

英特尔“放大招” AI芯片竞速升级

青岛财经日报/首页新闻讯

4月9日,芯片巨头英特尔发布新一代 Gaudi 3 人工智能(AI)芯片, 争夺 AI 芯片市场高地。

同日,谷歌宣布推出一款基于 ARM 架构的服务器芯片 Axion,其性能比通用 ARM 芯片高 30%,比英特尔生产的 x86 最新芯片性能提高 50%。

4月10日,美国科技巨头 Meta 宣布,正在部署一款自主研发的人工智能芯片,助力 AI 业务发展,并减少对英伟达和其他外部公司芯片的依赖。

英特尔推出新款 AI 芯片

英特尔声称,Gaudi 3 芯片与英伟达 H100 芯片相比,推理能力平均提高 50%,能效平均提高 40%,运行人工智能模型的速度是 H100 的 1.5 倍。英特尔表示,这款产品将与英伟达最新的 H200 大致相当,在某些领域甚至表现更好。

该公司已经在 Meta 的开源模型 Llama 上测试了 Gaudi 3。英特尔表示,Gaudi 3 可以帮助训练或部署人工智能大模型,包括用于文本到图像生成模型 Stable Diffusion 或 OpenAI 的语音识别模型 Whisper。

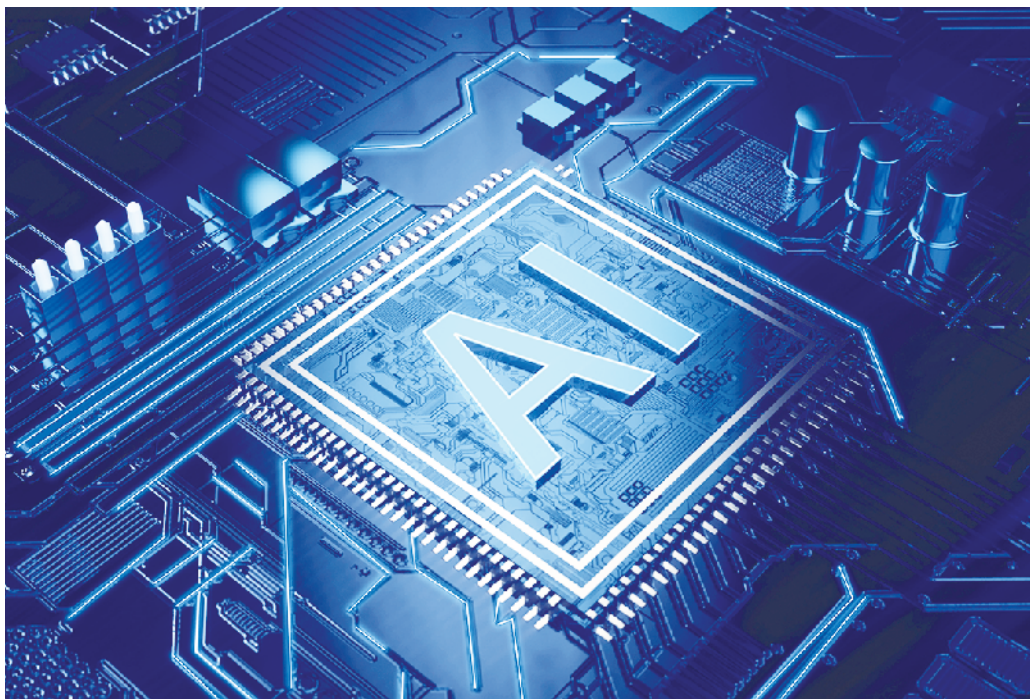
据英特尔介绍,Gaudi 3 采用 5 纳米工艺制造,这是一种相对较新的制造技术,这表明该公司正在使用外部代工厂生产芯片。除了设计 Gaudi 3,英特尔还计划在俄亥俄州的一家新工厂生产人工智能芯片,预计将于 2027 年或 2028 年投产。

英特尔表示,新的 Gaudi 3 芯片将在第三季度向客户大规模提供,包括戴尔、惠普和超微电脑在内的公司将使用该芯片。

英特尔云和企业解决方案事业部副总裁兼高级首席工程师达斯·卡姆豪特在电话会议上声称:“我们确实预计 Gaudi 3 能够与英伟达的最新芯片一较高下。从我们具有竞争力的价格,我们独特的开放式集成芯片网络,我们正在使用行业标准的以太网来看,我们相信这是一款强大的产品。”

人工智能热潮使英伟达的股价在过去一年增长了两倍多,市值超过 2 万亿美元。英特尔的股价同期只上涨了 18%。

因此,英特尔首席执行官 Pat Gelsinger 希望借助人工智能热潮重振该公司的业务,英特



尔曾因过去的失误和更广泛的个人电脑市场低迷而遭受重创。与此同时,长期竞争对手 AMD 已经占据了个人电脑和服务器市场部分份额,英特尔的一些最大客户现在也开始自主设计芯片。

谷歌加快芯片自研

为了减少对英伟达芯片的依赖,谷歌正在加快芯片自主研发的速度。4月9日,谷歌宣布推出为数据中心设计的、基于 ARM 架构定制的新型中央处理器 Axion,以期降低云计算的运营成本。

Axion 结合了谷歌在芯片领域上的专业知识和 ARM 最高性能的 Neoverse V2 处理器内核,能提供业界领先的性能和能效,将于今年晚些时候向谷歌云的客户提供服务。

根据谷歌云内部的数据,与当今云中最快的、基于 ARM 的处理器相比,Axion 提供的实例性能高出了 30%;而与当代基于 x86 的同类处理器相比,Axion 的性能高出 50%,能效高出 60%。

谷歌母公司 Alphabet 有四分之三的收入来自广告,但云计算部门的增长更快,已占到总收入的 11%。据市场研究公司 Gartner 估计,到 2022 年,谷歌占据云基础设施市场 7.5% 的份额,而亚马逊和微软共控制着 62% 左右的份额。

此外,谷歌也在更新其 TPU 人工智能芯片,近日,谷歌 TPU v5p 芯片通过谷歌云服务上线。谷歌 TPU 芯片作为英伟达 GPU 的替代品用于人工智能加速任务,尽管开发者只能通过谷歌云平台访问它们而不能直接购买。

谷歌云副总裁兼计算和机器学习基础设施

总经理马克·洛迈耶表示,TPU v5p 是下一代加速器,专门用于训练一些规模最大、要求最高的生成式人工智能模型。单个 TPU v5p pod 包含 8960 个芯片,是 TPU v4 pod 上芯片数量的两倍多。

Meta 欲降低对英伟达依赖

4月10日,Meta 宣布,正在部署一款自主研发的人工智能芯片,助力 AI 业务发展,并减少对英伟达和其他外部公司芯片的依赖。

这款芯片是 Meta 训练与推理加速器项目(MTIA)的最新版本,MTIA 是 Meta 专门为 AI 工作负载设计的定制芯片系列,可以对 Facebook 和 Instagram 上的内容进行排名和推荐。该公司去年发布了首款 MTIA 产品。

据 Meta 介绍,新一代 AI 芯片由台积电代工,采用了其 5 纳米工艺制程。MTIA 也是该公司更广泛产品开发的一部分,针对 Meta 独特的工作负载和系统进行了优化。与前一个版本相比,新的 MTIA 在计算能力和内存带宽上都提高了一倍多,同时保持了与其工作负载的紧密联系,可以更好地服务全球用户,提供个性化的推荐和优化的用户体验。

Meta 表示:“为了实现我们定制芯片的雄心壮志,我们不仅要投资于计算芯片,还要投资于内存带宽、网络、容量以及其他下一代硬件系统。”

Meta 转向 AI 服务使其对计算能力的需求不断增长,一方面,该公司要让 Facebook、Instagram 等应用接入 AI 功能,另一方面,Meta 正在开发自己的大语言模型,以期与 OpenAI 的 ChatGPT 展开竞争。

业界简报

宁德时代发布全球首个“5年零衰减”储能系统

青岛财经日报/首页新闻讯 近日,宁德时代对外发布天恒储能系统,为新型储能规模化应用和高质量发展提供助力。

天恒储能系统是全球首个“5年零衰减”储能系统,可大规模量产。容量“5年零衰减”是长寿命锂电池的一个里程碑,但对致力于满足新型电力系统需求的储能电站而言,功率零衰减同样至关重要。天恒储能系统,采用了仿生 SEI 和自组装电解液技术,为锂离子清除路障,实现 5 年功率和容量零衰减,辅机功耗全生命周期可控且不增长,真正做到由内而外的“冻龄”。

天恒储能系统在标准 20 尺集装箱内,实现 6.25 兆瓦时级的高能量,单位面积能量密度提升 30%,整体占地面积减少 20%,达到 6 兆瓦时级别储能系统的全球第一。同时,该储能系统搭载了储能专用长寿命电芯 L 系列产品,使得磷酸铁锂储能电池的能量密度达到 430 瓦时/升。

蚂蚁集团推出 AI 新功能 设计图可一键生成代码

青岛财经日报/首页新闻讯 昨日,蚂蚁集团自研的智能研发平台 CodeFuse 推出“图生代码”新功能,支持开发人员用产品设计图一键生成代码,大幅提升前端页面的开发效率。

新推出的“图生代码”功能能够根据设计图自动生成相应的代码,极大地减少了开发团队在开发网页、小程序和 APP 时的代码编写工作量。

以一个中型网页为例,原本需要一名开发人员耗费约 1 小时编写 200 行代码,而现在通过“图生代码”,开发人员只需进行检查和调整,大大缩短了开发时间。

CodeFuse 负责人表示,AI 的普及不仅可以减少开发人员的工作压力,让他们有更多精力投入到更有创造性的工作中去,更大的意义在于降低编程门槛,推动软件开发行业的创新和进步。CodeFuse 的使命是探索下一代 AI 研发生产力工具,致力于打造创新的解决方案。在自然语言生成代码、图生代码之后,CodeFuse 还将持续推出新功能,助力企业研发全链路的效率提升。

站在 AI 肩膀上的素质教育

脑机对接是“医治病人”还是“制造超人”?

□ 黄全愈

三、AI 教育能培养数学思维吗?

AI 的算法和算力有多强悍? 人类目前只研究了蛋白质结构的千分之一,但 AlphaFold 2 (与“阿尔法狗”师出同门)一瞬间就能算出剩下的蛋白质结构。

例如,“阿尔法狗”击败众多围棋高手的结果,与其说“阿尔法狗”击败人类,不如说是人(利用 AI 的数据、算法、算力的机器“人”)击败了人(单打独斗的科学家们)。

在某种意义上,人类的“1”和 AlphaFold 2 的“999”之间的差距,是算料(数据)、算法、算力之间的差别。

如今,我们已经被数字化包围了。文字、声音、图像、视频,甚至人的基因测序、细胞等都被数字化了。可以说,AI 最基础的是数学思维。

我们可以通俗地把 AI 的“算法”类比为数学公式,也可以形象化地把 AI 的“算力”类比为数学思维。

对强大的数学思维造成更强力破坏的竟是数学公式。因为人们只想简单地、直接地利用公式获得答案,忘了公式后面的思维和逻辑。人类把关系纷繁、思维深刻、逻辑严密的数学计算,简化为几个数字及字母合成的数学公式。

数学思维何其强悍,可惜,“AI 自习室”并不教你数学思维。

有个故事,能言简意赅地把“AI 自习室”与数学思维的关系剖析得一干二净。请听过的读者,从 AI 算法的角度,重新思考数学思维的深刻含义。

一位名叫卡稀的在职小学校长,给我讲过一个美国天赋教育老师解析数学公式的故事,我把它改为了类似的“鸡兔同笼”故事。

老师:“有个老人养有兔子和鸭子,某天,老

人看到 5 个头、14 只脚。他看到的是多少只兔子? 多少只鸭?”

同学们纷纷举手。

学生 A:“需要设两个公式。”

学生 B:“ $4(x)+2(y)=14$ ”

学生 C 写道: $x+y=5$

老师:“大家不急于算公式,我们先来猜答案。”

学生们都成了文二和尚……

老师:“你们不吭声,那我来问:5 只兔子和 4 只鸭,对吗?”

学生们:“不对,那就有 9 个头啦。”

老师:“好,我们再从脚上做文章。”

学生们你望我,我望你……

老师:“兔子不少于 4 只,怎么样?”

学生 B:“不对,看我的公式:脚的总数是 14,而 4 只兔子就有 16 只脚。除非老人喝醉了,把自己的脚也数进去。”

学生们哄堂大笑。

老师:“老人没喝醉,是我喝高了。那能不能是 3 只兔子?”

学生们陷入思考……

学生 C:“也不对。看我的公式:共有 5 个头。3 只兔子有 12 只脚。这样就只剩 2 个鸭头和 2 只鸭脚。除非每只鸭只有 1 只脚。”

学生们又是哄堂大笑。

老师:“假设它们都不缺胳膊、不少腿,该有几只鸭呢?”

学生们再也不沉默了,开始议论纷纷……

老师:“如果兔子少于 3 只,我们能在鸭子的数量上做文章吗?”

学生 D:“兔子少于 3 只,鸭子必须 3 只,否则凑不够 5 个头。”

老师:“3 只鸭有几只脚?”

学生们:“6 只鸭脚。”

老师:“OK,这样一来,兔子的数量是多少?”

学生 A:“每只兔子 2 只脚,兔子的脚共有……”

答案呼之欲出。

老师高兴地说:“好,请算公式吧。”

这时,再算公式,小菜一碟:2 只兔子和 3 只鸭。

学生 B:“老师,看您那么高兴,我倒费解了。这个 2 只兔子和 3 只鸭的答案,我们花了一节课,推理来推理去。其实,一开始就用公式,一分钟就能算出来。”

老师:“你提出了一个比‘2 只兔子和 3 只鸭’更有价值的问题。为什么一开始我们没用公式,而是用了一节课来走完整个推理过程?”

学生 E:“我们浪费了不少时间去推论那些不正确的答案。”

学生 C:“我不同意‘浪费’的说法。有时候,你不能证伪一个答案,你就不能证实另一个答案。”

学生 F:“但是,整个推理过程值不值得花一节课?”

这位美国天赋教育老师的教学,非常值得我们深思、反思。

类似“鸡兔同笼”的问题,往往被教成算术课。只讲已知的、现存的公式,忽略学生去证实或证伪自己的假设,就是忽略对学生怀疑精神的培养,甚至是把学生训练成 AI 式的工具。鼓励学生去证实或证伪某个假设,就是教数学思维,教的是“智慧的”、“会学”的思维能力,是动词的“渔”的探索能力,是 AI 算法和算力难以企及的最底层的底料——心智运动。

数学公式之所以强大,不是几个干巴巴的数

字和字母,而是后面的逻辑关系——“算力”,这才是思维强悍之所在,才是学生思想涌动的深层源泉。

有人表示:“以后我们只要知道怎么操作 AI (诸如,怎么生成文件、图画、视频等),只要不是‘小白’,就可以了。”

我们借用那位天赋教育老师解析数学公式的故事,就能理解“学习 AI”和“掌握 AI”的本质不同。数学课不是教算术,更不是用一个似懂非懂的公式去计算一个只有公式才能得到的答案。公式告诉你“做什么(what)? 怎么做(how)?”我们充其量是个计算器。真正理解了“为什么(why)这么做是对的? 为什么(why)那么做是错的?”我们就不简单了。我们要的是通过演绎推理和归纳推理来证实和证伪某些答案,以及在这个过程中培养的推理能力。

对一般人来说,学习 AI 去知道“做什么”以及“怎么做”,焦虑可以解除。但,成为 AI 的专家,必须理解 AI 背后的逻辑——“为什么”这么做是对的?“为什么”那么做是错的? 才能解除大众的焦虑。

AI 可以回答所有考试的问题,但许多“AI 自习室”却训练孩子去刷题,去回答所有考试的问题,这就有点反讽的意味了。其实,数学公式已经回答了现有的相关问题;那么,我们人类还需要什么呢? 需要解决和理解数学公式后面的数学思维和数学逻辑。很多“AI 自习室”只教孩子通过刷题去获得答案;忘却学习数学的初衷是去解构答案后面的数学推理与逻辑……

其实,强化孩子的数学思维,这不是“不能”而是“不为”的问题。千万不能为了效率,通过 AI 去强迫孩子使用一知半解的、死记硬背的公式,从而牺牲孩子的数学思维。

(未完待续)