Imitation Games: when who you are is reduced to what you do

by Nicola Labanca

In 1950, Alan Turing devised what he called an imitation game to test if a computer could imitate a human. The imitation game has since come to be called the Turing Test and is understood as a test of whether a human being asking questions can tell the difference between answers received from a computer and a human.[1]

The recent release of the generative pre-trained transformer (GPT) chatbot called ChatGPT has sparked a public discussion on the difference, if any, between machines and men. For the most part, the debate on these and other Artificial Intelligence (AI) applications resembles those that accompanied the introduction of previous technologies like the smart phone—mostly confined to a discussion of the societal benefits and costs of a new technology.[2] However, the one element of AI that is distinctive is the claim that it could mimic human intelligence, particularly with the fabrication of an artificial general intelligence (AGI).[3]

Turing's imitation game was designed to test if computer responses to natural language questions could mimic those of a human. I argue here that our societies have now

1 For further information about the Turing test, see e.g., https://en.wikipedia.org/wiki/Turing_test

2 See e.g., Mcnamee, R. (2023). "There is only one question that matters with AI." *Time*. Online article available at https://time.com/6268843/ai-risks-democracy-technology/ or Eloundou, T., Manning, S., Mishkin, P., Rock, D. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. (Ithaca: Cornell University Press). https://arxiv.org/abs/2303.10130v4

3 For further information on AGI, see https://en.wikipedia.org/wiki/Artificial_general_intelligence.

Nicola Labanca, labnicola@gmail.com
Labanca, N. Imitation Games: when who you are is reduced to what you do.
Conspiratio, Fall 2023, p. 156-178.

begun imitation games that go well beyond the Turing test.[4] The social imaginary forming around AI clearly exceeds the simple difference between human and machine answers to questions. It presumes and encourages an on-going process of symmetrization between material artifacts, humans, and animals—whether this concerns their internal workings, what they can produce, and even their moral status.[5] Such symmetrization has far-reaching implications, some of which I explore in this article.[6]

The social imaginary of AI is being enacted by scientists and laypersons through rituals that assume these technologies could become bearers of human desire, thoughts, and feelings.[7] When they are taken literally, these ritualized acts blind people to how their interactions with artifacts produce unwanted consequences. Towards the end of his life, Ivan Illich discerned a transition in Western culture and thought from the "instrument" to the "system" sometime during the late 20th century.[8] Following his insight, I will argue that AI does not belong to the category of a tool or instrument or technology. Rather, to understand AI as an exemplar of a system, I start by referring to the Aristotelian understanding of action and its relation to potential. I then suggest that this relation underwent a decisive change when the "instrument"

4 Biever, C. (2023). "ChatGPT broke the Turing test – the race is on for new ways to assess AI." *Nature* 619, 686-689

5 See e.g., Liao, S. Matthew (2020). "The Moral Status and Rights of Artificial Intelligence," in S. Matthew Liao (ed.), *Ethics of Artificial Intelligence* (New York, 2020; online ed, Oxford Academic, 22 Oct. 2020) or Wojtczak, S. (2022) "Endowing Artificial Intelligence with legal subjectivity." *AI & Soc* **37**, 205–213

6 In the last section, I briefly discuss studies of science and technology inspired by Bruno Latour premised on the symmetric treatment of man, machine, and animal.

7 See e.g., Chalmers, D.J. (2023). *Could a Large Language Model be Conscious?* (Ithaca: Cornell University Press). arXiv:2303.07103

8 Cayley, D. (2005). *The Rivers North to the Future. The Testament of Ivan Illich as told to David Cayley*. (Toronto: House of Anansi Press), Ch. 4

emerged as a category in the 12th century, which also changed the very nature of both human action and human potential. Over the age of instruments, the scope for human action to be delinked from human potentiality vastly increased, in proportion to the expansion of large-scale technologies. However, the emergence of cybernetic system science in the mid-20th century further and fundamentally altered their relation and their nature yet again. I argue that this alteration is the outcome of the idea and design of systems, which is predicated on the claim that purposive behavior can be equated to negative feedback loops, and that men, animals, and machines are therefore similar from the scientific point of view. Crucially, conceptualizing purposive behavior with the logic of feedback loops removes action from the sphere of human potentiality. Instead, humans are reduced to the outcome of circular actions—they are what they do within negative feedback loops. AI is imagined as human only because the cord between an original human potentiality and human action has been cut through described transformations. When humans are understood as no more than action within negative feedback-loops, they lose the potential to be forgiven or redeemed and are no longer accountable for what they do.

The intrinsic contradiction of Instruments

AI is still often imagined as an instrument. Instruments or tools or technologies are widely understood as means designed to achieve specific ends. They are devices thought to exhibit a degree of autonomy but in a way that allows them to be controlled by their users. This intrinsic contradiction is widely accepted. For example, all guns are designed to fire bullets (its autonomous ability) but what they shoot at is dependent on the shooter (its dependence). In this section of the article, I will show that such an instrumental understanding of equipment emerges in western imaginary around the

12th century; that it represents a mutation of an ancient view of the slave in Greek thought; and that this long history has shaped the supposed similarity between early forms of AI and human characteristics.

The intrinsic contradiction of so-called good and old-fashioned AI (GOFAI)

The best approximation of the technological rituals and myths shaping AI as an instrument is *symbolic AI* which dominated research and development in that field from the 1950s to the 1980s.[9] Symbolic AI is also called good and old-fashioned AI (GOFAI) because it is based on human-readable representation of problems through symbolic reasoning and logic.[10] It is typically programmed through if-then rules and statements that establish relationships between inputs and outputs. It produces such applications as expert systems, automated theorem provers, planning and scheduling systems, and so on. Symbolic AI is meant to be totally transparent. Accordingly, its developers can inspect the logic behind the machine's decisions by following the instructions line by line and investigating its errors down to the most minimal details of the code. Symbolic AI therefore offers a paradigmatic example of digital artifacts programmed for any end-user to produce outputs from the feasible class. However, in so far as it is thought to exhibit a kind of autonomous intelligence, symbolic AI is also supposed to enable the generation of specific outputs beyond the ability of the programmers or the users. In this sense, GOFAI contains the intrinsic contradiction characteristic of all instruments. By investigating the historical roots of the idea of an instrument, we can appreci-

9 Haugeland, J. (1985) *Artificial Intelligence: The Very Idea* (Cambridge, MA, MIT Press).

10 On this, see e.g., Boden, M. (2014). GOFAI. In K. Frankish & W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 89-107). (London: Cambridge University Press).

ate how symbolic AI came to be thought of as an instrument endowed with its own intelligence.

The slave as an animate instrument

Ivan Illich sketched an original and insightful account of the historicity of the instrument.[11] Deepening the idea that the instrument was not a ahistorical given, Giorgio Agamben showed that the idea of a means that can be employed to achieve specific ends took hold about eight centuries ago and further that the intrinsic contradiction of instruments was rooted in the ancient understanding of the slave as an *animate instrument*.[12] Agamben focuses on one of the three types of relationships that define a family in Aristotle's *Politics*: the despotic relation between master and owned slaves.[13] Aristotle defines the slave as a being who "while being human, is by its nature of another and not of itself." To justify the necessity of despotic command among animate beings Aristotle relies on analogies with inanimate things. Whereas some instruments are inanimate—like the shuttle used to weave or the plectrum to play the lyre, others are animate—like the lookout person in a ship. The slave serves as an exemplar of instruments that, by definition, *cannot* accomplish their proper work by anticipating the will of others. To contrarily admit the possibility of autonomous will in slaves would be like admitting a shuttle could weave cloth by itself or a plectrum could play the lyre.[14]

11 See Cayley, D. (2005). *The Rivers North to the Future. The Testament of Ivan Illich as told to David Cayley*. p. 171 and Heron, N. (2017). *Liturgical Power: Between Economic and Political Theology*. (New York: Fordham University Press). Chapter 4.

12 Agamben G. (2016). *The Use of Bodies*. (Stanford: Stanford University Press), chapters 1 and 7.

13 Agamben refers to Aristotle's "three types of relations [that] define the family: the despotic (*despotikè*) relation between the master (*despotes)* and the slaves, the matrimonial (*gamikè*) relation between the husband and wife, and the parental (*technopoietikè)* relation between the father and the children." See ibidem, p.3.

14 Ibidem, p. 10

According to Aristotle therefore, the master's despotic command over slaves is as natural as that of the soul making use of the body as an instrument or the musician using the plectrum to play the lyre.

Agamben invites the reader to go beyond simplistic criticisms of this view of slavery and to focus on the context in which Aristotle inscribes the question of slavery. Agamben wants to expose the "zone of indifference between the artificial instrument and the living body"[15] in which Aristotle situates the body of the slave. The contradictory character of *animate instrument* was originally attributed by Aristotle only to specific types of humans (i.e., slaves) and *not* to their tools. By extending Illich's studies, Agamben proposes the fruitful hypothesis that this hybrid character of animate instruments–at once both animate and inanimate–may have migrated from slaves to the *instruments themselves* by the 12th century. This migration could then explain how, while remaining under the command of their masters, instruments, including symbolic AI, could be understood as exhibiting some kind of autonomy and intentionality. For this reason, Agamben emphasizes the "constitutive connection between slavery and [modern] technology."[16]

How the rise of instrumental causality equates technology and slaves

The tools that people employed during their everyday life did not receive a lot of attention from ancient philosophers. Hammers, swords, wood, and forges did not have a general category under which they could be subsumed. Illich reminds us that ancient Greeks referred to their tools by employing the term *organon* in a way that did not allow distinguishing between the hammer, the arm holding the hammer, and the

15 Ibidem, p. 23.

16 Ibidem, p. 79

action of hammering.[17] In this sense, tools were not conceptually separated from their users.[18] It is only starting from the 12th century that instruments acquired an independent status by becoming the carriers of a new and special subtype of efficient causation that theologians named *causa instrumentalis*. Following Agamben, I suggest it is the notion of an animate instrument that represent the signifier for conceptualizing instrumental causality and thereby constitutes the connection between the ancient slaves and modern instruments.

Agamben and Illich emphasize it was Scholastic theology which developed the theory of the instrumental cause in the doctrine of the sacraments.[19] The idea of *causa instrumentalis* has been first formulated by Aquinas when describing the role of the celebrant and of the material elements (e.g., water, consecrated oil, etc.) employed in the administration of the sacraments. Aquinas emphasized the function of the sacrament as conferring grace. God is the primary cause for the efficacy of the sacrament, but it produces its effects by means of another element which acted as a secondary instrumental cause. It is worth noticing here that instrumental causality is ascribed not only to the employed material element but also and primarily to the celebrant or minister himself. Whilst the material element participates as an inanimate instrument, the minister participates as an animate instrument. Both instru-

17 See Cayley, D. *The Rivers North to the Future*. p. 172, but also Agamben G. *The Use of Bodies*. p. 13

18 In the *Eudemian Ethics*, Aristotle argues that: "the relations between soul and body, artisan and tool, and master and slave are similar, between the two terms of each of these pairs there is no partnership (*koinonia*); for they are not two, but the former is one and the latter is part of that one, not one itself; nor is the good divisible between them, but that of both belongs to the one for whose sake they exist. For the body is the soul's tool (*organon*) born with it, a slave is as it were a member or tool (*organon*) of his master, a tool (*organon*) is a sort of inanimate slave" (Aristotle, *Eudemian Ethics*, 1241b, 15).

19 See Cayley, D. *The Rivers North to the Future*. p. 183 and Agamben G. *The Use of Bodies*. p. 70.

ments possess the intrinsic contradictions pertinent to instruments as such. Each is supposed to unfailingly transmit and realize the will of the principal agent while also obeying their own specific natures—as for instance, water flows and celebrants speak. Moreover, the efficacy of the ritual is not compromised by the condition or intent of the instrument—polluted water and sinful ministers can baptize as effectively.

Agamben convincingly shows that Aquinas refers precisely to the Aristotelian notion of *animate instrument* to explain how the administrator of the sacraments "comports himself in the mode of an instrument."[20] The contradictory character of the type of causation instruments embody—the fact of realizing the end of the principal agent while acting as a secondary agent which obeys its internal mechanical functioning—was hence first conceived and explained in the context of arguing for ministers as animate instruments in the ritual of administering the sacraments. Ancient Greeks thought that the slave body was an integral part of and non-separable from the master's body.[21] In contrast, for Aquinas, the minister was distinct from God and an instrument for the administration of sacraments. Accordingly, the understanding of modern technology as instruments that are separable and distinct from humans and designed for their use continues to carry the sign of the intrinsic contradiction of animate instrument. From this perspective, symbolic AI can be understood as a paradigmatic example of *absolute* instrumentality where the will of God as the principal agent who commands the minister as a human instrument has been incorporated in the technology itself. Early forms of AI and technology *per se* owe themselves to the separation between instrument and user inaugurated by the notion of *causa instrumentalis* in the 12th century.

20 Agamben G. *The Use of Bodies*. p. 74-75

21 Ibidem p. 18.

How instrumentality has changed human potential and action

The imitation game attending the rise of instrumentality has trapped people of modern societies and their tools within dichotomous views that also informed early AI. Modern humans have been imagined in the model of the Christian God, as beings able to construct any kind of artifact or technology that can convey their will without distortions. At the same time, like the slaves of antiquity, modern people are also imagined as office and job holders–working instruments that faithfully carry out the will of others. However, contrary to slaves, modern workers preserve their autonomy when they exercise the possibility of leaving their jobs. In this modern version of the imitation game shaped by instrumentality, both workers and AI embody the contradictory characteristic of having the will of a principal agent while also behaving as being an instrument.

Aristotle presented his arguments concerning human making based on his ideas of potentiality and action. Agamben's reflections are very enlightening in thinking about how the ideas of human action and potentiality have changed because of instrumental causality. According to Agamben, human making presupposes a specific type of potentiality and relationship between potentiality and action that cannot be found in animals and things. Though it is a matter he has extensively discussed in many of his writings, it is perhaps most clearly laid out in his *Creation and Anarchy*.[22] There, he discusses in detail the relationship between human making and *resistance*[23] to action by starting from the concept of *potential* as developed in Aristotle and Western philosophy.

22 See e.g., Agamben, G. (2004). *The Open: Man and Animal*. (Stanford: Stanford University Press), Chapter 9; Agamben, G. (1998). *Homo Sacer. Sovereign Power and Bare Life*. (Stanford: Stanford University Press). Agamben, G. (2019). *Creation and Anarchy. The Work of Art and the Religion of Capitalism*. (Stanford: Stanford University Press)

23 Agamben, G. *Creation and Anarchy. The Work of Art and the Religion of Capitalism*. p. 17

The accepted understanding of the relationship between potentiality and actuality is that the former passes into the latter. Accordingly, when playing the piano, the pianist actualizes her potential to play the piano by exhausting the potential to play it. However, according to Agamben, Aristotle has a deeper understanding of potentiality.

To elaborate the meaning of human potentiality and action,[24] Aristotle specifically refers to the potential of those who have already acquired an art or knowledge such as architects, sculptors, and grammarians.[25] Those who are not pianists cannot play the piano. This inability or privation or lack is not what Aristotle is concerned with. Rather he focuses on the accomplished professional and master of an art, say a pianist, who possesses both the potential to play and to not play the piano. When he does not play the piano, Glenn Gould is actively suspending or withholding his potential to play it. This suspension, refusal, or restraint with respect to possible action is well-known. What is less well understood, is the symmetric case of playing the piano. When Glenn Gould is playing the piano, he is not only completing in action what he could potentially do. Rather, while playing he is also actively suspending or resisting his potential to play. In this sense, Aristotelian potentiality is "essentially defined by the possibility of its non-exercise."[26]Aristotelian potentiality is therefore at

24 For Aristotle, act is generally what a being is already at a given moment. For example, a table is a table in act, a child is a child in act, and so on. Potency and potentiality refer instead to that which, at the moment, is not, but which can become. For example, an acorn has the potentiality to become an oak, a child has the potentiality to become an adult. When the acorn becomes an oak, the acorn is an oak in act and no longer in potency. On this, see for example, Aristotle's discussions in his *Metaphysics* and *Physics*. In the arts and human making, potentiality is however more specific than what can be generally identified in nature, in so far as it presupposes the presence of a habit. On this point see e.g., Agamben G. (2016). *The Use of Bodies*. p. 60, and Agamben, G. (2019). *Creation and Anarchy. The Work of Art and the Religion of Capitalism*., p. 19

25 Ibidem, p. 16.

26 Agamben, G. *Creation and Anarchy. The Work of Art and the Religion of Capitalism*, p. 17

work also *during* activity as potential not to act. For this reason, Agamben insists that "mastery preserves and exercises *in action* not its potential to play but its potential not to play."[27] Thus, human action cannot be understood as a simple transition from potentiality to action. Instead, as exemplified by masters of an art, action is exercised by acknowledging and suspending one's potentiality to act during action, by resisting the ever-present potential to do.

Said differently, this kind of resistance is always present in human making in the form of a constant dialectic between a personal and an impersonal element. The impersonal element (i.e., the potential-to, the genius that drives toward work and expression) constantly exceeds the particular subject while the personal element (i.e., the potential-not-to) is constantly exercised by the individual who opposes the impersonal. Through the enactment of suspending action, humans express a kind of second order action that results from the expression of a second order potentiality, which opens to contingency. Through this enactment, they can act on action and thereby express a potential of the potential[28], i.e., the potential to not pass into the act while opening up to the unexpected. In this dialectical process between personal and impersonal elements, subject/object distinctions become deactivated. Thereby, for example, painting becomes painting of painting through the exposition and suspension of gaze in the act of painting, poetry becomes poetry of poetry through the exposition and suspension of language, and so on. What people do is thereby distinctively characterized by the constant possibility of *stepping in and stepping out* from action during action. This possibility entails that persons are never completely defined by their potentialities and actions. Persons

27 Ibidem, p. 19

28 Ibidem, pg. 24

constantly have the possibility to enter and exit the roles that seek to define them, and this can happen during the exercise of these roles.

When seen through the lens of this relationship between action and the potential to-not-do, the change induced by instruments undermines action by severing it from the potential to-not-do. It is obvious that, in general, the age of instrumentality has reduced the potential of individuals to re-direct and interrupt what they do by having multiplied the possibility of automatic actions at unprecedented scales. Compared to previous human tools, modern instruments are relatively autonomous and detached from the people employing them. This increased automatism can be gauged, in part, by a shift from the endosomatic energy to exosomatic energy needed for their operations. Those who operate a machine have little or no capacity to interrupt or re-direct their actions. This diminution of human potentiality is also reflected in the large-scale standardization of action in regimes of industrial production that generate the unwanted systemic effects that Ivan Illich named counterproductivity.[29] Accordingly, we can understand the contradictory social imaginary expressed with early forms of AI as "intelligent machines" to reflect the increased automation and instrumentalization of human action. However, the more recent types of AI raise the possibility that action may no longer be understood as the active suspension of the potential to-not-act and vice versa. This deeper depotentiation of humans deprives them of potentiality by reducing it to self-recursive action within negative feedback loops. This reduction may well be the signature of systems, of which the latest types of AI are an exemplar. [30]

29 See Illich, I. (1976). *Limits to Medicine. Medical Nemesis: The Expropriation of Health*. (London: Penguin Books), p. 215

30 Depotentiation must be exclusively intended here in relation to reduced possibilities for single persons to resist to action while using instruments and not as a

Contemporary AI and integration into systems

Contemporary forms of AI can perform a large variety of tasks. Some of these AI can recognize human speech to make travel reservations in accordance with passenger preferences of routes and prices. Other AI's can steer vehicles autonomously for many miles. Then there are AI programs that can diagnose lymph-node pathologies or call emergency services when they detect a road accident.[31] The examples can be multiplied at will. Crucially, rather than relying on human programmers, many of the emerging AIs perform their tasks using machine learning techniques that consist in the automatic and iterative processing of very large data sets by which they produce their own algorithms without having been instructed on what to do.[32] Recently, it is so-called deep learning techniques that seem to achieve the highest level of performance in a variety of arenas including decisions about medical therapies, selection of job candidates and loan applicants.[33] Deep learning techniques rely on artificial neural networks with weights associated with each network node that automatically adjust to reduce the size of the outcome error.[34] Moreover, improvements in their performance in-

reduction in the number of activities and new applications of instrumentality.

31 For more examples of existing applications see e.g., Russel, S. J. & Norvig, P. (2020). *Introduction to AI: a modern approach*. (New York: Prentice Hall). Chapter 1.

32 Contrary to so-called *supervised learning techniques* where training occurs over pre-labelled data against which machine algorithms performances can be measured and improved (e.g., in images classification), machine-learning techniques are unsupervised or self-supervised learning techniques and do not rely on otherwise very time-consuming data labelling activities. See e.g., Spathis, D., Perez-Pozuelo, I., Marques-Fernandez, L., Mascolo, C. (2022) "Breaking away from labels: The promise of self-supervised machine learning in intelligent health", *Patterns*, 3(2) pp.1-6.

33 Wani, M. A., Palade, V., (2023). *Deep Learning Applications*. Springer Nature, Vol 4.

34 For a comprehensive overview of deep learning techniques and applications, see Sarker, I.H (2021). "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions." *SN COMPUT. SCI.* 2, 420.

creases as the intricacy of the network increases—in terms of the number of nodes and density of the interconnections between them. But this implies that the better deep learning AI machines perform, the less we can understand them or specify the steps by which they arrived at the decisions taken.[35] In this way, deep learning machines become endowed with a specific opacity.[36] Paradoxically, some now argue that this opacity of how machines work mimics the obscurity of the mechanisms by which the brain "is capable of bi-directional travel between exemplars and abstractions."[37]

The very large amount of data that are now employed for this kind of machine training[38] and the possibility of employing models applying learned patterns from one task to another[39] makes it very likely that a *single* general-purpose AI will soon appear. For instance, one can imagine the development of one AI that generates content, translates and summarizes texts, produces reports from notes, drafts emails, responds to queries and questions, creates new text, images, audio and visual content all based on user inputs such as text

35 On this point, see https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained or https://www.relativity.com/blog/paradox-of-the-black-box-inverse-relationship-between-ai-accuracy-and-transparency/

36 Machine learning techniques can become black-boxed and impenetrable even to their programmers as they can employ thousands of millions of connections that interact with one another in complex ways. Boge, F.J (2022). "Two Dimensions of Opacity and the Deep Learning Predicament." *Minds & Machines* 32, 43–75 (2022)

37 See Buckner, C. (2018) "*Empiricism without Magic: Transformational Abstraction in Deep Convolutional Neural Networks," Synthese, 195*(12), 5339–5372.

38 Notice that this kind of training is purely operational and does not entail any kind of cognition. A widely quoted definition of how machines learn is for example the one produced by Tom M. Mitchell: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*." Mitchell, T.M. (1997). *Machine Learning*. McGraw Hill.

39 This technique is usually referred to as *Transfer Learning*. For further information see e.g., Bommasani et al. (2021) *On the Opportunities and Risks of Foundation Models*. (Ithaca: Cornell University). https://arxiv.org/abs/2108.07258

or voice prompts.[40]

If the forms of AI and technologies that were mainly in circulation until the mid 20th century could still be seen as instruments that were programmed by people and disabled when not needed, then the more advanced forms of AI must be understood as integrating people into artificial systems without the possibility of disconnection.

A person integrated into an artificial system is one whose interactions with the external world are completely and constantly mediated by an artifact or by an ensemble of artifacts. Such is the case with our communication and road networks. Artifacts integrating people into systems function as prostheses that cannot be put down or put away. Like a body part, such artifacts become prostheses that function as an integral part of the person. Integration into systems therefore entails a progressive blurring of boundaries between persons and artifacts. Such integration also promotes the in-distinction between a persons' actions and reactions to system imperatives. When persons are integrated into systems it becomes hard to distinguish whether human action is the cause or the effect of systemic outcomes.

Crucially, such systemic integration entails the impossibility for people to step out from these interactions and to establish their own criterion for truth and reality. They are induced into a regime of constant simulation where actions become both real and unreal.[41] Insofar as they are constantly mediated by artifacts, these actions are only representations.

40 These General-Purpose AI models are named *foundational models* and the form of AI being discussed here is generally named *Generative AI* while foundational models trained on text data to generate natural language response are specifically named *Large Language Models*. For further information on foundational models and General Purpose AI see e.g., https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/#_ftn54

41 On this point, see the deep insights provided by Baudrillard, J. (1994). *Simulacra and Simulation*. (Ann Arbor: Michigan University Press)

Yet, insofar as these are the only relationships by which people interact with other people and the external world, they are very real.

Contemporary AI entail a deeper stage of integration into systems compared to what was achieved through previous information technologies. So far, integration into systems was mostly achieved only through well-defined material artifacts. The transport and the communication network whereby people connected to other people and with the external world was quite stable and built by people themselves. The material infrastructures and the models whereby systems were built were still human made and people could still assume to exercise some form of control over them. However, contemporary forms of AI seem instead to suggest the possibility of technological self-generation—whether of new prose, poetry, and images, or computer programs, routes, and connections— which progressively integrate humans. While the manufacturer of previous AI were still humans, new AI seem to be capable of self-production and self-replication.

The intrinsic contradiction generated by integration into systems

Whilst early forms of AI are instruments, contemporary forms of AI are integral parts of wide information networks. They cannot be thought outside the systems they contribute to create. Humans enter thereby a curious relation of mutualism with the models by which AI is constituted. The actions of people feed and change these models, which, in turn, feed and change people's actions within recursive cycles. For example, the news you receive is shaped by your swipes which is shaped by the news you receive. As a result of such recursive processes the two parts of the system that feed each other become ever more integrated, ever more similar, and indistinguishable. What might appear as a learning capability of AI is actually a process of constant and mutual adaptation

achieved through mutual surveillance and monitoring. In this way, human life becomes a gigantic simulation game and the models people come to rely on do not bear any lived truth.

The intrinsic contradiction that integration into information technologies and AI can produce is well captured by Escher's *Drawing Hands* where two hands draw one another. In the imitation game enacted through system integration, one hand represents humans and the other the system into which they are integrated. Through this game, humans draw their environment while their environment draws humans within a never-ending recursive cycle. There is no one or any will that is driving the operation and what is being drawn cannot be erased. The will of the drawer is constantly under construction and hence the drawer cannot exercise any autonomous action. It is for this reason that Escher's *Drawing Hands* cannot erase what they produce. In such recursive loops, people cannot step out, they cannot suspend what they do, they cannot exercise a potential to not-do. When integrated into artificial systems, people are defined by their actions and action equals existence. When machines imitate humans, as in the emerging AI forms, each of the two parts become progressively indistinguishable and flattened into a formless substratum made of information bits.

How cybernetic views on purposeful behavior have shaped systems

The rise of the instrumental age was informed by the idea of *causa instrumentalis* that Aquinas appended to Aristotle's *causa efficiens*. The model of the slave proved decisive in the spread of techno-scientific civilization—instruments whose 'will' was to do the will of others. Ongoing techno-science transformations that have accompanied the rise of systems are instead very likely linked to the first operational definitions of *purposeful behavior* that was formulated by cyberneticians around the mid-20th century and that are lead-

ing to various reinterpretations of *causa finalis* in science.

It is well known that Aristotle's *causa finalis* was excised from the understanding of cause in modern science.[42] It is instructive to note how Aquinas clearly linked *causa finalis* exclusively to intelligent agents. As he stated, "…whatever lacks intelligence cannot move towards an end, unless it be directed by some being endowed with knowledge and intelligence; as the arrow is shot to its mark by the archer",[43] and further that "those things that are possessed of reason, move themselves to an end; because they have dominion over their actions through their free-will, which is the 'faculty of will and reason.' But those things that lack reason tend to an end, by natural inclination, as being moved by another and not by themselves."[44] By relegating this type of causation to metaphysics, modern science has been able to progressively reify notions of intelligent subjects and objectivity while separating these two domains and developing own replicable and universal methods of inquiry.

It is in this light that we must understand cyberneticians like Rosenblueth, Wiener and Bigelow,[45] who enabled the readmission of *causa finalis* into science by re-defining purposive behavior and teleology. In their operational definitions, teleological and purposive behavior are proposed as synonymous with behavior controlled by negative feedback loops. In their view, the category of purpose understood as

42 For an overview on how these causations mechanisms have been excised by science, see, for example, Losee, J. (2011) *Theories of Causality*. Routledge. For an interesting overview of how all Aristotle's causes are being readmitted by systems science see e.g., Ulanowicz, R.E. (1997) *Ecology: The Ascendent Perspective. Complexity in Ecological Systems*. (New York: Columbia University Press). Chapter 2.

43 Aquinas. *Summa*. I, Q. 2, Art. 3.

44 Aquinas. *Summa*. I-II, Q. 1, Art. 2.

45 See Rosenblueth, A., Wiener, N., Bigelow, J. (1943). "Behavior, Purpose and Teleology." *Philosophy of Science*, 10(1), pp. 18-24

feedback loop becomes a "fundamental category in science"[46] that is amenable to scientific analyses. However, these forerunners of systems science tried to dismiss any reference to notions of final causality in their reinterpretations. According to them, causality and purposive behavior both pertain to the realm of science but "the prediction of the future from the past belongs to the theory of causality" whilst "the determination of the past from the present belongs to the theory of purpose."[47] Nevertheless, this distinction made little sense to subsequent scientists since final causation always entails a cause subsequent in time to a given effect.

Indeed, it is not accidental that the mechanisms of negative feedbacks by which purposive behavior was modeled by early cyberneticians have nowadays become the bedrock on which reinterpretations and readmission of *causa finalis* to science is taking place.[48] Reinterpretations proposed by scholars like Rosen,[49] Prigogine,[50] Kauffman,[51] and Ulanowicz[52] reveal how *causa finalis* is returning in the guise of such notions as autocatalysis, self-organization, and enablement that, to different degrees, represent reformulations of the negative feedbacks and servomechanisms that was proposed by early cyberneticians to explain 'purposeful' behaviors in ma-

46 See Rosenblueth, A., Wiener, N. (1950). "Purposeful and Non-Purposeful Behavior." *Philosophy of Science*, 17(4), p. 321.

47 Ibidem, p. 321.

48 See e.g., Chase, M. (2011). "Teleology and Final Causation in Aristotle and in Contemporary Science." *Dialogue*, 50, pp 511-536.

49 Rosen, R., 1991. *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. (New York: Columbia University Press).

50 Prigogine, I. and Stengers, I. (1986). *La Nouvelle Alliance: Métamorphose de la Science*. (Paris: Gallimard).

51 Kauffman, S. (2009). *Towards a Post Reductionist Science: The Open Universe*. (Ithaca: Cornell University). https://arxiv.org/abs/0907.2492v1

52 Ulanowicz, R.E. (1997) *Ecology: The Ascendent Perspective. Complexity in Ecological Systems*. (New York: Columbia University Press). Chapter 2.

chines, animals and humans. By redefining final causation in ways that do not entail any reference to human intelligence and free-will, they attempt to readmit final cause to science while keeping religion and metaphysics aside. Overall, the great merit of early cyberneticists such as Wiener, Rosenblueth, and Ashby is to have demonstrated the possibility of expanding the realm of science to include any phenomenon amenable to schemes of circular causality by constructing mathematical theories of feedback, stability, and regulation.[53]

However, it is crucial to contrast Aquinas on *causa finalis* and its purported return as feedback loops in contemporary systems science. The latter simply assume that beings that move themselves are, in fact, moved by something else within recursive feedback loops. By stipulating that what moves is really moved, systems science assumes to produce purposive behavior through negative feedback loops, whether in man, machine, or animal. Both Norbert Wiener and Arturo Rosenblueth made this imitative in-distinction between man, animal and machine the cornerstone of their rebuttal of Richard Taylor's critiques of their notion of purposive behavior.[54] They insisted that "men and other animals are like machines from the scientific standpoint because we believe that the only fruitful methods for the study of human and animal behavior are the methods applicable to the behavior of mechanical objects as well" and that "as objects of scientific enquiry, humans do not differ from machines."[55]

53 Drack M, Pouvreau D. (2015). "On the history of Ludwig von Bertalanffy's "General Systemology," and on its relationship to cybernetics - part III: convergences and divergences." *Int J Gen Syst*. 2015 Jul 4;44(5):523-571.

54 See Taylor, R. (1950). "Comments on a Mechanistic Conception of Purposefulness." *Philosophy of Science*, 17(4), pp. 310-317. Taylor, R. (1950). "Purposeful and Non-Purposeful Behavior: A Rejoinder." *Philosophy of Science*, 17(4), pp. 327-332.

55 See Rosenblueth, A., Wiener, N. (1950). "Purposeful and Non-Purposeful Behavior." *Philosophy of Science*, 17(4), p. 326.

Note however that both Aquinas' explanations of *causa finalis* and its reinterpretation by systems science as purposeful behavior are exclusively focused on representations of free-will, intelligence, and action in terms of motion and on how such motion can be activated and maintained. In so doing, both Aquinas and contemporary system thinkers miss that it is free-will and purposeful action in humans alone that entail the constant possibility of *saying no*, i.e., the capacity *to not* act, as was present in the Aristotelian understanding of potentiality. In erasing any difference between humans and non-humans, contemporary systems science has also erased this key aspect of willed action. Now, just like machines, humans have no potential to-not-do when enmeshed in recursive systems.

How integration into systems is reducing potentiality to action

The trend to depotentiate human potentiality that started with the notion of instrumental causality in the 12th century is now strengthening with the diffusion of information systems exemplified by contemporary forms of AI. These new artifacts and the increasing number of human functions they imitate begin to integrate people within artificial systems. This is clearly visible in how *storage* of energy sources, materials, competencies, and skills of any kind is being progressively reduced locally because they are integrated into ever expanding information and distribution networks. Why have an editor here, when a book can be edited in Ethiopia?; why have a shirt factory here when shirts can be shipped the next day from Thailand?; why store shoes in my local warehouse when they can be assembled or be re-distributed and sold on-demand? Embedding local demand and supply within spatially distributed chains entails that the potential to *resist* the integrative dynamics of information systems is diminished.

The demand and supply of energy and material resources must be dynamically matched through such networks as they become ever more distributed and fluctuating. This situation makes interruptions to existing energy and material flows extremely dangerous for populations relying on them. Exemplified by the dependence on the Internet, the possibility of disconnection from these systems is becoming increasingly difficult and the information networks that mediate human actions must be kept constantly on. Within these networks, even silence from a node is changed into a message that triggers a reaction (see the number of messages we start receiving when disconnected from social networks like *Linkedin*, *ResearchGate*, etc. for too long). Contemporary forms of AI accelerate this process of integration. At the same time, these new technologies seem to acquire their own subjectivity and intentionality, as the processes whereby they apparently "learn" by relying on training over huge amounts of data become black-boxed and increasingly opaque. If the distality of instruments enabled a sharp distinction between artifacts and subjects employing them, contemporary forms of AI seem to be able to flatten any difference and induce a perverse symmetrization between human and non-humans.[56] Long advocated by sociologists of science and technology like Bruno Latour and cemented in varieties of actor-network theory,[57] the notion of symmetrization is now being ushered in around contemporary forms of AI where it is not unusual to discuss the moral status and rights of AI.

56 To my knowledge, Bruno Latour provides the most lucid example of how potentiality can be dismissed by social science e.g., in Latour, B. (1988). *The Pasteurization of France.* (Cambridge: Harvard University Press) p. 158-176

57 Murdoch, J. (2008). "Inhuman/nonhuman/human: actor-network theory and the prospects for a nondualistic and symmetrical perspective on nature and society." in Philo, C. (2008) Ed. *Theory and Methods. Critical Essays in Human Geography*. (London: Routledge).

If the above reflections are correct, then the integration of man-in-systems entails a curious destiny. People integrated into systems get entirely defined within self-recursive actions. When integrated into systems people cannot *not*, and the relation between original human potentiality and human action is severed. Humans lose the potentiality to not-do, to not use the system, to resist the system. Enclosed within feedback loops people can only act and their existence is a function of their action. However, when human potentiality becomes severed from actions that are fed back in a recursive loop, action becomes, paradoxically and contradictorily, indistinguishable from reaction. Shorn of human potentiality, human action comes to resemble messages without a sender. Actions under the condition of integration become an input for data models in need of constant updating and are regulated by these models if they deviate from expected reactions. This is the kind of simulation in which the machines that imitate humans consume reality. Crucially, when defined only by actions within systems, humans are left with no possibility for *forgiveness* and *redemption*. Further, when actions cannot be separated from reactions, persons cannot be considered *accountable* for what they do either. Forgiveness, redemption, and culpability presuppose the ability to be other than what one does and to be more than actions within negative feedback loops. This is the main quandary when humans are integrated as functioning units within systems.