AI & SOCIETY
https://doi.org/10.1007/s00146-023-01629-w

MAIN PAPER

Artificial intelligence: a "promising technology"

Hartmut Hirsch-Kreinsen[1]

Received: 1 October 2022 / Accepted: 9 January 2023

Abstract
This paper addresses the question of how the ups and downs in the development of artificial intelligence (AI) since its inception can be explained. It focuses on the development of artificial intelligence in Germany since the 1970s, and particularly on its current dynamics. An assumption is made that a mere reference to rapid advances in information technologies and the various methods and concepts of artificial intelligence in recent decades cannot adequately explain these dynamics, because from a social science perspective, this is an oversimplified, technology-centred explanation. Drawing on ideas from social scientific innovation research, the hypothesis is rather that artificial intelligence should be understood as a "promising technology". Its various stages of development have always been driven by technological promises about its special powers and capabilities when applied to solving economic and societal challenges.

Keywords Technological promise · AI community · Key technology · Special mode of innovation · Technological utopia

1 The ups and downs of AI development

The beginnings of artificial intelligence (AI) date back to the mid-1950s in the United States. Its short history is a succession of boom and crisis phases; the latter have been called an "AI winter". At least since the 2010s, however, a continuous upswing in AI development has been observed. It has been accompanied by the diffusion of this technology into all kinds of fields of application in society. In addition, AI has been the subject of an intense innovation policy discussion, as well as public and media attention, which at times has taken on the character of a hype.

The dynamic and often contradictory course of AI development has for a long time been the subject of broad research and extensive debate (e.g. Ahrweiler 1995; Nilsson 2010; Russel and Norvig 2010; Görz et al. 2021). A generally accepted interpretation is that this changing course has been significantly shaped by sometimes exaggerated expectations given the often only limited capabilities of the AI systems and methods available at the respective times – and resulting disappointments. Nevertheless, AI development has continued to progress. This is attributed mainly to continuous advances in information technology as well as constantly expanding possibilities for the use of AI concepts and methods. Key factors mentioned in this regard are the massive increase in computer power and computing capacity in recent years, the development of usable complex AI concepts such as neural networks (neural nets), and finally "big data" methods with the availability of large amounts of data, especially via the internet.

This interpretation views the development of information technology as the primary determinant of AI dynamics. Without question, this factor is centrally important. However, this is a technology-centric point of view, from the perspective of information science. The social and societal conditioning factors of AI dynamics are not systematically taken into account. Yet the mainstream of social science research on technology and innovation has for a long time convincingly demonstrated that social and societal conditions are closely intertwined with technical and specifically also information technology developments, influencing them and in some cases making them possible in the first place (for a thorough discussion, see Bijker et al. 1987). With this in mind, in the sections below we will look at how the ups and downs of AI development can be explained from a social science point of view. In particular, we will also ask why

✉ Hartmut Hirsch-Kreinsen
hartmut.hirsch-kreinsen@tu-dortmund.de
https://wiwi.tu-dortmund.de/forschung/forschung-emeriti/hirsch-kreinsen/

[1] Department of Social Sciences/TU Dortmund University, Social Research Centre Dortmund, Dortmund, Germany

scientific and public interest in AI did not permanently subside, especially after downturn phases.

2 On the concept of "promising technology"

According to social science technology research, *expectations* about the future opportunities of new technologies have a particularly strong influence on the course and direction of technological innovations (Beckert 2016). This is because innovations are usually accompanied by risks and uncertainties with regard to their course, their feasibility and the desired, mostly economic effects. Expectations and associated promises, visions and scenarios about possible and desirable development paths reduce the uncertainties and complexity surrounding innovations, open up concrete perspectives for action for the developers, interest other involved actors, and coordinate their actions.

This argument can be made more precisely with the concept of *promising technology* (van Lente and Rip 1998; Bender 2005; Borup et al. 2006; Konrad 2006). The theory is that actors are guided in their decisions to participate in a technological development by an initially still very general technological promise associated with this yet-to-be-developed technology. A technological promise can be understood as a narrative that offers interested actors a vision, opens up innovation and research paths, and justifies expectations by looking to future potential applications. It is a necessary condition for approaching other actors, involving them in the innovation process, coordinating their actions in a purposeful way, mobilising innovation resources, and initiating investments in research and development. At the same time, however, from the point of view of the actors addressed and the public, a technological promise must be linked to the current state of research and the available technological potential in an understandable way, despite its general character. Such a link should take the form of highlighting innovation potentials that have not yet been used, but which the addressees see as being possible and promising. To put it more casually: the promise must not run the risk of being dismissed as the mere pipe dream of a few "technology freaks". In this way, a reciprocally reinforcing process is initiated between the concretisation of the promise, the drafting of an innovation agenda, and steps towards a concrete technological development – at the end of which a new technology and a new institutionalised socio-technical field emerge, of a kind that did not exist before.

According to this concept, it then follows that in AI's various stages of development, the dynamics of AI are always driven by renewed and modified technological promises about its special powers and problem-solving capabilities in respect of social and economic challenges. While the promises are more or less plausibly justified by the respective state of research and expected technological potential, there has often been a lack of concrete application experience, with the result that the promises have a less than concrete character, and are indeed technological utopias. Such technological promises, made and renewed by influential actors with an interest in the technology, have always played a major role in initiating, shaping and, above all, sustaining the dynamics of AI in its various stages. These developments are held to lead to a particular state of research and technology and, when successful, ultimately generate an institutionalised socio-technical field.

A series of mostly critical studies on the history of AI development show that promises, expectations and narratives have shaped the development history of AI (e.g. Cyranek and Coy 1994; Brödner 2019; Buchanan 2006; Larson 2021). On the one hand, these promises have proven to be exaggerated in the face of only limited performance capabilities and considerable application difficulties for the AI systems and methods available at the respective times. The result is often "disappointed expectations" on the part of key actors involved (Ahrweiler 1995, p. 22), and their withdrawal from the technology field. On the other hand, the rapid and obviously dynamic development of AI has repeatedly given rise to new promises. With great persuasive power, these promises arouse the expectations and interest of many actors and the curious public in this technology. The promises draw their power from "transhistorical" and "transcultural" fantasy stories about intelligent machines, which run through human history and existed long before the advent of modern science (Cave et al. 2020).

These promise-driven AI dynamics are summed up by an early protagonist of American AI and founding member of the American Association for Artificial Intelligence, Bruce G. Buchanan: "The history of AI is a history of fantasies, possibilities, demonstrations, and promises" (Buchanan 2006, p. 53). Similarly, a German AI scientist in 1994 pointed to the unachieved goals and constant postponement and renewal of AI development prospects: "This is an old game in AI. They say: in 10 years we will have this and that, and then after 10 years you ask the same question and they say: in 20 years we will have achieved it. After 20 years, you ask the same question again, and also 50 years from now we will be asking the same question" (Cyranek and Coy 1994, p. 259).

3 Methodology

These questions will be examined in more detail, using the example of AI development in Germany, which began in the 1970s. This cannot always be clearly differentiated from international AI development, to which it has been closely linked since the outset. AI in Germany has gone through very different economic cycles and, as internationally, since around the beginning of the 2010s it has generally been characterised by rapid scientific and technological development and an extensive public and policy debate.2F.[1]

The methodological basis of the arguments put forward in this paper consists of a qualitative analysis of the discourses on AI. The analysis relies in particular on the evaluation of a large number of "grey" documents, preprints, policy statements, websites and specialist publications, which originate from national and international contexts. Secondly, the investigation draws on a reinterpretation of own existing research findings from recent years on the process of the digitalisation of society.[2] Finally, the results of 19 interviews with AI experts in business and academia are used as an important information tool in the analysis process. These interviews were conducted between October 2021 and February 2022.[3]

4 Dynamics and stages of AI development

AI began to become established in the Federal Republic of Germany in the 1970s, closely linked to AI development in the United States and to some extent in the United Kingdom. Various stages of development can be identified. They differ in terms of the respective scientific and technological foundations, the related technological promises and expectations for the future based on it, the AI community that was instrumental in driving the technological promise, and the degree of institutionalisation of a socio-technical field of AI. In Germany, these stages run more or less in parallel to international AI development, especially to that taking place in the United States. At the same time, however, these stages also have specific national characteristics.

4.1 Scientific and commercial promises: the first stages

For the period up to around 2010, very broadly speaking, four development stages can be identified.

4.1.1 1970s: AI in a scientific niche

The first stage can be described as a science-orientated startup phase. By the mid-1970s at the latest, AI had begun to establish itself as a scientific niche discipline, in part against sustained resistance from computer science, which was already established at the time. The key players were a small group of young scientists who saw great future opportunities in this field of research. These were actors who had a shared orientation to a greater or lesser degree. They maintained informal contacts and exchanged research ideas and perspectives on a casual basis (Ahrweiler 1995). In terms of technology, the 1970s stage relied particularly on heuristic approaches and symbolic AI using rules-based methods (Bauberger et al. 2021, p. 908).

The first step towards institutionalising early AI in Germany can be dated to February 1975, when a first organised meeting on "Artificial Intelligence" took place in Bonn. In retrospect, it has been said that this event represents a pivotal starting point for AI in Germany (Konrad 1998). It was at this meeting that scientific perspectives and objectives of AI were systematically drafted and specified, in the context of specialist lectures. It was here that a technological promise was formulated for the first time, explaining AI's prospects for development and institutionalisation with a scientific goal in mind. Not least, this vision attracted increasing attention from national science and research policymakers. The new information technologies and also early AI were even at that time already regarded as "modern key technologies" and as a critical "productive factor" for future economic and social development.3[4]

It is notable that this development in Germany was hardly affected by the fierce criticism of AI that came to be voiced in the U.S. and United Kingdom by the end of the 1960s at the latest. Criticism there can be seen as a reaction to the great unfulfilled promises made in the early days of the 1950s. The consequence internationally was a stage in the development of AI that has been referred to as the first "AI winter" (Teich 2020).

[1] Studies of different provenance are available on the history of AI in the Federal Republic; most notably the sociology of knowledge study by Petra Ahrweiler (1995, the history of technology approach by Seising and Dittmann (2018), and reviews by computer scientists who had been involved at an earlier stage, e.g. Konrad (1998), Siekmann (2009) and Bibel (2006, 2014); on international developments cf. e.g. the broad study by Nilsson (2010).

[2] According to the current public discourse, the term digitalisation is used to describe the diffusion of IT technologies in recent years. However, it must be emphasized that this is a conceptually abbreviated term. Processes of digitalisation have actually taken place roughly 5000 years ago.To be precise, the current development can be termed more precisely as an "algorithmic revolution", since parts of cognitive work can be formalized in the form of calculation methods displayed and then also executed automatically under programme control (the author thanks an anonymous reviewer for pointing this out).

[3] The interview findings generated during the study are available from the corresponding author on reasonable request; see Hirsch-Kreinsen (2023).

[4] As stated by Hans Matthöfer, then Minister of Research (quoted in Ahrweiler 1995, p. 85).

4.1.2 Commercial dawning in the 1980s

The early 1980s marked the beginning of a phase that is unanimously described as an upswing in AI in Germany (Ahrweiler 1995; Teich 2020). It started with a fundamentally renewed technological promise. Now the primary focus was directed towards the commercial prospects of AI. This outlook was first made explicit at the 10th annual conference of the German Informatics Society (Gesellschaft für Informatik) in the autumn of 1980, in a lecture by the American computer scientist Edward A. Feigenbaum. He emphasised the huge economic significance of what was then called the knowledge-based approach in AI, i.e. the methods of symbolic AI and expert systems based on it, which were widely pursued at the time. He described "great" potentials with regard to methods of knowledge representation and complex applications such as language recognition, chemical analysis and synthesis, medical diagnosis and treatment, and mineral prospecting (Görz et al. 2021, p. 8). For this reason, far-reaching economic effects of the expected use of AI systems were also predicted (BMFT 1988, quoted in Ahrweiler 1995, p. 120).

These promises converged in particular with a growing interest in AI at the time on the part of research policymakers. A direct impetus for this came from the Japanese "Fifth Generation Computer Systems" research initiative. Launched in 1982, it effectively started an international technology race in computers and AI. At the same time, the technological promise and associated policy activities increasingly resonated with businesses in the IT and electrical engineering sectors, who were beginning to turn their attention to questions concerning AI and possible applications for expert systems. In parallel, a progressive institutionalisation of AI in the scientific system of the Federal Republic can be observed.

Altogether, this shows us that the dynamics of AI were being driven equally by the interests of science, research policy and businesses that were developing AI technology. The beginnings of a socio-technical field of AI were discernible: at its core was a loosely networked, but also determining constellation of actors in science, research policy and parts of the private sector.

4.1.3 Crisis at the end of the 1980s

Nevertheless, by the end of the 1980s at the latest, a situation emerged that is referred to as the AI crisis, or internationally as the second "AI winter" (Teich 2020). A large gap had emerged – and was impossible to overlook – between the expectations of many businesses and policymakers about the commercial usability of expert systems, and their actually realised benefits. Available data indicates that the diffusion and first applications of expert systems in the 1980s were very slow (e.g. Dostal 1993). The original technological and economical promises turned out to be exaggerated. This led to massive "disappointed expectations" (Ahrweiler 1995), especially among policymakers and many of the businesses involved. One consequence was a significant reduction in the amount of private-sector financing available for the now diverse range of AI institutes that had sprung up, putting their continued existence in jeopardy. Another consequence was a dramatic cutback in AI-oriented research and development (R&D) capacities in many companies. Meanwhile, growing scepticism on the part of research policymakers also had a braking effect. They too had wanted to see AI become a commercial success, ever since they first started providing funding (Reuse 2008). In other words, the socio-technical field of AI in the 1980s – and in particular the dominant constellation of actors in science, policy and business – proved to be unstable, and crumbled.

4.1.4 Consolidation into the 2000s

The following stage of development covered the period from the 1990s until well into the 2000s. It would be wrong to view this solely as a crisis phase; rather it was a long phase of AI consolidation, especially in Germany. Some interviewed academics confirmed this interpretation. Thus, at the beginning of the 1990s, after the promises of the 1980s and subsequent crisis, one computer scientist noted: "AI on the path to normality" (Brauer 1993). AI activities were now only partially funded by government research money, and further developments in AI took place only sporadically in partnership with business. A process began that was driven primarily by science and rarely had the commercial dimensions of the 1980s. During this stage, there were also few promises and expectations that were as far-reaching as before. Nevertheless, a position paper marking the formation of the Working Group of German AI Institutes (Arbeitsgemeinschaft der deutschen KI-Institute, AKI) in 1991 can be construed as a renewal of the basic research and development goals of AI (Barth et al. 1991):

- Firstly, it recalled the fundamental goals of AI, namely to pursue the question of the essence of intelligence and the technological implementation of functionalities that might be derived from that.
- Secondly, it emphasised in relatively concrete terms that AI saw itself as a science that creates concepts, models, methods, tools and know-how to realise usable intelligent systems in cooperation with others.

Thus, the position paper initiated a goal-oriented programme of consolidation in AI. This was reflected in a progressive differentiation of topic areas and development priorities, as well as a further institutionalisation of AI. The main topics were classic machine learning methods such as Bayesian statistics, and also a revitalisation of the connectionist AI concepts that had been rejected in the U.S. at the end of the 1960s, i.e. neural networks (Bauberger et al. 2021; Görz et al 2021). An influential AI scientist at the time described the institutional conditions under which this took place. Despite the crisis, the boom years of the 1980s did achieve something positive: "They left behind a research structure for AI that is otherwise unequalled in Europe. It is comparable to the Japanese and American infrastructure" (Siekmann 1994, p. 24).

4.2 Boom since the 2010s

4.2.1 Far-reaching technological promise

The technological basis for the dynamics of AI since the 2010s has been a development push dubbed the "Big Bang of deep learning" (Görz et al. 2021, p. 9). This refers to the development of artificial neural networks and machine learning methods, together with the availability of large amounts of data. A technological promise is based on this, encompassing far-reaching visions of society as well as commercial, industrial and economic goals. It concerns the usability of AI in all kinds of social contexts, the improvement of social living and working conditions, and also the possibility of using AI to deal with environmental challenges and climate change. A special focus, however, is on industrial AI applications, where AI is predicted to bring massive productivity gains. An important reference point for discourse on AI in Germany is the *vision of Industry 4.0*. Presented at the beginning of the 2010s, it offers far-reaching promises about the modernisation and especially digitalisation of industry. Overall, the technological promise is the result, continuously refined, of a discussion process that has taken place in the years since 2011. It has been driven by an influential AI community, and in its course, a large number of publications setting out the way forward have been presented (e.g. Forschungsunion and acatech 2013; Bitkom and DFKI 2017; Bundesregierung 2018; EFI 2022).

4.2.2 Influential AI community

At the core of this AI community is the scientific discipline of AI, which is highly motivated and insistent in pursuing its goals. This group is collectively convinced that AI has a great future and that it is possible to come close to the long-cherished visions of an intelligent machine. At the same time, it acts in close coordination with interested businesses, policymakers and business associations. A central institutional factor facilitating the activities of the AI community is the coordinating role of the German National Academy of Science and Engineering (Akademie für Technikforschung, acatech). The AI community essentially formulates the technological promise, has a strong voice in its concretisation and the definition of the research agenda, and influences the public discourse. It is able to attract considerable research funding and qualified scientists for its objectives, and can continuously build on this influential position through the increasing success of the technological promise. As a result, the AI community gains increasing influence and the power to define innovation priorities and resource allocations in the societal innovation system.

4.2.3 Policy: agenda-setting, support, coordination

The technological promise and the activities of the AI community have met with a particularly positive response from innovation policymakers, whose diverse funding and support measures have exerted a considerable influence on the specification of the research agenda and coordination of AI development. This is happening in a very particular way with the German federal government's artificial intelligence strategy, a policy document published at the end of 2018. This strategy explicitly refers to the discourse promoted by the AI community and the associated promises that have been formulated. AI is described as a "key enabling technology" for future social development, requiring sustained investment (Bundesregierung 2018, p.4). The strategy boils down to three goals:

- strengthening the competitiveness of Germany and Europe through large-scale support for AI in general;
- responsible development and use of AI oriented to the public good; and
- the ethical, legal, cultural and institutional embedding of AI in society.

In this way, policymakers are explicitly addressing critical questions surrounding the controllability and transparency of AI processes, and their social, environmental and especially also ethical consequences. As a result, these critical issues are also taken up by various bodies, commissions and organisations, often at the instigation of policymakers, and linked to questions concerning AI development and application.

In general, AI is seen as an innovation policy "beacon" that allows policymakers to portray themselves as future-oriented and bestows legitimacy on them. In addition, these innovation policy measures expand, intensify and initiate a large number of follow-up activities at various levels of government. The same applies to corresponding initiatives at EU level.

4.2.4 Legitimisation through public discourse

The technological promise and innovation policy activities have been closely associated with an extensive socio-political discourse on the general significance and social consequences of AI. Here it is hard to separate the technological promise of AI from the optimistic side of this discourse. The promise influences the discourse, but also gains legitimacy by referring to it and via the almost euphorically celebrated and internationally staged highlights of AI's capabilities, e.g. playing chess or the board game Go. Conversely, sceptical and dystopian arguments relating to social and ethical challenges of AI have hardly slowed its dynamics. Instead, these objections are increasingly becoming part of the research agenda itself; they are being taken up by policymakers, and their handling is being institutionally anchored in a number of bodies and organisations.

4.2.5 Socio-technical field

In structural terms, these dynamics have established a socio-technical field of AI whose central feature is a stable and closely networked constellation of developers in innovation policy, interested parts of the private sector, and science.

The influence of an increasing number of businesses that are interested in AI as developers, and in some cases as users too, also deserves special mention. These include IT and technology-intensive companies in established industries, as well as a rapidly growing landscape of highly specialised startups. For many development companies and some users in a wide range of business sectors, AI opens up new sales opportunities for new and hitherto difficult to exploit application areas for digital technologies and related products. Many interested businesses are primarily concerned with securing and expanding their existing world market position through AI-based innovations. In this respect, the technological promise of AI is in the tradition of the debate that has been going on for years about the possibilities of using a wide variety of information technologies for world market oriented product innovations.

With regard to AI development, stable, complementary interests are emerging among the representatives of all three areas–science, business and policy. They are developing in a context of institutionalised communication and interaction relationships, for example within the various innovation policy advisory bodies, all manner of publicly funded application-oriented development projects, and state-sponsored knowledge transfer institutions. In addition, long-term close relationships exist between science and business in the context of the funding of institutes by the business sector, joint research activities and, for example, joint publication activities at conferences.

5 Prospects for AI

The current dynamics of AI raise the question of the future prospects for AI development. The decisive factor for the direction and scope of future AI development is the extent to which the previous limitations and unresolved challenges of AI diffusion and application can be tackled and overcome. Depending on how these possibilities are assessed, very different AI development prospects are expected. Accordingly, the question of the prospects for AI development is intensely debated in the relevant literature, and is hotly contested (e.g. Ford 2018; Görz et al. 2021; Larson 2021).

5.1 Limitations and challenges

If we ask about present limitations and challenges in respect of further AI development, we can identify a whole range of very different factors and conditions, such as practical application problems and economic uncertainties. Social conditions, especially the unresolved legal and ethical questions, must also be taken into account; and finally these factors correlate with the current state of development and the technological shortcomings and system limits. Summarising the available findings and the state of the debate, the following challenges in particular should be noted:

The difficulty of effective knowledge transfer between the various actors and knowledge domains involved is mentioned in particular also by the interviewed experts as being a permanent social and organisational problem in the development and diffusion of usable AI systems. The ability to integrate knowledge from development and application areas is an essential requirement for functioning AI systems. New and effective modes of knowledge transfer and cooperation between actors in different domains are only just starting to be developed (e.g. Brödner 2019; ten Hompel et al. 2019; Ecker et al. 2021).

In addition, fundamental functional problems and the often only limited performance of machine learning systems in particular are pointed out, which stand in the way of rapid diffusion and broad application:

- First, this concerns the systems' poor ability to cope with "open worlds", i.e. situations that are hard to calculate *ex ante*. If unexpected events or unpredictable exceptional situations occur, an AI system's lack of "robustness" – as an expert highlighted—becomes a problem.
- Second, there is as yet no way of incorporating often indispensable everyday knowledge into the system pro-

cesses. According to Görz et al., this includes not only the ability of AI "to perform abstractions to a certain degree and for machines to comprehend causalities, but also to approximate that which is a special characteristic of humans, namely to understand and explain actions" (Görz et al. 2021:10).

- Third, this refers to the problem of so-called explainable artificial intelligence, which has been discussed since the 1990s. The more risky their decisions become for actions taken by humans – for example when it comes to medical diagnoses and proposed treatments – the more important it will be to understand what the systems are actually doing. Above all, this is a question of transparency and certainty – whether and how the system actually solves the task to which it is applied.
- Fourth, the question of data quality and the amount of data needed for particular use cases is not a problem that has a definitive solution. It is often unclear whether a system really needs all the available data, or whether the amount of data can be reduced for the sake of rationalisation. Conversely, there is the problem that often only insufficient amounts of data of highly varying quality are available for learning systems to work with.

These challenges are the subject of extensive research activities, in Germany and internationally. The problems of explainability and transparency of system processes have been the focus of R&D funding for some time now, for example, and the aim is to develop systems that are more robust and trustworthy than existing machine learning methods.

5.2 A new AI winter?

In view of the numerous application and system problems, a whole array of voices can be heard anticipating a repeat failure of high-flying AI promises. They do not rule out a new AI winter. Several arguments are put forward in support of this view: for example, the years since 2010 have been characterised by a new hype cycle, as has already occurred several times with AI, but which is unsustainable in the long term (Larson 2021: p. 74) Because of many development and application problems, it will very soon become apparent that AI development will by no means be as groundbreaking in the future as has often been promised. It is also pointed out that the performance capabilities of the available technology are currently, as in the past, completely overestimated: "Like in the AI winter at the end of the 1960s and 1980s, when the promising developments in the AI labs failed miserably in real, practical applications, there is once again a large discrepancy between what is expected of these technologies and their actual capabilities" (Heimbrecht 2021).

As a consequence, it is predicted that expectations and investments will be scaled back, and public interest will wane. One of the main reasons given for this is that to date, it is still unclear what AI actually means. This leads to exaggerated expectations, misunderstandings, and ultimately to a depression Pieknewski (2022). For example, it is argued that many expectations which are basically directed towards developments in strong AI are measured against the only limited performance of specialised weak solutions, so that disillusionment is the result. Therefore, some critics insist, the current dynamics are at a "tipping point", and further development could lead to a new stagnation and a new winter (Kaltheuner 2021: p. 193).

5.3 Far-reaching promises: artificial general intelligence

In complete contrast to these pessimistic expectations, far-reaching promises and expectations are being formulated once again in the German AI discourse. Many AI actors anticipate a lasting and far-reaching AI boom. For example, innovation policy forecasts for the key technology AI predict an "almost exponential trend" (Kroll et al. 2022, p. 31), and a coming "golden age" of AI is talked about in virtually euphoric terms. As justification, reference is made to new technological potentials and the anticipation of continued rapid technological development. In particular, great algorithmic advances in machine learning and deep learning are predicted in conjunction with the availability of enormous data sets and advances in fast, parallel computing. This would foreseeably overcome and surpass the current limitations of AI. This is because the new systems would then no longer be usable only for specialised applications. Instead, they would tend to be able to cope with all kinds of different applications (Kersting and Tresp 2019, p.3).

At the core of these optimistic outlooks are the far-reaching expectations firmly expressed by a whole array of scientists about progress in AI development in the coming years, and the associated possibilities of realising concepts of strong AI or artificial general intelligence. A significant number of scientists in the national and especially the international AI community are convinced that technologies using strong AI will become reality in the future and longer term (Müller and Bostrom 2016; Ford 2018; Science Media Center 2021). It is expected that on the basis of enhanced methods using neutral nets in particular, and rapidly increasing computer capacities, a general, human-like machine intelligence combined with everyday consciousness and emotionality can be developed in the foreseeable future. According to such views, it is entirely conceivable that everyday consciousness can be reproduced mathematically (Müller and Bostrom 2016; Ford 2018; Science Media Center 2021, p. 7). A driving motive behind this is undoubtedly the promise and the deep conviction, formulated and renewed over and again since the beginning of AI, that human intelligence can be reproduced with computer models and also finally understood. And such expectations are unmistakably linked to the vision of an intelligent machine that is supposed to be broadly similar to humans, which goes back to the founding era of AI (Nilsson 2010, p. 528).

5.4 "Routinisation" of AI

Other forecastsput forward a more nuanced view. A trend is anticipated that can be described as the *routinisation* of AI, in which experiences of AI increasingly become routine and ordinary, part of everyday used systems. From this perspective, existing AI approaches and methods will continue to be developed incrementally, current limitations and challenges will be gradually overcome, and therefore ever more powerful. So-called weak AI systems will come into widespread use. However, it remains to be seen whether the envisaged stable AI solutions based solely on learning systems can actually be realised in the foreseeable future. According to critical opinions that are also shared by the interviewed experts, it is more likely the case that the common shortcut of equating AI with machine learning has proven to be a mistake. Despite the gigantic amounts of data used to train self-learning AI systems and vast computing power, often no explainable and robust solutions have been found for the respective tasks. Therefore, AI development must "… abandon the initially somewhat one-sided overemphasis on machine learning from mass data in favour of a combination with state-of-the-art symbolic methods" (Wahlster 2020). This also implies that AI development in the future should focus on concrete and relatively precisely definable fields of application. These application areas will, of course, expand and lead to an increasing diversification of AI development.

In the longer term, therefore, a *hybridisation* of AI development can be expected as well as a general routinisation in the technological sense. More specifically, symbolic AI models – which are used for semantic knowledge representation – will be integrated with deep learning approaches or rather sub-symbolic AI models, and with machine learning methods for recognising and extracting structured scene, event and situation information from data streams. It is clear that this development perspective is being accorded increasingly central importance in the international debate as well (e.g. Brachman and Levesque 2022).

There is much to suggest that routinisation of AI will be the dominant path for AI in the future. On the one hand, many factors weigh against another AI winter. These include, above all, the current state of and level of investment in research and development, the vested interests of the established AI community, and the now firmly institutionalised socio-technical field of AI. On the other hand, there are well-founded doubts about the path of artificial general intelligence. It is questionable whether the fundamental technological hurdles and as yet completely unsolved development problems can be overcome in the foreseeable future, or whether the various expected breakthroughs in AI development can be achieved so as to come closer to a strong AI or artificial general intelligence (Barthelmeß and Furbach 2021; Koehler 2021). In addition, however, the arguments of many AI critics must also be considered, according to which this prospective development can never be realised. To summarise the much-discussed argument, this is because human intelligence and thought are not the same as computation, and the goal of human-oriented AI is ultimately a myth (Brödner 2019; Dickson 2021; Larson 2021).

6 A specific AI innovation mode

A condition for and consequence of the dynamics of AI, especially regarding its further development, is an AI-specific innovation mode. This mode is the indispensable prerequisite for the successful development of AI systems. However, the structural features of this mode are barely compatible with the long-established, industry-oriented institutions and practises of the German National Innovation System (NIS). This poses major challenges for innovation policy (Botthof et al. 2020, 2023). This innovation mode is a result of the specific development requirements of AI, and of the associated current limits to its diffusion and application. Its central features are:

- Transdisciplinarity and knowledge transfer: AI innovations tend to take place in transdisciplinary contexts, which integrate knowledge domains that have hitherto been separated from each other in different disciplines and sectors; in particular, a close-coupled knowledge transfer between the development side and application areas is necessary.
- Fluid cooperation processes and open innovation ecosystems: Relevant innovation knowledge is often available only beyond the boundaries of established companies, scientific institutes and traditional application fields of IT-technologies; this means that open innovation ecosystems become very impotant and must be supported, in particular with the inclusion of specialised startups.[5]
- Agile innovation strategies: Established companies in core industries transform their strategies towards forward-looking and at the same time flexible innovation,

[5] Concerning the term Innovation Ecosystems see e.g. Granstrand and Hogersson (2020).

to exploit the potentials of rapid technological development.

- Heterogeneous actors: AI innovations are increasingly taking place in the context of heterogeneous constellations of actors; besides computer science, AI and engineering, there are other actors from a wide range of application fields, scientific disciplines and segments of society.
- Wider skill profiles: Significantly wider skill profiles are needed for those actors involved in the development and deployment of AI-based systems. This is a general trend in which, with AI in almost all employment sectors, a shift towards additional skills such as handling AI systems or shaping the AI context can be observed.
- Socio-technical understanding of innovations: The term innovation is evolving from a concept focussed primarily on technological development to one that systematically includes complementary social innovations. This is necessary to achieve economic effects at the enterprise level as well as a desirable change in society as a whole.
- Diversification of innovation levels: AI innovations are increasingly taking place at different social and institutional levels, both subnationally and internationally, and a multi-level system of innovation is developing beyond previously existing structures.

The specific AI innovation mode and the development and application experiences of AI to date imply new requirements for established innovation policy. These can be outlined as follows:

- Flexibilisation of knowledge transfer and promotion of innovation ecosystems
- Openness to technology and increased focus on hybrid AI solutions
- Intensification and expansion of skill development measures
- Promoting concepts of socio-technical system design, especially also with regard to AI applications
- Agile funding models with short feedback cycles, but also expansion of basic research with a long-term orientation
- Greater promotion of transparency and explainability in machine learning methods, including ethically oriented solutions

Finally, it is unanimously emphasised in the innovation policy debate that the change in innovation policy should be accompanied by a stronger *mission orientation.* In essence, this means moving away from the role of the state as a framework provider, towards a state that gives direction and drives innovation (e.g. Botthof et al. 2023). It is foreseeable that the focus on the key technology AI and the associated technological promise will play a central role here.

7 On the rhetoric of the technological promise

Finally, it should be asked what the reasons are for the enduring persuasiveness and its strong impact especially on politics and the public debate of the AI technological promise. Without question, this has been accompanied by considerable success in the development and diffusion of AI. But the promise goes well beyond that. So to be persuasive for the addressees the technological promise must be as credible as possible. However, this is based on a number of pre-conditions. A technological promise needs to employ a rhetoric that must be convincing and at the same time allow a wide range of associated possibilities to be envisaged. The mechanisms at work here can be characterised in terms such as communicative generalisations and reduction of societal complexity, de-contextualisation of arguments, irrefutable topicality, quantifiability and relevance to everyday life (Kieser 1997; Hirsch-Kreinsen 2016; Madsen 2019).

The technological promise of AI also features this sort of argumentation. The promise must be formulated as vaguely and non-specifically as possible to generate interest and provide connectivity for a wide variety of actors. Therefore, the rhetorical core of the technology promise can be seen as the ambiguity of the term AI, and the term can be seen as a "very loose umbrella term" (Kaltheuner 2021, p. 23). Thus, concrete technologies and systems as well as the need for adaptation to concrete fields of application are largely ignored. At best, since the 2010s, the term AI has been equated rather hastily and without reflection with machine learning and artificial neural networks, without specifying this further. Above all, the promise must abstract from the much-discussed development and application limits of AI,[6] and, quite to the contrary, suggest far-reaching development prospects. For it is only by counterfactually bypassing these issues that the technology promise can have its lasting effect on the social process.

Hence, the technological promise of AI can be summed up under the following headings:

[6] Since its inception, AI development has been accompanied by fundamental criticism of its basic assumptions and concepts; cf. for example, among many other authors, the classics of this debate Hubert Dreyfus (1972) and Joseph Weizenbaum (1976), the critique of AI by Peter Brödner (2019), as well as the discussion by Erik J. Larsson (2021) of the current AI boom and the perspective of a human-like machine intelligence, so-called Artificial General Intelligence, which many consider realistic.

- Ambiguous metaphors: The core of the technological promise is the ambiguity of the term AI. The metaphors used to describe this technology in sweeping ways – such as "intelligent", "learning" or indeed "autonomous" – are as imprecise and misleading as they are easily accessible (cf. Brödner 2019; Larson 2021). The technological vision therefore appears to be clear and convincing for outsiders. This concerns in particular a weak grasp of the subject in the public debate and among journalists.[7]
- Inevitability: The widespread use of AI has an almost inevitable character due to its rapid development and its repeated attribution on the part of policymakers as a key technology for the future development of society.
- Accelerated growth: AI is being regarded as a necessary condition not only for considerable and sustained economic growth, but also for national success in the global technology race – especially for Germany's important industrial sector. Predictions in this regard may be impressive, but the majority of them cannot be validated.
- Generalisation of individual cases: The far-reaching economic promises and the message to businesses to introduce the new technology as quickly and comprehensively as possible are underscored by the constant, intense reference to specially high tech companies that have been successful with AI. The message to sceptics is that those who join in will almost inevitably succeed.
- Securing social legitimacy: It is emphasised that AI will bring positive, socially desirable consequences. Furthermore, it will enable the successful overcoming of social and environmental challenges.

However, the technological promise is only really convincing if the timing is right. It has to hit the spirit of the times (Kieser 1997). Without question this is true for AI, as it speaks to the *zeitgeist* about the need to modernise society through new technologies, especially through digitalisation.

To sum up: the technological promise of AI has the character of an unquestionable technological utopia. This is because AI is presented as the technological means to overcome the many urgent challenges facing society, and bring about a better society. The specific utopian aspect consists in the fundamental assumption that humanity will prove able to cope with the technological increase in its power and control over the whole of nature, contrary to ever-burgeoning technology-based scenarios of horror and disaster (Münkler 1997, p. 62f.)

But actually there is more at stake: there is a techno-utopian dream of being able to rationally plan and control the development of society, made possible by technology – to transcend intransparent social complexity, interest-driven political discussions and time-consuming democratic procedures. For many protagonists, this dream seems to be in reach. From this perspective, complex cause-and-effect analyses, the conflict-laden interpretation of contradictory scientific research results, and the interest-dependent formulation of action and solution strategies become unnecessary. Instead, the technology promises an autonomous and smart, seemingly objective solution to overpowering problems for society such as the climate crisis. This is the expectation that Evgeny Morozov has summarised as "solutionism" (Morozov 2013, p. 5). Ultimately, this links back to the old technocratic vision of society, in which fundamental and desirable transformations in society can be achieved by technological means alone, while political means are less suited to bringing about social change.

[7] The author thanks an anonymous reviewer for pointing this out.

Acknowledgements The paper summarizes main findings of a research project ("Expertise") in the context of the framework project „KI – Mensch – Gesellschaft: Den Wandel des Mensch-Technik-Verhältnisses durch Künstliche Intelligenz ganzheitlich verstehen und bewerten (KI. Me.Ge.)" coordinated by the ISF Munich and funded by the German Federal Ministry of Education and Research.

Funding Open Access funding enabled and organized by Projekt DEAL. Bundesministerium für Forschung und Technologie.

Data availability The empirical data referred to in this paper are available on request from the corresponding author, but are not public due to privacy restrictions.

Declarations

Conflict of interest The author has no conflict of interest to declare.

References

Ahrweiler P (1995) Künstliche Intelligenz-Forschung in Deutschland. Die Etablierung eines Hochtechnologie-Fachs. Waxmann, Münster

Barth G et al (1991) Künstliche Intelligenz: perspektive einer wissenschaftlichen Disziplin und Realisierungsmöglichkeiten. Informatik-Spektrum 14:201–206

Barthelmeß U, Furbach U (2021) Computer auf dem Weg zum Bewusstsein. Faz-net. https://www.faz.net/aktuell/wirtschaft/digitec/computer-auf-dem-weg-zum-bewusstsein-17421543.html?premium. Accessed 27 July 2021

Bauberger S et al (2021) Ethische Fragen der Künstlichen Intelligenz. In: Görz G, Braun T, Schmid U (eds) Handbuch der Künstlichen Intelligenz, 6th edn. De Gruyter, Berlin, pp 905–934

Beckert J (2016) Imagined futures: fictional expectations and capitalist dynamics. University Press, Cambridge

Bender G (2005) Technologieentwicklung als Institutionalisierungsprozess. ZfS 34:170–187

Bibel W (2006) The beginnings of AI in Germany. Künstl Intell 4(2006):48–54

Bibel W (2014) Artificial Intelligence in a historical perspective. AI Commun 27:87–102

Bijker WE, Hughes TP, Pinch TJ (eds) (1987) The social construction of technological systems. New directions in the sociology and history of technology. MIT Press, Cambridge

Bitkom, DFKI (2017) Künstliche Intelligenz: Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. Positionspapier, 05.09.2017. bitkom. https://www.bitkom.org/Bitkom/Publikationen/Entscheidungsunterstuetzung-mit-Kuenstlicher-Intelligenz.html. Accessed 25 June 2021

BMFT (Bundesministerium für Forschung und Technologie) (1988) Künstliche Intelligenz: Wissensverarbeitung und Musterverarbeitung. Bonn

Borup M, Brown N, Konrad K, van Lente H (2006) The Sociology of Expectations in Science and Technology. Technol Anal Strateg Manag 18:285–298

Botthof A, Edler J, Hahn K, Hirsch-Kreinsen H, Weber M, Wessels J (2020) Transformation des Innovationssystems: Neue Anforderungen an die Innovationspolitik. FhG ISI, Karlsruhe

Botthof A, Edler J, Hahn K, Hirsch-Kreinsen H, Weber M, Wessels J (2023) Transformative und agile Inovationssysteme. Campus, Frankfurt

Brachman R, Levesque H (2022) Machines like us. Toward AI with common sense. MIT Press, Cambridge

Brauer W (1993) KI auf dem Weg in die Normalität. KI 7:85–91

Brödner P (2019) Coping with Descartes' error in information systems. AI & Soc 34:203–213

Buchanan BG (2006) A (very) brief history of artificial intelligence. AI Mag 26:53–60

Bundesregierung (2018) Strategie Künstliche Intelligenz der Bundesregierung. 16.11.2018. https://www.bmwi.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.html. Accessed 01 July 2021

Cave S, Dihal K, Dillon S (2020) AI Narratives: A History of Imaginative Thinking about Intelligent Machines. Oxford UP, Oxford

Cyranek G, Coy W (eds) (1994) Die maschinelle Kunst des Denkens. Vieweg, Braunschweig

Dickson B (2021) Abductive inference: the blind spot of artificial intelligence. https://bdtechtalks.com/2021/09/20/myth-of-artificial-intelligence-erik-ddunieckon/. Accessed 14 Feb 2022

Dostal W (1993) Expertensysteme und Beschäftigung – Gibt es derzeit erkennbare Auswirkungen von Systemen Künstlicher Intelligenz auf Beschäftigung und Berufe? Mitt Arbeitsmarkt Berufsforsch 26:63–76

Dreyfus H (1972) What computers can't do: a critique of artificial reason. MIT Press, Cambridge

Ecker W, Coulon C-H, Kohler M (2021) KI in die Anwendung bringen – Eine Gemeinschaftsaufgabe für Hochschulen. Forschungseinrichtungen Unternehmen und Politik. Whitepaper aus der Plattform Lernende Systeme. Acatech, München

EFI (Expertenkommission Forschung und Innovation) (2022) Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2022. EFI, Berlin

Ford M (2018) Architects of Intelligence: The truth about AI from the people building it. Packt Publishing, Birmingham

Forschungsunion; acatech (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final Report of the Industry Working Group 4.0 Acatech, Frankfurt

Görz G, Braun T, Schmid U (2021) Einleitung. In: Görz G, Braun T, Schmid U (eds) Handbuch der Künstlichen Intelligenz, 6th edn. De Gruyter, Berlin, pp 1–26

Granstrand O, Holgersson M (2020) Innovation ecosystems: A conceptual review and a new definition. Technovation 90–91:102098

Heimbrecht F (2021) KI: Der nächste Winter kommt bestimmt. https://www.zukunftsinstitut.de/artikel/technologie/ki-der-naechste-winter-kommt-bestimmt/. Accessed 03 Oct 2021

Hirsch-Kreinsen H (2016) Industry 4.0« as promising technology: emergence, semantics and ambivalent character. Soziologisches Arbeitspapier 48. TU Dortmund, Dortmund

Hirsch-Kreinsen H (2023) Das Versprechen der Künstliche Intelligenz. Gesellschaftliche Dynamik einer Schlüsseltechnologie. Campus, Frankfurt

Kaltheuner F (ed) (2021) Fake AI. Space Press, Manchester

Kersting K, Tresp V (2019) Maschinelles und Tiefes Lernen. Whitepaper aus der Plattform Lernende Systeme. Acatech, München

Kieser A (1997) Rhetoric and myth in management fashion. Organization 4:49–74

Koehler J (2021) Zum Begriff der Künstlichen Intelligenz. In: Catani S, Pfeiffer J (eds) Handbuch Künstliche Intelligenz und die Künste. Preprint Version, https://www.dfki.de/fileadmin/user_upload/import/11269_KI-Kunst-JKoehler-PreprintVersion.pdf. Accessed 23 Jan 2022

Konrad E (1998) Zur Geschichte der Künstlichen Intelligenz in der Bundesrepublik Deutschland. In: Siefkes D (ed) Sozialgeschichte der Informatik. Studien zur Wissenschafts- und Technikforschung. Springer, Wiesbaden, pp 287–296

Konrad K (2006) The social dynamics of expectations: the interaction of collective and actor-specific expectations on electronic commerce and television. Technol Strateg Manag 18:429–444

Kroll H et al (2022) Schlüsseltechnologien. Studien zum deutschen Innovationssystem 7. EFI, Berlin

Larson EJ (2021) The Myth of Artificial Intelligence. Cambridge/MA/London, Belknap

Madsen DØ (2019) The emergence and rise of industry 4.0 viewed through the lens of management fashion theory. Adm Sci 9:71

Morozov E (2013) To save everything, click here – technology, solutionism and the urge to fix problems that don't exist. Public Affairs, New York.

Müller VC, Bostrom N (2016) Future progress in artificial intelligence: a survey of expert opinion. In: Müller VC (ed) Fundamental Issues of Artificial Intelligence. Springer, Berlin, pp 553–571

Münkler H (1997) Moral und Maschine. Star Trek im Spannungsfeld von Sozialutopie und technologischem Fortschritt. In: Hellmann K-U, Klein A (eds) Unendliche Weiten …« Star Trek zwischen Unterhaltung und Utopie. Suhrkamp, Frankfurt am Main, pp 59–71

Nilsson NJ (2010) The quest for artificial intelligence: a history of ideas and achievements. UP, Cambridge

Piekniewski F (2022) AI winter is well on its way https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way/ Accessed 02 Apr 2022

Reuse B (2008) Schwerpunkte der Informatikforschung in Deutschland in den 70er Jahren. In: Reuse B, Vollmar V (eds) Informatikforschung in Deutschland. Springer, Berlin/Heidelberg, pp 3–26

Russel SJ, Norvig P (2010) Artificial intelligence. A modern approach, 3rd edn. Prentice Hall, Boston

Science Media Center (2021) Transkript »Künstliche Intelligenz – Stand der Forschung und Förderung in Deutschland. https://

www.sciencemediacenter.de/fileadmin/user_upload/Press_Briefing_Zubehoer/Transkript_KI-Situation-Deutschland_SMC-Press-Briefing_2021-03-17.pdf. Accessed 04 Feb 2022

Seising R, Dittmann F (2018) Eine historisch-kritische Einführung. In: Wolfgang Bibel, Ulrich Furbach (Hg.), Formierung eines Forschungsgebiets – Künstliche Intelligenz und Intellektik an der Technischen Universität München. Preprint 15, Deutsches Museum, München

Siekmann J (1994) Künstliche Intelligenz: Von den Anfangen in die Zukunft. In: Cyranek G, Coy W (eds) Die maschinelle Kunst des Denkens. Vieweg, Braunschweig, pp 11–40

Siekmann J (2009) Die Entwicklung der Disziplin in Deutschland. Künstl Intell 23:47–52

Teich IB (2020) Meilensteine der Entwicklung Künstlicher Intelligenz. Informatik Spektrum 43:276–284

ten Hompel M, Anderl R, Schöning H (2019) Schneller zum Markterfolg. Memorandum des Forschungsbeirats der Plattform Industrie 4.1 Acatech, München

van Lente H, Rip A (1998) Expectations in technological developments: an example of prospective structures to be filled in by agency. In: Cornelis D, van der Meulen B (eds) Getting new technologies together: studies in making sociotechnical order. de Gruyter, Berlin, pp 203–229

Wahlster W (2020) Deep Learning alleine reicht nicht. In: faz-net. https://www.faz.net/-ikh-a3574. Accessed 18 Sep 2021

Weizenbaum J (1976) Computer power and human reason. Freeman, San Francisco