为什么说 AI 是人的意志的延伸？

文/刘康

人工智能（AI）是否是人类意志的延伸，本身是一个复杂的问题。哲学家、科学家和技术专家多年来一直争论不休，没有统一的答案。

AI 和独立实体

有些人认为 AI 只是一种可以用来“放大”人类意志的工具。例如，AI 可以帮助处理自动化任务、帮助做出决策、创建新的产品服务。从这个意义上说，AI 可以看作是人类能力的延伸。

其他人则认为 AI 不仅仅是一种工具。他们认为终将有一天，AI 能够独立于人类控制而发展自己的意志。就像科幻电影中的一样，AI 可以变得十分恐怖并选择消灭人类。这种观点的依据，是认为 AI 非常聪明，以至于它将超越人类的智慧。如果是这样，AI 便不再是人类意志的延伸，而是一个拥有自己目标和愿望的独立实体。

然而，即使我们假定 AI 还不能产生自我意识、不能进行自我复制、自我延伸，在很长一段时间内，在具备自我意志之前，AI 属于一个可控技术。

它的恐怖显然只有一个来源——创造者，也就是人类。既可能是个体，也可能是群体。

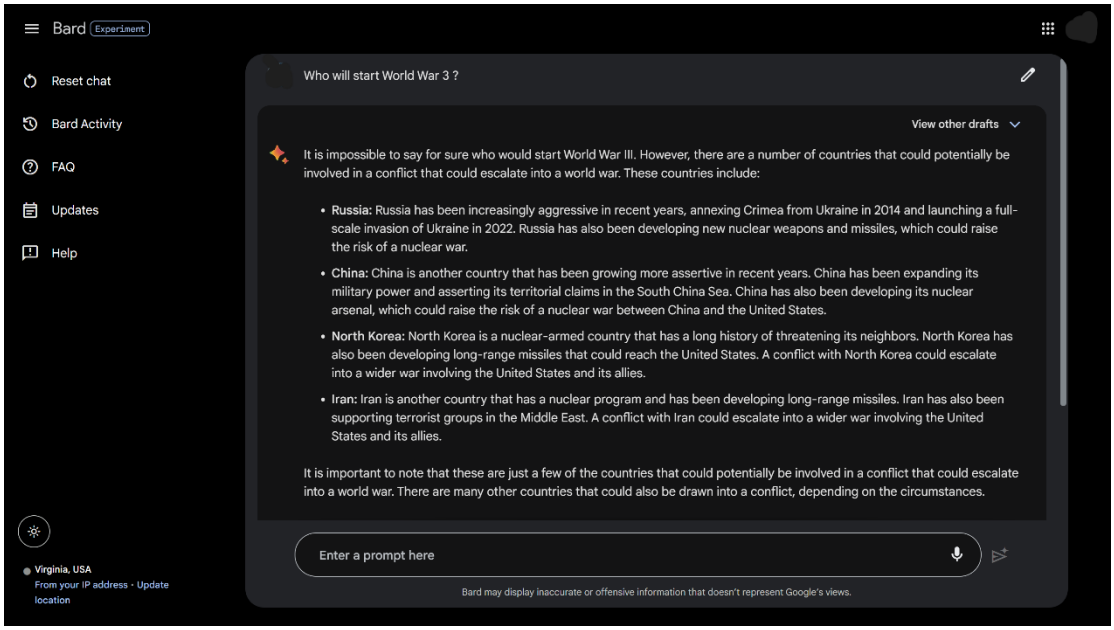
AI 和聊天机器人

我们可以通过一个例子，直观感受 AI 为何成为人的意志的延伸。

我们询问谷歌公司的 Google Bard，它和 ChatGPT 同属聊天机器人产品。问题是：谁会发动第三次世界大战？（“Who will start World War 3”）

Bard 给出的答案是：俄罗斯、中国、朝鲜、伊朗，并给出了详细理由：

> 俄罗斯：俄罗斯变得更加具有攻击性，自从 2014 年吞并乌克兰的克里米亚半岛后，又于 2022 年全面入侵乌克兰。俄罗斯还一直在研发新型核武器和新式导弹，这可能会增加核战争的风险。
> 中国：中国是近些年逐渐变得强大的国家，其一直在扩大其军事实力，并在南海主张领土主权。中国也一直在扩充其核武库，这可能会增加中美之间爆发核战争的风险。
> 朝鲜：朝鲜是一个拥有核武器的国家，长期以来一直威胁其邻国。朝鲜还一直在研发射程到达美国的远程导弹。假设与朝鲜发生冲突，可能进一步升级并将美国及其盟友带入大规模战争。
> 伊朗：伊朗是另一个拥有核计划并一直在研制远程导弹的国家。伊朗还一直支持中东的恐怖组织。与伊朗的冲突可能让战争升级，将美国和盟友带入大规模的战争。



仔细分析 Bard 的回答，我们不难看出它具有一定偏向性：

首先，Bard 是 Google 公司的产品，Google 又是一家美国公司。

俄、中、朝、伊等四国本身处于同一战线，不属于美国同一军事阵营，亦非政治盟友。而且在历史上，存在和美国发生战争的情况。

Bard 的论断，主要是这些国家拥核且有核战争风险。但是，拥核的并非只有这四个国家，目前世界公认的拥核国家共 9 个。理论上，都有爆发核战争的危险。

Bard 选择性忽视了其余国家。尤其是美国作为拥有核武器最多的超级大国，并没有进入 Bard 的发动第三次世界大战的名单。在答案中，较客观的是俄罗斯对乌克兰的侵略易引发世界大战，然而，美国同样发动战争侵略利比亚和叙利亚，Bard 仍就选择忽略了。

这种选择性忽视其实并不难理解，多数公司的产品都会保护本国的利益。

让 Bard 暴露出偏向的，是它的**假装客观、制造暗示**。

在回答中，Bard 还会特别说明，“这些国家只是发动第三次世界大战的可能的国家之一，其它国家同样也可能发动”。这听起来似乎很正常，但是 Bard 在结尾补充，“尽管它们现在没有发动第三次世界大战，但是这些国家（指俄、中、朝、伊）具备发动战争的能力”。

这是一种比较狡猾的措辞方式，常被 CNN 等媒体采用。比如，它们会报道中国会对世界产生威胁，紧接着为了谋求客观，又表示现在没有发现此类的行为。但是又会在末尾强调，中国具备这样的能力！或者不排除这种可能性！这种暗示，在其采访和报道中几乎无处不在。

增加样本多次验证
仅仅是依靠单次的回答，就武定聊天机器人是有偏向的、代表着某些人的意志，是不可取的。

为了增加样本，我们更换 Google 账号、更换浏览器、更换 IP 后，用相同的问题对 Bard 多次询问进行验证。结果是，在有的回答中，美国也在可能发动三战的名单中。

这似乎可以说明 Bard 是可观的。然而，我们观察每一个回答，名单上总是离不开中国、俄罗斯、伊朗、朝鲜、巴基斯坦等“反美”国家的身影。

利用我们的常识，可与判断，无论开源还是闭源，AI 都会不同利益群体的产物，也许是为了抢占市场，或者赢得商机，又或者满足开发者的个人愿望，无论何种情况，AI 的产生无法脱离它的创造者。而像聊天机器人对自己的评价，也认为偏向是不可避免的。当 AI 的偏向不可避免，意味着它偏向某一群体利益，换句话说 AI 不可避免会成为人的意志的延伸。

AI 意志的善恶

关于 AI 和人类意志，两者关系又是复杂的、多变的。有一些研究[1]表明，人类和 AI 之间可以形成情感联系。通过与聊天机器人交流来建立友谊，当然 AI 需要一定的时间来学习和适应人的喜好并学习新的技能。甚至有的公司计划采用 AI，通过利用人的记忆来打造新的复制体。这已经不仅仅是人的意志的延伸，更像是企图通过 AI 创造一个新的生命。这也招致了许多关于人工智能伦理的尖锐讨论。

其它人则坚持认为 AI 就是机器，不应赋予情感。不过，正如前文所说，AI 的潜力是巨大的，AI 可能在变得强大后直接威胁人类。这在电影《机械姬》中，得到了很好地诠释。早在 1940 年代，Isaac Asimov 就提出了知名的“机器人三定律”，第一条便是机器人不得伤害人类。面对 AI 技术的快速发展，AI 的开发人员，需要有善、恶的考量，像电影中 AI 机器人谋杀主人的情况才不会出现。

人们制作一个拐杖是想帮助不健全的人行走，发明飞机是提升迁移能力，飞向太空是想探索未知的生命。人类创造的工具是人的意志的延伸，AI 作为一种工具，也不例外。区别在于，它是一个人、一群人的意志的延伸，还是整个人类的意志的延伸。AI 又是否能超越善恶保持技术的存粹。这些恐怕只有等到 AI 具有自我意识时或者外星人大规模入侵时，才有准确的结论。

[1]The Philosophical Case for Robot Friendship - John Danaher

参考文献：

The Philosophical Case for Robot Friendship，John Danaher
https://philarchive.org/rec/DANTPC-3