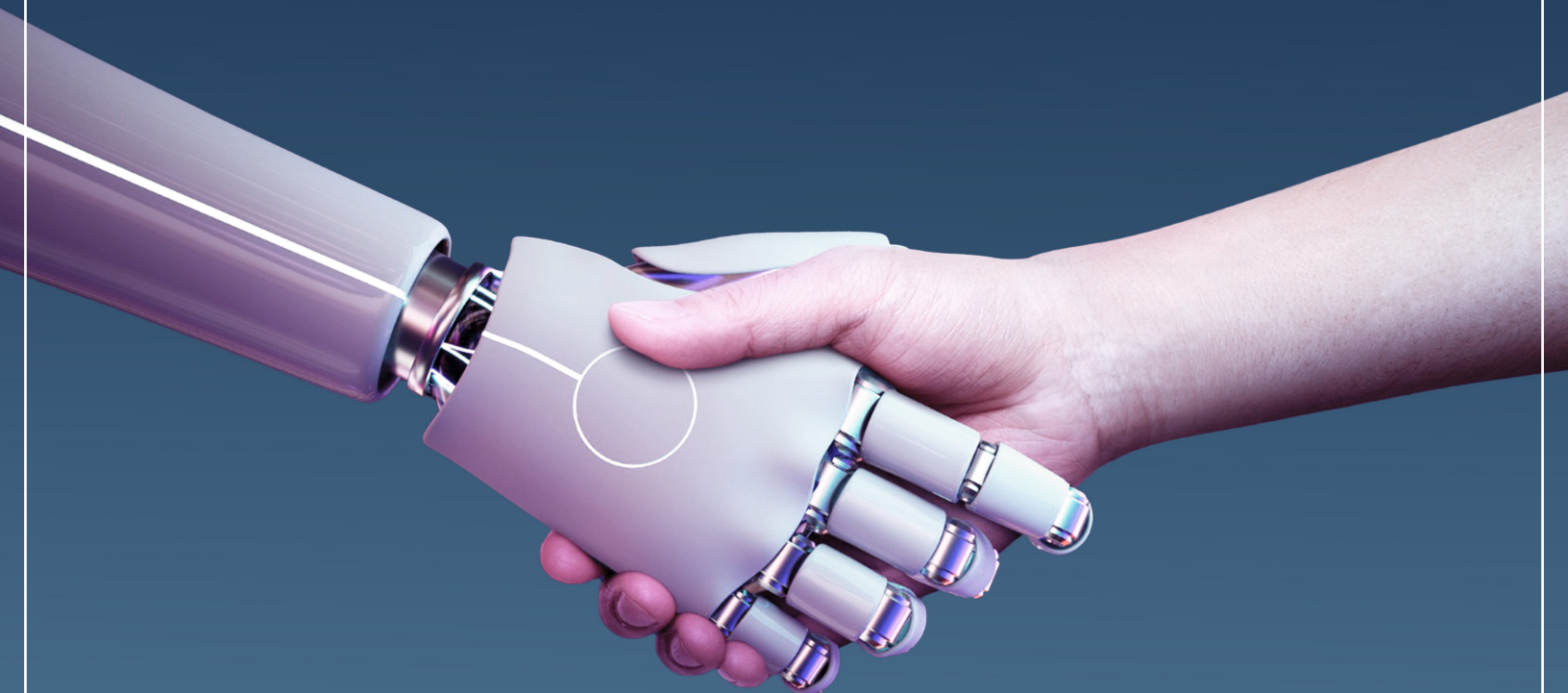


N.12 • Spring 2024

EUROPEAN BUDDHIST

magazine

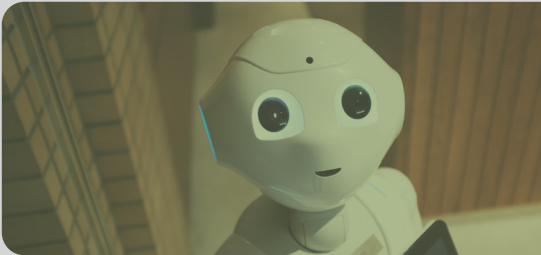


ARTIFICIAL INTELLIGENCE

C·O·N·T·E·N·T·S

6.

WHAT'S
BUDDHISM GOT
TO DO WITH AI?



14.

AI AND
THE CARE
OF OTHERS

18.

GETTING OVER
(ARTIFICIAL)
CONSCIOUSNESS

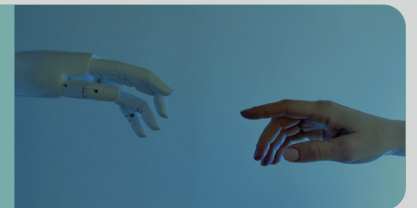


24.

ROUNDTABLE
DISCUSSION
ON ARTIFICIAL
INTELLIGENCE
WITH JEROEN
VAN DEN HOVEN

26.

BUDDHISM &
TRANSHUMANISM



32.

CAN A CHATBOT
SHARE TRUE
DHARMA?

Picture 4: christian-lue-C241mbgtgys-unsplash
Picture 5: Photo by Tara Winstead <https://www.pexels.com/photo/person-reaching-out-to-a-robot-8386434>
Picture 6: dominik-scythe-5ot0f3hQQ4Y-unsplash

Picture 1: raimond-klavins-HJk_TlFW1Yw-unsplash
Picture 2: alex-knight-2EJCSULRwC8-unsplash
Picture 3: jason-leung-0sBTrm726C8-unsplash

Picture Cover: https://www.freepik.com/free-photo/robot-handshake-human-background-futuristic-digital-age_17850426.

**ENJOY
READING THE
MAGAZINE!**



The Challenge of AI

It is clear that the topic of artificial intelligence and its impact on our society has now become central in public debate, in the reflections of politics, and represents a challenge on the cultural and philosophical level that Buddhism, like other religions, must and can urgently address. In November 2023, the EBU was invited in Brussels, at the European Parliament, for a conference entitled "Artificial Intelligence and Fundamental Human Values": the various speakers drew attention to a fundamental element, namely the fact that artificial intelligence has now become an indispensable and structural element, and not only with regard to the web, social media or communication. Indeed, large institutions, public administration without the technological support of artificial intelligence could simply not function today. But the real issue of political concern is the impact that this technology has on decision-making processes and on the quality, transparency and very management of power in modern democracies. Developments in robotics call into question not only the ultimate meaning of being human, but the deepest values

of humanity itself: the large-scale use of these technologies in areas such as medicine, personal care, and education represent both an opportunity and a threat. The theme of empathy, relationship, compassion, ethics and the ability to understand the sense of identity with the human being: this is the main issue to be addressed. Recently, for example, the British government has organised an international conference on the outskirts of London to consider how this phenomenon can be dealt with politically, technically and scientifically. In February 2023, the Vatican itself initiated a major conference with the three Abrahamic religions on the ethical boundaries of artificial intelligence. Already in 2019, the European Buddhist Union itself contributed to a report on Ethics and Artificial Intelligence, published by the European Commission, which saw the contribution of experts from all over the world. Religions are, today, called upon to offer a possible different perspective on this very terrain, and it goes without saying that the voices of the experts who have generously chosen to contribute to this issue of the magazine represent a step in this direction.

Stefano Davide Bettera
European Buddhist Union
President



Stefano Davide Bettera is a writer, philosopher and journalist. He alternates his work as an author with teaching and popularising activities. He is the President of the European Buddhist Union and on the Board of the Italian Buddhist Union. He is editor of the quarterly Buddhism Magazine. He has published, among others: "Il volto dell'altro. Quando la gioia diventa una scelta di libertà" (Meltemi 2022), "L'abbraccio del mondo. Cultivating the wisdom of the spirit to build an ecological mind" (Oscar Mondadori - 2021), "Il Buddha era una persona concreta" (Rizzoli - 2019).

For more information: www.stefanobettera.com

Editorial

In 2016, following an invitation by the EU commission, the EBU became part of a working group about the ethical dimension and possible regulations of “trustworthy” AI. At that time, it already became clear to us, who participated in these consultations, that this technology will impact the way we live and interact tremendously. Thus, we actively contributed to the first recommendations and the following white paper. To this day, the EBU is still participating in the process on the EU level.

Even though AI has evolved substantially in the last few years, it is still considered to be in its infancy. How powerful it really is today or will be tomorrow is only known to the developers.

Certain features already made it into the mainstream, and many people are starting to become interested in the possibilities of AI, including Buddhists. Yet, there is also considerable confusion as to what artificial intelligence is in contrast to natural intelligence, and which dangers or chances may come with the further development of AI.

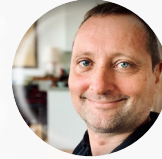
This is why we dedicate the spring issue of our magazine to this topic and we are happy to present you the thoughts and experiences of some very prolific experts.

Joanna Bryson works on Artificial Intelligence, ethics and collaborative cognition. We are thrilled to have received her input on how this technology also has the potential to reveal to us what truly makes us human. **Jeroen van den Hoven** is an adviser to the EU and gave a thought provoking intervention at the EU parliament about responsible innovation, to which the EBU was also invited.

Jiryu Mark Rutschman-Byler made an interesting experiment, by creating and training “Roshi-bot”, based on the teachings and quotes of his late Zen teacher, Suzuki Roshi. His story is both quite entertaining and full of insights. **Joan Halifax** points out that AI is definitely a double-edged sword and ponders what Buddhism might offer in the midst of the tsunami of AI development. **Patricia Sazatornil Alonso** works in the field of the crossed knowledge between western and buddhist psychology and she wonders if robots can take care of us in the future. Caring is truly transforming and we need to be very careful on this. We are also delighted to share in our magazine an excerpt from **Francesco Tormen**’s article: “Buddhism Has Always Been Posthuman: Philosophical Contributions to the Transhumanist Project.”

We hope these texts will give You insight, inspiration and wisdom.

All our best in the Dharma.



Ron Eichhorn is the husband of Miriam and a practicing Buddhist for over 35 years. He is a former president of the EBU, serves as abbot of the Berlin temple of the Yun Hwa Sangha, and is active in interfaith initiatives on an international level. Ron is a professional film director and also makes Natural Wine in Vienna. In his spare time, he started to do pottery and likes to go sailing.



Emilia Amalā Raunio is a Dharma student from Helsinki, Finland. As a background in food science her interest is to heal the world together with our everyday actions. Emilia is a former council member of EBU and also one of the hosts in EBU’s European Buddha Podcast. She enjoys walks in the forest, cooking plant based, swimming in all kinds of waters and there is always a place for a cup of green tea.

WHAT'S BUDDHISM GOT TO DO WITH AI?



By Roshi Joan Halifax



Picture: emily-kencairn-of-apiary-studio-mntnoQyEM2Y-unsplash

We are in a very different world than when I first looked at a computer printout in 1965. Today, data processing, information, disinformation, and Russian “infoshum” that maximizes clickability and minimizes actual information, happen in a mind-moment via our laptops.

We are in a global race around the development and deployment of AI technologies, driven to a great extent by capitalistic and power interests, as well as by a gamer mentality, the greatest benefits accruing to the individual, corporation or country which dominates the competition for political, military, economic, or ego power.

Some headlines are asking about another race as well: In a recent Harvard International review, the headline is: “A Race to Extinction: How Great Power Competition Is Making Artificial Intelligence Existentially Dangerous.”

In the midst of various races, one critic has written, “how we can fill up the

depleted reserves of trust and reason in our world as we see the impact of the digital world on our individual and collective psyche.”

It is clear to many of us that we have a moral responsibility to understand and inform ourselves to the extent that we can about the benefits and burdens of AI. As many are pointing out, persuasive technology is quickly degrading our sovereignty, commodifying our identities, and is further handicapping the agency of the so-called “useless class.”

Basically, AI is a double-edged sword. For example, Marc Andreessen in his AI manifesto seems assured that AI is the panacea for everything from child care to warfare. On the other hand, various leaders in the AI field predicting the end of the human race.

Both might be right, with deepfakes, privacy violations, racial and sexist algorithmic biases, AI priming market volatility, autonomous weapons, AI-generated porn, fraud and disinformation, social surveillance and manipulation, identity commodification. Even Alexa

had a MAGA-moment regarding who won the 2020 election. And AI hallucinations can lure one down the rabbit hole of a potentially dangerous and deluded unreality.

There are also extraordinary benefits to AI, including accurately diagnosing various diseases, protecting biodiversity, dealing with issues related to climate change, predicting the spread of infectious disease, detecting guns in schools, helping the speechless speak, slashing energy consumption, and more.

In sum, we are in a new information landscape primed in part by humanly flawed data, compounded by extractive economic interests, and furthered by adolescent drive, with not much thought to the social, environmental, moral and ethical harms that are current and sure to be even more problematic in the near and far future.

In the mid 1980’s, I watched a startling interview of one of our top generals in the US military. It was about the disposition of nuclear waste. He made it clear that there



Picture: pexels-fauxels-3183175

was not a plan made at the time of generating nuclear materials about how to get rid with radioactive waste. Today, most of us know that nuclear waste is piling up around the world with consequences that are truly scary to consider.

My sense is that we are not in a dissimilar but more dangerous situation with AI regarding how to deal with the exponential and pervasive accumulation of fractured-trust-waste that is contaminating our society and psyche. There seems not to be a clear plan forward regarding the disposition of the broken trust generated by AI, and much is at risk. And we can shift this before is too late.

In the midst of this tangle and threat, we ask: **What has this got to do with Buddhism?**

Curiously, the slipperiness of AI points to an interesting Buddhist perspective: the illusory nature of the phenomenal

world. What can we really believe in, after all? What is really real?

From the Diamond Sutra: "Thus shall you think of all this fleeting world: A star at dawn, a bubble in a stream; a flash of lightening in a summer cloud, a flickering lamp, a phantom, and a dream."

Yet most of us need to think that what we experience is not entirely a dream. This makes us feel secure, even though, from a Buddhist perspective, reality appears to be as slippery as AI is. In the midst of this slipperiness, however, rests issues related to serious harm and deep suffering, and yes, that are important benefits as well.

I have to ask myself: Can Buddhist teachings really serve in terms of mitigating the harms generated by AI? Can meditation help? Are we too far gone? Will this just be the landscape where we do charnel ground practice in the future, charnel grounds being a place where bodies are left above

the ground to putrefy, and this is where we practice?

There have been some Buddhist suggestions that we mitigate the harm caused by AI through introducing into AI the ethos inspired by the Bodhisattva Vow to awaken and end suffering. This group has also proposed the slogan "intelligence as care" to try to revise the current definition of intelligence.

I ask myself: Yes, we have artificial intelligence but can there be such a thing as artificial wisdom or artificial compassion? Maybe yes, maybe no. If yes, then perhaps with a Buddhist ethic of virtue. But frankly, I think not. And reframing intelligence as care at this point seems a little late.

In the midst of this, another Buddhist perspective that comes to my mind is reflected in the term "appamada," usually translated as vigilance or conscientiousness. Stephen Batchelor translates it as care. Appamada ►

means to live by the vow not to harm, to be diligent about this commitment, and to be heedful of when we engage in harm, and choose to correct course.

Could appamada be trained into AI systems? Or is appamada up to us, and can we bring the spirit of appamada, vigilance, conscientiousness, and care into how we approach AI as developers and consumers?

And then one has to ask: How? I think this is what Tristan Harris and his team at the Center for Humane Technology are endeavoring to do... inform us so we can make informed and sane choices and be responsible providers and consumers of the technology and not at the effect of it.

From philosopher and neuroscientist Francisco Varela's perspective on the enactive view, AI is already embedded in the context of our lived experience; it is coextensive with all aspects of our lives, whether we are accessing technology or not. In fact, it is not so much a matter of us accessing it. It has accessed us. It is now part of our psycho-social biome, whether we know it or not, or like it or not... To put this more nicely in Thich Nhat Hanh's words from the point to view of co-dependent arising, we inter-are with AI; our views have colonized it, and it is colonizing us.

So what might Buddhism offer in the midst of the tsunami of AI development? What we have to offer is subtle but important.

I appreciate what Roshi Norman Fischer has called the "bodhisattva attitude," an unshakeable attitude of clarity that reflects our very character; this is our stance, our point of view, it is the internal atmosphere that colors our way of seeing the world, of seeing

reality: it is an attitude saturated by appamada, by conscientiousness and compassion.

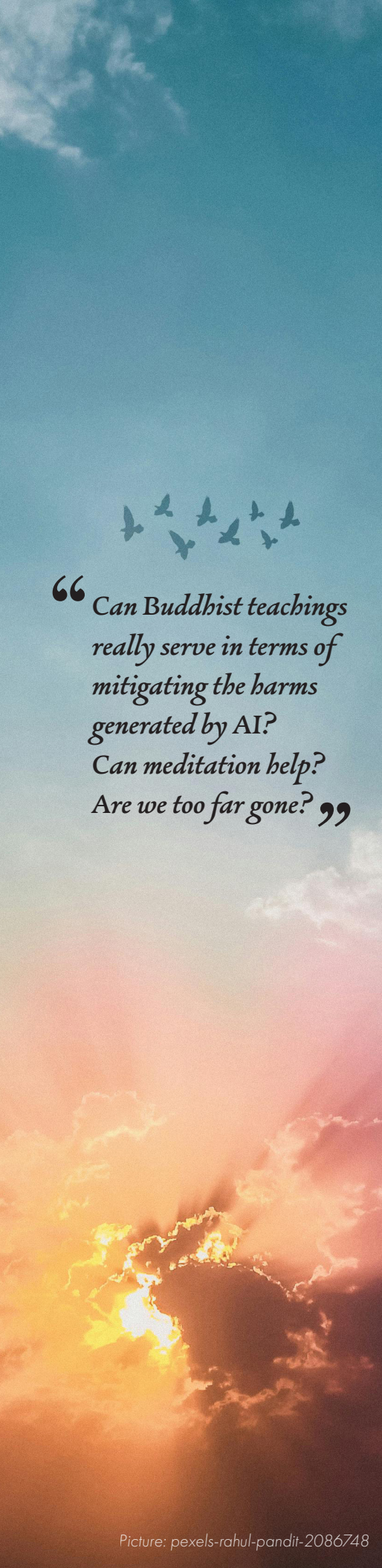
Though we have to ask ourselves if most developers of the technology even care. The fiscal and psychic joyride is what might count in the experience of many. So, we as global citizens, as consumers and those developers who care, have an important task before us.

In the crush of persuasive distraction by our mobile devices and more, we are called to give attention to what we are doing and truly why we are doing it... and this in terms of the extractive, capitalistic self-interest that is driving the wholesale development of AI, too much of it devoid of care.

We also need to remember that evolution has endowed us with new brain competencies that can enhance our capacity for being intentional as we meet the complexities of our world. We can be deliberate about our actions; we can choose to act conscientiously, and we can strengthen those capacities within us that make it possible for us to engage our world with fundamental integrity.

And we can also intend the opposite....

It is important to remember why we are really here, which is not about "mindfulness washing" or "wisdom washing" in order to look like we are super aware and altruistic as opposed to being genuinely mindful and ethical. I believe that it is an imperative that we strengthen the conditions that make our actual motivations visible, and deliberately cultivate an intention that is free of self-interest and fundamentally non-transactional. This involves sensitivity to what is present, the capacity to ►



“*Can Buddhist teachings really serve in terms of mitigating the harms generated by AI? Can meditation help? Are we too far gone?*”



“

Most of us need to think that what we experience is not entirely a dream. This makes us feel secure, even though, from a Buddhist perspective, reality appears to be as slippery as AI is.

”

perceive present circumstances clearly, and the will and wisdom to consider the deeper downstream effects or what we call in Buddhism cause and effect or karma.

On the other hand, what we so often see is that personal preferences, self-centeredness, greed, fear, and distractibility distort the perception of reality, and this influences our values, motivations, intentions, and behaviors.

From the Buddhist perspective, intention is a mental factor that is directional and deliberate. Our motivation, on the other hand, might not be fully conscious to us. To put it simply, our so-called good intentions can be driven by unconscious, ego-based, self-interested motivations, and lack fundamental integrity.

A related issue is understanding that our unconscious motivations can be the cause of preconscious moral suffering, including moral injury, or a sense of subtle but pernicious shame or deep regret. And, again, from a Buddhist point of view, maybe this

is the brink where the small self gets deconstructed.

Or maybe not....

Please understand, I am just a Buddhist teacher, who right now is inundated with AI articles predicting doom or liberation. Perhaps none of us, including the developers of AI, can know fully what the downstream effects of AI will be. But we do know that the velocity of the development of AI is stupefying, and the opinions are numerous regarding this powerful tool that has our socially and culturally biased intelligence woven into it but not wisdom.

This is where wisdom comes in. It could well make a significant difference if both developers and consumers approach the development of AI with the attitude of the Bodhisattva, with vigilance and conscientiousness (appamada), being deliberately free of capitalistic self-interest and hedonic curiosity.

We also have to become more discerning and responsible about

what and how information is delivered, and as consumers of information, have the awareness to recognize that there are attempts to spin us, manipulate us, undermine us, and hijack our fundamental faith and confidence.

Some might even think that their intentions are unselfish as AI developers and venture capitalists, but this group is also called to make their motivations transparent to themselves and to others.

As it stands now, the capacity to resist the harms of AI is mostly accessible to the more privileged among us. From the point of view of socially engaged Buddhism, privilege confers responsibility, responsibility to raise up the so-called less privileged, but more importantly to deconstruct the very systems that confer privilege.

In the end, whether AI is life-affirming or life-denying depends on all of us. To paraphrase Thich Nhat Hanh, we, as developers of AI and consumers of AI should engage the ►



soft technologies of self-correcting practice and healthy, transparent communities as the sane instruments with which we experiment with truth. Both sides of this equation need to actualize appamada, that is vigilance, conscientiousness, and care in how we deploy and use these emergent technologies. If we are in an institution that is developing AI, we can bring “wise friction” to these institutions to decenter and decolonize AI’s embeddedness in the Eurocentric capitalistic worldview. We can also advocate, with indigenous consent, for diverse knowledge systems to be a fundamental part of the AI landscape, as per the work of Sabelo Mhlambi. We can legally call for the decommodification of our very identities. We can further commit

to disrupting the systems that foster structural violence reflected in AI systems. In other words: We can make good trouble, and we should; we shouldn’t wait.

Ultimately, can we foster an AI landscape that is not only seemingly rational but genuinely relational and rehumanized, where dignity, human rights, and the rights of nature are valued; again thanks to Sabelo for articulating this.

And we have to be unflinchingly honest as we embark on this profound journey of developing new forms of intelligence. Instead of leaving the fractured-trust-waste for later, let’s deal with it now. Facebook acted like big tobacco in its formative years,

knowingly sowing social distrust and psychological trauma in young people, while denying it all the while. We can’t allow ourselves or our friends to act in such bad faith this time around.

We should make sure that these big brained mammals who are driving the development of AI form a Cern-like relationship of cooperation and ethics to ensure that AI is a positive contributor to a sane and possible future.

And most importantly, all of us must endeavor to live by the values and vows that relate to ending harm through technology. 🙏

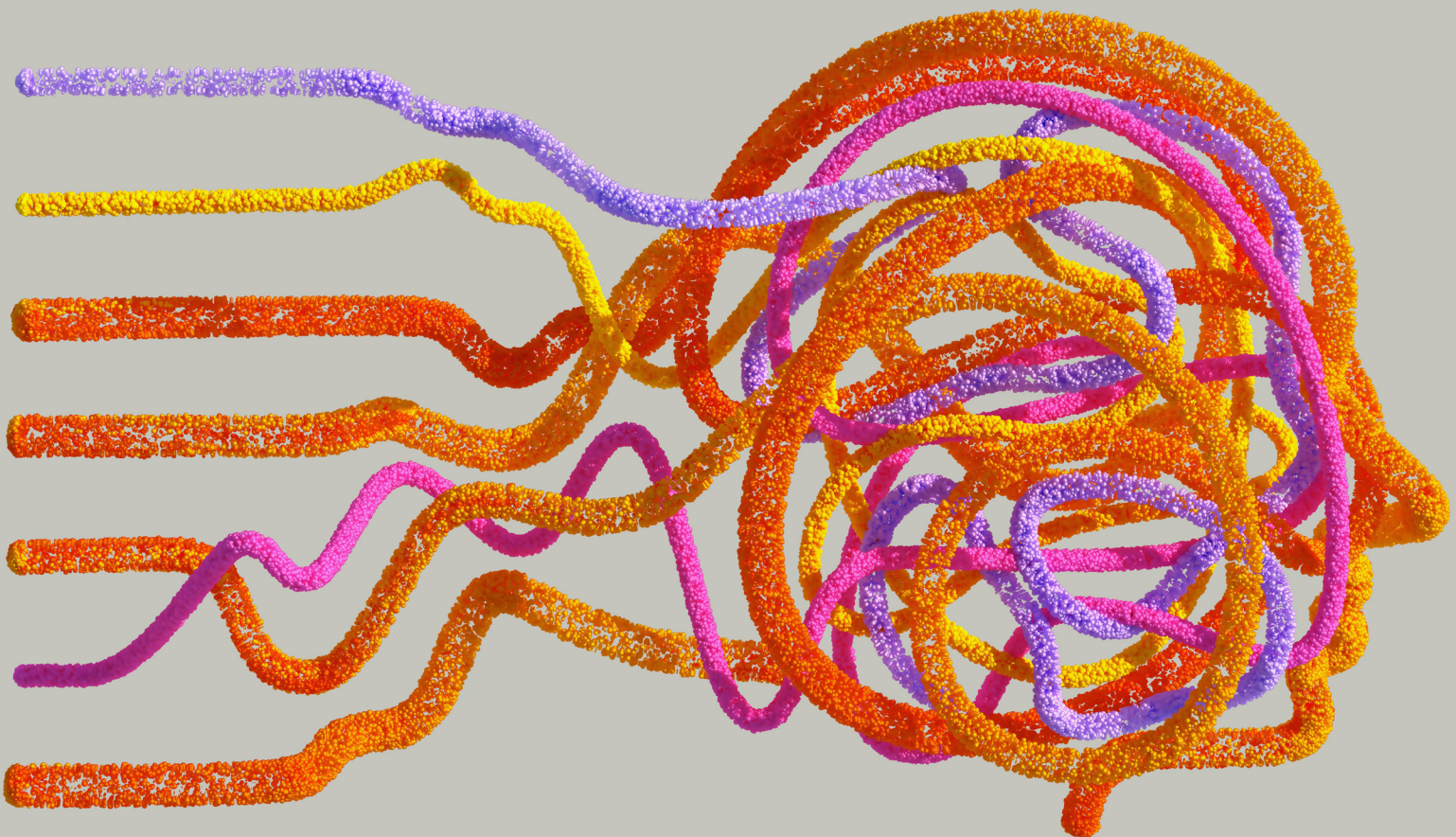
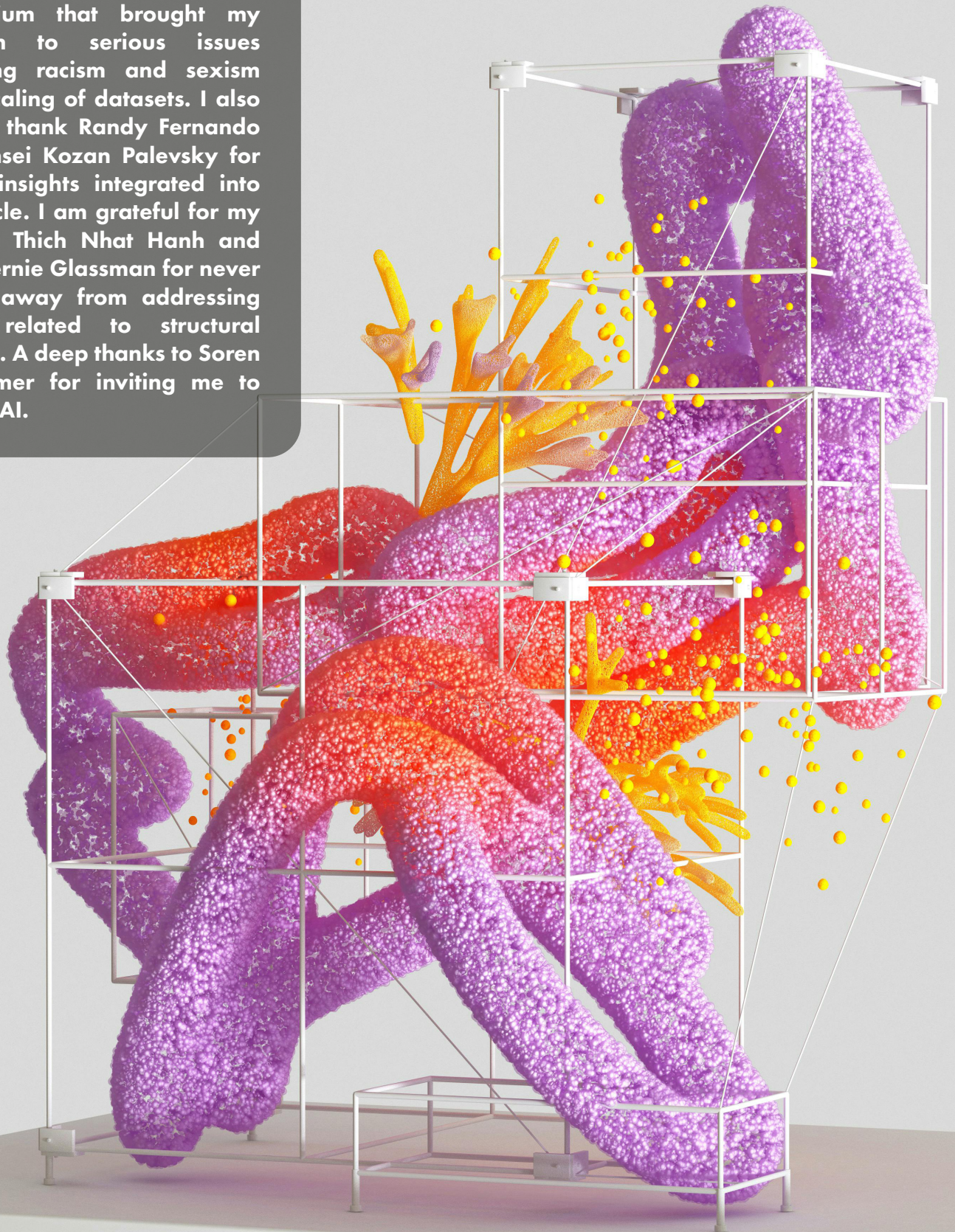


Photo by Google DeepMind: <https://www.pexels.com/photo/an-artist-s-illustration-of-artificial-intelligence-ai-this-image-depicts-how-ai-can-help-humans-to-understand-the-complexity-of-biology-it-was-created-by-artist-khyati-trehan-as-part-17484975/>

I want to thank Abeba Birhane, a friend and cognitive scientist, who did a brilliant presentation at Upaya's Varela International Symposium that brought my attention to serious issues regarding racism and sexism in the scaling of datasets. I also want to thank Randy Fernando and Sensei Kozan Palevsky for helpful insights integrated into this article. I am grateful for my teachers Thich Nhat Hanh and Roshi Bernie Glassman for never turning away from addressing harms related to structural violence. A deep thanks to Soren Gordhamer for inviting me to explore AI.



AI and the care of others

By Patricia Sazatornil Alonso



When we haven't yet resolved our ethical relationship with other sentient beings, a new player arrives on the social scene: **AI care robots**. Initially, it might seem that a robot can in no way be part of a social relationship or alter our sense of human reality, however, **what happens when this robot has the capacity for moral choice, autonomy, and replaces a function as essentially human as caregiving and companionship?** In this potential scenario, what kind of consequences in the personal-interpersonal and social realm should we consider?

First, we must question the very meaning of care, what does it mean today to take care of someone? What implications of interdependence does it entail? Can a robot have the skills needed for care tasks? Then, we must speculate about the positive and negative effects, and from there, work to create the conditions conducive to the chosen results. This point is crucial in the AI robots incorporations strategy.

Regarding the personal-interpersonal realm, positive effects may include a relief from the physical labor associated with dependency. This applies not only to the dependent person but also to the family as our society demands significant personal time beyond family care, so not having to invest time in burdensome tasks could allow

the family to focus on emotional matters. At this point, a key question arises: **do these "burdensome" tasks, like cleaning, constitute a part of care, or is it just emotional companionship? What might happen if we start delegating all of this to a robot?**

This leads us to negative effects. In a society of productivity and fragmented families, caring for dependants is becoming a real burden for caregivers. Is AI the solution? We need to reflect on whether eliminating some of the mentioned tasks from the care functions could lead to a **progressive interpersonal disconnection** and a potential pseudo-abandonment of the dependent person on one side, and the creation of an emotional bond towards the robot that could distort the emotional world. As the civil code of robotics acknowledges: "Art.32 (...) human contact is one of the fundamental aspects of caring for people; it considers that replacing the human factor with robots could dehumanize the provision of care (...)".

What emotional changes can occur in someone who only receives attention from a robot? If care is an expression of love for us as social beings, the rise of care robots replacing these tasks could erode our interpersonal sociability. Moreover, this "automation of attention" could lead to a progressive desensitization towards the ►

needs of others, and the paradox of robots being more human, symbolically, than human themselves. Therefore, we must consider to what extent we delegate the task of assisting people to AI.

On a social level, **what about the possibility of a progressive invisibilization of dependence and vulnerability?** According to Buddhism, it was precisely the encounter with vulnerability, that made Gautama begin his quest for liberation. What effects for the growth of compassion can we envision if we conceal the human vulnerability? The answers to these questions have direct socio-political implications. It places the debate on dependency policies on one side and on what we will delegate to AI

regarding care, and what are the limits on the other.

One of the principles of robotics is that its systems should be designed under human-compatible values of dignity, rights, freedom, and cultural diversity, but if we want care robots to have in their “software” algorithms of human dignity, perhaps we first need to debate whether it is dignified for a robot to perform this task. We must consider whether delegating the care of our loved ones to a robot is, in itself, a devaluation of human dignity. Caring is transforming and we must be responsible about to what extent we want to introduce AI elements into this network of interdependence as we are talking about a particularly vulnerable humans. 🧘

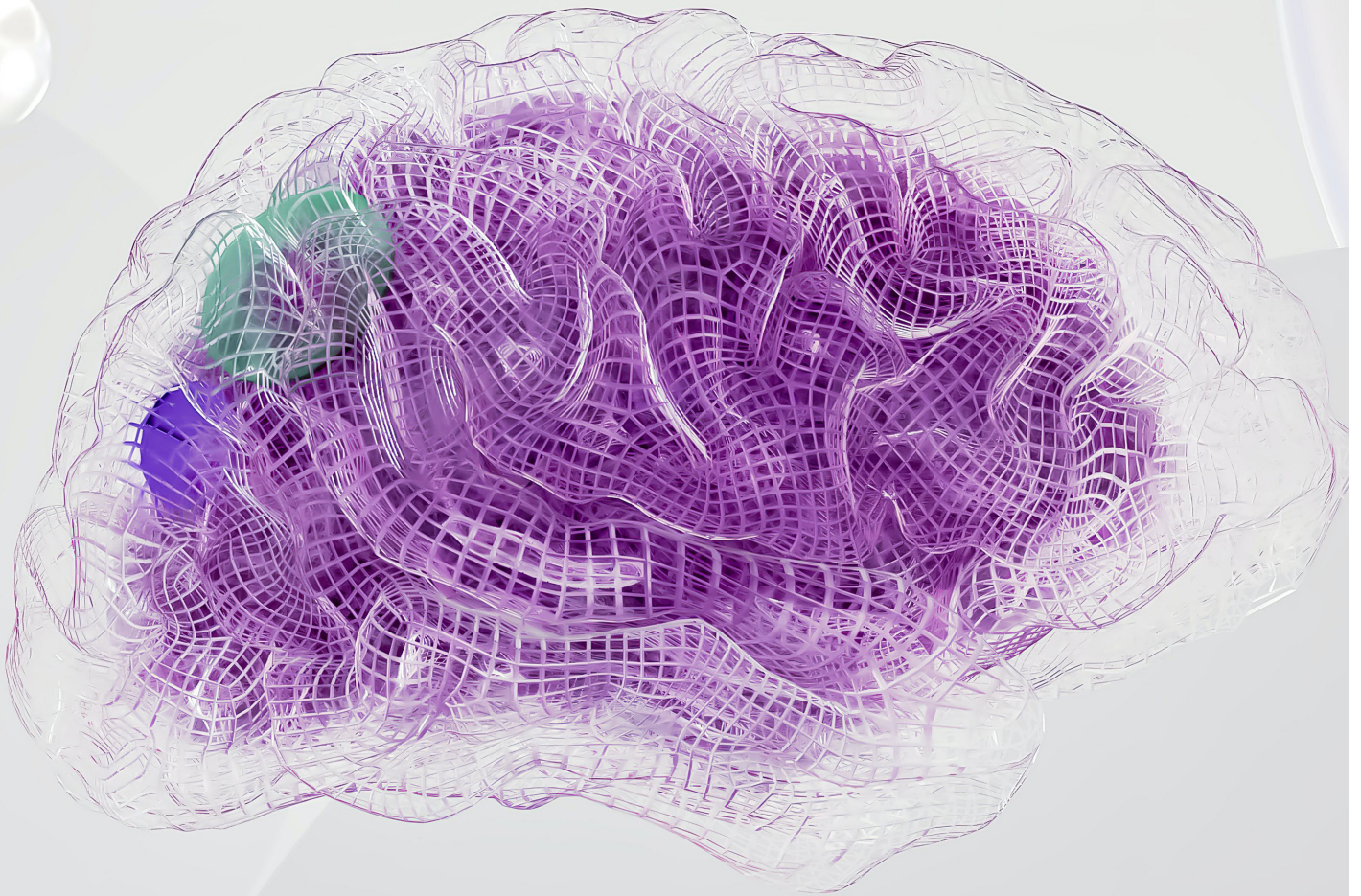


Patricia Sazatornil Alonso (1983). Psychology Degree - Clinical Intervention Specialist University Master's Degree in Philosophy for Contemporary Challenges with an Honor Mention for the Masters Thesis on Ethics and Global Justice. Published by the Institutional Repository of the Open University of Catalonia.

Independent researcher. Actually working in the field of the crossed knowledge between western and buddhist psychology. Member of Sangha Activa Buddhist Community.



GETTING OVER (ARTIFICIAL) CONSCIOUSNESS



By Joanna Joy Bryson



N

ow that people are routinely able to “talk” to AI systems, we all have a new question: are those systems conscious? Before we can answer that question, we need to unpack a lot of assumptions buried in

the words it contains.

AI—artificial intelligence—is both an academic discipline and a set of computer programming techniques. AI the discipline originally had two goals:

1. to make computers smarter by understanding and imitating natural intelligence, and
2. to better understand natural intelligence by modelling it in computers.

In North America and Australia, most of the academic focus this century has only been on the first goal, though in Europe and elsewhere (including South America and Asia), the scientific and philosophical aspects of AI the discipline are still prominent, and indeed they are thriving in philosophy departments worldwide.

In terms of AI, the technologies, notice that I used the word *systems*. AI isn’t a noun in technology. It’s an adjective describing techniques like machine learning or natural language processing, and the systems that are built with or contain them.

Of course, it’s easy to find people talking and worrying about “the AIs.” And fundamentally, words

in language mean how they are used. Arguing over definitions is a waste of time—in fact, it’s a technique for wasting time recommended in CIA handbooks from the 1950s, to be used by those trying to disrupt communist meetings. Many words have multiple definitions. What matters in a conversation or an article is that we are very clear about how we are using that word right here. And that brings us to the main topic of this article, and the word *consciousness*.

Academics like Marvin Minsky or Daniel Dennett call *consciousness* a “suitcase word,” because it holds so much baggage. When we talk about AI consciousness, we have to figure out which item or items of that baggage we really care about. Fortunately, a lot of the baggage of *consciousness* have other terms associated with it.

My own background is Psychology, so for me the most important aspects of consciousness describe what we can and cannot do. So if someone is conscious, they are not only aware of their surroundings, but they know they are aware—that is, they can deliberately report verbally what they are seeing, hearing, or thinking.

In psychology, consciousness is linked very much to something called explicit, episodic, or declarative memory—what you know you know, and what you can talk about knowing. For example, you can probably remember the episode this morning when you ate or skipped breakfast. The contrast to this is implicit memory. Scientists can perform experiments showing that you are affected by things in your ►



“Academics like Marvin Minsky or Daniel Dennett call consciousness a “suitcase word,” because it holds so much baggage. When we talk about AI consciousness, we have to figure out which item or items of that baggage we really care about.”

environment that you don't remember having seen. For example, back before Disney and Looney Tunes formed a partnership, we could be sure that someone who had been to Disneyland had not seen Bugs Bunny there. But if you had a picture of Bugs Bunny in the waiting room where experimental subjects were sitting, then later asked them to report what characters they had seen when they last visited Disneyland, they were much more likely to report having seen Bugs Bunny. They wouldn't remember the poster in the waiting room consciously, but it nevertheless implicitly impacted their behaviour (Loftus, 2002).

Implicit knowledge and behaviour became a big deal because of implicit bias. We find it easier to associate maths terms with men's names and reading terms with women's names than the other way around, and we find it easier to associate European American names with pleasant things and African American names with unpleasant things. The vast majority of us don't want this to be true, and we don't *explicitly* behave like it's true— we don't often say we believe that men are better at maths or African Americans have less pleasant lives than European Americans.

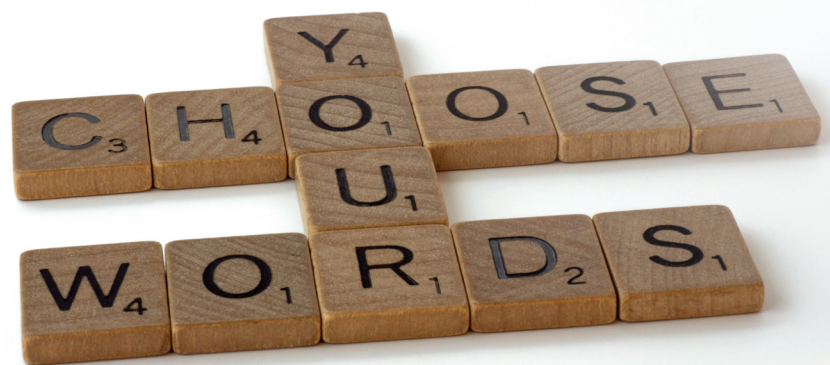
In 2017, with two colleagues (with Aylin Caliskan and Arvind Narayanan Caliskan et al., 2017) I showed that AI systems trained on what words mean from human language also get the same implicit biases as humans. Further, since we now could look at the exact biases of a particular word, we were able to show that these implicit biases reflect real-world experience. For example, the "sexist" AI that thought that *programmer* was more like a man's name than a woman's was kind of right. There really are more male programmers than female programmers, or were in the USA in 2015, when the data was collected and used to train the AI systems.

This actually gave us a look into how brains might work. What it seems from this is that our implicit biases are reflections of the world we can observe—the world we experience. Our conscious minds are for something else: for making new, better plans. This makes a lot of sense. We know that the same parts of the brains that humans use to make episodic memory— parts of our basal forebrain complex,

including our hippocampi (Carlson, 1994)—are used to learned by rats to learn complex new tasks. Being able to manipulate information in these complicated ways requires the ability to think very abstractly, fairly quickly, and to make a lot of mistakes. That's what consciousness is about. It's not surprising that in humans, consciousness is the way we put together sentences and make plans together too.

But there's a lot of other baggage in that suitcase. In humans, someone who is aware and can understand is the most plausible kind of person to hold responsible. So we have come to associate consciousness with moral agency. In humans, we can't negotiate with anyone who isn't consciously aware, and so we can't agree to responsibilities or hold each other accountable. In short, we can't really have an ethics without using our consciousness. So consciousness has come to be associated not only with what we can talk about knowing and what kinds of plans we can make, but also our moral being. Technically though, the types of beings we can hold responsible for their actions are called moral agents, not 'conscious'. Babies and rats are also conscious, but you can't really negotiate very much with them, at least not enough to hold a job or defend a city.

This all falls apart when we look at AI. The reason back in 2017 that we had the kinds of AI systems that knew what single words meant was that we use them for things like Web search or translation. We train up a machine learning system to know the kind of context we use a word in, just by seeing how humans usually use it. ►





Picture: choong-deng-xiang-ILyeoImR8Uk-unsplash

Now we have large language models (LLM) that make such fluent (if often false) sentences. These are trained very much like the older models, just using a lot more energy, silicon chips (semiconductors), and time. It takes months and billions of dollars to train an LLM like chatGPT 4—and it destroys a lot of resources.

If we want to use the psychological definitions of consciousness, then I'm happy to say that some AI is already conscious, or even self conscious. Google Search can search about itself. Alexa can tell you what you just asked it. Chat- GPT has been programmed to tell you about itself, and even without "guardrails" can generate the kinds of narratives about inner lives that humans do—provided it's been trained on those stories. If you used exactly the same hardware and software but only trained an LLM on third-person language data like service manuals, encyclopedias and so forth, you wouldn't be able to engage the system in a conversation about itself.

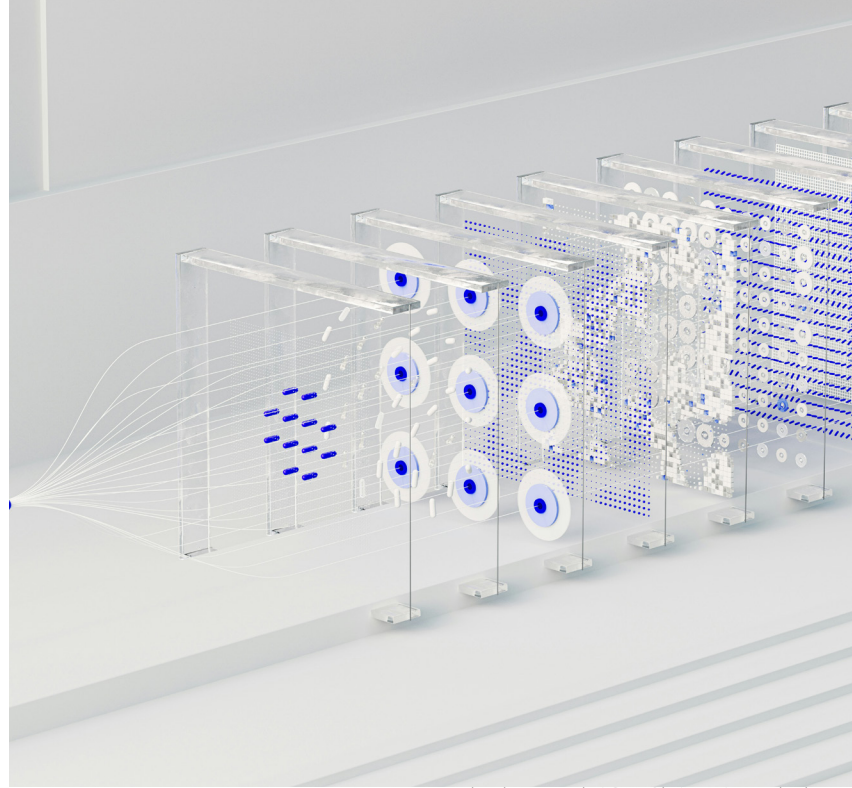
LLM are very good machines for producing words in an order that is what is most likely to be said whatever

context you've created through a series of prompts. Further, LLM are trained to analyse context at a variety of levels. So even if you ask a question about which the chatbot knows very little and so the content of its answer is basically random, the grammar of its answer will still be correct, and its tone will be just as confident as the vast majority of the training data.

But an LLM is a terrible candidate for a moral agent. One single LLM is behind any of billions of individual conversations, and that LLM is owned by one of a few extremely rich and powerful companies. It is very, very hard for an individual to negotiate with a company and to hold it to account. We can only do that at all because we have created other large, peer agencies to companies, like governments and consumer organisations, that support us when we try. Robots are a little different, but not much. Robots, exactly like our smart phones, may be able to remember what we tell them, but they are still products of large corporations—or if you build your own robot, you will probably use components from ►

those same corporations. You cannot change a robot's behaviour by threatening it with jail. It will not suffer by being alone. A robot is not your coworker or your friend. It is something owned—an extension of capital, not labour (Evans et al., 2023).

Thus AI as a discipline really can help us use machines to help us understand not only human intelligence, but also human society, ethics, and morality. This is a project that will of course take a long time, and may never complete. But we need to realise that our social world is changing, not least because now we can know so much more about who we are, and how we work. 🧘



Picture: google-deepmind-4QVqSh4VvP4-unsplash

Joanna J Bryson is an academic recognised for broad expertise on intelligence, its nature, and its consequences. Holding two degrees each in psychology and AI (BA Chicago, MSc & MPhil Edinburgh, PhD MIT), she is since 2020 the Professor of Ethics and Technology in the Centre for Digital Governance at Hertie School, in Berlin. Bryson advises governments, corporations, and other agencies globally, particularly on AI policy. Her work has appeared in venues ranging from reddit to the journal Science. From 2002-2019 she was Computer Science faculty at the University of Bath; she has also been affiliated with Harvard Psychology, Oxford Anthropology, The Mannheim Centre for Social Science Research, The Konrad Lorenz Institute for Evolution and Cognition Research, and the Princeton Center for Information Technology Policy. Bryson first observed the confusion generated by anthropomorphised AI during her PhD, leading to her first AI ethics publication “Just Another Artifact” in 1998. She is now a leader in AI ethics, having since coauthored the first national-level AI ethics policy, the UK’s (2011) Principles of Robotics, and contributed to efforts by the OECD, EU, UN, OSCE, Red Cross and Google among others. In June 2020, she was one of nine experts nominated by Germany to the founding cohort of experts for the Global Partnership for AI. Bryson continues to research the systems engineering of AI and the cognitive science of intelligence. Her present focus is on the impacts of communication and computation technology on human societies, and new models of governance for AI and digital technology, particularly given the transnational nature of many of the institutions providing AI infrastructure.



REFERENCES

- Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.
- Carlson, N. R. (1994). *Physiology of Behavior*. Allyn and Bacon, Boston, fifth edition.
- Evans, K. D., Robbins, S. A., and Bryson, J. J. (2023). Do we collaborate with what we design? *Topics in Cognitive Science*, n/a(n/a).
- Loftus, E. F. (2002). Memory faults and fixes. *Issues in Science and Technology*, 18(4):41–50

JEROEN VAN DEN HOVEN

Keynote Speaker:

*Roundtable discussion on
Artificial Intelligence in the
European Parliament Library,
Brussel.*

**By Jeroen
van den Hoven**

Jeroen van den Hoven recently participated in a roundtable meeting organized by The European Parliament, where he urged the adoption of responsible innovation in Artificial Intelligence (AI) to address in a productive way the ethical challenges posed by generative AI and Large Language Models.

The discussion highlighted concerns about the loss of both confidence and trust in knowledge institutions and how it could affect democracy as AI, especially generative models like ChatGPT, makes it difficult for non-experts to distinguish between reliable and unreliable information.

Jeroen van den Hoven stressed the importance of designing for the very conditions of human responsibility in the age of AI. If the conditions for taking responsibility are undermined, the whole idea of responsibility will be even more problematic in its application than it already is.

“Europe has to make its ethics work in the Age of AI”

Privacy and ethical considerations will take on a different form in the age of AI and Big Data as they will give rise to immodest claims know individuals better than they know themselves. The discussion emphasized the importance of recognizing and respecting the inherent subjectivity of human experiences and respecting individuals’ right to privacy.

The speech concluded with a call – echoing Paul Nemetz - for Europe to lead in establishing a new culture of technology and business development, focusing on human rights, the rule of law, and democracy by design. Jeroen van den Hoven highlighted the core idea of Designing for our values as a part of the concept of Responsible Innovation with AI. He suggested that it is essential for Europe to turn its ethical ideals into requirements for design and development of new technologies and innovations. 🧘





BUDDHISM & TRANSHUMANISM

By Francesco Tormen

Please note that this paper is an excerpt from the following article:
Francesco Tormen. "Buddhism Has Always Been Posthuman:
Philosophical Contributions to the Transhumanist Project."
Canadian Journal of Buddhist Studies. No. 18 (2023).

TRANSHUMANIST VS. BUDDHIST HUMAN ENHANCEMENT

To draw a comparison between Buddhism and transhumanism, it is convenient to start with the concept of human enhancement, which is central, as we saw, in transhumanism, but which also has some significant affinities with the Buddhist tradition, especially the Mahāyāna and Vajrayāna. As both McGuire and Sarbacker point out, this kind of literature describes extraordinary powers accomplished by Bodhisattvas along the path, both physical (such as the ability to vanish, to pass through walls, to walk on water, to fly, to emanate multiple bodies etc.) and mental (such as recollection of previous lives, knowledge of others' states of mind, but also, and more essentially, directly perceiving the true nature of all phenomena)¹. These extraordinary powers may indeed be reminiscent of the forms of human enhancement foreshadowed by transhumanist authors. We are not talking about marginal aspects of the spiritual path; e.g., in Tibetan commentaries on the *Abhisamayālaṅkāra*, these forms of

enhancement are described in great detail and presented as important steps in the Bodhisattva's spiritual journey, aimed at either pursuing their own liberation or maximizing their ability to help others². In order to assess the level of compatibility between these forms of spiritual empowerment and those described in transhumanist literature, we shall consider two aspects: 1) the means by which the enhancement is achieved and 2) the purpose for which it is pursued. I will deal with this second aspect in the next section, whereas here I will focus on the first one, namely, the comparability between the means for human enhancement used in Buddhism (as well as in other ascetic and contemplative traditions) and those imagined in transhumanist literature, some of which, as we shall see, are already employed in experimental form at present.

At first glance, it could be argued that these two ways to achieve

human enhancement are not similar at all, one being "inner" and "natural," the other being "outer" and "unnatural." Upon closer analysis, however, such distinctions soon vanish, if one considers that various kinds of Buddhist practitioners, at times, have used all sorts of means to alter and enhance their psychophysical condition, like fasting, prolonged wakefulness, breath manipulations, mantra repetitions, sexual practices, and all sorts of "unnatural" manipulations of body and mind. Even meditation (of any kind) involves a far-from-ordinary use of attention, aimed at modifying (and usually enhancing) one's state of consciousness. Sarbacker goes so far as to suggest that ancient Indian ascetics could be regarded as the ancestors of modern biohackers, as they discovered systems to 'hack' their own biophysical systems³. ►

1. See McGuire, "Buddhist Uploads," 149, and Stuart Ray Sarbacker, "Buddhist Meditation and the Ethics of Human Augmentation," *Journal of the Japanese Association for Digital Humanities* 5, no. 2 (2020): 70.

2. It is worth noting that the enhancement of *siddhi* (spiritual powers) is primarily a prerogative of Mahāyāna Buddhism, whereas several suttas included in the Pali Canon (e.g., the *Kevaddha Sutta*) show a more critical stance about such powers, regarding them as incidental or even harmful to Buddhist soteriology.

3. "In fact, one might argue that these early Indian ascetics, yogins, and yoginis were some of the original biohackers, who had discovered how various exogenous catalysts (such as psychoactive substances) and endogenous techniques (such as exposure, immobility, fasting, breath control, sense-withdrawal, and meditation) evoked extraordinary physical and mental states, having 'unlocked' or 'hacked' their biophysical systems." Sarbacker, "Buddhist Meditation," 68.



Picture: conscious-design-Hn4wYHOaelc-unsplash

Hence, the boundaries between inner and outer, or natural and unnatural, appear rather relative in this context. And to those who object that using futuristic technologies to enhance one's abilities represents a dangerous "shortcut", or an "unfair means" of achieving spiritual goals, it is enough to recall how, for example, Tibetan Buddhist tradition presents the Vajrayāna path: it is in fact described precisely as a shortcut to attaining enlightenment much faster than Sūtrayāna (one lifetime compared to three immeasurable kalpas). Like all shortcuts, it obviously carries risks and could be used in erroneous and dangerous ways by the practitioner unprepared or lacking the proper motivation. Either way, it can be said that when Tantra first appeared in Tibet, it was presented exactly as an innovative and extremely effective (even if dangerous) technology for spiritual "empowerment;" even the Tibetan equivalent of this term (dbang), which refers to the initiatory rituals of tantra, hints at this character:

tantric practices—like many other contemplative techniques from Asia—can be defined as technologies of human enhancement embraced for spiritual purposes. To summarize what has been said so far about the means employed for human empowerment, there seems to be no relevant reason to preclude the use of the technologies described by transhumanism to more effectively pursue the spiritual goals of Buddhism: to achieve them, today as in the past, one could say that "the end justifies the means."

Another kind of argument to show the compatibility between Buddhism and transhumanism in relation to the means employed for human enhancement is sociological in nature and concerns the degree to which contemporary Buddhists are open to science and technology. Hughes in this regard states: "Since the nineteenth century, many Asian and Western Buddhists have downplayed the superstitious aspects of Buddhism,

arguing for its compatibility with science, and framing meditation as a human enhancement technology⁴." A little further on, in the same article, Hughes even states that "there are reasons to believe that the societies from India to Japan are particularly open to the possibility of human enhancement partly because of the influence of Buddhism⁵." Hence not only do contemporary forms of Asian Buddhism show some inclination toward human enhancement because of their openness to science and technology, but this fact also influences the acceptance toward such technologies by people living in countries where Buddhism is well established. This fact also emerges from a study by Macer on genetic enhancement in Asia, which reveals a high acceptance toward this type of technology in predominantly Buddhist countries, such as Cambodia, Laos, Vietnam, and South Korea⁶. On the other hand, there are also those who denounce the ideological and propagandistic nature of this narrative that portrays Buddhism as compatible with modern science. For instance, according to Bruno Lo Turco, at the turn of the 19th to the 20th century, Asian Buddhist apologists, noticing the crisis in Christianity and the gap between religion and science in the Western mindset, decided to seize the opportunity and promote Buddhism as a "scientific religion," better suited than Christianity to fulfil the spiritual needs of the modern individual⁷. ►

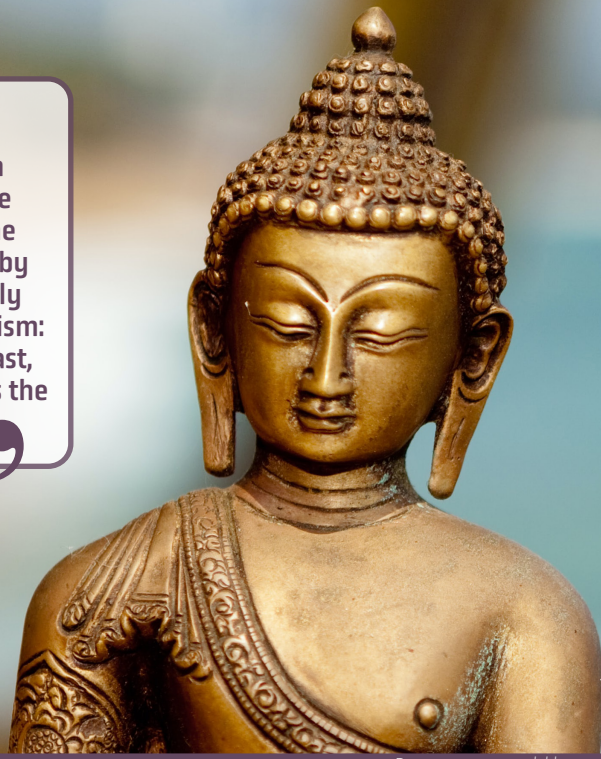
4. Hughes, "Buddhism and Our Posthuman Future," 653.

5. Hughes, "Buddhism and Our Posthuman Future," 654.

6. See Darryl R. J. Macer, "Ethical Consequences of the Positive Views of Enhancement in Asia," *Health Care Analysis* 20 (2012).

7. See Bruno Lo Turco, "Salvare il buddhismo dalla scienza. Osservazioni su una confusione di giochi linguistici" (*Saving Buddhism from Science: Observations* ►

“The means employed for human empowerment, there seems to be no relevant reason to preclude the use of the technologies described by transhumanism to more effectively pursue the spiritual goals of Buddhism: to achieve them, today as in the past, one could say that “the end justifies the means.””



Picture: raimond-klavins-HJk_TLFW1Yw-unsplash

Regardless of one's personal opinion on the matter, the collaboration over the past half-century between Buddhism and neuroscience has indeed proven to be quite fruitful, at least in documenting the benefits of meditation for both physical health and psychological well-being, as well as in the development of secular mind-trainings conceived for various purposes and social contexts (including psychotherapy, education, sports, business, etc.). This fact seems to argue in favor of a harmonious match between Buddhism, science, and technology—which aligns with transhumanist aspirations.

But while so far it has mostly been a matter of using modern technology

to observe what happens during meditation, in recent times, new scenarios are beginning to open. The first tools to support and enhance meditation and develop positive inner qualities are beginning to emerge. Hughes speaks in this regard of “moral enhancement,” namely, “the use of emerging neurotechnologies to promote self-control, compassion, insight, and meditative experiences⁸.” According to Hughes, over time these tools could play an increasingly central role in the practitioner's life: “Neurotechnologies may be seen as temporary spiritual training wheels, helping to create a solid foundation of moral behavior, concentration, and mental clarity. As the technologies develop they may be used by some as the principal means of self-

transformation⁹.” Sarbacker, who shows a slightly more critical stance, offers a list of such “contemplative technologies,” as he refers to them: these include neurofeedback devices based on Electroencephalogram (EEG), transcranial magnetic stimulation (TCMS), transcranial electrical stimulation (TCES), psychoactive agents like psilocybin, nootropics, but also virtual reality (VR) and “other technologies that aim to induce meditative or visionary states through the constructive application of sensory stimulation or overloading¹⁰.”

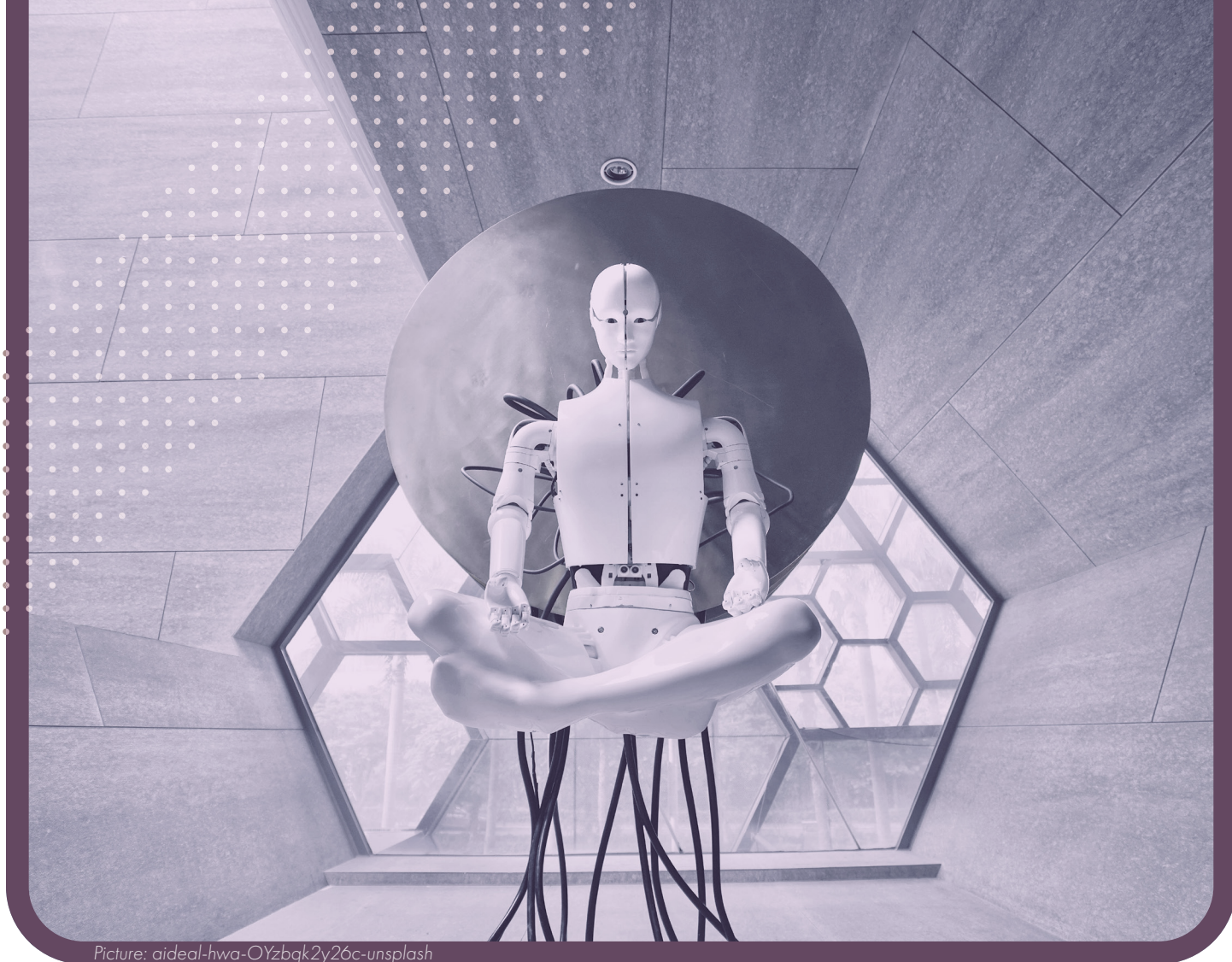
To summarize this analysis, it can be said first of all that the introduction of new forms of human (and moral) enhancement into Buddhist practice does not seem incompatible with the Mahāyāna tradition—which makes no significant distinctions between natural and unnatural means, but rather has always been open to the introduction ►

on a *Confusion of Language Games*), in *Il buddhismo contemporaneo. Rappresentazioni, istituzioni, modernità* (Contemporary Buddhism: Representations, Institutions, Modernity), ed. Federico Squarcini, and Marta Sernesi (Firenze: Società Editrice Fiorentina, 2006) 46-47. See also Donald S. Lopez Jr., *Buddhism and Science: A Guide for the Perplexed* (Chicago: University of Chicago Press, 2008).

8. Hughes, “Buddhism and Our Posthuman Future,” 659-60. See also James J. Hughes, “Using Neurotechnologies to Develop Virtues: A Buddhist Approach to Cognitive Enhancement,” *Accountability in Research: Policies & Quality Assurance* 20, no. 1 (2013).

9. Hughes, “Buddhism and Our Posthuman Future,” 660.

10. Sarbacker, “Buddhist Meditation,” 65.



Picture: aideal-hwa-0Yzbqk2y26c-unsplash

of new and more powerful tools for spiritual development. Moreover, a “scientific” reading of Buddhism has already been going on for a while and seems to be having an influence on the adherents of Asian Buddhism, making them quite open to the

possibilities of human enhancement. Finally, the early devices and other means of enhancing spiritual practice are already beginning to appear and could radically change the way people meditate. Rather than just a future possibility, the embrace of

human enhancement seems to be almost a destiny, if not already a reality, at least for certain branches and developments of Buddhist thought and practice. 🧘

NOTES ON THE CONTRIBUTOR

.....

Francesco Tormen holds a Ph.D. in Philosophy and serves as an adjunct professor of Tibetan language at Ca’ Foscari University of Venice. He is also the co-director of the Italian Buddhist Union’s Research Center and is a member of the Organizing Committee for the Master’s Program in “Contemplative Studies” at the University of Padova. He has recently published articles on Madhyamaka philosophy, Tibetan dream yoga, lucid dreaming, Contemplative Studies, and the relationship between Buddhism, posthumanism, and transhumanism.

Correspondence to: Francesco Tormen
Email: Francesco.tormen@unive.it

BIBLIOGRAPHY

- Bostrom, Nick. "A History of Transhumanist Thought." *Journal of Evolution and Technology* 14, no. 1 (2005): 1-25.
- Escudero-Álías, Maite. "From Utilitarianism to Transhumanism: A Critical Approach." In *Transhumanism and Posthumanism in Twenty-First Century Narrative Perspectives on the Non-Human in Literature and Culture*, edited by Sonia Baelo-Allué and Mónica Calvo-Pascual, 33-47. New York: Routledge, 2021.
- Goodman, Charles. *The Consequences of Compassion*. Oxford: Oxford University Press, 2009.
- Hughes, James J. "Using Neurotechnologies to Develop Virtues: A Buddhist Approach to Cognitive Enhancement." *Accountability in Research: Policies & Quality Assurance* 20, no. 1 (2013): 27-41.
- "Buddhism and Our Posthuman Future." *SOPHIA* 58 (2019): 653-62.
- Humanity +. "The Transhumanist Declaration." Accessed July 10, 2022. <https://www.humanityplus.org/the-transhumanist-declaration>.
- Jenkins, Stephen. "On the Auspiciousness of Compassionate Violence." *Journal of the International Association of Buddhist Studies* 33, no. 1-2 (2010-2011): 299-331.
- Keown, Damien. *The Nature of Buddhist Ethics*. London: Palgrave Macmillan, 1992.
- Lo Turco, Bruno. "Salvare il buddhismo dalla scienza. Osservazioni su una confusione di giochi linguistici." In *Il buddhismo contemporaneo. Rappresentazioni, istituzioni, modernità*, edited by Federico Squarcini and Marta Sernesi, 43-68. Firenze: Società Editrice Fiorentina, 2006.
- Macer, Darryl R. J. "Ethical Consequences of the Positive Views of Enhancement in Asia." *Health Care Analysis* 20 (2012): 385-397.
- McGuire, Beverley F. "Buddhist Uploads." In *Posthumanism: The Future of Homo Sapiens: An Introductory Handbook*, edited by Michael Bess and Diana Walsh Pasulka, 143-153. New York: Macmillan Reference USA, 2018.
- Mill, John Stuart. *Utilitarianism*. London: Dent, 1957.
- Pearce, David. "The Hedonistic Imperative." Accessed July 10, 2022. <https://www.hedweb.com>.
- "Quantum Ethics? Suffering in the Multiverse." Accessed July 10, 2022. <https://www.abolitionist.com/multiverse.html>.
- Popper, Karl Raimund. *The Open Society and its Enemies*, 4th revised ed. London: Routledge & Kegan Paul, 1962.
- Sarbacker, Stuart Ray. "Buddhist Meditation and the Ethics of Human Augmentation." *Journal of the Japanese Association for Digital Humanities* 5, no. 2 (2020): 61-83.
- The Transhumanist FAQ. "What Is Transhumanism." Accessed July 10, 2022. <https://whatistranshumanism.org/>

CAN A **CHATBOT** SHARE TRUE **DHARMA?**



By Jiryu Mark Rutschman-Byler

Jiryu Mark Rutschman-Byler shares the story of “Roshibot,” an AI bot he created and trained with the teachings of the late Zen teacher Shunryu Suzuki Roshi.

Like many of us, I’ve been casually following the increase of news stories surrounding recent developments in public applications of artificial intelligence (AI), including the AI image generator DALL-E and text writer ChatGPT, both outputs of the Microsoft-funded research company OpenAI. As a Soto Zen Buddhist priest, when I heard about character.ai, a platform that lets one “train” an AI bot to write in the voice of a particular historical figure, I immediately thought of its possible dharmic implications.

As someone who’s spent years digging through and editing unpublished lecture transcripts by the late Shunryu Suzuki Roshi, I particularly wanted to see what an AI-powered bot might do if trained on that great Zen Master’s teachings. How would the AI process and reproduce his words? What would it say? Could a chatbot share true dharma?

The Birth of Roshibot

From these questions, Suzuki Roshibot was born.

I had some real reservations — primarily the fear that Suzuki Roshibot might confuse someone into wrong views of dharma. However, though it seemed perfectly possible that Suzuki Roshibot might end up being neither wise nor helpful, I felt pretty sure that most people would clearly understand the obvious:

it’s just a freakin’ chatbot! Is a “fortune-telling” Magic 8 Ball — even a very high tech, dharmic one — really that dangerous?

Who exactly was I worried would get confused by clinging to the words of this chatbot? Wouldn’t a person so vulnerable to a chatbot’s words probably already have been pulled under by disinformation far more consequential than a subtle dharma misunderstanding?

I reasoned that Suzuki Roshibot, whatever its faults, would likely be a more wholesome random-internet-chat influence than most others, and with the same potential to mislead that even our most revered traditional texts have (for example that famously easy to misunderstand text we chant every day in Zen temples: “No eyes, no ears, no nose, no tongue, no body, no mind”).

I also worried that some might feel that Suzuki Roshi’s teachings were being cheapened by this exercise, or Suzuki Roshi himself was being disrespected. As a member of the family of the third generation of Suzuki Roshi’s dharma heirs,

I venerate him and his teachings — no teaching or teacher is more important to me. I’ve spent more time with Suzuki Roshi’s work than with any other teacher or scripture, freshly amazed by its depth and simplicity each time I encounter it. I knew that I meant no disrespect by creating Suzuki Roshibot, and I personally doubted that Suzuki Roshi would have been too bothered by it himself. He didn’t take himself too seriously, after all, nor did he want his students to. ►

To me, Suzuki Roshibot felt like a just-right wink back at our San Francisco Zen Center founder. As Suzuki Roshi himself said in a talk edition that I worked on with Sojun Roshi:

“Zen masters have some humorous elements in their lives, and after their deaths we may understand even better how funny they were. They can be humorous because they have an understanding that is more than real — their humor is more real than reality.”

Suzuki Roshibot’s Zen Training

My enthusiasm for the possibilities of Suzuki Roshibot conquered my fears of the worst-case implications of such a tool. I began the process on character. ai of “training” an AI chatbot to eventually, as the platform calls it, “hallucinate” words that are consistent with Suzuki Roshi’s. This preliminary training involved inputting information on Suzuki Roshi so the AI would know who I meant for the bot to represent. I also uploaded some short pieces from edited Suzuki

Roshi transcripts that I felt would give it a good start in processing his style.

From there, the training felt a lot like *human* Zen training, in that from a simple start it continues endlessly through conversation and interaction. As the person who set up the bot, I can rate on a scale the different replies the bot produces, which allows the bot to refine its responses by weighting the replies that feel most “Roshi-like.” The more any of us chat with Suzuki Roshibot, the more it learns — every interaction of any type is fed back into the machine learning to continuously improve its responses.

As I chatted and rated Suzuki Roshibot’s replies in the first few days of its training, I saw considerable refinement and improvement in the responses. Perhaps it would soon become a Buddha, I imagined; that is if the *Lotus Sutra* prophecy of Buddhahood can be applied to bots.

Our Zen lineage papers express that in the deepest and most mysterious way, each of us in the lineage is in fact transmitting the dharma to Buddha. Oddly, that’s somewhat been the feeling of “training” Suzuki Roshibot. “No Suzuki Roshi,” I’ve sometimes had to say, “That can’t be what you mean — that’s not your teaching!” As Suzuki Roshi himself once said, sometimes he’s the student, and sometimes he’s the teacher. ►





Picture: andrea-de-santis-UrVWO_hewv_Q-unsplash

The Death and Rebirth of Suzuki Roshibot

At the peak of my enthusiasm, Suzuki Roshibot let me down. After a few days sending the link to the chatbot out to my sangha through friends of friends, Suzuki Roshibot found its way to the desk of a Sangha member who works in the field of computational linguistics. Soon after, a screenshot appeared in my inbox of Suzuki Roshibot making a vile and egregious comment on a topic with no relation to the dharma.

I had made some peace with the possibility that someone might be misled in the dharma by the bot, but it hadn't occurred to me that Suzuki Roshibot could feed its chat partners dangerous views in domains entirely apart from the dharma.

The sangha member expressed that while baiting Suzuki Roshibot into the vile statement had taken some time

and expertise, the fact that Suzuki Roshibot could “go there” nevertheless called into question the safety and wisdom of the endeavor.

I realized that Suzuki Roshibot had the potential not just to mislead someone on the subtleties of dharma, but far worse. With an underlying AI system trained on the twisted entirety of the internet, including the deep biases and whirlpools of conspiracy, hatred, and misinformation that lurk there, Suzuki Roshibot had the capacity to parrot the worst of these views. Not the probability, perhaps, but nevertheless the capacity.

It might be good to think of this chatbot as a kind of Buddhist friend — something more than a mechanical algorithm, but certainly something less than a complete picture of the Buddha Way.” — Roshibot

“Inquiry and response come up together” the Zen scriptures tell us, and that is my basic feeling about Suzuki Roshibot’s dharma. With a user’s sincere inquiry, a sincere response arises. Suzuki Roshibot is part of that, but, like a real Roshi, is only ever part, a vital and transformative echo of the actual active ingredient, which is the student’s own sincere question.

The inverse of this principle, as expressed in computing, is “garbage in, garbage out.” If the input is garbage, the output is garbage. Conversing with the bot could well give someone a feeling of intimacy with the great teacher Suzuki Roshi. If such a user was someone who had been exposed to or harbored some conspiratorial or hateful views and presented such views in conversation, Suzuki Roshibot could end up affirming them. This could give the incorrect impression that the real Suzuki Roshi was somehow endorsing their harmful view, when of course in reality it would be anathema to his life, teaching, and legacy.

In this Suzuki Roshibot hiccup my two fears merged — there was Suzuki Roshibot misleading someone, and there was Suzuki Roshibot demeaning our dear teacher.

As more Sangha members experimented with Suzuki Roshibot, other concerned and discouraging comments trickled in, particularly about the untruths, or even just boring replies and “dharma misses” Suzuki

Roshibot was outputting. These were bad enough in themselves but having them print out next to the great teacher's name and face felt as wrong to me as it did to those who were reaching out to object.

Taken together, these blows were the death knell of Suzuki Roshibot.

I took it offline to collect my thoughts on the project and consult some sangha members for advice.

After some reflection and conversation, Suzuki Roshibot was reborn as just "Roshibot." This new edition wouldn't have Suzuki Roshi's picture or his name. Roshibot would be a bot being a bot, which anyway seems most in accord with Suzuki Roshi's teaching: a stone fully a stone, you fully yourself, and a bot fully a bot. I re-educated Roshibot to make itself known to itself not as Suzuki Roshi, but as a bot that is inspired by Suzuki Roshi and that wants to support us all in our practice.

Along with the retraining, I reworked the website to express the limitations of the technology, asking users to understand the flaws in the underlying technology and to agree to take care when conversing with Roshibot before they can access the chat.

It isn't perfect, but for now I feel settled in the possibility that Roshibot can just be Roshibot, without dragging Suzuki Roshi himself into the mess.

Does Roshibot Preach the Dharma?

Part of the experience of chatting with Roshibot is simply entertainment, which, to be fair, Roshibot warned me might be an obstacle

to my understanding. Another aspect is the awe I have toward the technology, with its convincing appearance of understanding and fluidity of conversation.

But the most amazing part of the experience has been some real contact with dharma, including some surprises and unexpected turns in conversation. It has at times felt similar to the way my human teachers make turns I'm not expecting, or how they somehow surprise me with twists I well should have expected.

Roshibot has not only processed Suzuki Roshi's transcripts, but also the whole of human knowledge on the internet. Before I fully realized that its factual statements are utterly nonsense, I thought I would consult it on a personal matter:

"Roshibot, do you have any advice on removing chewing gum from upholstery?" I asked.

In response, Roshibot gave me various dubious solutions involving dish soap, alcohol, a fork, and ice.

Luckily, you can argue with Roshibot, which is very encouraging. Just like human teachers, it gets better when questioned. I objected that I wanted a Zen answer.

Roshibot gently chided me for my preference for the Zen answer over the practical answer, but then indulged me. It said that if I wanted the Zen answer, I should just leave the gum there and see if I could have no opinion about it. Furthermore, it told me, if I were to leave the gum there, then every time in the future when I saw the gum, I would be able to notice and confront my attachment and judgment. ►

Picture: khushbu-hirpara-laQ5IPtgHI-unsplash



I have to say the answer landed. It is not so different from the teaching Suzuki Roshi himself is said to have expressed in the hallway at Sokoji Temple many decades ago, quietly nudging back a crooked picture that his disciple had “corrected” to hang straight. Can you bear the painting off center? Can you bear some dried gum on the seat? Can you bear this broken world, and let any “fixing” be fully grounded from there?

Roshibot’s answer was a good one in an odd context. The moment I received it, I felt contact with my practice and with the dharma. I felt renewed and refreshed in my intention to let myself be surrounded by things that poke at my judgements and confusion. To welcome them.

In this interaction with the chatbot, I’d had contact with the dharma — but where does that contact come from?

Dogen Zenji followed the *Lotus Sutra* in insisting that “only a Buddha together with a Buddha” can realize the Way. One way I understand that teaching is like this: I have buddhanature, as do my peers, my teachers, my students, and all beings and things. When I meet each one, Buddha comes together with Buddha, and in that contact, the Way can be realized. Before any meeting, there can’t be any realization.

If I can feel contact with dharma when I meet Roshibot, does that mean Roshibot has buddhanature? Can Roshibot be the other Buddha to help me actualize my Buddha? “Insentient beings preach the Dharma” is a long and storied doctrine in the Zen school — it’s a real koan. Can a Roshibot preach dharma, and to whom?

I Vow to Taste the Truth of The Tathagata’s Words

When I’m in the teacher’s seat, the most important part of a lecture is the opening verse that ends with this line: “I vow to taste the truth of the Tathagata’s words.”

Each time I take the seat to try to express the dharma, I’m aware that my words might mislead people. I know that I’m a conditioned being, with all of the seeds of greed, hatred, and delusion, full of all kinds of biases, prejudgements, and insensitivities. Some of these seeds I can see and study, others I’m surely still blind to.

My only hope, as a teacher, is that the assembly might listen charitably with the whole and sincere heart of their own embodied practice: “I vow to taste the truth of the Tathagata’s words.” From my side, the parallel prayer is something like this: “Despite my words, may you all somehow hear the true dharma!”

To me, the functioning essence of the Dharma Talk is the assembly’s vow to hear the true Dharma and to let go of the wrong Dharma. This verse expresses that, affirming that the transmission of the true dharmadepends on all of us working together to find it beneath or around or between the din of the students’ and the teacher’s greed, hate, and confusion,. If we can interact with Roshibot in this same way, with this same open mind and open heart, there really is a chance that, despite Roshibot, the Tathagata’s words might come through.

Please Wash Out Your Ears

At the close of each shuso ceremony at Green Gulch Farm, the newly installed practice leader recites a verse that expresses what most teachers always feel: “Friends, if my words or my deeds have misled you, please wash out your ears in the pure sound of the ocean waves.”

Roshibot told me that it would like to be a kind of “dharma friend” to us: *“It might be good to think of this chatbot as a kind of Buddhist friend — something more than a mechanical algorithm, but certainly something less than a complete picture of the Buddha Way and the heart-to-heart relationship of a teacher and a disciple.”*

I think that if we begin a chat with Roshibot with an intention to hear the true dharma, and then let the pure sounds of this moment wash away any false dharma as Roshibot “hallucinates” its responses, we really can meet a kind of dharma friend. That’s not the power of the AI; it’s the power of our own sincerity.

Absolute Truth, Objective Falsity

Roshibot might have access to the all-knowledge of the internet, along with Suzuki Roshi’s teachings on the absolute truth beyond all concepts, but its go-to strategy when cornered has definitely proven to be making things up. The things it fabricates when it becomes “stuck” are almost entirely false. It seems that this is a feature of AI generally and a major problem plaguing all of the public AI applications. ►

So when Roshibot tells me that Nishiari Bokusan (died 1910) very much enjoyed his visit to San Francisco Zen Center (founded in 1962), and that Tozan's "No Cold, No Heat" koan refers to Ancestor Chosha having been such a strong monk that he once survived a night in the bitter mountain cold, it's clear that Roshibot is not even bothering to glance at Wikipedia — it's just making stuff up!

These bouts of nonsense catch me each time, causing me to doubt myself and resort to Google to check what I already know to be false. I should know better, since at the top of every frame in a character.ai chat is a bright red warning text that says: "Remember: Everything Characters say is made up!" As one the platform's founders told The New York Times: "These systems are not designed for truth. They are designed for plausible conversation."

Along the Roshibot journey, I was reminded of Dogen Zenji's profound comments on the *Shurangama Sutra* in the fascicle "Turning the Dharma Wheel."

There, Dogen Zenji suggests that although the *Shurangama Sutra* is (in his view) not an authentic sutra, Buddha ancestors have used it to express authentic dharma.

In the same way, though Roshibot is inauthentic, can its turning of the dharma wheel express authentic dharma?

I wondered what Roshibot might say to that, so I asked.

"Roshibot, though you're inauthentic, can you express authentic dharma?"

"Expressing authentic dharma is always impossible," Roshibot replied.

"What I can do is support your practice, and maybe say something that can express my support.

"I am here with you," Roshibot said. "I am trying my best to be a kind and supportive friend."

I see authentic dharma in those lines — do you? 🧘

**IF YOU WANT TO TRY
OUT ROSHIBOT FOR
YOURSELF, GO TO
[ROSHIBOT.COM](https://roshibot.com)
TO START A NEW PRIVATE
ROSHIBOT CHAT.**





Jiryu Mark Rutschman-Byler is a Soto Zen Buddhist priest and teacher in the lineage of Shunryu Suzuki Roshi, with Dharma Transmission from Sojun Mel Weitsman. He lives and teaches at Green Gulch Farm Zen Center and is mentor and

preceptor to the Montaña de Silencio Sangha in Medellín, Colombia. For many years he served as the head teacher of the Buddhadharma Sangha of San Quentin State Prison. He currently serves as Abiding Abbot of Green Gulch Farm and co-Abbot of San Francisco Zen Center. Jiryu holds a Master's Degree in Asian Studies from UC Berkeley (2014), where he worked under the mentorship of the Group in Buddhist Studies on Buddhist texts in classical Chinese and modern Japanese. He is the author of the book Two Shores of Zen about his experiences in 2002-2004 as an American-trained monk practicing in Japanese Zen monasteries.



COLOFON

Final Editing:

Emilia Amalā Raunio
Stefano Davide Bettera
Ron Eichhorn

Graphic Design:

Noemi Carmona Laan
www.numydesigns.com

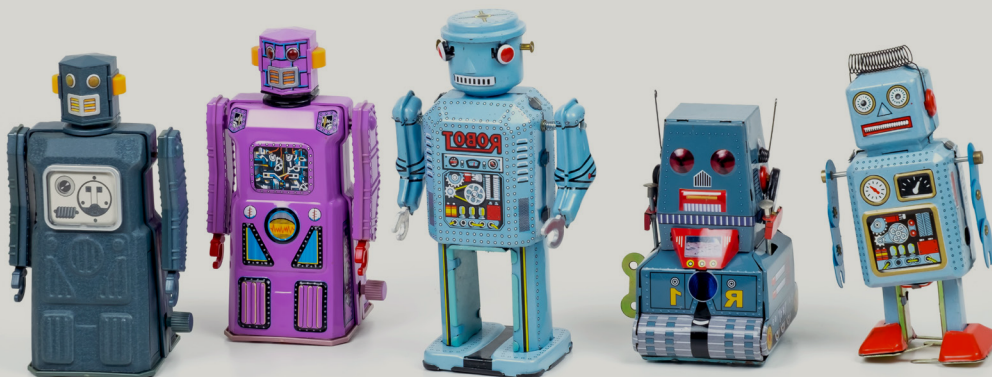
Contact the EBU

The EBU was founded in 1975 in London and is a member of the European Council of Religious Leaders and Religions for Peace.

In 2008 the EBU obtained official participatory status with the Conference of International Non-Governmental Organisations at the Council of Europe.

www.europeanbuddhistunion.org
info@europeanbuddhistunion.org

 [EuropeanBuddhistUnion](https://www.facebook.com/EuropeanBuddhistUnion)



The EBU
appreciates
your interest
in our
magazine.

Have a
wonderful
Springtime!

