth participation typical of argumentative exchanges among peers.

As we mentioned before, it is however unprecedented that the author of a piece of contribution to a discursive engagement is not immediately recognizable also as a responsible moral agent that is, or has been, a human member of the community of speakers.[6] To understand the implications of this decoupling, it is useful to consider how Habermas characterizes, in general, the relation between authors and interpreters of a text, to then suggest a consistent characterization of AI authorship in terms of communicative competence and agency.

In the process of reaching understanding, Habermas argues, the interpreters approach a text based on the assumption that they can understand what the author

[5] This is not to deny that animals seem able to express forms of creative communication and visual performance, but not within the specific realm of human visual and written languages, as generative AIs do.

[6] Naturally, several philosophical and theological traditions have contemplated and reflected upon the possibility of engagements with spiritual and divine interlocutors through the medium of language. This present account just looks at the issue within the scope of Habermasian post-metaphysical thinking. There are, however, interesting analogies and insights that can be drawn from an engagement between AI and theological studies. See also Brittain 2020, O'Gieblyn 2021, Oviedo 2022.

is saying because of a certain grasp of the context within which the text has been conceived and makes sense. This assumption rests on the notion that both interpreter and author raise validity claims on truth, values, and sincerity within a specific context, but that the reasons why they think they can do that are rationally accessible from context to context:

> [O]nly to the extent to which the interpreter also grasps the reasons why the author's utterances seemed rational to the author himself does he understand what the author meant. The interpreter, then, understands the meaning of a text only insofar as he understands why the author felt justified in putting forth certain propositions as being true, in recognizing certain values and norms as being right, and in expressing certain experiences (or attributing them to others) as being authentic. [...] Interpreters cannot understand the semantic content of a text if they do not make themselves aware of the reasons the author could have brought forth in his own time and place if required to do so. (Habermas 1990: 30)

Based on this picture, when the user approaches an AI-generated text as the product of an author, she will still have to rely on the presupposition that at the other side of the conversation there is an interlocutor that produces and understands meaning the same way the interpreter does. LLMs, however, do not operate the same way their human readers and listeners do, based on relatable reasons that make sense within their relationships to a lifeworld and that allow for reasoning about their mutual mental states (Trott et al. 2023). They rather produce an accurate simulation of what an appropriate utterance would be in the face of the textual prompt of the user based on the elaboration of the existing repertoire of appropriate utterances available to the machine learning process. The interpreter can still find in the text some plausible discourse around the topic at stake, but the understanding will happen "as if" the author had reasons the same way the interpreter does:

> For reasons to be sound and for them to be merely considered sound are not the same thing, whether we are dealing with reasons for asserting facts, for recommending norms and values, or for expressing desires and feelings. That is why the interpreter cannot simply look at and understand such reasons without at least implicitly passing judgment on them *as* reasons, that is, without taking a positive or negative position on them. [...] Reasons can be *understood* only insofar as they are taken seriously as reasons and *evaluated*. This is why the interpreter can elucidate the meaning of an obscure expression only if he explains how this obscurity came to be, that is, why the reasons the author might have given in his own context are no longer immediately illuminating for us (Habermas 1990: 30-31).

The conditions of this explanation, however, are different for the interpreter confronted with a text produced by a LLM, since the way obscure or dubious expressions have been generated radically differs from the kind of process that is usually found among human speakers. Anthropomorphic first-person statements do not arise from a personal connection with an individual lifeworld, hallucinations

are unforeseen outcomes of a stochastic process rather than a form of perceptual distortion (Hongbin et al. 2023). This brings into question to what extent LLMs, besides their evident authorial capacities, can be credited with the «know-how of subjects who are capable of speech and action, who are credited with the capacity to produce valid utterances, and who consider themselves capable of distinguishing, at least intuitively, between valid and invalid expressions» (Habermas 1990: 31).

Whereas investigating the issue from the perspective of internal intuitions seems unfruitful, it is instead important to acknowledge that AI systems can be designed as more or less capable of providing "reasons" for their outputs when required to do so, thus supporting the interpreter's job. The problem of the explainability of AI systems is increasingly subject to scrutiny (Preece 2018), with a growing awareness of the ethical dimension of this aspect of their design (McDermid 2021).[7] What the Habermasian perspective suggests here, is that explainability is not only relevant in consequentialist terms, to improve the accuracy and readability of the outputs, but also more substantially to bring the participation of LLMs in discursive practices closer to an expectation of reciprocity that is intrinsic to the process of human understanding. However, even if advances are being made in the field of LLMs explainability, we are still left with the conundrum of their lack of vital relationship with a contextual lifeworld, which is highly problematic within the paradigm of communicative rationality. This becomes apparent by looking more closely at how the linguistic performance of LLMs can – or cannot – be characterized within the Habermasian concepts of *communicative competence* and *communicative action*.

For Habermas, who originally developed the concept inspired by Noam Chomsky's theory of linguistic competence, *communicative competence* is the implicit know-how that speakers have of the implicit rules and presuppositions that make them capable to produce and understand utterances (Habermas 1970; Allen 2019). The most fundamental presupposition, the orientation towards reaching mutual understanding, is specified into validity claims over truth, normative rightness, and sincerity. Every speaker has the implicit expectation that, under suitable conditions, their claims to truth, normative rightness, and truthfulness should be acceptable to all (Habermas 1990: 31). LLMs react to their users' utterances by simulating a human use of language that ordinarily stems out of those presuppositions; by doing so they generate understandable new pieces of discourse. In human communication, each type of validity claim rests on a kind of world relation: relations to the objective natural world, to the intersubjective social world, and to the inner subjective world. LLMs, however, do not entertain the same kind

[7] The importance of explainability to define criteria of accountability and responsibility for AIs is also signaled by the attention it is receiving from policymakers. In the 2020 Assessment List for Trustworthy AI (ALTAI), a document prepared by a group of high level experts set up by the European Commission, accountability is defined as «the idea that one is responsible for their action – and as a corollary their consequences – and must be able to explain their aims, motivations, and reasons».

of world relations as human speakers do and this affects the way they can engage beyond the level of unquestioned everyday communication into the medium of discourse where validity claims are challenged and adjudicated. In the case of current generative AI, truth claims are structurally derivative from those embedded in the textual sources that fed the machine learning process, since the systems have no "experience" of the world or direct access to the natural world to raise and verify truth claims of their own. Normative rightness claims are based on judgments about the appropriateness of speech acts, but these need to rely on the social relationships that the speakers entertain as peers that inhabit a shared lifeworld. Finally, and possibly even more problematic, sincerity claims should be vindicated by the consistency between actions and claimed subjective states of the speakers, but LLMs have no "actions" to display beyond their writing and claims about subjective states are the expression of simulated anthropomorphic approaches to user interaction rather than the reflection of any subjective state we know of. LLMs seem to possess, then, only a structurally derivative communicative competence.

Similarly limited by their lack of lifeworld relations is the ability of AI systems to express proper communicative agency through their participation in human practices. For Habermas, *communicative action* is characterized by the use of discourse to coordinate the actions of its participants (Habermas 1984; Krüger 2019). To some extent, LLMs do adapt to the kind of discursive input they receive from their users, like when correcting previous statements that have been pointed out as erroneous, or when modifying the style of communication based on previous interactions. However, these systems do not self-regulate their own guidelines, which are externally established by their developers and often not even disclosed to the users. Moreover, because of the structurally derivative nature of their communicative competence, LLMs also cannot autonomously adjust their behavior based on contestations to their validity claims, given the absence of direct experiences of the world and of recordable subjective states that can serve as a basis to support and adjudicate those claims. Ultimately, LLMs are capable of manifesting some simulation of communicative agency, but they are not autonomous communicative agents.

Luciano Floridi has influentially argued that the behavior expressed by LLMs is a form of agency without intelligence or understanding (Floridi 2023). His perspective makes sense within a general effort to downplay the kind of "intelligence" that AIs are actually capable of. However, it is noteworthy that, at least within a Habermasian framework,[8] communicative agency without understanding is not even proper agency in the first place. For Habermas, the most paradigmatic form of human agency is indeed the outcome of an interplay between linguistic understanding and autonomous behavior, where each polarity is essential in

[8] I suspect also within several non-Habermasian frameworks, but supporting this conclusion goes beyond the scope of this paper.

defining and structuring the other. LLMs at present appear to be extraordinarily prolific linguistic authors, but not full communicative agents: a crucial decoupling that brings us to the puzzling question of how responsibility for their original utterances should be assigned when they participate in our discursive practices.

5. AI AUTHORSHIP AND MORAL RESPONSIBILITY IN DISCOURSE AND DELIBERATION

The itinerary developed so far allows some tentative suggestions as to how the moral status of LLMs within our discursive practices should be assessed and to what kind of new responsibilities emerge out of the already ongoing discursive interactions between humans and AIs.

From the perspective of their participation in communicative practices, the creative capacity displayed by generative AI systems suggests that their status goes beyond that of mere technical tools in the hands of human speakers. LLMs, then, can be acknowledged as original authors even within specialized discursive practices. In these contexts, they can positively contribute with their extraordinary authorial capabilities to support the ongoing process of mutual understanding among all participants. This kind of contribution could have empowering functions, especially for human participants in public conversations who are otherwise disadvantaged by disabilities, lack of linguistic prowess, or rhetorical education (Kasneci et al. 2023; Pavlik 2023).

At the same time, the limited explanatory capacity, derivative communicative competence, and lack of proper communicative agency of these systems stand in the way of any project to construe them as a new kind of morally responsible subjects that join the community of communicants as peers with their human counterparts. AI systems participate in discourse but not in communicative action. In their discursive interactions, LLMs cannot have, at present, interchangeable roles with human counterparts, a requirement of communicative agents that is fundamental for Habermas to ensure parity among the actors of discursive and deliberative practices. However, human members of the linguistic community can integrate AI authorial contributions into their own communicative agency and vicariously provide the connection with lifeworld relations that AIs lack. Suggestions, insights, images, and information discursively organized by AI systems can resonate with the members of the community of speakers and the living relations with the world, society, and themselves. In turn, those human speakers can operate as moral proxies and stand for the validity of claims raised by AI-generated discourse.

In this perspective, we can still interact within the same discursive practice with human and non-human participants, but the process of mutual understanding needs to adapt substantially to the kind of moral status that each participant

entertains within the community of speakers. For this to happen, it is necessary that all human participants are transparently made aware if they are discursively engaging with a human or an AI and if the author of the piece of discourse they are engaging with is human or not.

On these grounds, I argue that a twofold normative stance on the participation of LLMs in discursive practices can be taken:

> (i) First, based on their authorship properties, the contribution they may bring to the articulation of public discourse, and the enhancement of otherwise discursively disadvantaged participants to the conversation, the involvement of linguistically trained AI systems in our discursive and deliberative practices is acceptable, provided that the human members of the community of speakers (a) take the necessary steps to disclose the authorship of AI contributions and the identity of those who brought them into the conversation (*responsibility as attribution*) and (b) are ready to respond to the contestation of the validity claims that are raised through those contributions, especially when it comes to claims of assertoric truth and subjective truthfulness (*responsibility as answerability*).
>
> (ii) Second, in the field of institutionalized political procedures and formal processes of argumentation and negotiation, the moral call for the disclosure of AI authorship is even more comprehensive and urgent, as the legitimacy of the deliberative procedures is based on the condition of democratic citizens as co-authors of the law, which demands a substantial correspondence between the community of speakers and the community of those who are affected by the normative outcomes of deliberation.

The ensemble of agents involved in bringing about, distributing, and re-circulating the fragments of discourse produced by LLMs collectively shares a responsibility that the AI systems cannot bear themselves, as they are not full communicative agents within the community of speakers, although through their contributions they can participate in important discursive and deliberative practices. It is important to note that the decoupling of authorship and responsibility does not allow for a one-on-one transfer of accountability from the AI system to a singular human subject. The system developers are not responsible for what a Chatbot "says" as if it they said it themselves. Similarly, anyone who brings a text drafted by a generative AI into a public debate is not solely responsible for it in the same way as if they had written it by their own hands before entering the conversation. However, this phenomenon is not the cause of a collapse of responsibility, but rather the premise of a new form of diffused responsibility between company

owners, AI system developers, service users, and social media sharers.[9] Responsibilities will be adjudicated case by case within this relational network based on the agents that played a decisive role in getting that piece of AI discourse into that specific discursive practice.[10]

Without this kind of relational moral context supplied by human speakers, we are left with fragments of "rogue discourse" generated by AIs that by entering our discursive and deliberative practices may determine the effects of communicative agency in the absence of communicative agents that are responsible for them as full members of the community of communicants that share the same world with their peers.

6. CONCLUSIONS: CITIZEN AI?

The perspective of a pervasive presence of generative AI systems within the public sphere of liberal democracies inspires motivated concerns, especially at a moment in history when the advent of social media and the rise of populist movements haven't yet exhausted their momentum and have abundantly shown how deeply technological transformations can affect the political realm (Sunstein 2017; Dijk and Hacker 2018; Urbinati 2019).

It is to be noted, in this sense, that the interpretive framework sketched here, which sees AI systems are creative participants in highly sophisticated human practices without assigning them the full status of moral agents, can be applied also to other kinds of generative AI, beyond the case of LLMs. An obvious example are visual AIs like DALL-E by OpenAI, Midjourney by Midjourney Inc., and Stable Diffusion by Stability AI. The visual creative practices where these systems express their authorial capabilities are not as central in the Habermasian account as the medium of language and discourse are. However, in the digital public sphere, the importance of the production and circulation of images and videos in shaping cultural trends and embodying political agendas can be hardly overstated (Green 2010; Bottici 2014).

In the recent past, when considering the rise of genetical engineering, Habermas had already raised significant concerns about the risk that technological innovation could reshape our moral identities and introduce new forms of political subjectivity in undesirable ways. In a sense, his core preoccupation with this process of technological transformation being appropriated by the strategic and self-serving

[9] Moving from an Aristotelian framework, Mark Coeckelbergh comes to a similar conclusion by articulating a relational account in terms of distributive and collective responsibility from individuals and organizations involved in AI development and use (Coeckelbergh 2020).

[10] Notice that the effort to disclose the AI authorship could also be expressed through the conscious adoption of technical solutions like automatic watermarks for texts produced by LLMs that can be tracked with appropriate tools. As an example, see Kirchenbauer et al. 2023.

logic of capitalism at the expense of the egalitarian and communicative nature of the democratic ethos remains largely applicable to the case of AI:

> The self-understanding of this subject [that intervenes to artificially shape future individualities] now determines how one wants to use the opportunities opened up with this new scope for decision – to proceed *autonomously* according to the standards governing the normative deliberations that enter into democratic will formation, or to proceed *arbitrarily* according to subjective preferences whose satisfaction depends on the market. In putting the question this way, I am not taking the attitude of a cultural critic opposed to welcome advances of scientific knowledge. Rather, I am simply asking whether, and if so how, the implementation of these achievements affects our self-understanding as responsible agents.
>
> Do we want to treat the categorically new possibility of intervening in the human genome as an increase in freedom that requires normative *regulation* – or rather as self-empowerment for transformations that depend simply on our preferences and do not require any *self-limitation*? (Habermas 2003: 12)

Worries about the influence of power and capital over the infrastructure of democratic societies at the expense of agency coordinated through discourse and mutual understanding have been a central focus for Habermas during his entire career. This problematic influence takes everchanging forms and it doesn't seem unthinkable that the next incarnation of "Citizen Kane", who achieves political domination through the use of media, could be in the near future a "Citizen AI" in the shape of one of the global market players that are heavily investing into the development and release to the public of services based on machine learning.

To be fair, it is unclear what the technological advancements in AI technology will produce in the near future. We could very well see soon more direct engagements of LLMs with real-world interactions of some sort, even though it is at least dubious that these interactions will count as vital relations with a lifeworld in the same way as human experience and self-awareness do. It is also possible that the explainability of future AIs will rapidly improve, thus rendering these systems more reliable and responsive partners in our own practices.

In the meantime, it is certainly up to humans to make sure that the insights that emerge out of their embodied circumstances and relational experiences, together with their self-reflexive awareness of the discursive presupposition of their mutual understanding, keep nurturing their moral insight into the responsibilities at stake in all their conversations, including those with their brand-new kind of non-human partner.

REFERENCES

Allen, A. 2019. *Communicative Competence*. In A. Allen & E. Mendieta (Eds.), *The Cambridge Habermas Lexicon*. Cambridge: Cambridge University Press, pp. 47-48.

Arendt, H. 1998. *The Human Condition*. Chicago and London: The University of Chicago Press.

Arnold, T., Kasenberg, D., Scheutz, M. 2017. *Value Alignment or Misalignment—What Will Keep Systems Accountable?*. In «Workshops at the Thirty-First AAAI Conference on Artificial Intelligence».

Bender, E.M., Gebru, T., McMillan-Major A., Shmitchell S. 2021. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. In «FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency», pp. 610–623.

Bottici. C. 2014. Imaginal Politics. Images beyond Imagination and the Imaginary. New York: Columbia University Press.

Brittain, C.C. 2020. *Artificial Intelligence: Three Challenges to Theology*. In «Toronto Journal of Theology», 36, 1, pp. 84-86.

Calloni, M., Nicoletti, M., Petrucciani, S. 2021. Filosofia, pensiero post-metafisico e sfera pubblica in cambiamento. Intervista a Jürgen Habermas. | Philosophy, Postmetaphysical Thinking, and a Changing Public Sphere. An Interview with Jürgen Habermas. In «Rivista Italiana di Filosofia Politica», 1, pp. 137-154.

Christian, B. 2020. The Alignment Problem: Machine Learning and Human Values. New York: W.W. Norton & Company.

Christiano, P.F., Leike, J., Brown, P., Martic, M., Legg, S., Amodei, D. 2017. *Deep Reinforcement Learning from Human Preferences*. In «NIPS 2017 – Advances in Neural Information Processing Systems 30», Long Beach, CA, USA.

Coeckelbergh, M. 2020. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. In «Science and Engineering Ethics», 26, pp. 2051–2068.

Derner, E., Batistič, K. 2023, *Beyond the Safeguards: Exploring the Security Risks of ChatGPT*. Pre-print In: «arXiv:2305.08005».

Dijk, J.A.G.M. van, Hacker, K.L. 2018. *Internet and Democracy in the Network Society*. London and New York: Routledge.

Floridi, L. 2023. AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. In «Philosophy & Technology», 36, art. 15.

Gabriel, I. 2020. *Artificial Intelligence, Values, and Alignment*. In «Minds & Machines», 30, pp. 411–437

Gordon, J.-S. and Gunkel, D.J. 2021. *Moral Status and Intelligent Robots*. In «The Southern Journal of Philosophy», 60, 1, pp. 88-117.

Green, J.E. 2010. The Eyes of the People. Democracy in an Age of Spectatorship. Oxford: Oxford University Press.

Habermas, J. 1970. *Towards a Theory of Communicative Competence*, In «Inquiry», 13, pp. 360–75.

Habermas, J. 1977. *Hannah Arendt's Communications Concept of Power*. In «Social Research», 44, 1, pp. 3-24.

Habermas, J. 1984. *The Theory of Communicative Action*, vol. 1, Boston, MA: Beacon Press.

Habermas, J. 1990. Moral Consciousness and Communicative Action. Cambridge: Polity.

Habermas, J. 1994. Justification and Application. Remarks on Discourse Ethics. Cambridge MA and London: MIT Press.

Habermas, J. 2003. *The Future of Human Nature*. Cambridge: Polity.

Habermas, J. 2008. *Between Naturalism and Religion*. Cambridge: Polity.

Habermas, J. 2023. A New Structural Transformation of the Public Sphere and Deliberative Politics. Cambridge: Polity.

Hongbin, Y., Tong, L. Aijia, Z., Wei, H., Weiqiang, J. 2023. *Cognitive Mirage: A Review of Hallucinations in Large Language Models*. Pre-print In: «arXiv:2309.06794».

Kasirzadeh, A., Gabriel, I. 2023. *In Conversation with Artificial Intelligence: Aligning language Models with Human Values*. In «Philosophy & Technology», 36, 27.

Kasneci, E. et al. 2023. *ChatGPT For Good? On Opportunities And Challenges Of Large Language Models For Education*. In: «Learning and Individual Differences», 103

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., Goldstein, T. 2023. *A Watermark for Large Language Models*. Pre-print In: «arXiv:2301.10226v3».

Knox, W.B., Stone, P. 2011. *Augmenting Reinforcement Learning with Human Feedback*. In «Proceedings of the ICML Workshop on New Developments in Imitation Learning», Bellevue, WA, USA,

Krüger, H. 2019. *Communicative Action*. In A. Allen & E. Mendieta (Eds.), *The Cambridge Habermas Lexicon*. Cambridge: Cambridge University Press, pp. 40-46.

McDermid, J.A., Jia, Y., Porter, Z., Habli I. 2021. *Artificial Intelligence Explainability: the Technical and Ethical Dimensions*. In «Philosophical Transactions of the Royal Society A», 379, 2207.

O'Gieblyn, M. 2021. *God, Human, Animal, Machine. Technology, Metaphor, and the Search for Meaning*. New York: Doubleday.

Oviedo, L. 2022. *Artificial Intelligence And Theology: Looking For A Positive – But Not Uncritical –Reception*. In« Zygon», 57, 4, pp. 938-952.

Pavlik, J.V. 2023. *Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education*. In « Journalism & Mass Communication Educator», 78, 1.

Preece, A. 2018. *Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges*. In «Intelligent Systems in Accounting, Finance and Management», 25, 2, pp. 63-72.

Redaelli, R. 2023. *Different Approaches To The Moral Status Of AI: A Comparative Analysis Of Paradigmatic Trends In Science And Technology Studies*. In «Discover Artificial Intelligence», 3, 25.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C.A, Kersting, K. 2022. *Large Pre-Trained Language Models Contain Human-Like Biases Of What Is Right And Wrong To Do*. In «Nature Machine Intelligence», 4, pp. 258-268.

Schwitzgebel, E. 2023. *AI Systems Must Not Confuse Users About Their Sentience Or Moral Status*. In «Patterns», 4, 8.

Sinnott-Armstrong, W., Conitzer, V. 2021. *How Much Moral Status Could Artificial Intelligence Ever Achieve?* In: S. Clarke, H. Zohny, J. Savulescu (Eds.) *Rethinking Moral Status*. Oxford: Oxford University Press, pp. 269-289.

Sunstein, C.R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press.

Trott, S., Jones, C., Chang, T., Michaelov, J., Bergen, B. 2023. *Do Large Language Models Know What Humans Know?* In «Cognitive Science», 47.

Urbinati, N. 2019. *Me the People. How Populism Transforms Democracy*. Cambridge, MA: Harvard University Press, 2019.

Véliz, C. 2021. *Moral Zombies: Why Algorithms Are Not Moral Agents*. In «AI & Society», 36, pp. 487–497.

INFORMATION ON THE JOURNAL

Etica & Politica/ Ethics & Politics is an open access philosophical journal, being published only in an electronic format.

The journal aims at promoting research and reflection, both historically and theoretically, in the fields of moral, political and legal philosophy, with no preclusion or adhesion to any cultural current or philosophical tradition.

Contributions should be submitted in one of these languages: Italian, English, French, German, Portuguese, Spanish.

All essays should include an English abstract of max. 200 words.

The editorial staff especially welcomes interdisciplinary contributions with special attention to the main trends of the world of practice.

The journal has an anonymous double peer review referee system. Three issues per year are expected.

ETICA & POLITICA/ETHICS & POLITICS POSITION ON PUBLISHING ETHICS.

The Editors of Etica & Politica/Ethics & Politics have taken every possible measure to ensure the quality of the material here published and, in particular, they guarantee that peer review at their journal is fair, unbiased and timely, and that all papers have been reviewed by unprejudiced and qualified reviewers. The publication of an article through a peer-review process is intended as an essential feature of any serious scientific community. The decision to accept or reject a paper for publication is based on the paper's relevance, originality and clarity, the study's validity and its relevance to the mission of the journal. In order to guarantee the quality of the published papers, the Editors encourage reviewers to provide detailed comments to motivate their decisions. The comments will help the Editorial Board to decide the outcome of the paper, and will help to justify this decision to the author. If the paper is accepted with the request of revision, the comments should guide the author in making the revisions for the final manuscript. All material submitted to the journal remains confidential while under review.

Once the author receives a positive answer, he/she should send the final v ersi on of the article since proofs will not be sent to him/her. Etica & Po litica/Ethics & Politics will publish the pa per within twelve months from the moment of the acceptance, and the author will be informed of the publication. The journal is committed to such

standards as originality in research papers, precise references in discussing other scholars' positions, avoiding plagiarism.

E&P takes these standards extremely seriously, because we think that they embody scientific method and are the mark of real scholarly communication.

Since Etica & Politica/ Ethics & Politics is devoted solely to scientific and academic quality, the journal neither has any submission charges nor any article processing charges.

The following guidelines are based on existing Elsevier policies and COPE's Best Practice Guidelines for Journal Editors

1. PUBLICATION AND AUTHORSHIP

EUT (Edizioni Università di Trieste), is the publisher of the peer reviewed international journal Etica & Poitica/Ethics & Politics.

The publication of an article in a peer-reviewed journal is an essential step of a coherent and respected network of knowledge. It is a direct reflection of the quality of the work of the authors and the institutions that support them. Peer- reviewed articles support and embody the scientific method. It is therefore important to agree upon standards of expected ethical behaviour for all parties involved in the act of publishing: the author, the journal editor, the peer reviewer, the publisher.

Authors need to ensure that the submitted article is the work of the submitting author(s) and is not plagiarized, wholly or in part. They must also make sure that the submitted article is original, is not wholly or in part a re- publication of the author's earlier work, and contains no fraudulent data.

It is also their responsibility to check that all copyrighted material within the article has permission for publication and that material for which the author does not personally hold copyright is not reproduced without permission.

Finally, authors should ensure that the manuscript submitted is not currently being considered for publication elsewhere.

2. AUTHOR'S RESPONSIBILITIES

Etica & Politica/Ethics & Politics is a peer-reviewed journal, and Authors are obliged to participate in our double blind peer review process.

Authors must make sure that all and only the contributors to the article are listed as authors. Authors should also ensure that all authors provide retractions or corrections of mistakes.

3. PEER REVIEW AND REVIEWERS' RESPONSIBILITIES

Both the Referee and the Author remain anonymous throughout the "double blind" review process. Referees are selected according to their expertise in their particular fields.

Referees have a responsibility to be objective in their judgments; to have no conflict of interest with respect to the research, with respect to the authors and/or with respect to the research funders; to point out relevant published work which is not yet cited by the author(s); and to treat the reviewed articles confidentially.

4. EDITORIAL RESPONSIBILITIES

Editors hold full authority to reject/accept an article; to accept a paper only when reasonably certain; to promote publication of corrections or retractions when errors are found; to preserve anonymity of reviewers; and to have no conflict of interest with respect to articles they reject/accept. If an Editor feels that there is likely to be a perception of a conflict of interest in relation to their handling of a submission, they will declare it to the other Editors. The other Editors will select referees and make all decisions on the paper.

5. PUBLISHING ETHICS ISSUES

Members of the Editorial Board ensure the monitoring and safeguarding of the publishing ethics. This comprises the strict policy on plagiarism and fraudulent data, the strong commitment to publish corrections, clarifications, retractions and apologies when needed, and the strict preclusion of business needs from compromising intellectual and ethical standards.

Whenever it is recognized that a published paper contains a significant inaccuracy, misleading statement or distorted report, it will be corrected promptly. If, after an appropriate investigation, an item proves to be fraudulent, it will be retracted. The retraction will be clearly identifiable to readers and indexing systems.

6. POLICY ON EDITORIAL NORMS AND STYLE

In accordance with the pluralist principle and the breadth of interests that characterize the editorial line of Eti ca & Politi ca/ Ethi cs & Po l itics, the review leaves the contributors free to choose which style of quotation to adopt in the composition of their article. Of course, this style must conform to one of the two internationally recognized models in the area of scientific publications in the humanities, namely the Chicago A (equivalent to APA) or the Chicago B, both defined by The Chicago Manual of Style, 17th edition. Moreover, the reference to the complete source must contain all the information required by the ISO 690 standard, both if it is provided only in the final bibliography (Chicago A) and if it is provided in a note (Chicago B).

7. PAST ISSUES AND STATISTICS

Past issues with download and visitors statistics for each article are provided by EUT publisher through its official repository.:

http://www.openstarts.units.it/dspace/handle/10077/4673

EDITOR: