XVIII Congreso Español de Ciencia Política y de la Administración (AECPA)

GT 5.11 Regresión democrática y burocracias públicas

Do automated decision-making systems facilitate illiberal administrative behavior? A Pre-Registered Experimental Design

Irene Liarte
Postdoctoral fellow
Department of Economics, Public
University of Navarra, Pamplona, Spain
irene.liarte@unavarra.es

Guillem Ripoll
Distinguished Researcher
Department of Economics, Public
University of Navarra, Pamplona, Spain
guillem.ripoll@unavarra.es

Abstract

Automated decision-making (ADM) has become a central feature of contemporary public administration, commonly framed as a means to improve efficiency and consistency in casework. At the same time, many democratic systems are experiencing a rise in illiberal administrative practice, raising concerns about democratic backsliding. Despite the growing relevance of both phenomena, existing research has rarely examined them in conjunction. This study addresses this gap through a pre-registered experimental design guided by the following question: Does automated decision-making increase illiberal administrative behavior by activating displacement of responsibility and dehumanization of the citizen? Building on three bodies of literature—Bandura's moral disengagement theory, Hood's blame avoidance framework, and recent experimental research on algorithmic decision aids—we propose a 2X2 between-subjects factorial experiment. Public employees are randomly assigned to one of four cells crossing the source of a case recommendation (algorithmic system vs. human peer of equal standing) with the framing of the case file (dehumanized risk-flag vs. individuated narrative). Participants complete a casework simulation set in tax administration, an area chosen to avoid the interpretive noise of recent high-profile scandals. This proposal specifies hypotheses, treatments, outcome measures, and a Negative Binomial Regression analysis plan, discussing how administrative technology can actively produce—not merely permit—democratic backsliding.

Keywords: automated decision-making, moral disengagement, illiberal administrative behavior, democratic backsliding

1. Introduction

Public administrations have adopted automated decision-making (ADM) at a rapid pace, usually framed as a means to improve efficiency, consistency, and fairness in casework (). In practice, a growing number of scandals suggests that ADM can just as easily corrode the administrative practices it was meant to improve. Australia's Robodebt scheme pursued hundreds of thousands of welfare recipients for debts calculated by an algorithm with almost no human oversight (Rinta-Kahila et al., 2022). The Dutch tax authority's childcare-benefits system flagged thousands of families for fraud on the basis of a risk-scoring algorithm, disproportionately targeting those with a migration background and driving many into severe financial hardship. The UK's administration of Carer's Allowance produced comparable harm through a data-matching system that left claimants unknowingly accruing debt over years (Kiely & Swirak, 2025).

Some of these cases have been interpreted as failures in algorithmic design, data quality or institutional oversight, rather than as technical accidents (Kiely and Swirak, 2025). Studies have focused on organizational aspects, whilst less attention has been paid to a complementary possibility: that ADM might alter the way in which the officials managing it perceive their own responsibility towards the people about whom they make decisions.

This hypothesis links two strands of research that have, to a large extent, developed separately. On the one hand, academic studies show that algorithmic decision-making aids can facilitate unethical behavior within organizations through moral disengagement, as employees come to regard themselves as mere executors of a system's outcomes, rather than as making autonomous decisions (Jolly et al., 2026; Ebrahimi and Matt, 2024). On the other hand, research on public administration examines how civil servants and organizations avoid blame for unpopular or harmful decisions, often by delegating discretion to rules, protocols or mechanisms that do not leave a single person clearly in charge (Hood, 2011).

Both theories point to the same idea: the removal of a visible human figure responsible for decision-making changes who is regarded — and who regards themselves — as accountable. However, these two schools of thought rarely engage with one another, and neither has been directly tested against a mechanism that research into public administration increasingly regards as fundamental to democratic backsliding: illiberal administrative behavior (Ripoll, 2026; Bauer, 2024). Hoewever, recent experimental research in the field of public administration has examined whether an algorithmic source alters behavior and judgement compared to an equivalent human source, yielding largely null results (Alon-Barkat and Busuioc, 2023; Alon-Barkat et al., 2025). This raises a fundamental empirical challenge: if changing the source of a recommendation does not reliably alter compliance or the attribution of blame to third parties, why should a civil servant's own sense of responsibility change? We argue that compliance, the attribution of blame to third parties and personal responsibility are distinct constructs that are strongly influenced by social desirability bias in standard survey designs.

This study addresses these theoretical and methodological gaps by proposing a pre-registered experimental study to determine whether automated decision-making increases anti-liberal administrative behavior amongst civil servants and, if so, through which mechanisms.