Machine Intelligence, Bureaucracy, and Human Control

Justin B. Bullock†, Hsini Huang‡, Kyoung-Cheol (Casey) Kim‡,‡

‡Independent Researcher
†National Taiwan University, Taipei, Taiwan
‡University of Georgia, Athens, GA, USA

Address correspondence to the author at Kyoungcheol.Kim@uga.edu.

Abstract

Machine intelligence, used extensively throughout modern bureaucracies, is quickly evolving, giving rise to machine agents that accomplish tasks typically reserved for human beings. This shift affects task completion, human/machine coproduction, and the control of the bureaucracy itself. Using Max Weber’s ideal type bureaucracy as a guiding construct, we argue that machine agents may offer technical capacity for task completion beyond that of humans. Furthermore, the technical strengths of machine intelligence, including (1) speed, (2) dispassion, (3) predictability, and (4) rational rule-based functioning, mirror those found within Weber’s ideal type. Through this lens, the evolution of both bureaucratic structures and the decision-making agents within them presents at least three important challenges for modern scholars: (1) deciding the scope of tasks machine agents should complete, (2) adapting the bureaucracy to work with machine agents, and (3) retaining the role of humans and human control.

INTRODUCTION

The bureaucratic organizational structure, with its increasingly powerful technical and administrative tools, has grown to effectively coordinate and control human behavior. Marked by humans’ flexibility, ingenuity, hierarchy, and tendency to specialize within communities, this construct has created empires, global religions, and nation states, and remains the predominant shape of human institutions. The technical and communication advances of the last century have further amplified the power of this organizational form. As recently as 1997, Herbert Simon argued that while bureaucracies had begun to integrate electronic information communication technology (ICT) tools into their workflows, these tools did not yet influence the decision-making apparatuses of these organizations. This paradigm, though true for most bureaucracies at the time, was quickly beginning to change.

Soon after, Bovens and Zouridis (2002) suggested that the form had evolved. Street-level bureaucracies moved towards screen-level bureaucracies, and, in some cases, to system-level bureaucracies. This work found increasingly capable ICT tools to be integrated into bureaucratic decision-making processes, improving speed, capacity, rule application, and task completion capabilities. Systematic digitization and widespread personal computing capabilities produced these so-called screen-level bureaucracies, and organizational structures evolved alongside ICT capacity.

This path has created tools that surpass human technical capabilities across wider and broader task domains (Bullock 2019). The early 2010s brought a new revolution to the field of artificial intelligence (AI), as greater computational capacity allowed the application of probability theory, neural networks, and Bayesian inference to machine learning. These enhancements taught machines to learn and respond effectively and intelligently in increasingly complex environments, representing a qualitative shift in the utilization of intelligent ICT tools. Public administration scholars concerned with discretion (decisions made by an agent in an uncertain environment in which choice is constrained by time, resources, and rules) of human bureaucrats have taken note, arguing that this quality can take human or digital form; machine learning-based discretion, a subset of digital discretion, is called artificial discretion (Young, Bullock, and Lecy 2019). Both digital and artificial discretion feature decision-making and behaviors that differ from those of humans.

The rise of digital and artificial discretion means that bureaucracies, designed to forward public goals, must learn to manage, motivate, and control a new kind of agent. The development and definition of modern bureaucracy is gradually evolving toward a scenario in which both human and artificial bureaucrats exist and play important roles (Bullock and Kim 2020). However, scholars and managers must better understand the synergistic effects and impacts on both human and machine behavior in order to function collectively, effectively, and efficiently (Rahwan et al. 2019).

In 1921, Weber published Economy and Society, with its famous chapter “Bureaucracy: The Nature, Conditions, and Development of Bureaucratic Herrschaft” (Waters and Waters 2015), which brilliantly characterized bureaucracies, their component parts, and their interactions. In particular, Weber observed the conditions of the ideal type bureaucracy, with the position of the “Beamte,” or officer, the prerequisites and side effects of bureaucratization, the persistent character of the bureaucratic apparatus, its economic and social consequences, its authoritative power, the development of the rational bureaucratic structure of “Herrschaft,” and the rationalization of “Bildung” (training).
We borrow Weber’s language here, to integrate his lessons for modern governance and AI safety scholars. At least three important challenges for the enduring application of Weber’s ideal type bureaucracy lie ahead: (1) the scope of tasks machine agents should complete, (2) human and machine coproduction within a changing organizational structure, and (3) maintaining human control.

To explore these problems, we first provide a more detailed description of Weber’s characterization of ideal type bureaucracy. Then, we discuss the differing views on AI decision-making within bureaucracies. Next, we consider the role of machine learning as it exerts evolutionary pressures on the shape and structure of bureaucracies. Finally, we stage these three guiding challenges in a world in which organizations contain both human and machine agents. We conclude by reflecting on the implications of this bureaucratic evolution for governance in the modern world.

**WEBER, BUREAUCRACY, AND MACHINES**

Weber’s (1978, 956–58) *Economy and Society* lays out the six features of the ideal type bureaucracy (*Beamtentum*) listed below:

1. The modern bureaucracy “is based on the general principle of precisely defined and organized across-the-board competencies of the various offices (Behorde).”
2. The principles of “hierarchy of offices” and “proper channel” exist.
3. Modern administration is based on a) documents preserved as original copies or concepts, and b) a staff of subordinated Beamte and writers of all kinds. Beamte working in the Behorde, together with relevant resources including material goods and documents constitute a bureau (Buro).
4. The work of the “Beamte typically requires an in-depth specialist training to undertake all specialized tasks of the government officer (Amt).”
5. A full-fledged Amt occupies all the professional energy of the Beamte to process its tasks, regardless of limits to his mandatory working hours.
6. The duties of the Amt undertaken by the Beamte are based on general learnable rules and regulation, which are more or less firm and more or less comprehensible. The knowledge of these rules and regulations thus constitutes a special kind of “applied science,” which the Beamte possesses.

In Weber’s modern ideal type bureaucracy (Beamtentum), various organized sets of offices (Behorde) containing a variety of tasks bundled into individual positions (Amt) are fulfilled by professionally trained human workers (Beamte). These overriding characteristics elucidate the classic relationships among organizational structure, task design, and the knowledge and duties of humans fulfilling these tasks.

This article suggests the need to broaden the conceptualization of the Beamte to include machine agents as well as human, and explores how this could affect the structure of offices (Behorde) and the broader definition and function of modern bureaucracy (Beamtentum). To illustrate, we take in-depth look at Weber’s ideas about Amt and Beamte and their rational and technical superiority in the Beamtentum.

Weber (1978, 959) describes the Amt as “a calling and a profession... This can be seen in the requirement to complete a prescribed course of studies and to pass subject examination as a prerequisite for employment. This can (also) be seen with the inherent ‘nature of the duty’ with respect to the Beamte’s position.” Beamte is also defined by an “aim to gain and enjoy a high appreciation by people in power,” nomination by a higher authority, lifetime appointment, regular remuneration, and professionalization. As we proceed, we must consider ways in which machine agents may fulfill some of these traditional Beamte roles, but not others.

Beyond listing the required characteristics and necessary reforms to optimize modern organizations, Weber argues for the technical superiority of the bureaucratic system to other forms of organization and control. Using collegial and honorarium pay as examples, Weber (1978) notes that:

Ultimately, such administration systems face limitations as the qualitative complexity of their tasks increases, as happens in England today. Further work based on collegiality implies frictions and delays, and compromises between colliding interests and views. Therefore, the work performed is less precise and less dependent on superiors, hence it becomes less uniform and slower.

In further support of the bureaucratic system above all other administrative forms, he continues that:

A fully developed bureaucratic mechanism compares with other organizations in the same way a machine compares to non-mechanical means for producing goods. Precision, speed, unambiguity, knowledge of the files, continuity, discretion, unity, strict subordination, reduction of friction and of material and personal costs, are raised to the optimum point in the strictly bureaucratic administration. (p. 973)

However, Weber did recognize and explore the challenges to bureaucratic administration in most modern societies, which include (1) an acceleration of needed tasks, (2) maintaining the necessary objectivity and dispassion needed to accomplish those tasks, (3) the primary importance of predictability, and (4) the increasing complexity of justice and the application of law. The century between his observations and today has illustrated the salience of these points, as organizations have developed and evolved alongside political and technological changes. Even so, scholars of decision-making agree that organizations remain rational and reduce uncertainty (Gajdushek 2003), and that the role and discretion of the human Beamte remains paramount to implement complex rule-based systems well. Of discretion and the trade-offs across individualized and standardized (mechanical) approaches, Weber (1978, 979) says:

The idea that there can be a law without loopholes is generally strongly contested. Also, the idea that the modern magistrate is nothing more than a ‘judging machine’ is rejected with disgust. In such a judging machine, files and costs would be thrown into the top of the machine in order that it would spit out the verdict along with a mechanical reasoning at the bottom... In fact, though, as within the
domain of the findings of justice, there are areas where the bureaucratic judge is instructed by the legislature to use ‘individual’ paths to find justice. Moreover, in the area of the actual administrative functions, including all government activities that do not belong to the area of creating or implementing the law, or the finding of justice, the freedom and domination (Herrschaft) of individual approaches to tasks are simply taken for granted. In contrast to such emphases on individualized approaches stand the invariable norms that play a negative role, because they restrict the never-to-be-regulated ‘creative’ work of the Beamte, which is normally seen as a positive thing.

We should note that the “judging machines” of Weber’s time have little in common with those in today’s bureaucratic world (Desiere, Langenbucher, and Struyven 2019). Now machines can, to some significant degree, learn. They can be trained on input data and “general learnable rules and regulation.” Modern machine agents may go further, in effect and practice, to balance standardized and individualized approaches, as do human Beamte. In fact, machine agents may be superior to human Beamte in some domains, in terms of their abilities to learn, memorize, and adapt to tedious rules. However, how does this superiority translate in circumstances requiring individualized approaches to justice?

BUREAUCRACY, BEAMTE, AND BEHAVIOR

Bureaucracy is a systematic mechanism, steeped in rules and procedures through constitutional authority, used as a fundamental mode of control (Waters and Waters 2015; Weber 1922). Weber’s ideal type bureaucracy follows principles of rationalization, transferring of authority, specialization of jobs, a top–down organizational arrangement, and vertical and horizontal communications (Weber 1922). Division of labor, for example, appears in the context of policy implementation, in which high-level authorities may be responsible for strategic planning while ground-level workers focus on front-line execution. Meanwhile, limits of the legal system and restrictions in communication introduce ambiguity into rule- and procedure-based functioning at each level, leaving administrative agents to actualize their own discretion (Rourke 1972). However, as Weber’s bureaucracy also emphasizes exact goal and job specifications, assigned by the top authority via hierarchical channels through the Behorde (office), Büro (bureau), and Beamtenrath (bureaucracy), rules and discretion are often in tension. Scholars have considered this tension, suggesting that discretion in bureaucracy could either strengthen professional autonomy and performance or weaken them, depending on the structure of the bureaucracy and the technological tools available (Bovens and Zouridis 2002; Lipsky 1978; Maynard-Moody and Musheno 2000; Young et al. 2021).

Hierarchy and specialization within bureaucracies segregate levels of authority and channels of communication by rank and position, complicating control over the individual Beamte. First, the principal-agent problem applies: subordinates, motivated by individual interests, may obtain or act on hidden information to avoid accountability or shirk responsibility. Secondly, trade-offs occur between specialization and coordination, challenging top–down implementation of authority and bottom–up flows of communication and expertise. In addition, while Weber maintained that rationality equals efficiency, the relationship remains an open question in the field (Scott and Davis 2006, 52; Simon 1997). Weber (1978, 973) uses words like “precision,” “reliability,” “unambiguity,” “continuity,” “discretion,” “calculability,” “speed,” and “reduction of friction, material, and personal costs” without direct linkage to efficiency outcomes. In fact, a wider consideration of Weber’s full bibliography suggests that a bureaucratic organization matters more for predictability and formal rationality than for economic efficiency (Gajdushek 2003).

Following Weber’s conceptualization, machine Beamte, designed to be fast, dispassionate, and predictable and follow technical-rational functioning, may eventually achieve the actual technical superiority promised by ideal type bureaucracy. Machines may even be more suitable for many tasks; in Weber’s language, they may not be the most-preferred model, but sufficiently rational and simplified to govern in the civilized West (Scott and Davis 2006). Weber also emphasizes that the ideal type bureaucracy should deliver objective and specialized expertise of administrative work, which also resonates with the general expectation of intelligent, machine agents.

Meanwhile, the use of human agents as Beamte, embedded within the rule and procedure-based systems designed with hierarchy and specialization, introduces important limitations. Psychological and interpersonal interactions clearly influence the actual bureaucratic conduct of human agents, suggesting that human decision-making can be biased, noisy, and riddled with errors (Tversky and Kahneman 1974; Wilke and Mata 2012). Simon (1997) places blame on limitations in information processing and communication. Other scholars argue that bureaucracy operated by humans suffers from red tape, power abuse, and opportunism (Lee 1984; Rainey and Bozeman 2000). Continually improving AI systems may allow organizations to incorporate consistent, stable, dispassionate, intelligent agents that match Weber’s ideal type, though challenges and uncertainties remain.

The integration of machines into bureaucracy is not a new concept; prior insights may help explain the unique challenges presented by AI systems and machine agents. Digital computer and ICTs already support many administrative tasks, producing innovations in record documentation, filing, and storage and aiding efforts to alert, analyze, predict, and make decisions. E-government reforms in the past 20 years have changed the core of public administration by reintegrating public sector processes, designing need-based services, and digitizing the nation (Fountain 2001; Dunleavy et al. 2006; Margetts and Dunleavy 2013).

Many prior studies in this field focus on ICT tools to facilitate organizational administration and the factors and resource conditions required to accept or introduce them, such as organizational culture, openness, leadership, technological readiness, self-efficacy, and acceptance to innovation (Bannister and Connolly 2014; Borins 2001; Grimmelikhuijsen and Welch 2012; Grimmelikhuijsen and
Few studies view emerging digital technologies as drivers of organizational and structural change. Many debate whether new digital technologies curtail or enable discretion through different phases of digital governance (Buffat 2015). Other work tracks the evolution of bureaucratic work in public organizations, alongside that of ICT, from street-, to screen-, to system-level bureaucracies, resulting in the transmutation of task content and discretionary power (Bovens and Zouridis 2002; Bullock, Young, and Wang 2020). As Bovens and Zouridis (2002) suggest, the increasing role of rule-based ICT tools could radically reduce discretion among frontline bureaucrats, transferring it to system developers. In addition, computer automation itself may add to rather than ease some administrative burdens, as when dehumanized interactions with service providers result in an overload of unwanted emails (Bozeman and Youtie 2020).

Prior literature uses complexity and uncertainty to classify organizational tasks (Galbraith 1973). While in modern digital society, computers and information technologies cope better with complexity, they do not respond as well to uncertainty and adaptation. Conversely, humans can react quickly to exceptional situations using metacognition across knowledge domains, though they may compromise the basic value under the initial guidelines (Lee 1984). Following this logic, we argue that the introduction of machine agents is likely to greatly alter both the bureaucratic structure of controls and the behavior of the bureaucracy itself.

By assigning human and AI agents to specialized tasks and positions based on their own comparative advantages, we can imagine and expect substantial changes in the structure of hierarchy and specialization. For example, one might imagine both vertical and horizontal communication channels extending across human and machine agents according to their roles and functions. Thus, machine agents may take on tasks that include vertical and horizontal communication and transfer of authority, presenting challenges for their incentivization, motivation, and control within the decision-making system of a bureaucracy. For example, will machine agents be successfully monitored by their superiors? How will machine agents give and receive tasks and monitor the work of other machines and humans? How might this alter communication channels or affect future evolutionary pressures? These “facts on the ground” for bureaucratic evolution present a number of challenges for modern governance scholars. In the next section, we propose three ongoing challenges and discuss them in some detail.

THREE CHALLENGES FOR MODERN GOVERNANCE SCHOLARS

Challenge 1: Deciding the Scope of Tasks Machine Agents Should Complete

As illustrated above, machine behavior is becoming more complex and better able to flexibly adapt to more complex environments across a greater number of task characteristics. In particular, machine behavior, fueled by AI and machine learning, appears to often excel, relative to human behavior, at the same rationalizing characteristics that Weber identified for bureaucracies, namely speed, predictability, dispassion, and rule-following (Weintraub 1948). By direct inference, machine agents may be advantageously applied to bureaucratic tasks requiring improvements to these four aspects.

Even before the era of AI, ample studies linked the various types of information systems (e.g., transaction processing, management information, decision support, and expert systems) with the hierarchy of organizations. Computer system engineers have long aimed to develop appropriate and specific systems suitable for different operational, managerial, or strategic planning levels (Laudon and Laudon 2005). The relationship between information systems and organizational automation once related closely with workflow, information processing, and clear internal rules for reliable inputs and outputs (Salisbury 2003). This specific characteristic now distinguishes traditional computer algorithms from machine learning algorithms for AI models.

As intelligent machine agents become better and faster in terms of calculation power and self-learning ability, we are likely to see a radical redefinition and redistribution of tasks among human and machine agents within organizations. For example, in the United States and EU countries, government AI use has begun to spread across areas of enforcement, adjudication, public engagement, regulatory analysis, and internal management (Engstrom et al. 2020; Misuraca and Van Noordt 2020). Intelligent machine agents process general requests, standardize paperwork flows, and even detect possible fraud for drug authorities, health insurance administration, and student aid services.\(^3\) Local authorities are embracing this trend too. For instance, regarding the UK introduction of smart technologies into local public services, Vogl et al. (2019) observe that street-level bureaucracy collaborating with machine AI models is likely to be more responsive and outcome-oriented, and, hence, more specialized, better facilitating the different agents involved in service delivery.

However, since rule-based systems are incomplete and may require professional judgment, comparative advantages in speed, predictability, dispassion, and rule following may be offset by machine agents’ difficulties with issues of uncertainty or justice. Both uncertainty and justice are inherent in many tasks, but predictive policing (Meijer and Wessels 2019; Oswald et al. 2018) provides a clear example. Machine agents’ scope for task completion can be limited both by the level of discretion required for a specific task (Bullock 2019) or as a function of task uncertainty and complexity (Lee 1984). Put another way, in task domains with significant uncertainty, requiring professional judgment, or prioritizing individual paths to justice (as Weber argued), we should be careful to limit the use of machine agents in favor of human Beamte.

However, the list of tasks that machine agents can complete using factual, digital, multi-modal data is growing. Machine learning, in particular, now allows AI machine agents to intelligently and effectively infer value judgments, such as political orientation, consumer behavior, and preferences, based on a user’s digital searching, purchasing, and clicking records. These tasks are likely to have significant negative impacts, which should caution us against the use of AI. We should constantly review what facts and values AI might integrate comprehensively into its decision-making models. In this

\(^3\)https://studentaid.gov/h/aidan.
situation, the conduct of AI could be both inefficient (make inaccurate predictions) and unfair (Danaher et al. 2017).

As their utilization increases, machine agents execute more tasks at different levels in the hierarchy, another key aspect of administration. AI’s role in the military presents the intriguing and classic relationship among bureaucrats (both human and machine agents) and bureaucracy. The United States (The National Security Commission on Artificial Intelligence [NSCAI] 2021), China, South Korea, and many other nations have developed unique applications for AI in the realm of national defense. For example, South Korea has designed plans to allow an AI system to decide tactics and give orders for platoons, based upon machine learning grounded in field information (Lim 2020). This intelligent battlefield recognition system connecting air-sea-ground-based command centers and equipment will both collect and analyze data. Within the bureaucratic structure, this can be explained as human agents ordering machine agents on sub-levels.

Given machine agents’ general tendency of increases in speed, dispassion, predictability, and rule-based action, many expect, with some concern, that they will grow to constitute a substantively large portion of organizational decision-making and task completion. As machine agents advance and improve within the bureaucratic structure, the first challenge to the ideal type bureaucracy becomes to identify which types of tasks benefit from these increases, and separate out those tasks that require individual pathways to justice or professional judgment, or are characterized by high uncertainty. Current evidence suggests that this is being overlooked in practice. Many organizations deploy machine agents as if they were Beamte, without assurance that they are best qualified for the tasks at hand.

While some advise that machine agents will and should take the primary portion of administrative tasks, more scholars agree that a multi-agent decision-making model is most appropriate (Bullock and Kim 2020; Peeters et al. 2021). This suggests a complementary and coordinative relationship between human agents and AI agents to solve problems with high uncertainty and high complexity.

Therefore, this first challenge demands an immediate investigation to measure the level of qualification and specialization of each machine agent, and thus determine their ideal scope of tasks. Secondly, we should expect to see some reforms in machine–human collaboration in bureaucracy. To better comprehend the consequences of such reform, in terms of bureaucratic structure and actual operational practices, we must first examine the potential roles of machine agents as co-workers, subordinates, or supervisors. We elaborate on this co-production scenario below.

**Challenge 2: “Adapting” the Bureaucracy to Work with Machine “Agents”**

As we have highlighted, the complexity of modern society requires both task specialization and organizational hierarchy, which imply agent coordination and communication across various ranks and positions within bureaucracy. Popular AI-enabled communication machines (e.g., chatbots) can provide the necessary channels for this collaboration. Much of the prior literature has characterized AI use as an informative and supporting tool, used by human Beamte to support efficient public service delivery (Aoki 2020; Vogl et al. 2019). We argue that modern AI systems, previously understood as machine agents, deserve their own category, as a new type of agent for decision-making and task completion. This, again, suggests new types of co-working arrangements across human and machine agents.

Fundamentally, we anticipate human and machine agents to co-work within a multi-agent bureaucratic system (Bullock and Kim 2020), with both human-to-machine and machine-to-machine relationships. There are a number of ways to conceptualize these co-working arrangements. The Weberian hierarchy and specialization characteristics suggest that both human and machine agents will make decisions by orders, rules, or, communicating with each other, either on the same horizontal level within the bureaucracy, or along the vertical levels where supervisor and supervisee relationships take place.

However, the communication of machine–human co-working is likely to be incomplete on both axes, partly because of the transparency problem of AI. The process of AI decision-making remains unexplainable, that is, a black box (Bannister and Connolly 2020). The machine learning algorithms of AI could be unsupervised and adaptive, implying a self-learning mechanism not fully understood by humans. Due to the very characteristic comparative advantage of AI in calculation, humans cannot predict how AI will respond in certain computational outcomes. Machine-to-machine interactions may compound this problem, as one machine agent makes decisions based on the insights of another.

In response to the black box problem, many current trials are working to develop Explainable AI (XAI). One tentative solution involves regulatory coding. In addition to variations in accuracy and performance, engineers also fear potential unexplainable mutations of the machine learning system itself. A proposed response is “opening the black box,” or developing additional AI system(s) to check layer by layer through the initial neural network of the original. While some scholars find this promising, another camp argues that problems caused by AI cannot be solved by applying additional AI (Peeters et al. 2021).

Some recent AI research imagines actualizing its socio-technically informed perspectives, a step beyond dominant algorithm-centered solutions to transparency and XAI (Ehsan et al. 2021). Given the possibility of societal collaboration by incorporating AI agents (Daylamani-Zad, Agius, and Angelides 2020), experts are also using a multi-agent systems approach for intensive research on utilization (Carrera and Iglesias 2015).

Yet the process of incorporating the multi-agent system into the bureaucracy may bring unforeseen risks. Some research and practices have begun to call for regulatory frameworks and legislation regarding AI (Bannister and Connolly 2020). For example, the European Commission (2020) aims to propose new bills to regulate the uses of AI based on a risk assessment framework. However, like humans, theoretically speaking, AI systems are virtually non-identical. Even with the same set of initial training data and similar computing and programming infrastructure, variations and mutations may occur. To what extent the XAI approach and legal regulation could solve the transparency concerns over this organic evolution is thus far unknown.

Advances in human–AI co-working involve substantive transformations of classic organizational functioning, such as responsibility, loyalty, and motivation. AI is designed to be dispassionate in its administrative behavior; it only “cares”
about, say, reward maximization based on an explicit function, and, by default, ignores many higher concepts when making decisions and interacting with other agents. Humans, on the other hand, are influenced by responsibility, loyalty, and motivation, as well as broader factors such as value, emotion, and cognitive limits (Simon 1997), implying that AI interventions in the co-working environment could interfere with human agents’ decision-making. Furthermore, leadership serves as a major control and motivation mechanism for humans, inspiring individual and group values to reflect relevant and pertinent environmental conditions (Rainey, Fernandez, and Malatesta 2021). Such higher concepts and motivations may help humans maintain ultimate control.

Issues of bureaucratic rank may influence acceptance and compatibility of human–machine co-working. As the scope of tasks assigned to machine agents increases, we may think of the development of these co-working arrangements as potential human-enhancement technology (Korinek and Stiglitz 2019) or potential human-curtailment technology (Buffat 2015). For example, Huang et al. (2021) argue that the perception/acceptance of AI use in organizational routine tasks varies more widely among public managers than among non-managers. Young et al. (2021) found that the use of AI could affect automation bias of human beings or, in other words, increase reliance of humans on AI for making decisions. These findings introduce an intriguing causal relationship among machines, humans, and final decision-making in the context of human/AI co-working. While some novel research explores how the use of AI could impact human decision-making and cognitive reactions, scholarly interest in value, transparency, accountability, and coordination issues has lagged. Since administrative behavior and relevant factors are systematically interconnected (Simon 1997), more consideration is needed, and soon. Scholarly answers to these questions can prepare the bureaucracy to adapt its environment to house new types of agents that are differently motivated, incentivized, and operated than the human Beamte for which the modern bureaucracy was built.

This challenge builds upon the first, which examined the difficulty of identifying which classes or types of tasks a machine agent may complete within the bureaucracy. Machine agents lack the capacity to navigate significant uncertainty, engage in professional judgment, and find individualized pathways for justice, and so should not be assigned such tasks. The second challenge examines the consequences of integrating machine agents into a bureaucratic structure where specialization and hierarchy are present, focusing on the role of traditional management and their control tools. Notably, these types of agents engage in actions and make decisions in fundamentally different ways, which expand the multi-agent system of bureaucracy to include a host of new types of co-working arrangements.

From here, the third challenge becomes clear. If machine agents have technical strengths akin to that of the bureaucratic structure itself, namely speed, dispassion, predictability, and rule following, then how do we ensure that humans retain control of tasks within this environment?

**Challenge 3: Retaining the Role of Humans and Human Control**

One may argue that Weber is most concerned with this question of control. Based upon observations on the industrialization as well as the First World War with the military formation and operation, Weber’s ideal bureaucratic structure suggested a holistic view on how organizations, or groups of humans, can be rationally and logically organized towards a common purpose. The bureaucratic structure in a complex modern society demands hierarchy and specialization; this complexity demands that control be at the center of the rationalization process. Weber argued that human Beamte could be controlled through the rational application of structured communication, authority, and specialization. In this structure, a human Beamte is essentially constrained within a narrow domain of specialty. A top authority, generally speaking, sets the goals and agenda, and the Beamte execute those based upon their Amt (position) and Bildung (training).

This description of control, as laid out by Weber, highlights two crucial locations at which the control of a bureaucracy may be altered by the introduction of machine agents into its functioning. The first is the level of the individual agent within the organization. Weber’s explanation for human control relies on notions that appeal to human motivations, such as prestige and regular remuneration, which, as discussed above do not directly apply to constraining and guiding the behavior of machine agents. Another mechanism used to control human Beamte is the allocation of ranks and positions based on the specificity of tasks to be completed and responsibilities to be assumed. To take an extreme case, foot soldiers do very little in the way of directing or controlling a bureaucracy: they have discretion on a much narrower set of organizational tasks. We can apply this mechanism when considering the control of machine agents as well.

New developments in machine agents and AI tools aim to achieve flexible generalized capabilities based upon data training, and then self-learning models to solve general and specific tasks. By juxtaposing our above arguments with the view of technology determinists (Kline 2001), we see that technology leads to changes in social structure and culture, and not vice versa: the frontier thoughts and invention of AIs will arrive before humans know that they need them, and certainly before they know how to utilize them. It remains to be seen whether we can successfully direct and control intelligent machine agents capable of some arbitrary level of generality. Collectively, these concerns present a crucial challenge for maintaining control over both human and machine Beamte within a bureaucratic structure.

The concerns above also address the unknown territory of human–AI interactions in the supervisor–supervisee relationship. Changes in this dynamic could influence overall information processing and decision-making, through both top–down orders and bottom–up observations and feedback, from a cognitive behavior (Newell and Simon 1972; Simon 1997) or MIS system design point of view (Laudon and Laudon 2005). The increasingly intelligent machine, as a general exponential rule, may present more general intellectual challenges to humans’ collective abilities: how will we know when we are (and are not) being manipulated, and, to some degree, controlled by machine agents? In particular, AI models guide machine agents’ behavior to achieve perfect information processing and efficiency, and, therefore,
economical rationality, without the psychological reflection of values. Many fear an alternate scenario, in which machine agents not only surpass and challenge the skill of front-line task-oriented human Beamte, but replace the conventional vertical and horizontal controlling mechanisms. Once machine agents receive higher-level decision-making authority, within an environment of information, communication, decisions, and actions, they can strongly influence bureaucracy, a social organization of control.

Even worse, scholars debate whether AI has a value misalignment problem: the means-end rationality of an agent programmed to achieve a goal is necessarily heavy on the end side. This consequence could challenge the legitimacy of decisions and authority. Russell (2019) suggested that humans should inspect and permit every AI decision to ensure value alignment, but may only partially solve the control problem.

Agency theory considers these concerns of value and goal misalignment to be representative of the principal-agent challenge. Through this lens, recent work has connected such incompatibility with pressures of dehumanization and administrative evil (Young et al. 2021). As machine agents are given more autonomy and influence within bureaucracies, further increases in administrative evil may result, as a consequence of: (1) the technical inscrutability of machine agents, (2) increased quantification bias, (3) centralization of control by machine agents, (4) organizational value misalignment, and (5) AI exuberance.

The issue of maintaining human control on bureaucracy will be vital as more and more machine agents with independent artificial discretion are deployed. For instance, the US Air Force is already integrating AI agents (e.g., unmanned missile systems) with human soldiers, calling it human–AI teaming. Similarly, the UK and the US Navy are working on utilizing AI to control submarines, while China is developing and testing AI-enabled drone submarines. Some reports find that the success of human and AI pairing in the military requires trustable, reliable, and predictable machines (Konaev and Chahal 2021). This said, the controlling mechanisms should be arranged subtly, to address both AI’s mistakes and human errors in the face of various AI-caused biases, such as automation bias and overtrust in technologies.

Given our general arguments on bureaucratic discretion in organizations, the existing controlling mechanisms to enhance positive collaboration are still immature. While legal scholars argue for regulative controls on the use of autonomous vehicles, robots, and other powerful AI applications, the process of legalization and the scope of regulated domain raise new applicability concerns. This work, which calls for more active human–AI co-working, returns this question to the organizational level, in the context of both human-to-human and AI-to-AI interactions (Bullock and Kim 2020). We may also be able to retain human control if we allocate machine agents to suitable functions and roles, based on thorough analysis of the focal tasks and workflow.

To sum up, limitations in AI information processing make the retention of human control in bureaucracy essential. Agents (both human and machine) throughout different ranks and positions, from top authority to frontline staff, will execute their discretion. However, while the management and control issues of humans stem from our own subjectiveness, values, and emotions, machines face different control problems, based on value misalignment, goal-maximizing problem-solving, and dispassion.

**DISCUSSION AND CONCLUSION**

In 1921, Weber defined bureaucracy as a rationalizing force on complex social challenges. In his conception, bureaucracy used hierarchy and specialization in a technically superior fashion that encouraged speed, dispassion, predictability, and rational rule application. However, he noted concerns around its mechanical nature and encouraged individual approaches in areas of justice. His interpretation rings true in the modern understanding, which prioritizes human rights and holds that all legal codes are, to some degree, incomplete. This work upholds this line of thinking, and recommends that machine agents, acting as Beamte within bureaucracy, be relegated to task areas less concerned with justice, with little uncertainty, and requiring no professional judgment. While AI does out-perform professional experts in some classifications, bureaucracy must depend on human Beamte, and their nuanced responses to ambiguity, to retain a significant overlap of the ineffable components of human and organizational value. This, as we have argued, is a cautionary principle.

As established in the literature (Bovens and Zouridis 2002; Bullock, Young, and Wang 2020), bureaucracies have already begun co-evolving along with their newly incorporated technologies and agents, from street-level to screen-level and on to system level. The trend is plain: machines are playing an ever-increasing role in the functioning of bureaucracies, in technical and administrative tasks central to their Weberian specialization and communication channels. In this way, even before the full development of machine agents that could function as Beamte, humans have already ceded much control over bureaucratic functioning.

The ideal type bureaucracy assumes a high level of efficiency, with its mechanistic transfer of goals and orders and actions falling into line. It suppresses personal decisions based on values. Therefore, an increase in the use of machine agents, primarily for fact-based information processing, may improve efficiency in general. Additionally, humans themselves do not reliably make the best decisions for effective, efficient, and consistent administration (Kahneman, Sibony, and Sunstein 2021). Moreover, authority requires active involvement and cognitive acceptance by individual human beings, which cannot result from a mechanical application (Barnard 1974; Simon 1997).

In principle, the ideal type bureaucracy is composed of two aspects: the specialization of jobs and the hierarchical transference of authority to horizontal levels (Simon 1946, 1997; Waters and Waters 2015; Weber 1922). Supervisors or managers are responsible for coordinating horizontal flows as well, while sub-level entities are primarily expected to perform specialized jobs (Bullock and Kim 2020; Kim 2021; Simon 1997; Weber 1922). In addition, bureaucracy is based upon institutional grounds: it takes in laws and rules as the fundamental medium of administrative behaviors. In contemporary constitutional nations, especially, public organizations are expected to perform administrative behaviors primarily through the applications of laws and statutes, both to protect administration from excessive political obligations and accountability and to protect citizens from administrative
abuse or power. Currently, human beings are the only acceptable legal subjects. To imagine a better legal framework for issues of human–machine co-working and control, scholars and practitioners must first address the possible consequences and optimal utilization.

The field of public administration studies how bureaucracies solve social problems using efficient and effective governmental interventions. Government itself is often symbolized by the typical bureaucratic structure of governmental agencies (the intra-bureaucratic approach); it can be further understood by its relationship to other institutions, such as the legislative and judiciary branches in most democratic countries, as macro-level power controlling (the inter-bureaucratic approach) (Krause and Meier 2003; Miller and Whitford 2016). To that end, public administration scholars have distinguished public bureaucracy from private sector equivalents using this system of checks and balances, with its more coherent legal and constitutional controls (Kettl 2008).

The virtue of revisiting Weber’s ideal type bureaucracy for this line of inquiry is that it highlights a lack of discussion about (1) how institutional influence impacts bureaucratic practice in actualization, and (2) potential associated implications for structure, task allocation, and control. In fact, our approach suggests that the politics-administration dichotomy aligns with the orientation of the ideal type bureaucracy. It implies a complete detachment of mechanistic behaviors by sub-level bureaucrats—including the top authority, which may represent both political and managerial dimensions—from macro-level control and institutional influence. Weber considered bureaucracy to both control and be controlled by society and economy, as reflected in the structure of his book (Waters and Waters 2015). Although bureaucracy is about strictly formalized organization, it does not deny associative relationships with the surrounding institution.

At this point, a completely mechanistic ideal bureaucracy may never exist, unless the hypothetical singularity (the point in time at which technological growth becomes uncontrollable, with awful implications for humanity) arrives. If robots manage the physical dimension using completely simultaneous communication with a single decision-making entity, that being could decide to eliminate every additive artificial (Bostrom 2014). We stop short of this hypothetical, and focus on optimizing the use of AI (both AI–human and AI–AI interactions) within human-controlled bureaucracy. Since human bureaucratic agents are also individuals who subjectively interpret their institutional environments, institutional control and influence cannot be completely restricted to the top authority in contemporary bureaucracy.

Considering bureaucracy from the perspective of organizational study can yield significant and practical implications for the intervention of machine intelligence. Concerned scholars and practitioners should be sure to check the functioning and implications of AI deployment with extra care to prevent existential risks, beginning with top-level and critical positions and ultimately extending throughout the entire bureaucracy. Since humans have relatively limited analytic capabilities, effective monitoring and correction may require more active use of AI, to better understand their competition in the multi-agent system context. Human capabilities may also be increased in the long term, using as-yet-unknown technological developments.

Human history has taught that passive approaches, grounded in restrictions, will dampen our abilities to deal with unexpected and accidental events. As AI’s power, scope, and scale increase exponentially, this danger will grow. Active military application of AI is growing significantly, raising questions about hegemony and competition among powerful nations. Meanwhile, civilian AI applications may distort the public allocation of interests and values among those receiving service. Eventually, without dedicated scholarly attention, these issues will converge, with implications for human ethics and humanity itself.

References


