BETWEEN ECONOMIC BENEFITS AND ETHICAL RISKS

Examining the knowledge production of AI in

public administration

MOLLY LUNDGREN

Published by Media@LSE, London School of Economics and Political Science ("LSE"), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

ABSTRACT

This dissertation deploys a Qualitative Content Analysis (QCA) of policy documents to examine the Swedish government's initiative: "Promote public administration's ability to use AI". The study employs the theoretical frameworks of the '4Ks' (Gunter, 2017) and 'contributory epistemic injustice' (Dotson, 2012) to analyse policy formulation as a process of knowledge production. By integrating these frameworks, the research examines the characteristics of this knowledge production and assesses whether it exhibits contributory injustice. Three components of the '4Ks' – knowledges, knowings and knowledgeabilities – are reinterpreted as 'hermeneutical resources', a term defined by Dotson (2012) to reference meanings and interpretative frameworks. These components, along with the fourth 'K' – knowers, constitute the elements of knowledge production. By dissecting risk and benefits -assessments made within the policy documents, the aim was to outline and interpret the policy with regards to deployment and exclusion of hermeneutical resources and knowers. The analysis identifies different sets of hermeneutical resources and knowers depending on assessment aim within the policy, revealing a discrepancy between how knowledge was produced for 'benefits' in the policy versus that of 'risks'. Claims around benefits of AI in public administration were found to promote simplified assessments of quantitative economic benefits, omitting adequate evaluation of benefits for citizens. Conversely, citizens were centred in the risk assessment, which considered ethical risks while neglecting economic implications. In outlining the differing characteristics of these assessments, the discussion unpacks the imbalance between knowers deployed, and how the construction of benefits and risks obstructs critical response. The findings illustrate how knowledge production of the policy formulation is not a neutral assessment of facts, but rather a constructed product of preferences aligned with the policy goal of increasing the use of AI. However, the study also highlights theoretical limitations of the contributory injustice framework in assessing marginalisation in knowledge production, which furthermore underscores a need for future studies of ideological underpinnings of knowledge production in policy formulation.

INTRODUCTION

The digitalisation of the public sector has been highly prioritised by political leaders all over Europe for more than a decade, with the European Union driving policy developments forward emphasising the relation between digitalisation and economic growth (European Commission, 2019; Giannone & Santaniello, 2019; Mura & Donath, 2023). Numerous parts of the public sector have undergone extensive digital transformations over an extended period, but policy-solutions to the Covid-19 pandemic significantly catalysed further demand for digitalisation, establishing it as a critical response to challenges for public life (Moser-Plautz & Schmidthuber, 2023). As highlighted by the targets in the 'Digital Decade Policy Programme 2030' (Decision 2022/2481), the European Union aims to enable 100% of key public services to be accessible online by 2030, accelerating digitalisation in public administration. While some scholars point to benefits as improved forecasting, reduced administrative burdens, more sufficient decision-making and general quality improvement of public services (Androutsopoulou et al., 2019; Bullock, 2019; Margetts & Dorobantu, 2019), critics have voiced concerns about the relationship between reliance on 'objective' proponents of AI in public administration and the apolitical narratives surrounding digital transformations (Dencik & Kaun, 2020; Yeung, 2023). On a bigger scale, scholars of administrative law problematize the use of AI as an intermediary for citizen rights, since it implies an automated application of law that might exclude external circumstances and context in administrative cases, obstructing citizens from claiming their rights (Ranchordás, 2022; Coglianese & Lehr, 2017). Furthermore, critics of the apolitical narrative surrounding AI in public administration argue that the automation of government works to depoliticise state-citizen relations by outsourcing decision-making to private tech-vendors, increasing market power in social relations (Cardullo & Kitchin, 2019; Collington, 2022; Wilson, 2022).

Disregarding the debate on whether AI in public administration is beneficial or not, the implementation highlights some tensions within decision-making. Although public values of democracy and civil rights are at the core of public administration (Haque, 1999) and a digitization of government services might have the potential to enable citizen participation by simplifying state-citizen communication (Duberry, 2022), knowledge of AI is often situated with industry experts construing the issue as complex and opaque, limiting civic engagement (Buhmann & Fieseler, 2022; Schiff, 2024). Whereas AI implementation in the public sector generates comparatively little engagement in the public debate (Wilson, 2022), tech companies are increasing their presence in the

political sphere. A recent report show that tech companies now represent the biggest lobby group within the EU (Bank et al., 2021), and their mandate in policy-making through representation by consultancy firms has been criticized to skew decision-making (Transparency International EU, 2021). As highly technical matters as AI require specific knowledge, industry experts and consultancy groups have salient claims in contemporary policy-making on the matter (Ulnicane et al., 2021). This illustrates a general trend beyond AI and tech, wherein 'objective' and 'factual' decision-making is increasingly valued and the harnessing of expert knowledge legitimizes policy-processes (Parkhurst, 2017). The question is not whether policies ought to be based on expertise and facts or not, but rather how facts are assessed and how the process of knowledge production in policy-matters play out. The prominence of private sector representation versus the inherent citizen centred values of public administration raises questions about involvement in the knowledge production process of policy-making – who gets a say?

In Sweden, a forefront country in public digital transformations, the state has made substantial investments in research institutes to support digitalisation efforts (Fagerberg & Hutschenreiter, 2020), alongside increased expenditure on commissioned research from private consultancy firms (Larsson & Teigland, 2019; van den Berg et al., 2019). Whereas the variety of research commissioned to support Swedish digitalisation policy is not necessarily significant in terms of uniqueness, the results of these policy-making processes have undergone critique (Ljungqvist & Sonesson, 2022; Blix & Jeansson, 2019), pointing to conflicts of private and public interests in digitalisation of the education system as well as in digital health care. In a report from 2022, media scholar Katti Björklund criticise the factual grounds for the Swedish government's digitalisation strategy for national health care. Pointing to an overreliance in the strategy on a report from consultancy firm McKinsey & Company, which the author argues lacks substantial facts and evidential ground, Björklund (2022) calls for more critical examinations of welfare-digitalisation policies.

In a response to that call, this inquiry will focus on the field of AI in public administration. As the Swedish government advances the mission "Promote the public administration's ability to use AI" (Regeringsbeslut I2021/01825), this dissertation aims to explore the knowledge production of AI in public administration in Sweden. By examining evaluations and considerations in policy texts, this can hopefully enhance an understanding of what knowledge and expertise was included in the policy

- and moreover, enable a discussion of represented actors and assessments in the policy and the implications of such. In short, what knowledge is presented and how?

LITERATURE REVIEW

The integration of artificial intelligence (AI) in public administration is a broad field of study, given that both AI technologies and public administration cover vast ranges of usages, applications and definitions. Hence, the following literature review will regard varying definitions and appliances in the field of AI in public administration to sum the most general implications and notable considerations. To narrow the focus, policy-making aspects will also be addressed. However, considering the scope of this thesis, detailed technological discussions and sector-specific case studies will not be included, recognising that some significant scholarly contributions might be omitted.

Public administration, AI, and public values

Originally a sub-branch of computer science, artificial intelligence (AI) has evolved into an interdisciplinary research field that includes a vast range of studies such as law, psychology, economics, political science, and philosophy (Uzun et al., 2022). Definitions vary among experts and between fields, but the term AI generally refers to digital systems automating or replicating intelligent behaviour, with an emphasis on rationality, autonomy and human-like thinking as the essence of AI (Russell & Norvig, 2021; Scherer, 2015; Dirican, 2015). Highlighting the autonomy aspect, self-improvement and adaptive learning are key characteristics of AI. The term is often used interchangeably with machine learning, denoting a process of analysing large datasets to identify useful patterns or make decisions (El Naqa & Murphy, 2015). Rapid advancements in innovation of AI technologies have accelerated the adoption of AI in government functions, further catalysing the expansion of studies of AI in public administration (Ahn & Chen, 2022). Specifically with the outbreak of COVID-19, a proliferation of studies on AI can be seen in public administration and policy studies (Önder & Uzun, 2021).

Traditionally, the view of public administration restricted its scope to the executive functions of government with a narrow focus on the organisation, personnel, practices and procedures related to performance of public services, whereas the modern view of the term encompasses all branches of government: executive, judicial and legislative. It deals with management and administration of

government policies and law, such as education, public health, social security, welfare, public works etc (Thapa, 2020). As a discipline, the study of public administration covers various dimensions of these practices, but can broadly be summed as studies of organisation, management, methods and procedures, public finance, and administrative accountability. The discipline is commonly referred to as public management studies, and although there are debates among scholars around the differences between public administration studies and public management studies, the terms are often used interchangeably (Pollitt, 2016).

A notable theme in public administration literature is the concept of 'New Public Management' (NPM). The term refers to the transfer of private-sector managerial techniques to public services, along with reforms introducing outsourcing and competition for public services and goods, and NPM is commonly related to neoliberal ideas – highlighting the focus on economic growth (Brown, 2015). Studies of NPM emphasise efficiency and effectiveness as core values in contemporary public administration, along with increased use of information and technology, as well as performance measurement (Gruening, 2001; Buschor, 1994). Critics of NPM argue that the marketisation of public interest erode core responsibilities of the state, in favour of economic growth and the market (Elcock, 2012). Along with the incentives of efficiency and economic growth, cooperative institutional arrangements between private actors and the public sector have increased, denoted in literature as 'Public-Private Partnerships' (PPP) (Wang et al., 2017), and critique on the matter are commonly centred around how PPP relate to citizens interests and democratic values (Bertelli et al., 2020; Delmon, 2011; Hodge & Greve, 2017). In a response to NPM and PPP studies, 'public values' have received increased scholarly attention. Public value literature expands on the notions of public-private partnerships and the transfer of private-sector ideals and techniques into public services, but goes beyond PPP and NPM studies to explore foundational principles of contemporary public administration. The public value approach responds to critique of NPM by re-emphasising value-related concerns within public administration and highlight non-state actors and citizen's roles in public administration, while maintaining a view of the government as ultimate guarantor of public values (Alford & Hughes, 2008; Bozeman, 2007; Bryson et al., 2014; Denhardt & Denhardt, 2000).

Conceptually, a 'public value' can be assessed as something either valued by the public or good for the public, or both. Hence, there is not a distinct definition of what such values are, but as a theoretical viewpoint it has both served as a practitioner approach with emphasis on how public managers can

create public value and as a reference point for policy-inclined research (Wallmeier et al., 2019). As a critical viewpoint beyond neoliberalism, it has been criticized to reproduce what it criticizes (Dahl & Soss, 2014) and for its weak theoretical ground in assessing control and power outcomes in the public sphere; as Morrell (2009) argues, there is no "how to" (public values) perspective that can answer questions of outcome related to power and control without more ethical and fundamental perspectives. Nevertheless, it stands as a major contribution to contemporary theory on public administration, specifically within research on AI in public administration.

Measuring benefits

In relation to this, AI has been studied with regards to public values and the potential for AI to improve public services and foster citizen-centred value creation (Van Noordt & Tangi, 2023). The assumed benefits generally include - but are not limited to - increased efficiency in service delivery, in example through access to real-time answers or assistance in welfare-applications (Mehr, 2017); support of both civil servant decision-making as well as government policy with data analysis (Pencheva et al., 2020; Veale & Brass, 2019); automation of routine tasks (Ranerup & Henriksen, 2022); and personalisation of service (Van Noordt & Misuraca, 2020). However, there are comparatively few empirical assessments confirming the potential benefits (Van Noordt & Tangi, 2023) and the views on benefits of AI in public administration diverge. Although some studies emphasise increased efficiency (Wirtz & Müller, 2018; Wirtz et al., 2018), others highlight the complications of measuring generalised improvement due to varieties in both AI tools as well as government functions wherein they are applied (Krejnus et al., 2023; Gębczyńska & Brajer-Marczak, 2020), as well as the ambiguous and qualitative nature of public values as 'citizen satisfaction' or 'enhanced decision-making' (Filguieras, 2022). Furthermore, even quantifiable benefits in measures of cost-reduction or revenue are complex and difficult to assess (Sidorenko et al., 2019; He, 2019). A review of the economic benefits of AI reveal a tension between improved efficiency for governments (Chun, 2007, 2008; Zheng et al., 2018), economic growth through innovation (Gonzales, 2023) and the contrasting potential loss of jobs and digital divide (Aguilera & Ramos Barrera, 2016; Dwivedi et al., 2021). Moreover, efficiency and economic growth do not necessarily imply improvement of services for citizens (Dobroyulbova, 2021).

However, there are instances that highlight the possibilities of AI in public value creation for citizens and governments. In New South Wales, Australia, AI is utilised by the government's revenue office to identify vulnerable customers to provide other settlement choices in cases when customers cannot

pay penalties. Assessing vulnerability based off indicators such as the number of former major penalties, expected socioeconomic position, and frequency of contact with the revenue office, the programme aids government workers in haltering enforcement and predicting appropriate payment plans to protect economically disadvantaged citizens, which furthermore have proved to increase overall efficiency in investigations (Alhosani & Alhashmi, 2024). In an article by Brandão et al. (2024), AI is proposed as an approach to analyse public bidding documents, to ease citizen insight to public decision-making and expenditures. Results in experiments with the program also shows work reduction in specialists' detection of irregularities in public bidding.

Risks

There is a broad literature discussing potential risks of AI in public administration, and common themes include widening societal divides, mass surveillance, harm to vulnerable groups, and lack of accountability in decision-making (Medaglia et al., 2023; Madan & Ashok, 2023; Monarcha-Matlak, 2021; Dencik & Kaun, 2020). Moreover, some critics argue that the implementation of AI in public administration results in the favouring of NPM techniques, and furthermore neoliberal ideals (Dencik & Kaun, 2020; Yeung, 2023), since the technology deployed implies a favouring of quantitative metrics and 'efficiency' by which qualitative values and processes are subordinated.

With regards to creation of public values, scholars point to the differing interests and goals between the private and public sector (Fatima et al., 2022). As AI poses a need for technical expertise, its adoption in the public sector requires collaboration with private actors (Van Noordt & Tangi, 2023), since governments often lack resources to develop systems in-house and have barriers to keep up with the innovative pace of the private sector (Madan & Ashok, 2023). The policy-making process must then safeguard against external pressures from interest groups potentially constraining citizen-centred policies in favour of profit incentives (Yeung, 2023). In an interview-based study with Canadian government officials, consultants were shown to have a significant role in generating institutional pressures in favour of AI in public administration, and the authors stress the importance of understanding the sense-making among government officials in relation to different narratives surrounding AI (Madan & Ashok, 2024). As AI is more than merely technological systems but rather embedded in socio-technical contexts and moreover risk ridden, there is a need for inclusivity and transparency in the decision-making process regarding the adoption of AI systems, to enhance

citizen's perspectives in governing and allow insight to the matter (Sigfrids et al., 2022). Insofar, the literature on AI in public administration have revealed tensions in the implementation or diffusion process; between private and public goal-setting, between resources in both spheres and between prescribed benefits and empirical validation. This will now be discussed in relation to an earlier stage, with a perspective on the policy-making process.

Policy formulation and the use of evidence

Policy formulation refers to the activity of identifying, devising, and defining problems and solutions, occuring once an issue has been acknowledged as deserving government attention (Howlett & Mukherjee, 2017). With regards to process, Thomas (2001) outlines four notable aspects that are commonly apparent: appraisal, dialogue, formulation, and consolidation. Appraisal refers to the data collection stage, wherein research, expertise and stakeholder or input from the general public is sought. The dialogue stage denotes the interaction between actors engaged in policy formulation that ensues, including the exchange and deliberation of perspectives on policy goals and means to resolve them. This can take the form of more or less formal proceedings and can involve representatives from industry, labour organisations, consultancies or other interest groups. Following that, the formulation stage represents the actual 'work', wherein administrators and public officials consider costs, benefits, challenges and opportunities of various alternatives with the aim of formulating a proposal, which is then often followed by the consolidation phase where feedback from different stakeholders about proposals are considered.

These elements of the process all relate to knowledge (Howlett & Mukherjee, 2017), and the relation between knowledge and policy formulation has been covered within the field of policy studies with various perspectives. A vast body of research emphasise the trend of 'evidence-based policymaking' (EBP), which relates to the call for expert knowledge and evidentiary validation as a mean for legitimacy in policy (Parkhurst, 2017; Oliver et al., 2014). By aligning policy with empirical evidence, complex issues are argued to be addressed more sufficiently (Richards, 2017). While that is a fundamental aim of EBP, scholars point to the issues of realising it in practice (Oliver et al., 2014; Sanderson, 2002; Choi et al., 2005). Studies on EBP show that governments often seek simple facts in support of already established policy goals, disregarding critical science and sometimes even constructing facts in accordance to validate certain policy options (Smith & Leech, 2010; Wye, et al.

2015; Torriti, 2010). Whereas forementioned studies critically assess whether EBP is feasible in practice or not, the model has received critique as an aim overall. In a discourse analysis on EBP in school policing, Nolan (2015) argues that quests for evidence in policy privilege ideologically driven research that conforms to hegemony, assuming an objective scientific stance while excluding critical research. Central to Nolan's (2015) argument is the relation between 'neoliberal policy-makers' and exclusion of critical studies opposing the position of such.

Other studies on EBP have focused on different actors in the process and how 'policy networks' work (Smith & Joyce, 2012), an area where policy-consultants have received increased attention (Howlett & Migone, 2013). Commercial consultancy groups are increasingly commissioned by governments to aid policy formulation, a trend that is consistent with new public management's (NPM) approach to increase efficiency in public administration (Marciano, 2022). Moreover, consultancy aid is more prominent in policy formulation in countries that show extensive NPM reforms, such as the Nordic countries (Seabrooke & Sending, 2022). For instance, van den Berg and co-authors (2019: 188) observed that consultant spending in Swedish government policy increased by a factor of 3.5 between 2003 and 2011. This has been problematised in relation to how the economic incentives of the private sector might shape social policy, especially with regards to the extent of the Swedish public sector (Jobér, 2023).

Literature on policy formulation in relation to AI in Sweden is small, but a policy analysis by Toll et al. (2019) explores the discourse on AI in a Swedish policy initiative led by Vinnova, a government funded research institute. The initiative included the government, Vinnova, two interest groups for regions and municipalities, one private research institute and a commercial consultancy firm. The article contends that values related to efficiency and professionalism were most prominent among the benefits, whilst benefits related to citizen engagement were absent, corresponding to NPM's sidelining of democratic ideals of civic engagement in favour of efficiency. In a critical analysis of Swedish digitalisation policy by Gidlund and Nyhlén (2022), the authors state that Swedish digitalisation policy is largely characterised by technological developments as an unquestionable imperative, with few critical assessments and economic growth as the main aim for digitalisation. Furthermore, the educational background of members on Swedish digitalisation-expert advisory groups has been studied, and findings show that there is a substantial lack of people with a background in humanities or social sciences related to welfare professions that digital

transformations regard, indicating a potential homogeneity in perspectives on what knowledge might be valuable, as the advisory boards consists mainly of engineers and business and economics experts (Gidlund & Sundberg, 2021). No studies were found on what the actual knowledge of AI is in Swedish digitalisation policies, or how assessments of that knowledge are done.

THEORETICAL FRAMEWORK

As the literature has shown thus far, both in the context of AI in public administration and policy formulation, there are concerns regarding the appraisal of evidence highlighted in the sections above by the complex scope of empirical assessments on benefits of AI in public administration, and by the challenges related to evidence-based policymaking. Another concern regards the private-public tension as illustrated by the differing goals of the private and public sector in collaborations on AI in public administration, as well as the role of commercial consultancy experts in public policy formulation. Moreover, the literature review reveals a potential research gap in the examination of the relation between the two tensions. To critically examine these implications, this thesis will adapt a constructivist perspective, viewing policy as knowledge production and applying an epistemic approach.

Policy formulation as knowledge production

In a constructivist attempt to make sense of the assessment of evidence and facts, and the different actors and stakeholders in the policy formulating process, Gunter (2017) suggests examining policy formulation as 'knowledge production'. The author explains knowledge production in relation to policy formulation as the identification, selection, accession, usage and legitimisation of knowledge, and emphasises the role of experts as knowledge actors (p.337). This epistemological perspective on policy as knowledge construction has further been stressed by other scholars, arguing that policy is a comprise of value-laden actions and constructions whereby some epistemological assessments of the social world succeed over others (Nyhlén & Gidlund, 2022). Gunter (2017) states that there is a misconception in that knowledge is seen as something transferred into policy, curated beforehand, which the author contrasts by arguing that knowledge for policies is constructed within the practice of policy formulation.

The process of policy formulating as knowledge production is explained through the '4Ks'; knowledges, denoting data, ideas, arguments and theories that shape the understanding of the policy issue; knowings, meaning the means/methods by which knowledges are made available, i.e., through belief statements, analysis from research or the use of specific language or practices; knowledgeabilities, relating to how the text is synthesized to exhibit preferred knowledges and knowings through arguments, postures and inter-relating themes; and knowers, the actors involved in constructing issues as knowable. In relation to knowledge production, the state is perceived as a 'researcher' as well as a 'producer' and 'user', all at the same time, and have mandate over what ontologies and epistemologies gets represented in the problem identification. Furthermore, expertise is explained as a matter of recognition through inclusion/exclusion (p.341-342), and Gunter emphasises the role of language and framing of issues in shaping what is normalised and perceived as common sense, as well as the relocation of political issues to the private sector as part of the shift to depoliticisation.

This is linked to the introduction of NPM reforms, which brought managerial techniques from the private sector into the public, increasing the stake for external experts and consultants in the public sphere. As Gunter states: 'It seems that NPM was brought into governments by consultants and NPM needed consultants to make it work' (p.346). Moreover, the author contends that the issue for research on knowledge production is not what is known but rather why certain knowledges, knowings and knowledgeabilities come to gain salience while others are repressed and why certain 'knowers' are preferred in policy work over others. Therefore, to examine the knowledge production of policy formulation in relation to inclusion/exclusion of knowledges, knowings, knowledgeabilities and knowers, the epistemic (in)justice framework can offer an explanation.

Epistemic (in)justice in knowledge production: contributory injustice

Theorised by Miranda Fricker (2007), epistemic (in)justice describes the capacity at which someone is recognised as a 'knower'. According to Fricker (2007), epistemic injustice can be divided into two main categories: testimonial injustice which occurs at an individual level when prejudice causes a hearer to discredit a speaker; and hermeneutical injustice, which signifies structural asymmetries in collective hermeneutical resources due to a marginalisation of certain social experiences, i.e., the lack of words to describe sexual harassment before the term was coined due to the marginalisation of

women's experiences. In critique of Fricker's assessment of hermeneutical injustice, Kristie Dotson (2012) introduces a third form: contributory injustice. Dotson (2012) argues that there is more than one set of 'collective hermeneutical resources', and that contributory injustice occurs when epistemic agents refuse to acknowledge certain hermeneutical resources, thus compromising the epistemic agency of certain knowers. Hermeneutical resources here refer to meanings or frames for interpretation and reference (p.30), with the ontological implication that such can differ as meanings or frames are constructed within various social contexts. Thus, there can be different knowledges available, but due to situated ignorance, certain perspectives are dismissed further reinforcing structural asymmetries. Dotson (2012) states: 'Situated ignorance follows from one's social position and/or epistemic location, which works to institute epistemic differences, while obscuring those same differences' (p.31).

Although the theory is an expansion and a critique of the works of Fricker (2007), it still rests on the same notion of marginalisation in relation to knowledge production. Fricker's (2007) theory departs from two logics underpinning the marginalisation: socio-economic power and identity, meaning that one might be excluded based on lack of material power as education or financial assets; or as a result of prejudicial stereotypes framing the knower as unsuitable. Moreover, Fricker (2007) notes that the marginalisation does not necessarily need to stem from either of these variables, but usually does. Dotson (2012) does not explicitly address what logics underpin prejudice in contributory injustice, but also discusses the term in relation to cases of marginalisation due to socio-economic power and identity. Hence, studies of epistemic injustice have commonly focused on the exclusion, subordination, and marginalisation of citizens in epistemic settings, generally centring marginalised groups and their credibility, involvement and access to knowledge production/consumption, i.e., patients in healthcare (Liabo et al., 2022) or indigenous people in climate policy (Widenhorn, 2013). Fewer studies have adopted the framework in relation to EBP and policy formulation (Mormina, 2022). To apply the framework to policy formulation, it will be integrated with the '4Ks' (Gunter 2017) to account for the different parts of knowledge production whereby subordination and exclusion could imply contributory injustice. Thus, knowledges, knowings and knowledgeabilities will be considered hermeneutical resources, and 'knowers' will be assessed on the basis of their recognition in the policy.

RESEARCH STATEMENT

This thesis aims to examine the policy formulating of the Swedish government's mission: "Promote the public administration's ability to use AI" (Regeringsbeslut I2021/01825), in relation to the knowledge production and whether the subordination or exclusion of certain hermeneutical resources or knowers might constitute contributory epistemic injustice. This will be explored through the following research questions:

RQ 1: What characterises the knowledge production of the policy formulation in "Promote the public administration's ability to use AI"?

RQ 2: Does the policy formulation exhibit contributory injustice, and if so – how?

As described in the literature review, the implications of introducing AI in public administration includes a broad range of assumed benefits and risks, and by examining the policy formulation of such a project, this thesis hopes to dissect the assessment of those to extend the discussion on the ideas and actors that guide knowledge production on AI in public administration. Hopefully, this will serve to unpack questions about what perspectives are preferred and moreover what values are reflected in the policy, specifically with concern to the responsibility of the government as a guarantor of citizen-centred services. By exploring the policy formulation of AI in public administration and what knowledge directs the aims, this study hopes to develop an understanding of policies as a result of inclusion and exclusion. Moreover, by integrating policy formulation as knowledge production with the theory of contributory epistemic injustice, this can potentially contribute to expand constructivist frameworks for analysis of AI policy, as no earlier studies on contributory injustice in relation to AI policy could be found.

METHODOLOGY

To explore knowledge production in policy formulation, a qualitative content analysis (QCA hereafter) was employed. QCA is a methodological approach to systematically describe and interpret qualitative material (Schreier, 2012). The method emerged from the quantitative content analysis tradition within media and communication studies, with a similar systematic and rule-based approach to data analysis (Puppis, 2019; Mayring, 2014). This is also the main difference from other qualitative approaches such as discourse analysis, as the process follows a standardised step-by-step

sequence. Although it shares this with quantitative content analysis, it also differs from the quantitative tradition by emphasising analysis in relation to context and enables analysis of latent meanings of the text (Puppis, 2019: 368-369). Thus, it addresses quantitative aims of developing systematic descriptions of meanings, as well as the qualitative inquiry to generate, through interpretation, the very meanings that will be analysed systematically. This implication of QCA is that it entails two dimensions of examination; phenomenological descriptions of manifest content and hermeneutic interpretations of the latent content (Graneheim et al., 2017), which allows for both a mapping of aspects as well as an interpretivist account for their meanings. The systemic approach of QCA departs from a codebook, whereby categories to arrange the analysis according to are established either inductively or deductively or both (Mayring 2014: 104), implying that it lends itself useful to studies with set aspects of inquiry within the theoretical framework (Puppis, 2019: 376). Since this study concerns specific aspects of knowledge production in policy formulation, the systematic analysis of those is an advantage that QCA allows.

The methodological foundation of QCA

Rooted in the quantitative tradition, while commonly applied for distinct qualitative means, QCA does not entail a set ontological stance. A positivist approach is possible for research emphasising reliability and 'objective' descriptions, whereas a hermeneutical approach is possible through an interpretative understanding of the text with more emphasis on intentions, context and background. Therefore, it is the responsibility of the researcher to set out the approach taken and declare epistemological implications of chosen stance and research design (Schreier et al., 2019). Considering this in relation to the research aim of deepening the understanding of the role of knowledge production in policy formulation through a specific case, by the means of dissecting the assessments made in the policy, this thesis will adapt a hermeneutical approach to emphasise the constructive element of interpretation. Thus, the epistemological basis for this study is – despite entailing descriptive elements – an interpretative mode of QCA that implies a co-construction of meaning between the researcher and the text (Graneheim et al., 2017), which furthermore assumes more than one meaning available for interpretation. Moreover, the acknowledgement of the position of the researcher as involved in the construction of said meanings implies a subjective assessment of and in the analysis (Yanow, 2007). Acknowledging this and hence distancing the research from claims of examining one 'truth', reliability will not be considered as the aim is not that of objective description.

Following this, the validity aspect of this study is to be ensured through a focus on trustworthiness and credibility, which simplified relates to comprehensiveness, transparency, and thoroughness of the research (Graneheim et al., 2017).

Categories and interpretation

Whether positivist or hermeneutical, QCA requires specification of which aspects of the material are studied, guided by the research question. As such, it does not allow for the holistic depth of qualitative analysis that discourse analysis does (Schreier, 2012). Regarding this, a critique of QCA is that it does not go beyond descriptions of manifest meanings (Groeben & Rustemeyer, 1994; Rössler, 2005), to which other scholars (Früh, 2007; Krippendorff, 2004) have responded by arguing that the method does allow interpretation in relation to context, beyond manifest text. Moreover, the specification of systematically categorising the material which disables a holistic view is also one of the main benefits, specifically to policy-analysis. By focusing on certain aspects of the texts, the material for analysis is reduced (Schreier, 2012), which makes it compatible with policy documents, as they can encompass a vast range of aspects and tend to be long (Puppis, 2019). Critical Discourse Analysis, for example, delves deeply into the whole text, accounting for detailed aspects of the material (Janks, 1997).

Since the object of study for this dissertation was policy-documents amounting to over 130 pages, QCA is suitable in that it allows covering the full material while also reducing it by a selection of specific aspects. Another advantage of the method is the combination of descriptive and interpretative approaches, suitable for the research aims since RQ1 is more of a descriptive nature, whilst RQ2 is more of interpretative nature. The categorical focus of QCA is another point of critique, with some scholars (Rosenthal, 2015; Lueger, 2010) arguing that the reduction of material into specific aspects reflects a quantitative orientation ultimately incompatible with interpretative research, since it restricts the scope for interpretation. To address this limitation, Kuckartz (2014) suggests combining QCA with an empirical construction of categories and a thematic focus in analysis. Schreier (2012) similarly advocates for the inductive development of categories, emphasising the need for interpretative approaches in guiding the selection of aspects and exploring the relationships between categories to enrich interpretative depth. Therefore, to ensure credibility in that the analysis can explore the relational dimensions of the material and address potential restrictions of a deductively

developed coding frame, the coding frame was continuously refined through iterative trials and readings of the material, allowing for inductive adjustments as needed.

Research design

Being a rule-based method, QCA entails a set procedure of steps. Although outlines vary between scholars (for example, see Elo & Kyngäs, 2008; Mayring, 2014), the main parts of the process regards the following: 1. deciding a research question, 2. selecting material, 3. building a code frame, 4. trial of code frame, 5. evaluation and potential adjustment of code frame, 6. main analysis, and last, 7. interpretation and presentation of findings (Schreier, 2012; Mayring, 2014; Krippendorff, 2004). The following sections will address how steps 2-5 were followed to construct a research design for this study.

Material

Step 2: The material selected to examine the policy formulation of AI in public administration in Sweden are the two documents "Interim report: Promote the public administration's ability to use AI" (DIGG, 2020) and "Final report: Promote the public administration's ability to use AI" (DIGG, 2023), both made publicly available by the Swedish government. The documents constitute the full report of the formal investigation that was assigned by the Swedish government to the agency for digital government (DIGG) in 2021 (Regeringsbeslut I2021/01825), and amount to 44 and 119 pages, respectively. The interim report, initially commissioned for another assignment, became the basis for the subsequent investigation, hence why it was published before the formal investigation was assigned (Regeringsbeslut I2021/01825). Thereby, they do not represent the full extent of policy proposals of AI in public administration to be implemented, since the government recently assigned a new committee to continue the work (Dir 2023:164). However, as the mission represents the most recent comprehensive investigation of AI in Swedish public administration and the findings are cited by the government in other official documents (Regeringsbeslut Fi2024/01535; Dir 2023:164), the reports can be considered highly impactful, thereby enhancing the credibility of the research. This credibility is supported by relevancy and richness of material, as highlighted by Sandelowski (1995).

Coding Frame

Step 3: The initial coding frame was inspired by Puppis' (2019) and Mayring's (2014) recommendations for structure and developed deductively, departing from Gunter's (2017) '4Ks' of knowledge production as categories. The initial coding frame (app.a) entailed the original aspects of knowledge production as categories.

Step 4: A pilot-study was conducted using coding software MAXQDA, following recommendation by QCA scholars (Mayring, 2014; Schreier, 2012). The analysis did not utilise any of the software's AI tools; it was solely used to code the documents manually and print the markings in documents arranged by the program. However, after the first trial, the codes were adjusted with amendments in the form of subcategories to three of the four categories. These were the four categories, including respective subcategories:
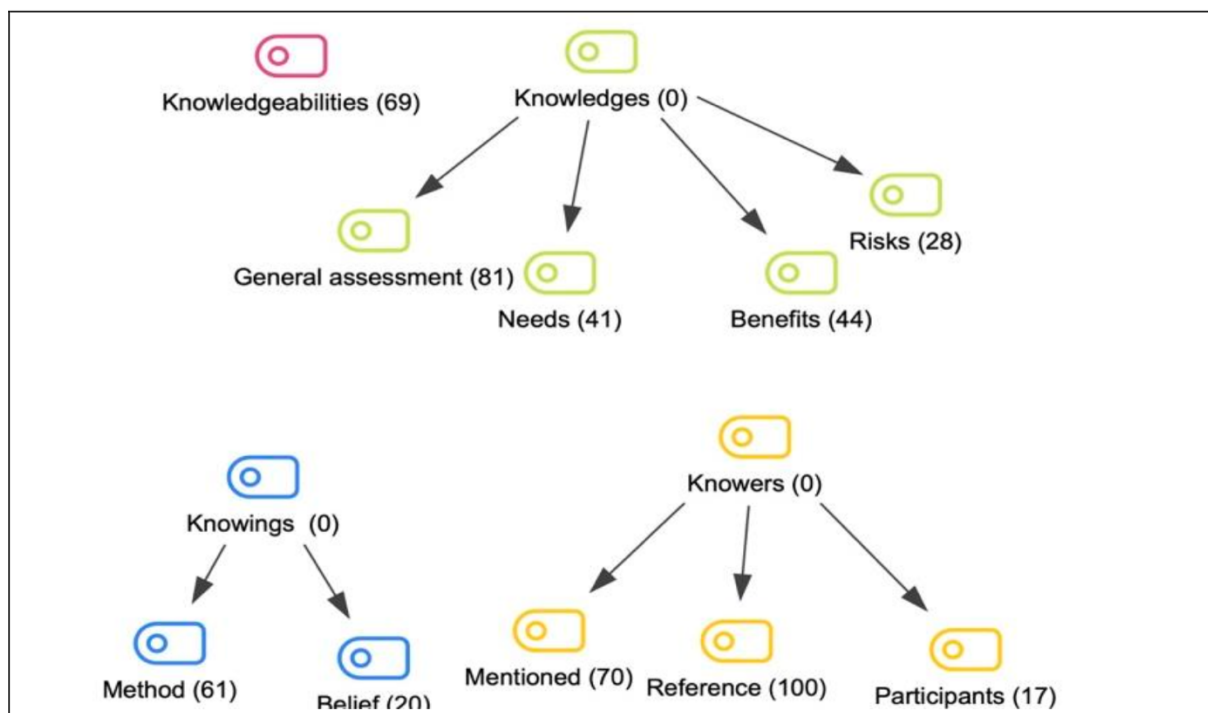

Figure 1: Outcome from trial of initial coding frame (source: MAXQDA)

Step 5: Instead of coding according to the '4Ks', the **final** coding frame (app.b) departs from the initial subcategories 'risks' and 'benefits' as main categories, to enable a mapping of the relation between the '4Ks' within a category, as the pilot-study proved difficulties in assessing interrelating themes. Moreover, the 'risks' and 'benefits' categories align with key elements of policy formulation (Thomas, 2001), making them a suitable starting point with respect to the research aim of examining knowledge

production in policy formulation, which furthermore corresponds with QCA being guided by the research question regarding which aspects to analyse (Schreier, 2012). Thus, the design of the coding frame was not limited to a deductive development, as the subcategories for 'knowledges' – which were later adjusted to be the main frame - relates to specific themes in the texts inductively assessed through interpretation (Elo & Kyngäs, 2008). This instead displays an **abductive** approach whereby the coding frame was construed, which implies a combined approach of inductive and deductive elements in category descriptions (Graneheim et al., 2017). Below is an extract from the coding-frame (Figure 2).

Category	Definition	Anchor examples	Coding rules
Benefits	Statements describing benefits related to AI in public administration	*"Artificial intelligence (AI) is an area with great potential to improve society."*	- Explicit statements regarding positive effects of having AI in public administration - Arguments/reasoning that support explicitly negative statements

Subcategories:

Benefits

Economic benefits → Benefits that regard economic benefits of implementing AI in public administration, i.e., *The total economic value of a full introduction of current AI technology in Swedish public administration is estimated to amount to approximately SEK 140 billion*

For citizens → Statements that directly address citizens as beneficiaries, i.e., *Swedish citizens and companies can benefit greatly from AI*

Figure 2: Coding-frame example (appendix B)

The material is written in Swedish, which was not a challenge to the coding since it is my first language. However, to ensure clarity for the reader and thereby enhance the analysis' validity in terms of trustworthiness (Graneheim et al., 2017), units included as examples in the analysis as well as all other units in the appendix were translated into English. Once coded, the units from the texts were compiled into Word-documents using MAXQDA's printing tool. Thereafter, they were translated with Microsoft Word's translation function. Moreover, the appendix includes page numbers and source for each unit, providing insight into the specific sections that have been translated. With that said, the translation is a result of both the software used as well as the assessment made by the researcher, which might affect how some wordings come across as well as how certain interpretations were made – since they are not presented in the original language of the texts, which speaks for the subjective character of the study.

ANALYSIS

The following analysis will be arranged according to the mapping of benefits and risks, in relation to knowledge, knowings, knowledgeabilities, followed by a mapping of knowers corresponding with the categories of benefits and risks. The interim report consists of eight main sections (excluding reference list and appendix): introduction; potential benefits of AI in public administration; external analysis; mapping and analysis of AI within public administration; Sweden in an international perspective; summary of needs; proposals and measures for increasing our AI capability; consequences. The final report consists of seven main sections (excluding reference list and appendix): introduction; architecture & infrastructure; model of trust; the AI-guide; providing information regarding relevant AI-projects; conclusions and recommendations. Both documents were coded, and the following analysis of the material consists of units from both documents compiled.

Benefits

The following table (for coding-frame, see app.b) presents the identified benefits of implementing AI in public administration:

Table 1. Benefits and subcategories

Subcategory:	Economic	General	For public administration	For citizens
Occurences:	19	17	8	1

The analysis will focus on the categories 'Economic' and 'For citizens', as the literature review underscores the tension between economic considerations and citizen-centred perspectives of public values. Additionally, these categories represent the smallest and largest groups, respectively. To ensure comprehensiveness, the other two categories will also be reviewed in relation to their meaning for the 'Economic' and 'For citizens' categories.

Economic Benefits

The most cited benefit of AI in public administration is economic benefits. The section 'potential benefits of AI in public administration' in the interim report concludes that the implementation of AI in Swedish public administration will imply cost-reductions, increases in productivity and revenues

that together will generate 140 billion (SEK) in yearly profits. For context, this amounts to 6% of the Swedish governments annual budget (app.c).

> *The estimated benefit of AI in public administration amounts to 140 billion annually (app.c)*

As for *knowings* related to the statement above, the *knowledge* was generated through a 'benefit-analysis' in collaboration with McKinsey & Company, a commercial consultancy firm.

> *In collaboration with McKinsey, a model for benefit analyses was developed which then formed the basis for overall assessments of the benefits in different subsectors. (app.f)*

The analysis departs from predictive calculations on seven possible implementation areas: cross-sectorial; social security and labour market; education and culture; health care; general administration; infrastructure; public protection (app.c). Cost-reduction, increase in productivity and revenue were assessed individually with respect to each sector, i.e., implementing AI in healthcare is said to generate 30 billion.

Considering the massive economic benefit as a *knowledge* (Gunter, 2017), it can be seen as argumentative. Although there are data being referenced, from the benefit-analysis, the analysis raises some questions. Firstly, just three of seven sectors are accounted for in examples of which variables were used in the section of the appendix describing the methodological procedure (app.f). 'Education and culture', for example, is said to generate 19 billion SEK, but the analysis fails to adequately display how. The following quote is the only description of what the potential benefit was derived from:

> *In education and culture, the potential of existing AI applications primarily lies in the ability to provide personalised education and follow-up, benefiting individual students and giving teachers better tools for their educational efforts. It also aims to improve the overall quality of education, including reducing the risk that student's needs go unnoticed. The economic benefits of better education are substantial for society as a whole. The direct productivity gains are estimated to be somewhat smaller. (app. c)*

Second, the examples of variables for the analysis calculated for social security and labour market display a highly simplified assessment. The benefit of 9bn SEK is estimated based on the prediction that AI tools could reduce costs of social benefits and increase revenue, through an increase in employment of 19,650 people by improving relevancy in job-matching.

> *In simple terms, the value of their additional production is calculated to correspond to the salary they receive – in this case the Swedish average annual salary of SEK 415,200 (2018) – whereby the value that accrues to the employees (and in practice also the employers through the profit share in the additional production) is calculated to be SEK 19,650 x SEK 415,200 = SEK 8.1 billion. 266 The total potential for benefits, which accrues to both the state and other actors, would thus be 1.4 + 8.1 = SEK 9.5 billion. (app.f)*

This implies that every person would acquire a job of average salary, not considering the average salary for a person who gets a job with the help of government services. Although it is outside the scope of this essay to consider economic variables in terms of validity, the simplification of the analysis speaks for an argumentative character whereby *knowledges* are constructed by *knowings* of belief (Gunter, 2017), rather than critical examination.

Third, in terms of *knowledgeabilities*, the 140bn SEK is not estimated with respect to costs. Although required investments are acknowledged in the report, there are no predictions of what these might be in terms of numbers and the conclusion is that they will generate long term benefits (app.f).

> *The cost of each organisation's implementation is not subject to the report's calculations as this assessment is not possible to make within the framework of the assignment. (app.f)*

Although costs of implementation are not considered, costs of *not* implementing AI in public administration are. The final report displays estimates in numbers as well as examples of how this was assessed (app.d), which works to strengthen the *knowledge* of economic benefits of AI, furthermore displaying the *knowledgeability* of economic impact in favour of proposed solution. As *knowledgeabilities* regard preferred *knowledges* and *knowings* (Gunter, 2017), the exclusion of knowledge about costs can be seen as a preference for economic benefits, rather than overall economic impact. This is furthermore evident in the use of references to strengthen the validity of the economic benefit assessment:

> *The results are in line with previous studies of AI's economic potential, or calculations of*
> *digitalization's potential for the economy and the public sector. (app.f)*

Here, the 'previous studies' in question is a reference to one study by consultancy firm PwC, on the value of AI for businesses (app.e). It is not 'studies', nor is it a study on the public sector. This testifies to a preference for a construction of knowledge in alignment with wanted results, as well as the preference for justification over examination, as no references to other sources were included in relation to potential costs or general economic impact.

Citizen-centred benefits

There is only one explicit mention of citizens in relation to benefits in the reports.

> *Through collaborative efforts and distinct investments in AI, Swedish citizens and*
> *companies can benefit greatly from AI. (app.c)*

There is no mention in what way Swedish citizen's can benefit greatly from AI, therefore this knowledge can be assumed to have been gathered from other, general knowledges about AI benefits mentioned in the reports. With reference to the interim report, the final report makes the following conclusion:

> *"In addition, it is estimated that there are indirect economic values that have not been*
> *quantified, as well as additional significant social and qualitative values in the form of better*
> *quality of life, justice, the environment, etc." (app.c)*

The interim report includes the same quote (app.c), but there is no justification or evidence to support the suggestion that AI would improve general life quality, justice, or the environment. Thus, it can be assumed that this knowledge is based on an overall assessment of benefits in the report. In the search for the assessment of benefits that could indicate support for the statement, every benefit identified in the interim report (app.c) will be considered.

First, the reports mentions that AI can improve balance between work and leisure by automating routine tasks (app.c). That *knowledge* is supported by a reference to a report by OECD (app.e), which can be seen as a method of *knowing* (Gunter, 2017). Second, the section 'external analysis' lists

demographical challenges that *"the use of technological tools can help countries deal with societal challenges that these macrotrends cause or are symptoms of"* (app.c), regarding societal challenges as climate change, technological development, increased urbanisation, shifts in global economic power and demographical changes. The method of *knowing* whereby this *knowledge* is acquired is through a referenced report from consultancy firm PwC (app.e). A quick assessment of the PwC report shows that this *knowledge* exhibits a belief-induced way of *knowing*. The PwC report is neither an empirical study, nor does it mention AI or give any examples to how AI can counteract climate change (app.e). The estimation in both the final and interim reports that AI can improve the environment is based on a belief rather than empirical grounds. That is not to say that it is false, but that the *knowledgeability* displayed is inclined towards more simplistic assessments.

Finally, the knowledge that AI can enhance justice is open to various interpretations, depending on concepts of justice. Regarding justice in legal terms, the interim report includes a section in the appendix called 'legal aspects', which covers legal challenges of AI in public administration, not possibilities for AI to enhance legal processes. With respect to just administrative processes, the report states that *"processes and decision-making can become more objective"* (app.c), and: *"For example, AI has the potential to increase the equivalence of assessments so that equal situations are assessed equally"* (app.c) Neither of these assessments are based on any further reasoning, referencing, or arguments. The first quote does reference the assessment made in section two of the report, 'potential benefits of AI in public administration', but that section does not cover qualitative benefits of AI in processes outside the economic benefits. It does include a statement saying that AI can improve the overall judicial-chain of processes and decisions (app.c), but the *knowing* is based on the implicit assumption of 'improvement' from the intergral framework used in the economic benefit-analysis. As such, there is no external framework for estimating qualitative variables applied, they are rather implicit in the framework provided by McKinsey. Thus, the *knowledge* that AI can work to enhance the quality of life, justice, and the environment, is constructed through the *knowing* of referencing the OECD report on work-life balance and belief.

Risks

In the coding of the category, risks were identified either with mention to challenges with AI in public administration or risks of AI in public administration. Then, they were categorised based off whether they regarded risk of not implementing AI in public administration, risk of implementing AI in public

administration and challenge to implement. The following table (for coding-frame, see app.b) represents the results in the risk category:

Table 2. Risks

Category of risk:	Risk of implementing AI in public administration	Risk of *not* implementing AI in public administration	Challenge to implement AI in public administration
Occurences:	14	6	3
Subcategories:	Ethical risks	Security risks Economic risks	Manegerial Judicial constraints

Comprising the biggest category of risk assessment, only the ethical category will be reviewed. This is due to the aim of examining the knowledge production around AI in public administration, and not public administrations general processes of implementation, which many of the other categories regard. Second, the 'ethical risks' category was the only one that regarded *risks of implementing* AI in public administration, as the others mainly concerned risks of *not* implementing AI or possible challenges *to* implement AI. Since benefits were identified on the premises of 'benefits of implementing AI in public administration', the same will go for risks.

Ethical risks

The interim report features an appendix section titled 'Ethical Aspects', which outlines three primary areas of risk: *transparency*, *bias* and *responsibility*. The introduction to the section acknowledges that additional risk areas exist, but fall outside the scope for the report and that the three selected risks are based on the challenges *"discussed within the framework of the survey"* (app.f). This illustrates how knowledgeability arranges knowledge according to preference (Gunter, 2017), as the 'framework of the survey' indicates a limited perspective on risks, as those would mainly be examined in accordance with prior knowledges and knowings assessed in the report. For example, the main knowers consulted in interviews and international visits are public administration personnel, or industry experts (app.e). No citizens or 'receivers' of AI services were heard, which limits the perception of potential risks to the assessment of decision-makers and public servants.

Transparency is examined in relation to technological complexity and how advanced technologies can obscure the decision-making process. It is also addressed in other areas, such as the need for

insight into 'training data' and the importance of transparency in AI-human interactions, emphasising the necessity of disclosing AI usage in administrative processes (app.d).

> *An AI application can be based on algorithms that are very advanced, and it can be difficult to determine why an application arrives at the result it does in the individual case. The phenomenon is often referred to as the "black box". (app.d)*

Furthermore, the transparency assessment centers citizen's and public administration personnel by addressing potential conflicts regarding state-citizen relations and trust.

> *Transparency is an important building block for building trust and a lack of transparency can damage citizens' trust in public administration. (app.f)*

Bias is discussed in terms of pre-existing biases in training data and how data can reflect structural inequalities in society with the risk of reinforcing them (app.d). Furthermore, the report highlights the human factor in programming algorithms, noting that implicit opinions or belief of programmers can might introduce bias (app.d).

> *Data represents society, but society itself contains imbalances. AI solutions then risk reproducing or even reinforcing the problematic aspects of society, without that necessarily being the intention. This raises a normative issue that thus needs attention, where the AI system becomes a "mirror" of existing but undesirable structures. (app.d)*

> *Bias and biases may be due to the fact that the data used to train the algorithms contains biases and biases, which thus leads to the AI solution "inheriting" these. This, in turn, can lead to discrimination against groups or people, for example. (app.d)*

As illustrated by the quotes above, the bias section follows the theme of addressing citizens (although indirectly here) by highlighting social issues related to potential harm of AI in public administration.

Lastly, **responsibility** is addressed in relation to accountability in incorrect assessments.

But if an AI application gives incorrect answers, who is responsible? The issue can be more complex than in the development of traditional IT systems, as there may be more aspects to consider. Could the deficiency be based on training data that has been used, for example? (app.d)

As illustrated in the sections above, public administration workers and citizens are centred in the assessment of risks, which is further displayed with regards to responsibility of workers in the quote below:

Situations can arise where individual employees are put in situations where they have to take responsibility for making mistakes linked to an AI application or instead blame the AI solution itself. (app.d)

In terms of knowledge production, the risk assessment exhibits some notable aspects. First, considering the quotes in the analysis above as *knowledges*, they illustrate a referencing mode of *knowing*, as they are all legitimised through references to and examples from scientific articles (app.e). The *knowledges* of ethical risks can moreover be caractherised as informative as opposed to argumentative, as the section consists of statements about risks identified in various research articles, which is not followed by any reasoning promoting specific proposals. Rather, the conclusion of the section argues that it is necessary to 'deal' with these risks without mentioning specific solutions. Public administration must be prepared for errors to be identified in this way, be able to learn from it and develop operations so that identified errors are reduced.

If public administration does not deal with identified errors in a satisfactory manner, trust risks being diminished.(app.f)

In order for public administration to be able to make use of the potential, it is therefore necessary to be able to handle the ethical aspects. (app.f)

Another key aspect is the direct addressing of citizens. As discussed above, citizens and public administration workers are perceived as the main recipients of potential risks, and the *knowledges*

presented in relation to this deals with qualitative values as 'trust' or 'equality' (app.d,f) rather than quantitative numbers or more specific case studies.

Based on the risk assessment in the interim report, the final report suggests a voluntary 'trust model' for public administrations implementing AI (app.f). The knowledge production of the 'trust model' includes two parts: the interim risk assessment and the development of the model. In the development of the model, several *knowers* participated. The report describes internal trials of the model on applied AI-systems, as well as trials alongside external actors who participated in the development (app.f,). The specific participants are not explicitly referenced, but contributing organisations in the development are listed in the appendix (app.e). The contributors listed are four government agencies, three universities and one municipal body (app.e). As participant *knowers*, they can be said to have a public profile, based on the notion that they represent public institutions rather than the private sector. Furthermore, the 'trust model' was also developed through national and international dialogues, with the aim of considering UNESCO's recommendations on ethical AI (app.f).

These dialogues and trials can be seen as ways of *knowing*, whereby the construction of *knowledge* comprises of multiple inputs from different *knowers*, which testifies to a diverse scope of *knowledgeability.* Furthermore, the emphasis on UNESCO's recommendations can be seen as a framework of *knowing*, as it guided the dialogues on the trust model (app.f). This, alongside the emphasis on the European Commission's high level expert group on artificial intelligence (AI HLEG) and their ethical guidelines for 'Trustworthy AI', which is referenced throughout the interim report (app.f), displays an institutional mode of *knowledgeability*, whereby knowledge is guided and legitimised through coherence with external institutionalised frameworks.

Knowers of benefits and risks

To examine the knowers of benefits and risks, all texts referenced to support the assessment of the analysed benefits and risks were assembled (see app.f). The authors, in terms of organisation wherein the text was produced, will be considered 'knowers' regarding their capacity as " knowledge workers who […] construct the world as knowable" (Gunter, 2017: 338). Moreover, they were categorised as 'academic' on the premise of being published in an academic journal. Then, the actors mentioned in the reports in relation to statements supporting the assessment of benefits have been identified. Lastly,

the category participants include explicitly stated actors contributing to the assessment. The findings of each category are presented in the table below (table 3):

Table 3. Knowers of benefits and risks

Knowers of:	Benefits	Risks
Sources referenced:	Consultancy firm reports: 7 Intergovernmental reports: 1 Governmental agency reports: 4	Academic articles: 10 EU reports: 2 Google report: 1
Mentioned in text:	PwC McKinsey & Company Gartner Socialstyrelsen (national board of health and welfare)	UNESCO AI HLEG OECD Diskrimineringsombudsmannen (non-discrimination ombudsman) Jämställdhetsmyndigheten (national agency for equality) Nick Bostrom
Participants:	DIGG McKinsey & Company RISE Tillväxtanalys (national agency for growth analysis) Digitaliseringsrådets kansli (national board for digitalisation) Statistiska Centralbyrån (national agency for official statistics) Policy advisory board (see app.e)	DIGG Diskrimineringsombudsmannen (ibid.) Jämställdhetsmyndigheten (ibid.) Intergritetsskyddsmyndigheten (national authority for privacy protection) Region Stockholm Statskontoret (national agency for public management) Gothenburg University Södertörn University Linköping University Lund University

This shows that the *knowers* of benefits in the report are either consultancy firms or Swedish state-actors, whereas the *knowers* of risks are of a seemingly more diverse range represented by universities, intergovernmental organisations and Swedish state-actors. However, the reports include other referenced articles and a working-group to support the policy formulation, and those could have contributed to the benefit/risk assessment without being mentioned. Thus, the mapping only includes explicitly stated knowers.

DISCUSSION

The purpose of this study has been to examine the knowledge production of benefits and risks in the policy formulation of the Swedish Government's assignment "Promote the public administration's ability to use AI". Through a qualitative content analysis of two policy documents, this thesis aims to create an understanding of the knowledge production, in terms of inclusion and exclusion, which will now be discussed in relation to the research questions.

RQ 1: What characterises the knowledge production of the policy formulation in "Promote the public administration's ability to use AI"?

In the benefit assessment, argumentative knowledge about economic benefits in terms of specific numbers were centred, and knowings as estimations and calculations were made. Furthermore, citizen-centred benefits were substantially fewer and mainly supported by knowings of belief. The benefit assessment did not include academic knowers in terms of references nor participants. While the knowledge production did include the usage of external sources outside the internally assessed estimations and calculations, to validate the outcome of those, the external sources consisted of mainly consultancy reports without empirical relation to the economic benefit estimation. **As such, the benefit assessment can be characterised by belief-induced means of knowing, argumentative economy-centred knowledges and a mix of state actors and commercial consultancy firms as knowers**. This relates to the earlier studies on Swedish digitalisation policy (Nyhlén & Gidlund, 2022; Gidlund & Sundberg, 2021) that describes such as economic growth-centred, as well as the trend of consultancy aid in policy formulation in the Nordics (Seabrooke & Sending, 2022).

This characterisation is not to devalue economic benefits - or say that there is no evidential assessment of such benefits (see Chun, 2007, 2008; Zheng et al., 2018), but it does confer to abovementioned discussion on the complexity of empirically validating economic benefits of AI in public administration (see Sidorenko et al., 2019; He, 2019). Moreover, as discussed in the analysis above, the *knowledgeability* of assessing benefits was limited to quantitative economic variables in the framework provided by McKinsey whereby qualitative social benefits were merely implied. This highlights the efficiency and growth-oriented aims of consultants (Marciano, 2022), which at many times is at conflict with the socially oriented main aims of the public sector (Jobér, 2023).

In the risk assessment on the other hand, citizen's and social issues were centred in the *knowledges* and statements were generally inclined towards an informative character. The *knowings* of risks were references to scientific articles as well as trials and dialogues with public sector and university representatives, which moreover accounts for the knowers of risks – together with intergovernmental organisations as OECD. **As such, the risk assessment can be characterised by a generally externally informed way of knowing, informative socially- centred knowledges, and a mix of institutional actors as knowers.**

In terms of *knowledgeability*, this speaks for a discrepancy in preferred *knowledge*, *knowings* and *knowers* depending on aim. Whereas both academics and private consultants were prominent in the knowledge production, they were utilised as knowers in distinctively different assessments. Consequently, the knowledge in the reports diverged between an economic focus on benefits and socially-oriented risks, as well as a higher prevalence of empirical studies as a way of knowing risks, in contrast to beliefs as underpinning the benefit assessment. As for outcome, the risk assessment did catalyse the proposal for a 'trust model', although that was proposed as a voluntary mean (app.f), whereas the benefit assessment underpinned the three other proposals to further accelerate the implementation of AI in public administration. This illustrates the issues with 'evidence based policy-making' discussed in the literature review, as governments might seek simple knowledge in favour of already established policy goals (Smith & Leech, 2010) - as with the incoherent references to consultancy reports in support of the benefit assessment; or construct facts to validate policy goals (Torriti, 2010) - as in the instance of the simplified estimations and calculations of economic-benefits. The preference in knowledgeability of economic benefits over social risks also testifies to the disregard of critical assessments (Wye et al., 2015), as the risk-assessment included many references to serious harms while not resulting in any proposals for action outside a voluntary 'trust model', which furthermore highlights the view of digitalisation as an unquestionable imperative in Swedish policy (Gidlund & Nyhlén, 2022).

RQ 2: Does the policy formulation exhibit contributory injustice, and if so – how?

Considering the different sets of knowledges, knowings, knowledgeabilities as hermeneutical resources (therefore, referencing them as hermeneutical resources henceforth), the following discussion aims to develop an understanding of whether the adoption of these hermeneutical resources in the policy formulation constitutes a case of contributory injustice.

In Dotson's (2012) definition of contributory injustice, an epistemic agent's situated ignorance in the use of structurally biased hermeneutical resources causes harm to the agency of a knower (p.31). Moreover, this is explained as a structural deployment of biased resources which hampers a knower's contributions to collective epistemic resources 'within a given epistemic community by compromising her epistemic agency' (Dotson, 2012: 32).

In the previous outline of the characteristics of hermeneutical resources and knowers in the knowledge production, the policy formulation displayed a discrepancy in deployment of hermeneutical resources depending on whether benefits or risks were considered. This does not testify to contributory injustice, as that would imply a disregard of opposing hermeneutical resources, and such were not disregarded since they were never acknowledged. For example, if a critical assessment of benefits had been made but disregarded, that would signify a contributory injustice. **As such, the assessment within each category cannot constitute ground for contributory injustice.** Moreover, contributory injustice is a framework based on dialogue (Dotson, 2012), although it does not explicitly depend on it, the examples provided by Dotson (2012: 33-34) reflect instances of disregard in dialogues within communities. Therefore, the framework might be incompatible with the methodology of QCA as method and policy documents as material, since that limits the scope for narrative analysis and cannot account for dialogues between knowers outside the scope of the policy document.

On the other hand, the theory might allow for an expansion by regarding the relation between the hermeneutical resources applied to each assessment as a dialogue. The discrepancy between the different sets of hermeneutical resources applied can be viewed as a structural bias against the hermeneutical resources presented in the risk assessment, as they encompass qualitative and citizen-centred values that were not considered within the scope of the benefit assessment, since they were limited by the hermeneutical resources deployed there – i.e., by the knowing of the framework analysing cost-reduction and revenue which could not account for qualitative values (see - analysis). Thus, the hermeneutical resources and knowers of the risk assessment suffers contributory injustice in that the critique was limited by the favouring of a different set of resources for the benefit assessment, a set that the knowers of the risk assessment could not adequately respond to, rendering their critique unfit. **In that sense, the policy formulation does exhibit contributory injustice in the**

thwarting of qualitative values and favouring of hermeneutical resources deployed in the benefit assessment.

However, the conclusion that contributory injustice occurred does raise questions to the theoretical framework. Epistemic (in)justice theory relies on the notion of *marginalised knowers* within knowledge production (Fricker, 2007; Dotson, 2012), and the conclusion that contributory injustice occurred in this policy formulation therefore requires an assessment of who these knowers are, and on what basis they could be considered marginalised. Simply put, the knowers of the risk assessment were mainly academic scholars whereas the knowers of benefits were mainly state-actors and commercial consultants (Table 3; app.e). How could academic scholars then be considered marginalised in comparison to state-actors and commercial consultants, without considering who they are as specific individuals? As it is outside the scope for this thesis to assess the dynamic between these groups of actor in terms of power, it might be incorrect to regard academic scholars as marginalised in their capacity as knowers, on the basis of the conclusions from the analysis. But considering some perspectives from the literature review, the marginalisation here might be underpinned by other factors - outside the scope of the identity or socioeconomic power dimensions considered marginalisation-basis within the theoretical framework of epistemic (in)justice (Fricker, 2007; Dotson, 2012).

AI in public administration is suggested to correspond with neoliberal ideals of NPM (Yeung, 2023) and the work of commercial consultants is generally characterised as being influenced by NPM values (Marciano, 2022; Jobér, 2023) Considering Nolan's (2015) suggestion that policy privileges research that conforms to ideological ideals of the decision-makers, **the marginalisation can be viewed as ideological.** Following this line of reasoning, the knowers of the benefit assessment were favoured on the basis of aligning with the ideological underpinning of the policy, in this case, neoliberalism, and the knowers of risks were thwarted by not corresponding to the values of efficiency and productivity that NPM and neoliberalism encompasses (Elcock, 2012). This would imply that the subordination of qualitative, citizen-centred hermeneutical resources and the exclusion of academic perspectives on benefits in favour of consultants, **constitutes an ideological marginalisation of contrary perspectives in the knowledge production.** While this answer resonates with ideas cited in the literature above, there is no theoretical expansion of neoliberalism covered in this study, which

speaks for a limitation in assessing the validity of this claim. Hence, the answer to RQ2 will be considered insufficient due to theoretical limitations.

CONCLUSION

The main aim of this study was to analyse a case of policy-formulation as a form of knowledge production. Prompted by Björklund's (2022) inquiry to critically examine digital transformations in welfare, regarding the use of experts and evidence in policy, this thesis set out to explore the Swedish government's mission: "Promote the public administration's ability to use AI". The research was conducted through a qualitative content analysis underpinned by the theoretical frameworks of policy formulation as knowledge production (Gunter, 2017) and contributory injustice (Dotson, 2012), to understand policy-formulation as a result of the deployment of certain hermeneutical resources. Furthermore, the aim was to consider this deployment in relation to the exclusion or subordination of alternative hermeneutical resources, thus emphasising policy-formulation as a constructive process of inclusion and exclusion of knowledge.

The analysis and discussion show that the knowledge production of the policy was characterised by a discrepancy in the form of different sets of hermeneutical resources and knowers applied to different assessments of the policy texts. While the benefit assessment was characterised by quantifiable metrics and a higher prominence of consultancy firms as *knowers*, the risk assessment indicated an inclination towards qualitative values and a higher use of academic knowers. Furthermore, the study underscores the earlier critique of evidence assessments in policy (Smith & Leech, 2010; Wye et al., 2015; Torriti, 2010) by pointing to inadequate support for claimed benefits, as well as the subordination of critical examinations. The findings shows that the policy favoured the assessment of economic benefits over citizen- centred benefits, which illustrates the epistemic tension between quantifiable measures and the abstract concept of 'public values'. This highlights the need for critical appraisals of policies, as the findings speak for a constructivist perspective on knowledge in policy – there were more than one set of knowledge available, but certain knowledge gained salience and some perspectives were omitted in the process of knowledge production. In that regard, this study has shown that the epistemological workings of the benefit assessment favoured simplified knowledge aligned with the policy aim over critical knowledge, through the omittance of a critical

examination of economic risks, as well as a risk assessment incompatibly construed with the assessment of benefits.

In an attempt to develop an understanding as to why the knowledge production illustrated these issues, the discussion suggests a case of contributory injustice in the form of an ideological marginalisation of counter perspectives in the knowledge production. Moreover, this relates to the other aim of the study, to expand on theoretical accounts for epistemic assessments of AI policy. Although the argument for contributory epistemic injustice somewhat aligns with an earlier study on knowledge production in policy-making (Nolan, 2015), the suggestion lacks theoretical ground for further support, as 'neoliberalism' was not included in the theoretical framework. This reflects a potential for further research, on knowledge production in policy formulation underpinned by neoliberal ideals, as many of the sources cited in the literature review touch upon related issues (specifically with regards to new public management), whilst not addressing the actual knowledge production of policy in relation to ideological underpinning. Therefore, further research could consider ideological frameworks related to knowledge production, and how the inclusion/exclusion of certain perspectives relates to this.

ACKNOWLEDGEMENTS

REFERENCES

Aguilera, A., & Barrera Ramos, M. (2016) Technological unemployment: An approximation to the Latin American case, *AD-minister* 29(1): 59–78.

Ahn, M. J., & Chen, Y-C. (2022) Digital transformation toward AI augmented public administration: The perception of government employees and the willingness to use AI in government, *Government Information Quarterly* 39(2): 101664.

Alford, J., & Hughes, O. (2008) Public Value Pragmatism as the Next Phase of Public Management, *The American Review of Public Administration* 38(2): 130-148.

Alhosani, K., & Alhashmi, S. M. (2024) Opportunities, challenges, and benefits of AI innovation in government services: A review, *Discover Artificial Intelligence* 4(1): 18. URL:https://doi.org/10.1007/s44163-024-00111-w [Last consulted 2024-08-02]

Androutsopoulou, A., Karacapilidis, N., Loukis, E., Charalabidis, Y. (2019) Transformingthe communication between citizens and government through aI-guided chatbots, *Government Information Quarterly* 36(2): 358–367.

Bank, M., Duffy, F., Leyendecker, V., & Silva, M. (2021) *The lobby network: Big Tech's web of influence in the EU*. Corporate Europe Observatory and LobbyControl. URL: https://corporateeurope.org/en/2021/08/lobby-network-big-techs-web-influence-eu [July 16, 2024]

Bertelli, A. M., Mele, V., & Whitford, A. B. (2020) When new public management fails: Infrastructure public–private partnerships and political constraints in developing and transitional economies, *Governance* 33(3): 477–493.

Björklund, K. (2022) *Digitaliseringslobbyn - på vems villkor digitaliseras vården?* Arena Idé. URL: https://arenaide.se/rapporter/digitaliseringslobbyn/ [Last consulted 2024-02-14]

Blix, M., & Jeansson, J. (2019) Telemedicine and the Welfare state: The Swedish experience, pp. 15-32 in A. Larsson & R. Teigland (eds). *Digital Transformation and Public Services*, London: Routledge.

Bozeman, B. (2007) *Public Values and Public Interest: Counterbalancing Economic Individualism*, Washington, DC: Georgetown University Press.

Brandão, M. A., Reis, A., Mendes, B., De Almeida, C. A., Oliveira, G., Hott, H., Gomide, L., Costa, L., Silva, M., Lacerda, A., & Pappa, G. (2024) PLUS: A Semi-automated Pipeline for Fraud Detection in Public Bids, *Digital Government: Research and Practice* 5(1): 1-16.

Brown, W. (2015) *Undoing the Demos : Neoliberalism's Stealth Revolution,* New York, NY: Zone Books.

Bryson, J., Crosby, B. C., Bloomberg, L. (2014) Public Value Governance: Moving Beyond Traditional Public Administration and the New Public Management, *Public Administration Review* 74(4): 445-456.

Buhmann, A., & Fieseler, C. (2023) Deep Learning Meets Deep Democracy: Deliberative Governance and Responsible Innovation in Artificial Intelligence, *Business Ethics Quarterly* 33(1): 146–179.

Bullock, J. B. (2019) Artificial intelligence, discretion, and bureaucracy, *The American Review of Public Administration* 49(7):751–761.

Buschor, E. (1994) Performance Management in the public sector: Past, current and future Trends, *Tékhne* 11(1): 4-9. URL: https://doi.org/10.1016/j.tekhne.2013.05.005 [Last consulted 2024-08-03]

Cardullo, P., & Kitchin, R. (2019) Being a 'citizen' in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland, *GeoJournal* 84: 1–13.

Choi, B. C., Pang, T., Lin, V., Puska, P., Sherman, G., Goddard, M., Ackland, M. J., Sainsbury, P., Stachenko, S., Morrison, H., & Clottey, C. (2005) Can scientists and policy makers work together?, *Journal of epidemiology and community health* 59(8): 632–637.

Chun, A. H. C. (2007) Using AI for E-government automatic assessment of immigration application forms, paper presented at the 19th National Conference on Artificial Intelligence, Vancouver, Canada, 22-26/07

Chun, A. H. C. (2008) An AI framework for the automatic assessment of E-government Forms, *AI Magazine* 29(1): 52–64.

Collington, R. (2022) Disrupting the Welfare State? Digitalisation and the Retrenchment of Public Sector Capacity, *New Political Economy* 27(2): 312–328.

Coglianese, C., & David, L. (2017) Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, *Georgetown Law Journal* 105(5): 1147-1223.

Dahl, A., & Soss, J. (2014) Neoliberalism for the Common Good? Public Value Governance and the Downsizing of Democracy, *Public Administration Review* 74(4): 496–504.

Decision 2022/2481. Decision (EU) 2022/2481 of the European Parliament and the Council establishing the Digital Decade Policy Programme 2030. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022D2481 [Last consulted 2024-08-01].

Delmon, J. (2011) *Public-private partnership projects in infrastructure: An essential guide for policy makers*, Cambridge, UK: Cambridge University Press.

Dencik, L., & Kaun, A. (2020) Datafication and the Welfare State, *Global Perspectives* 1(1): 12912.URL: https://doi.org/10.1525/gp.2020.12912 [Last consulted 2024-05-30]

Denhardt, R. B., & Denhardt, J. V. (2000) The New Public Service: Serving Rather Than Steering, *Public Administration Review* 60(6): 549–559.

DIGG. (2020) *Främja den offentliga förvaltningens förmåga att använda AI* [Promote public administration's ability to use AI]. https://www.digg.se/download/18.79c61f7c17db5871992f0ad/1647952779554/framja-den-offentliga-forvaltningens-formaga-att-anvanda-ai.pdf [Last consulted 2024-03-14].

DIGG. (2023) *Slutrapport: Uppdrag att främja offentlig förvaltnings förmåga att använda AI* [Final report: Mission to promote public administration's ability to use AI]. https://www.digg.se/download/18.5b30ce7218475cd9ed39384/1674479294670/Slutrapport%20Uppdrag%20att%20fr%C3%A4mja%20offentlig%20f%C3%B6rvaltnings%20f%C3%B6rm%C3%A5ga%20att%20anv%C3%A4nda%20AI%20I2021-01825.pdf [Last consulted 2024-03-14]

Dir. 2023:164. *Förstärkt AI-förmåga i Sverige*. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/kommittedirektiv/forstarkt-ai-formaga-i-sverige_hbb1164/ [Last consulted 2024-06-10]

Dirican, C. (2015) The impacts of robotics, artificial intelligence on business and economics, *Procedia – Social and Behavioral Sciences* 195(1): 564-573.

Dobrolyubova, E. (2021) Measuring Outcomes of Digital Transformation in Public

Administration: Literature Review and Possible Steps Forward, *NISPAcee Journal of Public Administration and Policy* 14(1): 61–86.

Dotson, K. (2012) A Cautionary Tale: On Limiting Epistemic Oppression, *Frontiers: A Journal of Women Studies* 33(1): 24.

Duberry, J. (2022) *Artificial Intelligence and Democracy*, Edward Elgar Publishing, URL: https://doi.org/10.4337/9781788977319 [Last consulted 20204-06-11]

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y.,

Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021) Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy, *International Journal of Information Management* 57(1): 101994.

El Naqa, I., & Murphy, M. J. (2015) What is Machine Learning? In I. El Naqa, R. Li, M. Murphy (eds) *Machine Learning in Radiation Oncology*, Cham: Springer

Elcock, H. (2012) Ethics and the public interest: A question of morality, *Teaching Public Administration* 30(2): 115-123.

Elo, S. and Kyngäs, H. (2008) The qualitative content analysis process, *Journal of Advanced Nursing* 62(1): 107-115. European Commission. (2019) *Digital Single Market*, URL: https://commission.europa.eu/publications/digital-single-market_en [Last consulted 2024-07-30]

Fagerberg, J., & Hutschenreiter, G. (2020) Coping with Societal Challenges: Lessons for Innovation Policy Governance, *Journal of Industry, Competition and Trade* 20: 279–305. https://doi.org/10.1007/s10842-019-00332-1 [Last consulted 2024-07-29]

Fatima, S., Desouza, K. C., Buck, C., & Fielt, E. (2022) Public AI canvas for AI-enabled public value: A design science approach, *Government Information Quarterly* 39(4): 101722.

Filgueiras, F. (2022) New Pythias of public administration: Ambiguity and choice in AI systems as challenges for governance, *AI & SOCIETY* 37(4): 1473–1486. https://doi.org/10.1007/s00146-021-01201-4 [Last consulted 2024-05-12]

Fricker, M. (2007) *Epistemic injustice : power and the ethics of knowing*, Oxford: Oxford University Press.

Früh, W. (2007) *Inhaltsanalyse* [Content analysis] (6th ed.), Konstanz: UVK.

Gębczyńska, A., & Brajer-Marczak, R. (2020) Review of Selected Performance Measurement Models Used in Public Administration, *Administrative Sciences* 10(4): 99.

Giannone, D., & Santaniello, M. (2019). Governance by indicators: The case of the Digital Agenda for Europe, *Information, Communication & Society* 22(13): 1889–1902.

Gidlund, K. L., & Sundberg, L. (2021) Undisclosed creators of digitalization: A critical analysis of representational practices, *Information Polity* 26(1): 3–20.

Gonzales, J. T. (2023) Implications of AI innovation on economic growth: A panel data study, *Journal of Economic Structures* 12(1): 13.

Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017) Methodological challenges in qualitative content analysis: A discussion paper, *Nurse Education Today* 56(1): 29-34.

Groeben, N., & Rustemeyer, R. (1994) On the integration of quantitative and qualitative methodological paradigms (based on the example of content analysis), pp. 308-326 in I. Borg & P. Mohler (eds) *Trends and perspectives in empirical social research*, Berlin: De Gruyter.

Gruening, G. (2001) Origin and theoretical basis of new public management, *International Public Management Journal* 4(1): 1-25.

Gunter, H. M. (2017). Consultants and policy formulation. In M. Howlett & I. Mukherjee (Eds.), *Handbook of Policy Formulation* (pp. 337-355). Cheltenham, UK: Edward Elgar Publishing

Haque, M. S. (1999) Relationship between Citizenship and Public Administration: A Reconfiguration, *International Review of Administrative Sciences* 65(3): 309–325.

He, Y. (2019) The Effect of Artificial Intelligence on Economic Growth: Evidence from Cross-Province Panel Data, *Korean Journal of Artificial Intelligence* 7(2): 9–12.https://doi.org/10.24225/KJAI.2019.7.2.9 [Last consulted 2024-06-01]

Hodge, G. A., & Greve, C. (2017) On Public–Private Partnership Performance: A Contemporary Review, *Public Works Management & Policy* 22(1): 55-78.

Howlett, M., & Migone, A. (2013) Policy advice through the market: The role of external consultants in contemporary policy advisory systems, *Policy and Society* 32(3): 241–254.

Howlett, M., & Mukherjee, I. (2017) Policy formulation: where knowledge meets power in the policy process, pp. 3-23 in M. Howlett & I. Mukherjee (eds) *Handbook of Policy Formulation*, Cheltenham: Edward Elgar Publishing.

Janks, H. (1997) Critical Discourse Analysis as a Research Tool, *Discourse: Studies in the Cultural Politics of Education* 18(3): 329-342.

Jobér, A. (2023) Private actors in policy processes. entrepreneurs, edupreneurs and Policyneurs, *Journal of Education Policy* 39: 1-20.

Krejnus, M., Stofkova, J., Repkova Stofkova, K., & Binasova, V. (2023) The Use of the DEA Method for Measuring the Efficiency of Electronic Public Administration as Part of the Digitization of the Economy and Society, *Applied Sciences* 13(6): 3672. https://doi.org/10.3390/app13063672 [Last consulted 2024-06-22]

Krippendorff, K. (2004) *Content analysis: An introduction to its methodology* (2nd ed.), London: Sage Publications.

Kuckartz, U. (2014) *Qualitative text analysis. A guide to methods, practice, and using Software*, London: Sage Publications.

Larsson, A., & Teigland, R. (eds) (2019) *Digital Transformation and Public Services: Societal Impacts in Sweden and Beyond* (1st ed.), London: Routledge.

Liabo, K., Cockcroft, E. J., Boddy, K., Farmer, L., Bortoli, S., & Britten, N. (2022) Epistemic justice in public involvement and engagement: creating conditions for Impact, *Health Expect* 25(4):1967-1978.

Ljungqvist, M., & Sonesson, A. (2022) Selling out Education in the Name of Digitalization: A Critical Analysis of Swedish Policy, *Nordic Journal of Studies in Educational Policy* 8(2): 89–102.

Lueger, M. (2010) *Interpretative Sozialforschung: Die Methoden* [Interpretative social research: The methods], Wien: Facultas.

Madan, R., & Ashok, M. (2023) AI adoption and diffusion in public administration: A systematic literature review and future research agenda, *Government Information Quarterly*, 40(1): 101774.

Madan, R., & Ashok, M. (2024) Making Sense of AI Benefits: A Mixed-method Study in Canadian Public Administration, *Information Systems Frontiers*: 1-35. https://doi.org/10.1007/s10796-024-10475-0 [Last consulted 2024-07-20]

Marciano, R. (2022) Beyond consultocracy and servants of power: Explaining the role of consultants in policy formulation, *Governance* 37(1): 39–58.

Margetts, H., & Dorobantu, C. (2019) Rethink Government with AI, *Nature* 568 (7751): 163–165.

Mayring, P. (2014) *Qualitative content analysis: theoretical foundation, basic procedures and software solution*, Klagenfurt: SSOAR, URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173 [Last consulted 2024-08-01]

Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023) Artificial Intelligence in Government: Taking Stock and Moving Forward, *Social Science Computer Review* 41(1): 123-140.

Mehr, H. (2017). *Artificial intelligence for citizen services and government*. Cambridge, MA: Harvard Kennedy School, Ash Center for Democratic Governance and Innovation. URL: https://ash.harvard.edu/wpcontent/uploads/2024/02/artificial_intelligence_for_citizen_services.pdf [Last consulted 2024-07-01]

Monarcha-Matlak, A. (2021) Automated decision-making in public administration, *Procedia Computer Science* 192(1): 2077–2084.

Mormina, M. (2022) Knowledge, Expertise and Science Advice During COVID-19: In Search of Epistemic Justice for the 'Wicked' Problems of Post-Normal Times, *Social Epistemology* 36(6): 671–685.

Morrell, K. (2009) Governance and the Public Good, *Public Administration* 87(3): 538–56.

Moser-Plautz, B., & Schmidthuber, L. (2023). Digital government transformation as an organizational response to the COVID-19 pandemic, *Government Information Quarterly* 40(3): 101815.

Mura, PO., & Donath, LE. (2023) Digitalisation and Economic Growth in the European Union, *Electronics* 12(7):1718.

Nolan, K. (2015) Neoliberal common sense and race-neutral discourses: A critique of "evidence-based" policy-making in school policing, *Discourse: Studies in the Cultural Politics of Education* 36(6): 894–907.

Nyhlén, S., & Gidlund, K. L. (2022) In conversation with digitalization: Myths, fiction or professional imagining?, *Information Polity* 27(3): 331–341.

Oliver, K., Lorenc, T., & Innvær, S. (2014) New directions in evidence-based policy research: a critical analysis of the literature, *Health Research Policy and Systems* 12(1): 34.

Parkhurst, J. (2017). *The politics of evidence: From evidence-based policy to the good governance of evidence*, London: Routledge.

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020) Big Data and AI – A transformational shift for government: So, what next for research?, *Public Policy and Administration* 35(1): 24–44.

Pollitt, C. (2016). *Advanced introduction to public management and administration* (pp.1-26), Cheltenham: Edward Elgar Publishing Limited.

Puppis, M. (2019) Analyzing Talk and Text I: Qualitative Content Analysis, pp. 367-385 in H. Van den Bulck, M. Puppis, K. Donders, L. Van Audenhove (eds) *The palgrave handbook of methods for media policy research*, London: Palgrave Macmillan.

Ranchordás, S. (2022) Empathy in the digital administrative state, *Duke Law Journal* 71(6): 1341-1390.

Ranerup, A., & Henriksen, H. Z. (2022) Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden's Social Services, *Social Science Computer Review* 40(2):445-461.

Regeringsbeslut I2021/01825. *Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens* [Mission to promote public administration's ability to use artificial intelligence]. Regeringskansliet.

https://www.regeringen.se/contentassets/fd078181a7614601845f5439efcafe81/uppdrag-attframja-offentlig-forvaltnings-formaga-att-anvanda-artificiell-intelligens/ [Last consulted 2024-07-02]

Regeringsbeslut Fi2024/01535. *Uppdrag till Myndigheten för digital förvaltning och Integritetsskyddsmyndigheten* [Assignment to the Digital Governance Agency and the Privacy Protection Authority]. Regeringskansliet. https://www.regeringen.se/contentassets/aee6dff026b849e8957604351614038a/uppdrag-till51digg-och-imy-att-ta-fram-riktlinjer-for-anvandningen-av-generativ-artificiell-intelligensinom-den-offentliga-forvaltningen.pdf [Last consulted 2024-07-02]

Richards, G. W. (2017) How research-policy partnerships can benefit government: a win–win for evidence-based policy-making, *Canadian Public Policy* 43(2):165–70.

Rosenthal, G. (2015) *Interpretative Sozialforschung: Eine Einführung* [Interpretative social research: an introduction] (5th ed.), Basel: Beltz Juventa.

Russell, S., & Norvig, P. (2021) *Artificial Intelligence, A Modern Approach* (4th ed.), London: Pearson

Rössler, P. (2005). *Inhaltsanalyse* [Content analysis], Konstanz: UVK.

Sandelowski, M. (1995) Sample size in qualitative research, *Research in Nursing & Health* 18(2): 179-183.

Sanderson, I. (2002) Evaluation, Policy Learning and Evidence-Based Policy Making, *Public Administration* 80(1): 1-22.

Scherer, M. U. (2015) Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, *Harvard Journal of Law & Technology* 29(2): 354-398.

Schiff, D. S. (2024) Framing contestation and public influence on policymakers: Evidence from US artificial intelligence policy discourse, *Policy and Society* 43(3): 255-288

Schreier, M. (2012) *Qualitative content analysis in practice* (1st ed.), London: SAGE Publications

Schreier, M., Stamann, C., Janssen, M., Dahl, T., & Whittal, A. (2019) Qualitative content analysis: Conceptualizations and challenges in research practice-introduction to the FQS special issue "qualitative content analysis I", *Forum, Qualitative Social Research* 20(3): 38.

Seabrooke, L., Sending, O. J. (2022) Consultancies in public administration, *Public Administration* 100(3): 457-471.

Sidorenko, E .L., Bartsits, I. N., & Khisamova, Z. I. (2019) The Efficiency Of Digital Public Administration Assessing: Theoretical And Applied Aspects, *Public administration issues* 2(1): 93-114.

Sigfrids, A., Nieminen, M., Leikas, J., & Pikkuaho, P. (2022) How Should Public Administrations Foster the Ethical Development and Use of Artificial Intelligence? A Review of Proposals for

Developing Governance of AI, *Frontiers in Human Dynamics* 4(1): 858108. https://doi.org/10.3389/fhumd.2022.858108 [Last consulted 2024-05-11]

Smith, K. E., & Joyce, K. E. (2012) Capturing complex realities: Understanding efforts to achieve evidence-based policy and practice in public health, *Evidence & Policy* 8(1): 57–78.

Smith, A., & Leech, R. (2010) 'Evidence. What evidence?': evidence-based policy making and School Sport Partnerships in North West England, *International Journal of Sport Policy and Politics* 2(3): 327–345.

Thapa, I. (2020) *Public Administration: Meaning, Scope and Its Nature* (Technical Report). Public Administration Campus, Tribhuvan University, URL: https://doi.org/10.13140/RG.2.2.33704.80641 [Last consulted 2024-05-10]

Thomas, H.G. (2001) Towards a new higher education law in Lithuania: reflections on the process of policy formulation, *Higher Education Policy* 14(3): 213–223.

Toll, D., Lindgren, I., Melin, U., & Madsen, C. Ø. (2019) Artificial Intelligence in Swedish Policies: Values, Benefits, Considerations and Risks, pp.301-310 in I. Lindgren, M. Janssen, H. Lee, A. Polini, M. P. Rodríguez Bolívar, H. J. Scholl, & E. Tambouris (eds) *Electronic Government*, Springer International Publishing. https://doi.org/10.1007/978-3-030-27325-5_23 [Last consulted 2024-06-17]

Torriti, J. (2010) Impact Assessment and the Liberalization of the EU Energy Markets: Evidence-Based Policy-Making or Policy-Based Evidence-Making?, *Journal of Common Market Studies* 48(4): 1065–1081.

Transparency International EU. (2021) *Big pockets, open doors: Big tech lobbying in Brussels*, URL: https://transparency.eu/deep-pockets-open-doors/ [Last consulted 2024-07-16]

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021) Framing governance for a contested emerging technology:insights from AI policy, *Policy and Society* 40(2): 158–177.

Uzun, M., Yıldız, M., & Önder, M. (2022) Big Questions of Artificial Intelligence (AI) in Public Administration and Policy, *Siyasal: Journal of Political Sciences* 31(2): 423–442.

van den Berg, C., Howlett, M., Migone, A., Howard, M., Pemer, F., & Gunter, H. M. (eds) (2019) Swedish Government Agencies' Hiring of Policy Consultants: A Phenomenon of Increased Magnitude and Importance?, pp. 177-220 in *Policy Consultancy in Comparative Perspective: Patterns, Nuances and Implications of the Contractor State* Cambridge: Cambridge University Press.

van Noordt, C., & Misuraca, G. (2020). Exploratory insights on artificial intelligence for government in Europe. *Social Science Computer Review* 40(2): 426-444.

van Noordt, C., & Tangi, L. (2023) The dynamics of AI capability and its influence on public value creation of AI within public administration, *Government Information Quarterly* 40(4): 101860.

Veale, M., & Brass, I. (2019) Administration by algorithm? Public management meets public sector machine learning, pp.1-30 in K. Yeung & M. Lodge (eds) *Algorithmic Regulation*, Oxford: Oxford University Press.

Wallmeier, F., Helmig, B., & Feeney, M. K. (2019) Knowledge Construction in Public Administration: A Discourse Analysis of Public Value, *Public Administration Review* 79(4): 488–499.

Wang, H., Xiong, W., Wu, G., & Zhu, D. (2017) Public–private partnership in Public Administration discipline: a literature review, *Public Management Review* 20(2): 293–316.

Widenhorn, S. (2013) Towards Epistemic Justice with Indigenous Peoples' Knowledge? Exploring the potentials of the convention on biological diversity and the philosophy of Buen Vivir, *Development* 56(1): 378–386.

Wilson, C. (2022) Public engagement and AI: A values analysis of national strategies, *Government Information Quarterly* 39(1):101652.

Wirtz, B. W., & Müller, W. M. (2018) An integrated artificial intelligence framework for public management, *Public Management Review* 21(7): 1076–1100.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018) Artificial Intelligence and the Public Sector— Applications and Challenges, *International Journal of Public Administration* 42(7): 596–615.

Wye, L., Brangan, E., Cameron, A., Gabbay, J., Klein, H. J., & Pope, C. (2015) Evidence based policy making and the 'art' of commissioning – how English healthcare commissioners access and use information and academic research in 'real life' decision-making: an empirical qualitative study, *BMC Health Services Research* 15(1): 430.

Yanow, D. (2007) Interpretation in policy analysis: On methods and practice, *Critical Policy Studies* 1(1):110–122.

Yeung, K. (2023) The New Public Analytics as an Emerging Paradigm in Public Sector Administration, *Tilburg Law Review* 27(2): 1–32.

Zheng, Y., Yu, H., Cui, L., Miao, C., Leung, C., & Yang, Q. (2018) SmartHS: An AI Platform for Improving Government Service Provision, paper presented at the 32nd AAAI Conference on Artificial Intelligence, Louisiana, US, 2-7/2

Önder, M., & Uzun, M. M. (2021) Roles of Artificial Intelligence (AI) on COVID-19

Pandemic Crisis Management Policies, *International Journal of Public Administration in the Digital Age (IJPADA)* 8(2):1-13.