

zhì liào
知了

AI:吃电巨兽

到2027年,生成式AI用电量将超过一些国家一年耗电

生成式AI,听起来挺酷,但你知道吗?它可是地地道道的“吃电巨兽”:现在爆火的ChatGPT,每日耗电量高达50万度!随着人工智能的发展,其用电量急速上升问题,正引发全球警惕,以后会发生“电荒”吗?能耗问题会不会成为AI发展的“绊脚石”?

主笔:于梅君

1 日耗电50万度! GPT成“吃电狂魔”

由美国OpenAI公司最新推出的视频生成工具Sora,自发布以来,便成为人工智能界最火爆的“网红”。该公司上一个全网刷屏的产品——ChatGPT,在推出仅两个月后,就达到1亿月活跃用户数。

不过,蓬勃发展的AI背后,是算力;而算力背后则是能耗。人工智能飞速发展的背后,其能源消耗问题日益凸显,成为全球关注的焦点。

据报道,全球数据中心的耗电量,已从十年前的100亿瓦增加到如今的1000亿瓦水平。ChatGPT聊天机器人每天耗电超过50万度,用于处理约2亿个用户请求,相当于美国家庭每天用电量的1.7万多倍。据OpenAI公司称,他们在ChatGPT上,每天花费高达70万美元。

近日,特斯拉创始人马斯克公开表示,未来两年,人工智能行业将由“缺硅”变为“缺电”。

据国际能源署称,全球数据中心目前约占用电量的1%至1.5%。到2025年,在全球数据中心用电量中,人工智能业务的占比,将从2%增至10%。

以谷歌为例,每年用于人工智能训练的电力为23亿度,相当于美国一个州首府所有家庭一年的用电量。即使在训练完成后,这些人工智能仍需大量计算能力运行,并耗费大量能源。

若谷歌在每次搜索中,都使用AIGC(生成式人工智能技术),其年耗电量将增至约290亿度。

最新研究表明,到2027年,生成式人工智能所消耗的能源,可满足荷兰这样一个国家一年的需求,即约85-134太瓦时(TWh)。

2 一次训练耗电2.4亿度电 能加热1000个游泳池水量

仅仅是GPT-4的GPU,一次训练就会用去2.4亿度电。AI为什么那么耗电?它们用掉的电都跑到哪里去了?

我们现在所说的人工智能(AI),主要是指生成式人工智能。它们需要大规模的数据中心来训练和推理。

这些数据中心由大量服务器组成,服务器消耗的电能,绝大部分转化成热能,最后通过水冷系统释放出来。可以说,AI的物理硬件,是个巨大的“电热水器”。

据统计,全球数据中心的能耗,约占全球总电力消耗的1%左右,预计这一比例,还会随着数据量的增长而增加。

能量无法创造也无法消灭,它只能从一种形式转化成另一种形式。对计算机来说,最主要的能量转化方式,就是从电能转化成热能。

大语言模型也是如此。它对电能和冷却水的需求,正造成越来越严重的环境问题。波士顿咨询集团曾发布报告称,到2030年底,仅美国数据中心的用电量,就将是2022年的三倍。

在OpenAI训练大语言模型GPT-4时,完成一次训练需要约三个月时间,使用大约25000块英伟达A100 GPU。仅仅是这些GPU(图形处理器),一次训练就要用2.4亿度电。这些电能几乎全部转化成了热能,可以将大约200万立方米冰水——大概是1000个奥运会标准游泳池的水量——加热到沸腾。

全球人工智能的蓬勃发展,可能会导致这一数字大幅上升。与此同时,大量的碳排放和数百万加仑淡水的消耗也不容小觑。

3 回答50个问题,得“喝”500毫升水降温

生成式AI,除了“吃电,还得喝水”——用水量也是个惊人数字。因为要训练这些AI,就得有强大的算力中心,而算力中心要运转,就得有散热系统,这都得用到水。

全球科技巨头,都在建设新的数据中心,争的就是谁的算力更强,结果就是对水资源的需求越来越大。

研究显示,训练GPT-3所需的水量,能够填满一个核反应堆的冷却塔。ChatGPT(在GPT-3推出之后)每与用户交流25到50个问题,就得“喝下”一瓶500毫升的水来降温。

谷歌发布的2023年环境报告显示,公司前一年的用水量同比增加20%,达到56亿加仑,其中绝大部分用于为数据中心散热。除了谷歌,另一个巨头Meta在美国亚利桑那州建了数据中心,仅2022年用水量就超过260万立方米。

正在大力开发人工智能的微软,也在最新的环境报告中透露,2022年公司用水量同比前一年激增34%,达到640万立方米,比2500个奥

林匹克标准游泳池的水量还多。

除了运营过程中耗水,制造人工智能所需芯片的过程中,也会消耗大量能源和水资源。

比如,硅片工艺需要“超纯水”清洗,生产一个2克重的芯片,大约需要32公斤水。制造8寸晶圆,每小时耗水约250吨,制造12英寸晶圆,每小时耗水可达500吨。

科学家预计,到2027年,全球人工智能可能需要4.2亿—66亿立方米的清洁淡水,这比4-6个丹麦或一半英国的年取水总量还要多。

耗水问题已成为制约数据中心快速发展的因素之一。为节约宝贵的自来水资源,很多企业尝试用各种方法为数据中心散热,例如,微软曾尝试部署海底数据中心,Facebook数据中心选址北极圈附近,阿里云千岛湖数据中心使用深层湖水制冷。

不过,上述方法总是会带来新的问题,目前国内数据中心的用水,主要还是用自来水,近几年政府更鼓励数据中心企业使用中水(循环再利用的水)。



◀ 阿里云千岛湖数据中心,节水率创全球数据中心最低纪录。

4 面对“能耗巨兽”,我们该如何应对

随着AI的进步,模型的增大,如何解决伴生的高耗能问题?

专家表示,当下的重点在于开源,风能、光伏等可再生新能源,被视为数据中心重要的能量池,核能尤其被寄予厚望。

在2023年的达沃斯世界经济论坛上,OpenAI的CEO山姆·阿尔特曼认为,核聚变可能是能源的发展方向,但要开发出真正可用的核聚变技术,可能还需要一些时间。

据国际能源署的报告,到2050年,可再生能源预计将满足全球近90%的电力需求增长。

节流方面,可以通过优化数据中心的设计和管理,提高电源使用效率。国内大厂也纷纷给出解决方案。比如腾讯,组建了专门的绿色能源团队,应用自然冷却、液冷、余热回收等新兴节能技术,降低数据中心能耗。华为、京东、亚马逊等也纷纷入局智慧能源领域。

如何让散失的热能不至于白白浪费?最容易实现的,是热回收技术。即将数据中心回收的热量,用于提供民用热水,冬季则提供民用采暖。现在有些企业已着手于回收废热再利用,比如阿里巴巴千岛湖数据中心等。

此外,通过优化算法来减少计算冗余,使用更高效的芯片,以及采用节能的数据中心设计,都是降低能耗的有效方法。

政府也在出台相关政策,鼓励企业研发更为节能环保的AI技术和基础设施。

早在2021年,我国就已出台公共机构碳达峰方案,新建大型、超大型数据中心,必须全部达到绿色数据中心要求。

北京印发的《关于进一步加强数据中心项目节能审查的若干规定》,也新增了关于引导数据中心充分利用再生水的内容。

知多一点

美国知名企业家马斯克近日在社交媒体上表示,到明年年底或2026年,新的人工智能模型可能将超越人类智力,“比最聪明的人还聪明”。这比他去年预测的2029年这一时间点提前不少。

人工智能发展速度 超出人类预期

近期,包括英伟达创始人兼CEO黄仁勋和谷歌DeepMind CEO哈萨比斯在内的科技大佬,都对AGI(通用人工智能)来临的时点做出预测。

哈萨比斯在最近的一次播客采访中表示,他认为AGI最快可能在2030年到来。

黄仁勋在近期举行的斯坦福大学经济论坛上表示,从某些定义来看,通用人工智能可能会在短短五年内实现;但从另一些定义来看,可能还很遥远。

黄仁勋认为,大模型的发展速度很快,但这并不等同于AGI,虽然它们很相关。

他说:“科技行业距通用人工智能还有数年时间的的原因之一是,尽管机器学习目前擅长识别和感知等任务,但它还无法执行多步推理,这对公司和研究人员来说是首要任务,每个人都在为此努力。”

这也说明,人工智能技术的发展速度超出人类预期。通用人工智能(AGI)的实现,被认为AI将具有人类水平的智能。

马斯克等“AI威胁论者”认为,AGI很快就会到来,甚至已经开始到来。如果AI能够智胜人类,或者变得具有欺骗性、与人类价值观不一致,也可能是危险的,因此,需要防止将这一技术掌握在单个公司或者权力聚焦的实体手中。

马斯克说,去年,人工智能发展的主要制约因素是缺少高性能芯片,如今这一限制正在得到缓解,而电力供应成了新的限制。未来一两年,人工智能发展的关键就在于电力供应。

马斯克去年成立人工智能初创公司xAI,该公司去年11月推出首款人工智能模型Grok。马斯克透露,Grok 2的训练将于今年5月完成。

AI要达到人类 智能水平尚存差距

“AI将比最聪明的人还聪明”,这一预测靠谱吗?清华大学战略与安全研究中心特约专家朱荣生表示,人脑消耗功率远低于人工智能运行时所损耗功率,人工智能要达到人类智能水平尚存差距。

但从长远看,人工智能在算法、数据的不断“加持”下,或最终超越人类智能,而达到这一目标所需的硬件和软件,分别为芯片和算法。不过,朱荣生表示,人工智能的发展,需要政策、资源与技术支持,但许多领域的人工智能“培育”尚缺乏相关支持,人工智能的发展受到遏制。

AI真的很快会比人类聪明吗