

DARIA KIM<sup>1</sup>/ MAXIMILIAN ALBER<sup>2</sup> / MAN WAI KWOK<sup>3</sup> / JELENA MITROVIĆ<sup>4</sup> / CRISTIAN RAMIREZ-ATENCIA<sup>5</sup>  
 JESÚS ALBERTO RODRÍGUEZ PÉREZ<sup>6</sup> / HEINER ZILLE<sup>7</sup>

## Clarifying Assumptions About Artificial Intelligence Before Revolutionising Patent Law

This paper examines several widespread assumptions about artificial intelligence, particularly machine learning, that are often taken as factual premises in discussions on the future of patent law in the wake of ‘artificial ingenuity’. The objective is to draw a more realistic and nuanced picture of the human-computer interaction in solving technical problems than where ‘intelligent’ systems autonomously yield inventions. A detailed technical perspective is presented for each assumption, followed by a discussion of pertinent uncertainties for patent law. Overall, it is argued that implications of machine learning for the patent system in its core tenets appear far less revolutionary than is often posited.

### I. Introduction

1. The controversy: A tool or more than a tool?
2. The purpose and scope of the analysis

### II. Clarifying the assumptions

1. A human states a problem – an ML system solves it
2. ML systems learn ‘complex rules’ based on ‘simple rules’
3. ML systems are autonomous, non-deterministic and unpredictable
4. ML is a ‘black box’
5. ML is a ‘general-purpose technology’
6. ML is a ‘general-purpose method of invention’
7. One ML algorithm – many inventions

### III. Synthesis and research outlook

1. ML as computational techniques of problem-solving
2. Implications for patent law

### Abbreviations

AI (artificial intelligence); ANN (artificial neural network); DL (deep learning); EA (evolutionary algorithm); GP (genetic programming); EPC (European Patent Convention); EPO (European Patent Office); GPT (general purpose technology); GPT-3 (Generative

Pre-trained Transformer 3); HLEG-AI (High-Level Expert Group on Artificial Intelligence); ML (machine learning); RL (reinforcement learning); RNG (random number generator)

### I. Introduction

#### 1. The controversy: A tool or more than a tool?

The European Commission’s initiative ‘Promote the Scientific Exploration of Computational Creativity’ launched in 2013 defines computational creativity as a ‘burgeoning area of creativity research that brings together academics and practitioners from diverse disciplines, genres and modalities, to explore the potential of computers to be *autonomously creative* or to collaborate as co-creators with humans’.<sup>8</sup> From an engineering perspective, the aspiration is ‘to construct autonomous software artifacts that achieve novel and useful ends that are deserving of the label “creative”’.<sup>9</sup>

Some researchers share the view that artificial intelligence (AI) is ‘no longer a tool’, not even ‘a very sophisticated tool’.<sup>10</sup> Stephen Thaler refers to an artificial neural network (ANN)-based system invented and patented by him in 1997 as the ‘creativity machine’ and claims that it can ‘perform imaginative feats that extend beyond technological invention into the realms of aesthetics and emotions’.<sup>11</sup> More recently, Thaler designated the

<sup>1</sup> MA, LL.M., Dr iur., Senior research fellow at Max Planck Institute for Innovation and Competition, Munich, Germany; [daria.kim@ip.mpg.de](mailto:daria.kim@ip.mpg.de).

<sup>2</sup> BSc, MSc, Dr, Chief Technical Officer at Aignostics; PostDoc at Charité Institute of Pathology, Berlin, Germany.

<sup>3</sup> BSc (Hons), MSc (Engineering and Data Science); Chinese University of Hong Kong, Department of Physics (formerly).

<sup>4</sup> Dr, Junior Research Group Leader supported by the BMBF program for Women in AI, Faculty of Computer Science and Mathematics, Chair of Data Science, University of Passau, Germany; the Institute for Artificial Intelligence Research and Development of Serbia.

<sup>5</sup> Dr, Assistant Professor at Department of Computer Systems, Universidad Politécnica de Madrid, Spain.

<sup>6</sup> Dr, Lecturer at the School of Computing Science, College of Science and Engineering, University of Glasgow, UK.

<sup>7</sup> Dr-Ing, Research and Teaching Associate at the Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany.

<sup>8</sup> <<https://cordis.europa.eu/project/id/600653>> accessed 8 October 2021 (emphasis added).

<sup>9</sup> *ibid.*

<sup>10</sup> Laura Butler, ‘World First Patent Applications Filed for Inventions Generated Solely by Artificial Intelligence’ (1 August 2019) <<https://www.surrey.ac.uk/news/world-first-patent-applications-filed-inventions-generated-solely-artificial-intelligence>> accessed 21 May 2021 (citing Adrian Hilton, Director of the Centre for Vision, Speech and Signal Processing at the University of Surrey).

<sup>11</sup> US patent 5,659,666 ‘Device for the autonomous generation of useful information’.

ANN-based system DABUS,<sup>12</sup> also previously patented by him,<sup>13</sup> as an inventor in two patent applications.<sup>14</sup> Thaler is not original in his intention to enhance the status of computational systems. The mathematician Doron Zeilberger, for instance, often credits a computer – which he named Shalosh B. Ekhad – as his co-author.<sup>15</sup>

In contrast, others have viewed mathematics as ‘a fundamental intellectual tool in computing’, computers as a tool in mathematical problem-solving, and Shalosh B. Ekhad as Doron Zeilberger’s pseudonym.<sup>16</sup> Commentators have expressed reservations regarding the ability of AI systems to invent autonomously,<sup>17</sup> arguing that the assertion that ‘machines have been autonomously generating patentable results for at least twenty years’<sup>18</sup> is ‘simply wrong’;<sup>19</sup> that AI techniques are tools in a (human) inventor’s hands;<sup>20</sup> and that ‘fully autonomous creation or invention by AI does not exist, and will not exist for the foreseeable future’.<sup>21</sup> In the realm of artistic expression it has been argued that ‘[c]omputers do not create art, people using computers create art’;<sup>22</sup> that ‘all “computer-generated art” is the result of human invention, software development, tweaking, and other kinds of direct control and authorship’; and that ‘the human is always the mastermind behind the work’.<sup>23</sup>

Notably, the anthropomorphisation of AI has been criticised within the AI community itself.<sup>24</sup> As one professor of computer science once observed, a researcher who employs a ‘wishful’ language endowing computers with understanding and intentionality when referring to computer programs and data structures is misleading the general public and, ‘most prominently [...] himself’.<sup>25</sup> Studies show that the anthropomorphised depiction of AI

can affect the human perception of AI systems, particularly the level of trust and the attribution of responsibility in automated driving<sup>26</sup> and even the allocation of credit for AI-generated artistic output.<sup>27</sup> AI researchers caution that ‘we must be careful’ about anthropomorphising AI, which is ‘math, statistical analysis and pattern-matching’,<sup>28</sup> and that the pervasive use of anthropomorphic language ‘inadvertently promotes misleading interpretations of and beliefs about what AI is and what its capacities are’, and can also ‘have epistemological impact on the AI research community itself’.<sup>29</sup>

## 2. The purpose and scope of the analysis

The assumption that AI has already replaced<sup>30</sup> – or is about to replace<sup>31</sup> – human creators and inventors prompted the arguments that the output ‘autonomously’ generated by such systems does not and should not merit IP protection.<sup>32</sup> The perception that AI systems are autonomous agents endowed with problem-solving, decision-making and inventive capabilities<sup>33</sup> provoked several reformist proposals concerning patent law, including abolishing the patent system altogether,<sup>34</sup> recognising AI as an inventor<sup>35</sup> and replacing a ‘POSITA’ (a person ordinarily skilled in the art) with a ‘MOSITA’ (a machine

12 <[http://imagination-engines.com/iei\\_dabus.php#DABUS%20Patent](http://imagination-engines.com/iei_dabus.php#DABUS%20Patent)> accessed 21 May 2021.

13 US patent 10,423,875 ‘Electro-optical device and method for identifying and inducing topological states formed among interconnecting neural modules’.

14 European patent application 18275163; European patent application 18275174.

15 <<https://sites.math.rutgers.edu/~zeilberg/papers1.html>> accessed 21 May 2021.

16 Zhiwen Hu and others, ‘Shalosh B. Ekhad: A Computer Credit for Mathematicians’ (2020) 122 *Scientometrics* 71, 93.

17 Dan L Burk, ‘AI Patents and the Self-Assembling Machine’ (2021) 105 *Minn. L. Rev. Headnotes* 301, 317 (pointing out that ‘[a]sserting that AI tools are either inventors or infringers is equally absurd and can only be based on ignoring the human hand at work behind the AI’). See also Rose Hughes, ‘Is It Time to Move On from the AI Inventor Debate?’ (*IPKat*, 2 December 2020) <<https://ipkitten.blogspot.com/2020/12/is-it-time-to-move-on-from-ai-inventor.html>> accessed 21 May 2021.

18 Ryan Abbott, ‘I Think, Therefore I Invent: Creative computers and the future of patent law’ (2016) 57(4) *B.C.L. Rev.* 1079, 1081.

19 Mark Summerfield, ‘The Impact of Machine Learning on Patent Law, Part 2: “Machine-assisted inventing”’ (*Patentology*, 21 January 2018) <[https://blog.patentology.com.au/2018/01/the-impact-of-machine-learning-on\\_21.html](https://blog.patentology.com.au/2018/01/the-impact-of-machine-learning-on_21.html)> accessed 21 May 2021.

20 Peter Blok, ‘The Inventor’s New Tool: Artificial Intelligence – How Does It Fit in the European Patent System?’ (2017) 39(2) *E.I.P.R.* 69, 73; Burk (n 17) 317 ff; Daria Kim, ‘“AI-Generated Inventions”: Time to Get the Record Straight?’ [2020] *GRUR International* 443, 455.

21 P Bernt Hugenholtz and others, *Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework* (European Commission 2020) 6, 116.

22 Aaron Hertzmann, ‘Can Computers Create Art?’ (2018) 7(2) *Arts* 1, 2, doi: 10.3390/arts7020018 (emphasised in the original).

23 *ibid.*

24 Donald Clark, *Artificial Intelligence for Learning: How to Use AI to Support Employee Development* (Kogan Page 2020) 30-31 (referring to the anthropomorphisation of AI as a ‘dangerous tendency’ and ‘a type of category mistake’).

25 Drew McDermott, ‘Artificial Intelligence Meets Natural Stupidity’ (1976) 57 *ACM SIGART Bulletin* 4, at 4.

26 Adam Waytz, Joy Heafner and Nicholas Epley, ‘The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle’ (2014) 52 *Journal of Experimental Social Psychology* 113.

27 Ziv Epstein and others, ‘Who Gets Credit for AI-Generated Art?’ 23(101515) *iScience* (2020), doi: 10.1016/j.isci.2020.101515.

28 Clark (n 24) 231.

29 Arleen Salles, Kathinka Evers and Michele Farisco, ‘Anthropomorphism in AI’ 11 *AJOB Neuroscience* 88, 93 (2020). See also ‘Open Letter of to the European Commission. Artificial Intelligence and Robotics’ para 2 <<http://www.robotics-openletter.eu/>> accessed 21 July 2021 (warning that an understanding of AI as ‘self-learning’ is ‘superficial’ and reflects ‘an overvaluation of the actual capabilities’ of AI systems). See generally David Watson, ‘The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence’ (2019) 29 *Minds and Machines* 417.

30 Abbott (n 18) 1081 (assuming that ‘machines have been autonomously generating patentable results for at least twenty years’).

31 See eg Daniele Fabris, ‘From the PHOSITA to the MOSITA: Will “secondary considerations” save pharmaceutical patents from artificial intelligence?’ (2020) 51 *IIC* 685, 690 (stating that ‘in the near future, AI’s computational power is likely to supplant human research ability’, and that ‘more and more inventions might soon be made by a machine rather than by a human being’ (referencing Ryan Abbott)). See also Liza Vertinsky and Todd M Rice, ‘Thinking About Thinking Machines: Implications of Machine Inventors for Patent Law’ (2002) 8(2) *B. U. J. Sci. & Tech. L.* 574, 586 (stating that the ‘notions of scarcity and uniqueness of inventive talent will dissipate as computers make the process of invention cheaper and more readily available to people without special skills’).

32 See eg Shlomid Yanisky-Ravid and Xiaoqiong (Jackie) Liu, ‘When Artificial Intelligence Systems Produce Inventions: The 3A era and an alternative model for patent law’ (2018) 39 *Cardozo L. Rev.* 2215, 2216 (arguing that the ‘traditional approach to patent law in which policy makers seek to identify the human inventor behind the patent is [...] no longer relevant’). See also Martin Stierle, ‘A *De Lege Ferenda* Perspective on Artificial Intelligence Systems Designated as Inventors in the European Patent System’ [2021] *GRUR International* 115, 116, note 18 (referencing divergent views on the relevance of the patent system in view of the inventive potential of AI).

33 Below (nn 44, 45, 152-156, 334-337) and the accompanying text.

34 Yanisky-Ravid and Liu (n 32) 2222. See also Tim W Dornis, ‘Artificial Intelligence and Innovation: The End of Patent Law as We Know It’ (2020) 23 *Yale Journal of Law & Technology* 97 (predicting that ‘[w]ith the advent of [AI] the end of patent law is near’ and stating that ‘humans no longer stand at the center of the creative universe – we are no longer the masters of innovation’).

35 See eg Ryan Abbott, ‘The Artificial Inventor Project’ (2019) 6 *WIPO Magazine* <[https://www.wipo.int/wipo\\_magazine/en/2019/06/article\\_0002.html](https://www.wipo.int/wipo_magazine/en/2019/06/article_0002.html)> accessed 25 September 2021 (arguing that AI ‘should be listed as an inventor’ in order to ‘protect the moral rights of traditional human inventors and the integrity of the patent system’).

ordinarily skilled in the art).<sup>36</sup> Claims that computers autonomously generate inventions have extended beyond the legal scholarship<sup>37</sup> and entered litigation.<sup>38</sup>

On closer inspection, the polarisation of the discourse on patent law and AI ostensibly stems from divergent understandings of AI techniques and state of the art research on AI. As for the normative perspective, there can hardly be room for disagreement on the point that, if inventions could indeed be generated by computers autonomously, the need for the patent system and its design would need to be fundamentally reconsidered. The question is, precisely, whether the anthropomorphised depictions of AI and its claimed capabilities can be taken for granted as a basis for legal analysis and policy recommendations.

In this article, we take a close look at a set of widespread assumptions about AI – more specifically, machine learning (ML) – that are often taken as factual premises in the discussions on the future of patent law in the wake of ‘artificial ingenuity’. The common denominator is that, if taken at face value, such assumptions prompt a radical update of the patent system, if not call into question its very existence. The objective is to draw a more realistic and nuanced picture of the human-computer interaction in solving (technical) problems than the one where AI ‘spits out’ inventions ‘after a button press’.<sup>39</sup> While it is generally recognised that ML is not a ‘magic’ tool,<sup>40</sup> several technical aspects require clarification to understand more precisely their implications for patent law.

The discussion focuses on ML techniques<sup>41</sup> – in particular, ANNs and genetic programming (GP)/evolutionary

algorithms (EAs)<sup>42</sup> – given their increasing application in technical problem-solving such as technical constraint-satisfaction and constraint-optimisation problems.<sup>43</sup> By an ‘AI/ML system’, this paper refers to the combination of software and hardware necessary to implement an AI/ML method.

As for the legal context, the analysis draws mainly on European patent law, with occasional recourse to other jurisdictions. At the same time, the technical aspects of AI discussed in this paper concern the basic tenets of patent law that can be relevant for any jurisdiction dealing with the implications of AI for patent law and policy. Besides the much-debated issue of the genesis and allocation of inventor’s rights in inventions allegedly ‘generated’ by AI, the scope of the analysis includes substantive patent law questions such as patent eligibility, the sufficiency-of-disclosure requirement and standard, the definition of a skilled person and the fulfilment of the inventive step requirement. More broadly, we reflect on the role of patent protection given the general-purpose-technology nature of ML. Part II discusses each assumption according to a threefold structure: first, an assumption and challenges it provokes for patent law are stated; second, a detailed technical perspective is presented; third, patent law uncertainties are revisited and clarified in light of the technical explanation. Part III situates the question of whether AI is a tool or ‘more’ than a tool within a broader context of research in cognitive science and puts the implications of AI for deontological and economic underpinnings of patent law into perspective.

## II. Clarifying the assumptions

### 1. A human states a problem – an ML system solves it

Narratives about ML might create an impression that a mere statement of *what* needs to be achieved without *how* that should be done can be enough for a computer to accomplish a task. For instance, EAs are vested with a

<sup>36</sup> Fabris (n 31).

<sup>37</sup> For an overview, see Hugenholz and others (n 21) 100 ff.

<sup>38</sup> Recent judgments at the time of writing include: *Thaler v Comptroller General of Patents Trade Marks and Designs* [2021] EWCA Civ 1374 (21 September 2021); *Stephen Thaler v Andrew Hirshfeld, Performing the Functions & Duties of the Under Sec’y of Com. for Intell. Prop. & Dir. of the United States Pat. & Trademark Off., et al.*, No. 1:20-cv-903 (LMB/TCB), 2021 WL 3934803 (E.D. Va. 2 September 2021); *Thaler v Taiwan IP Office (TIPO)*, Administrative decision No. 2021-Xing-Zhuan-Su-3 of Intellectual Property and Commercial Court of Taiwan (Aug 2021); *Thaler v Commissioner of Patents* [2021] FCA 879 (30 July 2021).

<sup>39</sup> Robin C Feldman and Nick Thieme, ‘Competition at the Dawn of Artificial Intelligence’ in Björn Lundqvist and Michael S Gal (eds), *Competition Law for the Digital Economy* (Edward Elgar Publishing 2019) 71, 77.

<sup>40</sup> See eg Tanja Dimitrov and others, ‘Autonomous Molecular Design: Then and Now’ (2019) 11 ACS Applied Materials & Interfaces 24825, 24832; Kat Jerich, ‘AI Isn’t Magic’ (*Healthcare IT News*, 9 September 2020) <<https://www.healthcareitnews.com/news/emea/ai-isnt-magic>> accessed 21 May 2021; Gil Press, ‘Deep Learning Pioneer Yoshua Bengio Says AI Is Not Magic and Intel AI Experts Explain Why and How’ (*Forbes*, 20 September 2019) <<https://www.forbes.com/sites/gilpress/2019/09/20/deep-learning-pioneer-yoshua-bengio-says-ai-is-not-magic-and-intel-ai-experts-explain-why-and-how/>> accessed 21 May 2021; Kalev Leetaru, ‘Today’s Deep Learning “AI” Is Machine Learning Not Magic’ (*Forbes*, 14 November 2018) <<https://www.forbes.com/sites/kalevleetaru/2018/11/14/todays-deep-learning-ai-is-machine-learning-not-magic/>> accessed 21 May 2021; Will Knight, ‘AI Is Not “Magic Dust” for Your Company, Says Google’s Cloud AI Boss’ (*MIT Technology Review*, 8 November 2018) <<https://www.technologyreview.com/2018/11/08/139151/ai-is-not-magic-dust-for-your-company-says-googles-cloud-ai-boss/>> accessed 21 May 2021; Alex Reichenbach, ‘Artificial Intelligence is Applied Calculus, Not Magic’ (*Medium*, 6 February 2018) <<https://medium.com/1517/artificial-intelligence-in-the-modern-era-bdf599737247>> accessed 21 May 2021; Andrew Farah, ‘AI Is Not Magic; It’s Manual Labor (& Math) – How We Built An Accurate People Counter’ (*Medium*, 6 June 2018) <<https://medium.com/density-inc/ai-is-not-magic-its-manual-labor-math-how-we-built-an-accurate-people-counter-e00408ea30de>> accessed 21 May 2021.

<sup>41</sup> There is no universally accepted classification of AI methods and techniques. Given that GP can be applied as an ML technique to the same problems as ANNs, this article refers to GP and EAs as ML techniques.

At the same time, it should be noted that, besides ML, GP can be applied in other fields such as automatic programming.

<sup>42</sup> As the name suggests, these methods follow the principles and mechanisms of biological evolution. Accordingly, the search for a better solution to a problem is often based on the principle that ‘the fittest survives’ – the solutions with better quality have higher chances to be used further, whereby ‘quality’ is assessed in terms of a solution’s proximity to the optimum. EAs comprise three main categories: evolution strategies, evolutionary programming and genetic algorithms. While all three use essentially the same conceptual framework, they differ in the representations of individuals and schemes for implementing fitness evaluation, selection and search operators. In *evolution strategies*, an individual can be represented as a vector of real numbers; in *evolutionary programming* each individual is taken as a pair of real-valued vectors; *genetic algorithms* use ‘binary representation [whereby] each individual will be represented by a number of binary bits, 0 or 1’. Xin Yao, ‘Evolutionary Computation. A Gentle Introduction’ in Ruhul Sarker, Masoud Mohammadian and Xin Yao (eds), *Evolutionary Optimization. International Series in Operations Research & Management Science*, vol 48 (Springer 2003) 27, 29–34.

<sup>43</sup> Constraint satisfaction refers to a subtype of search problem where the goal is to allocate values to the variables so that the pre-specified constraints are satisfied or proven impossible to satisfy, given that the search space is finite. A constraint-optimisation problem combines constraint satisfaction with an objective function to be optimised. Dana S Nau, ‘Artificial Intelligence and Automation’ in Shimon Y Nof (ed), *Springer Handbook of Automation* (Springer 2009) 249, 253.

genie's capability to solve technical problems – a human only needs to make a wish.<sup>44</sup> According to the High-Level Expert Group on Artificial Intelligence (HLEG-AI), AI systems are 'goal-directed, meaning that they receive the *specification of a goal* to achieve from a human being and *use some techniques* to achieve such goal'.<sup>45</sup>

## a) Patent law uncertainties

### (i) Inventorship

If a human could merely state a problem that AI systems could solve 'on their own', there would be a strong argument that not only the allocation of inventor's rights but the need for the patent system in the first place would need to be fundamentally reassessed. Some legal scholars argue that the patent system should be eliminated 'altogether'<sup>46</sup> in the age of artificial ingenuity. Others suggest applying rules on employee inventions to 'AI-generated' inventions, provided that an 'electronic personality' can be attributed to AI.<sup>47</sup> Before such policy proposals can be considered, it is necessary to clarify whether 'making a wish' might be sufficient to solve a problem by AI.

### (ii) Inventive step/non-obviousness

The perception that the most significant input by a human subsists in merely supplying a problem, which is then solved by AI, might prompt an analogy with the so-called 'problem inventions'. Under the EPO approach, the appreciation of an unrecognised technical problem can confer an inventive step, even if the claimed solution is 'retrospectively trivial and in itself obvious',<sup>48</sup> provided that the problem could not have been 'readily posed' by any skilled person.<sup>49</sup> To consider the aptness of the analogy, let us first clarify whether the human role in ML, in fact, subsists solely in posing a problem.

## b) The problem-solving capacity of ML systems from a technical perspective

### (i) An algorithm as an explicit and specific problem-solving route

The assumption that a mere statement of a goal suffices for a computer to solve a problem is largely inaccurate – otherwise, we would have had a ready-to-deploy vaccine against SARS-Cov-2 as soon as we realised it was

urgently needed.<sup>50</sup> While telling a computer 'what' to do without 'how' has long been an aspiration of research on automated programming, it is considered 'unrealistic, at least in the foreseeable future'.<sup>51</sup>

The 'how' in ML – and any computer-implemented process in general – is represented by computational operators and configurations that determine the sequence of computational states, through which the given inputs are transformed into the desired output. In computer science, an algorithm is defined as 'a finite set of rules that gives a sequence of operations for solving a specific type of problem'. It is distinguished by the following features: *definiteness* (each step is precisely, rigorously, and unambiguously specified); *input* (a set of data is provided before the procedure starts); *output* (the procedure generates 'a nonempty set of results'); *finiteness* (the procedure consists of a finite number of steps); and *effectiveness* ('operations must all be sufficiently basic that they can in principle be done exactly and in a finite length of time by someone using pencil and paper').<sup>52</sup> This definition holds for ML algorithms as they specify a sequence of steps on how a model for a particular problem should be developed. These steps are precise, rigorous and unambiguous and are clearly defined by logical reasoning, conditional structures and loops. While the process might seem to be running 'by itself' to an outsider, the designer of an ML system knows what a computer is supposed to do at each step. Neither conventional nor quantum computers can deviate from a given algorithm. Furthermore, it should be noted that ML algorithms can work 'off the shelf' only in limited cases and usually require special-purpose adjustment.<sup>53</sup>

### (ii) How are problems specified in ML?

The definition of ML as techniques allowing 'an AI system to learn how to solve problems that *cannot be precisely specified*'<sup>54</sup> is somewhat confusing. To translate a real-life problem into something that can be processed and solved by a computer, a problem essentially needs to be expressed in an abstract way – by using a formal notation (a mathematical model, functions, logic rules, etc.) – that a computer can decipher and implement. In optimisation

<sup>44</sup> Robert Plotkin, *The Genie in the Machine: How Computer-Automated Inventing Is Revolutionizing Law and Business* (Stanford Law Books 2009) 2-3.

<sup>45</sup> High-Level Expert Group on Artificial Intelligence (HLEG-AI), 'A Definition of AI: Main Capabilities and Scientific Disciplines' 7 (18 December 2018) <[https://ec.europa.eu/futurium/en/system/files/ged/ai\\_hleg\\_definition\\_of\\_ai\\_18\\_december\\_1.pdf](https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf)> accessed 21 May 2021 (emphasis added).

<sup>46</sup> Yanisky-Ravid and Liu (n 32) 2222. See generally Dornis (n 34).

<sup>47</sup> Jan Phillip Rektorschek and Tobias Baus, 'Protectability and Enforceability of AI-Generated Inventions' in Kai Jacob, Dierk Schindler and Roger Strathausen (eds), *Liquid Legal. Towards a Common Legal Platform* (Springer 2020) 459, 475 (suggesting that the extension of the concept of inventor to AI would require the establishment of an 'electronic personality' for AI worldwide; under that condition, 'an assignment to the "employer" could also be regulated [if AI systems] create inventions "in the exercise of their employment"').

<sup>48</sup> See Case T 2/83, decision of 15 March 1984, ECLI:EP:BA:1984:T000283.19840315, point 6 of the Reasons; case T 0764/12, decision of 18 February 2014, ECLI:EP:BA:2014:T076412.20140218, point 4.7.5 of the Reasons.

<sup>49</sup> Case T 0109/82, decision of 15 May 1984, ECLI:EP:BA:1984:T010982.19840515, point 5.1 of the Reasons.

<sup>50</sup> At the same time, it is acknowledged that AI techniques can be applied in the discovery and development of vaccines and treatments against SARS-Cov-2. See Arash Keshavarzi Arshadi and others, 'Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development' (*Front. Artif. Intell.*, 18 August 2020), doi: 10.3389/frai.2020.00065 (pointing out that 'leveraging computational models capable of filtering and generating reliable therapies can significantly speed up [...] discovery efforts'). According to the authors, employing ANNs and supervised learning methods 'has proven to be a vital game-changer when used for the purpose of virtual filtering and *de novo* design'. However, 'to achieve the desired performance in such intelligent methods, one requires the knowledge to recognize the most relevant biotargets in addition to a large-scale training dataset'. *ibid.*

<sup>51</sup> Michael O'Neill and Lee Spector, 'Automatic Programming: The Open Issue?' (2020) 21 *Genetic Programming and Evolvable Machines* 251. The authors hold that GP is unlikely to be 'sufficient, or at least on its own the most efficient method, to fully realise automatic programming, or at least the most appropriate or efficient method to achieve all the necessary functions that an automatic programming system requires'. *ibid.* 256.

<sup>52</sup> Donald E Knuth, *The Art of Computer Programming*, vol 1 (3rd edn, Addison-Wesley Professional 1997) 4-6. The condition of finiteness can be omitted by embedding a 'while loop', a command that will be continuously implemented as long as a condition remains true.

<sup>53</sup> As discussed in detail below at II.6.b.

<sup>54</sup> HLEG-AI (n 45) 4 (emphasis added).

problems<sup>55</sup> where one seeks to maximise or minimise certain values of interest, a problem is typically defined as an *objective function* that assigns numerical values to certain outcomes, such as a cost function in ANNs or a fitness function in GP/EA.

What is not ‘precisely specified’ in ML is the solution. To give a simple example, assume that we do not know which number multiplied by 4 produces 16 (a problem). To convey this problem to a computer, we express it as  $4 * x = 16$ . One approach would be to brute force by testing all possible values of  $x$  until value ‘4’ is reached. An ML-based approach would be to provide a relevant<sup>56</sup> annotated (labelled) dataset of mathematical operations and build a model that will derive  $x = 4$  from the observed cases. To guide the process of deriving  $x$ , one would need to specify a cost function providing a mathematical measure of how close or far-off a trained model is from the actual data. A third approach would be to use an optimisation algorithm to perform a ‘guided search’ by altering the solution based on the proximity to the number 16. No labelled data would be required for that, but one would need to model  $4 * x - 16$  as a fitness function to be minimised. Irrespective of which method is used, deriving the ‘non-precisely specified’ output in ML is akin to solving a mathematical equation when the exact solution is not known *ex ante*.<sup>57</sup>

### (iii) (Non-)explicit programming in ML

Arthur Samuel is credited with coining the term ‘machine learning’.<sup>58</sup> He is also falsely<sup>59</sup> said to have defined ML as a ‘field of study that gives computers the ability to learn *without being explicitly programmed*’.<sup>60</sup> To clarify, ML algorithms themselves are explicitly programmed. As noted above, the general definition of an algorithm as a procedure that is specified precisely, rigorously and unambiguously applies to ML algorithms as well.<sup>61</sup>

As discussed above, what is not ‘explicitly programmed’ in ML is the output. The optimisation of ML model parameters is governed by the inputs and configurations set up before an algorithm is implemented on a computer. In the case of ANNs, for instance, one needs to provide a relevant pre-processed dataset and configure the hyperparameters<sup>62</sup> that, in turn, determine the model parameters (weights and biases). Thus, one could say that

a human does not directly (‘explicitly’) program the output. This concerns the core distinction between two types of AI: rule-based<sup>63</sup> and learning-based systems. However, with both approaches, human input is decisive for the computational output.<sup>64</sup>

### (iv) ‘High-level’ statements in GP

GP is defined as ‘an automatic method for solving problems’<sup>65</sup> that starts with ‘a high-level statement of the requirements of a problem and attempts to produce a computer program that solves the problem’.<sup>66</sup> ‘High-level requirements’ refer to multiple, often conflicting optimisation constraints and considerations, including the requirements regarding robustness and stability of a solution.<sup>67</sup> Such requirements are expressed as a *fitness function* that provides ‘the search’s desired direction’.<sup>68</sup> Besides, mutation and recombination operators strongly impact the performance of EAs.<sup>69</sup> These configurations operationalise the iterative process of search and improvement until the optimal solution is found, or constraints are proved impossible to satisfy, or other termination criteria apply.<sup>70</sup> In other words, ‘high-level statements’ do not mean the absence of human guidance on how computation should be implemented.

In sum, the assumption that ML systems can solve problems without human guidance does not hold. If a problem is solved through computation, it means that the combination of inputs and instructions, irrespective of the level of granularity, is sufficient for finding the solution. This echoes Herbert Simon’s observation that solving a problem ‘simply means representing it so as to make the solution transparent’.<sup>71</sup>

## c) Patent law uncertainties revisited

### (i) Implications for the inventorship

In view of the above, it appears clear that ML systems do not replace human problem-solving skills, and the assumption that an ML system can solve a given problem by itself is unfounded.

As a general principle, a normative analysis should be undertaken if there is genuine uncertainty about the appropriateness of the existing legal order. Hence, the currently applicable definitional criteria for inventorship should serve as a starting point. As the study by

<sup>55</sup> Most ML methods are based on optimisation.

<sup>56</sup> In this case, the relevant instances could be those that produce number 4.

<sup>57</sup> On the predictability of ML output, see below at II.3.b.iv.

<sup>58</sup> Arthur L Samuel, ‘Some Studies in Machine Learning Using the Game of Checkers’ (1959) 44(1.2) IBM Journal of Research and Development 210.

<sup>59</sup> Maria Schulda, Ilya Sinayskiya and Francesco Petruccione, ‘An Introduction to Quantum Machine Learning’ (2015) 56 Contemporary Physics 172, 183, note 2 (observing that it is ‘interesting to note that although quoted in numerous introductions to machine learning, the original reference to the machine learning pioneer’s most famous statement is very difficult to find [as authors] either refer to other secondary publications, or falsely cite Samuel’s seminal paper from 1959’).

<sup>60</sup> See eg Issam El Naqa and Martin J Murphy, ‘What Is Machine Learning?’ in Issam El Naqa, Ruijiang Li and Martin J Murphy (eds), *Machine Learning in Radiation Oncology. Theory and Applications* (Springer 2015) 3, 6 (referencing the paper by Samuel (above (n 58)) (emphasis added)).

<sup>61</sup> Above (n 52) and the accompanying text.

<sup>62</sup> Model hyperparameters refer to configurational settings that determine the structure of an ANN (such as the number of hidden nodes and layers) and how the network is trained (such as the learning rate, the loss function, an activation function and execution framework). Hyperparameters are provided to an ANN before training; model parameters (weights and biases) are inferred from the training data.

<sup>63</sup> Rule-based AI systems generate predefined outcomes based on the given set of rules coded by a programmer.

<sup>64</sup> On (non-)determinism in ML, see below at II.3.

<sup>65</sup> John R Koza, Martin A Keane and Matthew J Streeter, ‘Automated Synthesis by Means of Genetic Programming of Complex Structures Incorporating Reuse, Parameterized Reuse, Hierarchies, and Development’ in Rick Riolo and Bill Worzel (eds), *Genetic Programming Theory and Practice. Genetic Programming* (Springer 2003) 221, 222.

<sup>66</sup> *Ibid.*

<sup>67</sup> John R Koza and others, *Genetic Programming IV. Routine Human-Competitive Machine Intelligence* (Springer 2003) 58.

<sup>68</sup> *Ibid.* 12.

<sup>69</sup> Holger H Hoos and Thomas Stützle, ‘Stochastic Local Search Algorithms: An Overview’ in Janusz Kacprzyk and Witold Pedrycz (eds), *Springer Handbook of Computational Intelligence* (Springer 2015) 1085, 1097 (further noting that ‘much research in EAs has been devoted to the design of effective mutation and recombination operators’).

<sup>70</sup> Termination criteria commonly used in ML include the maximum time, the number of created solutions and the predefined quality of a solution.

<sup>71</sup> Herbert A Simon, *The Sciences of the Artificial* (2nd edn, MIT Press 1981) 153.

Shemtov<sup>72</sup> on the concept of inventorship in the context of AI concludes, '[c]reative or intelligent conception of the invention, or contribution thereto, is a feature that runs either explicitly or implicitly throughout the definition of inventorship in all of the [examined] jurisdictions'.<sup>73</sup> To meet this requirement, the 'engagement in the conception phase [should go] beyond the provision of abstract ideas on the one hand, and mere execution of those provided by others on the other hand'<sup>74</sup> and should be made 'on an intelligent and creative level rather than financial, material or mere administrative level'.<sup>75</sup> The subsequent analysis will rely on this standard as a general guiding principle,<sup>76</sup> while it is acknowledged that jurisdictions might differ in fleshing out the concept of the inventor.

Conceptualisation is about abstract thinking, a capability even the 'most sophisticated AIs in the world today have trouble with'.<sup>77</sup> To solve problems through computational modelling, including in the fields of technology and engineering, one needs to grasp a problem and gain 'an extensive understanding of mathematical structures in order to match them to the problem at hand'.<sup>78</sup> One also needs to define accurate assumptions, boundaries and parameters that reflect and characterise inherent properties of a system or process being modelled, etc. Such cognitive activities can refer to the *conceptualisation* phase of computational problem-solving that precedes programming and algorithm implementation.<sup>79</sup> As noted above, ML algorithms usually require special-purpose application. Hence, in situations where ML is used in the inventive process, it appears appropriate to refer to the *decision-making* in applying ML techniques to a technical problem at hand as a proxy for the 'intelligent engagement' in invention conception.

Accordingly, as long as the use of problem-solving tools and techniques is not prejudicial to allocating the inventor's rights to a human, the person making *necessary decisions* as to how to solve a technical problem by applying ML can and should be deemed an inventor, or a co-inventor (where such process involves joint human

effort), provided that the co-inventorship criteria are fulfilled.<sup>80</sup> In contrast, the one who develops the basic algorithm of a general-purpose nature – such as backpropagation<sup>81</sup> – but is not involved in its application to a specific task should not be regarded as an inventor. Furthermore, even though the inventorship standard applied to humans cannot be directly transposed to human-computer interaction, the underlying principle that the mere *implementation of instructions* should not suffice for the inventor entitlement can counteract the claim that the computer should be recognised as a standalone inventor.<sup>82</sup>

As clarified in the preceding section, human input in ML is not confined to giving a problem to a computer. However, a normative doubt might arise about whether such input should be deemed sufficient relative to the computer's contribution, often depicted as overshadowing the human role in terms of complexity and significance. Section II.2 will address the assumption underlying such doubt in detail.

## (ii) Implications for the inventive step requirement

### *The inaptness of the analogy with 'problem inventions'*

Given that it is the *appreciation* of a problem – not how it is solved – that can confer an inventive step in the case of 'problem inventions',<sup>83</sup> the use of ML techniques in developing an invention (problem-solving) would not be a relevant factor for assessing obviousness. While one could argue that identifying a problem and solving it should not be appraised equally when assessing inventive step, this is not an AI-specific issue and, therefore, it is not analysed here.

### *Knowledge and skills in ML should be factored into the definition of a skilled person where relevant*

To fulfil the inventive step requirement, an invention must represent an achievement that lies beyond the reach of a hypothetical practitioner with 'average' skills. The examination is fact-specific and methodologies differ among national patent offices. In the EPO's practice, the crux of the assessment is whether a skilled person, having regard to state of the art, *would* (i.e. not just *could*) have arrived at 'something falling within the terms of the claims, and thus achieving what the invention achieves'.<sup>84</sup>

<sup>72</sup> Noam Shemtov, *A Study on Inventorship in Inventions Involving AI Activity* (EPO 2019).

<sup>73</sup> *ibid* 19. In particular, the study examined the applicable provisions in the US, China, Japan, Republic of Korea, the UK, Germany, France and Switzerland. In addition, the Russian Civil Code (art 1347) defines the 'author of an invention' as 'a citizen whose creativity has led to the creation of such a result'.

<sup>74</sup> Shemtov (n 72) 19.

<sup>75</sup> *ibid* (emphasis added).

<sup>76</sup> The argument that the inventorship requirement for 'intelligent contribution' might not always be fulfilled in practice is not discussed here, given that this is not an AI-specific issue. See Burk (n 17) 307 (pointing out that 'existing patent law already has the formalist quandary well in hand [as] patent doctrine has never hesitated to bestow inventorship, and subsequently patents, on serendipitous discoveries').

<sup>77</sup> David Newman and Oliver Blanchard, *Human/Machine. The Future of Our Partnership with Machines* (Kogan Page 2019) 201. See also Lene Pettersen, 'Why Artificial Intelligence Will Not Outsmart Complex Knowledge Work' (2019) 33(6) *Work, Employment and Society* 1058 (arguing that '[a] great deal of knowledge work concerns highly complex problem solving and must be understood in contextual, social and relational terms'; that such 'aspects have no generic nor universal rules and solutions and, thus, cannot be easily replaced by artificial intelligence or programmed into computer systems, nor are they constructed based on models of the rational brain').

<sup>78</sup> Tony Hürlimann, *Mathematical Modeling and Optimization. An Essay for the Design of Computer-Based Modeling Tools* (Springer 1999) 30.

<sup>79</sup> For an overview of computational problem-solving stages, see Kim (n 20) 449-450.

<sup>80</sup> In the absence of a statutory definition, the co-inventorship criteria might be defined in case-law. For instance, the German Patent Act (s 6) does not specify the eligibility criteria for the joint inventorship but states that, where two or more persons have jointly made an invention, the right to the patent shall belong to them jointly. According to the case-law, the co-inventor status shall not be recognised only where a contribution had almost no impact on the overall success and was therefore insignificant for the claimed solution, or where a contribution resulted from instructions of an inventor or third parties. At the same time, for the co-inventorship to be acknowledged, the contribution in and of itself does not need to fulfil all criteria of patentability. Bundesgerichtshof, decision of 18 June 2013 – X ZR 103/11, para 8.

<sup>81</sup> On 'general-purpose' components of ML, see below at II.5.b.

<sup>82</sup> The assumption that ML systems 'make decisions' and do not merely implement instructions is examined in detail below at II.3.b.

<sup>83</sup> Provided that a problem is not obvious to a skilled person; above (nn 48-49).

<sup>84</sup> Guidelines for Examination in the European Patent Office (March 2021) [hereinafter EPO Guidelines for Examination], pt G-VII, ss 4, 5.3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_4.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_4.htm)> accessed 21 May 2021.

As clarified, ML systems do not solve problems ‘on their own’ but can be applied as computational problem-solving techniques. Accordingly, skills and knowledge in ML should be factored into the definition of a skilled person<sup>85</sup> when assessing the obviousness of inventions that could or would<sup>86</sup> have been developed by applying ML. The gradual integration of ML techniques into a skilled person’s arsenal of tools is inevitable, given the broad applicability of ML techniques on the one hand and the expectation that the skilled person is ‘involved in constant development in the relevant technical field’<sup>87</sup> on the other hand. Given that ML (as computational problem-solving techniques based on optimisation and modelling) is broadly applicable across technological and engineering fields,<sup>88</sup> technical experts in such fields would often cooperate with ML researchers and data scientists when developing or improving technical solutions. Such situations could fall within the ambit of the current EPO methodology, as it allows the definition of a skilled person to be extended to an *interdisciplinary* team of skilled practitioners.<sup>89</sup>

Accordingly, in instances where a technical problem underlying the claimed invention could have been solved through ML, inventive step should be assessed through the lens of an interdisciplinary team availed *inter alia* of knowledge and skills in ML, in addition to knowledge and skills in the field of technology to which the problem underlying an invention pertains.<sup>90</sup> For that, the assessment would need to establish that a skilled person from the field of a technical problem<sup>91</sup> would be *prompted* ‘to seek a solution’ and ‘look for suggestions’<sup>92</sup> in the field of ML. The combination of knowledge – even from remote technical fields<sup>93</sup> – in the definition of a skilled person is not excluded, which can accommodate the diversity of

ML applications. Notably, not only can a skilled person search for ‘clues’ in a different technical field, but they can also transfer a technology, provided that such transfer involves ‘routine experimental work’.<sup>94</sup> Conversely, an inventive step can be acknowledged where technology transfer requires efforts beyond routine work, such as scientific research.<sup>95</sup> To summarise, the EPO assessment methodology, in principle, allows to integrate skills in ML into the definition of an interdisciplinary team of skilled persons. Accordingly, the patent examiner would need to establish objectively, first, whether the skilled person would be motivated to look for suggestions in the field of ML and, second, whether the ‘routine’ application of ML would allow the skilled person to arrive at ‘something falling within’ the claims of the invention at issue.

## 2. ML systems learn ‘complex rules’ based on ‘simple rules’

In his book *Genie in the Machine*, Robert Plotkin observes that some AI techniques can ‘discover *complex* rules and patterns [...] given only an abstract problem definition and *simple rules* for generating and evaluating possible solutions to the problem’.<sup>96</sup> For instance, providing an airflow equation to an AI system would suffice to produce a car frame.<sup>97</sup> Silver et al. report that they used ‘the simplest possible search algorithm’ to develop AlphaGo Zero, a program which eventually achieved ‘superhuman performance, winning 100–0 against the previously published, champion-defeating AlphaGo’.<sup>98</sup> They conclude that ‘it is possible to train [an RL system] to *superhuman* level, without human examples or guidance, given no knowledge of the domain beyond *basic* rules’.<sup>99</sup>

### a) Patent law uncertainties

#### (i) Inventorship

The accounts presented above can create an impression that remarkable achievements can be accomplished with surprisingly negligible intellectual effort. Besides, the term ‘learning’ can create an impression that computers gradually become independent.<sup>100</sup> Even though it was already clarified under Assumption 1 that ML researchers do not merely state a problem but essentially configure a problem-solving route to the solution, one could still doubt whether human input should be viewed as sufficient to merit the inventor entitlement.<sup>101</sup>

<sup>85</sup> The EPO Guidelines for Examination define a ‘person skilled in the art’ as ‘a skilled practitioner in the relevant field of technology who is possessed of average knowledge and ability and is aware of what was common general knowledge in the art at the relevant date’. EPO Guidelines for Examination, pt G-VII, s 3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_3.htm)> accessed 29 July 2021 (with further references).

<sup>86</sup> The ‘could-would’ language alludes to the EPO patent examination methodology of assessing an inventive step, the so-called ‘could-would’ approach. See EPO Guidelines for Examination, pt G-VII, s 5.3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_5\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_5_3.htm)> accessed 29 July 2021.

<sup>87</sup> *ibid.*

<sup>88</sup> See Kim (n 20) 446 (with further references).

<sup>89</sup> EPO Guidelines for Examination, pt G-VII, s 3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_3.htm)> accessed 29 July 2021.

<sup>90</sup> Jurisdictions differ in the extent to which different technical fields can be considered when a skilled person’s relevant knowledge and skills are defined for the purpose of inventive step/obviousness assessment. Regarding the US approach, see eg Brenda M Simon, ‘The Implications of Technological Advancement for Obviousness’ (2013) 19 Mich. Telecomm. & Tech. L. Rev. 331, 356 ff (discussing the limits on combining analogous prior art in situations where ‘cognitive technologies’ might be applied in the inventive process).

<sup>91</sup> Under the EPO’s problem-solution approach, ‘the starting point for defining the appropriate skilled person is the technical problem to be solved on the basis of what the closest prior art discloses’. Case T 422/93, decision of 21 September 1995, ECLI:EP:BA:1995:T042293.19950921, headnote, para 1.

<sup>92</sup> EPO Guidelines for Examination, pt G-VII, s 3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_3.htm)> accessed 29 July 2021.

<sup>93</sup> *ibid.* (pointing out that a skilled person might be ‘expected to look for suggestions in neighbouring and general technical fields [...] or even in remote technical fields, if prompted to do so’ (with further references)).

<sup>94</sup> Frédéric Bostedt and others (eds), *Case Law of the Boards of Appeal of the European Patent Office* (9th edn, EPO 2019) 205–206 (albeit referring to the transfer of technology from a neighbouring field in the context of defining a person skilled in the art in the field of biotechnology).

<sup>95</sup> Case T 0441/93, decision of 27 March 1996, ECLI:EP:BA:1996:T044193.19960327, point 47 of the Reasons.

<sup>96</sup> Plotkin (n 44) 80 (emphasis added).

<sup>97</sup> *ibid.* 105.

<sup>98</sup> David Silver and others, ‘Mastering the Game of Go without Human Knowledge’ (2017) 550 Nature 354, <https://doi.org/10.1038/nature24270>.

<sup>99</sup> *ibid.* 358 (emphasis added).

<sup>100</sup> This argument is addressed below at II.3.b.

<sup>101</sup> See eg Tim W Dornis, ‘Of “Authorless Works” and “Inventions without Inventor” – The Muddy Waters of “AI Autonomy” in Intellectual Property Doctrine’ 14 (2021) <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3776236](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3776236)> accessed 21 May 2021 (observing that ‘it is often overlooked that, in many scenarios of AI creativity and inventiveness, there no longer exists a sufficient nexus between a human actor’s input to and the result of a creative or inventive AI process’).

## (ii) Inventive step/non-obviousness

The alleged simplicity of human input in ML can bear on the assessment of obviousness of an invention. If only trivial knowledge and skills on the human's part could suffice to solve a technical problem through ML,<sup>102</sup> one would question whether the purpose of the inquiry into inventive step – to discern the achievements beyond those requiring average knowledge and skills – is still relevant, or whether *any* invention actually or potentially<sup>103</sup> resulting from the application of ML should *prima facie* be deemed obvious.<sup>104</sup>

Before these legal implications can be examined, let us clarify what exactly 'learned rules' mean in the ML context and how human and computer's contributions can be compared in terms of simplicity or complexity, if at all.

## b) The simplicity of human input vs complexity of ML outcome from a technical perspective

### (i) The term 'learning'

Whether computers can 'learn' depends on how one defines 'learning'. As a cognitive phenomenon, learning comprises 'inductive knowledge acquisition, behavior generation, and intelligence aggregation'.<sup>105</sup> The extent to which human and computer's learning processes can be comparable can hardly be ascertained, given that human learning alone cannot be considered to be fully understood. As for the analogy between ML and human learning, current ML approaches based on 'statistical induction over massive training sets' are viewed to 'radically differ' from human-like learning.<sup>106</sup> In this regard, the pervasive use of the anthropomorphic language, 'exacerbated by the ML literature itself',<sup>107</sup> has been criticised for 'inadvertently promot[ing] misleading interpretations of and beliefs about what AI is and what its capacities are' among the general public.<sup>108</sup>

In computer science, a 'learning automaton' refers to a system that can be trained 'to associate with each context the particular action which maximises payoff'.<sup>109</sup> If by learning, one means that a system can improve its performance over time based on an in-built feedback mechanism, one can say that computers can learn, but they do

so within the pre-specified rules and boundaries. However, a more appropriate terminology in the case of ANNs would be 'inferring' a numeric function or 'fitting' data into a model.<sup>110</sup>

### (ii) 'Learned rules'

There is no specific technical definition of a 'rule' in ML. The views cited above<sup>111</sup> assume a distinction between the 'rules' determining ML process and the 'learned rules' as ML output. The former can refer to instructions on how to build a model from the input data; the latter can be understood as an immediate output of ML (a numeric model derived from the input data based on the 'given rules'). Such model can be deemed as 'learnt rules' if it determines how a prediction is generated in supervised ML. In RL, what is 'learnt' (calculated) is a 'policy'<sup>112</sup> which determines which action to take.

Important is that the calculation of the numeric output through ML is guided<sup>113</sup> by the pre-programmed instructions and setup configurations. In this regard, the portrayals of ML techniques as 'self-learning',<sup>114</sup> or 'self-teaching' networks<sup>115</sup> is not accurate. ML systems are products of software engineering. For instance, even though the developers of AlphaGo did not impart to the system ready-made 'rules' as to when to make which move,<sup>116</sup> they provided inputs, including data and computational operators (mathematical functions),<sup>117</sup> that determined how the moves maximising the long-term reward were computed.

Furthermore, if the 'learned rules' refer to the correlations inferred from the training data, it would not be accurate to refer to such correlations as 'rules', as long as the causes behind the revealed regularities are not exposed or as long as it remains unclear why the inferred

<sup>102</sup> Plotkin (n 44) 106.

<sup>103</sup> See Blok (n 20) 71 (noting that where AI 'becomes a standard tool for routine work, the abilities of the average skilled person are assumed to have improved accordingly [and hence], patents will not be granted if the invention is obvious to the average skilled person equipped with the artificial intelligence application, even if the inventor did not use artificial intelligence').

<sup>104</sup> See generally Ryan Abbott, 'Everything Is Obvious' (2018) 66 U.C.L.A. L. Rev. 2.

<sup>105</sup> Yingxu Wang and others, 'Cognitive Intelligence: Deep Learning, Thinking, and Reasoning by Brain-Inspired Systems' (2016) 10(4) International Journal of Cognitive Informatics and Natural Intelligence 1, 2 (with further references).

<sup>106</sup> Pat Langley, 'The Computational Gauntlet of Human-Like Learning' 1 <<http://www.isle.org/~langley/papers/gauntlet.aaai22.pdf>> accessed 22 September 2021. The paper describes those characteristics of human learning that 'provide strong constraints on computer artifacts designed to exhibit human-like learning' and concludes that 'human learning displays qualities that are seldom addressed by current work on data-intensive induction'. *ibid* 4.

<sup>107</sup> Epstein and others (n 27) 3.

<sup>108</sup> Salles, Evers and Farisco (n 29) 93.

<sup>109</sup> PC Bressloff and J Stark, 'Neural Networks, Learning Automata and Iterated Function Systems' in Anthony J Crilly, Rae A Earnshaw and Huw Jones (eds), *Fractals and Chaos* (Springer 1991) 145, 146.

<sup>110</sup> For instance, in the course of supervised ML an algorithm infers a model that correlates inputs (independent variables) with outputs (dependent variables). In unsupervised ML an algorithm searches for the best fitting of the input data (akin to pattern finding). Both types of learning are guided by objective functions. Bhanu Yerra, 'Objective Functions Used in Machine Learning' (*Medium*, 2 March 2019) <<https://medium.com/@bhanuyerra/objective-functions-used-in-machine-learning-9653a75363b5>> accessed 21 May 2021.

<sup>111</sup> Above (nn 96-99) and the accompanying text.

<sup>112</sup> A 'policy' in RL refers to a 'mapping from perceived states of the environment to actions to be taken when in those states'. Richard S Sutton and Andrew G Barto, *Reinforcement Learning: An Introduction* (2nd edn, MIT Press 2018) 7 (further noting that, in some cases, the policy can be represented by 'a simple function or lookup table, whereas in others it may involve extensive computation such as a search process').

<sup>113</sup> On (non-)determinism in ML, see below at II.3.

<sup>114</sup> See eg Massimo Craglia (ed), *Artificial Intelligence. A European Perspective* (Publications Office of the European Union 2018) 64.

<sup>115</sup> Curtis EA Karnow, 'Foreword' in Woodrow Barfield and Ugo Pagallo (eds), *Research Handbook of Artificial Intelligence and Law* (Edward Elgar 2018) xviii, xx.

<sup>116</sup> Peter Remmers, 'Would Moral Machines Close the Responsibility Gap?' in Birgit Beck and Michael Kühler (eds), *Technology, Anthropology, and Dimensions of Responsibility*, vol 1 (JB Metzler 2020) 133, 139 (with further references).

<sup>117</sup> The core elements of RL are: reward signals (represented by stochastic functions); a value function (specifying which actions are beneficial in the long run); a policy (which can also be represented by a function); and a model (predicting next state and next reward, given the current state and an action). Based on these operators, a set of environment and agent states, a set of actions taken by the agent, probability of transition from one state to another under certain action and the immediate reward following the transition from one state to another are computed. Sutton and Barto (n 112) 7-9.



correlations may or may not work when applied to new data.<sup>118</sup> Consider the recent achievement of AlphaFold credited with ‘solving’ the protein folding problem in molecular biology. The problem of protein folding subsists in understanding the rules governing the relationship between the structure of a protein and its amino-acid sequence. Despite what the media headlines touted,<sup>119</sup> AlphaFold has not ‘cracked’ the protein folding problem – it revealed ‘nothing about the mechanism of folding, but just predicts the structure using standard machine learning [...] being trained on the 170,000 or so known structures in the Protein Data Base’.<sup>120</sup> Meanwhile, the ‘why’ behind protein folding remains ‘a black box’.<sup>121</sup>

### (iii) The alleged simplicity of given rules vs complexity of ‘learned’ rules

Robert Plotkin argues that computers do not always ‘operate according to the adage “garbage in, garbage out”’,<sup>122</sup> and AI systems can ‘follow [given] *simple* rules to discover *complex* rules’.<sup>123</sup> In his book, he refers to GP as an ‘artificial invention technology’.<sup>124</sup>

Solving a problem through GP consists of two phases, namely preparatory steps, and executional steps (the implementation of an algorithm).<sup>125</sup> As defined by Koza et al., preparatory steps are ‘*problem-specific* and *domain-specific* steps that are performed by the human user prior to launching a run of the problem-solving method’.<sup>126</sup> In contrast, executional steps are ‘automatically executed during a run of the problem-solving method’.<sup>127</sup>

Furthermore, Koza et al. introduce the concept of an ‘artificial-to-intelligence ratio’ (an ‘AI ratio’), where ‘I’ refers to the knowledge provided by a human expert in a particular field, and ‘A’ indicates the value added by the genetic programming.<sup>128</sup> Some problems can require ‘a small amount’ of ‘I’.<sup>129</sup> While others can involve ‘non-trivial amount’ of ‘I’.<sup>130</sup> Where ‘A’ is high and ‘I’ is low, GP is considered to have delivered a ‘human-competitive’

result.<sup>131</sup> As problems differ in complexity, they can require varying levels of knowledge and skills in GP.

However, issue can be taken regarding both the concept of ‘artificial-to-intelligence ratio’ and the notion of ‘human-competitive’ results. First, to measure an ‘artificial-to-intelligence ratio’, one needs to draw a line between the output of running a GP algorithm and ‘intelligence that is supplied by the human applying the method to a particular problem’.<sup>132</sup> The preparatory steps in GP include defining ‘the set of terminals [...], the set of primitive functions for each branch of the to-be-evolved computer program, the fitness measure (for explicitly or implicitly measuring the fitness of candidate individuals in the population), certain parameters for controlling the run, a termination criterion and method for designating the result of the run’.<sup>133</sup> As iterative search techniques,<sup>134</sup> genetic algorithms evolve a candidate population according to the given search operators until the fitness function is satisfied or other termination criteria apply.<sup>135</sup> In this view, ‘the problem-specific preparatory steps’ done by a human are *determinative* of the outcome of implementing an algorithm on a computer. If so, it appears pointless to split up the two phases to define which contributes more to the solution.

Second, even if one wanted to compare a human versus a computer’s contributions in terms of their ‘complexity’, what would be a suitable objective measure? For instance, should the complexity of a mathematical function, which can be both the input and the output of ML, be measured by its length, the amount of energy consumed by the human brain as opposed to computational power, or how many operations are necessary to derive it? One could say that it is equally simple for a computer to execute instructions, irrespective of whether it calculates  $y=2x$  or  $y=10x^3+4x-2x^2+777x^{21}$ . For a human, devising computational instructions can vary in complexity depending on the problem to be solved and the level of expertise.

In GP, the design of a well-suited problem-specific fitness function and selection of the problem representation and genetic operators are *not trivial* tasks. In the case of ANNs, even if a standard algorithm can be applied ‘off the shelf’, hyperparameters<sup>136</sup> need to be attuned by a researcher, and this might or might not be straightforward depending on the problem at hand.<sup>137</sup> Setting the right hyperparameters for an ANN and defining a value function in RL are decisive for the computational outcome. Furthermore, the selection of input data, which predetermines the ‘learned’ correlations, should not be downplayed. The selection and preparation of training datasets require expertise and a keen understanding of the problem at hand. Both AphaGo and AlphaFold involved a team of highly skilled engineers and researchers; in both cases, a computer alone could not have achieved results without complex design decisions on the human part.<sup>138</sup>

**118** As discussed later, the absence of ‘ready-made’ knowledge in an ANN causes its perception as a ‘black box’. See IL4.b.

**119** Ian Sample, ‘DeepMind AI Cracks 50-Year-Old Problem of Protein Folding’ *The Guardian* (London, 30 November 2020) <<https://www.theguardian.com/technology/2020/nov/30/deepmind-ai-cracks-50-year-old-problem-of-biology-research>> accessed 21 May 2021; Ewen Callaway, ‘It Will Change Everything. AI Makes Gigantic Leap in Solving Protein Structures’ *Nature*, 30 November 2020) <<https://www.nature.com/articles/d41586-020-03348-4>> accessed 21 May 2021.

**120** Philip Ball, ‘Behind the Screens of AlphaFold’ (*Chemistry World*, 9 December 2020) <<https://www.chemistryworld.com/opinion/behind-the-screens-of-alphafold/4012867.article>> accessed 21 May 2021. See also Brigitte Nerlich, ‘Protein Folding and Science Communication: Between Hype and Humility’ (*Making Science Public*, 4 December 2020) <<https://blogs.nottingham.ac.uk/makingsciencepublic/2020/12/04/protein-folding-and-science-communication-between-hype-and-humility/>> accessed 21 May 2021.

**121** Ball (n 120).

**122** Plotkin (n 44) 80.

**123** *ibid* (emphasis added).

**124** *ibid*.

**125** Koza and others (n 67) 10.

**126** *ibid* (emphasis added).

**127** *ibid*.

**128** Koza and others define ‘AI ratio’ as a ratio ‘of that which is delivered by the automated operation of the *artificial* method to the amount of *intelligence* that is supplied by the human applying the method to a particular problem’. *ibid* 4 (emphasised in the original).

**129** *ibid* 419, 482.

**130** *ibid* 412.

**131** *ibid* 112.

**132** *ibid* 4.

**133** *ibid* 10.

**134** *ibid* 29.

**135** On global and local optima and the role of randomisation in ML, see below (nn 342-344) and the accompanying text.

**136** For an explanation, see above (n 62).

**137** Below (n 325) and the accompanying text.

**138** This is the authors’ reading of the report by Silver and others (n 98).

In sum, if instructions provided by a human are decisive for the computational outcome, they are probably not ‘garbage’<sup>139</sup> but rather illustrate the adage that a genius subsists in making complex ideas simple.<sup>140</sup>

#### (iv) The tendency to put a spotlight on an algorithm

As noted by Pat Langley, research papers on automated discovery ‘typically give the algorithm center stage, but they pay little attention to the developer’s efforts to modulate the algorithm’s behavior for given inputs’.<sup>141</sup> Such tendency can be observed both in mass media and research papers. As one commentator points out, while RL – the approach applied in training AlphaGo – might be viewed as ‘one of the closest things that looks anything like [Artificial General Intelligence], beautiful demos of learned agents hide all the blood, sweat, and tears that go into creating them’.<sup>142</sup> For research papers, the tendency to bring an ML system to the forefront is, in a way, natural and unavoidable. A paper should report the results and therefore presents an algorithm and results achieved with it as they are, not the process that has led to the system design. Usually, the researcher would not describe how much thinking and effort was devoted to creating an algorithm before reaching the version that works, or how many attempts failed before a model was tuned correctly.

### c) Patent law uncertainties revisited

#### (i) Inventorship

At the outset, it is worth noting that the ‘simplicity’ of arriving at an invention, *de lege lata*, has not been prejudicial to the genesis or allocation of the inventor’s rights.<sup>143</sup> It might simply be impossible to objectively assess in terms of complexity the mental process of conceiving an idea – ‘the formation in the mind of the inventor, of [the] idea of [an] invention’.<sup>144</sup> The argument *de lege ferenda* that inventor’s rights should not be allocated to a human applying ‘self-learning’ AI systems appears to be somewhat fairness-based, given the seeming triviality of the human input relative to the alleged complexity of a computer’s contribution. However, as clarified in the preceding section, ML output is the function of *human decision-making* in applying ML. Such decision-making can merit the inventor entitlement under both the criterion of an intelligent conception of an invention<sup>145</sup> and fairness considerations.

<sup>139</sup> Above (n 122) and the accompanying text.

<sup>140</sup> Albert Einstein is credited with a saying: ‘Genius is making complex ideas simple, not making simple ideas complex’. This principle echoes the maxim ‘truth is simple’, see Thomas W Körner, *The Pleasures of Counting* (Cambridge University Press 1996) 192.

<sup>141</sup> Pat Langley, ‘The Computational Support of Scientific Discovery’ in Georgios Paliouras, Vangelis Karkaletsis and Constantine D Spyropoulos (eds), *Machine Learning and Its Applications. ACAI 1999. Lecture Notes in Computer Science*, vol 2049 (Springer 2001) 231, 234.

<sup>142</sup> ‘Deep Reinforcement Learning Doesn’t Work Yet’ (*Sorta Insightful*, 14 February 2018) <<https://www.alexirpan.com/2018/02/14/rl-hard.html>> accessed 21 May 2021.

<sup>143</sup> See Burk (n 17) 307 ff. On serendipitous inventions, see also below (n 160) and the accompanying text.

<sup>144</sup> As defined in the US case-law: *Townsend v Smith* 36 F.2d 292, 295 (CCPA 1929).

<sup>145</sup> On this criterion, see above (nn 73-75) and the accompanying text.

#### (ii) Inventive step/non-obviousness

When confronted with the allegations regarding the simplicity of human input in solving tasks through ML, one needs to be aware of the *hindsight* problem. As a cognitive phenomenon, the hindsight bias refers to a ‘tendency to judge events to be more predictable, knowable, and certain in hindsight than in foresight’.<sup>146</sup> In patent law, an invention might, at first sight, appear obvious because, once ‘a new idea has been formulated, it can often be shown theoretically how it might be arrived at, starting from something known, by a series of apparently easy steps’.<sup>147</sup> The EPO methodology stipulates that patent examiners ‘must be wary of *ex post facto* analysis [and] seek to make a “real-life” assessment of [the] relevant factors’.<sup>148</sup> The principle of avoiding the foreknowledge of an invention is embedded within the ‘problem-solution approach’,<sup>149</sup> which instructs to start the analysis from the closest prior art reference and formulate the technical problem addressed by an invention so ‘as not to anticipate the solution’.<sup>150</sup>

Given that the ‘simplicity’ of human input in ML cannot be alleged in general, the obviousness of inventions which could or would have been developed with the aid of ML can only be assessed on a case-by-case basis. As will be discussed later, solving problems through ML can involve considerable decision-making.<sup>151</sup> Hence, the mere existence of ML methods applicable to technical problems, by itself, does not render obsolete the purpose of inventive step assessment to discern whether practitioners from the relevant technical fields with average knowledge and skills could or would have arrived at the claimed invention.

### 3. ML systems are autonomous, non-deterministic and unpredictable

Some legal scholars assume that AI is ‘capable of defining or modifying decision-making rules autonomously’<sup>152</sup> and can ‘determine for themselves the means of completing their goals’.<sup>153</sup> AI systems are often depicted as performing randomly. For example, an ANN makes ‘a random guess’ regarding the output;<sup>154</sup> the training of AlphaGo Zero started reportedly from ‘completely random behaviour and continued without human intervention’.<sup>155</sup> One legal scholar submits that AI refers to the ‘so-called non-deterministic algorithms: computer programs whose function and output

<sup>146</sup> Neal J Rouse and James M Olson, ‘Counterfactual Thinking’ in Lynn Nadel (ed), *Encyclopedia of Cognitive Science* (John Wiley & Sons 2006) doi: 10.1002/0470018860.s00504.

<sup>147</sup> EPO Guidelines for Examination, pt G-VII, s 8 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_vii\\_8.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_8.htm)> accessed 29 July 2021.

<sup>148</sup> *ibid.*

<sup>149</sup> The effectiveness of this methodology is not at issue here.

<sup>150</sup> Case T 0422/93, decision of 21 September 1995, ECLI:EP:BA:1995:T042293.19950921, headnote, para 2.

<sup>151</sup> Below at II.6.b.

<sup>152</sup> Ugo Pagallo and others, ‘The Rise of Robotics & AI: Technological Advances & Normative Dilemmas’ in Marcelo Corrales, Mark Fenwick and Nikolaus Forgo (eds), *Robotics, AI and the Future of Law* (Springer 2018) 1, 6.

<sup>153</sup> Ryan Abbott, ‘The Reasonable Computer: Disrupting the Paradigm of Tort Liability’ (2018) 86 *Geo. Wash. L. Rev.* 1.

<sup>154</sup> WIPO, ‘Background Document on Patents and Emerging Technologies’ SCP/30/5, para 21 (28 May 2019) <[https://www.wipo.int/edocs/mdocs/scp/en/scp\\_30/scp\\_30\\_5.pdf](https://www.wipo.int/edocs/mdocs/scp/en/scp_30/scp_30_5.pdf)> accessed 21 May 2021.

<sup>155</sup> Silver and others (n 98) 355.

are not exclusively determined by human creators'.<sup>156</sup> A corollary of such assumptions is the allegedly 'surprising' effect of the output of AI/ML systems.

While the perceived autonomy of AI systems might have crucial implications for various regulatory frameworks, such as personal data protection, consumer protection, transparency and product liability, it poses distinct issues for patent law. In either case, legal challenges are triggered by the assumption that AI-based systems have the capacity to 'decide' and that humans have limited or no control over such decision-making.

### a) Patent law uncertainties

#### (i) Inventorship in relation to a patentable outcome of the allegedly autonomous or non-deterministic computational process

The purported autonomy of ML systems can cast doubt on whether a human applying AI to a problem should be credited for finding a solution. For instance, Thaler describes his patented invention – 'Device for the autonomous generation of useful information' – as a device that 'allows for the *totally autonomous* generation of new concepts, designs, music, processes, discovery, and problem-solving using recent developments in the area of artificial neural network (ANN) technology'.<sup>157</sup> Relying on Thaler's depiction of DABUS, one judge has recently held that he is 'simply recognising the reality by according artificial intelligence the label of "inventor"'.<sup>158</sup>

It is worth noting here that the (perceived) randomness<sup>159</sup> of a process leading to a patentable invention, *de lege lata*, has not been a material factor for the genesis or allocation of the inventor's rights. Consider, for instance, serendipitous solutions<sup>160</sup> or 'a-ha' moments (which are, nevertheless, regarded to be preceded by the orientation, preparation and incubation stages of creative thought<sup>161</sup>). Uncertainty *de lege ferenda* can arise if one assumes that the task of finding a solution to a problem 'is handed over to the non-deterministic evolution of the algorithm',<sup>162</sup> and that 'the nexus between the human contribution to the search process

and the ultimate finding of a solution is severed during the AI's evolution and transformation'.<sup>163</sup> The view that AI systems can yield solutions that '*cannot be predicted*' by programmers, operators, or any other entities involved<sup>164</sup> led some legal scholars to argue that the allocation of patent rights to a human might no longer be justified.<sup>165</sup> Accounts crediting ML systems with making 'brilliant moves' not understood by humans<sup>166</sup> and delivering 'flummoxing',<sup>167</sup> results that '[a]lmost no human pro would've thought of it'<sup>168</sup> might create an impression of a computer's 'cognitive superiority' and a dissipating causal link between the human input and computational output.

In other words, the contention is that 'non-deterministic' algorithms imply the *lack of causality* between human input and computational outcome. If determinism is defined as a 'doctrine that all phenomena are causally determined by prior events',<sup>169</sup> even switching on a computer can be viewed as a 'prior event' that causes computation. However, if a problem is then solved by the computer 'on its own', one would doubt whether allocating the inventor's rights to a human is justified. Even though it was already clarified under Assumptions 1 and 2 that ML systems do not solve problems 'by themselves', a technical perspective on a computer's 'autonomy' and the relationship between determinism, randomisation and human guidance in ML would be welcome.

#### (ii) The assessment of inventive step in situations where the outcome of a randomised ML process is claimed within an invention

##### *Expectations of success of a skilled person*

Even if human decision-making in ML cannot be excluded, the mixed contribution by allegedly 'autonomous' or 'non-deterministic' AI systems and humans in the process of solving a technical problem raises the question of how such 'autonomy' or 'non-determinism' should be factored into the assessment of inventive step. Under the EPO's approach, for an invention to be deemed obvious, it is sufficient to establish that a skilled person would have arrived at the invention following the prior art teaching with a *reasonable expectation of success*.<sup>170</sup> Notably, where a technique involves randomness (e.g. mutagenesis) and where chance can play 'a key role in the achievement of success, as *no form of control* can be exerted over the

<sup>156</sup> Dornis (n 101) 4 (with no references).

<sup>157</sup> US patent 5,659,666 'Device for the autonomous generation of useful information' (emphasis added).

<sup>158</sup> *Thaler v Commissioner of Patents* (n 38) para 126. Some scholars assume that '[w]e are facing a new era of machines "acting" independently, with no human being behind the inventive act itself' (Yanisky-Ravid and Liu (n 32) 2216), and that AI systems 'take over the active part of cognitive processes [and] are not under the full control of a human user' (Jan-Hendrik Heinrichs, 'Artificial Intelligence in Extended Minds: Intrapersonal Diffusion of Responsibility and Legal Multiple Personality' in Birgit Beck and Michael Kühler (eds), *Technology, Anthropology, and Dimensions of Responsibility*, vol 1 (JB Metzler 2020) 159, 171).

<sup>159</sup> In the sense of happening by chance rather than according to a particular method. Cambridge Dictionary, 'Random' <<https://dictionary.cambridge.org/dictionary/english/random>> accessed 21 May 2021.

<sup>160</sup> For instance, decaffeinated coffee was accidentally discovered in 1904 and patented in 1906 (US patent 897,840 'Preparation of coffee', granted on 1 September 1908). Drug discovery was dominated until 1980 by 'random screening' that relied primarily on serendipity. See Giovanni Dosi and Mariana Mazzucato, 'Introduction' in Mariana Mazzucato and Giovanni Dosi (eds), *Knowledge Accumulation and Industry Evolution. The Case of Pharma-Biotech* (Cambridge University Press 2006) 1, 4. See also Burk (n 17) 308 (discussing in the context of AI the doctrine of simultaneous conception and reduction to practice and conception as *recognition* of an invention in the case of serendipitous or unforeseen inventions).

<sup>161</sup> Dennis Coon and John O Mitterer, *Introduction to Psychology: Gateways to Mind and Behavior* (Cengage Learning 2012) 292.

<sup>162</sup> Dornis (n 101) 22-23.

<sup>163</sup> *ibid* 23.

<sup>164</sup> Yanisky-Ravid and Liu (n 32) 2225 (emphasis added).

<sup>165</sup> *ibid* 2228 (noting that human ownership over inventions developed by AI systems is 'questionable'). See also Vertinsky and Rice (n 31) 575 (assuming that 'machines [...] originate novel solutions not imagined by their human operators, transforming the invention process in ways not easily accommodated within the current US patent system').

<sup>166</sup> Cade Metz, 'How Google's AI Viewed the Move No Human Could Understand' (*WIRED*, 14 March 2016) <<https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/>> accessed 21 May 2021.

<sup>167</sup> *ibid* (further noting that David Silver, researcher at DeepMind who led the development of AlphaGo, 'had no more insight' into the moment than anyone else witnessing the match when move 37 arrived).

<sup>168</sup> Matt McFarland, 'What AlphaGo's Sly Move Says About Machine Creativity' *The Washington Post* (Washington, 15 March 2016) <<https://www.washingtonpost.com/news/innovations/wp/2016/03/15/what-alpha-gos-sly-move-says-about-machine-creativity/>> accessed 21 May 2021.

<sup>169</sup> Albert Ellis, Mike Abrams and Lidia Abrams, *Personality Theories. Critical Perspectives* (SAGE Publications 2009) 245.

<sup>170</sup> Bostedt and others (n 94) 200 (with further references).

[course of] events’,<sup>171</sup> it is not considered ‘appropriate to attempt to evaluate the expectation of success’.<sup>172</sup> Accordingly, the assessment of inventive step in situations where the output of an ML method involving randomisation is claimed within an invention raises a two-fold question: (i) whether a human can exercise control over achieving such output; and (ii) if so, how the use of randomisation can affect a skilled person’s expectations of success in solving a problem through ML (provided that the relevant skilled person is versed in ML<sup>173</sup>).

#### *A surprising effect of an invention*

Furthermore, in some jurisdictions, a ‘surprising’ or ‘unexpected’ effect of an invention over the prior art can be considered as an indication of an inventive step (non-obviousness). The Patent Cooperation Treaty Examination Guidelines instruct that, in situations where the claimed invention holds a considerable technical value attributed to one or more of the claimed technical features and provides an advantage over the prior art ‘which is new and *surprising* [...]’, the examiner should be hesitant in raising a negative determination that such a claim lacks inventive step’.<sup>174</sup> Under the EPO approach, an unexpected effect over the prior art might confirm an inventive step.<sup>175</sup>

The developers of a NASA antenna report that the antenna ‘evolved’ by applying an evolutionary algorithm had ‘an unusual organic-looking structure, one that expert antenna designers would likely not produce’.<sup>176</sup> As pictured by Plotkin, to produce a car frame, one ‘can just provide an airflow equation’ to an AI system.<sup>177</sup> Later, when assessing that car frame under the non-obviousness standard, a patent examiner would determine, ‘quite correctly, that [the] car frame would not have been obvious to an automotive engineer of ordinary skill [because] the frame has a shape that is *surprising* to automotive engineers and violates the principles [they] learned in engineering school [which] makes it a classic case of a nonobvious design’.<sup>178</sup> The contention is, thus, that ‘surprising’ technical solutions resulting from the application of ML can be deemed inventive (non-obvious) under the existing legal standard, even where they, in fact, resulted

from the application of a routine technique and ‘merely by applying ordinary skill’.<sup>179</sup>

#### **(iii) The fulfilment of the sufficiency-of-disclosure requirement in situations where an ML method involving randomisation is claimed as an invention**

The requirement of sufficiency of disclosure means that an invention needs to be disclosed in a patent application in a manner reproducible by a skilled person without undue burden.<sup>180</sup> At issue is whether randomisation in ML poses a challenge to fulfilling this requirement in situations where an ML-based method is claimed as an invention.<sup>181</sup> An analogy with ‘non-deterministic’ biotechnological inventions<sup>182</sup> was drawn in this regard. In particular, it was hypothesised that ML-based techniques face a similar challenge regarding *reproducibility and plausibility* of an invention as in the case of biological materials which ‘variability is unavoidable’.<sup>183</sup> The European patent system, for instance, provides for a deposit system for such inventions.<sup>184</sup> Before the necessity of a similar instrument can be contemplated for inventions claiming ML techniques, it should be clarified how the use of randomisation in ML can affect the reproducibility of ML methods.

Before the above-outlined legal implications can be examined further, let us clarify in what sense ML output might be ‘unpredictable’, if at all.

#### **b) A technical perspective on autonomy, (un)predictability and (non-)determinism of ML systems**

##### **(i) The term ‘autonomy’ is used mistakenly in the sense of ‘automation’**

In everyday discourse and technical literature on AI, the terms ‘autonomy’ and ‘automation’ are sometimes used interchangeably.<sup>185</sup> However, there is a significant conceptual and technical difference.

<sup>171</sup> Case T 0737/96, decision of 9 March 2000, ECLI:EP:BA:2000:T073796.20000309, point 11 of the Reasons (emphasis added). The invention at issue related to astaxanthin-producing yeast strains, methods for their preparation, methods for their cultivation and methods for isolating the astaxanthin from the yeast cells; the technical problem to be solved was defined as ‘the provision of *Phaffia rhodozyma* mutants producing increased yields of the carotenoid pigment astaxanthin’. *ibid* point 5 of the Reasons.

<sup>172</sup> *ibid* point 11 of the Reasons.

<sup>173</sup> On factoring ML skills into the definition of a skilled person, see above (nn 85–89) and the accompanying text.

<sup>174</sup> ‘Patent Cooperation Treaty (PCT) International Search and Preliminary Examination Guidelines’ PCT/GL/ISPE/10 para 13.17 (30 April 2020) (emphasis added).

<sup>175</sup> Case T 181/82, decision of 28 February 1984, ECLI:EP:BA:1984:T018182.19840228.

<sup>176</sup> Jason D Lohn, Gregory S Hornby and Derek S Linden, ‘An Evolved Antenna for Deployment on Nasa’s Space Technology 5 Mission’ in Una-May O’Reilly and others (eds), *Genetic Programming Theory and Practice II*, vol 8 (Springer 2005) 305, 311. See Plotkin (n 44) 80 (assuming that an AI system can ‘discover complex rules and patterns that its human programmer never imagined’).

<sup>177</sup> Plotkin (n 44) 105.

<sup>178</sup> *ibid* 106 (emphasis added).

<sup>179</sup> *ibid* (further observing that ‘[b]ut little does the patent examiner know you were able to produce this surprising result *merely by applying ordinary skill* because you used software that doesn’t shy away from trying designs human engineers of ordinary skills have mental roadblocks against trying’ (emphasised in the original)). Whether the assumption that a car frame can be generated by AI by ‘merely applying ordinary skills’ is not at issue here. On the decision-making involved in applying ML, see below at II.6.

<sup>180</sup> EPC, art 83. The reproducibility requirement was developed in the case-law of the EPO. See case G 1/03, decision of 8 April 2004, ECLI:EP:BA:2004:G000103.20040408, point 2.5 of the Reasons (with further references to case-law). Under the EPO’s approach, reproducibility of an invention is a material factor for the requirement of sufficiency of disclosure if the technical effect is expressed as a technical feature of the claim, since such features characterise the claimed subject matter. See case T 1079/08, decision of 19 April 2010, ECLI:EP:BA:2010:T107908.20100419. In contrast, where an effect is not expressed in a claim, its reproducibility would be considered under the inventive step assessment.

<sup>181</sup> Core elements of ML (algorithms, numeric models and computer programs) are excluded from patentability, insofar a patent application relates to such subject matter ‘as such’. EPC, art 52. See also EPO Guidelines for Examination, pt G-II, s 3.3.1 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g\\_ii\\_3\\_3\\_1.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_3_1.htm)> accessed 21 May 2021.

<sup>182</sup> WIPO (n 154) para 72.

<sup>183</sup> *ibid*.

<sup>184</sup> EPC, rule 31.

<sup>185</sup> An example of commonplace misuse of the term ‘AI autonomy’ is one of the top applications of ML – ‘autonomous driving’, which should be more accurately called ‘automated driving’. On this point, see Kim (n 20) 447, note 61.

Autonomy implies self-governance and self-determination,<sup>186</sup> the existence and the ability to exercise free will concerning own decision-making and behaviour.<sup>187</sup> To what extent humans and computers can be comparable in this regard has been a contested issue and perhaps cannot be stated conclusively.<sup>188</sup> While no universally accepted definition of autonomy exists,<sup>189</sup> philosophers distinguish between two aspects: autonomy as ‘the negative condition of freedom from the external constraints’<sup>190</sup> and ‘the positive condition of a self-determined will’.<sup>191</sup> One could say that neither humans nor computers are free from absolute constraints, such as gravity. However, as far as *man-made* rules are concerned, humans have in many situations at least a theoretical possibility not to comply as a manifestation of self-determination (whether or not non-compliance is a rational choice is a different question). In contrast, computers cannot ‘decide’ to violate human-imposed constraints – in other words, constraints in programming are *inherently inviolable*. In computer science, an ‘automaton’ means a machine or mechanism that either reacts to a ‘pre-determined set of rules or adapts to the environmental dynamics in which it operates’ to accomplish a goal.<sup>192</sup> In either case, it is bound by the pre-programmed instructions.<sup>193</sup> The interaction with the real world perceived through the sensors might cause the loss of control over an AI system by its designer or user, but the lack of control would be attributed to the *environment’s unpredictability* and not to a computer’s capacity for ‘self-determination’.

Automation means that a task can be carried out without direct human intervention *during* its implementation.<sup>194</sup> For that, however, the computational process needs to be *conceptualised* and *configured* by a human in the first place. AI systems can automate operations such as data processing or mathematical optimisation to the extent that such tasks can be modelled as computational processes executed without direct human intervention during the implementation phase.<sup>195</sup> The conceptualisation phase includes identifying a problem to be solved through

computation, problem abstraction (i.e. reducing a problem to the elements and relations necessary for understanding and solving it) and its formal representation (e.g. mathematical expressions). While computer programming becomes increasingly automated, conceptualisation is carried out by humans, and it is difficult to see how computers can substitute human intention, aspiration, insight and decision-making in the conception phase.<sup>196</sup>

In popular media, one comes across anthropomorphic depictions of AlphaGo, such as:

drawing on all its other training with millions of moves generated by games with itself, AlphaGo came to *view* Move 37 in a different way. It came to *realize* that, although no professional would play it, the move would likely prove quite successful. It *discovered* this *for itself* [...] through its own process of *introspection* and analysis. [...] AlphaGo learned to discover new strategies for *itself*, by playing millions of games between its neural networks, against themselves, and gradually improving. [...] In other words, AlphaGo *knew* this was not a move that a professional Go player would make.<sup>197</sup>

While it is understandable that popular media outlets use language that appeals to a broad audience, such depiction of AlphaGo can be confusing for the lay audience. The use of anthropomorphic language – ‘came to view’, ‘came to realise’, ‘discovered for itself’ – approximates the AI performance to cognitive processes and creates an impression of an independent agency. At the same time, researchers point out that vesting computers with ‘self-determination’, ‘independence’, ‘freedom’ and the ability to choose a course of action ‘for own reasons’,<sup>198</sup> is misleading<sup>199</sup> and ‘potentially dangerous’.<sup>200</sup> In technical terms, the system’s ‘realisation’ and ‘discovery’ subsist in *calculating the probability* of whether a certain move will be successful based on the given reward and value functions.

ML systems might come across as ‘non-deterministic’ or ‘random due to a *lack of understanding* of cause-and-effect relationships and a lack of resources for controlling sources of variability’.<sup>201</sup> Such perception of ML can illustrate the notion of ‘deterministic chaos’ where deterministic laws govern seemingly random systems.<sup>202</sup> In reality, the term autonomy used in the context of computers ‘does not mean that machines are free in the choices that they make [because] the conditions for deciding on how to proceed are carefully set by human actors’.<sup>203</sup>

**186** The word is derived from the Greek *autos* (self) and *nomos* (law). Online Etymology Dictionary, ‘Autonomy’ <<https://www.etymonline.com/word/autonomy>> accessed 21 May 2021.

**187** Sara Goering, ‘Autonomy’ in Carl Mitcham (ed), *Encyclopedia of Science, Technology, and Ethics* (Macmillan Reference 2005) 155, at 155.

**188** Among other aspects, the biological basis of free will remains an open-ended inquiry. On this subject, see eg Theodosius Dobzhansky, *The Biological Basis of Human Freedom* (Columbia University Press 1956); Chris Willmott, *Biological Determinism, Free Will and Moral Responsibility: Insights from Genetics and Neuroscience* (Springer 2016); Peter Tse, *The Neural Basis of Free Will: Criterial Causation* (MIT Press 2013).

**189** Remmers (n 116) 139 (observing that defining ‘autonomy’ is ‘a complicated affair’).

**190** Goering (n 187) 155 (with further references).

**191** *ibid.*

**192** John Oommen and Sudip Mistra, ‘Cybernetics and Learning Automata’ in Shimon Y Nof (ed), *Springer Handbook of Automation* (Springer 2009) 221, at 221.

**193** A computer might break the default rules given certain conditions, but it would still be bound by the pre-programmed instructions defining such conditions; an accidental ‘violation’ can occur if the programmer made a mistake in defining the constraints.

**194** See eg Shimon Nof, ‘Automation: What It Means to Us Around the World’ in Shimon Y Nof (ed), *Springer Handbook of Automation* (Springer 2009) 13, 22.

**195** Gopinath Rebala, Ajay Ravi and Sanjay Churiwala, *An Introduction to Machine Learning* (Springer 2019) 1 (defining AI as ‘a field of computer science that studies algorithms and techniques for automating solutions to complex problems that are hard to program using conventional programming methods’).

**196** As explained above at II.1.b, humans not only state a problem but conceive a problem-solving approach.

**197** Metz (n 166) (quoting David Silver, one of the developers of AlphaGo) (emphasis added).

**198** Remmers (n 116) 139 (emphasised in the original).

**199** *ibid.* (further noting that such depictions divert attention from ‘genuine issues’).

**200** Watson (n 29) 434.

**201** Alkan Donmez and Johannes A Soons, ‘Impacts of Automation on Precision’ in Shimon Y Nof (ed), *Springer Handbook of Automation* (Springer 2009) 117, 119 (emphasis added) (further pointing out that a closer examination can reveal that seemingly random results are determined by factors imperceptible to an observer).

**202** Britannica, ‘Chaos Theory’ <<https://www.britannica.com/science/chaos-theory>> accessed 21 May 2021.

**203** Merel Noorman and Deborah G Johnson, ‘Negotiating Autonomy and Responsibility in Military Robots’ (2014) 16 *Ethics Inf Technol* 51, 53.

### (ii) A ‘decision’ as the choice of an option from a set of possibilities

If AI systems do not have autonomy in decision-making in the sense of ‘self-determination’ free from the human-imposed constraints, what does a ‘decision’ mean in the ML context?

There is no particular technical definition of a ‘decision’ or ‘decision-making’ in ML. While the HLEG-AI defines AI as systems that can ‘reason’ and ‘decide’,<sup>204</sup> it notes that ‘AI researchers use mostly the notion of rationality [to refer] to the ability to *choose* the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources’.<sup>205</sup> Thus, in general terms, a ‘decision’ can then be understood as the choice of the most suitable option from a set of possibilities, given the instructions.<sup>206</sup>

### (iii) ‘Decision-making’ of AI systems is a function of the decision-making of their designers

In a policy document, one comes across statements that ‘AI-based products can act autonomously by perceiving their environment and *without following a set of pre-determined* [...] instructions, [whereby] their behaviour is constrained by the goal they are given and other relevant design choices made by their developers’.<sup>207</sup> ‘Without following a set of pre-determined instructions’ means in this context that the human does not provide explicit instruction, e.g. ‘in situation A, perform action B’, but a system calculates which action is more optimal based on the given mathematical functions and data. One could say that the inferred ‘rules’ or ‘decisions’ as to which action to take are not provided explicitly and directly by the human user. For instance, instead of being programmed to ‘turn left if there is a wall’, a system’s performance is based on the probability estimation, e.g. based on the training data, the probability of succeeding is 35 per cent if turning right and 65 per cent if turning left. Such probabilities result from complex mathematical calculations involving numerous features and variables guided by the objective function. While an ML model performs in a way it was trained to perform, its performance is not pre-defined *ex ante* explicitly and directly by the designers; instead, it is *numerically derived from training data* through mathematical optimisation, the process set up purposefully and methodically by humans.

<sup>204</sup> HLEG-AI (n 45) 7 (defining AI as ‘systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, *reasoning* on the knowledge derived from this data and *deciding* the best action(s) to take (according to pre-defined parameters) to achieve the given goal’ (emphasis added)). Furthermore, a ‘learning rational AI system’ is depicted as ‘a rational system that, after taking an action, evaluates the new state of the environment (through perception) to determine how successful its action was, and then adapts its reasoning rules and decision making methods’. *ibid* 3.

<sup>205</sup> *ibid* 1-2 (emphasis added).

<sup>206</sup> For instance, if the information derived through sensors is that the floor is dirty, the best action (‘decision’) by an ML-based system for automatic floor cleaning would be to activate the cleaning, otherwise the best action would be to stay still. HLEG-AI (n 45) 3. The detection of whether the floor is dirty or not depends on how an ANN model was trained. *ibid* 4.

<sup>207</sup> European Commission, ‘Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics’ COM(2020) 64 final 6, note 34 (emphasis added).

Thus, contrary to the assumption that ML-based systems can have ‘*freedom to decide* which path to take to achieve the given goal’,<sup>208</sup> there is no decision made by a computer in the sense of ‘freely choosing’ computational steps. A ‘decision’ as to whether this weight should be smaller or bigger than it was in the previous iteration depends on the optimisation algorithm.<sup>209</sup> A ‘decision’ in the model optimisation is, thus, nothing more than *purely mathematical calculations* based on the given functions. In GP and EAs, what is ‘decided’ is which candidate solution will survive to the next generation or be mated for recombination or mutation. Such choices are made based on the selection operators and the fitness values of the population individuals (solution candidates).

A computer does not have a choice whether to perform computations or not, or whether to deviate from the given instructions. Where an RNG is used, there is no freedom for a computer to select a ‘random’ number other than what was ‘seeded’ by a human.<sup>210</sup> Neither do quantum computers have the freedom to decide but can only respond to the signals from the external environment (and one might further ponder to what extent the environment is deterministic). In this regard, the depiction of ML techniques as ‘*self-learning*’ or ‘*self-teaching*’ networks<sup>211</sup> is inaccurate.<sup>212</sup>

The bottom line is that the ‘decision-making’ of AI systems is a product of the decision-making of their designers. Consider an example where one tries to understand why a self-driving car caused an accident. As one commentator explains, the true answer will be not because a car ‘decided’ to do so, but because ‘it applied a transparent and deterministic computation using the values of its parameters, given its current input, and this determined its actions’, whereby ‘those particular parameters are the result of the model that was chosen, the data it was trained on, and the details of the learning algorithm that was used’.<sup>213</sup>

### (iv) ML output is predictable

If ‘predictable’ means that something happens in a way that one expects,<sup>214</sup> ML algorithms and models cannot be viewed as delivering ‘unpredictable’ outcomes. First of all, the task for which an algorithm is deployed or a model is developed is known upfront. Second, the process of deriving the output is also known and transparent. At the heart of ML is the process of optimising an objective function and, even if the exact numeric values optimising the objective function might be unknown *ex ante*, it is certain that such values will be derived based on input data according to an algorithm and predefined configurations, such as model hyperparameters in ANNs, or mutation and recombination operators in GP. Given that

<sup>208</sup> HLEG-AI (n 45) 7 (emphasis added).

<sup>209</sup> Training an ANN is often based on the gradients of the error that an ANN makes.

<sup>210</sup> For a detailed explanation, see below (nn 220-223) and the accompanying text.

<sup>211</sup> Karnow (n 115) xx.

<sup>212</sup> On the (non)-determinism in ML, see below at II.3.b.v.

<sup>213</sup> Dallas Card, ‘The “Black Box” Metaphor in Machine Learning’ (*Towards Data Science*, 5 July 2017) <<https://towardsdatascience.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0>> accessed 21 May 2021.

<sup>214</sup> Cambridge Dictionary, ‘Predictable’ <<https://dictionary.cambridge.org/dictionary/english/predictable>> accessed 21 March 2021.

computers cannot violate instructions,<sup>215</sup> there is *nothing fundamentally unpredictable* about the computational output. The application of randomisation does not make the outcome unpredictable. In ANNs, a good training setup would yield quite similar models, no matter how initial weights are randomised. In GP, the output can be more sensitive to randomisation,<sup>216</sup> but it is ‘expected’ because it is known how the optima are to be found within the given search space.

If predictability is understood as ‘the state of knowing what something is like’,<sup>217</sup> one could say that where ANNs or EAs are applied in technical design and engineering, the exact *visual representation* when numeric values are converted into the attributes of a technical design might not be imagined before implementing an algorithm and might perhaps cause a surprising effect.<sup>218</sup> However, one would argue that such visual representation might be equally surprising if humans had completed the same computations on paper. To emphasise, everything that a computer does – be it inferring a function from the input data according to an objective function or evolving the fittest solution given the optimisation parameters – can be done by a human capable of performing mathematical calculations without a computer, except for it would take a prohibitively long time. If humans can perform the same computational operations and arrive at the same numeric values, one cannot say that such values are ‘unpredictable’.

#### (v) (Non-)deterministic algorithms in computer science

In computer science, an algorithm is deterministic if, given the *same input*, the output of implementing an algorithm is *the same*. A truly non-deterministic computational process would be one where choices among possible successive computational states are executed subject to no predefined conditions. As researchers point out, ‘in any random process the presence of deterministic computational structures can never be ruled out’ and ‘complete randomness is an information theory abstraction only’.<sup>219</sup>

Two methods are currently used to introduce randomness in computation – physical and computational.<sup>220</sup> Physical methods use naturally occurring sources of entropy (such as nuclear decay or atmospheric noise);<sup>221</sup> computational methods use an algorithm that generates a strand of numbers from an initial ‘seed’<sup>222</sup> serving as an input to a mathematical function.<sup>223</sup> The latter type is also known as ‘pseudo-random’ number generators since the ‘seed’ is chosen by a human, while a mathematical function controls the process of generating numbers. Such

method is a state-of-the-art technique most commonly applied in ML today. Both physical and computational methods are *deterministic* in the sense that the same input would produce the same outcome (which, in the case of natural sources of entropy, might be difficult or impossible to ensure in practice). In neither case can a computer ‘decide’ by itself to choose a different number.

The key reason for applying randomisation is that the route to the solution (optimum) can be shorter with different starting points.<sup>224</sup> Therefore, when implementing an algorithm, researchers usually try different initial conditions because the search result can depend on the starting point.<sup>225</sup> Very complicated problems (‘NP-Hard problems’) cannot be solved with the current computer power in a reasonable time. Non-deterministic algorithms do not guarantee to find the optimum, but they can provide a near-optimum solution within a suitable period. To emphasise, there is nothing extravagant about what an RNG does – programmers could select random numbers manually (by rolling dice, for example), but that would be overwhelming and inefficient, given how many random numbers might be required.

#### (vi) Randomisation and reproducibility

The output of implementing the same algorithm will differ if an RNG produces different numbers, which can occur if, for instance, an algorithm was executed at two different points of time and the timestamp served as the ‘seed’. However, if the function and the seed are *identical*, then the sequence of generated random numbers will be the *same*. Consider the example of the initialisation of weights in an ANN coded in Python as `w=numpy.random.randn`. Even though the code does not specify the exact number, the computer will extract a mathematical function from the library that would initialise the weights in response to this instruction. On the outside, this may look like a ‘random’ process, but from a library writer’s perspective, or the designer of the random generator program, the process is neither random nor unexplainable. Libraries get modified and updated, which can be one of the factors of model irreproducibility. However, if the same library version is used and other conditions are observed, the same outcome can be obtained. RNGs are also broadly applied in GP, particularly in the initialisation of the population, the recombination of solutions, selection and mutation.<sup>226</sup> All these instances are

<sup>224</sup> Optimisation algorithms are based on the same principle as the ‘warm-cold’ game, where one searches for a hidden object guided by feedback based on the proximity to the goal. Even though in principle, the desired goal can be reached from any position (provided that it is located within the search space), the choice of the starting location can affect the length and the route of the search process.

<sup>225</sup> The following example can illustrate that the initial position in randomisation is important. Assume our goal is to reach the highest peak of the Alps. The initial conditions are: we are dropped somewhere in the Alps (the random initialisation), we do not have the full map of the Alps, and we cannot see. Given these constraints, we can make our way by trying all directions that go upwards. We can use only the information we can perceive in our immediate vicinity to make the best guess. Even if we always make the best ‘local’ decision (the upward direction), the path could turn out to be leading us farther away from the actual peak.

<sup>226</sup> On the application of RNGs in GP/EAs, see Ivan Zelinka and others, ‘Do Evolutionary Algorithms Indeed Require Random Numbers? Extended Study’ in Ivan Zelinka and others (eds), *Prediction, Modeling and Analysis of Complex Systems. Advances in Intelligent Systems and Computing*, vol 210 (Springer 2013) 62.

<sup>215</sup> As explained above (nn 192-193) and the accompanying text.

<sup>216</sup> On global and local optima in GP, see below (342-344).

<sup>217</sup> Cambridge Dictionary, ‘Predictability’ <<https://dictionary.cambridge.org/dictionary/english/predictability>> accessed 21 March 2021.

<sup>218</sup> Above (nn 176-178) and the accompanying text.

<sup>219</sup> Wang and others (n 105) 7-8.

<sup>220</sup> Shri VK Jain, *Cryptography and Network Security* (Khanna Book Publishing 2017) 251-252.

<sup>221</sup> Luis L Bonilla, Mariano Alvaro and Manuel Carretero, ‘Chaos-based True Random Number Generators’ (2016) 7 J Math Industry 1.

<sup>222</sup> A typical example is a timestamp. If no function and seed are specified, the number generation will depend on a library.

<sup>223</sup> A common pseudorandom generator uses the function  $x_{n+1} = (a * x_n + c) \bmod (m)$ , where  $x_0$  is the seed and  $a, c < m$  are numbers defined by the RNG. As long as one has the seed  $x_0$ , it is possible to replicate the conditions for generating the same random numbers.

deterministic and reproducible as long as an RNG is known and the initial seed is fixed.

Conditions of reproducibility of ML models are specific to an algorithm and a technique applied. If one keeps due records of all relevant information – the algorithm's code, the RNG used and the seed for an RNG, etc. – the output of ML methods is *completely reproducible*. As for the training data, it might be sufficient to describe the training dataset (size, selection criteria, etc.) to build another model with comparable performance (in terms of accuracy, precision, etc.) as the original model, without providing the actual training dataset used to train the original model.

#### (vii) Randomisation, human control and researchers' expectations of success

ML algorithms, such as backpropagation and EAs, are based on optimisation through *iterative search* guided by an objective function. In other words, if the optimisation time were unlimited, the optimum could be found irrespective of the particular random numbers – even random search would find the optimum if it had infinite time, provided such optimum or optima exist within the given search space. Even though the training of AlphaGo included randomisation, success was determined by how the computational process was set up by researchers, including the value and reward functions, according to which the probability of successful moves was estimated.

The relationship between randomisation and researchers' expectations of success is not straightforward. Randomisation does not guarantee that the optimum will be found, no matter what. Applying the wrong ML method to a problem will not lead anywhere; an algorithm can get 'stuck' in a local optimum.<sup>227</sup> Randomisation is a part of ML, but it requires expert knowledge of ML and understanding of the problem at hand to configure a computational process.<sup>228</sup> In general, one can be more confident if a technique was successfully applied in the past to a problem with similar mathematical properties. Whether the use of randomisation can increase a researcher's confidence in solving a problem is *highly individual* to the respective application, given that success usually depends on multiple factors.

### c) Patent law uncertainties revisited

#### (i) Inventorship

In view of the foregoing, the contention that 'neither programming nor data training will directly effectuate or shape the outcome of the search'<sup>229</sup> does not hold. Quite the opposite, it is precisely the inputs and the overall computational set-up that 'shapes' the outcome of ML. Given that randomisation does not indicate the absence of human guidance and decision-making in ML, the assumption that inventing is 'handed over' to non-deterministic ML systems appears unfounded.

<sup>227</sup> This can happen because an algorithm searches for the global optimum without the 'full picture', making decisions based on the 'local' information, as the example of searching for the highest peak of the Alps illustrates (above n 225).

<sup>228</sup> As discussed more in detail below at II.6.b.

<sup>229</sup> Dornis (n 101) 22.

Neither does there seem to be a reason why *predictability* of ML output should be a material factor for allocating inventor's rights *de lege ferenda* in situations where application of ML might lead to a patentable invention. First, this criterion would be contrary to the very notion of invention as a solution to a problem<sup>230</sup> – a goal that is not 'immediately attainable'.<sup>231</sup> Second, predictability of results and human guidance are not linearly related: even a carefully set up process, such as a scientific experiment conducted according to a study protocol, can produce results that might not be envisaged by researchers *ex ante*.<sup>232</sup> The unpredictability of ML output could cast a normative doubt as to whether the inventor's rights should be allocated to a human if it could indicate the computer's ability to generate 'innovative outcomes independently, rather than merely by following digital orders'.<sup>233</sup> As clarified in this and the preceding sections, such contention is unfounded as far as current ML techniques are concerned. The fact that the precise numeric values are unknown before implementing an ML algorithm is not more relevant for the allocation of inventor's rights than in cases where other mathematical optimisation methods or model-based techniques are applied in technical problem-solving.

#### (ii) The assessment of expectations for success as a factor of inventive step/non-obviousness

##### *Expectations of success of a skilled person*

In contrast to biological processes where, 'as in a lottery game, the expectation of success always ranged irrationally from nil to high',<sup>234</sup> the use of randomisation in ML does not appear to exclude the possibility of assessing a skilled person's expectations of success. AI systems are about configuring multiple elements in view of the pursued objective – once the elements are *configured correctly*, one would have reasonable expectations of success. Randomisation can affect the duration and the computational route but not the *existence* of optimum or optima within the given search space. Thus, expectations of success would depend on the ability of a skilled person to configure the 'right' computational setup to find them. Accordingly, in situations where the claimed invention could have been developed by applying ML, it needs to be assessed whether a skilled person<sup>235</sup> following the prior art teaching *would* have applied an ML method in that particular way that would lead to the claimed invention. As will be discussed later, there can be substantial room for decision-making preceding the implementation of an algorithm,

<sup>230</sup> There is no statutory definition of an invention under international or European patent law. Under European patent law, an invention is understood as a technical solution to a technical problem, which is reflected in the so-called 'problem-solution' approach to the inventive step assessment.

<sup>231</sup> George Polya, *Mathematical Discovery: On understanding, learning, and teaching problem solving* (John Wiley & Sons 1962) 117.

<sup>232</sup> Burk (n 17) 316 (observing that '[t]he fact that the output may be unexpected or unpredictable does not remove human causality from their operation').

<sup>233</sup> Yanisky-Ravid and Liu (n 32) 2220.

<sup>234</sup> Case T 0737/96, decision of 9 March 2000, ECLI:EP:BA:2000:T073796.20000309, point 11 of the Reasons.

<sup>235</sup> On factoring ML skills into the definition of a skilled person, see above (nn 85-89) and the accompanying text.



which might pose a challenge regarding how the obviousness of such decisions can be objectively assessed.<sup>236</sup>

#### *A surprising effect of an invention*

Whether an effect is surprising or expected needs to be assessed vis-à-vis the *relevant prior art* on a case-by-case basis. Therefore, it cannot be alleged across the board that ML techniques generate inventive (non-obvious) solutions because they are ‘surprising’ in the abstract. Under the EPO approach, a surprising effect in the case of chemical inventions can be demonstrated by way of a comparative test with the structurally closest prior art to the subject matter of the invention:<sup>237</sup> the greater the structural difference, ‘the less unexpected are any differences in their effects’.<sup>238</sup>

An unexpected effect alone does not automatically confer the presence of an inventive step – other factors need to be considered, such as whether it would be obvious for the skilled person to combine the prior art teachings in expectation of an advantageous effect.<sup>239</sup> An invention would lack an inventive step not only where the results are ‘clearly predictable’ but also where a skilled person would follow the teaching of the prior art with a reasonable expectation of success.<sup>240</sup> For instance, where an invention claims numeric parameters of a technical device obtained by applying the state-of-the-art optimisation method, and where indications in the prior art suggest that ‘favourable results might be obtained by the method of calculation applied’, the effect of the claimed solution would not be considered ‘surprising’.<sup>241</sup> For that, the applied method needs to be disclosed in a patent application or identified otherwise in the course of the patent examination.

While the result of applying known equations<sup>242</sup> can be inferred relatively straightforwardly, ML techniques require multiple elements to be configured to accomplish a task. Even if it might be known from the prior art that a technical problem at issue *could* have been solved through ML, the assessment of inventive step would need to examine in light of individual circumstances whether ML *would* have been applied by an average skilled person in such a way that it would lead to the claimed invention.<sup>243</sup>

#### (iii) Implications for the sufficiency-of-disclosure requirement

As clarified, ML methods incorporating randomisation are, in principle, reproducible as long as their application is properly documented. In this view, the analogy between ML output and non-deterministic biological inventions is

inapt and the reproducibility of ML models can, in principle, be ensured without a deposit system.

Assessment of sufficiency of disclosure is based on a patent application as a whole. Even if two ANN models trained with the same algorithm and hyperparameters might slightly differ due to the initial randomisation of weights, such variations within the claimed process would not be material for fulfilling the disclosure requirement, as long as ‘the claimed process reliably leads to the desired products’.<sup>244</sup> Accordingly, for inventions comprising *inter alia*<sup>245</sup> an ML model – e.g. where the functioning of a technical system is enabled through predictions generated by a trained ANN model<sup>246</sup> – the exact numeric model would not need to be reproduced, provided that the disclosed process allows the skilled person to implement the claimed training process.<sup>247</sup>

### 4. ML is a ‘black box’

AI systems<sup>248</sup> are often called ‘black boxes’ – and even a ‘magical black box of code’<sup>249</sup> – that can generate output without providing ‘any information about what exactly makes them arrive at their predictions’.<sup>250</sup>

#### a) Patent law uncertainties

##### (i) Inventorship where an invention results from an ‘unexplainable’ or ‘unpredictable’ process

It is worth noting that, *de lege lata*, explainability of the inventive process has been relevant for inventorship. Such requirement might be unnecessarily burdensome and, in some situations, unfeasible to fulfil and enforce. From a *de lege ferenda* perspective, the lack of explainability of AI performance could call into question the allocation of

<sup>244</sup> Case T 281/86, decision of 27 January 1988, ECLI:EP:BA:1988:T028186.19880127, point 6 of the Reasons (also noting that art 83 of the EPC does not stipulate that ‘a specifically described example of a process must be exactly repeatable’). See also case T 301/87, decision of 16 February 1989, ECLI:EP:BA:1989:T030187.19890216.

<sup>245</sup> An ML model *per se* can fall within the excluded subject matter; see (nn 181, 297) and the accompanying text.

<sup>246</sup> For instance, an ML model can form part of a method and system for improved aircraft safety that predicts anomalous events. See eg a US patent application US20200317365A1.

<sup>247</sup> In a recent case where the claimed invention used an ANN for the transformation of the blood pressure curve measured at the periphery in the equivalent aortic pressure, the EPO Board of Appeal found that, as far as the training of the neural network was concerned, the patent application only disclosed that ‘the input data should cover a wide range of patients of different ages, gender, constitution type, health condition, and the like [but not] which input data for exercising the artificial neural network is suitable for exercising the invention, or at least one data set suitable for solving the present technical problem. Exercising the artificial neural network *can therefore not be reworked* by the specialist and the person skilled in the art can therefore not be carried out. The present invention based on mechanical learning in particular in conjunction with an artificial neural network is thus not sufficiently disclosed because the training according to the invention is *not executable* for lack of corresponding disclosure.’ Case T 0161/18, decision of 12 May 2020, ECLI:EP:BA:2020:T016118.20200512, point 2.2 of the Reasons (emphasis added) (translated by Daria Kim).

<sup>248</sup> Besides DL models, other AI systems can be prone to the problem of interpretability. For example, an evolutionary algorithm running thousands of iterations can produce results that can be difficult to track and decipher how they were produced. Support-vector machines can sometimes be difficult to explain.

<sup>249</sup> Tom Scott, ‘There Is No Algorithm for Truth’ (*The Royal Institution*, 24 October 2019) 7 min 25 sec <<https://youtu.be/leX541Dr2rU>> accessed 21 May 2021.

<sup>250</sup> Wojciech Samek and Klaus Robert Müller, ‘Towards Explainable Artificial Intelligence’ in Wojciech Samek and others (eds), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science*, vol 11700 (Springer 2019) 5.

<sup>236</sup> On human decision-making when applying ML as the true indication of knowledge and skills in ML, see below at II.6.

<sup>237</sup> Case T 181/82, decision of 28 February 1984, ECLI:EP:BA:1984:T018182.19840228, point 4 of the Reasons.

<sup>238</sup> *ibid* point 5 of the Reasons.

<sup>239</sup> Bostedt and others (n 94) 270 (with further references).

<sup>240</sup> Case T 149/93, decision of 23 March 1995, ECLI:EP:BA:1995:T014993.19950323, point 5.2 of the Reasons.

<sup>241</sup> Case T 36/82, decision of 25 October 1982, ECLI:EP:BA:1982:T003682.19821025.

<sup>242</sup> *ibid*.

<sup>243</sup> On the decision-making in applying ML techniques as a true indicator of expertise in ML, see below at II.6.b.

inventor's rights to a human if it could indicate an autonomous, 'supernatural agency'<sup>251</sup> of computers.

As explained under Assumptions 1-3, to solve a problem through ML, the computational process needs to be thoughtfully staged by those developing and applying ML techniques. To settle any remaining doubts, it appears pertinent to clarify why some ML models are considered 'black boxes', given that the process of building them is known and understood.

### (ii) The assessment of inventive step where an invention results from an 'unexplainable' ML process

The 'black-box' characterisation of ML raises the question of whether the likelihood that a skilled person applying ML *would* have arrived at the claimed invention can be objectively assessed if the workings of ML techniques cannot be explained.

### (iii) The sufficiency-of-disclosure requirement in situations where an 'unexplainable' ML method is claimed within an invention

It has been argued that 'the "black-box" nature of certain AI systems [...] may make it challenging to provide a sufficiently clear and complete disclosure for the invention to be carried out by a [person skilled in the art]'.<sup>252</sup> In this view, it needs to be clarified how the allegedly limited explainability of ML systems interacts with their reproducibility.

## b) A technical perspective on explainability of ML

### (i) What exactly is unexplainable about ML models?

There is no single specific definition of 'explainability'<sup>253</sup> of ML and, hence, the lack thereof. A 'black box' computational model is defined as the one 'for which the inputs and outputs are visible to the user, but its internal workings are not'.<sup>254</sup> In the case of ML models, however, 'visibility' of internal workings is usually possible as mathematical computations forming a model can be saved as a data file and examined. Moreover, commentators point out that, contrary to a popular view depicting ANNs 'as black boxes which mystically determine complex patterns in data[, ...] neural network designers typically perform extensive knowledge engineering'.<sup>255</sup> While this quote dates back a couple of decades, contemporary AI systems are not mystical either – they are complex. The issue of explainability concerns the *meaning of input-output correlations* revealed by an algorithm. What needs to be explained are factors behind the specific weighting of features, e.g. why feature *x* in input data correlates with an output/target *y*. When a 'black-box' ML model is

defined as a situation 'where it is not possible to trace back to the reason for certain decisions',<sup>256</sup> it means that factors ('reasons') determining the weighting of features that underlie a prediction ('decision'<sup>257</sup>) are not well understood. As tautological as it might sound, statistical correlations do not denote causation.<sup>258</sup> While model features are found mathematically, the semantics behind numbers are not revealed unless one 'translates' an ANN into a meaningful narrative.

The depictions of ML techniques being 'very successful from the accuracy point of view [but] very *opaque* in terms of understanding *how* they make decisions'<sup>259</sup> create an impression that AI systems 'know' the reasons but keep them secret. However, the truth is that there is no 'ready-made' knowledge within ANNs – it remains a collection of data points unless one construes a sensible narrative. Construing a narrative requires a skilled researcher with the relevant knowledge, willingness to treat an ML model as a 'glass box' and not a 'black box'<sup>260</sup> and access to the necessary information about the model design. The AlphaFold example mentioned earlier<sup>261</sup> highlights the importance of interpreting the numeric output of ML – while the results revealed by AlphaFold can certainly advance scientific understanding, the factors determining protein folding remain 'a black box'.<sup>262</sup>

Portrayals of AlphaGo making 'a brilliant move' that 'no human could understand'<sup>263</sup> are typical examples of mass media's attempt to sensationalise. AlphaGo's 'brilliant' move might not have been understood so far, but it does not mean that it is fundamentally incomprehensible, because, in essence, it is based on *mathematical optimisation*. AlphaGo was described as the software that 'modifies itself, [while] its analyses cannot be understood by humans even as it outstrips human performance'.<sup>264</sup> Such portrayal is not quite accurate – everything AlphaGo does is doable by a human, albeit with a significant difference in the speed of performing computations. The concept of 'understood' might refer here to the ability to generalise the steps of AlphaGo as the rules that explain why the moves optimised through RL ultimately led to the successful performance.<sup>265</sup> Given the complexity of computations, it can take years for a human to understand them. Yet, it is not impossible.

The key difference between humans and AI systems such as AlphaGo is that, while a human can contemplate

<sup>256</sup> HLEG-AI (n 45) 6.

<sup>257</sup> On the notion of 'decision' in ML, see above (nn 204-206).

<sup>258</sup> Spurious correlations found through ML are also known as a 'Clever Hans phenomenon', situations where a feature in the data is 'highly correlated with the correct outcome [...], but is not the cause for the answer being correct'. Eugen Lindwurm, 'Deep Learning, Meet Clever Hans' (*Towards Data Science*, 15 August 2020) <<https://towardsdatascience.com/deep-learning-meet-clever-hans-3576144dc5a9>> accessed 17 August 2021. See also Sebastian Lapuschkin and others, 'Unmasking Clever Hans Predictors and Assessing What Machines Really Learn' (2019) 10 *Nature Communications* 1096.

<sup>259</sup> HLEG-AI (n 45) 6 (emphasis added).

<sup>260</sup> Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1 *Nature Machine Intelligence* 206, at 206.

<sup>261</sup> Above (nn 119-121) and the accompanying text.

<sup>262</sup> Above (n 121).

<sup>263</sup> Metz (n 166).

<sup>264</sup> Karnow (n 115) xx.

<sup>265</sup> The key to AlphaGo's performance was that, while it was playing against itself millions of times, it calculated ('learnt') a strategy for implementing moves that more often than not turned out to be successful in the long run.

<sup>251</sup> As a side note, vesting autonomous agency in 'black box' ML models might illustrate a human tendency to attribute supernatural agency to inexplicable phenomena. See eg Terence L Nichols, 'Miracles, the Supernatural, and the Problem of Extrinsicism' (1990) 71 *Gregorianum* 23; Vernon Pratt, 'The Inexplicable and the Supernatural' (1968) 43 *Philosophy* 248.

<sup>252</sup> Hugenholtz and others (n 21) 112.

<sup>253</sup> Riccardo Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51 *ACM Computing Surveys* 1.

<sup>254</sup> Government Office for Science, 'Computational Modelling: Technological Futures' (2018) 12 <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/682579/computational-modelling-blackett-review.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/682579/computational-modelling-blackett-review.pdf)> accessed 21 May 2021.

<sup>255</sup> Steven Walczak and Narciso Cerpa, 'Heuristic Principles for the Design of Artificial Neural Networks' (1999) 41(2) *Information and Software Technology* 107.

many steps in advance in a board game, a computer can calculate several times more steps beforehand and choose the one with the highest likelihood of success in the long run.<sup>266</sup> A human would not have considered that landmark move, because the benefit of it comes many moves afterwards, and it is quite difficult for a human to calculate that many steps in advance. However, without the time limitation, and maybe with a piece of paper for note-taking, a human could, in principle, have arrived at that ‘brilliant move’ as well.

#### (ii) What causes computational complexity of ANNs?

Complexity is caused by the interaction of numerous (millions of) units (neurons) within a multi-layer non-linear structure of ANNs. As one can be puzzled when looking under the bonnet of a car, one might not readily make sense of the internal states of ANNs composed of lots of numbers. Yet, both a car and an ANN model are products of deliberate and well thought-through engineering. The process of implementing a model is purely mathematical – ANNs are ‘just a set of matrix operations and finding derivatives’.<sup>267</sup> For instance, there can be tens of thousands of input features for training a model for cat/dog image classification, while an ANN can contain tens of millions of weights and bias parameters derived when input features are passed through its layers, each containing hundreds of neurons.<sup>268</sup>

#### (iii) Explainability and causality

Computational complexity does not denote the lack of causality between human input and decision-making and the performance of an ML model. How a model is built is a transparent process. The answer to why an ML model was built and performs in a certain way ‘lies in the combination of the assumptions [made by researchers], the data it was trained on, and various decisions made about how to learn the parameters, including the randomness in the initialization’.<sup>269</sup> In other words, from the perspective of designers of ML systems and data scientists, there is nothing fundamentally unexplainable about how correlations between the data points are found.<sup>270</sup>

#### (iv) Model explainability and reproducibility

The limited explainability of ML models does not affect their reproducibility – an ‘unexplainable’ model can be consistently reproduced. For that, one needs to obtain a detailed description of the applied method, the type of model and algorithm used, exact values of the hyperparameters (where they are required for a model), the seed for an RNG, the number of epochs, evaluation metrics, computer configurations (software, operating system, hardware), and so on. Ultimately one needs to know all decisions made to develop the first model, otherwise model reproduction might be difficult and unreliable.

<sup>266</sup> On ‘policy’ in RL, see above (nn 112, 117).

<sup>267</sup> WIPO (n 154) para 31.

<sup>268</sup> *ibid* para 24.

<sup>269</sup> Card (n 213).

<sup>270</sup> *ibid* (pointing out that ‘[t]he only explanation for why the box is doing what it does is that all of the components are following the rules that govern their individual behaviour, and the overall behaviour emerges from their interactions’).

### c) Patent law uncertainties revisited

#### (i) Implications for inventorship

Given that a ‘black box’ notion in the context of ML does not denote the absence of a causal link between the human input, the way computation is performed and ML outcome, there appears to be no reason to deny the human contribution to the conception and development of an invention, in situations where an underlying technical problem was solved by applying ML.

#### (ii) Implications for the assessment of inventive step

As clarified, the ‘black box’ characterisation of ML models does not denote the absence of deliberation and careful decision-making on the part of researchers designing a model and interpreting computational output. Accordingly, in situations where a technical problem could have been solved through ML, it needs to be assessed whether a skilled person would have designed a model and interpreted the computational outcome in the way leading to the claimed invention.

#### (iii) Implications for the sufficiency-of-disclosure requirement

Given that ML models can be consistently reproduced even if they might not be readily understood, the limited explainability of ML models does not present an insurmountable challenge for the sufficiency-of-disclosure requirement. Where an invention comprises *inter alia* an ML model, the earlier discussed conditions<sup>271</sup> would need to be ensured to fulfil the disclosure requirement.

## 5. ML is a ‘general-purpose technology’

AI has been characterised as the ‘next general purpose technology’ (GPT),<sup>272</sup> given its broad applicability and the potential to enable innovation across economic sectors.<sup>273</sup> As a GPT, AI holds a promise to drive ‘sweeping transformative processes’<sup>274</sup> and to generate ‘a wave of complementary innovations in a wide and ever expanding range of applications sectors’.<sup>275</sup> As an ‘enabling technology’,<sup>276</sup> it can open up new opportunities rather than offer final products and solutions.

<sup>271</sup> Above at II.3.c.iii.

<sup>272</sup> Manuel Trajtenberg, ‘AI as the Next GPT: A Political-Economy Perspective’ (2018) NBER Working Paper No 24245 <[https://www.nber.org/system/files/working\\_papers/w24245/w24245.pdf](https://www.nber.org/system/files/working_papers/w24245/w24245.pdf)> accessed 21 May 2021; WIPO, ‘WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)’ WIPO/IP/AI/2/GE/20/1 (13 December 2019) 2 <[https://www.wipo.int/edocs/mdocs/mdocs/en/wipo\\_ip\\_ai\\_ge\\_20/wipo\\_ip\\_ai\\_2\\_ge\\_20\\_1.pdf](https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_ge_20/wipo_ip_ai_2_ge_20_1.pdf)> accessed 21 May 2021. But see Simone Vannuccini and Ekaterina Prytkova, ‘Artificial Intelligence’s New Clothes? From General Purpose Technology to Large Technical System’ (2021) SPRU Working Paper Series 1 <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3704011](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3704011)> accessed 21 May 2021 (arguing that, ‘for now, the claim “AI equals GPT” is premature, and eventually, taking into account potential future scenarios, it can turn out to be incorrect’).

<sup>273</sup> Timothy Bresnahan, ‘General Purpose Technologies’ in Bronwyn H Hall and Nathan Rosenberg (eds), *Handbooks of the Economics of Innovation*, vol 2 (Elsevier BV 2010) 761, 764.

<sup>274</sup> Trajtenberg (n 272) 2.

<sup>275</sup> WIPO, *Technology Trends 2019. Artificial Intelligence* (WIPO 2019) 142 (citing Dominique Foray).

<sup>276</sup> Yann Ménière, Ilja Rudyk and Javier Valdes, *Patents and the Fourth Industrial Revolution. The inventions behind digital transformation* (EPO 2017) 10.

### a) The importance of access to inputs for ML from an innovation policy perspective

From an innovation policy perspective, a significant implication of the GPT nature of ML is that *access to inputs* for building ML models – which in turn form the basis for developing sector- and case-specific applications – plays a paramount role in realising the potential of ML as an enabling technology across the economic sectors. In this regard, exclusive IP rights – to an extent applicable to inputs for developing ML models and models themselves – and the way such rights are exercised, can have a sizable impact on the innovation processes enabled through ML.

In the context of enabling technologies (e.g. research tools), the question of whether patents facilitate or hinder innovation has been subject to a long-standing debate.<sup>277</sup> The facilitating effect is associated with the role of patent rights as innovation incentives. The risk of a hampering effect has been hypothesised in the context of follow-on innovation, where a patented technology can be used as an input in knowledge creation. By definition, enabling technologies are broadly applicable innovation inputs.<sup>278</sup> Consequently, where usage rights in such technologies cannot be allocated efficiently, the suboptimal realisation of the innovative potential is posited to entail a welfare loss. Furthermore, AI techniques such as ML are peculiar in that they consist of multiple elements, which sets a precondition for rights fragmentation and high transaction costs of their convergence.<sup>279</sup>

Against this backdrop, one might contemplate that exclusive IP rights in AI techniques might have a ‘stifling effect’ on AI-enabled innovation<sup>280</sup> if the usage rights cannot be efficiently allocated. Accordingly, two key issues need to be clarified: (i) which constitutive elements of ML can function as ‘general-purpose’ building blocks; and (ii) whether such generic components might be subject to patent protection.<sup>281</sup>

### b) A technical perspective on ML as a GPT

The characterisation of ML as a GPT is accurate in that ML, as a variety of computational model-based techniques, is broadly applicable to problem-solving across technological and engineering fields.<sup>282</sup> At the heart of ML is

optimisation.<sup>283</sup> The ubiquitous<sup>284</sup> nature of optimisation explains why the scope of ML applications is expansive.

The so-called ‘no free lunch theorem’ in ML<sup>285</sup> postulates that there is no true ‘general-purpose’ ML method because no single algorithm can solve all ML problems better than any other algorithm. Some methods can be applied across a specific subset of problems sharing certain similar properties (e.g. a specific domain application, text and image recognition).

Some basic techniques and principles applied in ML can be viewed as ‘generic’. Examples include core algorithms such as gradient descent and backpropagation of errors that constitute the basis for the families of algorithms used in ML; basic mathematical tools, such as the derivation of a function; and methods of pre-processing training data, such as the methods for dimensionality reduction or outlier removal. Various software used to pre-process the training data is usually generic. It should be mentioned that the notion ‘generic’ does not imply that a method is guaranteed to work under any conditions; for instance, numerous dimensionality reduction methods exist, but only one might work for a given dataset and purpose. In GP and EAs, the crossover and mutation operators can be considered generic as they can be applied to a specific encoding of the solutions, regardless of the task to be solved. Such generic techniques are usually in the public domain.

For ANNs, two key components are training data and algorithms. Most algorithms are publicly available and shared as part of software libraries.<sup>286</sup> In contrast, access to training data, especially held by private companies, can be tricky. Collaboration with private companies often involves contractual restrictions regarding sharing datasets used for developing ML models. Without sharing datasets, however, the research community might not be able to reproduce results.

ML models are usually developed to perform a narrow task and their generalisability depends on multiple factors. Under the so-called ‘transfer learning’ approach, an ANN model initially trained for one task might be re-used as a foundation for developing another model for a related but distinct task. Whether or not transfer learning can be a relevant and efficient solution is individual to an ML method and its particular application. As far as access is concerned, companies tend to hold ML models

<sup>277</sup> Michael A Heller and Rebecca S Eisenberg, ‘Can Patents Deter Innovation? The anticommons in biomedical research’ (1998) 280 *Science* 698. For an overview of the debate, see Fiona Murray and Scott Stern, ‘Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An empirical test of the anti-commons hypothesis’ (2007) 63 *Journal of Economic Behavior & Organization* 648.

<sup>278</sup> On the enabling nature of GPT and the relationship between GPT and application sectors, see Bresnahan (n 273).

<sup>279</sup> Heller and Eisenberg (n 277) (identifying these factors as the preconditions for ‘anticommons’ in the context of biotechnological research tools).

<sup>280</sup> Josef Drexel and others, ‘Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate’ (2021) Max Planck Institute for Innovation and Competition Research Paper No 21-10, 21 <[https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI\\_PositionPaper\\_SSRN\\_21-10.pdf](https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_PositionPaper_SSRN_21-10.pdf)> accessed 21 May 2021 (noting that it is *prima facie* not clear whether such concerns can be justified, given that ‘ML models do not appear to be unique in the same way as, for instance, molecular research tools and resources, such as cell lines or DNA libraries’).

<sup>281</sup> Other forms of IP protection lie beyond the scope of this inquiry. A comprehensive study would need to examine all relevant legal determinants of control over and access to such components.

<sup>282</sup> Kim (n 20) 446, notes 43-50.

<sup>283</sup> As explained above, ML methods are based on optimising an objective function. Above at II.1.b.ii.

<sup>284</sup> Ramteen Sioshansi and Antonio J Conejo, ‘Optimization is Ubiquitous’ in Ramteen Sioshansi and Antonio J Conejo (eds), *Optimization in Engineering. Models and Algorithms*, vol 120 (Springer 2017) 1.

<sup>285</sup> This means that ‘no single machine learning algorithm is universally the best-performing algorithm for all problems’; an algorithm ‘may perform very well for one problem, but that gives us no reason to believe it will do just as well on a different problem where the same assumptions may not work’. Amol Mavuduru, ‘What “No Free Lunch” Really Means in Machine Learning. Demystifying This Often Misunderstood Theorem’ (*Toward Data Science*, 12 November 2020) <<https://towardsdatascience.com/what-no-free-lunch-really-means-in-machine-learning-85493215625d>> accessed 21 May 2021.

<sup>286</sup> See eg <<https://scikit-learn.org/stable/>> accessed 21 May 2021. Sharing software resources for ML can take place in various ways. Usually, for any ML resource, either a program (eg Weka) or an open-source code library (eg TensorFlow) is created. They are shared either on a designated website or through code sharing platforms (eg Github). Many packages for operating systems or programming languages, eg Python, are also available through package managers providing access to such resources form a central server distributing software that can be shared over platforms such as Github or Gitlab.

privately, either entirely or partially, while academics tend to make their work accessible.

ML is characterised by high modularity, as its application to a specific problem requires configuring multiple elements and combining various methods. If certain ‘ingredients’ are not available, they can usually be substituted with the available alternatives to achieve comparable results.<sup>287</sup> Substitutability of training data is, however, most difficult due to the limited access.

### c) Implications for patent policy and law revisited

As discussed, the core ‘building blocks’ of ML, except for training data, tend to be disclosed in academic publications,<sup>288</sup> which can alleviate concerns regarding their accessibility. Nevertheless, given the characterisation of ML as a broadly applicable enabling technique, the patentability of ML systems and their constitutive elements *de lege lata* merits a closer look, especially as regards the scope of protection.

Under European patent law, the essential elements of AI – mathematical methods<sup>289</sup> and computer programs<sup>290</sup> – fall within the categories of ‘non-inventions’ for the lack of technical character.<sup>291</sup> Besides, core AI applications based on predictive analytics or related to knowledge representation, reasoning, planning and scheduling<sup>292</sup> might qualify as business methods and, hence, come within the scope of the excluded subject matter as well.<sup>293</sup> The exclusion applies ‘to the extent to which a European patent application or European patent relates to such subject-matter or activities as such’.<sup>294</sup> While such approach appears, in essence, balanced and reasonable, some uncertainty persists between the clear-cut cases,<sup>295</sup> and the question of under what conditions computational methods should be considered ‘technical’ remains open-ended.<sup>296</sup>

In principle, ML algorithms and models can constitute only part of a patentable invention. According to the EPO Examination Guidelines, implementing a mathematical method on a computer would be sufficient to confer a

technical character on the subject matter as a whole<sup>297</sup> and, thus, overcome the exclusion. An analysis of whether the reference to implementation on a computer *should* suffice to overcome the exclusion under Art. 52(2)(a) of the EPC goes beyond the scope of this paper. The point worth highlighting is that the patent system’s goal of affording protection only to the technical subject matter and the policy concern regarding the scope of patent rights in enabling technologies<sup>298</sup> are related yet distinct issues. Adding the reference to a generic computer as a technical device to an independent patent claim directed to a ‘core’ ML method which is otherwise excluded as a mathematical method,<sup>299</sup> would not limit the scope of exclusive rights. In contrast, consider the distinction between patents for ‘core’ and ‘applied’ AI techniques: the former can claim a broadly applicable method of ML,<sup>300</sup> the latter can be confined to its particular application in a technical use-case. The difference in terms of the scope of protection is palpable.

In the context of European patent law, this issue is reminiscent of the debate on absolute vs purpose-bound patent protection for chemical substances (in particular, DNA sequences), which remains unsettled both from a *de lege lata*<sup>301</sup> and a *de lege ferenda*<sup>302</sup> perspective. In the US, the connection between the excluded subject matter and the scope of patent protection is reflected in the ‘building blocks’ doctrine applied by the US Supreme Court. By excluding ‘abstract ideas’, the Court in *Mayo*<sup>303</sup> and *Alice*<sup>304</sup> intended to keep innovation inputs – ‘the “buildin[g] block[s]” of human ingenuity’<sup>305</sup> – outside protection by exclusive rights. In this regard, the exclusion can be deemed instrumental for preserving ‘abstract intellectual concepts [...] not patentable, as they are the *basic tools of scientific and technological work*’.<sup>306</sup>

While discoveries, scientific theories and mathematical methods are excluded from patentability for the lack of technical character, they also happen to be broadly applicable knowledge inputs akin to research tools necessary to create further knowledge. For the reasons better explained by economists,<sup>307</sup> granting exclusive IP rights

**287** Vicky Yu, ‘Why Building a Machine Learning Model Is Like Cooking’ (*Towards Data Science*, 2 November 2020) <<https://towardsdatascience.com/why-building-a-machine-learning-model-is-like-cooking-4bed1f6115d1>> accessed 21 May 2021.

**288** On the rapidly growing academic output, see eg WIPO (n 275).

**289** EPC, art 52(2)(a).

**290** EPC, art 52(2)(c).

**291** EPC, art 52(2). See also EPO Guidelines for Examination, pt G-II, s 3.3.1 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e\\_g\\_ii\\_3\\_3\\_1.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e_g_ii_3_3_1.htm)> accessed 21 May 2021.

**292** WIPO (n 275) 148.

**293** EPC, art 52(2)(c).

**294** EPC, art 52(3).

**295** For instance, if a claim is directed solely to a method that increases the accuracy of the financial risk assessment, whereby the output is pure information, it would likely be devoid of technical character. In contrast, an ML model applied to detect a thermodynamic problem and integrated within a system of safety control on an aircraft would contribute to the technical solution.

**296** In its decision in case G 1/19 of 10 March 2021 (ECLI:EP:BA:2021:G000119.20210310), the EPO Enlarged Board of Appeal addressed the definition of technical character in the context of art 56 of the EPC (within the framework of the so-called COMVIK approach to the patentability of computer-implemented inventions). While the Board abstained ‘from putting forward a definition for “technical”’, it indicated that, for computer-implemented numeric simulation methods, ‘direct link with physical reality [...] cannot be a necessary condition [of technicality]’ (para 88), and that ‘it is neither a sufficient nor a necessary condition that a numerical simulation is based, at least in part, on technical principles that underlie the simulated system or process’ (para 142).

**297** EPO Guidelines for Examination, pt G-II, s 3.3 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e\\_g\\_ii\\_3\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e_g_ii_3_3.htm)> accessed 21 May 2021.

**298** Above (nn 277-280) and the accompanying text.

**299** EPO Guidelines for Examination, pt G-II, ss 3.3, 3.3.1 <[https://www.epo.org/law-practice/legal-texts/html/guidelines/e\\_g\\_ii\\_3\\_3.htm](https://www.epo.org/law-practice/legal-texts/html/guidelines/e_g_ii_3_3.htm)> accessed 21 May 2021.

**300** See eg a US patent application US2021256375 claiming a method for reduced computation real-time recurrent learning based on forward propagation.

**301** See eg Geertrui Van Overwalle, ‘The CJEU Monsanto Soybean Decision and Patent Scope: As Clear as Mud’ (2011) 42 IIC 1.

**302** See eg Matthias Lamping, ‘Purpose-Bound Patent Protection for Genes’ (2010) 1(4) European Journal of Risk Regulation 445, 447 ff (putting forward the reasons in favour of the purpose-bound protection, including that absolute protection ‘distorts market-oriented incentive mechanism’). But see also Michael A Kock, ‘Purpose-Bound Protection for DNA Sequences: In through the back door?’ (2010) 5(7) Journal of Intellectual Property Law & Practice 495 (arguing that absolute protection would benefit innovation).

**303** *Mayo Collaborative Services v Prometheus Laboratories, Inc.* 132 S.Ct. 1289 (2012).

**304** *Alice Corp. Pty. Ltd. v CLS Bank Int’l* 134 S. Ct. 2347 (2014).

**305** *ibid* 1-2.

**306** *Mayo Collaborative Services v Prometheus Laboratories* (n 303) 2 (referencing the 1972 case *Gottschalk v Benson*) (emphasis added).

**307** Cristiano Antonelli, ‘Technological Knowledge as an Essential Facility’ (2007) 17 Journal of Evolutionary Economics 451. The issue of exclusive rights in research inputs lies at the heart of the ‘anticommons’ debate in IP law; see above (n 277). See also Ronald F King, Ivan Major

in such inputs might not be an optimal innovation policy from a welfare perspective.<sup>308</sup> In this view, the present regime where an independent patent claim can be directed to an ML method as such, not limited to an application in a specific technical use-case, might need to be re-considered from a *de lege ferenda* perspective.<sup>309</sup>

Furthermore, the question of access to training data, which is often subject to factual control and trade secret protection, merits closer analysis. Overall, further research needs to examine how access measures should be optimally designed to remedy a market failure of data-sharing in the context of ML-enabled innovation.

## 6. ML is a ‘general-purpose method of invention’

Besides being characterised as GPT, AI – particularly DL – has been viewed as ‘a general-purpose method of invention’.<sup>310</sup>

### a) Uncertainty regarding the assessment of inventive step

From a patent law perspective, the main implication of ML being characterised as a ‘method of invention’ is that the use of ML, as a computational problem-solving method, should be factored in the definition of a relevant skilled person<sup>311</sup> when assessing the obviousness of inventions that could or would have been developed by applying ML.<sup>312</sup> The skilled person is ‘free to employ the best means already available for his purposes [...] provided this involves a choice from a multiplicity of possibilities’.<sup>313</sup> Otherwise, the lack of alternatives may ‘create a “one-way-street” situation leading to predictable advantages which remain obvious in spite of the existence of some unexpected “bonus” effect’.<sup>314</sup> Besides, where a skilled person would apply routine trial and error, the outcome is deemed obvious/lacking an

inventive step.<sup>315</sup> ML techniques involve trial and error<sup>316</sup> and iterative search.<sup>317</sup> According to Abbott, ‘as research is augmented and then automated by machines, the average worker will routinely generate patentable output’,<sup>318</sup> whereby a low obviousness standard would entail a danger of the stifling effect on innovation.<sup>319</sup> Furthermore, the attribute ‘general-purpose’ might suggest that ML algorithms can be applied ‘out of the box’ to different tasks.

As clarified under Assumption 2, the seeming simplicity of human input vis-à-vis computational complexity is not a reliable indicator of the obviousness of an invention. To clarify whether the application of ML techniques represents a ‘one-way street’ or involves only ‘routine’ trial and error, let us take a closer look at the decision-making involved in ML.

### b) A technical perspective on ML as a ‘general-purpose method of invention’

#### (i) ML as a special-purpose technique

Contrary to the ‘general-purpose’ attribute, ML applications need to be ‘*surgically altered* or *purpose-built*’.<sup>320</sup> This requires ‘lots of preparation by human researchers or engineers, *special-purpose* coding, *special-purpose* sets of training data, and a custom learning structure for each new problem domain’.<sup>321</sup> ML models are only ‘as good as the assumptions that they were created with and the data that was used to train them’.<sup>322</sup> As there is no ‘one-size-fits-all’ ML technique, choosing a method that can work better for a particular problem involves expert knowledge and insight. Most ML algorithms can work ‘out of the box’ – more precisely, ‘out of the library’<sup>323</sup> – only for a rather constrained set of relatively simple problems. In some cases – such as image or speech recognition – only minimum adjustment might be necessary, while in other cases almost every aspect of an ML technique might need to be configured or fine-tuned.

#### (ii) Which decisions are crucial for the successful application of ML?

The saying ‘garbage in, garbage out’ holds in ML, as the whole preparation is more important than implementing an algorithm. Decisions during the preparation phase require a good grasp of a problem and adequate knowledge and experience. Which decisions specifically might be crucial for the successful application of ML is individual to a particular ML technique and a use case.

In ANNs, one needs to configure hyperparameters<sup>324</sup> that would lead to the best performance of an algorithm

and Cosmin Gabriel Marian, ‘Confusions in the Anticommons’ (2016) 9(7) *Journal of Politics and Law* 64; Yi Zhou, ‘The Tragedy of the Anticommons in Knowledge’ (2015) 48 *Review of Radical Political Economics* 1; Norbert Schulz, Francesco Parisi and Ben Depoorter, ‘Fragmentation in Property: Towards a general model’ (2001) 158 *Journal of Institutional and Theoretical Economics* 594.

<sup>308</sup> Antonelli (n 307) 467. See also *Mayo Collaborative Services v Prometheus Laboratories* (n 303) 20 (holding that ‘the underlying functional concern here is a relative one: how much future innovation is foreclosed relative to the contribution of the inventor’).

<sup>309</sup> Notably, the intention of the drafters of art 52(2) of the EPC was not only to exclude pure mathematical methods but their application in solving a technical problem. As Pila writes: ‘The reason for that proposal was the delegation’s view that “[a]ny mathematical method as such (pure mathematics) and its application in solving a technical problem (applied mathematics) should [...] be considered to be intellectual activities and therefore excluded from patentability”. Put differently, the concern was that by qualifying the mathematical methods exclusion with use of the phrase “as such”, the EPC would enable the patenting of applications of mathematical methods, which, it was thought, should remain unpatentable for failure to constitute an invention.’ Justine Pila, ‘Article 52(2) of the Convention on the Grant of European Patents: What Did the Framers Intend? A Study of the *Travaux Préparatoires*’ (2005) 36 *IIC* 755, 766.

<sup>310</sup> Ian M Cockburn, Rebecca Henderson and Scott Stern, ‘The Impact of Artificial Intelligence on Innovation’ (2018) NBER Working Paper No 24449, 27.

<sup>311</sup> Above (nn 85-89) and the accompanying text.

<sup>312</sup> To illustrate, where ML is applied in the molecule design, the resulting molecule can constitute an invention. However, the molecule will be deemed ‘obvious’ in a patent law sense if a (hypothetical) skilled person with average knowledge and skills would have applied ML in a way to obtain the results falling within the scope of the claimed invention.

<sup>313</sup> Case T 192/82, decision of 22 March 1984, ECLI:EP:BA:1984:T019282.19840322, point 16 of the Reasons (emphasis added).

<sup>314</sup> *ibid.*

<sup>315</sup> PCT Examination Guidelines (n 174) 126-127.

<sup>316</sup> WIPO (n 154) para 21 (noting that the ‘mechanism for training a [neural network] is basically “learning from mistakes”’).

<sup>317</sup> Above (n 134) and the accompanying text.

<sup>318</sup> Abbott (n 104) 46.

<sup>319</sup> *ibid* 46-47.

<sup>320</sup> Rodney Brooks, ‘The Seven Deadly Sins of AI Predictions’ (2017) 79 *MIT Technology Review* 120 (emphasis added).

<sup>321</sup> *ibid* (emphasis added).

<sup>322</sup> Mavuduru (n 285).

<sup>323</sup> Numerous software libraries exist that provide access to collections of standard software programs and families of ML algorithms (classification, regression, clustering, etc.) that one can use out of the box for individual purposes. See eg <<https://scikit-learn.org/stable/>> accessed 7 November 2021.

<sup>324</sup> For a definition, see above (n 62).

and the most suitable model for a given problem. Usually, there is no prior knowledge about which hyperparameters are right for a particular task – as one commentator puts it: ‘Typically, the hyper-parameter exploration process is *painstakingly manual*, given that the search space is vast and evaluation of each configuration can be expensive’.<sup>325</sup> Other aspects that involve decision-making include choosing training datasets, encoders, RNGs, etc. While simpler models tend to ‘underfit’, more complex ones tend to ‘overfit’.<sup>326</sup> As suitable models are usually located somewhere in the middle of those two extremes, researchers need to try different models and compare them through cross-validation to find an optimal model for a particular problem.

In GP/EAs, one almost always needs to develop specific encodings for a problem representation and design a unique fitness function (a decisive part) and particular operators (recombination and mutation) that can work with that encoding.<sup>327</sup> In the case of decision trees, configuration choices concern the depth of a tree, the number of trees, the size of each leaf, etc.

All relevant factors need to be considered when choosing, adjusting or designing an objective function. Fitness functions in GP or EAs are highly specific to the application, and their design usually requires a great deal of expertise. While boilerplate functions most commonly applied with ANNs exist, human influence manifests in the decision-making regarding the hyperparameters and data preparation. Where an existing objective function is used, it might need to be adjusted to a specific problem.

In sum, for a problem to be solved successfully through ML, one should have a good grasp of computational model optimisation, the state-of-the-art algorithms and the application domain, be able to understand and model a problem, design an objective function, specify the right success metric, select and prepare<sup>328</sup> input data relevant for the objective, and make sense of the results. As with any software, the designer might not need to build all functionalities and constitutive elements of ML from scratch, but the results achieved by individual users applying ML can vary significantly depending on the purpose and skills. The importance of interpreting ML output should not be downplayed. At a basic level, the immediate output of ML is coded in binary digits. At a higher level, it can be represented as yes/no, numbers, options, words, etc. In any case, such output cannot be equated with ready-made knowledge or problem solution.

<sup>325</sup> Ram Sagar, ‘What Are Hyperparameters and How Do They Determine a Model’s Performance’ (*AIM*, 22 February 2019) <<https://analyticsindiamag.com/what-are-hyperparameters-and-how-do-they-determine-a-models-performance/>> accessed 21 May 2021 (emphasis added).

<sup>326</sup> In ML, ‘overfitting’ means that a model ‘learned “too much” from the training data set’, while ‘underfitting’ means that a model ‘has “not learned enough” from the training data’; in both cases, models have low generalisability and generate unreliable predictions. Anas Al-Masri, ‘What Are Overfitting and Underfitting in Machine Learning?’ (*Towards Data Science*, 22 June 2019) <<https://towardsdatascience.com/what-are-overfitting-and-underfitting-in-machine-learning-a96b30864690>> accessed 21 May 2021. Over- and underfitting can occur with any method that attempts to derive a prediction based on the input data.

<sup>327</sup> As discussed earlier, preparatory steps in GP are problem-specific; see above (nn 126, 133) and the accompanying text.

<sup>328</sup> Data preparation involves correcting or removing wrong data, including outlier removal and dimensionality reduction, which requires expertise.

### (iii) ‘Average’ knowledge and skills in ML

ML methods are characterised by modularisation. Like a LEGO constructor, they involve multiple building blocks that can be configured in various ways depending on the objective. As with any engineering work, computational modelling requires intuition and insight to make the right decisions and choices, which comes with experience. Setting up the configurations mentioned above is a true indicator of knowledge, skills and expertise in ML. Given a large variety of ML techniques, it is hard to define in the abstract what level of skills in ML can generally be considered ‘average’. This problem is compounded even more when it comes to a team of researchers, given that collaboration is quite common in ML.

Whether an average practitioner could accomplish the ‘right’ decision-making can only be considered vis-à-vis a particular problem solved through ML. For instance, while AlphaGo was developed based on state-of-the-art ML methods, the combination of techniques and configurations determined AlphaGo’s success.<sup>329</sup> Would any person with average skills and knowledge in ML be able to come up with such configurations? In the authors’ view, if a Master’s degree in computer science is considered a measure of ‘average’ knowledge and skills in ML, this is unlikely to be enough to design and implement AlphaGo. In addition, non-human factors, such as access to data and hardware (i.e. computational resources to implement training in parallel within a reasonable time), should be taken into account when considering what can be achieved with ML techniques.

### c) Implications for the assessment of inventive step revisited

Given that there can be considerable room for decision-making in applying even known ML methods, an invention developed through ML cannot be considered a mere ‘bonus’ obtained in a ‘one-way-street’ situation.<sup>330</sup> Besides, it cannot be alleged across the board that ML can solve *any* problem with no or only trivial adjustment. Otherwise, we would not have incurable diseases, given that AI techniques have been applied in drug discovery and development for decades.<sup>331</sup>

Even though the implementation of an iterative search-based algorithm<sup>332</sup> can be viewed as routine trial and error, it would be erroneous to reduce the entire problem-solving to the implementation phase.<sup>333</sup> Rather, it should be considered in light of individual circumstances of a

<sup>329</sup> Steffen Hölldobler, Sibylle Möhle and Anna Tigonova, ‘Lessons Learned from AlphaGo’ in Steffen Hölldobler, Andrey Malikov and Christoph Wernhard (eds), *YSIP2 – Proceedings of the Second Young Scientist’s International Workshop on Trends in Information Processing* (CEUR-WS 2017) 92 <<http://ceur-ws.org/Vol-1837/>> accessed 21 May 2021.

<sup>330</sup> On this criterion, see above (n 314) and the accompanying text.

<sup>331</sup> See eg Matthew A Sellwood and others, ‘Artificial Intelligence in Drug Discovery’ (2018) 10(17) *Future Medicinal Chemistry* 2025, at 2025 (noting that AI has been used in different forms, including ML, and to a varying extent in the search for novel molecules and structure-activity correlations since the 1960s).

<sup>332</sup> Above (n 134).

<sup>333</sup> As discussed above at II.2.b.iii. See also Gary F Marcus and E Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (Pantheon Books 2019) 6–9 (discussing examples of overstating AI performance and suggesting that, whenever one hears about ‘a supposed success in AI’,

case – against the relevant prior art – whether an average skilled person would have come up with a *particular set-up* of an ML system that would lead to the claimed invention. The question of whether and how such decision-making can be ‘reverse-engineered’ from the final result and objectively assessed by patent examiners calls for a more detailed inquiry.

## 7. One ML algorithm – many inventions

### a) The anticipation of technological singularity

Some legal scholars foresee ‘creative singularity in which computers overtake human inventors as the primary source of new discoveries’<sup>334</sup> and assume that one AI algorithm can generate multiple inventions – which could be closely related or unconnected<sup>335</sup> – each time it is implemented.<sup>336</sup> The view that AI can yield a ‘surprisingly large number of inventions’<sup>337</sup> echoes the notion of technological singularity. Besides, patent infringement would practically be excluded if alternative technical solutions could be ‘invented around’ effortlessly. The assumed prolificacy of AI algorithms calls into question the need for behavioural and economic incentives for innovation and the role of the patent system in this regard.

As discussed earlier, numeric output in ML can vary due to randomisation when one algorithm is implemented multiple times.<sup>338</sup> However, what matters is whether such varying output can constitute distinct solutions *substitutable* in their quality. For instance, would multiple runs of the same evolutionary algorithm used by the NASA researchers<sup>339</sup> generate disparate space antenna designs with comparable fitness and robustness in satisfying technical constraints?

### b) A technical perspective on an algorithm’s prolificacy

#### (i) The variability of the output depends on a technique

The variability of the numeric output in ML depends on a particular method. In ANNs, the chances are high that, irrespective of the initial randomisation of weights, a good training setup would yield very similar models. Thus, a robust algorithm would generate a model that would perform with a comparable level of accuracy<sup>340</sup> even if, due to the randomisation, the weights optimising the cost function can be slightly different each time the training algorithm is applied to the same training data. For instance, the second AlphaGo may not make every move exactly

one should ask: ‘Stripping away the rhetoric, what did the AI system actually do here?’).

<sup>334</sup> Abbott (n 18) 1080.

<sup>335</sup> Vertinsky and Rice (n 31) 594.

<sup>336</sup> Peter M Kohlhepp, ‘When the Invention Is an Inventor: Revitalizing Patentable Subject Matter to Exclude Unpredictable Processes’ (2008) 93 Minn. L. Rev. 779, 812; William Samore, ‘Artificial Intelligence and Patent System: Can a new tool render a once patentable idea obvious?’ in Woodrow Barfield and Ugo Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar Publishing 2018) 471, 479.

<sup>337</sup> Yanisky-Ravid and Liu (n 32) 2230.

<sup>338</sup> As discussed above at II.3.b.iv and vi.

<sup>339</sup> Lohn, Hornby and Linden (n 176).

<sup>340</sup> The key determinants of the accuracy of a prediction generated by a trained ML-model are a good and large enough input dataset (the more amount of cases, usually the better, except for outliers, which have to be removed from the input), the design of a model with balanced complexity (it cannot be too complex or too simple), and a selection of good features (features that are most capable of predicting the outcomes).

the same way as the first one, but the overall performance would most likely be the same.

In cases where GP/EAs are applied in technical design and engineering, the output variability would depend on different factors, including the number of iterations run by an algorithm, how much time is given to the algorithm to find the optimum, and the mutation rate. Where the RNG’s influence is enough to escape local optima,<sup>341</sup> multiple executions of the same EA would generate the same results. Alternatively, if the influence of the RNG is insufficient or the search time is limited, multiple executions of an algorithm can generate different solutions. The two main reasons why an algorithm might generate different solutions are that the seed makes the algorithm ‘get stuck’ in local optima, or the time of algorithm execution is insufficient to reach the optimum.

In research papers, authors usually report the averages/median and a corresponding spread measure (standard deviation/interquartile range). Since the same algorithm is typically implemented many times with different seeds, researchers draw several samples from the algorithm’s performance distribution. The spread of that distribution is measured as the standard deviation of that sample. The smaller the spread, the less sensitive an algorithm is to the choice of the random numbers, and the more robust an algorithm is. Where randomisation is applied to various aspects of ML, it cannot be guaranteed that a model would perform exactly like the one in the previous independent run of an algorithm, but the average performance is a good indicator of what results can be expected.

#### (ii) The output substitutability

To answer whether the same genetic algorithm can generate multiple distinct solutions with comparable fitness, we first need to define the terms ‘global’ and ‘local’ optima in optimisation problems. Two essential components of an optimisation problem are a search space of potential solutions and an objective function that evaluates the quality of each candidate solution.<sup>342</sup> Local optima are the extrema<sup>343</sup> that minimise or maximise the objective function for a given region of the search space, while a global optimum is the extrema (minimum or maximum) of the objective function for the entire search space.<sup>344</sup> *Multiple global and local optima* can exist in a given search space,<sup>345</sup> depending on how the search space and an objective function are defined. Whether a local optimum can be a satisfactory solution can hinge on the individual needs in a particular case – in many cases, the ‘near-best solution’ can suffice.

<sup>341</sup> On local and global optima, see below (nn 343-345).

<sup>342</sup> Rudolf Kruse and others, ‘Introduction to Evolutionary Algorithms’ in Rudolf Kruse and others (eds), *Computational Intelligence* (Springer 2016) 189.

<sup>343</sup> The term ‘extrema’ jointly refers to the maxima (the largest value) and the minima (the smallest value) of a function.

<sup>344</sup> Jason Brownlee, ‘Local Optimization Versus Global Optimization’ (*Machine Learning Mastery*, 29 January 2021) <<https://machinelearningmastery.com/local-optimization-versus-global-optimization/>> accessed 21 May 2021. See also Ragnar AK Frisch, *Maxima and Minima: Theory and Economic Applications* (Springer 2013) 5 (explaining that, with the global extrema, the comparison is established with all points in a given domain; with the local extrema, the comparison is confined within a chosen ‘locality’).

<sup>345</sup> *ibid.*



As noted earlier, in ANNs, different initial weights can result in slightly different models, but the chances are high that such models will perform with a comparable level of accuracy. In GP and EAs, different initial conditions can lead to different local or global optima. As far as the NASA antenna is concerned, the researchers did not perform a statistical sampling across multiple runs of the algorithms, as the goal was to design the antennas that would comply with the specifications of the NASA mission.<sup>346</sup> Theoretically, two algorithms applied by the NASA researchers could have identified more solutions comparable in terms of technical parameters optimisation. Such possibility would depend on the used algorithm, configurations and the modelling of the fitness functions. However, even if more solutions were identified, one could not say that an algorithm ‘invented’ multiple inventions. What happens is that an algorithm finds through the iterative search the optima that *already exist* within a search space.

### c) Implications for patent law

Even though some optimisation techniques can identify several optima within a search space, the assumption that an algorithm can yield a ‘large number of inventions’ appears to be an exaggeration. Quite to the contrary, in some cases – for instance, in the field of molecular design – ML models can produce ‘to a large degree invalid’ output.<sup>347</sup>

Computational model-based techniques can make problem-solving more efficient compared with brute force computation, random screening or blunt trial and error. However, this does not necessarily mean that the application of ML techniques might have a sizable cost-saving effect.<sup>348</sup> To date, the relationship between patents, AI and innovation remain underexplored. At the same time, without a detailed understanding of ML-based business models, one cannot argue that the investment amortisation function of patent law becomes redundant.

## III. Synthesis and research outlook

### 1. ML as computational techniques of problem-solving

#### a) ML systems are, technically, tools

The question of whether AI in its most advanced forms is a tool or ‘more than a tool’ obviously depends on the definition of a ‘tool’. Etymologically and in common parlance, tools are means that assist in performing an activity.<sup>349</sup> All tools are considered to share one attribute – they require human guidance.<sup>350</sup> As long as AI systems are not self-organising systems capable of performing automated tasks without pre-programmed instructions or

disobeying such instructions, they are, *technically, tools*. More precisely, techniques commonly referred to as AI, are computational model-based methods that can be applied in technical optimisation use cases.

Importantly, AI systems can replace humans in implementing computation not because it is unfeasible for a human to perform the same computational operations but because it is rather impractical. Humans are slower in computational tasks – some computations may take years for a human, while a computer can perform them in seconds. Automation is inevitable. Just as it is more practical to use cranes and excavators to move stones and build constructions, today we use search engines because it is more practical than gathering physical copies and searching them manually.<sup>351</sup> However, when computers are used to perform search and data processing tasks, it is the *implementation of computation* that is automated, not human thinking, cognition or problem-solving. Even where certain problem-solving segments – such as search, optimisation and modelling – can be automated because they can be expressed as computational tasks, their implementation on a computer *should not be equated with the automation of problem-solving* as such.

#### b) Looking beyond the hype

Popular depictions of AI are often saturated with hyped perceptions and might vest in AI systems qualities and capabilities that still need to be proved by research in AI. For instance, in the case of AI translation, some assume that GPT3 can write poetry; others argue that ‘despite the teasing of mainstream press headlines to the contrary, GPT-3 doesn’t signal the beginning of the end for humanity’.<sup>352</sup> Even though machine translation has undergone significant developments, it has still not become a mundane task. Even seemingly factual information, such as sports or stock market reports, might contain multi-word expressions<sup>353</sup> and figurative language<sup>354</sup> – i.e. rhetorical and stylistic markers – that the best NLP and machine translation systems currently available cannot always deal with.<sup>355</sup>

The perception of computers’ superiority in problem-solving might be prompted by the widely publicised victories of AI systems in games, which, according to some commentators, ‘left humans in its dust’<sup>356</sup> and

<sup>346</sup> Lohn, Hornby and Linden (n 176) 309.

<sup>347</sup> Dimitrov and others (n 40) 24829.

<sup>348</sup> As noted earlier, ‘transfer learning’ is an approach that can save costs in ML. Whether or not transfer learning can be implemented in a use-case depends on an ML method and a particular application.

<sup>349</sup> Cambridge Dictionary, ‘Tool’ <<https://dictionary.cambridge.org/dictionary/english/tool>> accessed 21 May 2021. For the etymological origins, see <<https://www.etymonline.com/word/tool>> accessed 21 May 2021.

<sup>350</sup> Carl Mitcham and Robert Mackey, ‘Tools and Machines’ in Carl Mitcham (ed), *Encyclopedia of Science, Technology, and Ethics* (Thomson Gale 2005) 1964, 1965.

<sup>351</sup> The definition of automation is not limited to physical tasks but includes tasks associated with cognitive processes. George A Schillinger, ‘Automation’ in Carl Mitcham (ed), *Encyclopedia of Science, Technology, and Ethics* (Macmillan Reference 2005) 146.

<sup>352</sup> Robert Dale, ‘GPT-3: What’s It Good for?’ (2021) 27 *Natural Language Engineering* 113, 118, doi:10.1017/S1351324920000601.

<sup>353</sup> Ivan A Sag and others, ‘Multiword Expressions: A Pain in the Neck for NLP’ in Alexander Gelbukh (ed), *Computational Linguistics and Intelligent Text Processing*, CILing 2002, Lecture Notes in Computer Science, vol 2276 (Springer 2002).

<sup>354</sup> Jelena Mitrović and others, ‘Ontological Representations of Rhetorical Figures for Argument Mining’ (2017) 8(3) *Argument & Computation* 267.

<sup>355</sup> Dale (n 352).

<sup>356</sup> Matt McFarland, ‘What AlphaGo’s Sly Move Says About Machine Creativity’ *The Washington Post* (Washington, 15 March 2016) <<https://www.washingtonpost.com/news/innovations/wp/2016/03/15/what-alphagos-sly-move-says-about-machine-creativity/>> accessed 21 May 2021. Bernard Marr, ‘Artificial Intelligence Masters the Game of Poker – What Does That Mean for Humans?’ (*Forbes*, 13 September 2019) <<https://www.forbes.com/sites/bernardmarr/2019/09/13/artificial-intelligence-masters-the-game-of-poker-what-does-that-mean-for-humans/>> accessed 21 May 2021.

'demonstrate the futility of competition between human-kind and machine'.<sup>357</sup> Others, however, have not viewed the capacity of Deep Blue to beat a human at chess as 'more portentous than that a tractor can beat a human at ploughing'<sup>358</sup> and consider AlphaGo's accomplishments as illustrating Level 1 of AI where 'engineered intelligence' can achieve 'technically defined goals' but is not capable of human-like reasoning.<sup>359</sup>

What the victory of AlphaGo can tell us about human intelligence is that a computer can efficiently perform calculations that would take a lifetime for a human without a computer. It also tells us that a human can be intelligent enough to model those processes: in essence a group of humans came up with a way to train AlphaGo by using certain rules and methods. However, it does not tell us what human intelligence is and how it is developed, e.g. whether the human brain computes some loss function and builds a numeric model when learning to play chess. One cannot deduce superiority because a computer can outperform a human in playing a specific game.

### c) Comparing 'apples to apples'

AI can augment human capacity to solve problems rather than replace a human. Even though races between humans and horses<sup>360</sup> can be entertaining, the two are obviously incompatible, while a horse's outperformance cannot be seriously taken as a sign of the apocalypse of the human race. If we wanted to compare 'apples to apples', the competition would need to be arranged between ML teams developing systems playing chess, Go, etc. The same can be said about the 'Humies' competition held annually as part of the Genetic and Evolutionary Computation Conference.<sup>361</sup> The concept 'human-competitive performance' is deceptive, as long as a human orchestrates the implementation of GP/EAs through task-specific preparation. It would be more sensible to compare the results achieved by humans using GP vis-à-vis those achieved by humans not applying GP or applying other techniques.

### d) A matter of conceptual perspective?

The question of whether AI might be 'more than a tool' pertains to the central inquiry in the philosophy of AI and cognitive science, namely: where algorithms can express mental states and processes, do computers executing such algorithms merely *simulate* mental states and processes, or do they *exhibit* them?<sup>362</sup> Research in the field of human-computer interaction and cognitive science produced several approaches illustrating how computer-mediated cognitive processes can be conceptualised.

These include the theories of 'extended mind'<sup>363</sup> and 'distributed cognition',<sup>364</sup> and concepts such as 'blended cognition',<sup>365</sup> 'cognitive systems',<sup>366</sup> 'cognitive artifacts',<sup>367</sup> 'instruments of mind',<sup>368</sup> 'things that make us smart',<sup>369</sup> and 'tools for thought'.<sup>370</sup> While an in-depth discussion of these frameworks goes beyond the scope of this paper, the diversity of these perspectives suggests that the questions of whether and where to draw a boundary between cognitive agents and cognitive artefacts depends on the frame of reference and can even be viewed as a matter of 'one's belief system'.<sup>371</sup>

What is clear is that AI systems are based on computational modelling, which is by definition an approximation and simplification of the phenomena that are being modelled (e.g. cognitive functions and processes).<sup>372</sup> Research in AI is concerned with a 'computational understanding of what is commonly called intelligent behavior, and with the creation of artefacts that exhibit such behavior'.<sup>373</sup> Thus, cognitive functions can be reproduced in computational artefacts to the extent to which they are comprehended by their (human) designers and engineers.<sup>374</sup> Currently, there are significant gaps in our understanding of human cognition, including questions such as 'what is the nature of consciousness', 'what is the neurological basis of creativity',<sup>375</sup> 'how exactly biological processes interact with cognitive phenomena',<sup>376</sup> and 'whether cognitive processes are necessarily computational in nature'.<sup>377</sup> Success in computational modelling of cognitive processes will depend on bridging those gaps.

<sup>363</sup> Heinrichs (n 158).

<sup>364</sup> Distributed cognition is an analytical framework that emerged in the mid-1990s. It provides concepts, methods, approaches to understanding human cognitive activity which is conceptualised in computational and information processing terms. Interested readers might take a look at the following literature: Edwin Hutchins, *Cognition in the Wild* (MIT Press 1995); Yvonne Rogers and Judi Ellis, 'Distributed Cognition: An alternative framework for analysing and explaining collaborative working' (1994) 9 *Journal of Information Technology* 119.

<sup>365</sup> Jordi Vallverdú and Vincent C Müller (eds), *Blended Cognition. The Robotic Challenge* (Springer 2019) vii (defining blended cognition as 'the study of how an intelligent system can use or even partially combine several methods to decide among possible action outputs or data evaluation and storage').

<sup>366</sup> David Vernon, *Artificial Cognitive Systems. A Primer* (MIT Press 2014).

<sup>367</sup> Don A Norman, *Cognitive Artifacts* (University of California 1990).

<sup>368</sup> Meurig Beynon and others (eds), *Cognitive Technology: Instruments of Mind: 4th International Conference* (1st edn, Springer 2001).

<sup>369</sup> Don A Norman, *Things That Make Us Smart. Defending Human Attributes in the Age of the Machine* (Basic Books 1993).

<sup>370</sup> Howard Rheingold, *Tools for Thought: The History and Future of Mind-expanding Technology* (Princeton University Press 2000).

<sup>371</sup> Richard D Patton and Peter C Patton, 'What Can Be Automated? What Cannot Be Automated?' in Shimon Y Nof (ed), *Springer Handbook of Automation* (Springer 2009) 305, 305-6.

<sup>372</sup> Enric Trillas, 'An Algebraic Model of Reasoning to Support Zadeh's CWW' in Janusz Kacprzyk and Witold Pedrycz (eds), *Springer Handbook of Computational Intelligence* (Springer 2015) 249, 251 (stating that, '[i]f it can metaphorically be said that if reality is in color, a model of this reality is a simplification of it in black and white').

<sup>373</sup> Stuart C Shapiro, *Encyclopedia of Artificial Intelligence* (2nd edn, John Wiley & Sons 1992) 54.

<sup>374</sup> For an in-depth exploration of how a better understanding of human processes can be leveraged for designing better computational artefacts, see eg Jordi Vallverdú and Vincent C Müller (eds), *Blended Cognition. The Robotic Challenge* (Springer 2019).

<sup>375</sup> See eg Anna Abraham, *The Neuroscience of Creativity* (Cambridge University Press 2018).

<sup>376</sup> Gregory R Mulhauser, *Mind Out of Matter: Topics in the Physical Foundations of Consciousness and Cognition* (Springer 1998).

<sup>377</sup> Tim Van Gelder, 'What Might Cognition Be, If Not Computation?' (1995) 92(7) *The Journal of Philosophy* 345-381.

<sup>357</sup> Aswin Pranam, 'Why the Retirement of Lee Se-Dol, Former 'Go' Champion, Is a Sign of Things to Come' (*Forbes*, 29 November 2019) <<https://www.forbes.com/sites/aswinpranam/2019/11/29/why-the-retirement-of-lee-se-dol-former-go-champion-is-a-sign-of-things-to-come/>> accessed 21 May 2021.

<sup>358</sup> Harry Collins, *Artificial Intelligence. Against Humanity's Surrender to Computers* (Polity Press 2018) 84-85.

<sup>359</sup> *ibid* 77.

<sup>360</sup> <<http://managainsthorse.net>> accessed 21 May 2021.

<sup>361</sup> <<https://www.human-competitive.org/awards>> accessed 21 May 2021.

<sup>362</sup> Barbara Von Eckardt, 'Cognitive Science' in *Encyclopedia of Computer Science* (John Wiley & Sons 2003), doi: 10.1002/0470018860.s00105.

In general, views on when ‘high-level machine intelligence’ (‘Strong AI’) can be achieved diverge substantially within the ML community.<sup>378</sup> Some researchers assume that ‘all human thought, or at least intelligent thought, can be reduced to computation, and [hence, computers can] exhibit human-like intelligence, and eventually intelligence even superior to that of humans’.<sup>379</sup> Others argue that ‘not all intelligent human thought and its consequent behavior can be reduced to simple computation, or even logic’.<sup>380</sup>

As far as the application of ML in technical problem-solving is concerned, a 2003 article on technical design emphasised that ‘computers are used only as an *aid*’ in technical design and engineering, which ‘involves extensive decision-making and subjective evaluation, activities that are aided greatly by using computers, but that are generally carried out by computers under human direction’.<sup>381</sup> In the authors’ view, this holds good for contemporary ML systems, while predicting future AI developments is akin to gazing into a crystal ball.

## 2. Implications for patent law

Whether AI is a tool or more than a tool can be a relevant factor for the definition of inventor and the allocation of the inventor’s rights from a deontological perspective. In light of the technical explanations provided in this paper, hardly any uncertainty persists in situations where ML techniques are applied in problem-solving either regarding the fulfilment of the requirement of intellectual engagement in the conception of an invention *de lege lata*,<sup>382</sup> or regarding the appropriateness of such requirement *de lege ferenda*. Like a sword that never kills by itself but is a tool in the killer’s hand,<sup>383</sup> computational modelling and computers executing models do not invent by themselves but are powerful problem-solving tools.<sup>384</sup> As emphasised throughout this

paper, the successful performance of ML is a function of knowledge, skills and expertise of humans who configure, modulate and apply ML techniques.

As far as the economic justification of patents is concerned, it is not *prima facie* evident that the application of ML in technical problem-solving renders innovation activity so effortless and low-cost that the investment amortisation function of patent rights becomes no longer relevant.<sup>385</sup> More economic insight into the role of patent protection for AI-incorporating or AI-induced inventions is needed as statistical data showing the increasing<sup>386</sup> number of AI-related patents alone do not lend themselves to drawing any conclusion on the causal relationship between patents and innovation.

On balance, implications of ML for the patent system in its core tenets, on closer examination, appear less revolutionary<sup>387</sup> than is often posited. Some of the identified issues – in particular, how ML techniques should be factored in the definition of a skilled person, the patentability of purpose-unbound ML methods and, more broadly, the role of AI-related patents in innovation<sup>388</sup> – need to be examined more in-depth. While this paper mainly drew on European patent law and the patent examination practice of the European Patent Office, the presented analysis concerns the core aspects of the patent system and, thus, our findings can inform policy and legal discussions on patent law and AI in other jurisdictions.

### ACKNOWLEDGEMENTS

The authors thank Preston Richard for his review and valuable comments on the draft. The feedback of two anonymous reviewers and the editorial assistance of GRUR International are much appreciated. Any errors in the legal analysis remain solely those of Daria Kim.

<sup>383</sup> A saying attributed to Lucius Annaeus Seneca.

<sup>384</sup> Dietmar PF Moeller, *Mathematical and Computational Modeling and Simulation. Fundamentals and Case Studies* (Springer 2004) 1 (referring to computational modelling as ‘one of the most powerful design tools’ used to ‘understand, control and improve the performance of complex systems and to solve problems’).

<sup>385</sup> It is worth noting that the amount of investment incurred in developing an invention has not been a material factor for patentability *de lege lata*. Thus, the relevance of the investment factor *de lege ferenda* would need to be considered across the board and not only in relation to inventions developed by applying AI.

<sup>386</sup> See eg WIPO (n 275); OECD, *Measuring the Digital Transformation. A Roadmap for the Future* (OECD Publishing 2019) 32 ff; Hidemichi Fujii and Shunsuke Managi, ‘Trends and Priority Shifts in Artificial Intelligence Technology Invention: A global patent analysis’ (2018) 58 *Economic Analysis and Policy* 60.

<sup>387</sup> This conclusion is consistent with the findings of Professor Straus. See Joseph Straus, ‘Will Artificial Intelligence Change Some Patent Law Paradigms?’ (2021) LXXXI *Zbornik znanstvenih razprav* 11 <<https://www.pf.uni-lj.si/media/zsr.2021.02.straus.pdf>> accessed 30 December 2021.

<sup>388</sup> As pointed out above, studies on this subject are rare. See Nathan Calvin and Jade Leung, ‘Who Owns Artificial Intelligence? A preliminary analysis of corporate intellectual property strategies and why they matter’ (2020) Working Paper 8 <[https://www.fhi.ox.ac.uk/wp-content/uploads/Patents\\_FHI-Working-Paper-Final-.pdf](https://www.fhi.ox.ac.uk/wp-content/uploads/Patents_FHI-Working-Paper-Final-.pdf)> accessed 25 May 2021 (observing that ‘AI and ML commercial activity has experienced massive growth and international investment even while the patentability of innovations remains uncertain, suggesting that the ability to patent AI is not necessary for innovation’).

<sup>378</sup> See eg Katja Grace and others, ‘Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts’ (2018) 62 *Journal of Artificial Intelligence Research* 729-754 <<https://www.jair.org/index.php/jair/article/view/11222/26431>> accessed 21 May 2021. The authors define ‘high-level machine intelligence’ as the capacity of unaided computers to accomplish any task better and more efficiently compared to qualified humans.

<sup>379</sup> Patton and Patton (n 371) 312.

<sup>380</sup> *ibid.* For a recent insightful discussion, see Michael I Jordan, ‘Artificial Intelligence — The Revolution Hasn’t Happened Yet’ (2019) *Harvard Data Science Review* <<https://hdsr.mitpress.mit.edu/pub/wot7mkc1>> accessed 28 December 2021 (followed by peer commentary).

<sup>381</sup> Barry Flachsbart, David Shuey and George Peters, ‘Computer-Aided Design/Computer-Aided Manufacturing’ in Edwin D Reilly, Anthony Ralston and David Hemmendinger (eds), *Encyclopedia of Computer Science* (4th edn, John Wiley & Sons 2003) 268, 270 <<https://dl.acm.org/doi/10.5555/1074100.1074238>> accessed 30 December 2021 (emphasised in the original).

<sup>382</sup> On the criterion of intelligent conception, see above (nn 73-75) and the accompanying text.