About heteronomy induced by machine learning

Professor Guy André Boy

Heteronomy/autonomy of humans and/or machines

In 1978, Sheridan and Verplank established 10 levels of automation from full autonomy of people (Level 1, people are in full control of what they do) to full autonomy of machines (Level 10, machines are programmed to execute tasks without any human control). More recently, ISO 22989 standard (Table 1), which tries to distinguish between autonomy, heteronomy et automation, claims that AI systems can be compared based on their degree of automation and whether or not they are subject to external control. Machine autonomy is at one end of the spectrum and a fully human-controlled system at the other, with degrees of heteronomy in between (which is exactly what Sheridan and Verplank proposed in 1978, except ISO 22989 proposes now 7 levels of automation instead of 10).

Table 1, Relationship between autonomy, heteronomy, and automation (ISO 22989).

		Level of automation	Comments
Automated system	Autonomous	Autonomy	The system is capable of modifying its operating domain or its goals without external intervention, control or oversight.
	Heteronomous	Full automation	The system is capable of performing its entire mission without external intervention
		High automation	The system performs parts of its mission without external intervention
		Conditional automation	Sustained and specific performance by a system, with an external agent being ready to take over when necessary
		Partial automation	Some sub-functions of the system are fully automated while the system remains under the control of an external agent
		Assistance	The system assists an operator
		No automation	The operator fully controls the system

Sheridan and Verplank, as well as ISO 22989, see the relationship between autonomy, heteronomy, and automation from a technology-centered perspective (i.e., ISO 22989 talks about autonomy and heteronomy of machines). Symmetrically, a human-centered perspective would advocate for **people's autonomy and heteronomy supported by machines**. This is the topic of this short note.

Human Machine Teaming (HMT) is a current hot topic in life-critical sociotechnical systems where machines have increasingly stronger machine learning (ML) capacities, which induce emergent human factors issues that require more attention, including resulting operational performance, trust and collaboration. How do people collaborate with ML-driven machines? Can they trust them? To what degree? On the machine side, is ML mature enough to be trusted? How do we assess ML maturity? What should people know about ML algorithms implemented inside these machines, in terms of their properties and behaviors? How deep should be this knowledge?

These questions are certainly relevant at this point in time when our sociotechnical environments are becoming increasingly digital, even virtual, and where autonomy has become a major issue. What do we mean by **autonomy** from both perspectives of engineering and operations? What is the difference between autonomy and automation? For example, we are developing increasingly autonomous cars toward an extreme vehicle that would not have a steering wheel anymore (i.e., a drone transporting people on roads!) However, autonomy should be thought on the human side also, especially when things tend to go wrong, that is when people are suddenly facing unexpected events that require appropriate changes of conduct.

Life-critical systems operations are commonly supported by procedures following and automation monitoring. However, when the system goes outside its definition context, both **procedures and automation** do not work any longer, **problem solving** capabilities are required. From this standpoint, automation rigidifies corresponding sociotechnical activities, as autonomy (involving problem solving) should be based on flexible support to finding appropriate solutions to known-but-unexpected as well as unknown problems.

There are two main philosophical routes at this point:

(1) the **technology-centered engineering** route that starts by developing means and dictates adaptation of people to machines to optimize performance with respect to some purposes, and therefore imposes new technology-driven rules upon people; and
(2) the **human-centered design** route that requires to think in terms of purpose before the development of technological means, to incrementally co-adapt technology and people, and to better understand and shape coordination rules required to keep trust and collaboration among increasingly autonomous human and machine agents.

These two approaches require appropriate definitions of autonomy, but even more of heteronomy.

Heteronomy is a property of an agent that consists in not having personal capacities and rules, but in obeying external laws. It is opposed to **autonomy** that, on the contrary, is a property of an agent that is capable of interacting with its environment according to its own capacities and rules of action. From full heteronomy to full autonomy, there are degrees to be defined fo each agent, whether a human or a machine.

Maturity and culture

There are relevant questions to be asked.

- Do we want a pilot to blindly obey the machine equipped with AI algorithms?
- Conversely, do we want the machine to be a slave of people in charge of a mission?
- What are the middle grounds?
- Can people and machine have leadership as well as followship capabilities depending on context? How this kind of dynamic function allocation would be accepted and mastered?
- More importantly and this a question of culture, where is the ultimate authority, on the human side or on the machine side?
- In what contexts does this previous question make sense?
- Are ML-driven machines new middle managers between high-level decision makers and human executants?

- What is the personality of an ML-driven machine?
- Could people understand and work with this kind of personality?
- What is the role of an ML-driven machine in an "advanced" sociotechnical society?
- What would happen if people reject orders from such machines or even do not understand what they are doing?
- Should external laws dictate that people must obey these emergent new rules (heteronomy) or keep their critical thinking and therefore authority to be in charge when their lives is in danger?

This is again a matter of maturity and culture that needs to be more investigated and mastered.

Maturity is currently measured in terms of technology readiness levels (TRLs). However, people and organizations need to be considered also, in terms of human readiness levels (HRLs) and organization readiness levels (ORLs) (Boy, 2021). HRLs should provide a scale for the maturity of practice, and ORLs should provide a scale for the maturity of the related organizations, both corresponding to the technology being developed and further used. These two types of maturity scales strongly depend on ethics and political models being chosen. More specifically, there are three types of generic **Systemic Interaction Models** (SIMs) to be considered (Boy, 2020):

(1) supervision (one agent dominates and the other [the others] follows [follow] its instructions);
(2) mediation (agents interact through a mediation space [via diplomats for example]); and
(3) cooperation which assumes that each agent is equipped with a model of the other [the others] (this model is constantly evolving through learning).

In addition, **authority** plays a fundamental role in the consistency and sustainability of the agency of these agents (i.e., the system of systems) being considered. The main problem within an organization is the allocation of authority (leadership functions) that can be **imposed** and/or **recognized**. The ideal case is when authority is recognized, no matter whether it is imposed or not. The worst case is when authority is imposed and not recognized. In order to study all possible cases, it is strongly suggested to cross SIMs and authority allocation, based on cultural factors.

As my colleague James Dator has already said, "culture is to people what water is to fish!" Consequently, it is sometimes difficult, and in some cases impossible to adapt to a brand-new culture. There is a continuum between full heteronomy and full autonomy depending on the culture where we come from and where we are. Problem is that algorithms are provided by some people who then are responsible for what machines do. When these algorithms provide deterministic, verifiable and explainable outcomes, then certification[1] makes sense and should be carried out. Otherwise, the very concept of certification should be replaced by another concept close to subjective qualification and judgement. Human beings who depend on activities of these ML-driven machines have no choice but live with what these machines are doing or reject them. Should we accept this approach when people's life or death are at stake? Are there contexts where this question can be positively considered? Answers to these

[1] Certification is taken in the verification and validation (V&V) sense.

questions should provide more information and tangibility on people's autonomy and heteronomy induced by ML-driven machines.

Conclusion and perspectives

We have seen the difficulty of mastering concepts such as autonomy, heteronomy, trust, collaboration, authority, responsibility, control and others at the same time. It is time to study specific representative cases and scenarios to elicit knowledge not only on these interrelated concepts, but also on their interrelations. In other words, ML-driven machines that we are planning to develop and further use should be incrementally tested with humans in the loop, within an agile design and development approach. Tests should be supported by appropriate measures of maturity: technology maturity (TRLs); maturity of practice (HRLs); and societal/organizational maturity (ORLs). In addition, TRLs should now be adapted to include machine leaning capabilities and be transformed into ML-TRLs.

References

Boy, G.A. (2020). *Human Systems Integration: From Virtual to Tangible*. CRC Press, Taylor & Francis, USA.

Boy, G.A. (2021 under review). Socioergonomics: A few clarifications on the Technology-Organizations-People Tryptic. FlexTech Chair Paper, CentraleSupélec – Paris Saclay University & ESTIA.

Sheridan, T.B., & Verplank, W. (1978). Human and Computer Control of Undersea Teleoperators. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.

Appendix

Levels of automation of Decision and Action Selection (Sheridan & Verplanck, 1978)

	Level	The computer…
Low	1	offers no assistance, human must take all decisions and actions
	2	offers a complete set of decision/action alternatives
	3	narrows the selection down to a few
	4	suggests one alternative
	5	executes that suggestion if the human approves
	6	allows the human a restricted veto time before automatic execution
	7	executes automatically, then necessarily informs the human
	8	informs the human only if asked
	9	informs the human only if it decides to
High	10	decides everything, acts autonomously, ignores the human