ONERA

THE FRENCH AEROSPACE LAB

[PhD Thesis TIS-DTIS-2020-12]

Artificial intelligence and transparency: predictability, sense of control and trust

Technological changes at work in aeronautical systems have profoundly changed the interaction between man and machine. As this development progressed, operators found themselves faced with increasingly complex and increasingly automated, or even autonomous, systems. Today, if human beings remain at the center of critical decisions in many fields (medical diagnoses, judicial decisions, decision to open fire in the context of armed conflict, etc.), this decision-making is often based on these decision-making "help" algorithms. It is therefore essential to reflect on our ability to design so-called collaborative artificial agents (Klein & al, 2014). This has become essential for the introduction of advanced functionalities based on Artificial Intelligence techniques in all areas where operators must keep control and responsibility for the evolution of operations.

It has been regularly suggested that one way to draw a more controllable interface is to consider supervision as a coaction between a human operator and an artificial agent following the same principles as a biological agent. Such socialization must necessarily be based on our knowledge of the mechanisms that govern interaction, cooperation between two human agents. In this context, several works have questioned the development of the feeling of control in joint tasks. These works sought to identify the information necessary for the development of this feeling of control, their nature, their temporality. They notably made it possible to highlight the importance of predictability of partner behavior in our ability to coordinate, but also with regard to the feeling of control and confidence developed during these cooperative actions. We can therefore imagine that the intelligibility of artificial agents could directly impact both the operators 'performance but also their acceptability as well as the operators' confidence in their use.

In this thesis project, we propose to use this knowledge in the design of more readable artificial agents. In particular, we believe that the theoretical foundations of agency, as well as the tools and measures stemming from this concept, can allow us to progress in terms of characterizing the readability of an artificial agent, and the interface design underlying this concept. To do this, we will focus more specifically on the neurocognitive mechanisms that underlie the development of the feeling of agency (ie, the feeling of controlling one's own actions and their consequences on the outside world), key mechanisms for understanding our ability to coordinate with others. We will question the information necessary for the development of this feeling of control and we will assess the impact of this information on the development of the feeling of control in a cooperative situation. These questions will be addressed using tools from the research field of agency, and in particular explicit (i.e., subjective relationships) and implicit (i.e., intentional binding) markers, on which classical psychophysical analyzes will be conducted. This will allow us to assess how the information communicated by the system improves the agency, and how the agency can in turn improve the processing of the signals communicated by the system. Finally, we will question the link between this feeling of control and (1) the performance of operators, (2) the level of confidence in the artificial agent, (3) the acceptability of such systems. Through this research program, the question addressed will be that of the readability of AI algorithms, and by extension the link between this readability and the acceptability / use of such tools.

ONERA/DTIS, Location: Salon-de-Provence, contact: Bruno.Berberian@onera.fr, +33 4 90 17 55 88

Thesis Director: Valerian.Chambon@ens.fr, +33 6 44 07 00 29 (ENS – Ulm)