# Article information:

Planning for AGI and beyond  
<https://openai.com/blog/planning-for-agi-and-beyond/>

# Article summary:

1. 我们的目标是确保人工智能（AI）系统比人类更聪明，以便为全人类带来好处。

2. 我们希望AGI的好处、访问权和治理权得到广泛而公平地分享。

3. 我们正在努力创造出更加一致和可驾驭的模型，同时也在进行全球性对话，就如何治理这些系统、如何公平分配它们带来的好处、如何公平分享访问权这三个问题进行讨论。

# Article rating:

May be slightly imbalanced: The article presents the information in a generally reliable way, but there are minor points of consideration that could be explored further or claims that are not fully backed by appropriate evidence. Some perspectives may also be omitted, and you are encouraged to use the research topics section to explore the topic further.

# Article analysis:

本文是OpenAI关于AGI和它之上的规划的一篇文章。文章中所提出的要点都是合理的：我们应该尽量避免AGI带来的风险，同时也要尽量发掘它所带来的好处。然而，文章存在一些可能存在的偏见或片面性。

首先，文章似乎将AGI作为一个必然存在的情况：即使当前进步可能会遭遇障碍，但我们不应该永远阻止它的发展。然而，由于AGI存在巨大风险（包括情感不当使用、严重意外、社会动荡等方面），因此我们应该对它有所保留。

此外，文章中强调“快速学习”和“小心迭代”作为成功对待AI部署技术所必要的方法。然而，文章未能考虑到AI部署中可能存在的风险——例如数据泄露、隐私侵害、数据不准确性、数据不平衡性、数据不安全性、数据样本不代表性以及AI决策中存在的歧视问题——因此无法真正说明AI部署中所必要采取的方法。

最后，文章强调了“使用AI来帮助人们评估复杂模型”作为一个重要方法。然而，文章未能考虑到使用AI来评估复杂模型也是一个风险——例如AI决策中存在隐含偏差或者AI决策中存在不准

# Topics for further research:

* AI 部署风险；
* 数据泄露和隐私侵害；
* 数据不准确性、数据不平衡性和数据不安全性；
* 数据样本不代表性；
* AI 决策中存在的歧视问题；
* AI 决策中存在的隐含偏差。

# Report location:

<https://www.fullpicture.app/item/978869e1a6591cb4bc6eaa8c3a6e39ec>