AI in society — the unexamined mind

人工智能走向社会—无法检验的思维

AI = artificial intelligence 人工智能

from The Economist February 17, 2018

第一段：

Science fiction is littered with examples of intelligent computers, from HAL 9000 in "2001: A Space Odyssey" to Eddie in "The Hitchhiker's Guide to the Galaxy". One thing such fictional machines have in common is a tendency to go wrong, to the detriment of the characters in the story. HAL murders most of the crew of a mission to Jupiter. Eddie obsesses about trivia, and thus puts the spacecraft he is in charge of in danger of destruction. In both cases, an attempt to build something useful and helpful has created a monster.

从《2001太空漫游》中的哈儿到《银河系漫游指南》中的艾迪，科幻小说中一直不乏智能计算机的戏份。这类虚构机器的一个共同特点是容易出问题，给故事中的人物造成损害。在一次飞往木星的任务中，哈儿杀害了大部分宇航员。艾迪则执着于细枝末节，险些导致它控制的太空船解体。这两个故事中，人们试图打造得力的机器助手，不料却创造出了怪物。

- Science fiction 科幻小说
- have sth. in common 有……共同点
- a tendency to … 有……趋势/倾向（容易……）
- detriment n. 损害，伤害
- obsesses about 执着，痴迷
- be in charge of 负责
- destruction n. 破坏，毁灭
- an attempt to do sth. 去做某事的尝试

  an attempt to build something useful and helpful

  attempts to create artificial intelligence

  attempts to develop self-driving vehicles

写作：One thing... have in common is...

One thing such fictional machines have in common is a tendency to go wrong, to the detriment of the characters in the story.

这类虚构机器的一个共同特点是容易出问题，给故事中的人物造成损害。

第二段：

Successful science fiction necessarily plays on real hopes and fears. In the 1960s and 1970s, when HAL and Eddie were dreamed up, attempts to create artificial intelligence (AI) were floundering, so both hope and fear were hypothetical. But that has changed. The invention of deep learning, a technique which uses special computer programs called neural networks to churn through large volumes of data looking for and remembering patterns, means that technology which gives a good impression of being intelligent is spreading rapidly. Applications range from speech-to-text transcription to detecting early signs of blindness. AI now runs quality control in factories and cooling systems in data centres. Governments hope to employ it to recognise terrorist propaganda sites and remove them from the web. And it is central to attempts to develop self-driving vehicles. Of the ten most valuable quoted companies in the world, seven say they have plans to put deep learning-based AI at the heart of their operations.

成功的科幻作品必然会利用真实的希望和恐惧。上世纪六七十年代哈儿和艾迪被创作出来时，创造人工智能（以下简称 AI）的尝试陷入了困境，因此希望和恐惧都是假想出来的。但这已经改变。随着深度学习技术（名为“神经网络”的

专用电脑程序处理大量数据以寻找并记住其中的模式）的诞生，给人“聪慧”印象的科技正在快速传播。从语音转录文字到检测早期失明症兆，各种智能应用应有尽有。如今AI已被用于控制工厂的质检程序和数据中心的冷却系统。政府希望运用AI识别恐怖分子的宣传网站并将之从网络上删除。另外，AI也是自动驾驶汽车研发的核心所在。在全球市值最高的十大上市公司中，有七家表示已计划将基于深度学习的AI技术作为核心事务。

- flounder v. 挣扎
- hypothetical adj. 假设的,假定的;有待证实的
- application n. 应用，申请（app）
- range from 范围是.....
- transcription 文本记录，翻译
- propaganda 宣传，宣传运动
- quoted company 上市公司

阅读/翻译：分裂结构+嵌套结构

The invention of deep learning, a technique which uses special computer programs called neural networks to churn through large volumes of data

looking for and remembering patterns, means that technology which gives a good impression of being intelligent is spreading rapidly.

随着深度学习技术（名为“神经网络”的专用电脑程序处理大量数据以寻找并记住其中的模式）的诞生，给人“聪慧”印象的科技正在快速传播。

第三段：

Real AI is nowhere near as advanced as its usual portrayal in fiction. It certainly lacks the apparently conscious motivation of the sci-fi stuff. But it does turn both hope and fear into matters for the present day, rather than an indeterminate future. And many worry that even today's "AI-lite" has the capacity to morph into a monster. The fear is not so much of devices that stop obeying instructions and instead follow their own agenda, but rather of something that does what it is told (or, at least, attempts to do so), but does it in a way that is incomprehensible.

真正的AI远不如虚构作品中惯常描写的那般先进。它肯定没有科幻小说中那种貌似有意识的动机。但它确实把希望和恐惧都变成了当前的现实，而不是模糊不定的未来。许多人甚至担心现在的“轻度AI”工具也足以演变为怪物。最令人

害怕的还不是AI设备不服从指令，而是它虽然听从指令（或至少尝试服从指令），却是以人们无法理解的方式执行。

- portrayal n. 描述,描写;画像
- motivation n. 动力;动机;诱因
- indeterminate adj. 模糊的，不定的
- incomprehensible adj. 费解的，无法理解的，不可思议的

阅读/翻译：平行结构

The fear is not so much of devices that stop obeying instructions and instead follow their own agenda, but rather of something that does what it is told (or, at least, attempts to do so), but does it in a way that is incomprehensible.

最令人害怕的还不是AI设备不服从指令，遵循自己的议程，而是它虽然听从指令（或至少尝试服从指令），却是以人们无法理解的方式执行。

第四段：

The reason for this fear is that deep-learning programs do their learning by rearranging their digital innards in response to patterns they spot in the data they are digesting. Specifically, they emulate the way neuroscientists think that real brains learn things, by changing within themselves the strengths of the connections between bits of computer code that are designed to behave like neurons. This means that even the designer of a neural network cannot know, once that network has been trained, exactly how it is doing what it does. Permitting such agents to run critical infrastructure or to make medical decisions therefore means trusting people's lives to pieces of equipment whose operation no one truly understands.

造成这种恐惧的原因是，深度学习程序会在处理的数据中发现模式，并据此重组自身的数字结构，从而完成学习过程。具体来说，这些程序是模拟神经科学家所认为的人脑学习机制——程序内的计算机代码就如同大脑的神经元，程序就是通过改变这些代码的连接强度来学习的。这意味着一套神经网络受训后，连它的设计者也无法了解其行事方式。因此，如果允许这些智能代理来管理关

键基础设施或做出医疗决策，就相当于把人命付托给了没人真正了解其运作方式的设备。

- innards n. 内脏，内部结构
- in response to 回应，响应
- emulate v. 模仿
- neuroscientist n. 神经学家
- neuron n. 神经元，神经单位
- neural adj. 神经的
- infrastructure n. 基础设施；基础结构

写作：The reason for... is that...

The reason for this fear is that deep-learning programs do their learning by rearranging their digital innards in response to patterns they spot in the data they are digesting.

造成这种恐惧的原因是，深度学习程序通过重新排列其数字结构（内部）来完成学习，以响应它们在所消化的数据中发现的模式。

阅读/翻译：断开+简化长难句

Specifically, they emulate the way neuroscientists think that real brains learn things, by changing within themselves the strengths of the connections between bits of computer code that are designed to behave like neurons.

具体来说，这些程序是模拟神经科学家所认为的人脑学习机制--程序内的计算机代码就如同大脑的神经元，程序就是通过改变这些代码的连接强度来学习的。

阅读/翻译：嵌套结构+分裂结构

This means that even the designer of a neural network cannot know, once that network has been trained, exactly how it is doing what it does.

这意味着一套神经网络受训后，连它的设计者也无法了解其行事方式。

写作：doing作主语/宾语

Permitting such agents to run critical infrastructure or to make medical decisions therefore means trusting people's lives to pieces of equipment whose operation no one truly understands.

因此，如果允许这些智能代理来管理关键基础设施或做出医疗决策，就意味着把人命付托给了没人真正了解其运作方式的设备。

第五段：

If, however, AI agents could somehow explain why they did what they did, trust would increase and those agents would become more useful. And if things were to go wrong, an agent's own explanation of its actions would make the subsequent inquiry far easier. Even as they acted up, both HAL and Eddie were able to explain their actions. Indeed, this was a crucial part of the plots of the stories they featured in. At a simpler level, such powers of self-explanation are something software engineers would like to emulate in real AI.

但如果AI代理能以某种方式解释自己运作的原因或动机，不但会增进人们对它们的信任，它们本身也会变得更有用。而且万一出现问题，它们对自身行动的解释对后续调查也会大有助益。就连胡作非为的哈儿和艾迪也能解释自己的行为。事实上这在它们存在的故事中还是情节的关键部分。在更简单的层面，这种自我解释的能力是软件工程师希望在现实AI中效仿的。

- subsequent adj. 随后的，继…之后的

• inquiry n. 调查；质询；探究

• crucial adj. 决定性的；重要的

语法：嵌套结构+分裂结构

if条件句的虚拟语气（见《句句真研》的第二部分第三章特殊句式）

If, however, AI agents could somehow explain why they did what they did, trust would increase and those agents would become more useful.

And if things were to go wrong, an agent's own explanation of its actions would make the subsequent inquiry far easier.