The Paradox of Autopoietic Artificial Intelligent Systems in Education

Philip Baron
University of Johannesburg, South Africa • pbaron/at/uj.ac.za

> Abstract • An autopoietic educational system may provide for immersive individualised learning experiences that cater to the unique interests of each student; however, universities generally require uniformity in education to enable group assessments and accreditation and thus do not readily accommodate individualised tuition. A further challenge is that for a system to be autopoietic, it should be able to create its own rules; this may be at odds with the educator's goals for their courses.

Autopoietic machines

« 1 » A lot has been written about the benefits of social learning (Bandura & Walters 1977). With improvements in digital pedagogy and the highly immersive and cooperative nature of certain forms of learner and machine interactions (§§35, 40f, 44), the scope of the word "social" may need to be redefined. While it is not a rule that being social requires verbal conversation, language and conversation are our primary modes of communication and are thus the main features in any educational system as a determinant of its success (Baron 2016; Baron & Herr 2019). The term "social" originally implied interacting with peers, family, or even strangers; however, with conversational models being built into artificial intelligent (AI) systems (such as Google assistant, Apple's Siri, and Amazon's Alexa), it seems that "social" might include chatbots too. In such cases, humans could consider such an interaction as social when the human deems it to be equivalent to a human interaction. This is in keeping with the principles of the Turing Test, albeit in the context of socialising. Thus, for machines to be social requires some meaningful interaction that relies on the AI's variety in its responses. It therefore makes sense that Claudio Aguayo proposes that his conception of autopoietic technology-enhanced learning (TEL) systems should be meaningful and nurturing (§§9, 13, 24, 26), and that for this to take place, a form of structural coupling would arise as the human and the AI get to "know" each other. This is indeed the aim put forward by Aguayo with respect to his autopoietic TEL system, which he proposes should be humanlike, providing meaningful social and sociocultural experiences for its users (§§8, 11, 31).

« 2 » Aguayo's proposal for meaningful social interactions has merit, as memorable interactions result in deeper learning – an aim of TEL systems. To achieve this goal of meaningful interactions in the context of learning, these TEL systems would require a high level of autonomous personalisation within the TEL system to cater for the many unique interactions that may arise between the system and the students. Proposing autopoiesis as a label for this goal is useful, since an autopoietic system creates its own rules, which could, in principle, cater for a wide variety of interactions. "Meaningful," to me, would be different from what it would be to the next person, owing to each person's unique background and interests. Hence, for all users/students to experience the interaction as meaningful, and by extension social, the autopoietic TEL would need to offer interactions that suit different personalities and their preferences as they drift along their unique paths.

« 3 » It seems that coding for such a wide scope of conversational interactions would be beyond the feasible scope of human labour and since Aguayo did not explain the technical parts of how such a software system would work, one may assume that it would comprise of natural-language generation and machine and deep learning. Autopoietic TEL would thus inevitably leave to computers the task of generating the growing content to maintain the structural coupling with the students. If this is the case, then the proposed autopoietic TEL system would be likely to utilise pre-trained large language models (LLMs) as are used in chatbots. A self-producing and decentralised software-based system that can maintain itself, set its own trajectory, adapt to users, set its own rules, and maintain its purpose of "promoting educational and learning outcomes" (§9, 17, 27) *seems* like a great idea. Aguayo (§11) even asks why would we not want that for a TEL system in education? I propose a few reasons why we should critically evaluate this goal.

Autopoiesis and useful knowledge

« 4 » Would an autopoietic system reliably provide the educators and students with the curriculum outcomes that are required to be part of the course, or would too much variability be introduced owing to the open-ended nature of the system? I attempt to describe this problem as follows. Aguayo (§14) provides an example of a limitation of a traditional TEL system that was programmed to cover only the topic of an endangered mollusc species. If the mollusc species becomes extinct, Aguayo argues that so too would this type of application. However, let us say, for example, that the TEL system was designed to be autopoietic as Aguayo has proposed and can now adapt and migrate to more relevant topics. For a topic to be meaningful for one student, the interaction would need to be recursive and mutually influential and thus constantly evolving in the purposeless[2] drift that ensues when there is structural coupling (Becvar & Becvar 2006). However, one should not assume that a target audience is homogenous just because all the students are enrolled in the same degree. The variety could be infinite, and since autopoiesis and structural coupling rely on highly unique attributes offered by each entity within the interaction, or as Ranulph Glanville (2008) termed it "in the between," I would assume that for the system to be meaningful and socially and culturally sensitive (§8), the system would need to provide unique social learning experiences for each student. This creates a scope problem; how does one pre-determine that the new topics provided by the autopoietic TEL system will be useful to all the students *and* the educator? Since the topics evolve out of the meaningful interactions between students and the autopoietic TEL, there would be uniqueness for each one, but are all these unique topics meaningful to the educators who are trying

2| Humberto Maturana (1988) described that a living system may appear as though it operates according to a set of goals, but this is based on the observer's conceptualisation as "[a]utopoietic machines are purposeless systems" (Maturana & Varela 1980: 85).

to get their students through their courses while still maintaining accreditation? With each student having drifted in different directions, I assume it to be a complex and a time-consuming task to examine a group of marine biology students who were all taught different things. And, this assumes that the self-learning TEL system maintained that same field of biology (molluscs), as some students may have shown a disregard for the mollusc topics and rather used the system to teach themselves about computer hacking instead.

« 5 » If the software designers attempted to restrict the TEL system to function within only a few set topics (negating Aguayo's goals), even within a restricted mode, it seems highly unlikely that this would solve the out-of-scope drift that I am predicting for an autopoietic system. For example, if the programmers only allowed the topics of molluscs, there are still many subtopics for study. One student might be learning about mollusc habitat while another might be more interested in learning about reproduction. This means that the curriculum is growing, which would surely make the educator's task far more complex than if the educator just used a traditional TEL that is programmed to only offer what the educator set as the content. Thus, my concern is how will students be evaluated to confirm they have all met the same learning outcomes when the autopoietic TEL can set its own rules and constantly evolve its content to adapt to the students? Q1

Reflexion and discernment

« 6 » I proposed contextual conversational teaching and learning in engineering education by using the students' own lived experience as informers to the curriculum by metaphorically depicting this as a nunatak[3] that provides the reference points for the curriculum (Baron 2018a, 2018b). This approach has been useful in addressing social justice in the classroom. Students perturb the system with their ideas and if the system cannot accommodate these perturbations – that is, the curriculum cannot adapt – the structural coupling is lost. Change is thus the process of structural transformation (Varela & Johnson 1976). If the students' ideas are useful and technically sound, then at times I must adapt and change my opinion about parts of the course – adapting the curriculum. This means that along the trajectory of the course I reflect on my own knowledge as my behaviour influences the structural coupling. By reflecting in community with the students (reflexion), I aim to maintain structural coupling with my students, which means adapting the curriculum *as a group* as we learn in community with one another.

« 7 » Aguayo (§13) does propose that his autopoietic TEL system can adapt to the learning affordances of the changing needs of the target audience while maintaining its purpose. This appears to be quite similar to the process I described above. It implies that the autopoietic TEL can reflect and act on these reflections to maintain its educational purpose. However, if an AI chatbot uses its already learned data to formulate responses, it is not easily convinced by end users to change its future responses (behaviour), as the end user's knowledge does not usually fall within the scope of the AI's learning model for its decision making. Put simply, it seems the AI chatbot models are not easily convinced by end users. I thus wonder if an autopoietic TEL system would be able to incorporate the students' perturbations in knowledge creation and differentiate useful from useless perturbations while still maintaining its purpose. Thus, how would the autopoietic TEL algorithms work to reflect and categorise the student suggestions as new knowledge, since the system is supposed to be aimed at addressing socio-cultural aspects too? Q2

Artificial hallucination

« 8 » It seems idealistic to assume that an uncontrolled and unconstrained TEL system, which essentially is a computer running sophisticated software, will do what is required of it, i.e., cover the required topics correctly while maintaining structural coupling with all the students. For example, Microsoft's Bing AI search and Open AI's ChatGPT have on many occasions provided misleading answers while confidently displaying them as "truth." Since these programs scour the internet and compute answers based on a statistical probability, some answers can be completely off the mark, as the answers are only as good as the sources used – the internet has a lot of questionable content. Thus, when the system is adapting and introducing new content to the student, there is a significant chance that the new content will have errors imbedded within the content. For example, when Bing's AI was asked to critique a certain model of vacuum cleaner, Bing cited that the vacuum cleaner had the disadvantage of being noisy and having a short power cord of only 16 feet; however, this appliance in question has no power cord, as it is a portable unit, and the top Amazon review talks about how quiet the it is.[4] When Bing was asked to summarise the financial statements of a company (Gap Inc), it provided some accurate responses but also fabricated numbers and arrived at conclusions based on its own fictitious numbers.[5] These shortcomings are serious and have been termed artificial hallucination (Ji et al. 2022). Bing is not alone in this, as other chatbots have been caught out too. It seems that LLM that are meant to exhibit human-like conversation styles, also have human-like problems, that is, they can appear arrogant and ignorant in their answers.

« 9 » A further problem is that such systems have no way of knowing their answers were bad. Having used ChatGPT to take my students' honours-level test prior to my students' taking the test, it was interesting to see that some of the questions were answered well while other answers were misleading and useless. If I were to incorporate such tools in my digital pedagogy, I would need to thoroughly review all the responses from the autopoietic TEL before the content could be used as authoritative content for educational instruction. Thus, how might an autopoietic TEL system be checked for accuracy? And if a content review is indeed a requirement, does this then negate Aguayo's goals (§§3, 11, 27, 28) for these systems to be efficient and cost-effective alternatives to the traditional but rigid TEL systems? Q3

3| An isolated outcrop of rock protruding through an ice sheet. The protrusion is a metaphor for where the educator can "attach" a curriculum topic to.

4| https://dkb.blog/p/bing-ai-cant-be-trusted

5| https://www.zerohedge.com/technology/microsofts-bing-ai-chatbot-starts-threatening-people

Conclusion

« 10 » Cajoling a human to follow an instruction has its challenges, but the contract of work for pay facilitates the management of workers to follow the instructions/rules set by their employers. If autopoietic TEL creates its own rules and purpose, I am unsure it will do what the educators want it to do; rather it may drift too far off the original goals set by the educator and it is unclear how educators can influence such unconstrained systems. If the TEL's goals are different to the educators' goals, one wonders if we will even be able to communicate with such a system (Riegler 2008). Thus, an autopoietic TEL with its own purpose as proposed (§§13, 14, 20–24) seems to me as though it will be at odds with the educator rather than in support of the educator's goals.

« 11 » Since Aguayo proposed that such systems should be able to maintain themselves and adapt, this opens the door to an unpredictable system. Once the system is set "free," how does its purpose align to our purpose? With AI processing much more data than any human, we could not predict all of its answers even if we designed the system. The system would continuously surprise us and since it has more knowledge than humans, this implies that AI should direct us. However, if AI cannot reflect and discern, then it seems that the only way to manipulate the TEL is to repopulate its database with our bias (purpose). Since its database is the internet, and we cannot rewrite the internet, this is a moot point, and thus the unconstrained AI is out of our control. However, if we do manage to gain control of the system, then it is no longer autopoietic, which is a paradox.

« 12 » Autopoietic TEL educational systems, however, could promote highly immersive and meaningful interactions. This could be very fulfilling for the student's independent learning. However, universities generally do not cater well to individualised learning; they thrive on hierarchy that focuses on conformity (Richards 2018). Students are expected to learn the same topics for which professional bodies are required to audit the learning outcomes, and these outcomes tend to be narrowly defined (as Heinz von Foerster so succinctly illustrated in Foerster 1972: 41, pointing out that "[a] perfect score in a test is indicative of perfect trivialization"). If there is too much variety, it becomes difficult for the educators to manage compliance.

« 13 » Even the best teachers are disliked by some students, and I predict that the same will happen for autopoietic TEL. Some students will just say that they did not find their interaction meaningful, because learning and socialising are personal choices. Since there is a choice, there is a lot of structural adaptation that takes place on the student's side that is beyond any autopoietic TEL's reach.

References

Bandura A. & Walters R. H. (1977) Social learning theory. Volume 1. Prentice Hall, Englewood Cliffs.

Baron P. (2016) A cybernetic approach to contextual teaching and learning. Constructivist Foundations 12(1): 91–100. ▸ https://constructivist.info/12/1/091

Baron P. (2018a) Ethical inclusive curricula design: Conversational teaching and learning. South African Journal of Higher Education 32(6): 326–350.

Baron P. (2018b) Heterarchical reflexive conversational teaching and learning as a vehicle for ethical engineering curriculum design. Constructivist Foundations 13(3) ▸ https://constructivist.info/13/3/309

Baron P. & Herr C. M. (2019) Cybernetically informed pedagogy in two tertiary educational contexts: China and South Africa. Kybernetes 48(4): 727–739. ▸ https://cepa.info/8297

Becvar D. S. & Becvar R. J. (2006) Family therapy: A systemic integration. Sixth edition. Pearson Education, Boston MA.

Foerster H. von (1972) Perception of the future and the future of perception. Instructional Science 1(1): 31–43. ▸ https://cepa.info/1647

Glanville R. (2008) A cybernetic musing: Five friends. Cybernetics and Human Knowing 15(3–4): 163–172.

Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y., Madotto A. & Fung P., (2022) Survey of hallucination in natural language generation. ACM Computing Surveys 55(12): 248.

Maturana H. R. (1988) Ontology of observing: The biological foundations of self-consciousness and the physical domain of existence. In: Donaldson R. E. (ed.) Texts in cybernetic theory: An in-depth exploration of the thought of Humberto Maturana, William T. Powers, and Ernst von Glasersfeld. American Society for Cybernetics (ASC). ▸ https://cepa.info/597

Maturana H. R. & Varela F. J. (1980) Autopoiesis: The organization of the living. In: Maturana H. R. & Varela F. J. (eds.) Autopoiesis and cognition: The realization of the living. Reidel, Dordrecht: 73–140. ▸ https://cepa.info/7624

Richards L. D. (2018) Changing the educational system: The bigger picture. Constructivist Foundations 13(3) ▸ https://constructivist.info/13/3/331

Riegler A. (2008) The paradox of autonomy: The interaction between humans and autonomous cognitive artifacts. In: Dodig-Crnkovic G. & Stuart S. (eds.) Computing, philosophy, and cognitive science. The nexus and the liminal. Cambridge Scholars Publishing, Cambridge: 292–301. ▸ https://cepa.info/292

Varela F. J. & Johnson D. (1976) On observing natural systems. CoEvolution Quarterly 3(2): 26–31. ▸ https://cepa.info/4370

Philip Baron has been published across several academic disciplines and has an active social-media presence with over 1000 educational videos published on YouTube, https://www.youtube.com/@ecologicaltime. Philip has postgraduate degrees in psychology, religious studies, and electrical engineering.

Funding: No external funding was received while writing this manuscript.
Competing interests: The author declares that he has no competing interests.

Received: 26 March 2023
Revised: 4 April 2023
Accepted: 5 April 2023