I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14021

Public Procurement and Innovation for Human-Centered Artificial Intelligence

Wim Naudé
Nicola Dimitri

JANUARY 2021

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14021

Public Procurement and Innovation for Human-Centered Artificial Intelligence

Wim Naudé
University College Cork, RWTH Aachen University and IZA

Nicola Dimitri
University of Siena

JANUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Public Procurement and Innovation for Human-Centered Artificial Intelligence

The possible negative consequences of Artificial Intelligence (AI) have given rise to calls for public policy to ensure that it is safe, and to prevent improper use and misuse. Human-centered AI (HCAI) draws on ethical principles and puts forth actionable guidelines in this regard. So far however, these have lacked strong incentives for adherence. In this paper we contribute to the debate on HCAI by arguing that public procurement and innovation (PPaI) can be used to incentivize HCAI. We dissect the literature on PPaI and HCAI and provide a simple theoretical model to show that procurement of innovative AI solutions underpinned by ethical considerations can provide the incentives that scholars have called for. Our argument in favor of PPaI for HCAI is also an argument for the more innovative use of public procurement, and is consistent with calls for mission-oriented and challenge-led innovation policies. Our paper also contributes to the emerging literature on public entrepreneurship, given that PPaI for HCAI can advance the transformation of society, but only under uncertaint.

JEL Classification: H57, D02, O38, O32

Keywords: artificial intelligence, data, innovation, public procurement, ethics

Corresponding author:
Wim Naudé
Technology and Innovation Management (TIM)
RWTH Aachen University
Kackertstraße 7
52072 Aachen
Germany

E-mail: naude@time.rwth-aachen.de

1 Introduction

Artificial intelligence (AI) is[1] the "simulation of human intelligence processes by machines, especially computer systems." These processes include learning, reasoning and self-correction. Although the term Artificial Intelligence (AI) was first used in 1956, when attempts to build AI was based on computational logic, modern AI is based on a statistical-probabilistic approach and is specialized in narrow domains (Moor, 2006; Naudé, 2021). It is also known as narrow AI, to distinguish it from the not-yet-existing Artificial General Intelligence (AGI) which would be indistinguishable from human intelligence (Stanford University, 2016). The use of Machine Learning (ML) and in particular Deep Learning (DL) based on big data has been particularly successful in generating technologies that have become as good and, in many cases, better than humans at pattern recognition, prediction and natural language processing (NLP). As a result, AI models are now routinely and widely used to provide services and products such as online search engines, chatbots and virtual assistants, recommender systems, reputation systems, news curation and aggregation, hyper-personalization of marketing, translation, credit scoring, predictive policing and spam filters, amongst others. Progress has also been made in developing autonomous vehicles and medical diagnostic tools.

Given the ubiquity and growth in data for ML, allowing for AI models to improve their performance over time, and the wide range of terrains where pattern recognition prediction and NLP are required, it is clear why there has been high expectations put on AI, and this is why it has even been described as a new general-purpose technology (Trajtenberg, 2018). The high expectations of AI are reflected not only in the declarations of scientists and government officials, but also in the growing investments and research in the field of AI specifically, and data science more generally. With these expectations have however also come warnings and concerns about the possible long-term existential threats of AI (Bostrom, 2014) as well as the shorter-term negative consequences. The latter include intrusive surveillance[2] and erosion of privacy, AI weapons, job losses due to automation, higher inequality, discrimination and biased policy making (Frey and Osborne, 2017; Korinek and Stiglitz, 2017; Feldstein, 2019; Russel et al., 2015). The GitHub site "Awful AI" contains a repository of some of the negative consequences of AI.[3]

[1] See `www.Whatis.com`; Note however that there is no single universal definition of AI (Van de Gevel and Noussair, 2013).

[2] As Smith and Neupane (2018, p.25) point out, "AI algorithms supercharge surveillance by processing data faster than previously possible and detecting patterns too subtle for human analysts to uncover."

[3] For Awful AI, see `https://github.com/daviddao/awful-ai`.

These negative consequences of AI have given rise to calls for proper AI governance[4] (Dafoe, 2018) to ensure that AI-based applications are safe, and that the development and diffusion of AI do not suffer from improper use and misuse. What would count as improper use and misuse, and safety-risks such as that could be caused by accidents or unintended consequences of AI systems, would be determined by the extent to which AI systems are human-centered. Human-centered AI (HCAI) draws on ethical principles and puts forth actionable guidelines for reducing the risks mentioned (Shneiderman, 2020). As such, HCAI is concerned with Ethical AI[5] and Responsible AI.[6]

In recent years there have been a growing number of initiatives to elaborate on Ethical and Responsible AI, see e.g. Boddington (2017), Etzioni and Etzioni (2017), Floridi (2018) and Yu et al. (2018). These include the Asilomar AI Principles of the Institute of the Future of Life[7] (2017), the European Union's April 2019 Ethics Guidelines for Trustworthy AI (EC, 2019), the OECD's May 2019 Principles on Artificial Intelligence (OECD, 2019), the G-20's June 2019 Human-Centered AI Principles (G-20, 2019) and the Institute of Electrical and Electronics Engineers' Ethically Aligned Design Principles (IEEE, 2019). In addition to these cross-country and multinational initiatives, the European Union Agency for Fundamental Rights (FRA) has documented more than 290 AI policy initiatives in individual EU Member States between 2016 and 2020.[8]

A shortcoming of these is that they tend to lack strong incentives for developers and users of AI to adhere to them (Askell et al., 2019; Calo, 2017; Hagendorff, 2020). To incentivize the ethics for HCAI, Eitel-Porter (2020, p.1) argues for "strong, mandated governance controls including tools for managing processes and creating associated audit trails." Floridi et al. (2018, p.704) argues that governments should "incentivise financially the inclusion of ethical, legal and social considerations in AI research projects."

In this paper we contribute to the debate on HCAI by arguing that public procurement and innovation is a potentially relevant tool with which to incentivize HCAI. We provide a simple

[4] "The field of AI governance studies how humanity can best navigate the transition to advanced AI systems, focusing on the political, economic, military, governance, and ethical dimensions" (Dafoe, 2018, p.5). In 2016, an open letter by an eminent group of scientists ignited the field of AI governance by calling for "expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial: our AI systems must do what we want them to do" - see `https://futureoflife.org/ai-open-letter/`.

[5] It is also sometime still referred to as "Machine Ethics", which has been defined as being "concerned with ensuring that the behavior of machines toward human users, and perhaps other machines as well, is ethically acceptable" (Anderson and Anderson, 2007, p.15)

[6] Responsible AI has been described as AI systems that "have an acceptably low risk of harming their users or society and, ideally, to increase their likelihood of being socially beneficial" (Askell et al., 2019, p.2).

[7] See `https://futureoflife.org/ai-principles/`

[8] See `https://tinyurl.com/y8juagop`.

theoretical model to show that such procurement of innovative AI solutions underpinned by ethical considerations, can provide both the tools and audit trails as well as the financial incentives that scholars such as Eitel-Porter (2020) and Floridi et al. (2018) have called for.

With the concept *public procurement and innovation* (PPaI) we encompass three dimensions of the relationship between public procurement and innovation (see Obwegeser and Müller (2018, p.5)). The first is *public procurement for innovati*on (PPfI), which deals with the question, how can public procurement drive innovation? The second is *public procurement of innovation* (PPoI), which deals with the question, how can public services be innovated? And the third is *innovative public procurement* (IPP) which deals with the question, how can public institutions procure innovatively? Our paper is relevant for all three of these dimensions of innovation and public procurement although we will lay more stress on the first, namely PPfI, and in particular one of its relatively new tools in the EU, the pre-commercial procurement of innovation (PCP).

Our argument in favor of public procurement and innovation (PPaI) to advance HCAI is also an argument for the more innovative use of public procurement. Most often in the past, public procurement has been used ex post to support new innovations by creating a demand for the product (being a customer for new products). This is also the way in which Lin (2020, p.26) conceives of the relationship between public procurement for innovation support. However, this is a rather restricted view: we argue in this paper that through PPfI of innovation, governments can also ex ante support research and development of goods and services that do not yet exist. This potential instrument for steering HCAI has been neglected in both the AI as well as innovation literature.[9] Miller and Lehoux (2020, p.2) quoting from Uyarra et al. (2017, p.828) confirms that the underlying mechanisms of how public procurement of innovation can shape the incentives of private agents in innovation is still "under-theorized." Our paper contributes by addressing this neglect.

Furthermore, our argument for PPaI as a policy towards HCAI is consistent with the calls for mission-oriented and challenge-led innovation policies, as made for instance by Mazzucato (2013, 2018). Mazzucato (2020, p.101) describe challenge-led policies as "policies that use investment and innovation to solve difficult problems." Certainly, the issue of HCAI is a challenge and a difficult problem. It is pre-eminently a challenge facing humanity and requires challenge-led innovation policies. PPfI and PPoI can fulfil this purpose, as it "uses public

[9]For instance, Bloom et al. (2019) provides an innovation policy toolkit, discussing eight policy tools: direct R&D grants, R&D tax credits, patent boxes, skilled immigration, supporting universities' research, competition policy, intellectual property rights, and mission-oriented policies. They do not discuss the role of public procurement of innovation.

procurement strategically to address a need which cannot be met by conventional solutions" (Lenderink et al., 2019, p.7).

Finally, our paper also contributes to the emerging literature of public entrepreneurship (see e.g., Hayter et al. (2018)), in that the use of public procurement of innovation for HCAI is an example of public entrepreneurship as it entails innovation, it aims to contribute towards transformation of society, and it is subject to uncertainty. These three elements – innovation, transformation and uncertainty, are according to Hayter et al. (2018, p.676) what characterises public entrepreneurship.

The rest of the paper will proceed as follows. In sections 2 and 3 we review the relevant literature dealing respectively with human-centered AI and innovation policy. In section 4 we provide a simple theoretical model wherein we show how public procurement of innovation can incentivize the development of HCAI. Section 5 concludes.

2 Human-Centered Artificial Intelligence

We start by reviewing the state of the literature on HCAI, in particular the role of ethical principles and actionable guidelines. This literature is part of the broader emerging literature on AI governance.

The downsides of AI, some of which was mentioned in the introduction, include intrusive surveillance and erosion of privacy, AI weapons, cybercrime, fake news and misinformation, job losses and tax losses due to automation, higher inequality, discrimination and biased policy making (Smith and Neupane, 2018). These are due to the misuse, accidents as well as systemic risks attached to the use of AI. It is clear that these downsides or risks posed by AI are of both an intentional nature (as in cybercrime, fake news and AI weapons) or unintentional nature (as in discrimination, inequality and accidents). In this paper we are largely concerned with the unintentional harm that AI can cause, and which can be more broadly analyzed as being a "systemic" or "accident" AI risks (Dafoe, 2018).

Systemic AI risk refers to "the risks of undesired outcomes - some of which may be very traditional - that can emerge from a system of competing and cooperating agents and can be amplified by novel forms of AI. For example, AI could increase the risk of inadvertent nuclear war, not because of an accident or misuse, but because of how AI could rapidly shift crucial strategic parameters, before we are able to build up compensating understandings, norms, and institutions" (Dafoe, 2018, p.28).

One aspect of systemic risk is that AI models can unintentionally result in biased decisions and recommendations due to being based on biased data (or data with gaps and absences) as well as design bias, where the values of the designer influences the model design and functioning. AI models also suffer, due to the nature of ML from lack of transparency and accountability resulting in the problem that decisions and outcomes from AI models very often cannot be easily explained. This is also known as the "black-box problem" (Castelvecchi, 2016; Nguyen et al., 2014). With bias and lack of explainability, "inequality in application" is a consequence: AI does not benefit everyone equality or fairly (Calo, 2017). Moreover, these shortcomings "disproportionately affect groups that are already disadvantaged by factors such as race, gender and socio-economic back-ground" (Crawford and Calo, 2016, p.312).

As far as accident risks are concerned, an AI accident occurs if "a human designer had in mind a certain (perhaps informally specified) objective or task, but the system that was designed and deployed for that task produced harmful and unexpected results" (Amodei et al., 2016, p.2). "Normal" accidents from AI should be expected. According to Maas (2018, p.4) it even "appears plausible that many AI applications may be even more susceptible to normal accidents than past 'textbook' case technologies such as nuclear power or aviation." There are at least three remedies for systemic and accident risks due to AI that have been given attention in the literature, and that forms part of the challenge of establishing a HCAI.

A first remedy is regulations and laws - to outlaw and police misuse – such as for instance through data privacy laws, combating cybercrime and other clearly malicious uses of AI – but also to lay down regulations for improving the safety of AI and limiting accidents. This remedy may be necessary but not sufficient. As Askell et al. (2019, p.7) remarks, the difficulties are that "AI regulation seems particularly tricky to get right, as it would require a detailed understanding of the technology on the part of regulators" and that "regulation that is reactive and slow may also be insufficient to deal with the challenges raised by AI systems."

A second remedy is to improve the safety of AI, and hence limit accidents and unintentional consequences, through for instance addressing technical issues in design, monitoring and operation (Maas, 2018; Leike et al., 2017) and by subjecting AI systems to regular monitoring and redesign (van de Poel, 2020). The problem is that issues of AI technical safety are complex,[10] and moreover as Leike et al. (2017, p.1) pointed out, "This nascent field of AI safety still lacks a general consensus on its research problems."

Thirdly, AI risks – both system and accident risks - can be addressed through the adoption and adherence to ethical principles and guidelines for the development and use of AI – for Ethical and Responsible AI. Such principles are needed for combatting the misuse and underuse of AI, for instance by making clear when the use of AI would constitute a fair or unfair usage, as well as for informing the technical safety of AI, for instance in identifying questions or problems that the design of an AI should incorporate.[11] The requirement of fairness has led to the concept of "Fair AI," a subset of Ethical AI, which refers to refers to "systems that both quantify bias and mitigate discrimination against subgroups" (Feuerriegel et al., 2020, p.379). HCAI extends beyond merely reducing misuse and negative side-effects and ensuring fairness. HCAI need to be able to contribute to "a more diverse, fair, and equitable society" (AI and Inclusion Symposium, 2017, p.2) and encourage and foster the diffusion and actual use of AI for promotion of societal development (Vinuesa et al., 2020). Thus, ethical AI will increase trust, and increased use of AI. Safe and ethical AI will also help firms[12] to avoid costly mistakes and so facilitate the uptake and diffusion of AI (Baker-Brunnbauer, 2020). Thus, HCAI approaches aim to establish the societal and commercial requisites for the greater use of AI (Floridi et al., 2018).

Eitel-Porter (2020, p.3) argue that most proposals for HCAI have similar pillars, namely fairness,[13] accountability, transparency, explainability and privacy. And according to Floridi et al. (2018, p.689) the ethical principles[14] proposed by the *AI4People* initiative are of beneficence, non-maleficence, autonomy, justice and explicability. In June 2019 the *G-20* proposed five *Human-Centered AI Principles* (G-20, 2019). These principles require AI to be consistent with (i) inclusive growth, sustainable development and well-being, (ii) human- centered values and fairness, (iii) transparency and explainability, (iv) robustness, security and safety, and (v) accountability (G-20, 2019, p.4).

[10] Everitt et al. (2018) considers the safety challenges involved over the long-term in the case of an AGI. In this paper we are only concerned with the shorter and medium risks arising from narrow AI.

[11] Leike et al. (2017) poses eight questions or problems that need to be answered during the design of AI to ensure that it will be safe for humans, for example "How do we ensure that an agent behaves robustly when its test environment differs from the training environment?"

[12] It is no surprise that a growing number of multinational firms have been coming up with their own ethical AI principles. For example, Baker-Brunnbauer (2020) relates the case of German automotive company *Continental* who in June 2020 "announced its intention to develop a code of ethics for its internal development and usage of AI that is based on the EC Trustworthy AI guideline."

[13] See Feuerriegel et al. (2020, pp.381-382) for a discussion of mathematical notions of fairness in AI.

[14] These are similar, with the exception of explicability, to principles of bioethics (Floridi et al., 2018).

Although the establishment of widely accepted and agreed on pillars and principles for ethical and responsible AI is a necessary and important milestone for the development and use of HCAI, it has a problem. The problem with this third remedy is, as we already stated in the introduction, that these principles and ethical guidelines lack strong incentives for developers and users of AI to adhere to them (Askell et al., 2019; Calo, 2017; Hagendorff, 2020). According to Hagendorff (2020, p.99) "Do those ethical guidelines have an actual impact on human decision-making in the field of AI and machine learning? The short answer is: No, most often not." He concludes, from a review of the field of AI ethics that AI ethics "lacks mechanisms to reinforce its own normative claims." Calo (2017, p.6-7) concurred with this assessment, noting that "... even assuming moral consensus, ethics lacks a hard enforcement mechanism. A handful of companies dominate the emerging AI industry. They are going to prefer ethical standards over binding rules for the obvious reason that no tangible penalties attach to changing or disregarding ethics should the necessity arise. " In fact, AI developers and users may under highly competitive conditions face the incentive to "underinvest in ensuring their systems are safe" (Askell et al., 2019, p.1). They may "skimp" on AI ethics also when lured by the potential winner-takes-all effects of eventually inventing an AGI, as Armstrong et al. (2016) emphasized.

To incentivize the ethics for HCAI, Eitel-Porter (2020, p.1) argues for a strong, mandated governance controls including tools for managing processes and creating associated audit trails." Floridi et al. (2018, p.704) argues that governments should "incentivise financially the inclusion of ethical, legal and social considerations in AI research projects." This then, is a key challenge facing HCAI: how to incentivize adherence to ethical principles? In the rest of the paper, we argue that the nature of this challenge is such that public procurement and innovation is very relevant.

In the next section, we argue that the literature on innovation policy suggests that public procurement can make an important contribution towards steering society towards its goals, in this case, the goal of a HCAI.

3 Innovation Policy and Artificial Intelligence

In the previous section we came to the conclusion that it is unlikely that HCAI will be forthcoming automatically from the market. This requires government intervention, and in particular, through steering the incentives for the private sector to develop and disseminate

AI models that conform to the ethical and responsibility requirement of HCAI. For this, innovation policy will be required. In this section we discuss how directional innovation policies may be appropriate and effective, what the contribution of one specific innovation policy – innovation procurement – may be, and finally how public procurement and innovation, and specifically public procurement for innovation may be utilised.

3.1 Directional innovation policies and AI

According to Edler and Fagerberg (2017, p.4) "Innovation is understood as the introduction of new solutions in response to problems, challenges, or opportunities that arise in the social and/or economic environment." According to Bloom et al. (2019, pp.167-168) the empirical evidence from the literature suggests that social returns to innovation "are much higher than private returns, which provides a justification for government-supported innovation policy." More generally, an innovation such as AI generates both positive and negative externalities and markets do not adequately capture these in market prices. As such there is, as Korinek (2019, p.4) points out "no theoretical reason to believe that the free market will direct innovative efforts to the most socially desirable innovations. . . . the market may thus guide innovation in the wrong direction." AI innovation will thus not automatically result in HCAI.

The implication is that HCAI is a challenge for innovation policies. Innovation policies are the "public interventions to support the generation and diffusion of innovation, whereby an innovation is understood as the transformation of an invention into marketable products and services, the development of new business processes and methods of organization, and the absorption, adaptation and dissemination of novel technologies and know-how" (WTO, 2020, p.24).

There are many innovation policy tools aimed at the "generation and diffusion of innovation," from both the supply and demand side.[15] Tools such as R&D subsidies and tax credits have typically been described as supply-side tools (Edler and Fagerberg, 2017), whist regulations and public procurement have been described as demand-side tools (Aschhoff and Sofka, 2009). In recent years there has been subtle shift towards the increasing use of demand-side tools, in particular as these are seen as being potentially useful for industrial policy purposes and to facilitate economic restructuring (Crespi and Guarascio, 2019; Edler and Fagerberg, 2017). This growing use is part of what is termed the "directional turn" in innovation policy – not to encourage innovation for the sake of innovation, but for addressing a pressing societal

[15]See e.g. Edler and Fagerberg (2017) who presents and discuss 15 major innovation policy instruments.

need or challenge (Miller and Lehoux, 2020).

Present concerns with directional innovation and industrial policies are based on recognition that government intervention was in the past important for the development of important technologies, such as "jet engines, radar, nuclear power, the Global Positioning System (GPS), and the internet" (Bloom et al., 2019, p.166). Clemens and Rogers (2020) studied how government procurement influenced innovation in prosthetics during the U.S. Civil War and World War I, conclude that the nature of government procurement can influence significantly whether innovation will be on costs or quality – which is a finding with much relevance for the challenge of HCAI which is primarily a challenge relating to the quality of the technology. Another oft-quoted example of government steering the development of new technologies is of the USA's efforts during the Second World War, which started with the creation of the National Defense Research Committee (NDRC) in 1940. These efforts resulted in many impactful new technologies and their diffusion during and after the war, including radar, mass-produced penicillin, radio communications, and pesticides such as DDT (Gross and Sampat, 2020). After the Second World War, three notable innovations where government incentives were important, included the creation of the internet through the USA's DARPA-program,[16] the creation of modern biomedical research through funding grants in the UK, and the creation of Google's search engine algorithm by a government grant (Lenderink et al., 2019). Interesting discussions of the specific role of government procurement in the establishment of the computer industry, the semi-conductor industry, and the US commercial aircraft industry is contained in (Geroski, 1990).

Azoulay et al. (2018, pp.69-70) discusses the USA's DARPA-program as a catalyst for many modern-day technologies, not only the internet but also such as the personal computer, lasers and Microsoft Windows. DARPA's approach may be argued to be eminently suitable to the challenge facing the development and diffusion of HCAI. Fouse et al. (2020) discuss how DARPA, through its various initiatives, had an "outsized" influence on the development of AI, and continues to do so. For instance, DARPA is at the time of writing reported to be "investing more than US $2 billion in AI through its initiative, AI Next" (Fouse et al., 2020, p.4).

As discussed by Azoulay et al. (2018) DARPA's projects typically are focused on generating new technological solutions where three characteristics of the technology and markets overlap, namely 1) there has to be a societal challenge; 2) the technology must be new and on the

[16] DARPA is the acronym for the USA's Department of Defense's "Defense Advanced Research Projects Agency" which was launched in 1958 and first called the Advanced Research Projects Agency (ARPA) (Azoulay et al., 2018).

beginning of the technology *S-curve*, and 3) there are significant frictions in the markets for ideas and technology that will hinder a spontaneous market solution. In the case of HCAI, all three these conditions are present, as it presents a significant societal challenge, the technology is new (in particular the alignment with human values) and the high degree to which the technology, being data-based, are subject to knowledge spillover and data network effects. These three conditions are also amenable to be influenced through public procurement for and of innovation as we will discuss in greater detail in the next sub-section. For instance, public procurement can effectively stimulate from the demand-side the production of socially desirable technologies (Miller and Lehoux, 2020); it is better suited to technologies that are still early in the product-life cycle (such as AI) (Geroski, 1990); and it overcomes knowledge spillover externalities by creating new markets and networks.

In the next sub-section, we discuss the emerging literature on public procurement and innovation.

3.2 Public procurement and innovation

Public procurement refers to "the direct purchase of goods and services by the public sector" (Crespi and Guarascio, 2019, p.783). The potential impact of public procurement is enormous. Gerdon and Molinari (2020) document for instance that there are more than 250,000 public authorities in the EU alone, who spent more than EUR 2 trillion annually on procurement.

Public procurement in general, has long since been a tool to promote innovation (and industrialization). Most often in the past, it has been used *ex post* to support new innovations by creating a demand for the product (being a customer for new products). This is also the way in which Lin (2020, p.26) conceives of the relationship between public procurement for innovation support. However, this is a rather restricted view: through the more targeted approaches of public procurement for and of innovation, governments can also *ex ante* support research and development of goods and services that do not yet exist – for present purposes, HCAI solutions. In this government can fulfil the function of a lead user role in innovation, as for instance argued by von Hippel (1976), and it can also create new networks and assist in the diffusion of innovations for instance by helping SMEs to adopt and use AI[17]

[17]In the case where the public sector procures a new innovation for its own use, it is termed *intrinsic procurement* and when it is procured for use outside of the government, it is termed *extrinsic procurement* (Czarnitzki et al., 2020).

Czarnitzki et al. (2020). There has been a growing interest in recent years by governments to support innovation through public procurement – the WTO (2020, p.67) documents that 81 percent of OECD countries have adopted initiatives to stimulate innovation through public procurement.

Public procurement and innovation (PPaI) are, as we explained in the introduction, a term encompassing three modes of innovation from the perspective of public procurement. As discussed by Obwegeser and Müller (2018, p.5), the first is public procurement for innovation (PPfI), which deals with the question, how can public procurement drive innovation? This refers broadly to all public procurement initiatives aimed at innovation. There are many types and modalities of public procurement for innovation depending on focus, use, output, and interaction with suppliers (see Lenderink et al. (2019)). A full discussion of the perturbations and typologies of procurement of and for innovation falls outside the scope of this paper – the reader is referred to Lenderink et al. (2019) for an extensive overview.

For the present purposes though, public procurement for innovation includes procurement of products and services that do not yet exist. In the European Union the explicit use of public procurement in order to procure such products or services is relatively recent (Czarnitzki et al., 2020). Edler and Georghiou (2007) relate how the public procurement for innovation, after having been neglected in the EU, got onto the EU's agenda around 2003. This eventually resulted in the European Commission's *Handbook on Public Procurement for Innovation* in 2007. One of the innovations in public procurement that the EU introduced herein was the legal instrument of *Pre-Commercial Procurement of innovation* (PCP). PCP "concerns the procurement of R&D services prior to commercialisation, where new solutions for a specific social need or challenge are developed in competition with risk-benefit sharing between the public organisation and potential suppliers" (Lenderink et al., 2019, p.8). PCP is "a competitive and selective" instrument, as firms have to compete and come up with the best solution on their own (Aschhoff and Sofka, 2009, p.1236), and it "emphasizes the need for new development of a solution to solve a specific requirement" (Obwegeser and Müller, 2018, p.11).

The second mode of innovation from the perspective of public procurement is public procurement of innovation (PPoI). This deals with the question, how can public services be innovated? For instance, government departments such as health, education and energy, to name but a few, could increase efficiency through adopting AI solutions, for instance AI diagnostic tools for hospitals and energy monitoring systems for reducing emissions from government buildings. A third mode of innovation from the perspective of public procurement is innovative public procurement (IPP). It deals with the question how can public institutions procure innovatively? The demands of steering societal outcomes through public procurement, including to steer advanced technologies such as AI, will require that public procurement itself be conducted in a more advanced manner – that the public sector innovate itself in how it procures innovation. This is not least due to the complexity of AI and the need for government agencies to understand the technology, but also due to the need to continually monitor and upgrade AI technologies, which are never static, a result of AI models constantly learning and changing behavior as they get exposed to more data.

PPaI have been shown to be an effective tool to steer innovation.[18] Czarnitzki et al. (2020) studies the case of Germany, where the government already since 2009 allowed for public procurement of innovation. They find from an analysis of 3410 cases, between 2010 to 2012, that the use of this instrument by the German government "increased turnover with new products and services in the German business sector by EUR 13 billion in 2012, which represents 0.37% of GDP. Standard procurement tenders without innovation-related components, by contrast, show no detectable relationship" (Czarnitzki et al., 2020, p.2). Moreover, based on calculations from German public procurement and R&D spending, they conclude that "The quantitative potential of public procurement of innovation is about ten times larger than the amount of public R&D subsidies distributed to the business sector" (Czarnitzki et al., 2020, p.3). Other interesting studies of public procurement and innovation include Miller and Lehoux (2020) who study the use of public procurement of innovation in the Canadian healthcare sector.

We can also report examples of how government procurement is already having an impact on the development and diffusion of AI. Simonite (2020) reports that in the USA, the Centers for Medicare and Medicaid Services (CMS) will facilitate the dissemination of particular AI models for use in medical diagnostics in US hospitals. Through this, the government is paying the health service to use particular AI algorithms. Beraja et al. (2020) studies the case of facial recognition AI in China and notes the significant role that the government's steering of the technology played. They argue from this that because data is central to AI, and furthermore that "the state is a key collector of data" as well as that there are economies of scope from sharing data across firms, that the government is well positioned to stimulate AI innovation. Specifically, they find that "following the receipt of a government contract to supply AI software, firms produce more software both for government and commercial

[18]Geroski (1990, p.189) argues that PPI can be even better than R&D subsidies and tax credits to steer innovation, because "procurement programmes link the production of innovations to their use" while "subsidies focus only on the production of innovations."

purposes when the contract provides access to more government data" (Beraja et al., 2020, p.1).

Finally, Stojčić et al. (2020) argue that public procurement of and for innovation can help build innovation capability, unlike R&D grants and tax credits which depends on existing innovation capability. In the novel case of AI ethical embeddedness, the innovation capability may not be broadly present amongst firms. Therefore, the suitability for public procurement of and for innovation in the context of steering human-centered AI is particularly appropriate.

3.3 How can public procurement and innovation steer AI?

We have noted on a few occasions in the preceding sub-sections why public procurement of and for innovation (PPoI, PPfI) can be relevant and appropriate for the steering of AI towards HCAI. Having established the why of public procurement for and of innovation, we will deal in the rest of this section with the how. We will in particular focus on the use of PCP.

For ordering the discussing it is useful to start with the salient features of PCP, and then indicate what elements the contracting authority should emphasize in the case of steering towards HCAI, where after we will note some additional requirements – innovations in procurement – that may be particularly useful for the success of PCP with respect to human-centered AI.

3.3.1 Pre-Commercial Procurement of Innovation

First, we discuss the salient features of PCP as a tool to steer HCAI. Pre-Commercial Procurement (PCP) for innovation is the very first European Union (EU) legal provision, available to the public sector, to steer and procure innovative solutions, which are not available in the market, and that require R&D activity. As such, it is very relevant to steer innovation towards human-centered AI. The 2007 PCP Communication was followed by additional, clarifying, documents released to specify its correct interpretation and for its implementation. Yet the main message has been clear from the very beginning, namely that PCP can be used to procure R&D services only, needed to develop an innovative solution which is unavailable in the market. More specifically, PCP can be used to develop prototypes, up to few units of the final product, to make sure the solution is as desired by the procurer. However, PCP cannot be used by a contracting authority (CA) to procure the needed quantity of the final product. Indeed, to do so, at the end of PCP the CA should follow-up with a commercial procurement, opening a competitive tendering, or some type of negotiated, procedure within the legal framework defined by the 2014 EU Public Procurement Directive (EC, 2014). In such procurement any eligible firm could compete for the commercial contract, and not just those which participated to the PCP.

The underlying idea, behind the separation of the R&D phase from the commercial procurement phase, is that the European Commission (EC) did not want to restrict the commercial procurement of potential innovative solutions only to PCP participants. As a matter of fact if, while a PCP procedure is taking place, other firms in the market developed interesting new solutions they should be allowed to participate in the commercial procurement of the final product. This would be compliant with the principle of fair treatment with respect to potential participants, with the principle of open competition, in the best interest of CA and of the entire society.

After more than a dozen years from its introduction PCP is gradually diffusing across EU countries, and AI based solutions are naturally being proposed within such purchasing procedures. A distinguishing feature of PCP is that the initial need is formulated directly by the public authorities, as much as the US Department of Defence (DoD) does it with DARPA, which implies that the solutions must satisfy technical but also ethical requirements, in compliance with the mission of the State. For this reason, AI based solutions originated within the PCP would avoid innovations that are potentially harmful for the society. The 2007 Communication elaborates on the fact that PCP should be founded on a risk sharing principle, between companies and CA, concerning the possibility of project failure. Furthermore, EC suggests that as an incentive to participate in the PCP, the Intellectual Property Rights (IPR) behind the new solutions should be left with the companies invited by CA, rather than being appropriated by CA. Though the reason behind the suggestion is clear, with some AI based innovative solutions contracting authorities may consider keeping the IPR, in exchange of a compensation to the companies. This could take place when the contracting authority wants to have under its own control the diffusion and continuous updating of the new product, which in the case of ethical and responsible AI solutions that it wants to disseminate to SMEs and update over time (as was shown to be necessary in section 2.1) would be necessary.

Article 31 of the 2014 Public Procurement Directive the EC introduced an additional legal provision for procuring innovative solutions, the so called *Innovation Partnership* (IP). Unlike

PCP in the IP the contracting authority may not unbundle the procurement of the R&D phase from the commercial phase, that is the procurement of the desired quantity of the final product. As a result, currently public officers in EU have two alternative legal instruments to procure innovative solutions, which have not been developed yet and require R&D activity: PCP and IP.

3.3.2 Using PCP to steer innovation towards HCAI

In what follows we discuss some of the main elements of PCP and case studies to articulate how public procurement can steer AI solutions in EU public sector. In the 2007 Communication the EC suggests a model procedure for PCP composed of three phases to implement the PCP. Contracting Authorities (CA) are by no means mandated to follow such model, but it can certainly be a very helpful benchmark for them. A PCP is typically preceded by one, or more, sessions of market consultation. These sessions are publicly announced and organized by CA to inform all the potentially interested subjects of its needs and of the problem – e.g., the need for an AI solution that is consistent with certain ethical principles - that will be the object of the PCP.

Then a call for manifestation of interest to participate in a PCP for an ethical AI solution is openly announced. After having received the initial manifestation of interest CA can opt to invite a restricted number of companies for the first phase. In such phase CA asks the invited companies to propose an initial draft of the would-be final solution. Once received the proposals, a subset of them will be selected for the second phase where a prototype of the solution will be presented. Then, a subset of firms which submitted a prototype will be invited to the third phase, where the selected companies are asked by CA to produce a few units of the final solution, to check if it works as desired. To prevent the emergence of dominant/unique solutions, the 2007 Communication suggests that at least two companies should be invited to the third phase.

As said, if the solution is satisfactory, and the CA interested in procuring a solution such as the one emerged from the PCP, then after the conclusion of the PCP the contracting authority will have to open a follow-up procurement procedure, for products which no longer requires R&D. For this reason, such procedure would fall under the 2014 Public Procurement Directive, and the CA will now be able to describe precisely in the tender documentation what it wishes to purchase. Of course, in this case, those companies which were able to reach the third phase of the PCP will have some advantage, but no certainty to obtain the contract. Still in this case, since the procured solution would be innovative, according to the EC terminology it will be called PPI, that is public procurement of an innovative solution, which however does not require R&D activity. It is still considered innovation procurement because it purchases an object which was not in the market yet.

The above short description should clarify how AI based innovative solutions could be driven in the desired direction by the public sector, in all the PCP phases. From the original formulation of the problem, to the gradual screening and selection of proposals taking place across the three phases, until PCP ends, CA can navigate the whole process towards solutions which are beneficial for the society. In this navigation process, a number of elements are particularly relevant for human-centered AI solutions – and some of them may require innovation in public procurement.

The first is that the CA should from the outset approach the use of PCP for an ethical, human-centered AI solution with the intention to foster cooperation between firms and research organizations, as well as and foster multi-disciplinary cooperation. According to Askell et al. (2019, p.1) "competition between AI companies could decrease the incentives of each company to develop responsibly by increasing their incentives to develop faster. As a result, if AI companies would prefer to develop AI systems with risk levels that are closer to what is socially optimal- as we believe many do- responsible AI development can be seen as a collective action problem." As said, the PCP benchmark is based on a procedure with sequential selection of companies, hence in principle competitive. Yet, despite this, in using PCP, the CA can design calls for HCAI solutions that requires and facilitates such cooperation. Askell et al. (2019) discuss various ways in which this can be incentivized, for instance by creating high shared gains from cooperation and penalizing non-cooperation. Finally, if the PCP model turns out to be unsuitable for convincingly establish cooperative behaviour among invited firms, then the Innovation Partnership could be used instead as legal instrument.

The second is that the CA should insist on ethical design approaches be followed through all the stages of the procurement process. In this respect Crawford and Calo (2016) proposed two broad approaches to the ethical design of AI, which may also be applied by a CA using PCP. The first is to require developers of AI solutions to use frameworks such as value sensitive design and responsible innovation, and the second is to require a social-systems approach to building AI, which would entail consideration of "the social and political history of the data" (Crawford and Calo, 2016, p.313). The latter approach would be in line with the need for data justice as discussed below.

The third is that the CA should promote greater openness in AI development – for instance by requiring open scrutiny of solutions that it purchases and requiring certain degrees of openness and outside participation in and evaluation of the solutions developed as part of the procurement action. Bostrom (2017) concludes that over the medium-term, greater openness in AI development would be unambiguously positive by speeding up the development and diffusion of AI.[19] He is more concerned that over the long run, the openness would discourage innovation to the extent that the effort may be motivated by having a monopoly in the market from a proprietary innovation. However, through PPI this long-run disincentive effect of openness may be compensated for by government not only be reducing the costs and risks of developing a new AI solution, but also offering a firm the chance to build its capacity for developing and using AI, and even for a company to benefit in complementary areas to the AI being procured (Bostrom, 2017).

Fourth, and related to the need for openness, is the requirement that the PCP call specifically take care of the reliance of AI models on large datasets. Beraja et al. (2020) for instance distinguish between "data-rich versus data-scarce government contracts." We want to argue that PCP contracts for HCAI should be of the data-rich variety. A data-rich PCP contract should in particular address the *data parity problem* (Calo, 2017) which refers to the fact that "a few large firms have disproportionate access to big data necessary to train AI models. Mostly, small businesses do not have the scale and scope and resources to gather data and benefit from data network economies. This could mean that machine learning applications will bend systematically toward the goals of profit-driven companies and not society at large" (Calo, 2017, p.19).

The data parity problem is mirrored in society as a digital divide, with data gaps, and data absences creating and perpetuity inequality in access to the use and knowledge of AI, and in benefiting from AI. In recent times the imperative of addressing these as a requirement for HCAI has resulted in the notion of data justice, which goes beyond data gaps and data absences to recognition of structural inequalities. According to Dencik et al. (2019, p.874) "The framework of data justice broadens the terms of the debate in a way that accounts for a host of issues that are compounded in the datafied society, as evidenced in recent scholarship relating to democratic procedures, the entrenchment and introduction of inequalities, discrimination and exclusion of certain groups, deteriorating working conditions, or the dehumanisation of decision-making and interaction around sensitive issues. These

[19] According to Bostrom (2017, p.137) "openness would improve static efficiency, by making products available at marginal cost (e.g., in the form of open-source software) and allowing a given level of state-of-the-art technical capability to diffuse more quickly through the economy."

discussions suggest a need to position data in a way that engages more explicitly with questions of power, politics, inclusion and interests, as well as established notions of ethics, autonomy, trust, accountability, governance and citizenship." The discussion of data justice and the relevance of societal concerns about fairness, diversity and equity in the application of potential general-purpose technologies such as AI reflects the fact that ultimately, AI is sociotechnical in that the value that it offers "depend on not only technical hardware but also human behavior and social institutions for their proper functioning" (van de Poel, 2020, p.7).

Fifth, the contract authority (CA) should make sure that all documentation[20] and audits are captured, as this "is important to provide an audit trail in case of subsequent issues with the model" (Eitel-Porter, 2020, p.4). This is relevant not only for adjudicating the quality of the solution offered, but to support efforts to continually review the solution so as to ensure that "ethical parameters are not breached over time" (Eitel-Porter, 2020, p.5). This is not so much due to the ethical norms changing, but because of the nature of machine learning (ML), which adapts and changes as it learns more (gathers more data). The important point is that ethical AI solutions may not automatically remain ethical given the possibility of learning, due to AI essentially being an intelligent agent (Riedl, 2019). Anderson and Anderson (2007) refer to this feature of AI to argue that the ultimate challenge is to create explicit ethical machines, which are able to learn and formulate "its own ethical choices based on its own set of principles."

Having ensured that the above five elements are followed, the CA purchasing an AI solution would need finally to be able to verify that the prototype solution offered during stage 3 of the PCP process conforms to the specifications set out in the call – in other words, the technical AI solution needed as well as the requirement that ethical principles are observed. In this regard, it is important to set up adequate measures for verification. As Brundage et al. (2020, p.1) stresses "The capacity to verify claims made by developers, on its own, would be insufficient to ensure responsible AI development. Not all important claims admit verification, and there is also a need for oversight agencies such as governments and standards organizations to align developers' incentives with the public interest." Values[21] in AI models need to be "intended, embodied, and realized values" (van de Poel, 2020, p.5). For present purposes, drawing on van de Poel (2020, p.9) we can state that for a procured AI solution to embody a particular value, denoted by V, it should be the case that at least two conditions

[20]This should include all documentation including AI impact assessments (Gerdon and Molinari, 2020).

[21]See the discussion in van de Poel (2020) on the meaning of values and how to assess whether the desired values are indeed intended, embodied and realized.

be met: first, the AI must be designed for V, and second that in using the AI solution, the values V would be promoted. Thus, AI models need to be designed in order to meet a particular value, for instance human health, and second, utilising the AI model should then indeed promote that value, e.g. human health.

What measures can be used for verification and assessing whether desired values are intended, embodied and realized? Brundage et al. (2020) discuss: third party auditing, red teaming exercises, the piloting of bias and safety bounties, and the sharing of information about AI accidents. In all of these actions, governments can benefit from having independent oversight bodies tracking the work of contractors, and subject these to risk-analyses and public scrutiny. Tzachor et al. (2020) advocates the use of "red teaming" by oversight bodies to stress test any AI solutions that contractors may come up with, where "red teaming is a way of challenging the blind spots of a team by explicitly looking for flaws from an outsider or adversarial perspective" (Tzachor et al., 2020, p.366). According to Tzachor et al. (2020) one legacy of the COVID-19 pandemic is the realisation that circumstances can require the application of new AI solutions under extremely short time horizons and with urgency, and that "doing ethics with urgency" becomes important. Such doing ethics with urgency will require according to the authors that government should be able to "rapidly conduct robust testing and verification of systems." (Ibid, p.366)

Even if these various measures and methods for verification and assessment is successful in finding that the conditions for the AI solution to embody human-centered values are met, a further problem is that as was mentioned, AI models may have unintended results. Moreover, value may change over time. To deal with such unintended results and changes in values, require continuous involvement and monitoring by the contracting authority. According to van de Poel (2020) the design of technologies can never remain static, and they may need to be redesigned at some future stage in light if possible unintended effects as value as possible shifts in values.

At this stage, once the CA is satisfied with the final prototype of the HCAI solution from the PCP process, it will have to ensure that the public sector more broadly has in place complementary mechanisms for dealing with remaining safety risks posed by AI. As we pointed out in section 2.2.1, normal accidents are one of the salient risks posed by AI. It is unfortunately the case that the risks of accidents "cannot simply be 'designed out" of the technology -at least not without giving up on many of their benefits" (Maas, 2018, p.5). Maas (2018) concludes that this requires two complementary approaches: one, that governments ensure that disaster and accident insurance schemes are appropriate to cover AI risks, and second that more research be encouraged into AI safety and AI ethics - see also Anderson and Anderson (2007).

Finally, for the PCP process to be ultimately successful in contributing to innovations in the field of AI, the government and its contracting authorities will need to upgrade and continually invest in its own AI expertise (Calo, 2017). With this they could even use data science and AI techniques within the public sector to explore and test the ethical consequences of AI. For example, Yu et al. (2018) survey various technical solutions to support evaluation of ethical concerns in AI. These include useful resources and initiatives on which the public sector's scientists and advisors could draw, such as the *GenEth* ethical dilemma analyser that can be used to explore ethical dilemmas; the *MoralDM* tool helping to resolve ethical dilemmas through first-principles reasoning and analogical reasoning; and crowdsourcing of self-reported preferences such as through the *Moral Machine* project.[22]

4 A Model of Public Procurement and Responsible AI Innovation

As we outlined in section 2.3, a contracting authority (CA) in the EU could use the provisions of the EU's PCP instrument to procure an AI-solution that conforms to high standards of ethical and responsible AI, as was identified in section 2.1. By issuing an explicit call for such a solution and offering to fund the R&D costs of the successful company, the public sector could incentivize the development and diffusion of human-centered AI (HCAI). This incentive is important, in the absence of which companies may skimp on adequate ethical and safety standards, as we argued in the introduction and section 2.1.

We can now formalize these ideas. First, we can note that in the absence of the public sector's financial incentives, companies may not take the risks to provide the socially optimal AI solution. Let us suppose that a company X is willing, and has the competence, to develop an AI solution which could result in a socially desirable, human-centered AI system. The project requires resources to be invested and success in its development process is intrinsically uncertain. More specifically, if $C > 0$ is the amount of money invested in the project by company X and $\pi(C)$ the success probability, with $\pi'(C) \geq 0$ and $\pi''(C) \leq 0$, then typically $0 \leq \pi(C) < 1$ for any level of C. That is, no matter how large the company's effort, success will remain uncertain.

[22] See `https://www.moralmachine.net`.

Here $\pi(C)$ formalizes the technology available to develop the project and exhibit non-increasing returns of scale.[23] Suppose now that, if development is successful and the project is commercialized that the company's revenue would be R. Therefore, its profit would be a random variable defined as (see Dimitri (2012)):

$$\Pi(C,R) = \begin{cases} R - C & \text{with probability } \pi(C) \\ -C & \text{with probability } 1 - \pi(C) \end{cases} \tag{1}$$

and the related expected profit is

$$E\Pi(C,R) = R\pi(C) - C \tag{2}$$

If the company maximizes (2) then the first order condition is

$$R\pi'(C) - 1 = 0 \tag{3}$$

and the optimal investment level C^* given by

$$\pi'(C^*) = \frac{1}{R} \tag{4}$$

as long as $E\Pi(C^*, R)$ is non-negative. Expression (4) immediately suggests that, due to the concavity of the success probability, the larger the revenue the higher is C^*. However, there is no guarantee that, regardless of the value of R, that the company would invest at all. Indeed, a non-negative expected profit requires that

$$\pi'(0) > \frac{1}{R} \tag{5}$$

which if not satisfied would discourage the company to undertake any investment.

For example, suppose $\pi(C) = a\left(1 - e^{-bC}\right)$ with $0 < a < 1$ and $b > 0$. It is immediate

[23] In reality, several technologies exhibit increasing returns of scale, at least for some investment levels. However, to keep the exposition simple, with no major loss of generality, in what follows we do not consider this possibility.

to see that the larger are both, a and b, the higher is $\pi(C)$ and so the less challenging is the development process of the AI project. Interestingly the parameters a and b play two different roles. Indeed while $a = \lim_{C\to\infty} \pi(C)$, that is it defines the maximum level that success probability can take, b only defines how $\pi(C)$ changes as C changes. This can be seen by considering for example the elasticity of the function $\pi(C)$, namely the percentage change in $\pi(C)$ due to 1% change in C, which is given by

$$\frac{\pi'(C)C}{\pi(C)} = \frac{be^{-bC}C}{(1-e^{-bC})} \tag{6}$$

As this is independent of a, (5) would become

$$ab > \frac{1}{R} \tag{7}$$

which suggests that the company decides to invest if the project is sufficiently rewarding, or technically not too difficult, or both. In this case the optimal level of investment C^* would be

$$C^* > \frac{log(abR)}{b} \tag{8}$$

which, consistently with (7), is positive for $abR > 1$.

However, suppose now that (7) is not satisfied and the company decides not to invest. Failure to develop the AI project may damage the society and so A could consider supporting the company, however only if the ethical level of the solution is sufficiently high. Broadly speaking, suppose that company X has to compete for funding with other companies as per the EU's PCP, and that the ethical level of the solution under development, summarized by the indicator $e \geq 0$, is positively related to the probability $\theta(e)$ of being funded by A. We assume $\theta(e=0) = 0$ if differentiable and $\theta'(e) > 0$ and that $\lim_{e\to\infty} \theta(e) = 1$. Finally, we now assume that the R&D total costs for the company are given by $C + ce$, where ce is total cost due to reaching the ethical level e and $c > 0$ is the marginal cost for doing so. Under these assumptions, in the analysis conducted so far X undertook no, or very low, investment to take care of the ethical features of the solution under development. Therefore, company X needs to decide the level of both C and e, and meticulously document e as per the CA's requirements.

For this, suppose A and X agree to write the following contract: "if the AI project is funded by A, with probability $\theta(e)$, and successfully developed then A pays to X the sum p while if the project fails to be developed then A pays to X the sum of q, with $p > q$. Moreover, the intellectual property rights will remain with X, which will still enjoy reward R". As we shall see below, such admittedly simple contract will induce the company to try developing a solution that contains consideration of the ethical requirements.

Before proceeding, it is worth noticing that conditional on A funding X the above contract operates as a mixture of pull $(p - q)$ and push (q) incentives. That is, although we assume A will only pay at the end of the project, as it may be with the EU's PCP, q are guaranteed to X and it is as if they would be paid upfront, hence as a push incentive. For this reason we could also imagine that should X lack resources for engaging in the process then A may indeed pay upfront the sum q. Nonetheless, the additional sum $p - q$ will be paid only conditionally upon success of the project, hence as a pull incentive.

Then the company's profit $\Pi(C, R, p, q, c, e)$ is defined as the following random variable:

$$\Pi(C, R, p, q, c, e) = \begin{cases} (p + R) - C - ce & \text{with probability } \theta(e)\pi(C + ce) \\ q - C - ce & \text{with probability } \theta(e)(1 - \pi(C + ce) \\ 0 & \text{with probability} 1 - \theta(e) \end{cases} \tag{9}$$

Therefore, its expected profit will be

$$E\Pi(C, R, p, q, c, e) = \begin{cases} \theta(e)\pi(C + ce)(p + R - q) + \theta(e)(q - C - ce) & \text{if } e > 0 \\ 0 & \text{if } e = 0 \end{cases} \tag{10}$$

and the optimal level of investment C^{**} chosen by the firm would now be such that

$$\pi'(C^{**} + ce) = \frac{1}{(R + p - q)} \tag{11}$$

Condition (11) looks like (4) except that the right-hand side now has the additional term $p - q > 0$ at the denominator. This basically increases X's expected revenues providing a stronger incentive for the company to invest resources and engage in developing a solution, provided it is ethically acceptable to A. This will now take place if

$$\lim_{(C+ce)\to 0} \pi'(C+ce) = ab > \frac{1}{R+p-q} \tag{12}$$

Condition (12) indicates that A could always find a large enough amount $p-q$ such that (5) is satisfied. Moreover, notice that on the left hand side of (12) we are taking $\lim\limits_{(C+ce)\to 0} \pi'(C+ce)$ because expression (10) is equal to 0 for $e = 0$, and so $\pi'(0)$ would not even be defined. Moreover, with an acceptable estimation of $\pi(C)$, authority A, from (11), could also determine the effort $(C^{**} + ce)$ of X.

Considering again the previous example $\pi(C) = a(1 - e^{-bC})$, the optimal investment level for X would now be

$$C^{**} + ce = \frac{log(ab(R+p-q))}{b} \tag{13}$$

If $C^{**} + ce > 0$ it is interesting to notice that $\frac{log(ab(R+p-q))}{b} < p-q$, namely the company will invest less than the additional incentive term $p-q$, for all levels of $p-q$ if $abR < 1$, that is if the company was not investing at all before A's intervention. However, if $abR > 1$ then the inequality would be true for $p-q > z$, with z solving $\frac{log(ab(R+z))}{b} = z$.

So, the element $p-q$ will work as a co-funding term for X to engage in the AI project. To summarize, what is important for A to motivate X to invest more in R&D activities are not the values of p and q separately, but rather only their difference. Indeed, this is true even if $q = 0$, as long as p is large enough.

It is worth stressing that the optimal level of R&D investment obtained from (13) must include the component $ce > 0$, needed by X to receive funding from A, which represents the expenditure undertaken by X to introduce ethical features into the solution. The desirable level of the indicator e resolves a trade-off between increasing the probability of being funded by A, as well as of finding the innovative solution, and of increasing the costs. It would be found by maximising (10) also with respect to e.

To summarise, the above equations describe a simple framework where public procurement can induce HCAI solutions, by conditioning the funding of a company to the presence of ethical features, for example in the outcome of a PCP as well as of Innovation Partnership procedures. This use of public procurement to promote HCAI as we modelled it here, is complementary to the approach in Naudé and Dimitri (2020), who model the use of public procurement to reduce the likelihood of arms races for AI.

The mechanism that we outlined above for incentivizing R&D for HCAI is relevant not only for addressing the general reluctance of companies to invest in HCAI, but more generally the fact that the concerns about technologies such as AI's have led to increasing calls in recent years to make "Technology Impact Assessments" (TIAs) obligatory for entrepreneurs before implementing and investing in new technologies (Korinek, 2019). This will have particularly negative effects however, on innovation, over and above the impact of the risks already illustrated here. It may be worth pointing out that one reason why innovation in the digital economy have not been declining as innovation elsewhere, is due to the fact that until now, such innovation has been relatively unencumbered by bureaucratic influence, i.e., it has been rather more "permissionless." Rather than risk dis-incentivizing innovation, we want to argue that PCP, organised for example within a framework such as the one expressed by equations (9) and (10), can obtain the same results, but through rewarding entrepreneurs who do perform technology impact assessments. In other words, PCP may reduce the fixed costs which a requirement such as TIA can cause, by requesting the solution to incorporate the requested features, possibly co-funding the engaged companies for this.

Finally, given the potential of PCP to steer HCAI as this simple model show, we can conclude this section by pointing out that although public procurement for innovation as instrument to steer AI is still relatively neglected, there are already a small sample of PCP AI based undertaken in EU since 2007 when PCP was launched. It is useful to briefly mention these projects, and to recommend further research to document the lessons learned from them.

A first example is that during the years 2015-16 a consortium of five hospitals, in different European countries implemented an EU funded PCP project to develop a highly interoperable, AI based, telemedicine system for the cockpit of intensive care units in hospitals. The project, called *THALEA*, was successfully completed and as a result the Dutch Company called *NewCompliance*, which developed its solution for the operating room cockpit, was able to attract venture capital and sell the solutions to several hospitals in EU and in the US. Moreover, it established partnerships with some major companies in the field to improve the solution, scaling up its business volume and size.

Among the PCP AI projects which have not been funded by the EU, a very timely and successful project is the PCP sponsored by the Danish Market Development Fund for self-driving-robots, based on ultraviolet light, to disinfect hospitals. The project was initially promoted in 2014, and completed in 2017, by a group of Danish hospitals, hence much earlier than 2020 when the COVID-19 pandemic hit the world. As a result of the PCP project, the Danish company *Blue Ocean Robotics* attracted millions of dollars in venture capital, and sold its robots across several countries, to be used in hospitals across the world to help disinfecting hospitals during the COVID-19 pandemic.

Another interesting example, of a non-EU funded PCP, is an AI-based smart mobility PCP project developed in the period 2014-2016 and promoted by the Dutch Province of North-Brabant, to try solving shock-wave traffic jams problems in the area. Traffic jams often occur when a vehicle suddenly stops, inducing also the following vehicles to suddenly use their breaks even more strongly, originating traffic jams. As the outcome of PCP, the company *Be-Mobile* produced an application merging and analysing real time data on traffic, reported by various sources, data on vehicles speed, and to advise drivers on which roads/lanes to take. This innovative solution has been so successful that *Be-Mobile* was able to meaningfully scale-up in terms of employers and raised funds.

5 Concluding Remarks

There are high expectations of AI as a general-purpose technology that can contribute towards economic growth, innovation and sustainable development. There are, however, also concerns about the negative consequences and uses of AI. These negative consequences of AI have given rise to calls for public policy to ensure that AI-based applications are safe, and that the development and diffusion of AI does not suffer from improper use and misuse. What would count as improper use and misuse, and safety-risks such as that could be caused by accidents or unintended consequences of AI systems, would be determined by the extent to which AI systems are human-centered. Human-centered AI (HCAI) draws on ethical principles and puts forth actionable guidelines for reducing the risks mentioned. As such, HCAI is concerned with Ethical AI and Responsible AI.

In recent years there have been a growing number of proposals for Ethical and Responsible AI. A shortcoming of these is that they lack strong incentives for developers and users of AI to adhere to them. In this paper, we argued that public procurement and innovation ($PPaI$) is a potentially relevant tool with which to incentivize HCAI. Indeed, if the governance of AI deals with "how decisions are made about AI," and what "institutions and arrangements could help in this decision-making (Dafoe, 2018, 2020), then public procurement and innovation as presented in this paper is a key element of AI governance – however an element that we argued is still neglected.

We provided a simple theoretical model to show that the procurement for an innovative AI solution, underpinned by ethical considerations, can provide both the tools and audit trails as well as the financial incentives necessary to steer AI towards HCAI. Additionally, we highlighted five elements that the EU's PCP instrument should specifically take into account when aiming to steer HCAI, as well as complementary initiatives, such as supporting more multi-disciplinary research into ethics and into making ethics computable. This could inform further innovations in public procurement itself.

In this, we believe our paper made a number of contributions to the literature. First, our argument and model illustrating that public procurement for innovation can promote HCAI adds to the literature on innovation policy. This potential instrument for steering HCAI has been neglected in both the AI as well as innovation literature. Our paper contributed to addressing this neglect by providing arguments based on an analysis of the literature as well as providing a theoretical model. Secondly, our paper contributed to the literature on innovation policy by locating the role of public procurement for innovation of ethical, HCAI within the recent literature on mission-oriented and challenge-led innovation policies. Finally, we think our paper also contributed to the emerging literature of public entrepreneurship, in that the use of public procurement of innovation is an example of public entrepreneurship as it consists of innovation, it aims to contribute towards transformation of society, and it is subject to uncertainty.

Acknowledgements

We are grateful to the participants of various workshops of the Research Unit for the Diffusion of Quality AI at Paderborn University for their comments on earlier versions of this paper, and in particular to Thomas Gries and Margarete Redlin. The financial assistance of the *Volkswagen Stiftung*, through planning grant *AZ 97042* from their project on *Artificial Intelligence and the Society of the Future* is gratefully acknowledged. The usual disclaimer applies.

References

AI and Inclusion Symposium (2017). An Evolving Reading List. *Available at https://drive.google.com/file/d/0ByG0FdgytPz7YkxRMHF1d19ibE0/view.*

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete Problems in AI Safety. *ArXiv:1606.06565 [Cs]*, June:http://arxiv.org/abs/1606.06565.

Anderson, M. and Anderson, S. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, 28(4):15.

Armstrong, S., Bostrom, N., and Schulman, C. (2016). Racing to the Precipice: A Model of Artificial Intelligence Development. *AI & Society*, 31:201–206.

Aschhoff, B. and Sofka, W. (2009). Innovation on Demand: Can Public Procurement Drive Market Success of Innovations? *Research Policy*, 38(8):1235–1247.

Askell, A., Brundage, M., and Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. *ArXiv*, July.

Azoulay, P., Fuchs, E., Goldstein, A., and Kearney, M. (2018). Funding Breakthrough Research: Promises and Challenges of the ARPA Model. *Innovation Policy and the Economy*, 19:69–96.

Baker-Brunnbauer, J. (2020). Management Perspective of Ethics in Artificial Intelligence. *AI and Ethics*, 11 September:https://doi.org/10.1007/s43681–020–00022–3.

Beraja, M., Yang, D., and Yuchtman, N. (2020). Data-Intensive Innovation and the State: Evidence from AI Firms in China. *NBER Working Paper No. 27723, National Bureau of Economic Research.*

Bloom, N., van Reenen, J., and Williams, H. (2019). A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives*, 33(3):163–184.

Boddington, P. (2017). Towards a Code of Ethics for Artificial Intelligence. *Heidelberg: Springer.*

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. *Oxford: Oxford University Press.*

Bostrom, N. (2017). Strategic Implications of Openness in AI Development. *Global Policy*, pages 1–14.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., and et al (2020). Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims. *ArXiv [cs.CY]*, April:https://arxiv.org/abs/2004.07213.

Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. *SSRN*, At https://ssrn.com/abstract=3015350.

Castelvecchi, D. (2016). The Black Box of AI. *Nature*, 538:21–23.

Clemens, J. and Rogers, P. (2020). Demand Shocks, Procurement Policies, and the Nature of Medical Innovation: Evidence from Wartime Prosthetic Device Patents. *NBER Working Paper No. 2667, National Burea of Economic Research*.

Crawford, K. and Calo, R. (2016). There is a Blind Spot in AI Research. *Nature*, 538:311–313.

Crespi, F. and Guarascio, D. (2019). The Demand-Pull Effect of Public Procurement on Innovation and Industrial Renewal. *Industrial and Corporate Change*, 28(4):793–815.

Czarnitzki, D., Hünermund, P., and Moshgbar, N. (2020). Public Procurement of Innovation: Evidence from a German Legislative Reform. *International Journal of Industrial Organization*, 71:1–16.

Dafoe, A. (2018). AI Governance: A Research Agenda. *Centre for the Governance of AI, University of Oxford*.

Dafoe, A. (2020). AI Governance: Opportunity and Theory of Impact. *Effective Altruism Forum*, 17th September.

Dencik, L., Hintz, A., Redden, J., and Treré, E. (2019). Exploring Data Justice: Conceptions, Applications and Directions. *Information, Communication & Society*, 22(7):873–881.

Dimitri, N. (2012). R&D Incentives for Neglected Diseases. *Plos One*, 7(12):e50835.

EC (2014). On the Coordination of Procedures for the Award of Public Works Contracts, Public Supply Contracts and Public Service Contracts. *European Commission Directive 24. Brussels.*

EC (2019). Ethics Guidelines for Trustworthy AI: High-Level Expert Group on Artificial Intelligence. *European Commission, Brussels.*

Edler, J. and Fagerberg, J. (2017). Innovation Policy: What, Why, and How. *Oxford Review of Economic Policy*, 33(1):2–23.

Edler, J. and Georghiou, L. (2007). Public Procurement and Innovation - Resurrecting the Demand Side. *Research Policy*, 36:949–963.

Eitel-Porter, R. (2020). Beyond the Promise: Implementing Ethical AI. *AI and Ethics*, 6 October(https://doi.org/10.1007/s43681-020-00011-6).

Etzioni, A. and Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. *Journal of Ethics*, 21(4):403–418.

Everitt, T., Lea, G., and Hutter, M. (2018). AGI Safety Literature Review. *International Joint Conference on Artificial Intelligence (IJCAI)*, arXiv: 1805.01109.

Feldstein, S. (2019). The Global Expansion of AI Surveillance. *Working Paper, Carnegie Endowment for International Peace*, Sept.

Feuerriegel, S., Dolata, M., and Schwabe, G. (2020). Fair AI: Challenges and Opportunities. *Business and Information Systems Engineering*, 62(4):379–384.

Floridi, L. (2018). The Ethics of Artificial Intelligence. *In Franklin, D. ed. Megatech: Technology in 2050. London: Profile Books. Chapter 13, pp. 155-163.*

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., and Vayena, E. (2018). AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28:689–707.

Fouse, S., Cross, S., and Lapin, Z. (2020). DARPA's Impact on Artificial Intelligence. *AI Magazine*, Summer:3–9.

Frey, C. and Osborne, M. (2017). The Future of Employment: How Susceptible are Jobs to Computerization? *Technological Forecasting and Social Change*, 114:254–280.

G-20 (2019). G20 Ministerial Statement on Trade and Digital Economy. 9 May.

Gerdon, S. and Molinari, V. (2020). How Governments can use Public Procurement to Shape the Future of AI Regulation - and Boost Innovation and Growth. *World Economic Forum Blog*, 8th June.

Geroski, P. (1990). Innovation, Technological Opportunity, and Market Structure. *Oxford Economic Papers*, (42):586–602.

Gross, D. and Sampat, B. (2020). Inventing the Endless Frontier: The Effects of the World War II Research Effort on Post-War Innovation. *Harvard Business School Strategy Unit Working Paper No. 20-126*.

Hagendorff (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines*, 30:99–120.

Hayter, C. S., Link, A., and Scott, J. (2018). Public-Sector Entrepreneurship. *Oxford Review of Economic Policy*, 34(4):676–694.

IEEE (2019). Ethically Aligned Design. *1st Edition at* `https://ethicsinaction.ieee.org`.

Korinek, A. (2019). Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence. *NBER Working Paper no. 26130, National Bureau of Economic Research*.

Korinek, A. and Stiglitz, J. (2017). Artificial Intelligence and its Implications for Income Distribution and Unemployment. *NBER Working Paper no. 24174. National Bureau for Economic Research*.

Leike, J., Martic, M., Krakovna, V., Ortega, P., Everitt, T., Lefrancq, A., Orseau, L., and Legg, S. (2017). AI Safety Gridworlds. *ArXiv:1711.09883 [Cs]*, November 27:http://arxiv.org/abs/1711.09883.

Lenderink, B., Johannes, I., and Voordijk, H. (2019). Innovation and Public Procurement: From Fragmentation to Synthesis on Concepts, Rationales and Approaches. *Innovation: The European Journal of Social Science Research*, (https://doi.org/10.1080/13511610.2019.1700101).

Lin, J. (2020). Industrial Policy Revisited. *In WTO, World Trade Report 2020: Government Policies to Promote Innovation in the Digital Age. Geneva. World Trade Organization*, page 26.

Maas, M. (2018). Regulating for Normal AI Accidents: Operational Lessons for the Responsible Governance of Artificial Intelligence Deployment. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 223–228.

Mazzucato, M. (2013). The Entrepreneurial State: Debunking the Public vs. Private Myth in Risk and Innovation. *Anthem Press: London, UK*.

Mazzucato, M. (2018). Mission-Oriented Innovation Policies: Challenges and Opportunities. *Industrial and Corporate Change*, 27(5):803–815.

Mazzucato, M. (2020). Mission-Oriented Innovation and Industrial Policies. *In WTO, World Trade Report 2020: Government Policies to Promote Innovation in the Digital Age. Geneva. World Trade Organization*, page 101.

Miller, F. and Lehoux, P. (2020). The Innovation Impacts of Public Procurement Offices: The Case of Healthcare Procurement. *Research Policy*, 49:1–13.

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, 27(4):87–91.

Naudé, W. (2021). Artificial Intelligence: Neither Utopian nor Apocalyptic Impacts Soon. *Economics of Innovation and New Technology*, 30(1):1–23.

Naudé, W. and Dimitri, N. (2020). The Race for an Artificial General Intelligence: Implications for Public Policy. *AI & Society*, 35(2):367–379.

Nguyen, A., Yosinski, J., and Clune, J. (2014). Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. *ArXiv*, https://arxiv.org/abs/1412.1897.

Obwegeser, N. and Müller, S. (2018). Innovation and Public Procurement: Terminology, Concepts, and Applications. *Technovation*, 74-75:1–17.

OECD (2019). Recommendation of the Council on Artificial Intelligence. *OECD/LEGAL/0449, Adopted 22 May 2019. Paris: Organization for Economic Cooperation and Development*.

Riedl, M. (2019). Human-Centered Artificial Intelligence and Machine Learning. *Human Behavior and Emerging Technologies*, 1(1):33–36.

Russel, S., Dewey, D., and Tegmark, M. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine, Association for the Advancement of Artificial Intelligence*, Winter:105–114.

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, 36(6):495–504.

Simonite, T. (2020). The US Government Will Pay Doctors to Use These AI Algorithms. *Wired Magazine*, 11 October.

Smith, M. L. and Neupane, S. (2018). Artificial Intelligence and Human Development: Towards a Research Agenda. *White Paper : International Development Research Centre.*

Stanford University (2016). Artificial Intelligence and Life in 2030. *Stanford University at* ***https://ai100.stanford.edu***, September.

Stojčić, N., Srhoj, S., and Coad, A. (2020). Innovation Procurement as Capability-Building: Evaluating Innovation Policies in Eight Central and Eastern European Countries. *European Economic Review*, 121:1–18.

Trajtenberg, M. (2018). AI as the Next GPT: A Political-Economy Perspective. *NBER Working Paper no. 24245. National Bureau for Economic Research.*

Tzachor, A., Whittlestone, J., Sundaram, L., and ÓhÉigeartaigh, S. (2020). Artificial Intelligence in a Crisis Needs Ethics with Urgency. *Nature Machine Intelligence*, 2:365–366.

Uyarra, E., Flanagan, K., Magro, E., and Zabala-Iturriagagoitia, J. (2017). Anchoring the Innovation Impacts of Public Procurement to Place: The Role of Conversations. *Environmental Planning*, C35:828–848.

Van de Gevel, A. and Noussair, C. (2013). The Nexus between Artificial Intelligence and Economics. *Springer Briefs in Economics.*

van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*, September:https://doi.org/10.1007/s11023–020–09537–4.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S., Tegmark, M., and Nerini, F. F. (2020). The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *Nature Communications*, 11(233).

von Hippel, E. (1976). The Dominant Role of Users in the Scientific Instrument Innovation Process. *Research Policy*, 5(3):212–239.

WTO (2020). E-commerce, Trade and the COVID-19 Pandemic. *Information Note . Geneva: World Trade Organization*, 4 May.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V., and Yang, Q. (2018). Building Ethics into Artificial Intelligence. *ArXiv*, 7 Dec.