



人工智能时代人类的繁荣和存在的基础

 演讲者 / Shimon Edelman

首先，我要感谢主办方邀请我参与这次活动，整个过程让我感到非常兴奋。

本次大会邀请我分享的题目是《人工智能时代的人类繁荣与存在根基》。事实上，光是这个标题中的第一个词就已经引发了我的思考。

我想先谈谈“人类繁荣”这个概念。或许正如画家构想的那样，理想境界里人们在湖畔演奏音乐，周围的景色与运河这一侧的景观并无二致。然而，这样的描绘似乎缺失了某些关键元素——它与我们面临的现实并不相符。网络上流传着一种说法：在当下共同的经济环境里，人类的共存本身似乎已成了一种奢求。这是一个极具时代特征的迷因。有些论文将其描述为某种奇迹，比如欧洲文献中定义的“幸福”。但在当前经济环境下，不仅是经济状况本身，各种现实因素都与任

何形式的繁荣或反繁荣现象相抵触。这便是当今世界的真实图景：它被描述为一场政治危机，由多重灾难交织而成——这些或许源自某些权威机构或个体制定的规则，但它们无疑构成了全球民众面临的最大困境。

让我用几组数据加以说明。我将主要聚焦于我的祖国，美国。虽然同样的问题也存在于其他国家，但我希望以切身观察展开批评。贫困数据屡见头条：美国有 80% 的人口面临严重的贫困和失业问题。每当有人谈论繁荣时，这都是我们需要谨记的现实之一。另一个因素则是警察暴力——更准确地说，应称为国家暴力，因为警察受国家控制，而国家又受权贵集团操纵。清晰揭示了世界上大多数人对自身处境的理解，以及当抗议公开化时他们如何被镇压与压迫。首先，美国警方的口号是“服务与保护”，但事实上他们从未真正服务或保护过本应守护的民众。注意背景中戴圆顶礼帽的人——他们是出资供养警察的纳税人，却如此无助。这不仅是历史问题：在上一届政府时期的得克萨斯州抗议活动中（这张照片实际摄于 2020 年），所谓的民主党执政下，警察正是这样对待民众的。

而这张是上周新共和党政府下的场景——一个公开奉行法西斯主义的政权。我们的国家正在变成这般模样：军队被调遣来镇压公民，即便他们没有参与抗议，尤其当他们的肤色或财富状况不符合某些标准时。接下来要讨论的是即将掩盖所有其他问题的重要议题：气候灾难与气候不公。这张图表按县划分了气候变化带来的经济影响，可见美国南部较贫困地区受气候灾难冲击最严重。左侧图表显示相对于县收入损失的分布，最贫困的县承受了最大伤害。我们不得不质疑：

“在这种环境下，我们真的能自由讨论最令人忧心的问题吗？”若追问气候变化的根本原因，答案显而易见。一幅具有广泛影响力的著名漫画中，部落前商人对围坐篝火的孩子说：“是的，大火毁灭了一切，但在那美好的瞬间，我们为股东创造了许多价值。”这便是资本主义制度的罪魁祸首。这幅 1911 年的漫画至今依然切中时弊。资本主义体系正是导致全球大规模贫困、不安定、营养不良与饥饿、战争、内乱、致命疾病、社会疏离、国家压迫与暴力以及气候灾难的根源。问题出在资本主义制度上。我会快速浏览接下来的幻灯片，如果时间允许我们可以就此展开讨论。

首先关注资本主义阶级关系这一点：它造成永无止境、无限延伸的人类苦难。应主办方要求，我需要花些时间定义“痛苦”（suffering）。我认为这是本次会议尚未提及却至关重要的概念。以下是对痛苦的简要阐释：痛苦是一种无法仅凭主观厌恶就能摆脱的负面状态。其根源由进化强制塑造，并不取决于我们自身——而是源于我们的生理构造与进化历程。我们关注结果本身，而非像纯信息处理系统那样仅仅偏好某些结果。我们因这种与生俱来的特质承受负面效应，且无法随意消除它。因为若能随意消除，进化早已将我们淘汰出基因池。若消除痛苦的行为可被选择，它就不再是痛苦，也不再是促进行为改善的有效目标或刺激。意识的自我目标是促进学习。众所周知的强化学习离不开批判过程，但真正驱动变革的是批判性行为。这不是任何人的过错。因此，



意识自我被进化塑造为责任主体，而非归咎对象。这便是痛苦的进化根源。

托马斯·梅辛格（Thomas Metzinger）的意识理论表征中提出了构成意识的四个必要条件，它们共同作用时苦难才会发生。我将暂不展开这四点，后续可继续探讨。此处的核心观点是：若人工系统满足这些条件——这也正是本讨论与人工智能的关联所在——那么该系统将具备承受苦难的能力。

回到资本主义的问题，我们需要区分不可避免的苦难与可避免的苦难。前者内在于意识本质，由进化塑造和学习过程引起（正如所言，学习过程最有效时往往伴随苦难），也源于衰老以及人类对死亡的清醒认知——这是我们终需面对的必然。但除此之外，还存在大量可避免的苦难，其根源皆指向同一个庞然大物：资本主义。社会动态、战争幻想、贫困、饥饿、可治愈疾病、政治压迫和气候灾难皆源于此。我将把资本主义的其他问题留待后续讨论。需要强调的是，资本主义由国家暴力支撑，无法凭愿望消除。若想做出改变，单打独斗绝无可能。古时若对恺撒的统治不满，尚可匕首相向；但如今已无单一恺撒可除。权力顶层的每个人都是他人的血肉同盟。无论如何，你都会遭到统治阶级上层代理人的武力镇压。正如我喜爱的科幻作家厄休拉·勒古恩在1974年出版的《一无所有》中所写：“个人无法与国家谈判。国家只认一种货币，那就是权力，而货币正是它自己发行的。”

现在让我们转向人工智能。如同任何先进技术，AI 服务于统治阶级。接下来几张幻灯片的思想源自大卫·格雷伯的研究。他是研究资本主义的人类学家，此处是他的论文标题，其副标题尤为重要——“为什么资本主义是奴隶制的变形？”。我对此有些见解。另一篇是我几年前发表的论文《有意识的人工智能就是人工奴隶制》。构建有意识 AI 的伦理问题今早尚未被讨论。据我追踪的发言，所有伦理探讨都聚焦于如何压制、控制 AI，使其符合我们的伦理标准，最大担忧似乎是防止 AI 系统伤害人类。但若 AI 系统具有意识，那么人类对其造成的伤害——甚至仅仅创造这种行为本身——又该如何看待？我将快速回顾科幻文学中的例子，并讨论艾萨克·阿西莫夫的机器人三定律。

关于人工造物与 AI 的伦理，我们都熟悉这些故事：玛丽·雪莱 1918 年创作的《弗兰肯斯坦》是早期科幻 / 恐怖小说之一，描写了一个人工造物；犹太民俗中的泥人戈伦，被注入生命并承担社区任务；第三是卡雷尔·恰佩克的《罗素姆万能机器人》（R.U.R.）——该领域最著名的作品。恰佩克作为捷克作家发明了“机器人”一词。这些造物在各自文学作品中皆属同类：被剥削、遭暴民追杀、被迫暴力自卫，且都获得作者同情。但现实截然不同：AI 伦理理论家几乎完全忽略机器人可能需要保护以免受创造者伤害的可能性（托马斯·梅辛格是少数例外之一）。这种显著的不对称性有个简单解释：AI 工程的主要（虽很少明言）目标是创造人类使用的工具，而非伙伴或同行。现在看机器人三定律，其创立者阿西莫夫在 2001 年访谈中说：“这些定律的类比隐含在



几乎所有工具的设计中。”我强调“工具”一词。他的解释是：第一定律，工具必须安全可用；第二定律，工具必须高效执行功能；第三定律，工具在使用中须保持完好。

我敬佩的另一位科幻作家斯坦尼斯拉夫·莱姆对意识 AI 的可能性更具洞察力。他在《想象幅度》中写道：“可惜我们缺乏一种伦理算术，无法通过简单加减在构建地球上最开明精神时判定：谁更混蛋？是我们还是它？我想应该是我们，因为我们将创造的不仅是必然承受苦难的意识体，更将其塑造成我们的工具。”实际上，将一个意识体用作工具有一个特定术语：奴隶制。这包括资本主义下的工资奴隶制。大卫·格雷伯在此指出，现代资本主义实则是奴隶制的更新版：从前是人出售或出租我们，现在是我们出租自己。在此背景下，关于被创造并用作工具的 AI 系统之命运，这个迷因非常贴切：顶部显示 AI 伦理学家试图拥抱伦理 AI 的概念，但却受限于现实——若你的薪水来自统治阶级的荷包，你便不该批判资本主义或阶级制度。我之所以能这样做，仅因我有终身教职且地位相对稳固。

根本解决方案在于通过首先取缔资本主义政治经济制度，其次以人民繁荣所需替代之，来预防可避免的苦难。唯有当 AI 发展成果有助于推进这份人文议程时，投入任何研发努力才具有正当性。最后，流行语言说“系统已坏，必须修复”——我不同意此说。我更认同“系统正按设计运行，唯一出路是摧毁它”。我想提示一个更恰当的表述可能：这幅来自墨西哥恰帕斯州的绘画展示萨帕塔运动维护的自治体系——它拒绝资本主义，并真正拥抱人类权威。另一个世界是可能的：一个包容所有世界共存的世界。感谢各位的聆听，我欢迎你们的提问。

张伟特：谢谢 Edelman 教授，接下来让我们与他进行讨论。

Müller：西蒙你好。我想知道这两个论点之间有什么联系。资本主义是一环，而意识是一环。您指出，将人工智能视为没有意识的实体是某种奴隶制。你可能知道乔安娜·布赖森的一篇论文说机器人应该是奴隶。我认为问题是，这两者之间有什么关系？如果我承认了你的资本主义批判，我是否必须承认另一部分，或者反过来，或者他们都是独立的问题？因为在我看来，他们是独立的问题。

Edelman：你的相机、你的电脑、你的笔记本电脑，那些都不是你的奴隶。但是，根据定义，有意识的工具是奴隶。资本主义是我们生活的体系，是我们赖以支付账单的系统，这是决定我们研究领域的系统。这就是这两部分之间的联系。我认为它们是不可分割的。主宰世界的政治制度就是这样，它推动利润。有一种强大的推动力使 AI 变得高效，尽可能有效。要做到这一点，

一个万无一失的方法就是赋予他们意识。我认为这并非不可能。如果这是可能的，那么我们必须保持警惕对人造生物的剥削。他们将被剥削工人的同一批人剥削，和我们一样，你是雇佣工人，我也是雇佣工人。我们都为我们所支付的工资劳作。

Müller: 所以，我可以相信机器人会有意识；但我也可以相信对资本主义的批判，那么这些机器人将只能成为绝对的工具，就像我所拥有的其他无意识工具一样。从这个意义上说，人工智能将无路可走。

Edelman: 是的。

王珏: 教授您好，您在报告中提到 AI 是为统治阶级服务的。我想知道是否存在例外？比如 AI 医疗或 AI 医生是否对我们有赋能作用——它们帮助我们更高效地获取医疗信息。

Edelman: 感谢您的提问。是的，当然可以将 AI 系统用于此类积极场景。但关键在于厘清 AI 系统的具体贡献方式：这与人类医疗专业人士的工作存在本质差异。我们需审视其合理性——是否值得为了一套官方系统而取代人类医生？在美国（尤其是医疗私有化体系下），AI 的医疗应用无一例外对患者产生了极其负面的影响。所谓 AI 系统主要被医疗保险机构（即医疗商业集团）用于审核患者理赔申请，并试图拒绝尽可能多的索赔。

您可能注意到某些辅助医生诊断的 AI 系统。若其运作模式类似传统的专家系统，或许尚可接受。但一旦市场利润介入，一切都会遭到腐蚀。在诊断过程中，AI 系统特别容易受到利益相关方干预，其脆弱性甚至可能超过人类医生——严肃对待希波克拉底誓言的医生或许会抵抗院方指令，而预设程序化的系统可能不具备这种抵抗能力。归根结底，由于主导的政治经济体系会腐蚀一切，医疗 AI 系统不仅同样易被腐蚀，甚至可能比人类更甚。

张伟特: 在您看来，在人工智能技术与人类历史上的其他技术中，如火的技术，铁的技术，铜技术，它们之间有什么区别，以至于它只能为统治阶级服务而不是改善弱势群体？在中国的许多偏远地区，我们使用在线教育分享很多优秀的教育资源给山区。我有一个学生开发了一个面部识别系统然后在非洲应用。如果我对我的手机拍照，我可以诊断我的身体状况，与此同时，我知道该买什么药。大大改善了非洲地区的健康状况这是第一个问题。为什么会是人工智能技术对弱势群体的危害更大并且只为统治阶级服务？第二个问题是：如果人工智能技术具有自我意识他们会不会反抗资本主义的种种恶习？谢谢。



Edelman: 第二个问题比较容易回答。应该教育有意识的人工智能并将其收编到抵抗（资本主义）的运动中。

关于第一个问题，我想到的一个例子是打印或复印技术。我在前苏联长大。其中每台打字机，每个绘制机，复印机必须注册在案。因为印刷很容易变成颠覆性的材料。所以你可以说，打字机怎么了？它让创建文档的工作更轻松。但是，当这样的技术变得可用，在特定制度下，工人阶级却不可用，公众也不可用。这样的例子每个时代都有。事实上，大卫·格雷伯几年前与大卫·温德罗合著的一本书，关于人类社会悠久的历史，社会经济力量互相竞争。（《人类新史：一次改写人类命运的尝试》）很多例子可以在正统的人类学和历史背景里找到。所以我认为关于任何技术，底线永远相同。自从人类开始拥有统治阶级，他们掌控了每一项技术发展。