Philosophy & Technology
https://doi.org/10.1007/s13347-019-00345-y

EDITOR LETTER

What the Near Future of Artificial Intelligence Could Be

Luciano Floridi[1,2]

Published online: 19 March 2019

1 Introduction

Artificial intelligence (AI) has dominated recent headlines, with its promises, challenges, risks, successes, and failures. What is its foreseeable future? Of course, the most accurate forecasts are made with hindsight. But if some cheating is not acceptable, then smart people bet on the uncontroversial or the untestable. On the uncontroversial side, one may mention the increased pressure that will come from law-makers to ensure that AI applications align with socially acceptable expectations. For example, everybody expects some regulatory move from the EU, sooner or later. On the untestable side, some people will keep selling catastrophic forecasts, with dystopian scenarios taking place in some future that is sufficiently distant to ensure that the Jeremiahs will not be around to be proven wrong. Fear always sells well, like vampire or zombie movies. Expect more. What is difficult, and may be quite embarrassing later on, is to try to “look into the seeds of time, and say which grain will grow and which will not” (*Macbeth*, Act I, Scene III), that is, to try to understand where AI is more likely to go and hence where it may not be going. This is what I will attempt to do in the following pages, where I shall be cautious in identifying the paths of least resistance, but not so cautious as to avoid any risk of being proven wrong.

Part of the difficulty is to get the level of abstraction right (Floridi 2008a, 2008b), i.e. to identify the set of relevant observables (“the seeds of time”) on which to focus because those are the ones that will make the real, significant difference. In our case, I shall argue that the best observables are provided by an analysis of the *nature of the data* used by AI to achieve its performance, and of the *nature of the problems* that AI

✉ Luciano Floridi
luciano.floridi@oii.ox.ac.uk

[1] Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

[2] The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK

may be expected to solve.[1] So, my forecast will be divided into two, complementary parts. In Section 2, I will discuss the nature of the data needed by AI; and in Section 3, I will discuss the scope of the problems AI is more likely to tackle successfully. I will conclude with some more general remarks about tackling the related ethical challenges. But first, let me be clear about what I mean by AI.

2 AI: a Working Definition

AI has been defined in many ways. Today, it comprises several techno-scientific branches, well summarised in Corea (Aug. 29, Corea 2018) in Fig. 1.

Altogether, AI paradigms still satisfy the classic definition provided by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon in their seminal "Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", the founding document and later event that established the new field of AI in 1955:

> For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. (Quotation from the 2006 re-issue in (McCarthy et al. 2006))

As I have argued before (Floridi 2017), this is obviously a counterfactual: *were* a human to behave in that way, that behaviour *would* be called intelligent. It does not mean that the machine *is* intelligent or even *thinking*. The latter scenario is a fallacy and smacks of superstition. Just because a dishwasher cleans the dishes as well as, or even better than I do, it does not mean that it cleans them *like* I do, or needs any intelligence in achieving its task. The same counterfactual understanding of AI underpins the Turing test (Floridi et al. 2009), which, in this case, checks the ability of a machine to perform a task in such a way that the *outcome* would be indistinguishable from the outcome of a human agent working to achieve the same task (Turing 1950).

The classic definition enables one to conceptualise AI as a growing resource of interactive, autonomous, and often self-learning (in the machine learning sense, see Fig. 1) *agency*, that can deal with tasks that would otherwise require human intelligence and intervention to be performed successfully. This is part of the ethical challenge posed by AI, because artificial agents are

> sufficiently informed, 'smart', autonomous and able to perform morally relevant actions independently of the humans who created them […]. (Floridi and Sanders 2004)

Although this aspect is important, it is not a topic for this article, and I shall return to it briefly only in the conclusion.

[1] For a reassuringly converging review based not on the nature of data or the nature of problems, but rather on the nature of technological solutions, based on a large scale review of the forthcoming literature on AI, see "We analysed 16,625 papers to figure out where AI is headed next" https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/

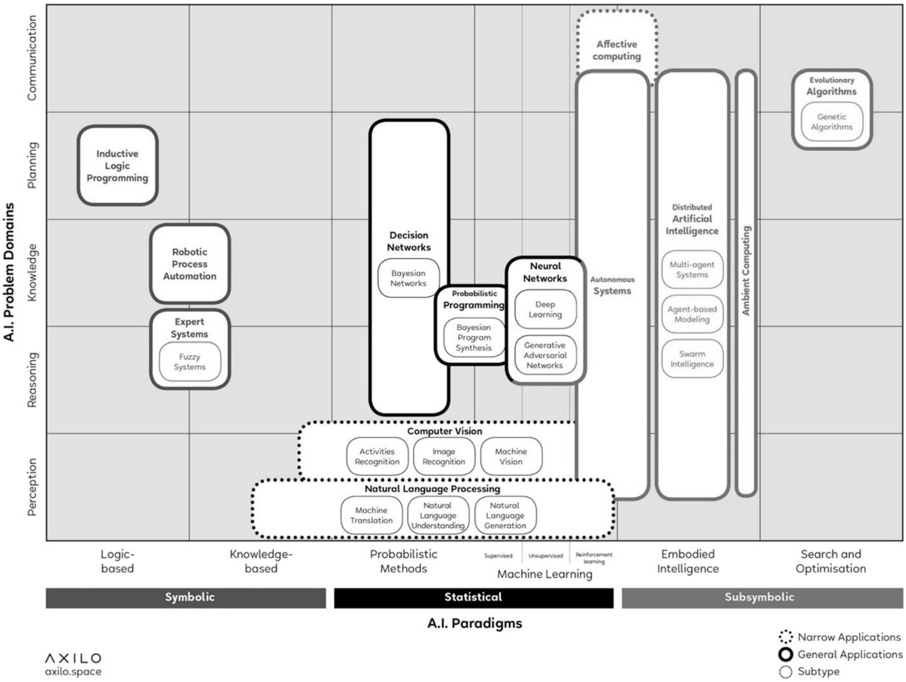

Fig. 1 AI knowledge map (AIKM), source: (Corea Aug. 29, Corea 2018)

In short, AI is defined on the basis of outcomes and actions and so, in what follows, I shall treat AI as a *reservoir of smart agency on tap*. The question I wish to address is: what are the foreseeable ways in which such a technology will evolve and be used successfully? Let us start from the data it needs.

3 AI's Future: from Historical Data to Hybrid and Synthetic Data, and the Need for Ludification

They say that data are the new oil. Maybe. But data are durable, reusable, quickly transportable, easily duplicable, and simultaneously shareable without end, while oil has none of these properties. We have gigantic quantities of data that keep growing, but oil is a finite resource. Oil comes with a clear price, whereas the monetisation of the same data depends on who is using them and for what. And all this even before introducing the legal and ethical issues that emerge when *personal* data are in play, or the whole debate about ownership ("my data" is much more like "my hands" and much less like "my oil"). So, the analogy is a stretch, to say the least. This does not mean that is entirely worthless though. Because it is true that data, like oil, are a valuable resource and must be refined in order to extract their value. In particular, without data, algorithms—AI included—go nowhere, like an engine with an empty tank. AI needs data to *train*, and then data to *apply* its training. Of course, AI can be hugely flexible; it is the data that determine its scope of application and degree of success. For example, in

2016, Google used DeepMind's machine learning system to reduce its energy consumption:

> Because the algorithm is a general-purpose framework to understand complex dynamics, we plan to apply this to other challenges in the data centre environment and beyond in the coming months. Possible applications of this technology include improving power plant conversion efficiency (getting more energy from the same unit of input), reducing semiconductor manufacturing energy and water usage, or helping manufacturing facilities increase throughput.[2]

It is well known that AI learns from the data it is fed and progressively improves its results. If you show an immense number of photos of dogs to a neural network, in the end, it will learn to recognise dogs increasingly well, including dogs it never saw before. To do this, usually, one needs huge quantities of data, and it is often the case that the more the better. For example, in recent tests, a team of researchers from the University of California in San Diego trained an AI system on 101.6 million electronic health record (EHR) data points (including text written by doctors and laboratory test results) from 1,362,559 paediatric patient visits at a major medical centre in Guangzhou, China. Once trained, the AI system was able to demonstrate:

> […] high diagnostic accuracy across multiple organ systems and is comparable to experienced pediatricians in diagnosing common childhood diseases. Our study provides a proof of concept for implementing an AI-based system as a means to aid physicians in tackling large amounts of data, augmenting diagnostic evaluations, and to provide clinical decision support in cases of diagnostic uncertainty or complexity. Although this impact may be most evident in areas where healthcare providers are in relative shortage, the benefits of such an AI system are likely to be universal. (Liang et al. 2019)

However, in recent times, AI has improved so much that, in some cases, we are moving from an emphasis on the *quantity* of large masses of data, sometimes improperly called Big Data (Floridi 2012), to an emphasis on the *quality* of data sets that are well curated. For example, in 2018, DeepMind, in partnership with Moorfields Eye Hospital in London, UK, trained an AI system to identify evidence of sight-threatening eye diseases using optical coherence tomography (OCT) data, an imaging technique that generates 3D images of the back of the eye. In the end, the team managed to

> demonstrate performance in making a referral recommendation that reaches or exceeds that of experts on a range of sight-threatening retinal diseases after training on *only 14,884 scans* [my italics]. (De et al. 2018), p. 1342.

I emphasise "only 14,884 scans" because "small data" of high quality is one of the futures of AI. AI will have a higher chance of success whenever well-curated, updated, and fully reliable data sets become available and accessible to train a system in a specific area of application. This is quite obvious and hardly a new forecast. But it is a

[2] https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

solid step forward, which helps us look further ahead, beyond the "Big Data" narrative. If *quality* matters, then *provenance* is crucial. Where do the data come from? In the previous example, they were provided by the hospital. Such data are sometimes known as *historical*, *authentic*, or *real-life* (henceforth I shall call them simply historical). But we also know that AI can generate its own data. I am not talking about *metadata* or *secondary data* about its uses (Floridi 2010). I am talking about its primary input. I shall call such *entirely* AI-generated data *synthetic*. Unfortunately, the term has an ambiguous etymology. It began to be used in the 1990s to refer to historical data that had been anonymised before being used, often to protect privacy and confidentiality. These data are synthetic only in the sense that they have been *synthesised* from historical data, e.g. through "masking".[3] They have a lower resolution, but their genesis is not an artificial source. The distinction between the historical data and those synthesised from them is useful, but this is not what I mean here, where I wish to stress the completely and exclusively *artificial provenance* of the data in question. It is an ontological distinction, which may have significant implications in terms of epistemology, especially when it comes to our ability to explain the synthetic data produced, and the training achieved by the AI using them (Watson et al. forthcoming). A famous example can help explain the difference.

In the past, playing chess against a computer meant playing against the best human players who had ever played the game. One of the features of Deep Blue, the IBM's chess program that defeated the world champion Garry Kasparov, was

> an effective use of a Grandmaster game database. (Campbell, Hoane Jr, and Hsu Campbell et al. 2002), p. 57.

But AlphaZero, the last version of the AI system developed by DeepMind, learnt to play better than anyone else, and indeed any other software, by relying only on the *rules* of the game, with no data input at all from any external source. It had no historical memory whatsoever:

> The game of chess represented the pinnacle of artificial intelligence research over several decades. State-of-the-art programs are based on powerful engines that search many millions of positions, *leveraging handcrafted domain expertise and sophisticated domain adaptations*. [my italics, these are the non-synthetic data]. AlphaZero is a generic reinforcement learning and search algorithm—originally devised for the game of Go—that achieved superior results within a few hours […] *given no domain knowledge except the rules of chess* [my italics]. (Silver et al. 2018), p. 1144.

AlphaZero learnt by playing against itself, thus generating its own chess-related, synthetic data. Unsurprisingly, Chess Grandmaster Matthew Sadler and Women's International Master Natasha Regan,

> who have analysed thousands of AlphaZero's chess games for their forthcoming book Game Changer (New in Chess, January 2019), say its style is unlike any

[3] https://www.tcs.com/blogs/the-masking-vs-synthetic-data-debate

> traditional chess engine. "It's like discovering the secret notebooks of some great player from the past," says Matthew.[4]

Truly synthetic data, as I am defining them here, have some wonderful properties. Not only do they share those listed at the beginning of this section (durable, reusable, quickly transportable, easily duplicable, simultaneously shareable without end, etc.). They are also clean and reliable (in terms of curation), they infringe no privacy or confidentiality at the *development* stage (though problems persist at the *deployment* stage, because of the predictive privacy harms (Crawford and Schultz 2014)), they are not immediately sensitive (sensitivity during the deployment stage still matters), if they are lost, it is not a disaster because they can be recreated, and they are perfectly formatted to be used by the system that generates them. With synthetic data, AI never has to leave its digital space, where it can exercise complete control on any input and output of its processes. Put more epistemologically, with synthetic data, AI enjoys the privileged position of a maker's knowledge, who knows the intrinsic nature and working of something because it made that something (Floridi 2018). This explains why they are so popular in security contexts, for example, where AI is deployed to stress-test digital systems. And sometimes synthetic data can also be produced more quickly and cheaply than historical data. AlphaZero became the best chess player on earth in 9 hours (it took 12 hours for shogi, and 13 days for Go).

Between historical data that are more or less masked (impoverished through lower resolution, e.g. through anonymisation) and purely synthetic data, there is a variety of more or less *hybrid* data, which you can imagine as the offspring of historical and synthetic data. The basic idea is to use historical data to obtain some new synthetic data that are not merely impoverished historical data. A good example is provided by Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014):

> Two neural networks—a Generator and a Discriminator [my capitals in the whole text]— compete against each other to succeed in a game. The object of the game is for the Generator to fool the Discriminator with examples that look similar to the training set. […] When the Discriminator rejects an example produced by the Generator, the Generator learns a little more about what the good example looks like. […] In other words, the Discriminator leaks information about just how close the Generator was and how it should proceed to get closer. […] As time goes by, the Discriminator learns from the training set and sends more and more meaningful signals back to the Generator. As this occurs, the Generator gets closer and closer to learning what the examples from the training set look like. *Once again, the only inputs the Generator has are an initial probability distribution (often the normal distribution) and the indicator it gets back from the Discriminator. It never sees any real examples* [my italics].[5]

The Generator learns to create synthetic data that are like some known input data. So, there is a bit of a hybrid nature here, because the Discriminator needs to have access to the historical data to "train" the Generator. But the data generated by the Generator are

[4] https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/

[5] https://securityintelligence.com/generative-adversarial-networks-and-cybersecurity-part-1/

new, not merely an abstraction from the training data. So, not a case of parthenogenesis, like AlphaZero giving birth to its own data, but close enough to deliver some of the very appealing features of synthetic data nevertheless. For example, synthetic human faces created by a Generator pose no problems in terms of privacy, consent, or confidentiality at the development stage.[6]

Many methods to generate hybrid or synthetic data are already available or being developed, often sector specific. There are also altruistic trends to make such data sets publicly and freely available (Howe et al. 2017). Clearly, the future of AI lies not just in "small data" but also, or perhaps mainly, in its increasing ability to generate its own data. That would be a remarkable development, and one may expect significant efforts to be made in that direction. The next question is: what factors can make the dial in Fig. 2 move from left to right?

The difference is made by the genetic process, i.e. by the rules used to create the data. *Historical data* are obtained by *recording rules*, as they are the outcome of some observation of a system behaviour. *Synthesised data* are obtained by *abstracting rules* that eliminate, mask or obfuscate some degrees of resolution from the historical data, e.g. through anonymisation. *Hybrid* and truly *synthetic data* can be generated by *constraining rules* or *constitutive rules*. There is no one-to-one mapping, but it is useful to consider hybrid data as the data on which we have to rely, using constraining rules, when we do not have constitutive rules that can generate synthetic data from scratch. Let me explain.

The dial moves easily towards synthetic data whenever AI deals with "games"—understood as any formal interactions in which players compete according to rules and in view of achieving a goal—the rules of which are *constitutive* and not merely *constraining*. The difference is obvious if one compares chess and football. Both are games, but in chess, the rules establish the legal and illegal moves before any chess-like activity is possible, so they are generative of all and only the acceptable moves. Whereas in football, a previous activity—let us call it kicking a ball—is "regimented" or structured by rules that arrive *after* the activity. The rules do not and cannot determine the moves of the players, they simply put boundaries to what moves are "legal". In chess, as in all board games whose rules are constitutive (draughts, Go, Monopoly, shogi…), AI can use the rules to play any possible legal move that it wants to explore. In 9 hours, AlphaZero played 44 million training games. To have a sense of the magnitude of the achievement consider that the *Opening Encyclopedia 2018* contains approximately 6.3 million games, selected from the whole history of chess. But in football, this would be meaningless because the rules do not make the game, they only shape it. This does not mean that AI cannot play virtual football; or cannot help identifying the best strategy to win against a team whose data about previous games and strategies are recorded; or cannot help with identifying potential players, or training them better. Of course, all these applications are now trivially feasible and already occur. What I mean is that when (1) a process or interaction can be transformed into a game, and (2) the game can be transformed into a *constitutive-rule* game, then (3) AI will be able to generate its own, fully synthetic data and be the best "player" on this

[6] https://motherboard.vice.com/en_us/article/7xn4wy/this-website-uses-ai-to-generate-the-faces-of-people-who-dont-exist

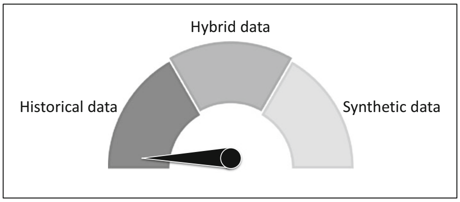

Fig. 2 Shifting from entirely historical to truly synthetic data

planet, doing what AlphaZero did for chess (in the next section, I shall describe this process as *enveloping* (Floridi 2014a)). To quote Wiener:

> The best material model of a cat is another, or preferably the same, cat. (Rosenblueth and Wiener 1945), p. 316.

Ideally, the best data on which to train an AI are either the fully historical data or the fully synthetic data generated by the same rules that generated the historical data. In any board game, this happens by default. But insofar as any of these two steps (1)–(2) is difficult to achieve, the absence of rules or the presence of merely constraining rules is likely to be a limit. We do not have the actual cat, but only a more or less reliable model of it. Things can get more complicated once we realise that, in actual games, the constraining rules are simply conventionally imposed on a previously occurring activity, whereas in real life, when we observe some phenomena, e.g. the behaviour of a kind of tumour in a specific cohort of patients in some given circumstances, the genetic rules must be extracted from the actual "game" through scientific (and these days possibly AI-based) research. For example, we do not know, and perhaps we may never know, what the exact "rules" for the development of brain tumours are. We have some general principles and theories according to which we understand their development. So, at this stage (and it may well be a permanent stage), there is no way to "ludify" (transformation into a game in the sense specified above, I avoid the term 'gamifying' which has a different and well-established meaning) brain tumours into a "constitutive-rule game" (think of chess) such that an AI system, by playing according to the identified rules, can generate its own synthetic data about brain tumours that would be equivalent to the historical data we could collect, doing for brain tumours what AlphaZero has done for chess games. This is not necessarily a problem. On the contrary, AI, by relying on historical or hybrid data (e.g. brain scans) and learning from them, can still outperform experts, and expand its capabilities beyond the finite historical data sets provided (e.g. by discovering new patterns of correlations), or deliver accessible services where there is no expertise. It is already a great success if one can extract enough *constraining* rules to produce reliable data in silico. But without a reliable system of *constitutive rules*, some of the aforementioned advantages of synthetic data would not be available in full (the vagueness of this statement is due to the fact that we can still use hybrid data).

Ludification and the presence or absence of constraining/constitutive rules are not either-or, hard limits. Recall that hybrid data can help to develop synthetic data. What is likely to happen is that, in the future, it will become increasingly clear when high-quality databases of historical data may be absolutely necessary and unavoidable—

when you need the actual cat, to paraphrase Wiener—and hence when we will have to deal with issues about availability, accessibility, legal compliance with legislation, and, in the case of personal data, privacy, consent, sensitivity, and other ethical questions. However, the trend towards the generation of as-synthetic-as-possible (synthesised, more or less hybrid, all the way to fully synthetic) data is likely to be one of AI's holy grails, so I expect the AI community to push very hard in that direction. Generating increasingly non-historical data, making the dial move as far as possible to the right in Fig. 2, will require a "ludification" of processes, and for this reason I also expect the AI community to be increasingly interested in the gaming industry, because it is there that the best expertise in "ludification" is probably to be found. And in terms of negative results, mathematical proofs about the impossibility of ludifying whole kinds or areas of processes or interactions should be most welcome in order to clarify where or how far an AlphaZero-like approach may never be achievable by AI.

4 AI's Future: from Difficult Problems to Complex Problems, and the Need for Enveloping

I have already mentioned that AI is best understood as a reservoir of agency that can be used to solve problems. AI achieves its problem-solving goals by detaching the ability to perform a task successfully from any need to be intelligent in doing so. The App on my mobile phone does not need to be intelligent to play chess better than I do. Whenever this detachment is feasible, some AI solution becomes possible in principle. This is why understanding the future of AI also means understanding the nature of problems where such a detachment may be technically feasible in theory and economically viable in practice. Now, many of the problems we try to solve through AI occur in the physical world, from driving to scanning labels in a supermarket, from cleaning flows or windows to cutting the grass in the garden. The reader may keep in mind AI as robotics in the rest of this section, but I am not discussing only robotics: smart applications and interfaces in the Internet of Things are also part of the analysis, for example. What I would like to suggest is that, for the purpose of understanding AI's development when dealing with physical environments, it is useful to map problems on the basis of what resources are needed to solve them, and hence how far AI can have such resources. I am referring to *computational resources*, and hence to degrees of *complexity*; and to *skill-related resources*, and hence to degrees of *difficulty*.

The degrees of complexity of a problem are well known and extensively studied in computational theory (Arora and Barak 2009; Sipser 2012). I shall not say much about this dimension but only remark that it is highly quantitative and that the mathematical tractability it provides is due to the availability of standard criteria of comparison, perhaps even idealised but clearly defined, such as the computational resources of a Turing Machine. If you have a "metre", then you can measure lengths. Similarly, if you adopt a Turing Machine as your starting point, then you can calculate how much time, in terms of steps, and how much space, in terms of memory or tape, a computational problem consumes to be solved. For the sake of simplicity—and keeping in mind that finely grained and sophisticated degrees of precision can be achieved, if needed, by using tools from complexity theory—let us agree to map the complexity of a problem (dealt with by AI in terms of space–time = memory steps required) from 0 (*simple*) to 1 (*complex*).

The degrees of difficulty of a problem, understood in terms of the skills required to solve it, from turning on and off a light to ironing shirts, need a bit more of a stipulation to be mapped here because usually, the relevant literature, e.g. in human motor development, does not focus on a taxonomy of problems based on resources needed, but on a taxonomy of the performance of the human agents assessed and their abilities or skills demonstrated in solving a problem or performing a task. It is also a more qualitative literature. In particular, there are many ways of assessing a performance and hence many ways of cataloguing skill-related problems, but one standard distinction is between gross and fine motor skills. Gross motor skills require the use of large muscle groups to perform tasks like walking or jumping, catching or kicking a ball. Fine motor skills require the use of smaller muscle groups, in the wrists, hands, fingers, and the feet and toes, to perform tasks like washing the dishes, writing, typing, using a tool, or playing an instrument. Despite the previous difficulties, you can see immediately that we are dealing with different degrees of difficulty. Again, for the sake of simplicity—and recalling that finely grained and sophisticated degrees of precision can be achieved, if needed, by using tools from developmental psychology—let us agree to map the difficulty of a problem (dealt with by AI in terms of skills required) from 0 (*easy*) to 1 (*difficult*).

We are now ready to map the two dimensions in Fig. 3, where I have added four examples.

Turning the light on is a problem whose solution has a very low degree of complexity (very few steps and states) and of difficulty (even a child can do it). However, tying one’s own shoes requires advanced motor skills, and so does lacing them, thus it is low in complexity (simple), but it is very high in difficulty. As Adidas CEO Kasper Rorsted remarked in 2017:

> The biggest challenge the shoe industry has is how do you create a robot that puts the lace into the shoe. I’m not kidding. That’s a complete manual process today. There is no technology for that.[7]

Dishwashing is the opposite: it may require a lot of steps and space, indeed increasingly more the more dishes need to be cleaned, but it is not difficult, even a philosopher like me can do it. And of course, top-right we find ironing shirts, which is both resource-consuming, like dishwashing, and demanding in terms of skills, so it is both complex and difficult, which is my excuse to try to avoid it. Using the previous examples of playing football and playing chess, football is simple but difficult, chess is easy (you can learn the rules in a few minutes) but very complex, this is why AI can win against anyone at chess, but a team of androids that wins the world cup is science fiction. Of course, things are often less clear-cut, as table tennis robots show.

The reader will notice that I placed a dotted arrow moving from low-complexity high-difficulty to high-complexity low-difficulty.[8] This seems to me the arrow that

[7] https://qz.com/966882/robots-cant-lace-shoes-so-sneaker-production-cant-be-fully-automated-just-yet/

[8] I am not the first to make this point, see for example: https://www.campaignlive.co.uk/article/hard-things-easy-easy-things-hard/1498154

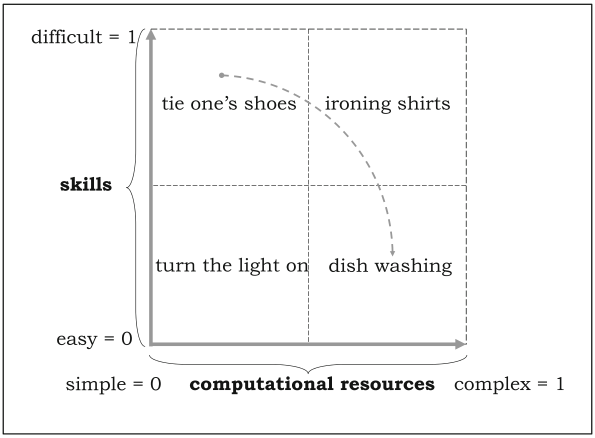

Fig. 3 Translating difficult tasks into complex tasks

successful developments of AI will follow. Our artefacts, no matter how smart, are not really good at performing tasks and hence solving problems that require high degrees of skilfulness. However, they are fantastic at dealing with problems that require very challenging degrees of complexity. So, the future of successful AI probably lies not only in increasingly hybrid or synthetic data, as we saw, but also in translating difficult tasks into complex tasks.

How is this translation achieved? By transforming the environment within which AI operates into an AI-friendly environment. Such translation may increase the complexity of what the AI system needs to do enormously but, as long as it decreases the difficulty, it is something that can be progressively achieved more and more successfully. Some examples should suffice to illustrate the point, but first, let me introduce the concept of *enveloping*.

In industrial robotics, the three-dimensional space that defines the boundaries within which a robot can work successfully is defined as the robot's *envelope*. We do not build droids like Star Wars' C3PO to wash dishes in the sink exactly in the same way as we would. We envelop environments around simple robots to fit and exploit their limited capacities and still deliver the desired output. A dishwasher accomplishes its task because its environment—an openable, waterproof box—is structured ("enveloped") around its simple capacities. The more sophisticated these capacities are, the less enveloping is needed, but we are looking at trade-off, some kind of equilibrium. The same applies to Amazon's robotic shelves, for example. It is the whole warehouse that is designed to be robot-friendly. Ditto for robots that can cook[9] or flip hamburgers,[10] which already exist. Driverless cars will become a commodity the day we can successfully envelop the environment around them. This is why it is plausible that in an airport, which is a highly controlled and hence more easily "envelopable" environment, a shuttle could be an autonomous vehicle, but not the school bus that serves my village, given that the bus driver needs to be able to operate in extreme and difficult circumstances (countryside, snow, no signals, no satellite coverage etc.) that are most unlikely

[9] http://www.moley.com/
[10] https://misorobotics.com/

(mind, not impossible) to be enveloped. In 2016, Nike launched HyperAdapt 1.0, its automatic electronic self-lacing shoes, not by developing an AI that would tie them for you, but by re-inventing the concept of what it means to adapt shoes to feet: each shoe has a sensor, a battery, a motor, and a cable system that, together, can adjust fit following an algorithmic pressure equation.[11] Enveloping used to be either a stand-alone phenomenon (you buy the robot with the required envelop, like a dishwasher or a washing machine) or implemented within the walls of industrial buildings, carefully tailored around their artificial inhabitants. Nowadays, enveloping the environment into an AI-friendly infosphere has started to pervade all aspects of reality and is happening daily everywhere, in the house, in the office, and in the street. We have been enveloping the world around digital technologies for decades, invisibly and without fully realising it. The future of AI also lies in more enveloping, for example, in terms of 5G and the Internet of Things, but also insofar as we are all more and more connected and spend more and more time "onlife", and all our information is increasingly born digital. In this case too, some observations may be obvious. There may be problems, and hence relative tasks that solve them, that are not easily subject to enveloping. Yet here it is not a matter of mathematical proofs, but more of ingenuity, economic costs, and user or customer preferences. For example, a robot that iron shirts can be engineered. In 2012, a team at Carlos III University of Madrid, Spain, built TEO, a robot that weighs about 80 kg and is 1.8 m tall. TEO can climb stairs, open doors and, more recently, has been shown to be able to iron shirts (Estevez et al. 2017), although you have to put the item on the ironing board. The view, quite widespread, is that

> 'TEO is built to do what humans do as humans do it,' says team member Juan Victores at Carlos III University of Madrid. He and his colleagues want TEO to be able to tackle other domestic tasks, like helping out in the kitchen. Their ultimate goal is for TEO to be able to learn how to do a task just by watching people with no technical expertise carry it out. 'We will have robots like TEO in our homes. It's just a matter of who does it first,' says Victores.

And yet, I strongly doubt this is the future. It is a view that fails to appreciate the distinction between difficult and complex tasks and the enormous advantage of enveloping tasks to make them easy (very low difficulty), no matter how complex. Recall that we are building autonomous vehicles not by putting robots in the driving seat, but by rethinking the whole ecosystem of vehicles plus environments, that is, removing the driving seat altogether. So, if my analysis is correct, the future of AI is not full of TEO-like androids that mimic human behaviour, but is more likely represented by Effie,[12] Foldimate,[13] and other similar domestic automated machines that dry and iron clothes. They are not androids, like TEO, but box-like systems that may be quite sophisticated computationally. They look more like dishwasher and washing machines, with the difference that, in their enveloped environments, their input is wrinkled clothes and their output is ironed ones.

[11] Strange things happen when the software does not work properly: https://www.bbc.co.uk/news/business-47336684

[12] https://helloeffie.com/

[13] https://foldimate.com/

Perhaps similar machines will be expensive, perhaps they may not always work as well as one may wish, perhaps they may be embodied in ways we cannot imagine now, but you can see how the logic is the correct one: do not try to mimic humans through AI but exploit what machines, AI included, do best. *Difficulty* is the enemy of machines, *complexity* is their friend, so envelop the world around them, design new forms of embodiment to embed them successfully in their envelop, and at that point progressive refinements, market scale, and improvements will become perfectly possible.

5 Conclusion: a Future of Design

The two futures I have outlined here are complementary and based on our current and foreseeable understanding of AI. There are unknown unknowns, of course, but all one can say about them is precisely this: they exist, and we have no idea about them. It is a bit like saying that we know there are questions we are not asking but cannot say what these questions are. The future of AI is full of unknown unknowns. What I have tried to do in this article is to look at the "seeds of time" that we have already sowed. I have concentrated on the nature of data and of problems because the former are what enable AI to work, and the latter provide the boundaries within which AI can work successfully. At this level of abstraction, two conclusions seem to be very plausible. We will seek to develop AI by using data that are as much as possible hybrid and preferably synthetic, through a process of ludification of interactions and tasks. In other words, the tendency will be to try to move away from purely historical data whenever possible. And we will do so by translating as much as possible difficult problems into complex problems, through the enveloping of realities around the skills of our artefacts. In short, we will seek to create hybrid or synthetic data to deal with complex problems, by ludifying tasks and interactions in enveloped environments. The more this is possible, the more successful AI will be, which leads me to two final comments.

Ludifying and enveloping are a matter of *designing*, or sometimes re-designing, the realities with which we deal (Floridi 2019). So, the foreseeable future of AI will depend on our design abilities and ingenuity. It will also depend on our ability to negotiate the resulting (and serious) ethical, legal, and social issues (ELSI), from new forms of privacy (predictive or group-based (Floridi 2014c)) to nudging and self-determination. The very idea that we are increasingly shaping our environments (analog or digital) to make them AI-friendly should make anyone reflect (Floridi 2013). Anticipating such issues, to facilitate positive ELSI and avoid or mitigate any negative ones, is the real value of any foresight analysis. It is interesting to try to understand what the paths of least resistance may be in the evolution of AI. But it would be quite sterile to try to predict "which grain will grow and which will not" and then to do nothing to ensure that the good grains grow, and the bad ones do not (Floridi 2014b). The future is not entirely open (because the past shapes it), but neither is it entirely determined, because the past can be steered in a different direction. This is why the challenge ahead will not be so much digital innovation per se, but the governance of the digital, AI included.

Acknowledgements I would like to thank all members of the Digital Ethics Lab, OII, Univeristy of Oxford, for many discussions about some of the topics covered in this article, and Nikita Aggarwal, Josh Cowls, Jessica Morley, David Sutcliffe, and Mariarosaria Taddeo for their hugely helpful comments on a last draft.

References

Arora, S., & Barak, B. (2009). *Computational complexity: a modern approach*. Cambridge: Cambridge University Press.

Campbell, M., Joseph Hoane, A., Jr., & Hsu, F.-h. J. (2002). Deep blue. *Artificial Intelligence, 134*(1–2), 57–83.

Corea, Francesco 2018. "AI knowledge map: how to classify AI technologies, a sketch of a new AI technology landscape." Medium - artificial intelligence https://medium.com/@Francesco_AI/ai-knowledge-map-how-to-classify-ai-technologies-6c073b969020.

Crawford, K., & Schultz, J. (2014). Big data and due process: toward a framework to redress predictive privacy harms. *BCL Rev., 55*, 93.

De, F., Jeffrey, J. R. L., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C. O., Raine, R., Hughes, J., Sim, D. A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P. T., Suleyman, M., Cornebise, J., Keane, P. A., & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine, 24*(9), 1342–1350.

Estevez, David, Juan G Victores, Raul Fernandez-Fernandez, and Carlos Balaguer. 2017. "Robotic ironing with 3D perception and force/torque feedback in household environments." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Floridi, L. (2008a). The method of levels of abstraction. *Minds and Machines, 18*(3), 303–329.

Floridi, L. (2008b). Understanding epistemic relevance. *Erkenntnis, 69*(1), 69–92.

Floridi, L. (2010). *Information: a very short introduction*. Oxford: Oxford University Press.

Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology, 25*(4), 435–437.

Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

Floridi, L. (2014a). *The fourth revolution - how the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Floridi, L. (2014b). Technoscience and ethics foresight. *Philosophy & Technology, 27*(4), 499–501.

Floridi, L. (2014c). Open data, data protection, and group privacy. *Philosophy & Technology, 27*(1):1–3.

Floridi, L. (2017). Digital's cleaving power and its consequences. *Philosophy & Technology, 30*(2), 123–129.

Floridi, L. (2018). What the maker's knowledge could be. *Synthese, 195*(1), 465–481.

Floridi, L. (2019). *The logic of information*. Oxford: Oxford University Press.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379.

Floridi, L., Taddeo, M., & Turilli, M. (2009). Turing's imitation game: still an impossible challenge for all machines and some judges—an evaluation of the 2008 Loebner contest. *Minds and Machines, 19*(1), 145–150.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*.

Howe, Bill, Julia Stoyanovich, Haoyue Ping, Bernease Herman, and Matt Gee. 2017. Synthetic data for social good. arXiv preprint arXiv:1710.08874.

Liang, Huiying, Brian Y. Tsui, Hao Ni, Carolina C. S. Valentim, Sally L. Baxter, Guangjian Liu, Wenjia Cai, Daniel S. Kermany, Xin Sun, Jiancong Chen, Liya He, Jie Zhu, Pin Tian, Hua Shao, Lianghong Zheng, Rui Hou, Sierra Hewett, Gen Li, Ping Liang, Xuan Zang, Zhiqi Zhang, Liyan Pan, Huimin Cai, Rujuan Ling, Shuhua Li, Yongwang Cui, Shusheng Tang, Hong Ye, Xiaoyan Huang, Waner He, Wenqing Liang, Qing Zhang, Jianmin Jiang, Wei Yu, Jianqun Gao, Wanxing Ou, Yingmin Deng, Qiaozhen Hou, Bei Wang, Cuichan Yao, Yan Liang, Shu Zhang, Yaou Duan, Runze Zhang, Sarah Gibson, Charlotte L. Zhang, Oulan Li, Edward D. Zhang, Gabriel Karin, Nathan Nguyen, Xiaokang Wu, Cindy Wen, Jie Xu, Wenqin Xu, Bochu Wang, Winston Wang, Jing Li, Bianca Pizzato, Caroline Bao, Daoman Xiang, Wanting He, Suiqin He, Yugui Zhou, Weldon Haw, Michael Goldbaum, Adriana Tremoulet, Chun-Nan Hsu, Hannah Carter, Long Zhu, Kang Zhang, and Huimin Xia. 2019. "Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence." Nature Medicine.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine, 27*(4), 12.

Rosenblueth, A., & Wiener, N. (1945). The role of models in science. *Philosophy of Science, 12*(4), 316–321.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419), 1140–1144.

Sipser, M. (2012). *Introduction to the theory of computation* (3rd ed.). Boston, MA: Cengage Learning.

Watson, David S., Jenny Krutzinna, Ian N. Bruce, Christopher E.M. Griffiths, Iain B. McInnes, Michael R. Barnes, and Luciano Floridi. forthcoming. Clinical applications of machine learning algorithms: beyond the black box. *British Medical Journal*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.