https://doi.org/10.1038/s42256-022-00582-5

Can AI invent?

Alexandra George & Toby Walsh

Check for updates

Artificial intelligence systems are used for an increasing range of intellectual tasks, but can they invent, or will they be able to do so soon? A recent series of patent applications for two inventions that are claimed to have been made by an artificial intelligence program are bringing these questions to the fore.

Central to humanity's ability to improve the quality of our lives has been our ability to innovate. In the past, such innovation has come from human sweat and ingenuity. But will we soon have machines that can invent, and will this speed up innovation?

Some recent legal cases are putting a spotlight on these questions. The cases revolve around a neural-network-based artificial intelligence (AI) system called DABUS (Device for Autonomous Bootstrapping of Unified Sentience), which its creators claim makes inventions worthy of patenting. Patent applications have been lodged around the world for two inventions for which DABUS is named as the sole inventor: a food container with a fractal surface to aid packing, and a warning light with a pulse train of fractal dimension to attract attention. These patent applications have so far been rejected in almost every jurisdiction[1], mostly on the legal grounds that an inventor should be a human being.

None of the legal cases has tested the claim that DABUS was the sole inventor. However, the patent applications have raised interesting and important questions about what it means to be an inventor, and thus what it means to invent.

In the context of patent law, the product or process needs to be "novel" (not found in the 'prior art base' of existing inventions) and "inventive" (not obvious to a notional person skilled in the relevant art), and it must have industrial application.

The question can therefore be asked whether AI-systems are, or will be, able to invent. We use here a broad definition of AI systems that encompasses rule-based systems, search-based systems and learning-based systems.

The DABUS patent

The DABUS case offers a good case study. The patent application forms have listed the inventor as "DABUS, The invention was autonomously generated by an artificial intelligence." In considering the strength of this claim of inventorship, we run into three stumbling blocks.

The first is that machine learning programs such as DABUS require significant expertise to set up. One major task is modelling the machine learning problem. How do we represent the inputs and outputs from which the program will learn? With DABUS, a human programmer, Stephen Thaler, provided a relatively small number of atomic concepts such as "container," "surface" and "fractal" that the system then glued together in a novel way. Perhaps the human programmer was inspired to include the concept "fractal" by the success of fractal surfaces in other settings, such as fractal antennae and fractal heat exchangers. In any case, this human modelling was critical to the inventive process and would suggest that DABUS was not an autonomous inventor.

The second stumbling block is that DABUS uses a form of supervised learning in which a human (whom Thaler dubs a 'mentor') provides guidance. The mentor identifies promising concepts produced by DABUS to be explored further, and cuts off less promising ones. According to Thaler: "The fractal container was ... formed largely through generative learning involving reward signalling by a mentor for concepts having significance to humans" (ref. [2]). This human mentoring was critical to the invention. There is a vast array of concepts that could be explored, and human wisdom was crucial in guiding the program to explore just a small part of this space. This would again suggest DABUS was not the sole inventor.

The third stumbling block is that DABUS produces a graphical or relational output of concepts that requires the expertise of a human user to interpret. For example, the output produced for the fractal container invention was, "food drink in fractal bottle increase surface area making faster heat transfer for warming cooling convenience pleasure"[2]. Thaler wrote:

> Important to note is that both the graph and pidgin language expressions of these inventions became readily understandable by a human mentor through accumulated experience with the DABUS system. In other words, person and machine became familiar with each other through cumulative mentoring, allowing the human partner to make sense of the developing concepts, as well as the system's idiosyncrasies.

Thus, to run DABUS, a human has to model the problem domain in which invention will take place, carefully steer the program to produce inventive output and then interpret the inventiveness of the output. From a technical perspective, it is therefore unconvincing to claim (as the DABUS patent applications do) that the invention was autonomously generated by an artificial intelligence system. It would perhaps be more accurate to suggest that DABUS was a tool used by the inventor. At a stretch, it might be concluded that DABUS and Thaler jointly invented the objects.

The three concerns we raise here about DABUS also apply to other AI systems, and not even just to supervised learning systems. For instance, the need for interpretation of output applies to any AI system. Similarly, identifying promising directions to explore applies just as much to search-based systems as it does to learning-based systems. We therefore ask whether any AI system can ever have been said to have created a patentable invention.

AI invention and surprise

To address this question requires consideration of the long and distinguished history that has brought us to the current state of AI systems. In 1843, Ada Lovelace wrote[3]:

> The Analytical Engine [arguably the first general-purpose programmable computer] has no pretensions to originate anything.

Fig. 1 | Pictures generated to illustrate the models. Images for the inventions "liquid container with fractal wall" and "glove with a fractal gripping pattern," made with the help of the text-to-image generative AI model Stable Diffusion (https://huggingface.co/CompVis/stable-diffusion), as described in Box 1. Images created through Stable Diffusion Online and reproduced under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license.

> It can do what-ever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths. Its province is to assist us to making available what we are already acquainted with.

This objection has haunted the field of AI from its very inception. Alan Turing attempted to refute this assertion in what is generally considered to be the first scientific paper about AI[4]. In 1950, Turing wrote:

> Who can be certain that 'original work' that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles. A better variant of the objection says that a machine can never 'take us by surprise'. This statement is a more direct challenge and can be met directly. Machines take me by surprise with great frequency.

Although machines might surprise humans, does this necessarily mean that a machine has autonomously invented? Colloquially, such surprise might suggest that an AI system has indeed invented something by autonomously contributing to an outcome that had not been foreseen by its human mentor. However, perhaps this reflects a lack of curiosity or inventive thought on the part of the mentor. From a patent perspective, such subjective surprise would not necessarily meet the threshold 'inventive step' requirement of being non-obvious to the notional person skilled in the art. That hypothetical person is equipped with the common general knowledge of a typical person skilled in the art, and has ordinary levels of creativity in inventiveness.

Abstract inquiries arguably leave us none the wiser as to whether an AI system could be said to have invented autonomously. As another approach to tackle this question, it may be instructive to review of the history of real-life AI inventions.

A brief history of AI inventions

Since Turing's attempt to refute Lovelace's objection to machine creativity, there have been many attempts to build programs that invent, so as to refute Lovelace with direct practical evidence[5]. We provide a brief and selective timeline of such AI inventions that illustrates the rich body of research in this area. We focus on inventions in technical domains, leaving aside the considerable body of research using AI to come up with original creations in domains such as music and painting (which would fall within the scope of copyright law's 'authorship' concept, rather than patent law's inventiveness).

As mentioned earlier, AI provides a collection of different tools and technologies ranging from rule-based systems in which knowledge is hand coded, through systems such as genetic algorithms in which solutions are found through search, to neural networks where knowledge is learnt from data. In many of these different subdisciplines of AI, we can see examples of systems that have been used to help invent.

In rule-based systems, one of the first AI systems that deserves consideration is Douglas Lenat's EURISKO system. This was applied to a number of domains, including very large-scale integration (VLSI) design[6]. EURISKO — named after a Greek word meaning 'I discover' — invented several novel, 'high-rise' three-dimensional circuits that were later fabricated. An example was a combined NAND and OR gate that could be effectively packed together. A provisional US patent application for such a circuit was filed in 1980 (SN/144,960) but the application was abandoned in 1984 for reasons that are not public[7].

Moving on to genetic algorithms: one of the first successes was the use of genetic programming in 1997 to evolve the design of an analogue circuit for a low-distortion, low-bias amplifier[8]. Subsequently, Koza et al. used genetic programming to evolve 15 previously patented electronic circuits[9]. In 2002, a patent application was filed for several improved proportional–integral–derivative (PID) and non-PID controllers that were discovered using genetic programming[10]. The patent was granted in 2005 but lists only Koza and colleagues as inventors. In 2003, genetic programming was used to evolve the design of an unusual antenna shaped like an unwound paper clip. This was then flown on NASA's Space Technology 5 (ST5) spacecraft[11] in 2006, and it was likely the first AI invention in space.

Finally, considering neural networks: Stephen Thaler filed a patent application (US 5659666) in 1994 for the Imagination Engine, a neural network for stimulating creativity. In a later patent application, he extended this to the Creativity Machine (US 7454388B2). Thaler used this system in the invention of the cross-bristle design for the Oral-B CrossAction Toothbrush, launched in 1998. This was likely the first consumer product invented with the aid of AI.

More recently, Massachusetts Institute of Technology researchers using a deep neural network identified a powerful new antibiotic compound called halicin[12]. The molecule had previously been investigated as a potential drug to treat diabetes. It appears to kill many treatment-resistant bacteria by a novel process of disrupting the flow of protons across cell membranes. Although AI has previously been used in the discovery of new antibiotics, this was claimed to be the first time AI has identified completely new kinds of antibiotics from scratch and without any background human expert knowledge[13]. MIT has filed a patent application (PCT/US2020/049830) both for the machine learning method used to discover halicin and for halicin itself along with 15 other compounds with antimicrobial properties (though, notably, it lists five human inventors and does not list the AI system as an additional inventor). Multiple companies using AI-based strategies for drug discovery and development have been set up in the last decade with billions of dollars in funding.

We make a number of observations about this brief history. First, AI has been used as a tool to aid human inventiveness in a wide variety of domains. Second, many different AI approaches have been used to help invent. Third, the amount of human input has varied widely in these different applications. There is a spectrum of autonomy ranging from significant to very little human involvement. Fourth, AI has been used to invent for four decades and, mirroring the evolution of AI as a whole, there appears to have been an increase in the use of AI for invention in recent years. Fifth, in some emerging application areas, such as drug design, AI shows considerable promise. In silico predictions can be made at much greater speed than in vitro predictions. Because developing new drugs takes years of time and billions of dollars, anything that can speed up the discovery of new drugs and bring this cost down is very welcome. However, observing and appreciating the contribution made by AI systems to human inventiveness is different from claiming that an AI system has autonomously invented.

BOX 1

An experiment using a large language model to generate a patent summary

To demonstrate the ability of large AI language models to generate text similar to that used in patent applications, we gave extracts of Stephen Thaler's patent application describing the fractal container (WO2020079499) to Jurassic, a 178-billion-parameter large language model from AI21 Labs[17]. This is slightly bigger than the 175-billion-parameter GPT-3 model. Like GPT-3, the model was trained on the Common Crawl dataset, which includes numerous patents: indeed, the largest website included in the Common Crawl is patents.google.com. The Jurassic language model is given some sample input text and generates output text in a similar style. We ran it 25 times and report here the most promising output.

Input: The first invention relates to a food container suitable for both liquid and solid food products.

The invention is particularly suitable for, but not limited to, containers for liquids, such as beverages, and other flowable products. The present invention provides a food or beverage container having a container wall of fractal form. The fractal profile of the wall permits coupling by inter engagement of a plurality of said containers together. This feature can provide a number of practical advantages, including being able to do away with separate and additional tie elements to hold together a plurality of containers. An increased surface area can assist in the transfer of heat into and out of the container, for example for heating or cooling the contents thereof.

The second invention relates to

Output: PVC, latex or silicone rubber gloves, especially disposable gloves. The invention provides a glove having a flexible gripping portion formed from a fractal pattern. The gripping portion is formed from a continuous fractal pattern. The flexible gripping portion is sufficiently strong and rigid to perform its intended function.

A search of the USPTO patent database identified no patent applications submitted since 1976 that contain "glove" and "fractal" in their abstract or claims. It is possible, then, that a sufficiently novel and inventive glove with a flexible fractal gripping pattern could be patented.

How might AI invent?

The abstract idea behind many AI programs that invent is usually relatively simple: a human delineates a design space of concepts, and the program then explores this space. The space is typically very large, even infinite. Considerable effort therefore needs to be invested in identifying whether a part of the design space reached is worth exploring further, as well as in assessing the promise of any new concept.

DABUS outsourced these two tasks to a human. Could a sufficiently sophisticated AI system make such decisions itself? A related question is whether the writing of a patent application might one day be performed by a machine. Recent advances in large language models such as GPT-3 have produced surprising and impressive results[14]. To explore this idea, we show how such AI models can already write relatively convincing text summarizing a new invention (Box 1). In addition, we show that a modern AI text-to-image system can generate images to illustrate the invention (Fig. 1).

Although it appears that the AI was probably not the sole inventor in the DABUS case, we must entertain the idea that an AI might one day come up with an invention that could be patented with little human input. Ultimately, the day might arrive when patent registry offices would be flooded with patent applications for AI-generated inventions. To combat this, perhaps we will eventually even see some AI systems taking the place of, or assisting, human patent examiners.

Conclusions

Our analysis suggests that, to date, there has been significant human input in devising the objects claimed to have been invented by AI system DABUS. This puts into doubt the claims that these inventions could be said to have been autonomously generated by an artificial intelligence system. Nevertheless, there is plenty of evidence suggesting that AI systems are increasingly being used to help make inventive steps.

Just as AI is transforming other aspects of our lives, it is likely to transform how humans invent. We need to give careful thought to how the innovation system adapts to these changes. For example, to reduce the risks of a chilling effect and patent trolling, the criteria for patentability – and thus the definition of a patentable invention – may need to be modified to prevent the system from becoming clogged with minimally creative innovations identified by AI systems. Alternatively, it may be necessary to design a bespoke new form of intellectual property to incentivize and protect inventions made by AI systems[15]. It would also be interesting to consider in more depth the different levels and types of autonomy present in such AI systems, and whether these justify targeted legal responses[16].

Alexandra George [1] & **Toby Walsh** [2,3]

[1]Faculty of Law & Justice, UNSW Sydney, Kensington, New South Wales, Australia. [2]School of Computer Science & Engineering, UNSW Sydney, Kensington, New South Wales, Australia. [3]CSIRO Data61, Eveleigh, New South Wales, Australia.

e-mail: a.george@unsw.edu.au; tw@cse.unsw.edu.au

Published online: 07 December 2022

References

1. Vincent, J. *AI systems can't patent inventions, US federal circuit court confirms* (2022); https://www.theverge.com/2022/8/8/23293353/ai-patent-legal-status-us-federal-circuit-court-rules-thaler-dabus
2. Thaler, S. L. *J. Artif. Intell. Conscious.* **8**, 81–111 (2021).
3. Lovelace, A. in *Scientific Memoirs, Selected from the Transaction of Foreign Academies of Science and Learned Societies and from Foreign Journals* vol. 3 (ed. Taylor, R.), 691–751 (1843).
4. Turing, A. M. *Mind* **59**, 433–460 (1950).
5. Boden, M. *The Creative Mind: Myths and Mechanisms* 2nd edn (Taylor and Francis, 2004).
6. Lenat, D. *Artif. Intell.* **21**, 61–98 (1983).
7. Abbott, R. in *Cambridge Handbook of the Law of Algorithms* (ed. Barfield, W.) 339–373 (Cambridge University Press, 2020).
8. Bennett, F. H., Koza, J. R., Andre, D. & Keane, M. A. in *Evolvable Systems: From Biology to Hardware* (eds. Higuchi, T., Iwata, M. & Liu, W.) 453–469 (Springer, 1997).
9. Koza, J. R., Keane, M. A. & Streeter, M. J. *Sci. Am.* **288**, 52–59 (2003).
10. Keane, M. A., Koza, J. R. & Streeter, M. J. Apparatus for improved general-purpose PID and non-PID controllers, 2005. *US Patent* **6**, 847–851 (2002).
11. Lohn, J. D., Linden, D. S., Hornby, G. S., Kraus, W. F. & Rodriguez-Arroyo, A.in *Proceedings of the 2003 NASA/DoD Conference on Evolvable Hardware (USA, 2003)*155–163 (IEEE Computer Society, 2003).
12. Stokes, J. M. et al. *Cell* **180**, 688–702.e13 (2020).
13. Marchant, J. *Nature* https://doi.org/10.1038/d41586-020-00018-3 (2020).
14. Brown, T. et al. in *Advances in Neural Information Processing Systems* vol. 33 (eds. Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H.) 1877–1901 (Curran Associates, 2020).
15. George, A. & Walsh, T. *Nature* **605**, 616–618 (2022).
16. Tamò-Larrieux, A., Ciortea, A. & Mayer, S. *Technol. Soc.* **71**, 102088 (2022).
17. Lieber, O., Sharir, B. L. & Shoham, Y. Jurassic-1: Technical details and evaluation. Technical Report (AI21 Labs, 2021).

Acknowledgements

T.W. is supported by the Australian Research Council via a Laureate Fellowship FL200100204.

Author contributions

T.W. ran the computational studies. A.G. and T.W. both wrote and revised the manuscript.

Competing interests

The authors declare no competing interests.