



北京大学

习思想概论 课程论文

科教兴国、人才强国之我观——AI 大爆发环
题目： 境下的科技发展现状与挑战

花名册序号：	40
姓 名：	雷逸鸣
学 号：	2300011454
院 系：	物理学院
主管老师：	曹得宝
助 教：	赵方

二〇二四年 5 月

摘要：2024 年 5 月 OpenAI 发布了其最新的多模态 AI——GPT-4o，其响应速度、对幽默的感知以及处理大量信息的能力再次震惊世界。^[1]创新始终是推动一个企业、一个国家、一个民族向前发展的重要力量。2015 年两会期间，习近平总书记在参加上海代表团审议时首次提出“创新是引领发展的第一动力”的重大论断。^[2]在 AI 大爆发的今天，我国目前在以芯片为基础的人工智能领域发展状况如何？未来路又在何方？接下来，我们将共同探讨以上话题。

关键词：芯片 AI 习思想 创新 科教兴国、人才强国战略

一、引言

2022 年 11 月，OpenAI 首次发布了 ChatGPT 大模型，5 天后便有了 100 万用户；上线两个月后，更是用户破亿。^[3]紧接着，OpenAI 在 2023 年又相继发布了 GPT3.5 和 GPT4.0。2024 年 2 月，OpenAI 的 CEO Sam Altman 更是提出了要筹集 7 万亿美元研发芯片的战略。^[4]从 GPT 到 DALL·E 再到 Sora，OpenAI 一次又一次地刷新着人们对人工智能领域的认知。原于人工智能的发展，在不远的未来，AI 极有可能成为最大的生产力。

正如习近平总书记强调的那样：“我国经济已由高速增长阶段转向高质量发展阶段，正处在转变发展方式、优化经济结构、转换增长动力的攻关期，迫切需要新一代人工智能等重大创新添薪续力。”^[5]发展人工智能，将为我国构建现代化经济体系、实现高质量发展提供重要支撑。而芯片又是人工智能得以高效发展的基石。我们将从芯片发展出发，以光学芯片为例，发表我对于我国目前的芯片行业发展的一隅之见，继而通过对脑机的分析，探讨我国人工智能领域的发展方向。

二、集成电路发展现状

1 集成电路现状概述

在近 60 年里，人们通过缩小集成电路器件面积的方法，在相同尺寸的晶圆上制造出了拥有更多电路的芯片，达到了降低成本和提升性能的目的。现在，我们已经进入到 5nm 技术节点，在向着 3nm 节点迈进^[6,7]。

在这些年的不断发展过程中，人们在金属氧化物半导体场效应晶体管（MOSFET）的基础之上，通过发明 FinFET，论证了实现 20nm 以下节点硅基集成电路的可行性^[8]。自 FinFET 开始，人们的目光转向了集成电路在物理上的三个维度。由此，人们已经发展出了器件级三维集成、电路级三维集成、芯片级三维集成等方向^[9]。

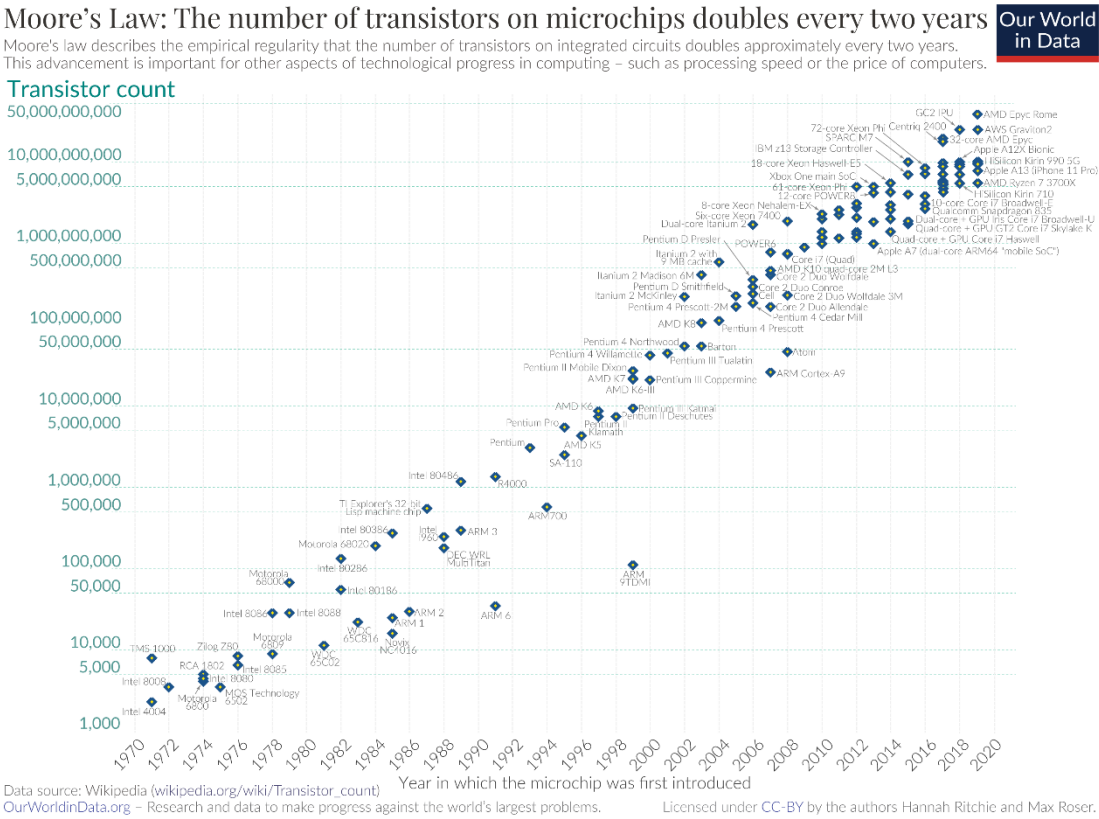


图 1 摩尔定律——集成度与发展时间关系图

但是，集成电路业终将面临器件尺寸无法继续缩小的问题^[10]。如今，集成电路业正从缩小尺寸提高性能的方式向着增大芯片面积来提高集成度的方式逐步转型，但这同时也带来了如何提高良率的问题^[11]。因此，集成电路业必然要从传统的 CMOS 大规模集成的形态向其他的技术形态转变，人们需要寻找新的发展路径、应用方式、使用场景，这也意味着我们将要进入“后摩尔时代”^[12]。

2 人工智能计算机

传统的计算机一般是基于逻辑电路，根据人们事先编写好的代码运行特定程序的机器^[13]。随着传统计算机性能的提高，一套全新的基于神经网络的人工智能算法进入了大众视野。这一算法不断地发展，其对算力的需求也不断提高，计

算机的算力逐渐成为制约其发展的主要因素。为了满足人工智能算法对算力的需求，人们开始着手寻找一种能适应人工智能算法的新型计算机。

2.1 基于传统计算机的尝试

人们最先想到的便是基于传统计算机的冯诺依曼体系进行改良。而人工智能算法（神经网络）的特点是大量的并行简单运算和参数的存储与调用，因此改良的核心在于如何在并行计算的基础上尽可能地减少数据存储调用耗费的时间。于是，人们发明了存算一体（存内计算）技术（例如：RRAM、PCM、MRAM、FeRAM）。在过去的 20 年中，存内数字计算一直专注于识别有更低能耗和面积消耗的新型逻辑门概念。使用纳米磁体进行数字运算、量子点等概念已经通过实验证实，二进制电阻开关器件也在数字计算中表现出了多种优势（例如通过互连线路直接接入电路、通过电信号重新配置电路的能力、纳米级的小型化等）^[10]。

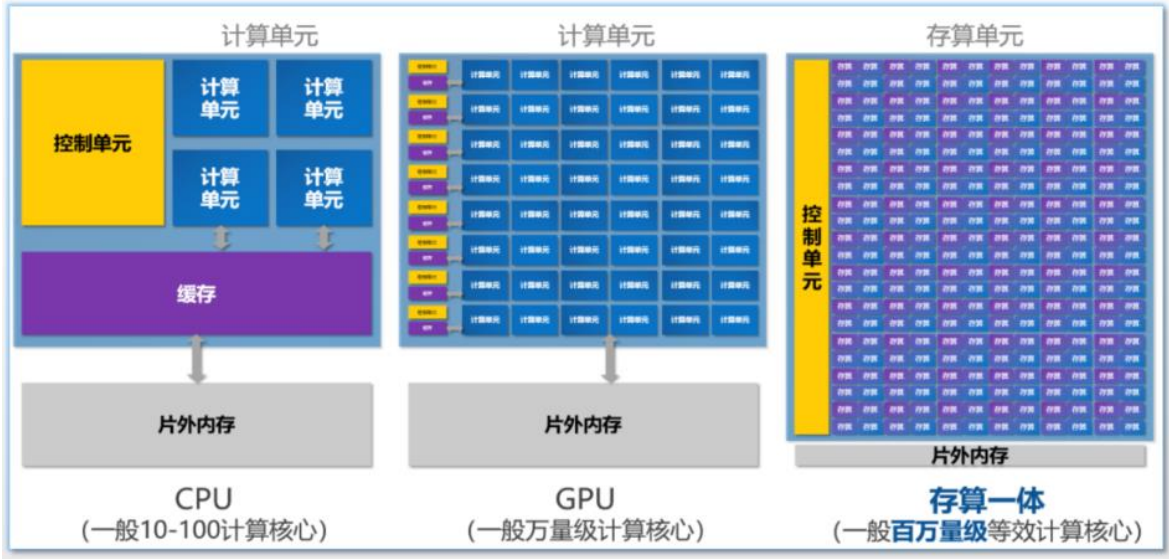


图 2 存算一体与 CPU、GPU 核心量级对比图

与此同时，存内计算也面临着一些挑战。例如：由内存不稳定性带来的计算准确性问题。由于其无定形态的亚稳态性质，电阻会随着时间的推移而漂移^[14]。

种种这些挑战依然等待着人们一一破解。

2.2 基于光的波动性的光神经网络

光的干涉行为的空间构型与神经网络的逻辑构型有很大的相似度。光的波动性给予了我们在其传播过程中进行数据的运算从而提升网络计算性能的可能性。

为了能够用光学信号完整的实现神经网络的功能，我们需要在信息调用、运算和输出等方面构造与神经网络模型对应的光路。

3 光神经网络

3.1 信息在光路中的存储

光携带的信息包括光的相位信息和振幅信息。因此，我们可以选择将信息存储在光的相位中或振幅中。对于前者，我们可以使用分振幅反向设计光路，使得在光经过后，所分出的两路光振幅之比恰好与我们希望存储的信息一致；而对于后者，我们则可以通过控制温度，改变不同光路介质的折射率，间接达到调控光的相位的目的。

3.2 信息在光路中的运算

信息在光路中的运算主要通过不同光路的光之间进行相互作用完成，达到这一目的的方法有很多种，例如：

通过在两路光波导间添置耦合环，使两者发生相互作用^[16-18]；通过在两层神经云中添置开放光场，直接让多束光发生干涉^[19]；在光路中加入光电转换器，通过电学方法在运算中加入激活函数，并达到增强光信号的目的^[20,21]；使用电磁感应透明度（EIT）实现用纯光学元件实现激活函数^[22]（EIT 是一种在介质中发生的相干光学非线性现象，其中一束光的透明度被另一束光控制）；通过反向设计使光在通过元件后光强分布达到预先训练模型的标准^[23]；通过柱面透镜实现光路中的张量运算^[24]。

3.3 信息在光路中的输出

信息在光路中的输出主要通过光电转换器将光信号转化为电信号。

对于将信息存储在振幅中的情况，可以直接将光波聚焦在接收屏上，读取光的强度；对于将信息存储在相位中的情况，我们可以通过在光进入光路前分离半个振幅的光强，并将其投射到接收器处作为背景光强获得光的绝对相位。

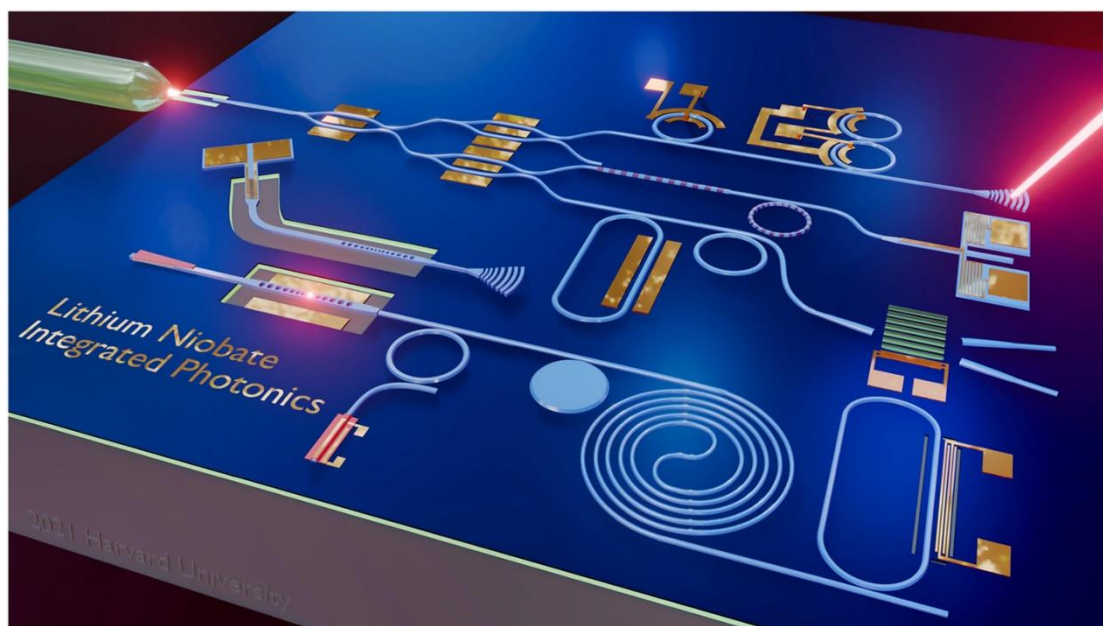


图 3 铌酸锂集成光学芯片示意图

4 芯片发展的可能新方向

在后摩尔时代，光神经网络已在不同的人工智能模型中得到广泛应用，但他仍面临各种挑战，其性能明显优于传统电子处理器的实际应用仍未得到证明。例如，如何改进材料性质及工艺使纯光学激活函数能够集成在光学芯片上；如何在低功耗芯片上实现高速大规模可重构性；如何设置光路结构使得可以在光学芯片上实现反向传播，而不仅仅将以训练好的模型刻蚀在芯片实体上实现。^[25,26]

同时，光神经网络也是一个富有发展潜力的领域，他是一个有望超越冯诺依曼体系实现与人工智能一对一适配的技术。他对于神经元的模拟不仅仅是数值上的，他可以为神经网络提供基于光信号的模拟平台，实现快速、高效的信息处理与传递，满足人工智能对算力的迫切需求。

三、脑机

1 脑机发展背景

2024年1月29日，马斯克在社交媒体上宣布，其脑机接口公司（Neuralink）已完成首例人类大脑设备植入手术，接入者“恢复良好”。^[27] 消息一出，支持

与反对的声音接踵而至，脑机在理论层面上是否可行，在道德层面上是否合理，数字生命的到来还有多久，等等话题在网络上被不断炒作，众说纷纭。“脑机接口火了，但有“钱景”吗？”，“脑机接口研究先驱：马斯克投资脑机接口是营销，那条路走不通”网络上各大营销号也纷纷抓住热点，发表观点。

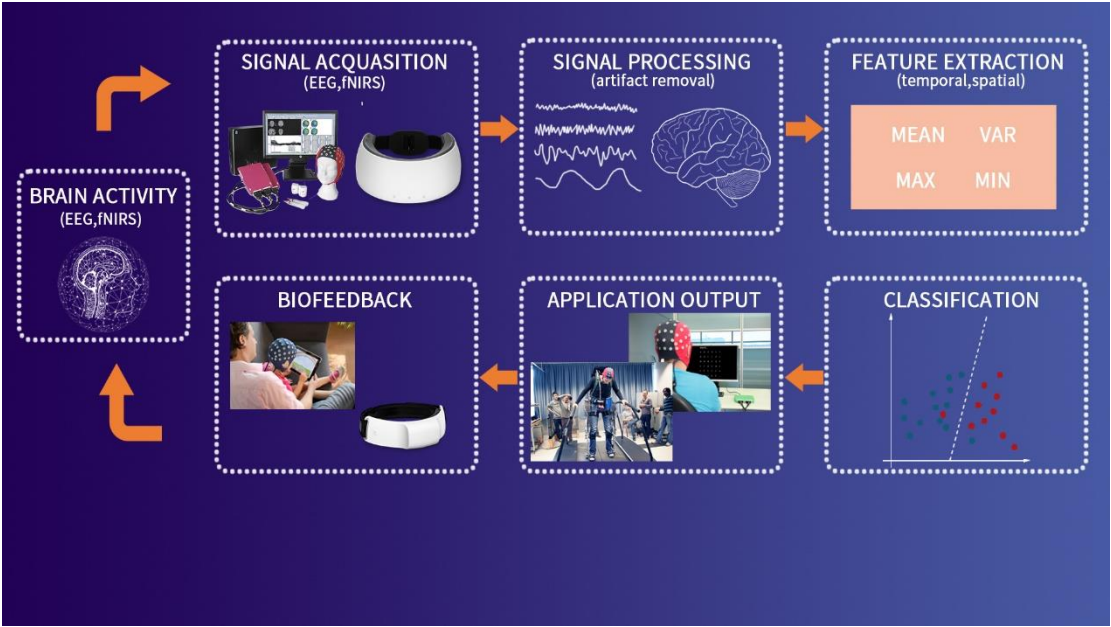


图 4 脑机技术实现路径示意图

根据中国电子技术标准化研究院的《脑机接口标准化白皮书 2021》，全球脑机接口市场规模在2019年约为12亿美元，并预计在2027年达到37亿美元，年复合增长率达15.5%。而中商产业研究院发布的《2023-2028年中国康复机器人行业分析及发展预测报告》显示，2022年中国康复机器人市场规模达11.1亿元，同比增长13.3%。中商产业研究院分析师预测，2023年康复机器人市场规模将进一步增长至13.9亿元，同比增长25.2%；2024年市场规模将超20亿元。

可以发现，以脑机为代表的人工智能市场在社会上分担着相当大的一部分市场规模，同时，大众对该领域的认知仍存在着较大空白与误解。这一来势凶猛的“科技热潮”究竟是什么，接下来让我们共同剖析。

2 一切的起源——人工智能技术

1950年，计算机科学家图灵提出了“图灵测试”。自此，机器的智能拥有了一个客观的衡量标准。1956年，“人工智能”被提出，成为了一个全新的计算

机研究领域。沃伦·麦卡洛克和沃尔特·皮茨在 1986 年提出了神经网络的计算模型^[28]。经过 20 年的发展，于 2006 年，该模型通过深度学习被用于大数据挖掘领域。

2.1 神经网络模型

神经网络模型来自神经信息处理单元——神经元，该单元能够对外界刺激做出反应，并调节自身状态。神经网络意在构建大量的神经元，使其产生社会化行为，通过在模型中加入大量的参数，达到拟合任意函数的目标。

在神经网络的视角下，视频、音频以及文字信息等内容，均可被看作是函数的输入或输出（高维张量）。而对模型的训练过程，便是通过大量的数据对网络参数进行调整，达到输入数据经网络的作用，输出我们希望得到的结果。由于神经网络是基于大量参数的拟合训练完成的，对于复杂的神经网络，对于其有效性的机制往往不可解释。

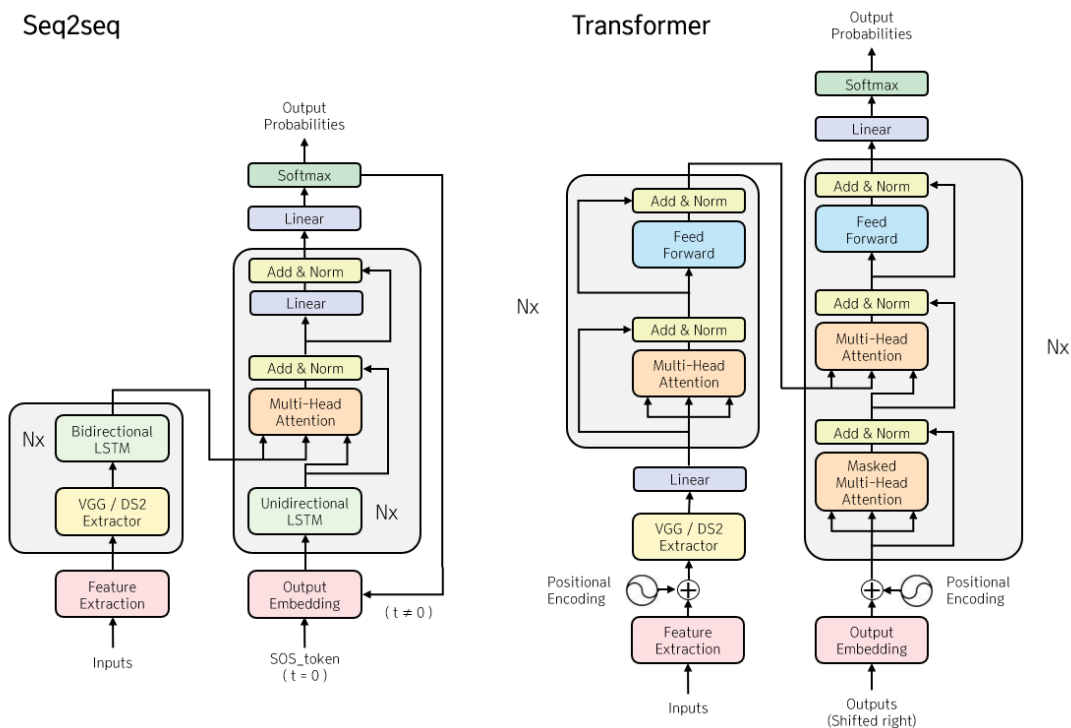


图 5 著名自然语言处理大模型 ChatGPT 所使用的神经网络架构——Transformer

在现代科学技术领域，神经网络为诸多领域做出了重要贡献。例如，2016 年 Google DeepMind 开发的深度神经网络在围棋领域击败世界冠军李世石；优素福·纳德，卢克·法里托，朱利安·席利格使用神经网络破译了 1900 多年前，

因维苏威火山爆发碳化的纸莎草卷轴；近年来，由 OpenAI 开发的 GPT 模型也在不断地刷新着我们的认知。而我们将要讨论的“脑机”也是其中之一。

2.2 脑机

脑机接口（BCI）是在生物体与机器设备间构建的信息通路。也就是说，我们可以通过脑机接口达到将指令信息由生物体传达至电脑，也可以通过电脑获取周围环境的信息（如听觉，视觉信息）将其传递给生物体。

脑机的信息获取方式可以分为两种：侵入式脑机接口和非侵入式脑机接口。侵入式脑机接口通过将电极植入到颅腔内，直接与脑神经进行信息交换；而非侵入式脑机接口则是在颅骨外安装电极，并通过收集外界信号分析脑神经信息。

3 脑机接口的可行性分析

目前，脑机有几大应用领域：以面向辅助运动功能为代表的信息传出型脑机，以修复、重建感觉功能（听觉、触觉等）为代表的传入型脑机等。

3.1 可行的标准

对于传出型脑机，我们可以将是否达到了辅助运动（如：机械臂是否可以被残障人士使用达到弥补缺陷的目的）视为验证脑机接口是否可行的标准。由于信息是向外传输的，我们可以直接通过结果（帮助了使用者行走、辅助了使用者进食）验证其可行性。而对于传入型脑机，由于其目的是让使用者感受到外界信息，而使用者的感受是无法直接获取的，因而对于其可行性的无法给出直接的判断。（例如：通过语言，我们无法判断使用者通过脑机接口观察到的颜色与“正常人体”观察到的颜色是否一致。甚至我们无法判断两“正常人体”在观察到同一颜色时，其感官对于颜色的感受是否一致。）

3.2 实现方法

以上讨论可以视为脑机可行性的必要条件，而实现脑机的方法又分为两类：有目的性地获取信息和无目的性的获取信息。前者需要提前确定要获取的信息的

位置及对应含义，并在大脑中的特定部位安装脑机接口对信息进行捕获与解释。而后者不关心大脑的运行机理，使用神经网络对信息进行解释，由于神经网络往往不具备可解释性（即，它往往可以不需要理解底层原理而达到所需目标），后者往往在实现目的后也并不清楚模型的运行原理。

由以上分析我们可以得出结论，对于以应用为主要目的的脑机，部分脑机的可行性可以直接判断，同时也存在无法直接验证其可行性的情况。对于脑机的原理，也存在不具有可解释性的场景。

3.3 伦理

客观上讲，脑机的诞生为残障人士带来了前所未有的便利，但同时也引来了许多伦理上的问题。例如：

由于脑机是时刻有能力捕捉人脑神经信号的，因此先前健康状态下在脑中封闭的想法不再只能通过语言和行为来表达最终的决定，而变成了可以实时接收翻译的、不再隐蔽的、可以预先读取的信号。这对于人的隐私造成了威胁。^[29]

商业公司将有能力将各种表达快乐体验的大脑信号编辑成一个又一个的程序，置入芯片。这些芯片通过信号刺激在大脑里制造不同的三维情节，使使用者体验极乐感受。对于这类程序的沉迷将使人类陷入虚拟世界的麻痹之中，与现实脱离，最终影响人类对客观物理世界的探索。^[29]

脑机连接之间的正负反馈，可能混淆谁是系统真正的主体。以至于难以分清是人脑在控制机器，还是机器在控制人脑，如果人脑被机器控制，我们或许将丧失自由的意识。^[29]

诸如此类的伦理问题都将影响到脑机的可行性。

4 数字生命

数字生命是一个常常出现在影视作品中，同时也为营销号所追捧的概念。本文中，对于数字生命的定义为将人类的意识以数字化的信息储存，使其在生理死亡后仍可以与环境进行信息交流的生命形态。

4.1 意识

在对数字生命的定义中，我们无可回避的谈及了个体对其内部和外部存在的

感知或认识，我们将其称为意识。对于意识，我们无法对其进行准确的定义，不论是《韦伯第三版词典》中“内在心理或精神的知觉或感知；个体对内在自我的直观感知知识。关于外部物体、状态或事实的内在知觉。知觉的关注；兴趣、关注。”^[30]；还是《剑桥词典》中“理解或意识到某物的状态”的定义^[31]；抑或《牛津活词典》定义意识为“感知到周围环境并作出反应的状态”“个体对某物的认知或感知”，以及“心智对自身和外部事件认知的事实”。^[32]

对于以上定义，我们都无法回避使用“感知”“知觉”“认识”“感知”等描述性的词语，同时我们也无法回避通过举例的方式对于概念进行描述。

正因为我们无法对意识进行明确的定义，我们对于意识的存在性也无法给予严谨的证明。因此，对于数字生命本身的讨论无可回避地是建立在不可解释的模型的基础上的。

四、结语

芯片作为人工智能发展的重要基石，我们若想在人工智能的新一轮国际竞争中走在世界前列，引领全球。首先，我们必须建立起独立自主的芯片研发、制造生产线。把自己的生命线牢牢掌握在自己手中。同时，在芯片研发方面，我们可以充分发挥中华民族的创造力、创新发展活力，不拘泥于西方发展老路，开创中国特色新路，坚持把芯片质量放在自己力量的基点上，坚持把人工智能发展进步的命运牢牢掌握在自己手中。最后，在人工智能研发领域，我们应该有自己的判断，不盲目跟风，要结合实际、通过独立判断，找到更有效的发展模式。既要着眼全局发展，探索规模化、体系化的智能研发，也可聚焦某一具体问题建立扁平化、轻便化的协同机制，不拘一格，务求实效。

参考文献

- [1] OpenAI Unveils New ChatGPT That Listens, Looks and Talks. New York Times. 2024/05/13
- [2] 惟改革者进 惟创新者强 惟改革创新者胜——习近平总书记关于创新发展的思想足迹. 习近平经济思想研究中心. 2023/04/20
- [3] 史上最快走红！ChatGPT 用户破亿. 推付费版. TVBS 新闻网. 2023/02/03 [2023-02-03].
- [4] OpenAI CEO Sam Altman seeks as much as \$7 trillion for new AI chip project: Report. CNBC. 2024/02/09.
- [5] 以人工智能赋能高质量发展. 红旗文稿. 2024/04/13
- [6] Auth C. 45 nm high-k + metal gate strain-enhanced CMOS transistors[C]//IEEE 2008 Custom Integrated

- Circuits Conference (CICC). Piscataway: IEEE Press, 2008: 379-386.
- [7] Yadav N, Jadav S, Saini G. Geometrical variability impact on the performance of sub-3 nm gate-all-around stacked nanosheet FET[J]. *Silicon*, 2022, 14, 10681-10693, Early Access.
 - [8] Huang X, Lee W C, Kuo C, et al. Sub 50-nm FinFET: PMOS[C]//1999 IEEE International Devices Meeting (IEDM). Piscataway: IEEE Press, 1999: 67-70
 - [9] DONG J, ZHANG X. Prospects of Advanced Integrated Circuit Technologies in Post-Moore Era[J]. *Science and Technology Foresight*, 2022, 1(3): 42-51.
 - [10] Williams R S. What's next?[J]. *Computing in Science & Engineering*, 2017, 19(2): 7-13.
 - [11] 姜晓鸿, 郝跃. 集成电路局部缺陷及其相关的功能成品率和电迁徙问题的研究[D]. 西安: 西安电子科技大学, 1998.
 - [12] 黄如. 后摩尔时代集成电路技术发展探讨[R]. 北 京: 中国科学院, 2021.
 - [13] Bhavin J. Shastri, Alexander N. Tait, T. Ferreira de Lima, Wolfram H. P. Pernice, Harish Bhaskaran, C. D. Wright & Paul R. Prucnal. Photonics for artificial intelligence and neuromorphic computing[J]. *Nature Photonics*, 2021, 15, 102–114
 - [14] Daniele Ielmini & H.-S. Philip Wong. In-memory computing with resistive switching devices[J], *Nature Electronics*, 2018, 1, 333-343
 - [15] Liao K, Dai T, Yan Q, Hu X, and Gong Q. Integrated Photonic Neural Networks: Opportunities and Challenges[J]. *ACS Photonics*, 2023, 10, 7, 2001-2010
 - [16] Lu, L., Zhao, S., Zhou, L., Li, D., Li, Z., Wang, M., Li, X., Chen, J. 16 × 16 non-blocking silicon optical switch based on electro-optic mach-zehnder interferometers. *Opt. Express* 2016, 24 (9), 9295–9307.
 - [17] Zhou, H., Zhao, Y., Wang, X., Gao, D., Dong, J., Zhang, X. Self-configuring and reconfigurable silicon photonic signal processor. *ACS Photonics* 2020, 7 (3), 792–799.
 - [18] Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. *Science* 2018; 361(6406):1004–8.
 - [19] Fu, T., Zang, Y., Huang, H., Du, Z., Hu, C., Chen, M., Yang, S., Chen, H. On-chip photonic diffractive optical neural network based on a spatial domain electromagnetic propagation model. *Opt. Express* 2021, 29 (20), 31924–31940.
 - [20] Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* 2017, 11(7): 441–6.
 - [21] George JK, Mehrabian A, Amin R, Meng J, De Lima TF, Tait AN, et al. Neuromorphic photonics with electro-absorption modulators. *Opt Express* 2019, 27(4): 5181–91.
 - [22] Zuo Y, Li B, Zhao Y, Jiang Y, Chen YC, Chen P, et al. All-optical neural network with nonlinear activation functions. *Optica* 2019, 6(9): 1132–7
 - [23] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, et al. Nanophotonic media for artificial neural inference. *Photonics Res* 2019, 7(8): 823–7.
 - [24] Caulfield HJ, Dolev S. Why future supercomputing requires optics. *Nature Photonics* 2010, 4(5):261–3.
 - [25] De Marinis L, Cococcioni M, Castoldi P, Andriolli N. Photonic neural networks: a survey. *IEEE Access* 2019; 7: 175827–41.
 - [26] Wu. J., Lin. X., Guo. Y., Liu. J., Fang. L., Jiao. S, Dai. Q. Analog Optical Computing for Artificial Intelligence[J]. *Engineering* 2022, 10: 133-145
 - [27] Simmone Shah. How Implanted Brain Chips Like Neuralink Could Change Our Lives. *TIME. Science Innovation*. Jan. 30. 2024
 - [28] McCulloch, Warren S.; Pitts, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943-12-01, 5 (4): 115–133

- [29] 大森健康漫谈. 不再是科幻,人类直面“脑机连接”八大伦理困境. 2020-05-14
- 陈建新, 袁贵礼著, 中国当代大学生的法律意识透析, 社会科学论坛, 2002, 6 (4): 88
- [30] 《韦伯第三版词典》1966 年版, 卷一, 第 482 页
- [31] CONSCIOUSNESS – meaning in the Cambridge English Dictionary. dictionary.cambridge.org. [2023-06-07].
- [32] consciousness – Definition of consciousness in English by Oxford Dictionaries. Oxford Dictionaries – English.
-