

Carolina Villegas-Galaviz

cvillegas@comillas.edu / cvillega@nd.edu

PhD Student, Universidad Pontificia Comillas, ICADE.

Research Fellow, University of Notre Dame.

José-Luis Fernández-Fernández

Professor, Universidad Pontificia Comillas, ICADE.

The untouchables for AI

AI is on the rise. Data analytics continues to be the main source for managerial decision-making. In some cases, it might even be irresponsible not to use the power of AI to inform certain decision-making: as in some diseases diagnoses, if there is a tool to improve the precision, it is nonsense not to use it. However, in some other cases, the AI outcome can create an impact such as the use of the model may exacerbate the harm: as in the case of the COMPAS algorithm where the use of the tool disregards unfair discrimination and even create a new problem of injustice.

The purpose of this paper is to delimit the scope of AI while arguing that in some circumstances we should refrain from the use of this technology. First, based on previous literature, we defend that once a path is proposed, those who use AI are directly affected by the proposition. For example, in court, if an algorithm proposes to sentence someone as guilty, it is often difficult for judges to contradict the AI model. In this type of scenario, those who deploy AI defer decision-making to algorithms even when they are the ones responsible for deciding. Secondly, we argue that in some contexts the impact of the decision or action is so big that there should always be a human in the loop. For this second case take the example of how Amazon is algorithmically rating and firing its drivers, without human intervention. Here the drivers only receive an email, sent by a bot, telling them they are fired. In this type of context, the impact of the action should be a cause to refrain from the use of AI, any employee should be treated with dignity and the impact of a job loss is so disrupting that should be done in a certain way: giving voice to the persons affected and treating them in a respectful and attentive way.

Our two arguments are not exhaustive, but the purpose of our article is to start the conversation of when and where to put limits to the use of AI, while identifying those scenarios where society should refrain from the use of algorithms. However, we do not want to overlook all the benefits of the good use of AI.