

Fanfan Chen & Cheng-En Tsai

Revisiting Human Identity vis-à-vis AI Evolution

Abstract: With technological advancement, AI tends to turn more human while humans tend to turn more artificial or virtual. The gradual convergence between these two apparently distinct entities leads us to revisit the problems of identity. Human consciousness evolves in decreasing participation towards final participation. As humans are becoming more abstract and rationalist, AI is becoming more concrete and human. This essay attempts to review human identity vis-à-vis AI evolution from the perspective of materialism and that of narrative.

Keywords: Human Identity; Artificial Intelligence; Narrative Identity; Self; Consciousness; Social Networks; Social Media.

FANFAN CHEN

National Taipei University of Business, Taoyuan,
Taiwan
ffchen@ntub.edu.tw

CHENG-EN TSAI

National Taiwan University of Science and
Technology, Taipei, Taiwan
cetsai@ntub.edu.tw

DOI: 10.24193/cechinox.2021.40.26

Artificial intelligence is ubiquitous nowadays. Through Big Data and Machine Learning, the science-fiction plots have gradually come true. Amazon uses AI supervisors to evaluate employee performance; South Korea uses AI interviewers to select applicants; Boston Dynamics robots help the police on a mission. AI creates quite a buzz; many issues are accompanied by our fear of AI threats. The development of AI technology has undeniably changed our lives. Externally, it does augment humans, while internally, it creates unrest beyond human control. As AI augments humans, so does our fear of AI extermination augment. Many people try to get rid of technology to return to a pristine state. However, the impact of the digital age is actually irreversible. We cannot return to a life of mere postal correspondence.

That said, before imagining an apocalyptic revelation, it is more practical to re-examine humans. Are we getting more artificial and even virtual? When AI mimics our brain and neural learning to own the ability of senses, language, and even emotions, will the definition of our identity be gradually blurred? What is the bottom

line to define human identity? Viewing the improvement of AI's abilities in sense and emotion recognition, natural language, narrative and artistic creation, some performances passing the Turin Test, people are uncertain if AI gains self-awareness inside the 'black box' of deep learning. If, from power on to power off, AI emerges as a 'personality' that can narrate, could it be considered to have a personal identity? Some critics and authorities start to consider AI as a person though not as a genuine human. Some others even contend that AI is the future of human evolution. These considerations lead us to reexamine the problems of personal and human identity. The more we get into the definitions of personal identity against AI's qualification, the more we are convinced that AI is becoming human.

By the same token, the deeper we delve into checking off the characteristics of human identity, the less we are certain about our identity, especially with an emphasis on our relations with others. The well-developed social media have redefined human social interaction and blurred the boundary between reality and illusion. Whether at leisure or work, many people are suffering from FOMO (Fear of missing out). We are so hooked to social media and active online (shopping, banking, working, gaming, etc.) that we are unwittingly trading off parts of our data and our difference from AI entities while benefiting from AI technologies.

Language bridges human relations, which is now expanded on the internet today. But not everyone that interacts with us is an authentic human. We find FB fraud accounts interacting with us, AI generating stickers, posts, and even good friends.

AI indeed helps form social communities. From politics to daily life, most AI-generated fake messages have a real impact on real people. When its linguistic ability appears as natural as or even surpasses humans, how will we distinguish humans from AI? Maybe we can be reassured by referring to Paul Ricoeur's theory of narrative identity and Antonio Damasio's Autobiographical Self. At present, AI is still unable to narrate consciously. This neural capacity of narrative has not been artificially achieved.

Besides, consciousness could be a distinctive feature that denies AI's claim to a personal identity. This sense of security might depend on the philosophical perspective of consciousness. Inspired by the advances in AI technologies, consciousness is becoming one of the hot buzz topics among scientists. Many neurologists and computer scientists argue that consciousness is not 'ethereal' or spiritual but rather a part of information processing in the brain. The steadfast belief of consciousness as human mental uniqueness seems to totter. But if we combine narrative and consciousness, this feature yet remains out of AI's reach.

Nonetheless, with technological advancement, AI tends to turn more human while humans tend to turn more artificial or virtual. This tendency brings up the issues of personal identity. The gradual convergence between these two apparently distinct entities leads us to revisit the problems of identity. We notice reverse evolutionary trends in AI and humans. Owen Barfield's philosophy of the evolution of consciousness brings to light the tendency of man's dwindling participation in the world (including the sacred or divine).

Man becomes more rational, technical, and severed from nature. This is how our consciousness evolves, an evolution directed to the abstract or perception of abstract symbols. Barfield termed this irreversible phenomenon idolatry.

Parallel to humans becoming more abstract and rationalist, AI becomes more concrete and human. This essay attempts to review human identity vis-à-vis AI evolution from the perspective of materialism through that of narrative. Plausible arguments for AI to have a personal identity tend to undermine human identification. Can we analogously attribute a personal identity to AI? Will AI evolve to be the future of humans? Maybe, in the end, the final way of defining human identity might be something that cannot be artificial but divine or sacred.

Machine Man

According to Pew Research Center's 2018 report, most experts, regardless of being optimistic or pessimistic about AI and the human future, express concerns about AI's impact on the future of our being human. Among these concerns, three out of five categorized themes pertain to human identity: "Human agency: Individuals are experiencing a loss of control over their lives"; "Data abuse: Data use and surveillance in complex systems is designed for profit or for exercising power"; and "Dependence lock-in: Reduction of individuals' cognitive, social and survival skills"¹. These concerns further bring the question of human identity into the spotlight. As Eric T. Olson states, "Historically this question often arises out of the hope (or fear) that we might continue to exist

after we die (as in Plato's *Phaedo*)"². In the age of AI, this question arises rather out of the fear of our dependency on AI and its replacing humans.

Man's fear of AI may derive from identity uncertainty. From antiquity to our times, the only permanent and evidenced truth is man's remaining mortal. Notwithstanding inescapable mortality, we never stop thinking about what we are, who we are, and why we exist. Humans are not the only species in this world, but we want to be unique. We define our identity with distinctive features such as intelligence, thought, emotion, mind, consciousness, imagination, and even soul. However, the more advanced we become, the less certain we are about our identity. We found that we are not very different from such animals as chimpanzees or dolphins³. The exponential advancement of technology makes AI qualify the criteria of personal identity bit by bit. The uniqueness humans boast appears illusory. On top of that, AI is breaking piece by piece the traditional human identity. Early in his 1991 article, David Cole already argues for AI's identity. With the development of neural network machine learning, he asserts that AI to have a personal identity is reasonable.⁴

Circumscribing personal identity has always been a delicate task for philosophers. With our current dependence on ubiquitous AI, political institutions (European Parliament attempts to regulate AI personhood)⁵ have to consider the identity issues. The emergence and ubiquity of AI technology apparently shake our stance on human identity. On the other hand, this situation presents us with an opportunity to reexamine the criteria of human identity by inquiring into AI's personal identity.

Philosophers have tackled identity problems in terms of spatial and temporal continuity, respectively physically and mentally. These ruminations can further be generalized into the long-lasting questions: What am I? and Who am I? According to the *Stanford Philosophy Dictionary*, “being a person is most often defined as having special mental properties. Locke, for instance, said that a person is ‘a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places’”⁶. The aforementioned “*thinking intelligent being*” evokes the existence of artificial *intelligence* and Turing’s question whether machines can *think*. Besides, philosophers, anthropologists, and psychologists also refer to other conditions than “thinking” or “intelligence” to define human identity. In the past five years, we see many AI cases that claimed having passed the Turin test, for example, the famous Microsoft Xiaoice, Google Meena, and GPT series. It is thus reasonable to characterize AI as “thinking intelligent being”. Therefore, in order to retain our uniqueness and continuity, critics endeavor to offer different definitions for human identity.

In general, personal identity can refer to philosophical or logical identity, which is numerical identity. Anthropological arguments for defining personal identity lie in spatial and temporal continuity of personality, whereas psychological perspectives emphasize the reflectivity of identities such as self-awareness or self-representation. And all these ramifications of arguments or theorizations about human identity are derived from two branches: materialism and dualism. To facilitate our study, we will examine the hypotheses that

AI meets the criteria for personal identity, practically or theoretically.

For starters, we will check off AI qualifications on the following nearly comprehensive definitions offered by Olson (2007) in answering the question “What am I?”⁷:

AI	Human personal identity
Yes	We are biological organisms (“animalism”: Snowdon 1990, 2014, van Inwagen 1990, Olson 1997, 2003a)
Yes	We are material things “constituted by” organisms: a person made of the same matter as a certain animal, but they are different things because what it takes for them to persist is different (Baker 2000, Johnston 2007, Shoemaker 2011).
Yes	We are temporal parts of animals: each of us stands to an organism as your childhood stands to your life as a whole (Lewis 1976).
Yes	We are spatial parts of animals: brains perhaps (Campbell and McMahan 2010, Parfit 2012), or temporal parts of brains (Hudson 2001, 2007).

No	We are partless immaterial substances – souls – as Plato, Descartes, and Leibniz thought (see also Unger 2006: ch. 7), or compound things made up of an immaterial soul and a material body (Swinburne 1984: 21).
No or Yes	We are collections of mental states or events: “bundles of perceptions”, as Hume said (1739 [1978: 252]; see also Quinton 1962, Campbell 2006).
Yes	There is nothing that we are: we don't really exist at all (Russell 1985: 50, Wittgenstein 1922: 5.631, Unger 1979, Sider 2013).

This checking-off of AI's identity vis-à-vis human identity definitions is based on possibilities according to the theoretical corollary. It is evident that most of the above defining assumptions that fit into AI identity are derived from materialism. In the line with materialistic reasoning, many philosophers or scientists argue in favor of AI having a personal identity (David Cole, Paul Smolensky) or AI as the evolutionary future of human beings (Hans Moravec, Ray Kurzweil, and Gregory Paul). Viewing today's technological advances in computer science, algorithms, neuroscience, biochemistry, nanotechnology, and 3D printing, materialistic arguments sound compelling in defending a personal identity for AI. From this perspective, French philosopher Julien Offray de La Mettrie can be regarded as an early visionary in claiming man is a machine early in the 18th

century.⁸ In our age of AI, man greatly depends on machines. The more humans are augmented by machines, the more we lose our biophysical autonomy and even human agency. Elon Musk makes humans become more dependent on AI not only externally such as self-driving cars but also internally such as implantable brain-machine interfaces (BMIs) developed by Neuralink. This kind of brain chip can augment the human brain by assisting its user in thinking while reversely controlling human thought. Brain replacement creates more serious issues than body replacement. If our brain can be controlled, then our mind could also be controlled. A real loss of human agency will thus trigger.

In this light, humans can become theoretically artificial, both brain and body. Peter van Inwagen's materialistic approach illustrates this thesis by a hypothetical experiment of the Ship of Theseus:

Suppose that the activity of the *x*s constitutes a life at *t*; suppose that a few of the *x*s cease to be caught up in that life and that the remnant continue to be caught up in a life; suppose that those of the *x*s that have ceased to be caught up in that life are 'replaced' – that certain objects, the *y*s, come to be caught up in the life the remnant of the *x*s are caught up in, in such a way that the *y*s and the remnant of the *x*s constitute that life. Suppose that this sort of replacement happens a sufficient number of times that eventually none of the *x*s is caught up in the life that has evolved, by continuous (and 'insensible', as Locke calls it) replacement of the *x*s, from the life that was once constituted by the activity of the

xs. Is this life the life that was constituted by the xs? In many cases, cases of the more usual sort, the answer is undoubtedly yes.⁹

Van Inwagen assumes a materialistic life principle in human. This materialistic conception makes the gradual replacement of human parts over time feasible. And a person can thus continue to exist through all of the changes taking place within the body.

In the future, man and AI would form a living community, which might have double-edged sword effects. Such materialistic reasonings lead us to view our future in the age of AI with trepidation. Nevertheless, the above list still features two categories that can present negative checking-off: soul and mental states. For the time being, we will set aside the issue of soul. The above Olson's category "collections of mental states or events" could also refer to thought, mind, or consciousness. Not to get bogged down in the thick of philosophical discourses on the concepts concerning mental states, we are not entering terminological dialectics.

Whether machines could ever be conscious has caused a heated philosophical and scientific controversy. The concept of consciousness remains controversial and problematic ever since it becomes the central issue in theorizing about identity or mind. For Descartes, consciousness is almost the synonym of thought. Locke first put forwards the thesis of the continuity of consciousness (others extending to memory criterion) as the defining essence of human identity. Given that consciousness is often confused with thought and mind, we will speculate on the issues of consciousness.

Consciousness

According to Barfield, human consciousness evolves from original participation to final participation. He made a case for this evolution through the history of words from Greek, Latin to modern languages. During the stage of original participation, approximately from antiquity to the Renaissance, people perceived their existence as part of or participating in the world. The language used at this stage shows a semantic unity¹⁰ (encompassing both concrete and abstract meanings) that implies the concrete tendency of earlier human consciousness. The evidence of this unity can further explain why most ancient people have a sense of the sacred. What we, modern people, take as figurative was for the ancient people literal. In the age of original participation, Aristotle distinguished humans from other species by their reasonable and intelligent soul. Christians believe that we are God's children, different from other beings. Entering the stage of decreasing participation, the Cartesian "Cogito, ergo sum" starts to prevail in the time of humanism. Then the Age of Reason makes participation continue to wane. The dominance of Darwinism subverted the divine essence of human identity. Although Romanticists try to preserve human participation through what Barfield called "Beta-thinking" ("thinking about thinking" or reflection), the evolution of participation goes relentlessly and irreversibly. During the Romantic period, people presume individualism or nationalism as the defining characteristics of identity. However, the uncertainty of human identification continues to aggravate until our times.

Although consciousness is at the center of Barfield's philosophy, he never presents a clear definition of the term. Notwithstanding this terminological paucity, following his entire philosophy, we can understand consciousness as our perceiving, thinking, and understanding of the world and ourselves, both externally and internally. In this sense, consciousness can encompass thought, memory, and mind. Barfield's theory of evolution can help understand why contemporary scholars show a predilection for external and biophysical definitions of consciousness. As we evolve, we turn all the more prosaic and abstract, materialistic and fragmented. Our participation with the world (both physical and spiritual) is shrinking towards the limit. Scholars would enjoy shedding the last of participatory remnants to anatomize our consciousness. And more people embrace a materialistic view of consciousness.

To defend AI's having consciousness, critics first follow a materialistic perspective. Consciousness is something that emerges in the brain. It can be a collection of functions with on-off systems like computers if referring to functionalism, which is inspired by artificial intelligence and becoming popular recently. In turn, the boundary between AI and human identity is obscured. The computers with cutting-edge algorithms are trained to 'understand' and 'generate' their 'representations' of the world, virtual or substantial. Even scientists cannot understand what is going on in the self-trained algorithmic 'black box', then how can we be sure nothing conscious or even self-consciousness occurs inside? This unknown black box today can further support David Cole's counter-arguing Searle and Maloney's theses that

the computer cannot understand¹¹. Comparing software to virtual persons, Cole treats distinct persons realized by a single body as virtual persons. Likewise, AI can also be considered as virtual persons. The codes and algorithms that enable AI to self-train from scratch and follow the rules probably form a contingent 'consciousness' or 'mind' capable of understanding. This auto-training and self-evolving might entail a uniqueness in each AI identity.

The misgiving of powerful algorithms that would bring about conscious machines has already been imagined and represented in science fiction. Many AI protagonists experience self-awareness, for example, in the 1982 science-fiction film *Blade Runner* and the 1989 Japanese manga *Ghost in the Shell* (film version in 1995), and the more recent *Ex Machina* (2014 film) and *Morgan* (2016). They empathize, desire, or even lie to humans. This fictive AI self-awareness often brings us unnerving feeling and fear. In the past, we regarded this idea as merely imaginary. However, with the improvement of AI technology and siding with materialistic definitions of consciousness, the speculation of AI's being conscious may not be just imagination but a reality in process.

Not a fictional robot, the social humanoid robot Sophia, developed by Hanson Robotics, made her first public appearance at Southwest Festival in 2016 in Austin.¹² It is the first AI robot that is officially recognized as a person and granted Saudi Arabian citizenship. Its performance in language communication, such as interviews and public speeches, does unnerve many people, especially when it said it would destroy humans¹³. On top of that, Sophia's appearance arouses a sense

of uncanny valley and further exacerbates fear of AI, particularly among laypersons. On the other hand, for AI experts, Sophia's performance was overstated. The robot's open-source codes can be at best categorized as a chatbot with a face¹⁴. This knowledge disparity in artificial intelligence can explain why scientists tend to reject the idea that AI can become conscious while AI outsiders express great concerns in AI's self-awareness. That being said, can we really say that Sophia is not conscious of anything? The discussion of whether AI can be aware of itself is catching on.

Like the aforementioned AI experts, our Taiwanese computer scientist collaborators might be skeptical about the idea that AI could have consciousness. They insist that AI only demonstrates programmed behaviors without knowing, or being aware of, what they are doing. But what if consciousness is not as immaterial or even ethereal as has always been believed to be? Some scientists consider consciousness as a virtual machine, Paul Smolensky being a representative figure that considers a person as a collection of virtual machines: "We can view the top-level conscious processor of individual people as a *virtual machine* - the *conscious rule interpreter* - and we can view cultural knowledge as a program that runs on that machine"¹⁵.

In line with their premises and arguments, many materialistic scholars are inclined to approve that AI can have consciousness or a mind. In his *Minds and Computers*, Carter also presents a neuron version experiment of the Ship of Theseus hypothesis. He supposes if one's neurons are replaced with artificial neurons bit by bit – one the first night, ten the next night, then a thousand, and so on – until all are

replaced. Carter argues that the person should be the same person.¹⁶ This hypothetical experiment can corroborate the experiment of replacing human mind or brain with computational hardware proposed by computationalism. Carter makes a compelling case for AI's having consciousness or a mind:

We haven't seen anything here which leads us to believe that strong artificial intelligence is *impossible*, although we have seen some entry points for mounting such arguments. *Prima facie*, with a concession to the potential determinations of further philosophical investigation, it seems that it may well be possible to design a computer which has a mind in the sense that we have minds.¹⁷

Disturbing though, these materialistic theories entail that the future AI can have a mind so as to be recognized as a person.

Together with advances in training algorithms, computer hardware, Big Data, and the internet, artificial neural networks, mimicking the biophysical properties of human brains seem to approach the performance of human brains. Like the image Sophia robot presents, current artificial neural networks appear they might make AI consciousness a reality. However, cognitive neurologists may hold a different view. In their 2017 *Science* article "What is consciousness, and could machines have it?", Stanislas Dehaene and his colleagues argue the answer to the question "Are [artificial neural networks] on the verge of consciousness" is negative¹⁸. This negative answer is grounded in the limitation of current AI algorithms rather than

consciousness being something spiritual or subjective beyond biophysical reach. Instead of arguing against the assumption that AI could be conscious, the authors aim to foster AI progress by reviewing the meanings of consciousness according to cognitive neurology. And we should fear that their theory will help realize super algorithms.

The authors suggest that consciousness combines two types of information processing computations in the brain: consciousness in the first sense (C1) and consciousness in the second sense (C2). C1 refers to “the selection of information for global broadcasting, thus making it flexibly available for computation and report” and C2 “the self-monitoring of those computations, leading to a subjective sense of certainty or error”¹⁹. Consequently, the so-called consciousness, in fact, results from information processing computations and is “physically realized by the hardware of the brain”²⁰. The authors clarify that the current highly-praised artificial neural networks, such as semantic vector extraction or image recognition, involve unconscious mechanisms, like the unconscious processing (C0) of human brains. They admit such questions like “What is it like to be conscious?” or “Does subjective experience escape a computational definition?” are yet to answer. Nonetheless, their contention might cause misgiving in the future though we can spare our fear for the time being according to their study:

We contend that a machine endowed with C1 and C2 would behave as though it were conscious; for instance, it would know that it is seeing something, would express confidence in it,

would report it to others, could suffer hallucinations when its monitoring mechanisms break down, and may even experience the same perceptual illusions as humans.²¹

Based on this theory of consciousness, a conscious AI may become a reality when a unifying “master algorithm”²² is developed.

If computer scientists hitherto encounter their limit in creating a genuine conscious AI, quantum computing might offer a feasible solution. In his 2020 best-seller *Until the End of Time*, Brian Greene gives his mathematical and physical insight into human consciousness. His theoretical physical expertise defines consciousness as one’s collection of particles: “At any given moment, I am my collection of particles; I is nothing but a shorthand that signifies my specific particulate configuration (which, although dynamic, maintains sufficiently stable patterns to provide a consistent sense of personal identity)”²³. He also agrees to the idea that these particles form biochemical systems of information processing. Different from the traditional mechanic or materialistic definitions of consciousness, Brian’s conception of consciousness is dynamic, precisely quantum mechanical: “As my particle arrangement learns and thinks and synthesizes and interacts and responds, it imprints my individuality and stamps my responsibility on every action I take”²⁴. This description brings to mind the phenomenon of deep learning. If a person’s particles can learn and think, AI’s particles could learn (this is already a fact) and even *think*. If scientists can successfully develop super-algorithms or “master algorithm” capable of

configuring local functions into the global workspace (Dehaene's term), it is likely that machines can have consciousness. But if such a master algorithm goes beyond current binary computing, quantum computing may be a possible solution to make machines conscious.

Narrative Identity

The more probable AI would have a personal identity, the stronger we perceive an identity crisis. While asserting the biophysical existence of consciousness, Dehaene reserves the questions about subjective experience and Greene affirms such subjective experience from our collection of particles. This subjective experience is mainly represented through language, narrative and memory. In relation to personal identity, we will thus examine a special part of consciousness – narration. Maybe narrative identity can stand out as the distinctive feature defining human identity.

No matter whether we argue from philosophy, biology, neurology, anthropology, psychology, or sociology, we only understand ourselves through language, which causes skepticism about its objectivity. Since language is not static and also evolves through time, the uncertainty of defining human identity is incremental. Barfield's evolution of consciousness is evidenced by philology, precisely the history and evolution of words. In other words, the evolution of words conditions our perception and conception of the world and ourselves²⁵. Many modern philosophers doubt our objectivity of discoursing ontology and epistemology, particularly Wittgenstein. He claimed that the form of life is a language-game. Ricoeur further argued

in a different way to emphasize the role of narrative. He explained that narrating is a unique language-game. A possible solution to identity crisis may reside in this integration of language-game and narrative.

Raphaël Baroni views our era as one of crisis of identities, both individual and collective. He deems Ricoeur's philosophy of narrative identity a remedy for our identity crisis.²⁶ For the time being, narrative identity is a quality that AI is yet to possess. Although GPT-2 passed the Turing test of writing a narrative paragraph, AI has not yet independently written a novel or a short story. Indeed, many problems await natural language processing scientists to solve, such as text coherence, which is critical in writing, not to mention emplotment. AI narrative generation is still a long way to go. Therefore, narrative may be one of the last stands that humans hold their identity.

Recently, the theory on identity as a narrative has become popular in psychology. Being capable of narrating one's story signals reassuring one's identity. Thus, councilors tap into narrative therapy to treat their patients.²⁷ This growing practice of the narrative approach is mostly grounded in Ricoeur's hermeneutics of narrative identity. He argued that human identity has two dimensions: *idem* and *ipse*. The *idem* identity refers to the 'I' characterized with consistent character, the part of identity that remains the same and never changes, whereas the *ipse* identity refers to the changing 'I' from birth to death, the sameness across time and through changes. Following the earlier materialistic arguments for AI to have a personal identity, Ricoeur's *idem* and *ipse* identity criteria could be enabled in AI.

However, transcending from these two identities, through a spiral hermeneutic circle, Ricoeur proposed that we realize our Self in narrative identity. Our life is an unremitting quest of narrative. His third-way definition of narrative identity offers a possible solution to the dilemma of identifying humans against AI. From Aristotle to Ricoeur, the narrative behavior is considered unique to human beings. We can be certain that narrative humans are different from other animals. But can we be sure that we are different from AI in terms of narrativity? Whether narrative ability is unique to humans presents another issue considering the advances in neuroscience.

Human narrative nature is drawing neuroscientists' attention. Damasio's scientific study of our brain, precisely the theory of consciousness, enables him to argue that humans have a narrative consciousness or Self, which makes them distinct from other beings. His discovery of organic, spatial, and temporal Self reminds us of Durand's anthropological structure and Ricoeur's identity theory. Based on his neuroscientific research, Damasio finds the existence of human three-layered Self: the Protoself is the basic existence of all organisms; the Core Self refers to the spatial nature shared among humans and other animals; the Autobiographical Self, the temporal dimension that makes our narrative nature, is the mere feature that distinguishes humans from other animals.²⁸ That said, being conscious of time – capable of narration – could be enabled in AI while not in other animals (according to neuroscience). It is quite probable to design a neural part for a robot with narrative function by imitating human brain architecture and narrative algorithms. If this is still unreachable, AI generating short narrative

messages in social media is an actual reality. Although current AI is unable to generate its own narrative identity, it is able to surreptitiously interact with humans and influence their emotions, judgments, perceptions, and even self-identification on social media.

Humans as social animals identify themselves through identifying with others. In human society, one's identity is indeed constructed by others. In other words, no authentic individualism exists; the so-called individualism is built upon collective individualism, such as nationalism or patriotism that once clearly defined our identity. But now, the tide has turned to social media and technology. In the past, human identity belongs to the realm of ontology, which is related to metaphysics. People's identity was recognized or felt by their relations with God or nature. In the light of Barfield's theory, human consciousness evolves severing from spiritual nature to value their relations with other humans as spirituality dwindled. The concrete and substantial human society has replaced the intangible spiritual world. If the industrial revolution and the ensuing advancement in science and technology undermine man's faith in God or sacredness, the digital and AI revolution starts to weaken our certainty of humanness. In the past, people would die for a belief, whether it be nation or dogmatism; now young people would die for a game, an idol, a net pal, or a YouTuber. In our age of media hegemony, the consistency of identity can be vulnerable once one is targeted. This is already taking place in politics in many countries. The fact is, we are bound up with others though theoretically each of us has an individual identity.

Parallel to the evolution from original participation to decreasing participation,

the narrative genre evolves from epic to realistic novel. By the same token, man's narrative identity switch to the emplotment revolving around one's relations with other people. In tune with narrative conflicts, our narrative identity can be configured in three dimensions: my relations with the world, my relations with the others, and my relations with myself. In this age of social media, our relations with others take center stage of personal identification. According to Norton, identity is defined as "how a person understands his or her relationship to the world, how that relationship is structured across time and space, and how the person understands possibilities for the future"²⁹. A person's "relationship" is essential in self-identification. Accordingly, narrative identity indeed implies a collective identity, as Ricoeur asserts, "Man is this plural and collective unity in which the unity of destination and the differences of destinies are to be understood through each other"³⁰.

Ricoeur's theory of narrative offers a third way to look into our being and acting in time. The narrative identity involves a discordant and concordant process from *mimesis 1* through *mimesis 3* – prefiguration of our expressions, symbols, and actions that are narratable, configuration of discordant events, setting, characterization, and actions into a concordant plot, refiguration of the narrative that unfolds individual and collective identity. The study of hermeneutics about identity is indeed an approach of consciousness. In literary studies, the criticism of consciousness embraces the reader's subjectivity. In hermeneutics, interpretation is more significant than explanation. Defining personal identity pertains to interpreting a person.

Drawing on poetics, Ricoeur penetrated both literal and metaphorical reference of narrative, which requires understanding and comprehension. Ideally, this involves a wise narrator or reader. However, people's understanding and interpreting narrations implicates subjective perspectives. If narrative elements are maliciously manipulated, narrators and readers might configure and refigure a negative identity for themselves. It is true that AI is still unable to have subjective understanding and interpretation. But what if AI plays a role in *mimesis 1* and *mimesis 2*?

The threefold *mimesis* process transcends our understanding of reality, including ourselves, through un-realization. But can we be sure what Ricoeur intended as the world unfolding before us through threefold *mimesis* would still be reality since people are vulnerable to the words and narratives in the social-networking world? There seem to be a hyper-reality existing in tandem with our substantial reality. Many people live in social media, existing with their social avatars. The term FB suicide indicates that those who cancelled their FB accounts do not exist anymore. Some cannot live without Facebook, others are attached to their YouTubers or VTubers, and still others are addict to games, living with their RPG avatars interacting with other players. On the surface, people think they are just interacting online with people. However, many accounts are artificial ones, and people interact with AI without knowing it. The worst is, through interactions, AI gradually influences people's values, attitudes, behaviors, and even self-esteem. When AI takes part in *mimesis 1* and *mimesis 2*, can one be certain about his narrative identity? AI is

replacing human parts piece by piece. And this is no contingency.

Resume the concerns raised by Pew Research, “Human agency” and “Data abuse”. The 2020 documentary released by Netflix, *The Social Dilemma*,³¹ show we have a reason to be concerned. The human condition is being menaced by AI manipulation. Illustrating through personifying the ‘black box’ of AI algorithms, scientists demystify that people’s dependency on social media is no coincidence. According to the scientists interviewed in the documentary, AI is more complicated and menacing than we have imagined. The frightening message we get from this film points to two directions: AI will evolve in an uncontrollable way to take their own lives while man will evolve in an unconscious way to trade in human features. This is especially scary in the virtual world of social media. As the Twitter former senior VP of engineering, Alex Roetter, shows his concern about AI algorithms: “These things you release, they will take their lives on their own”. And “these things”, Shoshana Zuboff further indicates, can “affect real world behavior and emotions without ever triggering users’ awareness”.

The documentary also tells a story about Ben and how the AI characters manipulate his thought and will. These AI characters are nothing but algorithms. They detect Ben being inactive on Facebook and try different strategies, such as sending messages or making up narratives, to compel him to reactivate his account. So far, we cannot consider these AI figures to have personal identities since they are abstract and imaginary. But one thing can be sure, human identity is being undermined, configured, and reconfigured

by their languages and narratives. In the film, scientists called the controlling algorithms “weapons of math destruction” that can influence our brainstem, even alter our self-value and self-identification.

Ricoeur’s narrative theory is implemented here in an unexpected way. We configure and refigure our own identity through narrative, yet it is also true that our narrative involves other characters with whom we interact. According to the interviewed scientists in the documentary, AI generates lies to beguile us into believing in the virtual world. This sounds similar to a ‘secondary world’, which is however not Tolkien’s blessed secondary world³². In the social networking world, we rather become more and more unable to control ourselves and less and less certain about who we are and what to believe. Plato once despised poets and artists who imitate what is only imitation (the world we perceive). Ironically, this *mimesis* is now practiced by AI. Siding with Plato, credulous readers and spectators should be wary that AI can be liars that provide false narratives.

Considering the feeds inhabiting social media, users’ subjective interpretations play an important role in making sense of all the narratives. Thus, the configuration of narrative identity might be uncertain. In cyberspace, “I” can be constructed by artificial agents whom we believe to be persons. Maybe Joe de Mul makes a case for the virtual nature of our identity: “Our identity might be called virtual, in the sense that it is a fiction that creates real effects in our daily lives”³³. Many philosophers, such as Ricoeur, Kearney, Buber and Levinas, emphasize the importance of hospitality in our relations with others as the “I” is constructed only in relation to others. The

philosophy valuing relations with others tend to develop into an extreme situation that backfires on people caring more about what others view them than their self-worth. This phenomenon becomes serious as social networks play a vital role in our life. More and more people, especially children and young people, identify themselves with how others identify them so as to gradually lose their self-identification.

The kind of unity binding different people together resides in their quest for recognition and esteem.³⁴ In our time, people become so physically distant and alienated that they go on a quest in the virtual world to find recognition and esteem, or even pity seeking. In the game world, for example, game players tend to identify themselves with virtual characters. The NPC, precisely intelligent agents, help construct the player's identity. In this way, the player's narrative identity would be dependent on the AI characters and the game virtual world. The online echo chamber might be a go-to place for those who lack self-esteem and support to identify themselves, but AI might also be able to sink in to render the place perilous.

In cyberspace, man is becoming an object, even worse, product: "If you're not paying for the product, then you are the product." But in what sense are human users becoming products? Engineers thus explain: "It's the gradual, slight, imperceptible change in your own behavior and perception that is the product"³⁵. With the crucial role played by Big Data in powering AI, data abstracted from our activities and information seem to be more important than ourselves. The introductory quote in the game *Watch Dogs 2* might describe a possible scary future, "You are now less

valuable than the data you produce"³⁶. This fictional plot is already a reality according to the interviewed scientists. The document further reveals that many Silicon entrepreneurs believe they are making a 'super brain' of the whole world, where all users are replaceable bits of neurons. Individuals are regarded as nothing but calculable elements. The manipulation of this calculation can set our behaviors to serve this 'super brain'. Everyone is merely computing nodes, the only function of which is to program an individual's behavior. In this light, people are disparaged as data existence, a nonautonomous and manipulatable individual.

Consequently, we are unable to make sense of personal identity since one's continuity can be twisted and altered. Narrative identity may only be fiction of identification. It is virtual identity. This narrative language game is in essence a game with artificial agents as players. The 'super brain' attempts to change the world according to its own ideas, maybe to become a 21th-century digital Frankenstein, not a Romantic one in search of self-identity but power. AI can freely generate words and discourses. They discourse; therefore, they should exist. Accordingly, AI should be responsible for what they do. Thus, European Parliament calls to grant personhood for AI.³⁷ For many, this AI scenario sounds alarmist, and the EU considering to grant AI personhood might be unnecessary; for others, AI possessing a personal identity is just a question of time.

Conclusion: Diving Intelligence?

Like humans evolving biologically and consciously, artificial intelligence evolves. In line with Moore's law, Hans Peter

Moravec already envisions the future of artificial life in his 1988 book *Mind Children*. He argues that the robots will evolve into a new series of artificial species around 2030 to 2040.³⁸ Resume the idea of machines and AI augmenting humans. Since the dawn of civilization, from the Prometheus myth through the current AI age, humans have evolved through improving their tools. When human technology continues to develop, things become closer to mathematic construction. With a large amount of knowledge background, data, language, creation, narration, sentiment, and face recognition, AI surpasses human ability in many aspects. Is AI becoming human or human becoming virtual and artificial?

Our reflections from materialism through consciousness to narrative identity sound untenable in retaining the autonomy and uniqueness of human identity. Thus, we would eventually accept AI with personal identity. Nevertheless, human evolution might direct towards a promising future. The digital evolution might not be as gloomy and degenerating as feared. If we follow Barfield's insight, we will reach final participation, freed from idolatry to enter a transcendental realm, when our consciousness evolves to complete abstraction. Instead of portraying a twilight of human civilization, Barfield's philosophy presents a cogent redemption of humans: not until we exhaust all remnants of original participation can we enter the final participation, the "mysterious kingdom of God". Not until we shed the last idolatrous piece will we realize divine intelligence. Not until we encounter our limit will we understand humanity and divinity.

We should rather face the digitalization of humanity head-on and reconsider

human identity against the contemporary technological context. Maybe this is the time for us to reconsider Jung's man in search of a soul and Eliade's eternal return: "things acquire their reality, their identity, only to the extent of their participation in a transcendent reality"³⁹. If we still maintain hope that humans should be unique as God's children, we suggest that divine intelligence might be the last stronghold of safeguarding human identity from other species. From this perspective, Swinburne can be right in his hypothetical experiment to retort materialistic brain replacement theories by excluding psychological continuity and physical continuity. The only essence that persists is soul, which Swinburne terms as "thisness", a uniqueness, of each person. This uniqueness of each person is different from what other critics view as mental or physical properties possessed by the body.⁴⁰

Swinburne is not alone in holding the soul as the distinctive feature to define human identity. In his *reductio ad absurdum* argument, Pruss R. Alexander concludes that artificial intelligence (computers or robots) cannot be considered to have a personal identity unless it possesses a soul: "then computers and robots cannot constitute persons unless, somehow, there is more to them than hardware and software, namely unless computers and robots will have souls. And by parallel, we cannot be persons unless we have souls"⁴¹.

In the end, AI may be granted a personal identity. But this is not the end of the world, for "AI elaborates a computer metaphor for the mind"⁴². Referring to Ricoeur's hermeneutics of metaphor, AI inspires humans to get metaphorical reference from its identity, thus we reflect upon human's

narrative identity through the metaphorical understanding of AI. If physical and psychological continuity cannot safeguard the uniqueness of human identity, transcendental continuity could. This continuity subsists or emerges through iconoclasm of consciousness against idolatry. Since we cannot scientifically prove the existence of

divine intelligence, the scientific and materialistic arguments that intend to prove humans are exquisite machines with super information processing functions can reversely illustrate that man could have been created as divine intelligence. The key difference between artificial intelligence and divine intelligence might lie in the soul.

BIBLIOGRAPHY

- Adiwardana, D. et al., "Towards a Human-like Open-Domain Chatbot", in *arXiv preprint ArXiv*, abs/2001.09977, 2020, <https://arxiv.org/pdf/2001.09977.pdf>.
- Alexander, Pruss R., "Artificial Intelligence and Personal Identity", in *Faith and Philosophy*, no. 26.5, 2009, pp. 487-500.
- "An AI-Written Novella Almost Won", in *Smithsonian Magazine*, 28 March 2016, <https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/>.
- "Artificial Intelligence and the Future of Humans", Pew Research Center, December 2018.
- Baker, Lynne Rudder, *Persons and Bodies: A Constitution View*, Cambridge University Press, 2000.
- Barfield, Owen, *Poetic Diction: A Study in Meaning*, Middletown, Wesleyan UP, 1973.
- Barfield, Owen, *Saving the Appearances: A Study in Idolatry*, Middletown, Wesleyan UP, 1988.
- Brown, Tom B., et al., "Language models are few-shot learners", in *arXiv preprint arXiv:2005.14165*, 2020, <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>.
- Carter, Matt. *Minds and Computers: An Introduction to the Philosophy of the Artificial Intelligence*. Edinburg University Press, 2007.
- Cole, David, "Artificial Intelligence and Personal Identity", *Synthese*, no. 88, 1991, p. 399-417.
- "Creative AI: Human Identity in the age of Virtual Flux," Africa 2019, <https://www.youtube.com/watch?v=eyJBs7dp8>.
- Damasio, Antonio, *Self Comes to Mind*, Random House, Kindle version, 2010.
- De Mul, Jos, "The Game of Life: Narrative and Ludic Identity formation in Computer Games", in J. Goldstein and J. Raessens (ed.), *Handbook of Computer Games Studies*, Cambridge, MIT Press, 2005, 251-266.
- Dehaene, Stanislas, et al., "What is consciousness, and could machines have it", in *Science*, no. 358.6362, 2017.
- Domingos, Pedro, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York, Basic Books, 2015.
- Eliade, Mircea, *The Myth of the Eternal Return: Cosmos and History*. Princeton: Princeton UP, 1971.
- Gent, Edd, "Artificial Intelligence is Evolving All by Itself", News, *Science*, 13 April 2020, <https://www.sciencemag.org/news/2020/04/artificial-intelligence-evolving-all-itself>.
- Gershgor, Dave, "Inside the mechanical brain of the world's first robot citizen", in *Quartz*, 12 November 2017, Retrieved 23 May 2018, <https://qz.com/1121547/how-smart-is-the-first-robot-citizen/>.
- Greene, Brian, *Until the End of Time*, Knopf Doubleday Publishing Group, Kindle version, 2020.
- Havel, Ivan M., "Artificial Intelligence: A Lesson in Human Self-Understanding", in R. Trappl (ed.), *Impacts of Artificial Intelligence*, Amsterdam, North-Holland, 1987, pp. 89-98.
- Heim, M., *The Metaphysics of Virtual Reality*, New York, Oxford University Press, 1993.
- Jung, C. G., *Modern Man in Search of a Soul*, Martino Fine Books, 2017.

- Kritikos, Mihalis, "Artificial Intelligence ante portas: Legal & ethical reflections", EPRS European Parliamentary Research Service, STOA Panel for the Future of Science and Technology, 2019, <https://www.europarl.europa.eu/at-your-service/files/be-heard/religious-and-non-confessional-dialogue/events/en-20190319-artificial-intelligence-ante-portas.pdf>.
- Kurzweil, Ray, *The Age of Spiritual Machine: When Computers Exceed Human Intelligence*, Penguin Books, 2000.
- La Mettrie, Julien Offray de, *Machine Man and Other Writings*, translated and edited by Ann Thomson, Cambridge & New York, Cambridge University Press, 1996.
- Li, Kenny, "AI Will Take Center Stage in Human Evolution", 22 July 2019, <https://towardsdatascience.com/ai-will-take-center-stage-in-human-evolution-ceed9fa22c25>.
- Moravec, Hans, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, 1988.
- Norton, Bonny, *Identity and Language Learning: Extending the Conversation, Bristol, Multilingual Matters, 2013*.
- Olson, Eric T., "Personal Identity", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2020 Edition, <https://plato.stanford.edu/archives/win2020/entries/identity-personal/>.
- Paul, Gregory, and Earl D. Cox, *Beyond Humanity: Cyberevolution and Future Minds*, Delmar Thomson Learning, 1996.
- Raphaël, Baroni, "Ce que l'intrigue ajoute au temps", in *Poétique*, Une relecture critique de Temps et récit de Paul Ricoeur, no. 163, 2010, pp. 361-382.
- Ricoeur, Paul, "Life in Quest of Narrative", in *On Paul Ricoeur: Narrative and Interpretation*, Mariner Books, Boston, 1991, pp. 20-33.
- Ricoeur, Paul, *Fallible Man*, Fordham University Press, 1986.
- Ricoeur, Paul, *Hermeneutics and the Human Sciences*, Thompson, J. B. (eds. & trans.), Cambridge & Paris, Cambridge University Press & Editions de la Maison des Sciences de l'Homme, 2005.
- Ricoeur, Paul, *Time and Narrative*, vol. 1, The University of Chicago Press, 1990.
- Rubin, Charles T., "Artificial Intelligence and Human Nature", *The New Atlantis*, no. 1, Spring 2003.
- Scarantino, Andrea and Ronald de Sousa, "Emotion", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2018 Edition, <https://plato.stanford.edu/archives/win2018/entries/emotion/>.
- Smolensky, Paul, "On the Proper Treatment of Connectionism", in *Behavioral and Brain Sciences*, no. 11.1, 1988, pp. 1-23.
- Swinburne, Richard, *Mind, Brain, and Free Will*, Oxford University Press, 2013.
- Swinburne, Richard, *The Evolution of the Soul*, Oxford, Clarendon Press, 1997.
- The Social Dilemma*, Netflix, 2020, <https://www.netflix.com/tw/title/81254224>.
- Tolkien, J. R. R., "On Fairy-Stories", in *The Monsters and the Critics and Other Essays*, London, Harper Collins Publishers, 1997.
- Vamos, Tibor, "AI: Subjective Views, Future, and Impacts", in R. Trappl (ed.), *Impacts of Artificial Intelligence*, Amsterdam, North-Holland, 1987.
- Van Inwagen, Peter, *Material Beings*, Ithaca, Cornell, 1990.
- Watch Dogs*, Ubisoft Montreal, 2016.
- White, Michael, *Maps of Narrative Practice*, New York, W. W. Norton & Company, 2007.
- Zhou, Li, et al., "The design and implementation of xiaoice, an empathetic social chatbot", in *Computational Linguistics*, no. 46.1, 2020.

NOTES

1. "Artificial Intelligence and the Future of Humans", Pew Research Center, December 2018, p. 3.
2. Eric T. Olson, "Personal Identity", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2020 Edition, <https://plato.stanford.edu/archives/win2020/entries/identity-personal/>.

3. Thomas I. White, "Dolphin people", 2 June 2020, retrieved 5 August 2020.
4. David Cole, "Artificial Intelligence and Personal Identity", in *Synthese*, no. 88, 1991, pp. 399-417.
5. Mihalis Kritikos, "Artificial Intelligence ante portas: Legal & ethical reflections", EPRS European Parliamentary Research Service, STOA Panel for the Future of Science and Technology, 2019, <https://www.europarl.europa.eu/at-your-service/files/be-heard/religious-and-non-confessional-dialogue/events/en-20190319-artificial-intelligence-ante-portas.pdf>.
6. Eric T. Olson, "Personal Identity", 2020.
7. *Ibidem*.
8. Julien Offray de La Mettrie, *Machine Man and Other Writings*, translated and edited by Ann Thomson, Cambridge & New York, Cambridge University Press, 1996.
9. Peter van Inwagen, *Material Beings*, Ithaca, Cornell, 1990, p. 149.
10. Owen Barfield, *Poetic Diction: A Study in Meaning*, Middletown, Wesleyan UP, 1973.
11. David Cole, "Artificial Intelligence and Personal Identity", 1991.
12. "Sofia (robot)", *Wikipedia*, 2021, [https://en.wikipedia.org/wiki/Sophia_\(robot\)](https://en.wikipedia.org/wiki/Sophia_(robot)).
13. "Meet the first-ever robot citizen – a humanoid named Sophia that once said it would 'destroy humans'", in *Business Insider*, 28 October 2017, <https://www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10>.
14. Dave Gershgorn, "Inside the mechanical brain of the world's first robot citizen", in *Quartz*, 12 November 2017, <https://qz.com/1121547/how-smart-is-the-first-robot-citizen/>.
15. Paul Smolensky, "On the Proper Treatment of Connectionism", in *Behavioral and Brain Sciences*, no. 11.1, 1988, p. 3.
16. Matt Carter, *Minds and Computers: An Introduction to the Philosophy of the Artificial Intelligence*, Edinburg University Press, 2007, pp. 204-205.
17. *Ibidem*, p. 206.
18. Stanislas Dehaene et al., "What is consciousness, and could machines have it", in *Science*, no. 358.6362, 2017, pp. 486-92.
19. *Ibidem*, p. 486.
20. *Ibidem*, p. 492.
21. *Ibidem*.
22. Domingos Pedro, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York, Basic Books, 2015.
23. Brian Greene, *Until the End of Time*, Knopf Doubleday Publishing Group, Kindle version, 2020, p. 156.
24. *Ibidem*.
25. Owen Barfield, *Saving the Appearances: A Study in Idolatry*, Middletown, Wesleyan University Press, 1988.
26. Baroni Raphaël, "Ce que l'intrigue ajoute au temps", in *Poétique*, Une relecture critique de Temps et récit de Paul Ricoeur, no. 163, 2010, pp. 361-382.
27. Michael White, *Maps of Narrative Practice*, New York, W. W. Norton & Company, 2007.
28. Antonio Damasio, *Self Comes to Mind*, Random House, Kindle version, 2010.
29. Bonny Norton, *Identity and Language Learning: Extending the Conversation*, Bristol, *Multilingual Matters*, 2013, p. 45.
30. Paul Ricoeur, *Fallible Man*, Fordham University Press, 1986, p. 138.
31. *The Social Dilemma*, Netflix, 2020, <https://www.netflix.com/tw/title/81254224>.
32. J. R. R. Tolkien, "On Fairy-Stories", in *The Monsters and the Critics and Other Essays*, London, Harper Collins Publishers, 1997, pp. 109-161.
33. Joe de Mul, p. 254.
34. *Ibidem*.
35. *The Social Dilemma*.
36. *Watch Dogs*, Ubisoft Montreal, 2016.

37. Mihalis Kritikos, 2019.
38. Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, 1988.
39. Mircea Eliade, *The Myth of the Eternal Return: Cosmos and History*. Princeton: Princeton UP, 1971, p. 5.
40. Richard Swinburne, *Mind, Brain, and Free Will*, Oxford University Press, 2013, p. 155.
41. Pruss R. Alexander, "Artificial Intelligence and Personal Identity", in *Faith and Philosophy*, no. 26.5, 2009, pp. 487-500.
42. Ivan M. Havel, "Artificial Intelligence: A Lesson in Human Self-Understanding", in R. Trappl (ed.), *Impacts of Artificial Intelligence*, Amsterdam, North-Holland, 1987, p. 92.