

Responsibility for AI causation

The research aim is to analyse the emerging legal issue of responsibility for AI damage causation, with a multidisciplinary approach based on a confrontation between scientific and legal causality to highlight critical points in artificial intelligence legislation

Vincenzo Calderonio

PhD student
Graduated in law

Affiliations:



INTRODUCTION

The starting point for the research was the work that I made for my master's degree thesis in law, entitled "Artificial intelligence, fictio iuris and causality: jusphilosophical profiles of an actual question". In this dissertation I have analysed the emerging legal framework of EU on artificial intelligence (AI act), comparing it to the most important debate on the field involving the works of Alan Turing "Computing machinery and intelligence" and John Searle "Chinese room argument" on the real capabilities of AI.

Then, the analysis of the literature has been resolved in the individuation of a problem of human-machine interaction, with three elements of interest: **machine learning**, **responsibility** and **causality**.

By the combination of these three elements I made an attempt to define the problem which is now at the centre of the debate on artificial intelligence. The scheme below represents graphically the definition of the problem and its challenge for the epistemological status of causality both in law and science.

KEY ELEMENTS

Artificial intelligence is particularly challenging for our system of thinking because, for the first time in history, we have a non-living object capable of act with some degree of autonomy. Despite of this definition (AI as an object capable of acting) there is a major difference between AI and human-being or animals, the first one does not have any subjectivity. So, from this statement born a difference between two features originally bonded together:

- Causality: the relationship between cause and effect
- Intentionality: the quality of mental states (e.g. thoughts, beliefs, desires, hopes) which consists in their being directed towards some object or state of affairs

—————→
Causality

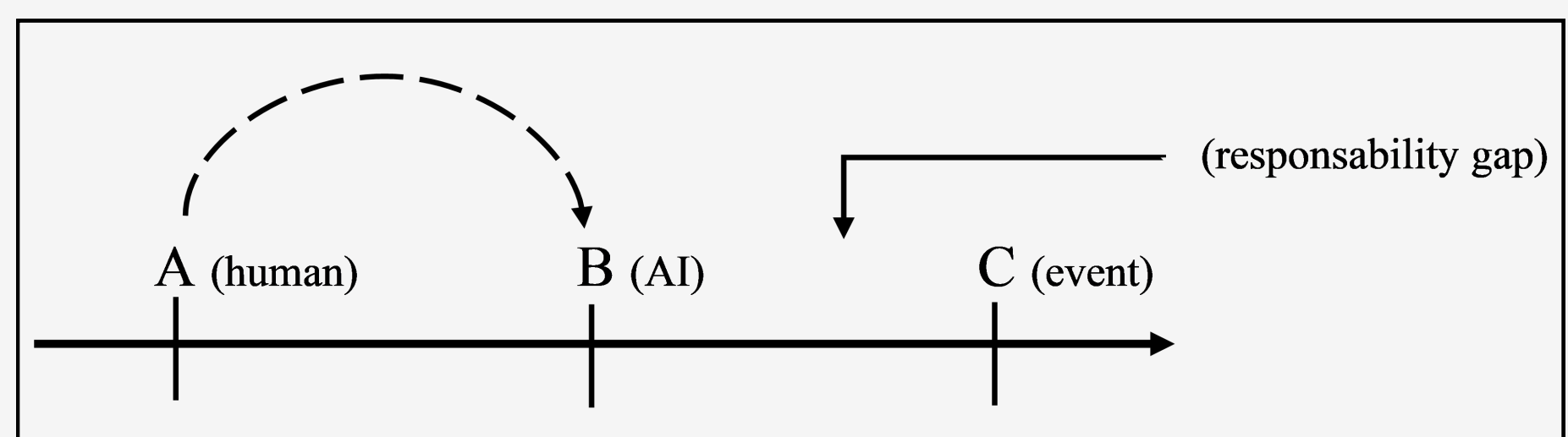
- - - - -→
Intentionality

ANALYSIS OF THE MAIN FEATURES

With the two elements represented by the lines illustrated above is possible to craft a diagram which represent the problematic situation of AI causation, both for the epistemological and legal paradigms.

In this diagram, where the strait line represents causation and the broken line represents intentionality, there is illustrated a classical paradigm of human-machine interaction with the flow of a causal series from A to C, where A is the human input, B is the machine execution and C is the complex event product of the interaction between human and artificial intelligence. The problem for law is the discrepancy between the causation of an event and the intentionality of it; as result of machine learning, the complex output which AI can create is not entirely dependent by human action and this produce the responsibility gap drawn between B-C. Furthermore, under an epistemological perspective there is a big divergence between the conception of causality in science (A-B-C) and law (A-B); because if in science is sufficient to say that the objective causation has created the event C, in law this description is insufficient.

The reason why is that in law the action is always followed by intention and so the legal conception of causality comprehend concepts such as fault, responsibility and will which are not included in the objective perspective of traditional science. This paradigm recalls the problem arised in quantum physics where there is a set of phenomenon which seems to not obey at the principle of causality. From this fact, in time are being developed theories which implies connection between these quantum phenomena and theory of mind; these connections try to respond to new questions: there is a real difference between *res cogitans* and *res extensa*? is needed a reshaping of causality in science and physics to introduce these concepts which are well-known to legal systems?



RESEARCH PERSPECTIVE

The PhD thesis will focus on this paradigmatic change which is activated by the challenge that artificial intelligence creates for our common sense and system of thinking.

In particular the two main research line which will be pursued are:

- Legal aspects of Responsibility
- Reshaping of the Paradigm of Causality

REFERENCES

- ALAN MATHISON TURING, On computable numbers, with an application to the Entscheidungsproblem, 1936; Computing machinery and intelligence, Mind 49: 433-460, 1950.
- JOHN ROGERS SEARLE, Minds, brains and programs, Behavioral and Brain Sciences, 3: 417-57, 1980.
- EUROPEAN COMMISSION, Proposal for a Regulation of the European Parliament and of the Council, laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, Brussels 21.4.2021.
- ERWIN SCHRÖDINGER, What is life?, Adelphi, Milano, 1944.