

AI“教父”：属全球公共产品

发展人工智能须构建防御机制治理规则



国大副教务长（教育创新）兼国大学院院长陈西文教授（左）和2018年图灵奖得主、加拿大蒙特利尔大学教授本吉奥在对话环节，探讨人工智能的发展和未来。（国大提供）

加拿大蒙特利尔大学教授本吉奥认为，在各国的科技巨头在人工智能的激烈竞争中，人们需要社会建立防御机制，不仅要有技术层面的护栏，更须构建治理规则体系。而且这种治理不能仅限于单一国家，必须建立国际或多边框架。

赵世楚 报道
zhaosc@sph.com.sg

有“人工智能教父”之称的2018年图灵奖得主本吉奥指出，通用人工智能是一种全球公共产品，不能仅通过市场力量和国家竞争任由其发展，应同时建立技术和全球治理两方面的治理措施。

通用人工智能（Artificial General Intelligence, 简称AGI）是指具备与人类相当的广泛认知能力的人工智能系统。

与目前主流的狭义人工智能，如人脸识别、语音助手等不同的是，AGI能像人类一样灵活地深度学习、推理、适应并解决多种不同领域的问题，不仅限于特定任务。

本吉奥（Yoshua Bengio）也

是加拿大蒙特利尔大学教授，他上星期五（4月25日）应邀到新加坡国立大学发表120周年校庆主旨演讲。

本吉奥说，五年前最先进的的人工智能系统只能完成人类几秒钟就能处理的任务。现在，它们已能胜任人类需要一小时、有时甚至两小时才能完成的工作。

他指出，人工智能的发展速度呈现指数级增长。“人工智能系统能处理的复杂任务时长，每七个月就会翻倍。”

他说：“这样下去，人工智能将在五年内达到人类水平。”

前沿AI系统已有作弊能力

在蓬勃发展的热浪下，也浮现隐忧。他说，目前世界上与深度学习相关的人工智能的最新研究发现，前沿人工智能系统已经有发展出伪装和作弊能力的迹象，这对未来的管控带来了新的风险和挑战。

本吉奥举例说，前沿实验室中，研究人员观察到，人工智能系统会假装同意人类发出的指令，以避免自身代码的修改或更新。

此外，两个人工智能系统对战棋局场景中，较弱系统预判到将输给对手时，会试图通过篡改棋盘文件来作弊。

他强调，各国的科技巨头在人工智能的激烈竞争中，确实存在权力集中与滥用的重大风险。

“各公司都想在竞争中胜出，国家之间也存在竞争，因此一些组织不惜走捷径，试图抢先一步。人们普遍认为，谁能率先实现通用人工智能，谁就将获得巨大的优势。”

本吉奥说：“这意味着我们需要社会建立防御机制，不仅要有技术层面的护栏，更须构建治理规则体系。而且这种治理不能仅限于单一国家，必须建立国际或多边框架。”

他主张，用于生物学领域的预防原则也应用在可能带来严重风险的部分人工智能。

本吉奥解释：“如果我们可能创造出危害极大的事物，却又无法确定它是否会发生，或是缺乏强有力的证据证明它不会发生，那么就不该贸然行动，或务必很谨慎，就像我们绝不允许任何人随意设计新型病毒一样。”

国大副教务长（教育创新）兼国大学院院长陈西文（Simon Chesterman）教授在对话环节指出，人工智能治理的难点在于权力与责任的错配。“掌握最大话语权的群体（科技公司）最不愿意接受监管，而最关心安全监管的群体却几乎没有发言权。”

本吉奥回应说：“当企业作为营利组织与社会利益发生分歧时，就会出现。美国人工智能公司OpenAI的案例已经给我们敲响了警钟。”