Delft University of Technology

The visible hand of innovation policy

Cantner, Uwe; Werker, Claudia

DOI
10.4324/9780429398971-31

Publication date
2023

Document Version
Final published version

Published in
Routledge Handbook of Evolutionary Economics

Citation (APA)
Cantner, U., & Werker, C. (2023). The visible hand of innovation policy. In *Routledge Handbook of Evolutionary Economics* (pp. 369-380). Taylor & Francis. https://doi.org/10.4324/9780429398971-31

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher
is the copyright holder of this work and the author uses the
Dutch legislation to make this work public.

28

THE VISIBLE HAND OF INNOVATION POLICY

Uwe Cantner and Claudia Werker

28.1 Introduction

While artificial intelligence (AI) has been a major game changer in technological, economic, and societal development (OECD, 2019), questions on how agency and power are distributed between human and artificial intelligence have not been addressed conclusively so far (Willson, 2016). A case in point is innovation policy, because – when it comes to AI – innovation policy focuses on implementing the usual measures leaving any ethical questions to expert councils (OECD, 2019). The problem with this line of action is that ethical questions involve decisions on values, i.e. "… things worth striving for" (Taebi, Correljé, Cuppen, Dignum, & Pesch, 2014, p. 119) and even more so on shared values which require to integrate the values of all relevant stakeholders (Werker, 2021). And, it is completely unclear how experts can be sufficiently legitimized to decide on shared values. In contrast, policy makers do have the legitimacy to define values and shared values, because they represent the elected government and because they bring together various stakeholders either in formal or informal settings.

Innovation policy measures targeting AI have been widely adopted in the OECD countries (see for this and the following OECD, 2019). Yet they have been mostly motivated by the systems approach. Innovation policy from a systemic perspective deals with system failures such as missing innovative agents, relationships, or institutions (e.g. Edquist, 2011; Klein Woolthuis, Lankhuizen, & Gilsing, 2005). Using the system failures approach for AI means that measures have included the stimulation of AI research, the fostering of AI talent, the support of the development and adoption of AI solutions as well as of AI-driven businesses (OECD, 2019).

While the systems failure approach and related policy broadly acknowledge and address problems stemming from the emergence and use of AI, we suggest that it insufficiently captures three issues constituting AI as a game changer for innovation policy. Each of these issues leads to a research question: The first issue is that innovation policy makers have been outsourcing ethical issues to expert councils (see first paragraph of introduction). Accordingly, we pursue the following research question:

DOI: 10.4324/9780429398971-31

RQ1*How can policy makers ensure that all stakeholders involved in and affected by AI, i.e. being legitimized to have a say about innovation processes and outcomes, contribute to developing shared values that guide technological progress of AI?*

The second issue is that AI is not only a technology with one trajectory to exploit but emerges as a method of invention (Cockburn, Henderson, & Stern, 2018). Therefore, there is not only the question of how much to exploit a specific technological trajectory of AI but also the question of how to change trajectories to explore more promising routes – and in fact how to define what promising routes are. Accordingly, we pursue the following research question:

RQ2*How can policy makers know how to influence the intensity and the direction of AI, i.e. how to implement the shared values?*

The third issue in need of attention is deep learning, i.e. combining different elements of machine learning where the understanding or prediction of the world takes place without any further human intervention (Taddy, 2018). This lack of human involvement in deep learning means that not only innovation policy is many steps away from having any say in its development but so are other stakeholders in the innovation system. Accordingly, we pursue the following research question:

RQ3*How can policy makers ensure that the human factor is sufficiently embedded in deep learning of AI?*

To answer these questions we create a concept of innovation policy that uses three elements: (1) We start from the stakeholders who are legitimized in the process and result of innovation (de Saille, 2015). (2) We integrate this with the approach of Schumpeterian catalytic innovation policy (Cantner & Vannuccini, 2018) which is "… capable to maneuver the parallel necessities to influence both the intensity and the direction of innovative activities" (Cantner & Vannuccini, 2018, p. 834). (3) Moreover, we use the Responsible Research and Innovation (RRI) systems approach which is able to systematically involve all stakeholders in AI processes (Werker, 2021).

The chapter is organized as follows: We start by introducing three essential elements of AI. Then, we discuss the technology AI as a game changer for innovation policy. After that, we introduce the visible hand of innovation policy at the interface of artificial and human intelligence. We conclude with our main insights and draw up a couple of major open research questions emerging from them.

28.2 Three essential elements of AI

Intelligence displayed by humans is the starting point when it comes to defining AI. Human intelligence emerges from learning and problem solving (Gardner, 1999). The capabilities of human intelligence serve as benchmark to assess the abilities of artificial intelligence. The weakest version of AI has only limited application areas, stronger versions might match human abilities and the strongest ones would even go beyond them (Kaplan & Haenlein, 2019).

Generally speaking, AI can be defined "… as a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17). Full end-to-end AI solutions are able to absorb human-level knowledge such as machine reading, thereby carrying out tasks previously done by human beings (cf. this and the following Taddy, 2018). They require three essential elements.

1 Machine learning (ML): ML routines are able to detect patterns and make predictions. Simple ML algorithms analyze historical data and are therefore "… basically limited to predicting a future that looks mostly like the past" (Taddy, 2018, p. 2).
2 A domain structure which mirrors the context of your AI by breaking complex problems into composite tasks. This domain structure relies heavily on domain expertise, e.g. in a business setting on business and economic expertise. This expertise provides the "rules of the game". As long as they are clear the tasks can be solved with ML. With the help of a domain structure combinations of ML algorithms can be used, i.e. "… dynamic processes designed and implemented by humans in conjunction with technical affordances and within broader political, social and cultural environments that are shaped by the continual interactions of strategies, structures and tactics" (Willson, 2016, p. 148).
3 Data generation, i.e. "… steady stream of new and useful information flowing into the composite learning algorithms" (Taddy, 2018, p. 4). This is not simply data collection but a strategy to develop the massive bank of data required to get the system up and running and to keep producing data so that the system can learn. Data generation in this sense requires the use of big data and internet of solutions (Kaplan & Haenlein, 2019).

One of AI's elements, i.e. ML, shows the typical features of a general purpose technology (GPT). (cf. this and the following Taddy, 2018). ML "… in its current form has become a general purpose technology. These tools are going to get cheaper and faster over time, due to innovations in the ML itself and above and below in the AI technology stack" (Taddy, 2018, p. 3). So, ML shows all three defining features of general purpose technologies, i.e. pervasiveness, innovation spawning, and scope for improvement (Helpman & Trajtenberg, 1994). These three characteristics of a GPT are not given by nature but the result of an ongoing mechanism of interaction and further innovation, the so-called dual inducement mechanism (Bresnahan & Trajtenberg, 1995). Given a sector supplying a GPT and a number of application sectors the dual inducement mechanism runs between the two types of sectors as follows: An increase in the quality of the GPT (the so-called "technological dynamism") incentivizes the actors in the application sectors to increase their technological level (the "innovation complementarities" property of GPTs), and this, in turn, induces the GPT sector to advance its technology, and so forth.

The dual inducement mechanism entails two features (Bresnahan & Trajtenberg, 1995). First, it connects various actors and therefore provides for *breadth in the application* of the GPTs' technological core; this, in a technological sense, contributes to an alignment of respectively widespread innovation activities. Secondly, the mechanism induces a *direction of innovation activities* and offers orientation by precluding alternative ways of further development; this, in a behavioral sense, aligns innovation activities and reduces some of the uncertainty inherent to innovation activities. Both features on the one hand provide for efficiency in innovation activities and contribute to their intensity – with all benefits to income and welfare. On the other hand, however, over time both features tend to build up dependency and to precluding alternative options and paths of development. A technological dynamism that turns out to be efficiency enhancing in the beginning over time may be causal for inflexibility in choice because of presumably high costs of switching to an alternative – the typical feature of so-called *lock-in situations* (Arthur, 1989; Cantner & Vannuccini, 2017).

28.3 AI as a game changer for innovation policy

AI is inherently different from other technologies because it comes with three major issues: (1) Innovation policy systematically outsources the core questions regarding developing shared values for AI to expert councils. (2) AI goes way beyond a general purpose technology (GPT). (3) AI involves deep learning which endangers human involvement in the exploitation and exploration of its potential. We consider these issues in the following.

28.4 Legitimacy of stakeholders and ethical guidelines for AI

28.4.1 Legitimacy of stakeholders and outsourcing of ethical issues of AI to expert councils

The legitimacy of stakeholders – often represented by the political actors – to intervene in innovation is nicely summarized by market and system failures. These failures have to do with the very process and context under which innovations are generated and introduced. This perspective, however, leaves out the very characteristics of the innovations generated and the very impact those innovations have on social and environmental structures and settings (e.g. are they environmentally friendly or not, are they health friendly or not, do they have socially positive or negative consequences?). Hence, when we talk about the legitimacy of stakeholders/policy makers to intervene, we consider this not only – quite traditionally – related to processes but also to outcomes of activities in general and of innovation activities in particular. On this basis, three arguments can be put forward that justify the legitimacy to intervene when innovation outcomes and their relation to values are concerned.

This legitimacy can be attributed, first, to a shift in perspective of the relation between the sphere of science and the sphere of society in the following sense (de Saille, 2015). In a Polanyi world, science (as a major source of innovation) takes place in a neutral space (republic of science) where political, moral, and social questions and hence values do not play a role. This neutralization and separation meanwhile is not considered proper anymore. Science is rather seen as embedded in the political, social, and economic world with causations running both ways and where values are relevant. Secondly, this legitimacy arises from a prominent concept, Schumpeter (1942/1975) "creative destruction" due to which any new idea and innovation has also destructive effects which affect values – and we observe this today in the social as well as the natural/environmental domains. Third, legitimacy can be justified also in an intergenerational context (Cantner & Vannuccini, 2018). Were future generations able to negotiate with the concurrent generations about which new ideas and innovations to implement quite some innovation directions would not have taken the way which in the end has been pursued – the Friday-for-Future initiative suggests such a connection. Politicians today could serve as the attorney of these future generations and in doing so need to have on board proper values – the German "Klimakabinett" could be interpreted in this way.

Along with these three arguments, proper policy approaches move away from a top-down government to a more reciprocal structure of governance (cf. this and the following de Saille, 2015). Herein policy has been moving away from considering other stakeholders as being ignorant towards respecting their questions as legitimized value-based questions about technological development.

Policy makers seem to be aware of the ethical issues coming with AI (cf. this and the following OECD, 2019). Yet, they ignore the question of legitimacy of stakeholders, because in the OECD countries, they outsource ethical questions of values to expert councils. We suggest that this is not the right way to deal with ethical AI issues, because those legitimized to take action to deal with AI issues do not take responsibility.

28.4.2 Ethical guidelines of expert councils have numerous problems

AI ethics initiatives in the form of expert councils come with numerous problems. First, they are very broad and simplistic while not giving much practical guidance on how to deal with real-world issues emerging around AI (cf. this and the following Copeland, 2019). They only help for rather simple human expert hand-craft machine learning models but do not cover any advanced AI solutions where AI algorithms decide what factors are relevant or where ML models iterate rapidly based on constantly incoming data streams. Moreover, AI ethics initiatives "… ignore fundamental normative questions about what kind of society we want" (D'Ignazio & Klein, 2019) and instead focus "… on more procedural technical and legal concerns …" (Kitchen, 2019).

The concerns about outsourcing ethical AI issues to expert councils become particularly clear when following the reasoning of Mittelstadt (2019) who compares AI with the medical community. He suggests that in contrast to the medical developments

> AI development lacks (1) common aims and fiduciary duties, (2) professional history and values, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms. … We must therefore hesitate to celebrate consensus around high-level principles that hide deep political and normative disagreement. Shared principles are not enough to guarantee 'Trustworthy AI' or 'Ethical AI' in the future. Without a fundamental shift in regulation, translating principles into practice will remain a competitive, not cooperative, process. (Mittelstadt, 2019)

We therefore suggest that while AI is an extremely difficult technology to be accompanied by policy measures, it lies without any doubts within the core responsibility of innovation policy.

The oversimplified ethical rules of expert councils suggest that AI solutions are designed carefully in advance giving ample room to detect, discuss and overcome possible negative effects. However, as the Collingridge control dilemma for rapidly evolving technologies, such as AI, has shown, negative effects are often only known after the technologies have been in full use.

> The dilemma runs thus: 'attempting to control a technology is difficult … because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow' (Collingridge, 1980: 19). … alongside these explicit references to his work, Collingridge's thinking also has implicit influence on RRI – for instance in the recognition of the significance of corrigibility in the form of 'responsiveness'. Stilgoe et al. (2013: 1572) describe this as the 'capacity to change shape or direction in response to stakeholder and public values and changing circumstances'. (Genus & Stirling 2018, p. 63)

28.5 AI is more than a general purpose technology

AI comprises not only ML, i.e. the first element mentioned in the previous section, but relies heavily on the two other elements that turn it into something "intelligent", i.e. domain expertise to build a domain structure and constantly instreaming high-quality data (cf. this and the following Taddy, 2018). Particularly, full end-to-end AI solutions require experts who can break complex human problems into composite tasks which ML can solve. And this goes way beyond computer games in which ML are often applied to test their power. Currently, many big data applications still use human expert hand-craft machine learning models, i.e. they still rely heavily on frequent human input (Copeland, 2019).

AI as a whole is often addressed as GPT (Cockburn et al., 2018). The fact that AI inherently changes the way innovation processes are carried out, i.e. in its exploitation AI emerges as a method of invention (Cockburn et al., 2018), shows that it is a GPT. Consequently, it reduces uncertainty and increases efficiency. However, this comes at a cost, namely in terms of decreasing flexibility in the direction of innovation activities and their increasing dependency on the GPT. Lock-in situations can arise and then become severe problems when a need to change direction in order to exploit new innovation opportunities by giving up well-known terrain becomes increasingly costly.

We suggest that AI goes well beyond a GPT with broad applicability, because it is agenda setting for ongoing as well as future innovation activities. Under these circumstances, the need to change directions because the current ones are exploited is a lesser problem compared to the following one: The need to change may relate to the fact that outcomes of innovation activities do not meet socially acceptable criteria anymore. Overcoming such situations is a problem of collective action depending on the distribution of preferences and attitudes and hence the heterogeneity of actors in being affected by stalemate situations that are socially unacceptable. For overcoming that and for inducing a self-organizing process of leaving a lock-in, political intervention may be a proper solution (Cantner & Vannuccini, 2017). In this sense, innovation policy has not only to deal with the intensity of innovative change and common lock-in situations resulting from the fact that current directions are technologically exploited. It also has to deal with lock-in situations that mirror socially non-acceptable situations.

28.6 Deep learning driving AI endangers human involvement in decision processes

As AI is agenda setting but might not mirror socially acceptable solutions the challenge is to figure out whether and how the direction AI is taking and developing is sufficiently transparent and manageable by stakeholders including policy makers based on shared values.

AI has been driven by machine learning. Recently, it has been shifting from "only" machine learning solutions, i.e. understanding or predicting the world based on historical experience, towards actual deep learning, i.e. combining different elements of machine learning where the understanding or prediction of the world takes place without any further human intervention (Taddy, 2018). This lack of human involvement in deep learning means that not only innovation policy is many steps away from having any say in its development but so are all stakeholders in the innovation system.

28.7 Using the visible hand of innovation policy at the interface of human and artificial intelligence

28.7.1 The visible hand ensuring the legitimacy of stakeholders

In order to use the visible hand of innovation policy to ensure the legitimacy of stakeholders, we suggest to follow a number of the suggestions made by (Mittelstadt, 2019), in particular:

1 to create accountability, implementation, and review structures at the organizational level and the sectoral level
2 to include stakeholders case-by-case, i.e. not to develop overarching principles but start bottom-up in developing ethical rules for specific cases
3 to consider establishing AI development as a profession such as medical doctors or lawyers
4 to install ethical principles on the organizational rather than on the individual level, i.e. ethics of business practices and
5 to see ethical guidelines as a process guiding technology development as it evolves.

28.7.2 The visible hand changing direction in AI development

To use the visible hand of innovation policy to change directions in AI development we suggest addressing the lock-in situations emerging from AI mirroring socially non-acceptable situations. Addressing AI as a GPT – even when using a system approach as in the OECD countries (OECD, 2019) – would lead to subsidizing activities emerging from it in rather unspecific ways. Yet the crucial question here is what kind of measures policy makers can in principle implement to change the directions within a technology development. The related question of timing and of how to determine why and when such a change might be necessary will be discussed below in the context of how to involve human intelligence in the process.

In the context of a "usual" GPT the question of how to implement redirection measures has been already answered by the Schumpeterian catalytic R&I policy approach. This approach is

> … capable to manoeuvre the parallel necessities to influence both the intensity and the direction of innovative activities. In other words, to know when to intervene on the incentive to exploitation of given technological trajectories, and when to intervene easing the transition from an exploited technological trajectory to others, richer in opportunities (hence, on the incentive to exploration). (Cantner & Vannuccini, 2018, p. 834)!

The *Schumpeterian catalytic R&I policy* can be summarized by providing a broad definition. First, it is called *Schumpeterian* in order to emphasize that this type of policy is not to be considered a repair shop restoring the incentives of private actors to innovate; it is rather considered a means to push forward specific innovative solutions for (pressing) problems by creating new markets and thereby redirecting and activating private entrepreneurs. These new markets are needed to redirect the innovation activities from a known and established innovation trajectory that is not desirable anymore – as in the case of exploited technological potentials or – more relevant in this chapter – as with AI innovation outcomes that do not meet the criterion of social consensus.

The term *catalytic* is used to label this type of policy style since policy-making should intervene in the domain of innovative activities as a catalyst intervenes in a chemical reaction. In contrast with the "market creation" approach, a catalytic public intervention is less persistent; it intervenes directly with its "visible hand", but it is smart enough to retreat its hand when the "reaction" leading to the enhanced innovative activities in new directions reaches the self-sustaining threshold.

In a sense, a catalytic R&I policy is a form of "balancing" intervention: Policy-making should focus on the framework conditions that can favor the establishment of "critical masses" (Witt, 1997) of choice in one or the other direction. New directions offer economic agents to autonomously engage in the exploitation of new or other opportunities that have the potential to meet social consensus. These critical masses can be reached through a temporary direct intervention (e.g. through innovative public procurement), or by helping to define the blurring boundaries of competition between alternative directions. By this, intervention implies that public policy has to adopt sophisticated criteria to discriminate between alternatives in the context of uncertainty and potential overall social consensus.

The Schumpeterian catalytic R&I policy shows features that go beyond being Schumpeterian and being catalytic. It furthermore is *situation-sensitive,* as it combines a "continuity" rationale – justified by the presence of challenges to policy interventions into the innovation realm that remains stable and persistent over time – with a "discontinuity" rationale – motivated by the specific trends of innovative activities in a given historical period (e.g. societal needs, grand challenges, lack of social consensus as in case of certain AI developments).

The policy further is *experimental*, as – within its rationale and given a rather broad "mission" – it should create alternative competing arenas and platforms with new potential for innovation that may achieve a social consensus. The experimental nature of catalytic R&I policy is particularly important, as it subsumes one important dimension related to innovative activities: The incentive for a (guided) bottom-up self-discovery (Foray, 2013; Hausmann & Rodrik, 2006). Self-discovery can be conceived as a criterion for action that has the potential to compensate and mild the risk of governmental failures, as the role of the public is to design the mechanism easing the directional exploration of new trajectories. The design of the experimental arena itself is key to the success of R&I policy.

Last but not least, the policy approach is *wary*, as it has to be built on the awareness that even a limited intervention may lock-in the system into inferior technologies, standards, and social values. Lock-ins are rarely irreversible in the real world (Cantner and Vannuccini, 2017), but the costs deriving from the inflexibility they generate have to be kept lower than the benefits of directional exploration. The case of a lock-in when lacking social consensus – as potentially in the AI case – is concerned is not straight forward. Whenever the costs to switch to another technological trajectory or opportunity are higher than the costs related to staying in one currently addressed which includes the costs accruing from not having a social consensus such a lock-in situation can come up.

28.7.3 The visible hand as guardian of human involvement in the era of deep learning

To use the visible hand as a guardian of human involvement in the era of deep learning we use an approach focussing on jointly developing values about the process and outcomes of

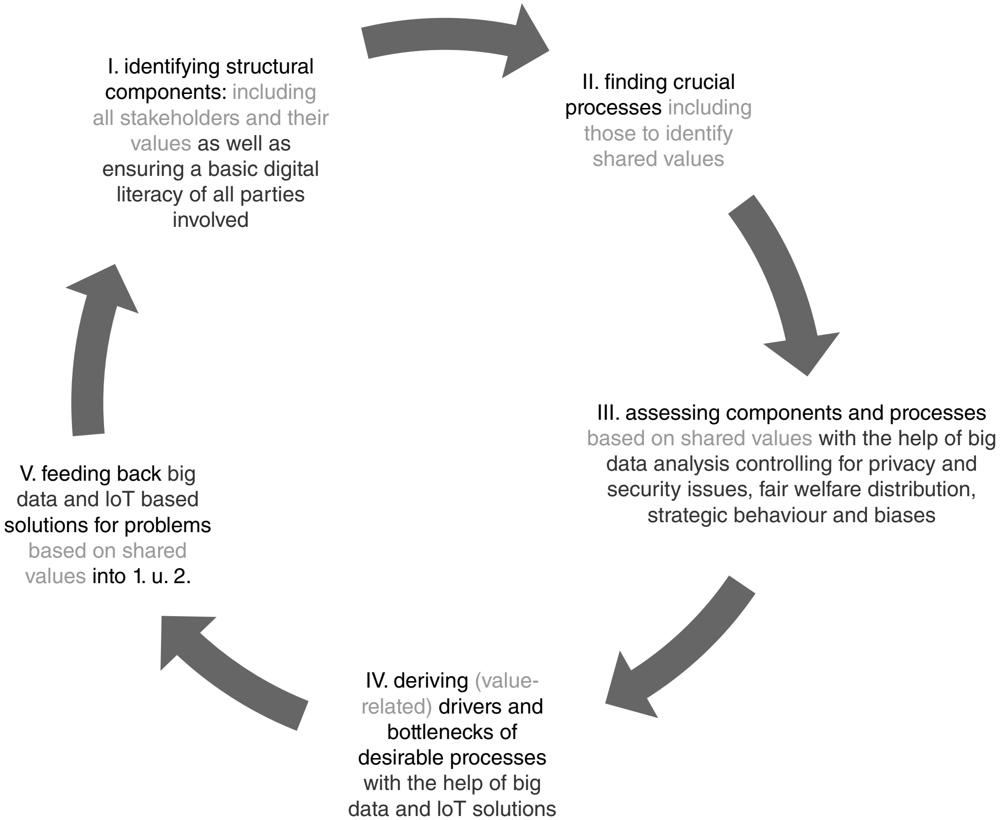

Figure 28.1 A scheme for assessing RRI systems (see Werker, 2021, p. 278).

research and innovation by including all stakeholders, also those not directly involved in innovation decisions (Werker, 2021). In fact, this approach follows the call for integrating the values of all relevant stakeholders. (European_Commission, 2013; Taebi et al., 2014). A systematic involvement of all stakeholders as suggested by the RRI systems approach appears to be giving a practical answer to how to come to shared values (Werker, 2021). Citizen science, hence the involvement of the society in invention and innovation process, could be interpreted in this sense. As such the RRI systems approach is closely related to the mission-oriented policy which stresses the importance of the "development of social capabilities, coordinate initiatives and public-private partnerships, foster synergies, and promote the introduction of new combinations that create Schumpeterian rents" (Mazzucato & Penna 2016, 316).

An RRI system contains "all relevant stakeholders of RRI and the way their values affect their activities, relationships and supporting institutions" (Werker, 2021, p. 304). In Figure 28.1, you find a "… scheme capturing the structural elements and the processes of innovation systems comprising the following five steps":

i identifying the structural components, i.e. innovative agents, relationships, and institutions (Klein Woolthuis et al., 2005)

ii finding crucial activities, such as knowledge advance by and diffusion amongst stakeholders, entrepreneurial experimentation, legitimation, market formation, development of institutions, and influence on the direction of search by different selection mechanisms, such as business models, technology development, market and institutional forces (Bergek et al., 2008; Edquist, 2011)
iii assessing components and processes by uncovering desirable ones (Bergek et al., 2008; Edquist, 2011; Klein Woolthuis et al., 2005)
iv deriving drivers and bottlenecks of desirable components and processes (Bergek et al., 2008; Edquist, 2011; Klein Woolthuis et al., 2005)
v feeding back solutions for problems into the structural components (I.) and processes (II.) including their functioning and co-evolution (Werker, 2021, p. 305f).

In the course of AI development, all stakeholders have to get some understanding of how the collection of large amounts of data, often life data, can take place and how this data can be used in AI solutions (Werker, 2021). Those stakeholders less digitally educated than other stakeholders in an RRI system will most likely fall behind (Sogeti, 2013).

> This opens ample opportunity for governmental, academic and civic agents to step up by educating and involving these disadvantaged stakeholders (see step III in Figure 28.1). As big data has been driving big science, i.e. data-driven solutions in research, e.g. at CERN (Sogeti, 2013), we might expect that the values emerging from, in this case, the academic sector, might already include goals of inclusiveness and enabling people by educating them. As long as an RRI system is not dominated by profit-oriented organizations only there is a good chance that the RRI process will lead to shared values providing a level playing field in the RRI system. (Werker, 2021, p. 314)

Using AI applications all stakeholders have to be aware of privacy and security issues as well as concerns regarding welfare, discrimination, and strategic behavior. After having determined the structural components and processes of the RRI system it is crucial to address this potential issue in the assessment of them in steps III to V of Figure 28.1.

Dealing with the lack of human involvement in AI requires an approach that goes beyond traditional approaches. AI's potential can only unfold by providing structure and rules around messy business scenarios (Taddy, 2018) and societal processes at large, not only including agents from the industrial but also from the governmental and scientific sectors. As such, AI might not only form the problem but also (parts) of the solutions to human involvement in decision processes, particularly regarding monitoring processes. AI might help in collecting and potentially including the values of all stakeholders affected by AI solutions.

On this background, we suggest to use the responsible research and innovation (RRI) systems approach, because involving all parties in an RRI process is at its very heart (cf. this and the following Werker, 2021). Particularly, innovation policy in RRI systems aims at developing shared values of innovative agents actively carrying out RRI as well as of stakeholders who are only subject to their effects (Taebi et al., 2014). Generally speaking, this approach is well suited to deal with the additional complications and opportunities of digital transformation. It helps to point at and to deal with privacy issues as well as the risk of discrimination and manipulation severely increasing in the digital age. Moreover, it shows how big data analytics and Internet of Things solutions offer multiple opportunities of following RRI processes more closely, thereby offering chances to sufficiently integrate shared values in the RRI.

28.8 Conclusions

AI and the algorithms involved in its use

> … invoke questions about how to conceptualise issues such as agency and power within a technologised everyday. … Science and Technology Studies, software studies and actor network theory all provide some fruitful insights and methods particularly in relation to the specificity of particular algorithms, yet largely fail to address many of the broader issues and questions of the everyday that are raised. (Willson, 2016, p. 148)

In this chapter, we addressed these broader issues and questions related to AI, particularly questions on how agency and power are distributed between human and artificial intelligence. We suggest using the visible hand of innovation policy in three ways when dealing with AI: (1) by involving clearly legitimized stakeholders in the design of ethical guidelines – and avoiding outsourcing this important task to expert councils; (2) by using policy measures that can distinguish between exploration and exploitation of AI; and (3) by a coordinated approach of involving stakeholders in several steps ensuring the implementation of their shared values in AI-driven decision processes. Such an approach can neither rely on policy actions nor market relationships alone but has to acknowledge their joint use (Etzkowitz, 2006).

While our integrated approach highlights three major issues and principle ways of dealing with them, so far it is still not clear how innovation policy as suggested can exactly take place. Guidelines as subsumed when discussing the involvement of legitimized stakeholders might help implement a bottom-up case-by-case approach when it comes to designing ethical guidelines for AI. Following a catalytic Schumpeterian approach, as suggested, will help change trajectories within AI, thus changing gear from exploration to exploitation. And the scheme provided in Figure 28.1 is a first step towards guarding human involvement in the era of deep learning. Combined with the ethical guidelines it might lead to a way of using AI while at the same time implementing shared values in all RRI systems of a society.

References

Arthur, B. W. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, *99*(394), 116–131.

Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., & Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research Policy*, *37*(3), 407–429. doi:10.1016/j.respol.2007.12.003

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics*, *65*(1), 83–108. doi: 10.1016/0304-4076(94)01598-T

Cantner, U., & Vannuccini, S. (2017). Innovation and lock-in. In H. Barthelt, P. Cohendet, P. L. Simon, & S. Henn (Eds.), *The Elgar Companion to Innovation and Knowledge Creation* (pp. 165–181). Cheltenham: Edward Elgar.

Cantner, U., & Vannuccini, S. (2018). Elements of a Schumpeterian catalytic research and innovation policy. *Industrial and Corporate Change*, *27*(5), 833–850. doi:10.1093/icc/dty028
Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The Impact of Artifical Intelligence on Innovation*. Cambridge: NBER.
Collingridge, D. (1980). *The Social Control of Technology*. London: Pinter.
Copeland, E. (2019). Does the public sector really need a code of AI ethics. Retrieved from https://www.nesta.org.uk/blog/does-public-sector-really-need-code-ai-ethics/
D'Ignazio, C., & Klein, L. (2019). *Introduction*. Data Feminism.
de Saille, S. (2015). Innovating innovation policy: The emergence of 'Responsible Research and Innovation'. *Journal of Responsible Innovation*, *2*(2), 152–168. doi:10.1080/23299460.2015.1045280
Edquist, C. (2011). Design of innovation policy through diagnostic analysis: Identification of systemic problems (or failures). *Industrial and Corporate Change*, *20*(6), 1725–1753. doi:10.1093/icc/dtr060
Etzkowitz, H. (2006). The new visible hand: An assisted linear model of science and innovation policy. *Science and Public Policy*, *33*(5), 310–320.
European_Commission. (2013). *Options for Strengthening Responsible Research and Innovation*. Luxembourg: European Commission. Retrieved from ISBN 978-92-79-28233-1.
Foray, D. (2013). The economic fundamentals of smart specialisation. *Ekonomiaz*, *83*(2), 83–102.
Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York: Basic Books.
Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, *47*(1), 61–69. doi:10.1016/j.respol.2017.09.012
Hausmann, R., & Rodrik, D. (2006). *Doomed to choose: Industrial policy as predicament*. Retrieved from
Helpman, E., & Trajtenberg, M. (1994) A Time to sow and a time reap: Growth based on general purpose technologies. In: *Vol. 4854. NBER Working Paper Series* (pp. 45). Cambridge: NBER.
Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. doi: 10.1016/j.bushor.2018.08.004
Kitchen, R. (2019). The ethics of smart cities. Retrieved from https://www.rte.ie/brainstorm/2019/0425/1045602-the-ethics-of-smart-cities/
Klein Woolthuis, R., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, *25*(6), 609–619. doi:10.1016/j.technovation.2003.11.002
Mazzucato, M., & Penna, C. C. R. (2016). Beyond market failures: The market creating and shaping roles of state investment banks. *Journal of Economic Policy Reform*, *19*(4): 305–326. doi: 10.1080/17487870.2016.1216416
Mittelstadt, B. (2019). AI Ethics – Too Principled to Fail? *SSRN*, 15. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293
OECD. (2019). *The Digital Innovation Policy Landscape in 2019*.
Schumpeter, J. A. (1942/1975). *Capitalism, Socialism and Democracy*. New York: Harper & Row.
Sogeti. (2013). *No More Secrets with Big Data Analytics*. Retrieved from Groningen: http://www.google.de/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0CD0QFjAA&url=http%3A%2F%2Fvint.sogeti.com%2Fwp-content%2Fuploads%2F2013%2F11%2FSogeti_NoMoreSecrets.pdf&ei=6zqwU4WvJImw7AbysoHQCg&usg=AFQjCNFMAmRVE6v4loN3Kjsg7Vp_sBguKg&bvm=bv.69837884,d.bGE
Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.
Taddy, M. (2018). *The Technological Elements of Artificial Intelligence*. Cambridge: NBER
Taebi, B., Correljé, A., Cuppen, E., Dignum, M., & Pesch, U. (2014). Responsible innovation as an endorsement of public values: The need for interdisciplinary research. *Journal of Responsible Innovation*, *1*(1), 118–124. doi:10.1080/23299460.2014.882072
Werker, C. (2021). Assessing responsible research and innovation (RRI) systems in the digital age. In E. Yaghmaei & I. Van de Poel (Eds.), *Assessment of Responsible Innovation: Methods and Practices*. Abingdon: Taylor & Francis.
Willson, M. (2016). Algorithms (and the) everyday. *Information, Communication & Society*, *20*(1), 137–150. doi:10.1080/1369118x.2016.1200645
Witt, U. (1997). "Lock-in" vs. "critical masses" — Industrial change under network externalities. *International Journal of Industrial Organization*, *15*(6), 753–773. doi:10.1016/S0167-7187(97)00010-6