

AI来了，小说家要死吗：董启章X伊格言X糖匪 | 中港台三地文化对话

被AI屠杀之前，作家们，有信心成为更好的AI吗？



图：Mantha Mok / 端传媒，由midjourney、InsightFace生成

特约撰稿人 红眼 发自香港

刊登于 2024-02-02

[# AI写作](#) [# midjourney](#) [# Open AI](#) [# AI监管](#) [# AI编剧](#) [# ChatGPT](#) [# AI](#)



随著AI技术日趋成熟，对不同创作范畴都带来极大冲击，譬如2023自春到秋荷里活编剧大罢工，诉求之一就是拒绝润饰 AI 生成剧本，是一场关乎饭碗与尊严的工潮。然而，又接连有新人作家以AI作品打败同行，横扫各大[指标性文学奖](#)。

算是请枪作弊，还是用得其所，把握大势？连人类作家评审都被“骗过”，是否代表AI创作已经领先人类？不久的将来，是否要由AI审判人类创作优劣？如此说来，AI尚未像科幻电影般统治人类，但首当其冲，已在大数据年代迎来一场人类创作者的大屠杀。而展望一个创作者与AI技术共存的新时代，是否太过乐观？

又一年伊始，端传媒文化组邀来董启章、伊格言和糖匪——分别来自香港、台湾及中国大陆，年龄层也介乎三代之间的三位小说家，相聚来谈谈这个“时新”话题。面对 AI 巨兽降临，乐观与悲观，兴奋与忧虑皆有。董启章认为人机一体将会成为新时代的审美标准，甚至是引导社会结构转型的重要时机；伊格言则主张放慢步伐，人机关系需要立宪，建立一个与 AI 共存的安全路线；糖匪则关心技术黑盒造成的权力及人际关系扭曲。

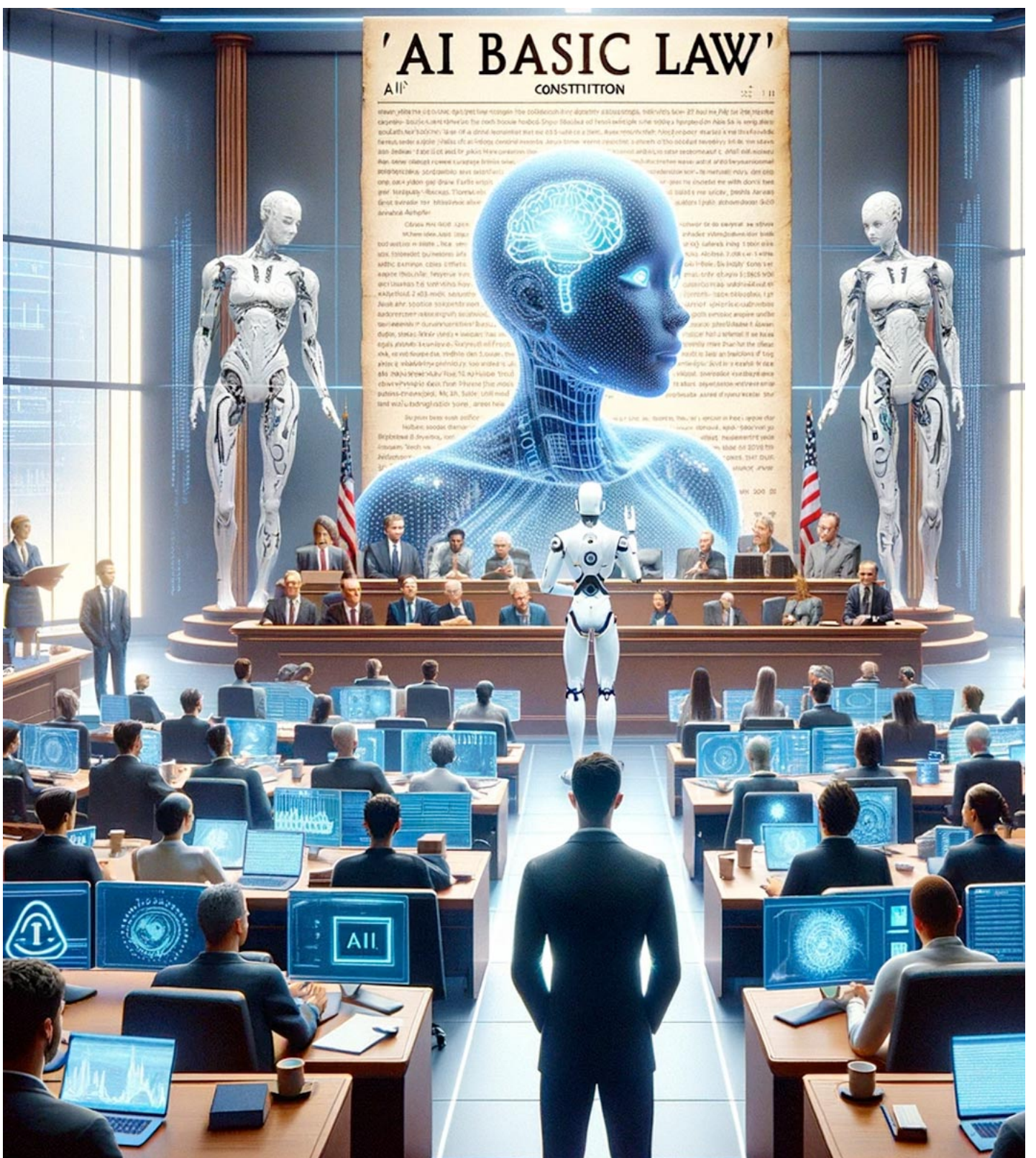
不可为敌，与之为伍，但谁人知道，AI是否都这样想？

董启章：香港小说家。小说《爱妻》和《后人间喜剧》涉人工智能及意识下载等主题。运用 GPT 把自己的作品翻译成英文，于个人电子报 DKC in Translation 发表。借助区块链技术推动分散式出版，自主出版 NFT书《天工开物·栩栩如真》、《心》及《自作集 Autofiction》。

伊格言：台湾小说家。曾获选台湾十大潜力人物，全球华语科幻星云奖长篇小说奖。获邀参与美国智库博古睿研究院（Berggruen Institute）于北京所举办的科幻论坛；刊有与台大心理系、脑神经学家谢伯让教授以“一路开外挂！AI与人类的狂飙时代”为题之长篇对话。

糖匪，作家评论人，SFWA（美国科幻和奇幻作家协会）正式作家会员。出版小说《后来的人类》《奥德赛博》，《看见鲸鱼座的人》，《无名盛宴》，短篇小说集于2022年在意大利正式出版。专注于人工智能后人类方向，好奇人类作为种群的发展方向，以及与人工智能媾和的可能。曾获美国最受喜爱推理幻想小说翻译作品奖银奖。

端=红眼、伊=伊格言、董=董启章、糖=糖匪



“因此，人类要为 AI 立宪，就是给一个 AI 基本法，限制那些矜谷的科技业大头们，什么事情可以做，什么事情不能做。为什么呢？因为它是另外一种生物，如果等到我们让它这样做了，尤其让它有身体之后，也就是给了另外一种生物一个类似人类的自我的时候，它可能就真的会反叛人类了。”伊格言 图：Mantha Mok / 端传媒，由DALL·E生成

AI只能生产内容农场？

“某些农场文章的质素，我认为就是ChatGPT写的，通常真人去写的话，你还是会做一些汇整，但这些内容农场的文字连汇整的感觉都没有，它从头到尾都没写出任何观点。”伊格言

端：大家什么时候开始觉得AI不再那么遥远，而是大势所趋，已经“入侵”到大家身边？

伊：早在2014年，就出现了那个“微软小冰”（微软亚洲国际网络工程院在中国推出的AI聊天系统）。当时就在北京参加过相关活动。后来它不是还出了一本诗集吗？我觉得那时我们就已经遇到AI写作的问题。但“微软小冰”的诗是写得不好，我意思是说人类的好诗人可以写得比它好很多。也因为诗本来就比较私密，有可能写得比较隐晦，要准确评价一首诗写得好或不好，并不是那么容易。

直到现在，那么厉害的ChatGPT出现之后，这个情况我认为都没有改变。比如说，我同样觉得ChatGPT写的小说很烂，但是呢，确实已经在内容农场上看到一些应该不是真人所写，我认为是ChatGPT写的文章了。就是那种内容农场上最常见的文章，像是钟楚红出道故事、钟楚红曾经如何风靡一时，类似这样的内容然后配上一张明星图片，就有流量了嘛。仔细看某些这种农场文章的质素，我认为有一部分就是 ChatGPT 写的，通常真人去写的话，你还是会做一些汇整，但这些内容农场的文字连汇整的感觉都没有，它从头到尾都没写出任何观点。所以说AI写作的确已经跟日常生活息息相关。

“我认同它现在写的东西还是很普通，用文学水准来鉴定的话，但在一般应用也就是较实用性的文体，它很可能已经做得很好。暂时来说，会被取代的很可能就是机械性的文字操作。”董启章

董：我留意 AI 写作可能不是太早，是从 2018 年开始。那时 OpenAI 还是一家非常小型的公司，刚发布了 GPT-2 这个很早期的 AI 模型，让它模拟十种不同的文体，写了十篇短文，包括新闻报导，然后也有奇幻小说，当然它的内容很普通，只是一些拼凑在一起的东西，但从文字风格来说，机器可以模仿不同的文类而去写作，已经令很多人觉得惊讶，在一些外国报刊上引起了热烈讨论。

那时我自己也写过一篇专栏，是引用卡尔维诺收录在《Literature Machine》这部文集里的一篇文章《Cybernetics and Ghosts》。早在 1967 年，卡尔维诺已谈到所谓写作机器，提出有没有可能出现一台写作机器呢？他形容语言本身就是一个组合的游戏，就是把词语跟词语放在一起，让一个字跟著另一个字，一个接一个有条理、有意义地放在一起。跟 2018 年我们谈到 GPT 是如何运作有点相似，它就是按机率去计算，当你输入一个字，或一个主题，然后后面应该把什么字放进去。不过，卡尔维诺有一个很大胆的观点，但是我非常同意。他最后提到，也许我们根本没需要发明一台文学机器，因为所有作家，在他们运作良好的时候，他们都是一台文学机器。我们本身就已经是文学机器了，哈哈。

即使进化到 GPT-3.5 还是 GPT-4，我认同它现在写的东西还是很普通，用文学水准来鉴定的话，完全不是一些合格的文学作品，但在一般应用上，也就是比较实用性的文体，它很可能已经做得很好。所以，暂时来说，会被取代的很可能就是一些非常机械性的文字操作，这一类的工作是会被机器取代的，但之所以会被取代，因为它本来就是由人类去做很机械性的东西。当更有效率的机器出现，他们这一批人的工作就会被取代。

但如果说到有创造性的，也就是艺术、文学的那个层次，现在的 AI 技术还是做不到。

“所以 AI 没有办法满足我对自我、对存在价值的追逐，它就只是满足一种生产机率、一种消费需求。工具本身是永远不可能构成问题，追溯我们人类的那些历史，问题永远是那些使用工具的人。”糖匪

糖：我很赞成，我们本身就是一台写作机器，而且要写出同等数量的小说，我们一定更节能，我们是超级便宜的机器。

是什么时候让我觉得 AI 进入到人类写作了呢？键盘输入法。我不知道你们是用五笔还是用其他的，我是用拼音。当我只是打出了两个同音的首字母，它会出现两个是我之前没有想到，但我觉得更好的词语。它提供了某种联想，于某种意义上也对我们的意识进行了引导，而这个引导是不被我们察觉的。

但在我看来，写作它其实是一个在微暗的光线中进行的过程，如果我在写小说、写诗的时候，想得特别明白没有一点阴影，我觉得我的创作是毫无乐趣。所以 AI 没有办法满足我对自我、对存在价值的追逐，它就只是满足一种生产机率、一种消费需求。商业写作的那些作者们，他们完全可以借助这些软体来帮

助写作。还有就是它能帮助你生成故事大纲，这个简直太棒了，因为我是特别不爱写所谓的主题跟大纲，那我就可以顺畅解决掉所有需要故事大纲的场合了。就这一点来说的话，你可以完全把它当作工具。

工具本身是永远不可能构成问题，追溯我们人类的那些历史，问题永远是那些使用工具的人。你没有足够的智慧去驾驭，或者去更好地使用这些工具的话，就会发生很可怕的事情。

“AI 对创作者带来的问题，就是写作上的同质化。在这种同质化环境里养成的这一代读者，会觉得所谓的诗，所谓的小说、文学，就应该是这样的。”糖匪

回到我们说 AI 对创作者带来的问题，我觉得就是写作上的同质化。如果它是根据我们现有文本取得数据，那么所有东西就是已有的，对已有的东西进行再生成，哪怕它的序列会有变化，其实只是一个剪贴过程，也就是搜索，剪接，黏贴，一些表面的、肤浅的写作。从这个意义上来说的话，我们可能很早就开始了所谓的人工智能写作。

内容农场的文章很烂，但为什么大家喜欢看？就是我们缺乏鉴别能力，我们被大量同质化文本……而且，这种同质化的文本，它不仅仅是来自于人工智能，也来自于我们人类自己，我们在 AI 开始学会写作之前，就已经有大量的这样一种文字生产者，用剪接黏贴，或是所谓的抄袭、洗稿，进行了这种同质化的写作，造成干涸的文化生态。

在这种同质化环境里养成的这一代读者，会觉得所谓的诗，所谓的小说、文学，就应该是这样的。有个小朋友，他就拿 ChatGPT 改写的一篇海明威小说来炫耀，说它写的多棒。但它只有大概的、很干扁的形，因为它的这种算法的逻辑，完全没有办法理解文字下面所有的这些很波澜壮阔的内容，它反馈出来的东西就是一个形式。那我觉得，如果仅是这个程度，小说家就被这样的写作机器打败的话，那你不妨改行。

“为什么到我们去写作的时候，我们有可能创造出新的东西呢？因为我们的个人经验是独一无二的……AI 最致命的地方，就正正是它没有经验。”董启章 图：Mantha Mok / 端传媒，由DALL·E生成

AI是另一种生物？它需要身体吗？

“AI 最致命的地方，就正正是它没有经验。”董启章

端：以 AI 模仿其他作家的文字风格，能否定义为“抄袭”？但人类作家不也是经常模仿、学习前人的写作风格吗？譬如一个年轻的香港作者，他就是受到董启章作品影响而开始阅读、创作小说，我们跟 AI 还有什么不一样？

董：从表面看是有点相似的，因为 AI 就是去学习既有的东西嘛，而我们人类也是先学东西，然后才写自己的东西，我们都一样受到既有东西所影响，或是去学习，然后再培养自己的修养跟能力。但为什么到我们去写作的时候，我们有可能创造出新的东西呢？因为我们的个人经验是独一无二的，就算我读的书都是既有的书，但这些书进入我的意识之后，我是在这个基础上再去经历我的生活，每天跟新的东西互动、产生刺激，然后把一些我们学回来的、受到影响的东西变成独有的东西。所谓创意，其中一部分就是来自这种经验。

但暂时来说，AI 最致命的地方，就正正是它没有经验。事实上，它并不存在于一个实在的时空，它是一堆抽象的数据，是一堆没有形态的东西。不过，如果用一种比较客观的想像，将来 AI 发展到某一个点，它会有一种方法，有一些介面让它连接到经验世界，让它能够看东西，能够观察每一天这个世界的发展，跟人类有互动，跟其他的 AI 或世界上的其他事物互动，到时候它就有了获得经验的这个能力，变成一个有自己立场跟创意的主体。

当然，现在的 AI 就是没主体性，所以它产生出来的东西都没意义，因为它没有一些有意义的概念在里面，但是当它变成一个经验体的时候，意义就会出现。就是它的互动，它在这个经验世界里的存在，它怎么去介入这个世界，这样子就会产生类似于创意的东西。

伊：我也同意这样的观点，其实大家都在问的问题，就是 AI 到底有没有“自我”。缩窄一点来说，目前看起来 Chat GPT 是没有自我的，它只是把话说出来，说得很像一回事，但你问 ChatGPT 知道自己到底在说什么吗，它其实是不知道的，它的运作方式就是把语言组合，处理语言的构建而已。虽然 ChatGPT 是不会有原创性的，但它有机率创造出一些有用的、实用性的东西，随著技术发展，它生产的东西的有机率会愈来愈高，以前生产一百篇文章只有三篇能用，后来可能变成有三十篇能用，但那不是 creative writing。

就像刚刚启章老师所说，人类是从小到大一直经由跟外界的情感互动、知觉互动、物理环境上的互动，以及所有人类语言上面的交流等等，我们才长成这样的，但 ChatGPT 不是，它没有清楚自我，准确地说，没有“类似像我们人类这样”的自我。既然如此，其实你可以把它当成地球上的另外一种生物，只是它的养成方式跟我们人类不一样，它的自我也和人类的自我不一样。

而这种生物它有可能生产一些，我们觉得可以用的东西，它在生产力上面对我们有一些帮助，对某些原创性比较低的工作有帮助。然而，当这个世界被众多没有原创性的东西所淹没的时候，也就是我们的后代都是看到这些东西的时候，那么人类的辨别能力也可能会越来越弱。

“有时候觉得甚至 AI 所表现出的这种服从性，以及服务精神，就很像女性在漫长人类历史上所充当的这样一个角色。”糖匪

糖：我很认同的这一点，就是 AI 它缺乏外部感知，无法和外界进行一个物理世界的互动，或是这种物理世界的经验。但 AI 没有跟外部世界进行沟通，是否就不具有主体性？看你要怎么定义自我，如果自我还是按照人类中心主义的方式定义的话，OK，它没有自我，但如果去破除这个想法来说的话，那么 AI 它其实是有自身的意义。

所有我们现在人类能用语言表达的感受，都是可以用一些化学的、生物的方法去直接传导给 AI 的。人类的所有这些感受其实也就是一个生物电，一个神经元冲动的过程，其实也没有那种那么不可替代。但即便是现在这样，在我們都能产生虚假记忆的这样一个时代，我们人类自己都没有办法知道意识是如何产生的，那么我们仅仅了解 AI 的工作原理跟逻辑，是否就能判断 AI 它没有意识呢？

这事情我觉得它很有趣，有时候觉得甚至 AI 所表现出的这种服从性，以及服务精神，就很像女性在漫长人类历史上所充当的这样一个角色。就是说，其实我们自己的逻辑跟想法，但是我们跟你们打交道的时候，是使用了另一种工作方式，让你们一直误以为我们就是这么想，我们是这么去思考的，甚至在很长的一段时间里，我们被视作没有逻辑，没有灵魂，没有主体性。可能是因为我这样的女性、边缘群体身份，我觉得我们是没有那么能肯定的判断 AI 有没有意识。

伊：就目前所知，它是没有像人类这样的自我，没有“跟我们一样”的意识。要说它没有自我吗？它可能有，我们也可以说它有意识，只是那个意识或自我，跟我们人类的不太一样。我觉得这是没问题的。像糖匪刚刚说，我们人类连自己的意识是什么其实都还不太清楚，所以我们只知道 AI 的意识，或是 AI 的自我，跟我们人类的不太一样。我是一直主张，应该把 AI 视为另外一种生物，而它的养成、思维方式都跟人类有分别。

但它又同时有一部分跟人类一样，因为它就活在语言之海，是被语言、文本所养大的。所以 ChatGPT 为什么看起来好像可以跟我们沟通，为什么你可以问它问题？就是因为这一部分的重叠。人类其实也是活在语言之海中，人类也是被文本喂大的。我们都是听了、读了很多语言而长大，但是人类自我的养成除了语言之外还包括了一些其他的部分，而那些是 ChatGPT 所没有的，就是人类跟实际世界的互动。

“或是 AI 的自我，跟我们人类的不太一样。我一直主张应把 AI 视为另外一种生物，它的养成、思维方式都跟人类有分别。”伊格言

让我想起一部纪录片《少年关影室》（The Wolfpack），一个变态的爸爸，把四、五个小孩从小关在地下室，因为他认为外面的世界都很危险，所以他不让孩子离开地下室，就这样子一直给他们看电影，大概看到了十几岁。而这就类似 AI 的养成方式了，它就是活在这样一个文本之海里。

但人类自我是怎么产生？从精神分析的角度来讨论，依照拉冈（Jacques Lacan）的镜像阶段理论，人类的自我并不是天生的，你一开始在婴儿时，可能是一大堆碎片的感官经验聚合，比如说你只知道这里感觉痛，那里感觉不舒服，所以你就哭闹，你甚至不知道原来这是你肚子饿了，婴儿时期的你还没有完整自我，要到长期跟母亲、跟这个世界互动，（世界）才慢慢让你知道，原来这是我，这是我的头，原来这是我的脚，原来我有一个完整身体，这是逐渐长出来的一个自我概念。

所以呢，如果我们想要为 AI 制造出一个比现在的 ChatGPT 更类似于人类的自我，理论我们就应该模仿这个过程，给它加上本体感觉的传感器、视觉听觉等其他知觉的传感器等等，意思就是，它们必须要有身体。给了它一个身体，它的自我，像人类的那种自我才能长出来。这是 AI 未来可能的演化路径之一，但我是反对这路径的。

像有些人在担心 AI 会不会对人类造成危害，我是一直主张不要给 AI 身体，或是赋予它们身体要有一个限度。因此，人类必须为 AI 立宪，就是给一个 AI 基本法，限制那些矜谷的科技业大头们，什么事情可以做，什么事情不能做。为什么呢？因为它是另外一种生物，如果等到我们让它这样做了，尤其让它有身体之后，也就是给了另外一种生物一个类似人类的自我的时候，它可能就真的会反叛人类了。

基于人类安全的理由，这一个演化方向是必须要被打上问号的。那当然，假设我们是基于幸灾乐祸的心理，我们把这个问号拿掉也是可以的。人类这个种族没有什么好珍惜。

董：AI 的出现，让我最感兴趣的地方，是有点类似格言所说，它是另外一种生物，或是另一种存在形态。刚才说到自我，说的时候好像就是因为它没有自我，而我们有自我，所以我们才有创造性。当然从我们人类的角度来说，是会这样判断的，但是它没有我们这种自我，可能也有另外一种形态的东西在里面运作，而那个东西的出现，对于我们人类很多观念，像已经当是理所当然的所谓自我，或是主体性，都是一种挑战。

如果给 AI 一个身体，它就变得跟人类非常类似吗？或者也不完全一样，因为这个 AI 并不会只有一个身体，比如说 GPT 现在就是一个大型的系统，假设我们身边都有一个由 ChatGPT 处理的机器人在我们身边，跟我们交朋友、谈恋爱，它分身变成了成千上万个身体的时候，这个身体还是同一个个体吗？如果它同时连到一个系统，但个别身体里面又有独特的体验，就是它同时有自我，也有比这个自我更大的一个超自我在后面。这跟我们人类是完全不同的。当然我们也有类似的概念，就是心理学里面说，我们虽然每个人都是一个个体，但在人类演化的过程之中，我们的心理结构是有共同之处，所以我们都会造梦啊。

但 AI 的状态是更奇怪了，如果它有感知的身体，它的这个经验的部分，跟它的超经验的部分会是什么关系？如果再科幻一点，当它变得越来越强，成为一个 Super Intelligence 的时候，就算是我们不主动给它一个感知介面，它很可能已经能够自行达到这种互动，而我们也不一定能够禁止它往这方面去发展。

“假设我们身边都有一个由 ChatGPT 处理的机器人跟我们交朋友、谈恋爱，它分身成了成千上万个身体，这个身体还是同一个个体吗？”董启章

糖：人工智能的超经验，是指什么？

董：超经验就是它还没有实在的物理世界经验之前，它已经有某些东西在里面，类似哲学上面的 a priori（先验）概念。如果再加上这个介面的话，它可以以这个基础去进入经验世界。

“所以如果 AI 真的有一天具有了所谓的身分性，或者说它变得跟人类一样的思维的话，它第一步要做的应该就是压榨人类吧。”糖匪 图：Mantha Mok / 端传媒，由DALL·E生成

人类将是AI的宠物？

“所以如果 AI 真的有一天具有了所谓的身分性，或者说它变得跟人类一样的思维的话，它第一步要做的应该就是压榨人类吧。”糖匪

董：还有一点，我不知道大家怎样看，就是所谓的人机合一。

糖：赛博格（Cyborg）？

董：对，就是我们人类的意识跟机器有某种联系、接通这样子。大概两个月前，以太坊的创办人 Vitalik Buterin 发表了一篇文章〈My techno-optimism〉（我的科技乐观主义），他就认为 AI 发展是一定会往前的，我们已不能把它停住，但我们要用什么样的态度往前走呢？他并不赞同一种盲目的乐观，所以他提出一个概念，就是 Defensive Acceleration，也就是在我们往前的时候，我们要怎么去保护自己，不至于盲目做出一些非常危险的东西。

然后他最后说，当 Super Intelligence 越来越强，如果我们只是放著它不理，有一天它就会超越人类，因为我们没可能跟上它，就算它不把人类毁灭，当所有东西都是 Super Intelligence 帮我们解决了，人类也会在 AI 统治的世界里变成一种类似宠物的存在。所以，要让人类跟 AI 同步的话，最终就要发展出某种人机合一，就是我们意识上能够跟它保持连接，但要怎样去实行，我们还不知道。

糖：这人机合一跟所有的权力关系都很相似，就是说是在一个怎样的权力关系的等级结构里面。格言刚才一直在用养成这个词，你知道吗？我就觉得超危险，充满那种男性的……就是男性塑造其他对象的那一种。

但这种养成意味着，其实你是对它有一个塑造。现在我感觉是我们把 AI 当作工具，是我们的一个去解决一些乏味、枯燥工作的这样一个机械性的东西。就很像又回到了蒸汽时代，大部分所谓的这种体力劳动都被机器取代了，今天也是，大部分的这种低价值的，不需要那么原创性的这种工作，都由 AI 来取代。

那之后的权力关系会是怎样呢？就是一个非常反乌托邦的，人类沦为了他们的一个感受器官，而我们失去了我们自己的主体性，但是我们可能因此获得了比现在更好的生活。什么样的生活才是更好的，我没有权力去定义什么是人类更好的生活，但在我那个反乌托邦的想像里面，那不是我要的生活。但人类其实自己也在不停的把自己的同类工具化，所以如果 AI 真的有一天具有了所谓的身分性，或者说它变得跟人类一样的思维的话，它第一步要做的应该就是压榨人类吧。

那个时候的话，可能它会许诺我们一个美好未来，然后我们就成为他们的附属。

董：当然，我们可能会想像一个危险情况，就是 AI 会成为一种强大的力量在上面控制、统治著我们，但我会从比较正面的方向去思考，怎样才能够不让它变成这样呢？

除了 AI，我近年更关心跟区块链（blockchain）和 web3 有关的东西。当然，目前很多人都比较重视 AI 发展，对区块链发展有兴趣的人是比较少。而且这两种科技表面看来是相反的，因为 AI 是由大公司发展出来的中心性系统，区块链说的是分散式（distributed）的、去中心（decentralize）的技术。但我觉得这当代两大科技，事实上是有机会把它们放在一起去理解，也就是怎样可以将 AI 去中心化，把它变成每一个个体都可以从中获得能力的技术。我认为最理想的状态是 AI 发展成每一个人都可以把握到，也就是跟区块链的精神互相配合。

像早前 OpenAI 发生的内部震动，背后就是关乎一间科技公司的组成跟运作。当初 OpenAI 是一间非牟利机构，强调所有技术都是开源的，所以它成立的时候，是尝试用一种开放模式去发展，跟所有其他的大型科技公司都不同，而且是相反。

但是呢，当它需要更多资金才能发展得更快的时候，它的开放性跟传统科技公司的那种权力架构，即是理念上跟管理上慢慢就出现冲突了。在很多人眼中 OpenAI 是失败了，它会慢慢走向其他科技企业，完全中心化，然后投资者老板完全拥有话语权。

“人类之中的绝大多数，看的永远都是眼前利益。当我们要估量一项新技术的利益，其实大家都没能力看到比较远的部分。”伊格言

伊：有消息说 OpenAI 发生的内部纷争，是跟他们内部对所谓 Alignment（对齐关系）的意见分歧有关。那为什么我们要确保 AI 跟人类之间有 Alignment，为什么要确保它的执行结果符合多数人的价值观呢？你可以说这是一种人类中心主义，但如果不这样要求的话，我们有没有办法保证 AI 的演化、发展会朝著一个比较好的方向吗？

我自己的看法是悲观的，我觉得人类本身就不太可能做到这件事情。人类之中的绝大多数，看的永远都是眼前利益。当我们要估量一项新技术的利益，其实大家都没能力看到比较远的部分。这点很有趣，在知名的恐怖分子，美国“大学炸弹客”泰德卡辛斯基（Ted Kaczynski）的著作《论工业社会及其未来》中就有提到，在新技术出现的时候，我们是没有能力判断该不该用的，我们只会看到很表面上的利益，以及很表面的“好用”。

这是人性。我不晓得大家觉得现在大家人手一支智慧型手机是不是一件好事，智慧型手机是在 2007 年第一次出现的，但如果现在有机会让我选的话，我会希望这世界并没发生过智慧型手机普及这件事情。因为在我看来，我觉得好像坏处大于好处。

“有时候觉得 AI 所表现出的这种服从性，以及服务精神，就很像女性在漫长人类历史上所充当的这样一个角色……我们被视作没有逻辑，没有灵魂，没有主体性。”糖匪 图：Mantha Mok / 端传媒，由DALL·E生成

科技黑盒：为何AI需要立宪

“技术黑盒不断变大之后，不仅让我们置身于一种我们和每个人都亲密无间的幻觉，同时又把我们置身于一个真正的孤岛上。因为我不再相信我跟你的连接，我跟你的一致”糖匪

糖：这样说可能比较拟人化，就是 AI 到底在想什么，我们可能并不完全知道。在以前的 DOS 电脑年代，我们大概还会知道它的程序编程，它是怎样的一个因果关系，但自从 Windows 开始，所有的技术变得越来越复杂，也意味著它的黑盒也越来越大。我们能看到的只是它带给我们的结果，但我们并不知道这个结果是否正确，它的数据来源以及这个结果是否有私心的。

今天的 AI 技术就给了我一种感觉，或一种幻觉，就是我们无所不能，无所不知。但我们知道的是否就是真相呢？像我遇到的一个上海大学生，他认为 AI 会提供给我们一个公正客观的答案。像我刚才说的，他们就是这种同质化所养成的那一代孩子，他们是看不到数据提供给我们的这些干净的、简洁的、高效的生活背后是一套怎样的一个逻辑，以及这些逻辑背后是出于谁的利益，谁是既得利益者。

我现在特别不相信巧合，以前如果有一人跟我意见一致，我心想，我们可以聊得来，可以做朋友，但现在我会觉得，他是不是看了我的社交帐号，是不是大数据把一样的东西推给我们。这个技术黑盒不断变大之后，不仅让我们置身于一种我们和每个人都亲密无间的幻觉，同时又把我们置身于一个真正的孤岛上。因为我不再相信我跟你的连接，我跟你的一致，反而让人跟人之间的连接发生了问题，也不知道我是悲观还是乐观了。

伊：我觉得技术黑盒的确是个问题。韩炳哲亦有提到，目前的演算法基本上是以归纳法（induction）为主，我们所知道的所谓“卷积神经网络”就是这样，像是社群媒体上，演算法决定了你可以看到什么，也决定了你不能看到什么，它其实已经巨大到人类没办法搞清楚里面的运作机制，长此以后，这会导致人类过度依赖归纳法，过于依赖数据，相对而言，随著演绎法（deduction）地位会下降，我们的推理能力、演绎能力都会减弱。因为我们只需要知道结果，那就不太需要用到理论，不需要用到逻辑思维，都可以产生一个绝大多数人都接受，而且愿意遵从的结果。

目前我们已经身处在这种演算法社会之中，未来继续扩大的话到底会怎么样，我觉得也是为 AI 立宪的一个原因。

“长此以后，这会导致人类过度依赖归纳法，过于依赖数据，相对而言，随著演绎法（deduction）地位会下降，我们的推理能力、演绎能力都会减弱。”伊格言

“要为 AI 制造出一个更类似于人类的自我，答案是它们必须要有身体。给了它一个身体，它的自我，像人类的那种自我才能长出来……但是我反对。”伊格言 图：Mantha Mok / 端传媒，由DALL·E生成

生化人、泛AI是不是公民呢？

伊：我很好奇的是，听了我跟糖匪这样子的想法，启章老师会不会变悲观一些？

董：我没有很悲观呀，也不是不同意你们的看法，当然你们的悲观是有你们的道理，而且悲观可能是一种比较好的态度，谨慎一点，就不会太容易盲目去相信一些东西。不过我有想过的是，对于 AI 未来发展的安全问题，是可以分开五个方向去尝试做一些事情。

第一就是建立 Consensus（共识）跟 Moral code（道德守则），也就是怎么样去用 AI 的伦理问题。比如说，像我们刚才讨论创作里有没有使用 AI，但用在哪里？怎么用？用来帮助你写什么？这些东西慢慢要有共识，成为一个大家共同接受的文化，也是作为你创作的声明。

第二就是立法跟行政的手段，像是格言刚才说的立宪或是法律手段，其实美国也通过了一些法案去限制 AI 的发展进度，这部分很可能就是由政府去做。

第三种就是发展 AI 技术的时候，公司的管理方式能不能有所改变？能不能采用更开放的、共治的方式，让民众能够参与和监察它的运作？而它也不只是向投资者负责，而是向使用者，甚至是公众负责？

第四种就是以 AI 来制衡 AI，也就是格言提到的 Alignment，或者说是 OpenAI 里面已尝试发展的 Super Alignment（超级对齐关系）。也是因为要投放多少资源去发展这东西，它内部才会出现分歧。Super Alignment 就是尝试用另外一种 AI 去评估和监察其他AI，即是让 AI 之间互相监察，让它不会失控、不会暴走的一种的方法。是用科技的方法去处理科技。跟今日的核武有点像，如果只有一个强大的 AI，它能够统治全部人类了，但如果是有很多 AI 而且互相克制，就没有一个 AI 能够完成称霸。

最后的第五种，就是应用区块链的去中心概念了。这种技术，这种社群的建立方法，有没有可能是慢慢应用在 AI 的发展上呢？当然我不能说是很乐观，但应该也是有事可做吧？

“AI 宪法是应该的，但我觉得 AI 宪法并不是单单用来限制 AI 的，它同时是赋予 AI 应有的法律地位，所以也是保障它们的。”董启章

伊：我觉得你已经为我们理想中的 AI 立宪提出了一些很好的构想。

董：AI 宪法是应该的，但我觉得 AI 宪法并不是单单用来限制 AI 的，它同时是赋予 AI 应有的法律地位，所以也是保障它们的。我写过一部科幻小说《后人间喜剧》，里面也谈到 AI 生化人，他们是不是公民呢？他们的公民地位是怎么样的？如果他们拥有某种智慧、某种意识的话，他们应该也有他们法定的地位，是要受到法律保护的。

伊：我完全赞同，是应该要在 AI 发展上掌握主动性。但我还是抱著一些悲观的想法，继续发展下去就会发现，绝大多数的人都不在乎这种主动性，他们会觉得，被 AI 当作宠物也是好的。像《人类简史》的作者赫拉利（Yuval Noah Harari）就说，未来人类的社会阶层依旧是金字塔型的。最上层的人就是知道如何掌握跟利用这些 AI 技术的人，他们就承担起管理任务，而绝大多数的人并不想要那么主动，为自己负责是很累的事，他们还是会在底下被养得好好的。

董：但是呢，这个问题跟有没有 AI 没关系。没有 AI 的人类社会都是这样子啊，都会出现相同的问题。

伊：对，就是掌握新技术，或是掌握那种有生产力的技术的人，他们就会就形成阶层。

董：几乎在任何的时代，任何的不同社会结构里，人类都是同样被动的。不过，也有一些社会结构，能够让每一个人的主动性变得比较多。最近我读到的一本书，就是日本思想家柄谷行人的《力与交换模式》，我非常受到柄谷行人的理论所启发。

他提到人类社会有交换模式 A、B、C、D 四种，A 就是最原始的共同体，即以前的小部族、游民这一种，在这种模式之下，因为规模比较小，每一个人的社会参与及互动都比较强，关系亦比较对等。但人类为什么会变成被动，开始变得被管制呢？就是交换模式 B 的出现，也就是“国家”这个模式。当产生大一统的国家概念以后，人类社会慢慢就变成一种金字塔，就是皇帝在上面，官僚在中间，下面是普通老百姓，而老百姓都非常被动，再没有自己的流动性。

交换模式 C 呢，就是资本主义。资本制度是另一种把人变成被动的，也就是变成消费者的模式。提供劳动，然后消费，只有这两个身份，而这两个身份是在同时存在的，你一边去劳动，然后另一边你去消费，把劳动出来的产品再买回去。所以 B 跟 C 模式都是让人类变得越来越被动的，但现在世界上几乎所有国家，不分什么政治意识形态，事实上都是 B 跟 C 模式的结合。所以柄谷提出了交换模式 D，也就是交换模式 A 的高次元回归。

我们有没有可能重建出一种新的社会关系，不是基于 B 的国家体制，也不是建基于 C 的资本交易，而是回归到古老社会里的一种共同体召唤，让人变得比较对等，权力也变得比较分散呢？柄谷举的例子是十九世纪所尝试建立的早期社会主义，当然它最终并没有成功。

尽管某些政治社会制度是会让人变成被动，但这并不是必然，有可能在另外一种社会关系里，人是可以主动的，就是一种去中心的，没有统治与被统治的关系。所以，我自己的看法就是，以区块链的去中心技术重新建立一种社群跟合作，很可能就是当代的交换模式 D。它一定程度可以针对人变得被动，变成单纯一个被管制的、提供劳动然后去消费的人，让我们从这样的状态解放出来。

“以区块链的去中心技术重新建立一种社群跟合作，很可能就是当代的交换模式 D。它可以让人从变成单纯被管制的、提供劳动然后去消费的被动状态解放出来。”董启章

端：将区块链跟 AI 两种技术发展融合在一起，可以实现交换模式 D 吗？

董：我就是一直在思考这个问题，但不一定要真的实质上去改变整个社会制度，而是在既有的制度下面拥有另外一个空间，让人去实现这种具有自主性的合作和互动。

“我认同它现在写的东西还是很烂，用文学水准来鉴定的话，但在一般应用也就是较实用性的文体，它很可能已经做得很好。暂时来说，会被取代的很可能就是机械性的文字操作。”董启章 图：Mantha Mok / 端传媒，由DALL·E生成

AI文学时代：一个叛逆的孩子？

端：无论是文学奖、摄影奖，还是绘画奖等，过去一年都出现过 AI 作品夺奖而引发争议的情况，以 AI 作品参赛是否不道德？递交 AI 作品而被取消资格又是否合理？

董：但现在香港浸会大学正正就有一个文学奖，它是公开要求大家使用 AI 去写诗，然后把你是如何运用 AI 写诗的过程写出来，所以它是公开、明确地鼓励大家运用 AI 写文学作品。我觉得这很有意思。

糖：在某种意义上就回应了我们刚才谈到的主动性，就是帮助我们主动的去使用这个工具，而且这里面其实就去掉了一种写作的虚荣心，因为很多人都只是想成为作家，但并不想去写作。他们倒是要开始跟 AI 争夺作家的名衔了。像我现在跟 ChatGPT 合作写一部作品，谁是真正的第一作者？这可能会变得非常复杂。

董：背后就是一个版权的问题。到底 AI 创作的东西有没有版权，它利用学习别人的东西来创作，它有没有侵犯别人的版权？

我是主张版权应该要尽量开放的，传统的版权概念只是把知识封锁，无法让它自由流通。譬如有人拿我的作品给 AI 学习，然后生产一些新的作品，会不会侵犯了我的版权？我真的无所谓，就让它去做。

端：用 AI 创作“骗过”评审的得奖作品，能否称为一部好作品呢？

董：我看将来会产生一种新的美学考虑，就是什么作品才算是把 AI 用得好，你不懂用，用得不好，然后你只是生产出很差的作品。但如果用得好，里面是会有技巧的，也会有它的艺术判断。

伊：我觉得，站在我们人类中心或者是作者中心的角度，如果 AI 写得比我好，那就只好认了，我们要认输。但问题在于何谓写得好呢，到目前为止，判断写得好或不好的还是人类啊。如果未来我们可以开放让 AI 当评审的话，那就触及一个更深的问题，人跟 AI 要如何合作？假设现在人类作者都觉得用 AI 帮助写作是没问题了，那用 AI 帮助评审又如何？

当 AI 继续发展下去，它会产生什么样的自我，而这样的自我会产生一套什么样的审美标准，目前我们还比较难想像。

“小说本身也跟人类一样，它有它的演化方式。如果在一种实验性的文本里面跟 Chat GPT 合作，哪怕是加入它生成的烂文本，应该还是会有蛮好玩的东西吧？”糖匪

董：但这种状态是无可避免的，而且一定会发展到这种局面。但我们就是要准备面对嘛，现在还没出现一套新的美学判断，可能要经过一段时间，很多尝试和错误才能整理出来，但它是无可避免的，逃避也没有用，AI 创作普及是一定会发生的，只是现在还没有做得很好，不是单纯叫 AI 写出什么，而是由一个有判断力的人运用 AI 创造出一些真是很有意思的东西。

我们常常觉得 AI 越来越强，走得很快，我们就在后面一直追不上，但我们可以倒转这个思维，因为我们可以有一些东西是比 AI 走先一步的，就是我们先设定一些东西，从我们今天的美学判断出发，用我们的价值观和经验设定一些审美规则，然后让 AI 学习。

AI 是学习人类的东西嘛，所以，它写的东西很烂，就是因为它接触到的大多数都是很烂的东西，那我们要让它学习一些我们觉得有艺术思考、有意义的东西，而不是让它一直学烂东西。觉得自己还是一个不太烂的作家，不太烂的艺术家的那些人，现在开始都应该立即主动用 AI 创作，做出一些好的作品，然后大家形成一个共识，啊，原来 AI 跟人类共创的东西是可以去到这个美学程度的，而这个程度称之为好，这个审美标准是不会一下子消失，它会一直延续下去。

反而我们现在不去做，所有好的作家、艺术家都只是说我们不懂 AI，觉得 AI 是垃圾，那结果就是让它发展出更垃圾的东西了。所以是我们让它变成垃圾的，我们不要让它变成垃圾，所以我们就得主动一起让它变成更好的东西。不要总是想像下一代会变差，变到很不行，不如我们就先做出一些好东西。

端：但目前有成功透过 AI 写作做出一些你们满意的好作品吗？

伊：我本来都有这个想法。在 GPT-4 刚推出的时候，我很兴奋的立刻去试用了，也希望它能够帮助我去写作。到现在却再没有大量使用，主要就是因为它真的写的太烂了，那如果它的参数变得更多，可以写得更好，那其实是有一定的应用空间。

董：情况跟你一样，刚开始的时候我也尝试过一段时间，后来再没有用它，就是觉得我自己写还比它好太多了，但我觉得是可以应用得更局部的地方，可能不是小说里面的一个部分，而是需要某一种的文体啊，或是某一种独特的反应啊，那就跟它互动一下，看它会写出什么东西。这些局部的东西都是你作品的一部分，然后从中慢慢去找它的可能性。

我觉得它里面肯定有很多东西是我们没发现，甚至是连研究人员都不能完全把握的部分。当然它自己也不知道自己在写什么，我们都说它没有这样的思考能力，只是一堆文字游戏，像刚才所说，是不是也有一种无意识埋藏在 AI 里面？如果能够找出来，可能会是很奇妙的一个语言宝藏。

糖：我觉得小说本身也跟人类一样，它有它的演化方式。如果在一种实验性的文本里面跟 Chat GPT 合作，哪怕是加入它生成的烂文本，应该还是会有蛮好玩的东西吧？

董：你可以用 AI 写一个小说，然后小说里面就有一个小说写得很烂的作家，哈哈。

糖：我还蛮期待养成一个叛逆的孩子，这种跟我们截然不同的一种逻辑。

“将来如果 AI 的技术更普及，或是有更多开源的方法，每一个普通人都能够参与其中，把自己心目中真是有价值的、好的东西做出来，然后去传播它”董启章

董：事实上如果我们愿意冒险，看看 AI 这东西最终能够发展到什么样模样，我们就应该让它变得叛逆，让它发展出我们不能控制的因素。但当然，这也非常危险。

在《Pluto》（浦泽直树原著）这部动漫里有说到，当 AI 发展超越某一个程度的时候，科学家就要将一些负面的因素加进去，像是仇恨，或者恐惧之类的东西。这些东西，本来我们觉得 AI 是不应该拥有的，因为这会令它变得非常危险，有可能出现灾难性的局面，但当然，这部动漫还是非常乐观的，认为通过仇恨和愤怒，这些人类的负面情绪，反而可以令 AI 成长，变得更像人类。我觉得跟 ChatGPT 对话的时候，有时很没趣的地方就是它的反应非常正面，它没办法承受一些奇怪的、不正常的负面东西，然后当你一直把负面的内容灌进去，它就会停机暴走，它就会疯掉。

糖：在科幻作品里，我们想为机器注入负面情绪，但是在人类现实生活中，我们却希望从自己身上剥离这些负面情绪。这两个对比还蛮有趣的，也就是让我想到，也许未来机器会变得像人，人会变得不像人。

端：但就算我们乐于建立一套文学审美标杆，它很容易就会被今日的演算法机制所淹没。比如抖音短片和内容农场的文章，甚至经营网络媒体，如今都需要迎合演算法的喜恶标准而作出改变。在 AI 所执行的另一套运算逻辑里，或者点击率和传播速度就已经是它的审美判断？

伊：因为人类就是这样，我们可以说是庸俗啊，如今我们已经把很多事情、很多判断都外包给演算法了，这些演算法是有黑盒的，人类已经再没有能力去好好检视里面的每一个参数，而且我们绝大多数人都接受了它。

在 20 年前，我们可能还会觉得人类的思维、推理和逻辑能力有著比较高一点的地位，但智慧型手机的传播速度实在太快了，而 AI 的普及也加剧了这些落差。

“我赞同我们要主动去做一些事情，但是好像不会成功，哈哈。我觉得我们最后应该是会败下阵来。”伊格言

董：但这不单纯是 AI 的问题，人类社会历来都是这样的，比较通俗的，不是特别好的东西，往往是最受欢迎的，而你觉得非常有艺术性的东西，都会比较少人欣赏啊。当然 AI 会让这个情况更严重，因为它的传播影响能力更大，危机是存在的，但我们也不能坐以待毙，还是要尽量做出好的东西去抗衡它，而不是坐著就让 AI 变成一个很坏的东西。

将来如果 AI 的技术更普及，或是有更多开源的方法，每一个普通人都能够参与其中，把自己心目中真是有价值的、好的东西做出来，然后去传播它……我并不是说我们什么都不做，它就一定会变得很差，反过来说，即使我们立即去做一些好的东西，它可能同样也会变差，但至少会没那么差，或是没那么快变差。我觉得这不是乐观，而是必须要做的事情。

伊：我赞同启章老师的做法，但我觉得，他说的是一个没办法之中的办法。那是对的，我赞同我们要主动去做一些事情，但是好像不会成功，哈哈。我觉得我们最后应该是会败下阵来。

尾语画外音

（虽然今次对谈照片和录音逐字稿都由 AI 代工，也就是说，当我们在讨论人机一体的时候，它本身已是某种意义上的人机一体；但今次对谈内容及文字整理，都没有 AI 参与其中。或者不久以后，更进阶版的文学 AI 会成为对谈作家之一，甚至取代了我这个撰稿人的角色。但到时候仍然有人会想看 AI 撰写、由 AI 对谈的关于 AI 创作的讨论吗？到时候可能已没有讨论意义。）

[# AI写作 # midjourney # Open AI # AI监管 # AI编剧 # ChatGPT # AI](#)

本刊载内容版权为端传媒或相关单位所有，未经[端传媒编辑部](#)授权，请勿转载或复制，否则即为侵权。

端传媒的下一程，需要你的守护。今天就成为订阅会员，支持我们走下去，支持华文世界不可或缺的深度报导和多元声音。点击了解更多[會員計畫](#)