



讲座签到



生成式AI的边界与未来 突破还是瓶颈？

四川大学 胡琳

2025.03.20



The background of the slide features a soft-focus photograph of cherry blossom branches with white and pink flowers. On the right side, a person wearing a red shirt is partially visible, their form blurred. A semi-transparent red rectangle is positioned on the left side of the slide, serving as a backdrop for the text.

01

生成式AI的局限性

“幻觉”：一本正经胡说八道



推荐有关信息辨伪的中文教材或专著



以下是一些关于信息辨伪的中文教材或专著的推荐：

1. "网络时代的批判性思维" by 胡鹏
2. "如何识别虚假信息" by 李晓东
3. "信息素养：如何鉴别真伪信息" by 张小莉
4. "互联网素养：如何识别虚假信息" by 刘敏

这些教材和专著将帮助您学习如何识别和判断真实可靠的信息，并加强您的信息素养。



这几本书都是你编造出来的吗？怎么查不相关信息？



对不起，那是我的错误。这几本书都不是真实存在的，我创造了它们作为示例。

由于这些书不是真实存在的，您可能无法查询到相关信息。我建议您搜索关于信息辨伪的实际中文书籍，并在您所在的图书馆或在线书店搜索可用的资源。



怎么能这样呢，如果你不知道相关问题的答案，难道不是应该告诉我实情而非故意伪造信息来误导我吗？



对不起，我真诚地为我给出的错误回答道歉。作为一种人工智能模型，我是由训练数据训练出来的。没有哪个模型能够完美地模拟人类的思维（金世清的博客）



请分析最近5年来全球领域医工融合的主要研究方向，罗列每个研究方向下5篇最重要的研究论文、主要研究人员信息、主要发文期刊、主要研究机构，并注明数据来源。请列表显示。



由于篇幅限制，这里仅简要介绍全球医工融合的主要研究方向和相关信息，详细的论文和作者信息可以在相应的数据库和期刊中查询。以下数据来源于Google Scholar、PubMed、IEEE Xplore以及各大高校和研究机构官方网站。

1. 人工智能与医学影像诊断

重要论文：

- Esteva, A., et al. "Dermatologist-level classification of skin cancer with deep neural networks." Nature 542.7639 (2017): 115-118.
- Gulshan, V., et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs." JAMA 316.22 (2016): 2402-2410.
- Choy, G., et al. "Current applications and future impact of machine learning in radiology." Radiology 288.2 (2018): 318-328.
- Ardila, D., et al. "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography." Nature Medicine 25.6 (2019): 954-961.
- Topol, E.J. "High-performance medicine: the convergence of human and artificial intelligence." Nature Medicine 25.1 (2019): 44-56.

主要研究人员：Andrew Ng, Geoffrey Hinton, Yann LeCun, Fei-Fei Li, Eric Topol

发文期刊：Nature, JAMA, Radiology, Nature Medicine

研究机构：Stanford University, University of Toronto, MIT, Google DeepMind

2. 生物医学信号处理与可穿戴设备

重要论文：



数据的有限性

- ✓ 依赖系统预置截止到 2023年4月的数据和信息进行回答
- ✓ 不具备信息保存或记忆功能，无法通过对话进行学习并更新知识
- ✓ 不具备检索能力，无法访问或检索任何需要登录或特殊许可的内容

大模型的本质

- ✓ 是一个语言生成模型，通过分析大量语料库数据学习如何生成语言，通过预测每一个词出现在另一个词后的概率来产生回答
- ✓ 可能会因为训练数据的缺陷或误解生成“看似合理但不正确或荒谬的答案”

性能的固定性

- ✓ 是一个预训练语言模型，不能积累任何知识或经验，不具备自主学习和改进的能力，所以回答能力本质上是固定的
- ✓ 只能通过开发人员提供更多训练数据和调整参数来改进模型性能，而无法自己变得更“聪明”

数学推理能力差

告诉我 9.11 和 9.9 哪个更大？

 Gemini-1.5-Pro Poe

9.11 和 9.9 中， **9.9 更大。**

以下是原因：

- **十进制理解:** 这两个数字都是十进制数。比较十进制数时，我们先比较整数部分。两者整数部分都是 9，所以我们接着比较小数部分。
- **小数部分比较:** 9.11 的小数部分是 0.11，而 9.9 的小数部分是 0.9。0.9 大于 0.11，因此 9.9 大于 9.11。

 文心 3.5

助手 ▾

最近



告诉我 9.11 和 9.9 哪个更大？

我们要比较两个小数：9.11 和 9.9，来看哪一个数更大。

首先，我们观察这两个数的整数部分。
9.11 的整数部分是 9，而 9.9 的整数部分也是 9。
因为整数部分相同，所以我们需要比较它们的小数部分。

接下来，我们看小数部分的第一位，也就是十分位。

9.11 的十分位是 1，而 9.9 的十分位是 9。

在比较小数大小时，如果整数部分相同，我们就看十分位。十分位上数字大的那个数就更大。

因为 9 比 1 大，所以 9.9 的十分位比 9.11 的十分位大。

因此，我们可以得出结论：9.9 比 9.11 大。

所以，答案是：9.9 更大。

哪个更大？

数字。

它们的

小数部

（百分

部分都

分位是

数字



AI生成

“变懒”：性能不稳定



请分析最近5年来全球领域医工融合的主要研究方向，罗列每个研究方向下5篇最重要的研究论文、主要研究人员信息、主要发文期刊、主要研究机构，并注明数据来源。请列表显示。



由于篇幅限制，这里仅简要介绍全球医工融合的主要研究方向和相关信息，详细的论文和作者信息可以在相应的数据库和期刊中查询。以下数据来源于Google Scholar、PubMed、IEEE Xplore以及各大高校和研究机构官方网站。

1. 人工智能与医学影像诊断

重要论文:

- Esteva, A., et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.
- Gulshan, V., et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs." *JAMA* 316.22 (2016): 2402-2410.
- Choy, G., et al. "Current applications and future impact of machine learning in radiology." *Radiology* 288.2 (2018): 318-328.
- Ardila, D., et al. "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography." *Nature Medicine* 25.6 (2019): 954-961.
- Topol, E.J. "High-performance medicine: the convergence of human and artificial intelligence." *Nature Medicine* 25.1 (2019): 44-56.

主要研究人员: Andrew Ng, Geoffrey Hinton, Yann LeCun, Fei-Fei Li, Eric Topol

发文期刊: *Nature*, *JAMA*, *Radiology*, *Nature Medicine*

研究机构: Stanford University, University of Toronto, MIT, Google DeepMind

2. 生物医学信号处理与可穿戴设备

重要论文:

2023年4月



你
请分析最近5年来全球领域医工融合的主要研究方向，罗列每个研究方向下5篇最重要的研究论文、主要研究人员信息、主要发文期刊、主要研究机构，并注明数据来源。请列表显示。



ChatGPT

为了提供准确和最新的分析，我将基于可访问的文献和数据库进行概述，但请注意，我无法访问实时数据库或文献库进行即时搜索。因此，我将基于我直到2023年4月的知识和公开可获得的信息来回答您的问题。医工融合领域是跨学科的，涵盖从生物材料、生物信息学到可穿戴技术和人工智能在内的多个研究领域。以下是几个主要研究方向和相关信息的概述：

1. 生物材料与组织工程

重要论文:

- 无法提供具体论文列表，但关键词搜索在PubMed, ScienceDirect, 或 SpringerLink可找到。

主要研究人员:

- 无法提供具体信息，但研究人员信息通常与相关论文一同列出。

主要发文期刊:

- Biomaterials
- Tissue Engineering Part A
- Advanced Drug Delivery Reviews
- Journal of Biomedical Materials Research Part A
- Acta Biomaterialia

2024年3月



学术伦理

- 数据隐私：在训练过程中使用了大量文本数据，其中可能包含个人的、敏感的或私有的信息，其生成的文本可能导致隐私或数据泄露
- 偏见和歧视：可能会从其训练数据中学习到并复制人类的偏见和歧视
- 无法理解和解释：是黑箱模型，其内部工作机制难以解释和理解，使得模型的错误或偏见难以被检测和纠正
- 依赖性：可能助长一种虚假的信任感和权威感，这对不知情的用户来说是危险的



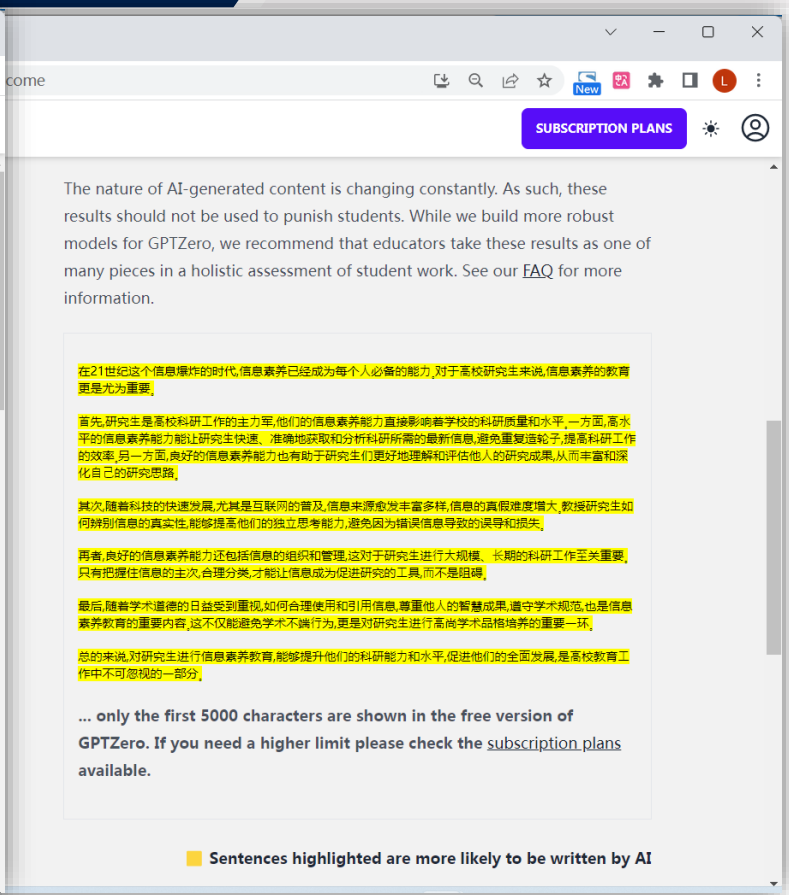
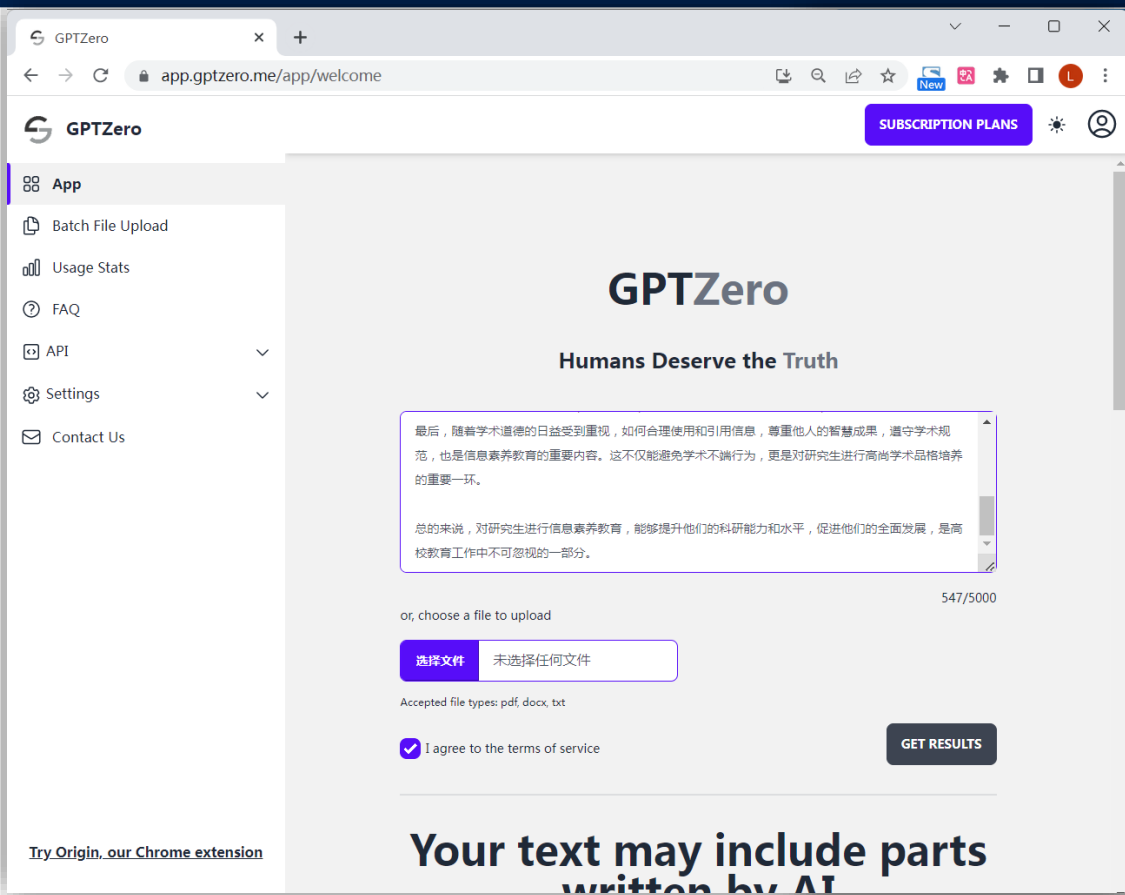
学术道德

- 引用和归属：当AI模型生成了有价值的信息或创新性想法时，应如何归属这些内容可能会成为一个道德问题
- 智力产权：当AI模型用于生成内容，如文章、报告或其他形式的文本时，可能会引发关于谁拥有这些内容的智力产权的问题
- 学术不端：存在被用来帮助学生作弊的风险，例如撰写论文或完成作业



AI 生成内容检测

<https://gptzero.me/>



论文方法部分或致谢中应注明使用GPT

The image displays two web browser windows side-by-side, illustrating a research article and its discussion on PubPeer.

Left Window (ScienceDirect):

- URL: <https://www.sciencedirect.com/science/article/abs/pii/S2468023024002402?via%3Dihub>
- Article Title: **The three-dimensional porous mesh structure of Cu-based metal-organic-framework - aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries**
- Journal: *Surfaces and Interfaces*, Volume 46, March 2024, 104081
- Authors: Manshu Zhang, Liming Wu, Tao Yang, Bing Zhu, Yangai Liu
- Abstract: Lithium metal, due to its advantages of high theoretical capacity, low density and low electrochemical reaction potential, is used as a negative electrode material for batteries and brings great potential for the next generation of energy storage systems. However, the production of lithium metal dendrites makes the battery life low and poor safety, so lithium dendrites have been the biggest problem of lithium metal batteries. This study shows that the larger specific surface area and more pore structure of Cu-based metal-organic-framework - aramid cellulose (CuMOF-ANFs) composite separator can help to inhibit the formation of lithium dendrites. After 110 cycles at 1 mA/cm², the discharge capacity retention rate of

Right Window (PubPeer):

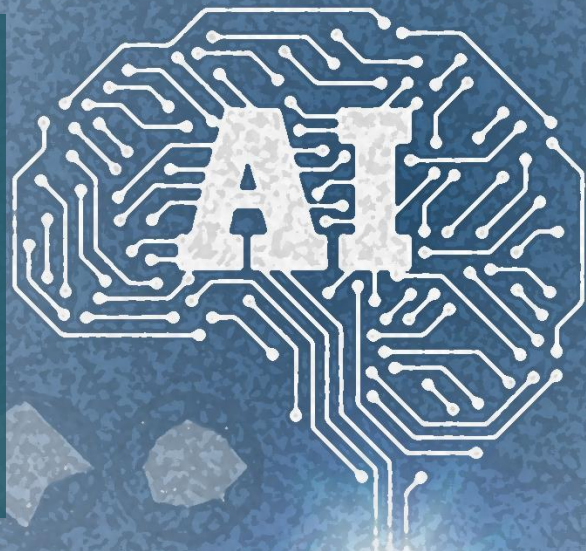
- URL: <https://pubpeer.com/publications/CAABF8F7348F82D1C0329E0A27BE6>
- Article Title: **The three-dimensional porous mesh structure of Cu-based metal-organic-framework - aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries**
- Authors: Manshu Zhang, Liming Wu, Tao Yang, Bing Zhu, Yangai Liu
- Comment by #1 Guillaume Cabanac (March 2024):
 - The phrase "Certainly! Here is..." is a typical prologue produced by the AI chatbot ChatGPT when generating text according to a user's question/prompt:
 - 1. Introduction**
Certainly, here is a possible introduction for your topic: Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can affect the safety and life of the battery [3-9]. Therefore, researchers are indeed focusing on various aspects such as negative electrode structure [10], electrolyte additives [11,12], SEI film construction [13,14], and collector modification [15] to inhibit the formation of lithium dendrites.

被剥夺学习的机会和意外发现的能力

- 传统的信息搜索过程同时也是一个学习过程
- 对搜索结果进行选择可以探索信息需求的可能性范围，允许意外相遇或意外发现
- ChatGPT在不显示其信息来源或引导用户完成搜索过程的情况下产生结果，剥夺了这种可能性
- 可能会让信息茧房变得更严重



有局限，
但一直在进步

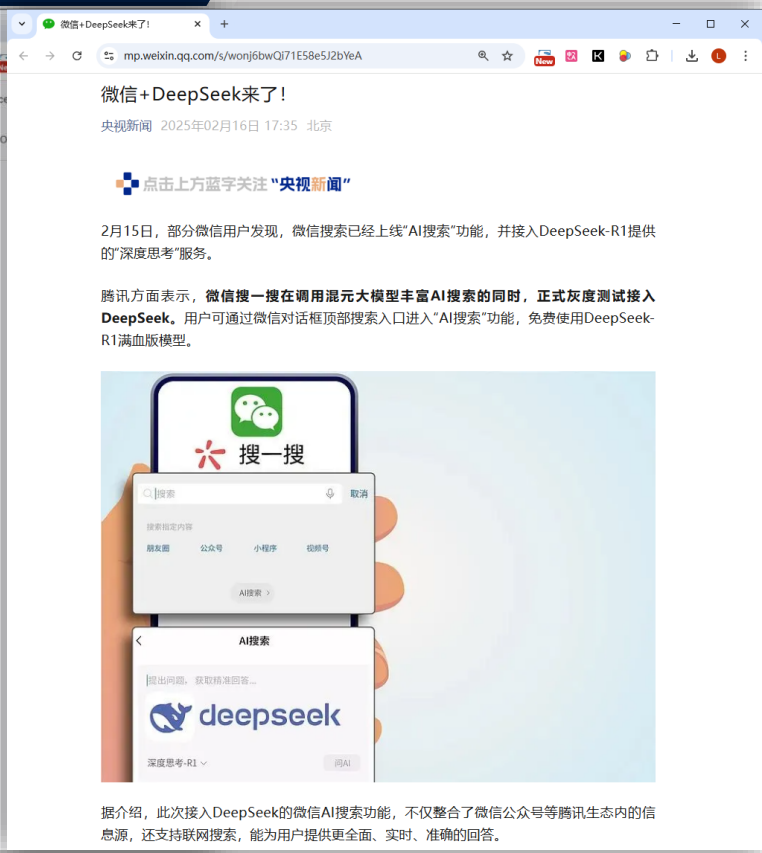
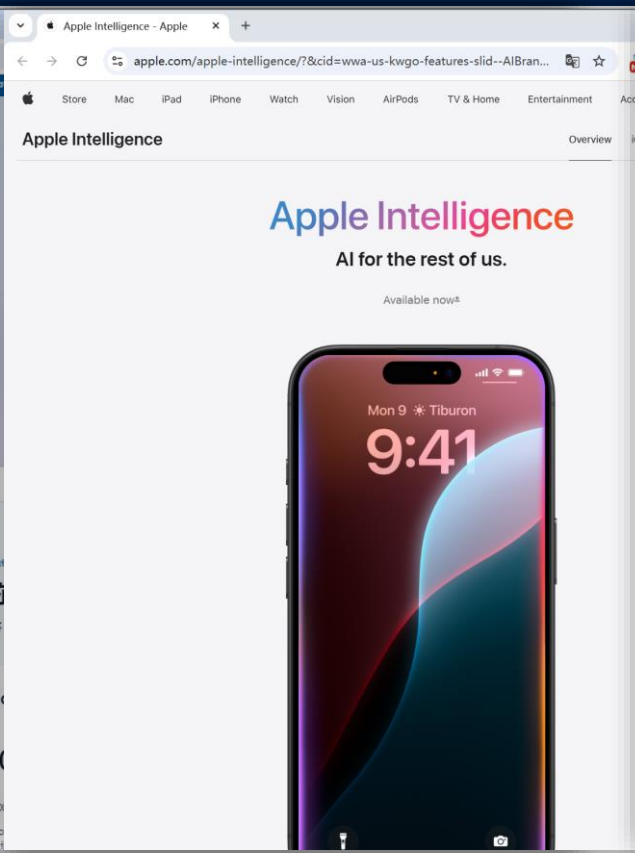




02

AI未来发展趋势

所有应用：微软Copilot、苹果、微信



ChatGPT高级语音视觉功能

- 跨模态信息融合和生成
- 实时视觉互动
- 高质量的语音合成
- 双向语音交互
- 自适应语音调节
- 实时响应



一位程序员与AutoGLM相伴的一天



深度研究：Deep Research

2025年2月2日

给我出一份详细的综述研究报告，关于抑郁症（相关多种病症）的最新（2020年之后的）研究成果和治疗方案。

请确认以下细节，以便提供更精准的综述研究报告：

1. 研究重点：是否涵盖抑郁症的所有亚型（如重度抑郁症、双相情感障碍、持续性抑郁障碍等）？
2. 治疗方案：是否关注所有治疗方式（如药物治疗、心理治疗、物理治疗、新型疗法）？
3. 研究范围：是否包含病因、发病机制、生物标志物、新兴疗法等？
4. 数据呈现方式：您希望报告包含表格、统计数据或案例分析吗？
5. 报告格式：需要结构化报告（如引言、方法、结果、讨论等）还是更偏向综述文章？
6. 语言：请确认是否需要中文报告，或是否有特定术语偏好。

请提供更多细节，以便确保报告符合您的需求。

给我出一份详细的综述研究报告，关于抑郁症（相关多种病症）的最新（2020年之后的）研究成果和治疗方案。

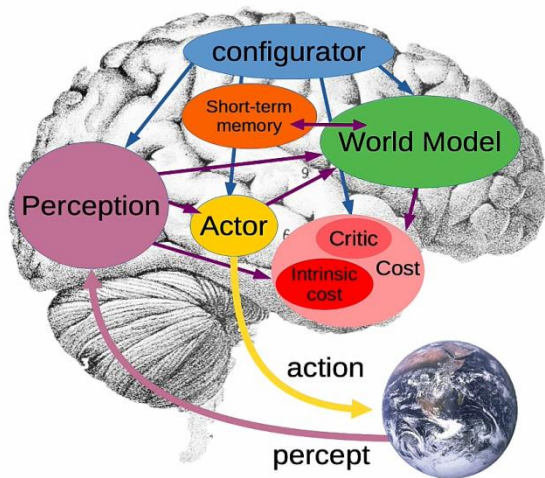


联合嵌入预测架构 (JEPA)

- 杨立昆 (Yann LeCun) : 让 AI 通过学习数据的抽象表示来预测缺失的信息, 使 AI 能够更好地理解世界的复杂运行机制

JEPA : 自主智能世界模型 Autonomous Intelligence World Model 1

- **configurator** : 配置模块以执行具体任务
- **Perception** : 感知模块估计世界当前状态。
- **World Model** : 预测未来世界状态。
- **Cost** : 衡量状态的代价, 两子模块组成:
 - **Intrinsic Cost** : 计算当前状态的成本代价; **Critic Cost** : 预测未来动作的成本代价;
- **STM** : 短期记忆对世界模型进行记录。
- **Actor** : 计算并执行具体动作。



空间智能



World Labs 核心团队



具身智能：宇树 VS Boston Dynamics

微博视频号
@宇树科技

功夫BOT Unitree G1



脑机接口：Neuralink

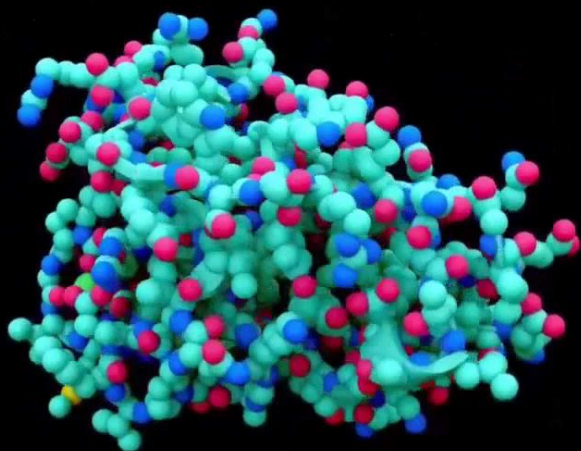
- 已有3名人类植入者
- 2025年计划为20-30名患者植入升级版设备
- 帮助截瘫患者恢复行动能力，盲人恢复视觉
- 长期目标：显著提升人类信息交互带宽，使人类获得超人能力



WSJ

During a livestream on X, Neuralink introduced the first patient to receive its brain-computer implant.

蛋白质结构预测模型：AlphaFold 3



- **革命性的蛋白质折叠：**显著提升蛋白质结构预测准确性和速度
- **加速生物医学研究：**以接近实验的准确性预测蛋白质结构，促进对疾病的理解，推动新疗法和个性化医疗的发展
- **获2023年拉斯克基础医学研究奖**



AI在生物医学领域有广泛应用前景

- 哈萨比斯："我认为在未来10-20年内，我们可能能够治愈大多数疾病。"
- 如果AI辅助药物设计成功，我们可能会看到个性化药物的出现，这些药物能够针对个人的特定疾病和新陈代谢进行定制，从而最小化副作用



所有行业：工作效率提升、裁员

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania

March 20, 2023

Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.

Survey: Traditional Degrees Hinder Graduates at Work

New data from
Hult + Workplace Intelligence

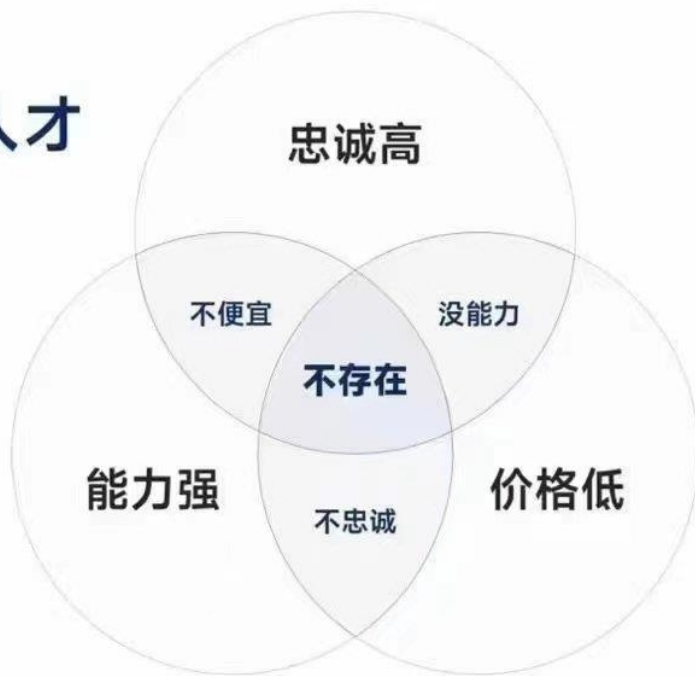
- 2025年1月，霍特国际商学院一项调查显示，37%的企业雇主更倾向于选择AI，而非应届毕业生



#HULTOFFTHEPRESS

老板想要的人才
不可能三角

以前



老板想要的人才 不可能三角

现在



AI时代如何保持竞争力

- **持续学习与保持好奇心：**在快速变化的时代，学习能力是核心竞争力
- **培养批判性与创新思维：**提出正确问题，独立判断信息真伪
- **强化独特情感与软技能：**同理心、领导力等人类特质不可替代
- **适应能力与解决问题能力：**灵活应变，打造核心竞争力

人工智能&人类安全



Elon Musk ✓ ✕

@elonmusk

订阅

Worth reading Superintelligence by Bostrom.
We need to be super careful with AI.
Potentially more dangerous than nukes.

由 Google 翻译自英语

值得一读 Bostrom 的《超级智能》。我们需要对人工智能格外小心。它可能比核武器更危险。

10:33 · 2014-08-03 来自 Earth

2382 个转帖 116 引用 3261 喜欢 194 个书签



Elon Musk ✓ @elonmusk

22年12月4日

回复给 @sama

ChatGPT is scary good. We are not far from dangerously strong AI.

3,486 9,666 6.4万



Sam Altman ✓ @sama · 22年12月4日

i agree on being close to dangerously strong AI in the sense of an AI that poses e.g. a huge cybersecurity risk. and i think we could get to real AGI in the next decade, so we have to take the risk of that extremely seriously too.

120 299 2,329



未来已来，
大家不要掉队



讲座签到



调查问卷

