

用芦苇笔在泥板上
书写苏美尔楔形文字

世界上只有75个人 能流利阅读的文字 竟然学会了

沉睡数千年的“世界第一文明”有望被唤醒

回望历史,数千年前的人类给我们留下了卷帙浩繁的楔形文字文书,以及承载其中的文化思想;展望未来,这些独具远古文化特色的泥碑文本,在人工智能的强大计算能力下,重新组合和解码,以现代科技的数字化形式,重现古代人类文明曾经的辉煌。

无论是科学发现还是艺术创作,人工智能(AI)正向着原来被认为只属于人类的领域进军。在《科学》杂志日前评选出的2022年度十大科学突破中,“创造性人工智能的快速发展”备受关注。从预测蛋白质结构到设计新药、疫苗,AI都取得了“攻城略地”式的进步,甚至在人文艺术领域

频频展现高光。

楔形文字是人类最古老的文字之一,揭开它的神秘面纱,一直是研究人员孜孜以求的目标。如今,人工智能正在帮助破译美索不达米亚人刻写在陶泥板上的楔形文字,这些古老的泥板文书将为我们揭开人类历史早期一些不可思议的故事。

破译楔形文字困难重重

芬克尔是大英博物馆古代美索不达米亚语言文化文献的助理保管员,也是世界上仅有几位能够流利阅读这种久已失传的文字的人之一。他打开一个抽屉,取出一块陶泥制成的泥碑,上面已有了裂纹,还留有灼烧过的痕迹。它不是一块平常的泥碑——上面以世界上最古老的书写文字楔形文字记载了古人对未来的预言。

芬克尔的身后,一位摄影师正对着泥碑上凹陷的蚀刻文字精心拍摄。他们的工作正在揭开一场破译古代文字伟大变革的序幕。这场革命将利用当今人工智能强大无匹的计算能力,唤醒沉睡5000年的古代人类历史,揭开被称为“世界第一文明”埋藏已久的秘密。

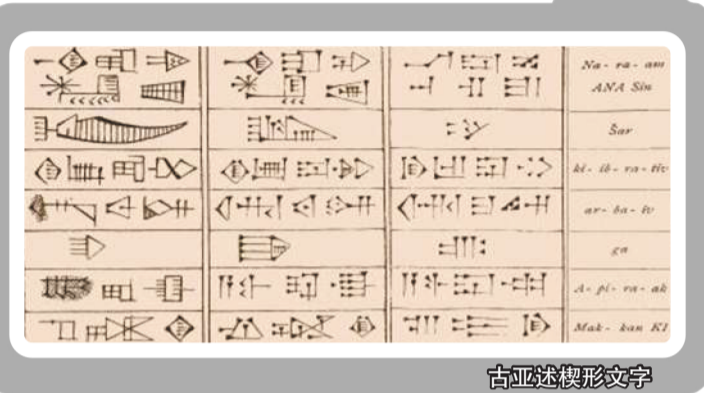
尽管早在165年前,就有人破译了这种古代文字的符号,但大多数楔形文字符号文献资料从未被翻译成现代语言。如今,随着人工智能的发展,计算机算法正在接受阅读和翻译楔形文字的训练,将大量泥碑留下的支离破碎的内容重新组合,并预测其中缺失的部分,为了解古代美索不达米亚人的历史提供一幅完整的图景。

楔形文字从简单的记事工

具演变成为一种语言表达形式,经历了相当漫长的时期。楔形文字本身并不是一种语言,而是一种书写符号系统,好比英语中用的字母,同样也用于法文和德文中一样。苏美尔人从世界上消失之后,他们的楔形文字书写方式流传了下来,成为许多其他语言的书写符号。

我们如今能够了解这段遥远的历史,要归功于用来书写楔形文字的陶泥板。这是当时的一种廉价材料,随处可见,并且可以耐久保存。“有人说,人类历史的前半部都记录在了这些楔形文字泥碑上。”德国慕尼黑大学的恩里克·杰梅内兹说道。

这些泥碑中隐藏着的远古时代的秘密不断被破译。2017年,学者发现,一块被称为“普林顿322”的3700年前的泥板文书是世界上最早出现的三角学知识。2021年,对1894年在伊拉克出土的一块泥碑的新的分析研究发现,早在毕达哥拉斯(古希腊哲学家、数学家)出生几个世纪之前,巴比伦人就已开始对三角学进行计算和研究。遗憾的是,如今世界上只有75个人能够流利阅读楔形文字,大多数泥板文书只能束之高阁,在博物馆里静静蒙尘,鲜为人知。



古亚述楔形文字



贝希斯敦铭文是楔形文字版的“罗塞塔石碑”。

楔形文字太复杂,太难读懂,长期以来一直困扰着人们。“碑文中文字的含义往往模棱两可,而一个词语可能会有不止一种书写方式。”杰梅内兹说。还有一个难题是,很多楔形文字泥碑并不完整,不是破裂缺角,就是碎成多个小块,通常边缘部分还有磨损,导致一段记事不是没有开头,就是没有结尾,或是中间部分缺失。

亚述帝国国王阿什巴尼帕尔的藏书室是世界上现存最古老的皇家图书馆,收藏了来自美索不达米亚各地的大量书写文献,3万块泥碑记载了各种文献资料,包括宗教仪式、医学百科全书、天文观测,以及王室开拓

疆域的功绩等。可惜的是,公元前612年,在城市被攻陷洗劫中,这座古老的皇家图书馆被焚为废墟。大英博物馆保存的大火中幸存下来的一些残碑上面,焦黑的灼烧痕迹仍然历历在目。

要想还原历史,就要将这些碎片像做拼图游戏一样拼合起来,没有任何参考依据,难度很大。而且记载有楔形文字的泥碑分散在世界各处,收集起来也很困难。例如,吉尔伽美什史诗的开头部分保存在一块很小的陶泥碑文书残片上,光是找到并确认这块泥碑,就花了一百多年时间。如今,有了人工智能来帮助做这项工作,一切都将有不同。

他们先用45500个音译短语对不同的人工智能算法进行训练,每个短语由多达19个单词组成,然后对人工智能将苏美尔语单词翻译成英语的能力进行了测试。2021年公布的结果表明,某种特定人工智能算法的翻译准确性可达95%。

研究人员尝试给深度学习人工智能喂入上万块阿卡德语书写的楔形文字泥碑的音译单词信息,发现人工智能根据上下文填补缺失部分内容的准确率能达到80%。人工智能的另一项潜在应用是确定泥版文书的年代,即以已知年代为线索,训练人工智能算法预测其他文件中缺失年代信息的能力。

但是,从符号音译来解读楔形文字是一回事,要解读楔形文字本身又是另一回事。楔形文字不仅随着时间的推移在演变,其拼写也在变化,不同时期的楔形文字还被用于不同的语言。最棘手的是,符号之间没有间隙,要推断出哪组楔形符号构成某个单词非常困难。

为克服这重重困难,研究人员正在开发类似于文本识别的人工智能计算机系统,向破译楔形文字的目标迈进。研究人员训练了一种叫做“Deepscribe”的机器学习系统,用来解读波斯波利斯卫城档案的数千块楔形文字文本。



人工智能对楔形文字泥碑进行扫描翻译示意图

未来有望实现 AI即时翻译

在人工智能的帮助下,未来将有望实现将符号识别系统和现代语言翻译系统一体化的目标。这将意味着也许有一天,我们就可以用手机拍下博物馆里一块泥碑文书的照片,并即时读出上面的内容。

要实现这一宏大目标,需要一个庞大的数字化文本数据库来支持人工智能算法的训练和学习。目前世界各大博物馆中约50万件楔形文字文本中,只有一半经过了音译翻译或文本翻译,经数字化处理的文本只有大约10万件。“楔形文字数字图书馆计划”和“电子巴比伦文学项目”的进展正在改变目前这一现状。

以色列阿里尔大学的萨伊·戈尔丁表示:“楔形文字文本数字化可为学者带来大量新信息和新联系。一旦能够将这些信息放在一个庞大的联通网络中,下一个重大突破的契机就会到来。我们就可以对那个时代古人的生活有一个全面的了解。”

开始阶段需要大量辛苦的工作,首先要获得各地博物馆和私人藏品中所有楔形文字泥碑文书的高清图像。目前,作为“电子巴比伦文学项目”的组成部分之一,大英博物馆收藏的阿什巴尼帕尔图书馆4万件楔形文字泥碑残片的拍摄工作正在进行之中。据预计,2023年将完成所有楔形文字藏品的拍摄工作,并向大众公开。以往深藏于大英博物馆内的楔形文字藏品不是普通大众都有机会观赏到的,但经过数字化处理的藏品,可以让人们舒服地坐在自己家中,欣赏苏美尔阿什巴尼帕尔图书馆的泥碑文献和其他早已失传的楔形文字文本。

本报综合

AI训练“复原”巴比伦文学

经过训练的人工智能可以对碑文进行翻译。研究人员根据文本符号的音译,对人工智能进行深度学习训练。经过学习训练的人工智能可以预测哪些楔形符号可能出现在缺失的残片中,还可以在数字化的庞大残片数据库中搜索特定的楔形文字符号。

2019年,人工智能帮助在多个残片中找到了吉尔伽美什史诗中的缺失部分,还发现了一种新的古代文学体裁,其特点是滑稽、夸张和模仿,比如一堆会开玩笑的驴粪,主要是用来帮助小学生学习读写技能的。2021年,在世界上首次使用人工智能完全自主识别楔形文字片段的

过程中,研究人员发现了名诗《正义的受难者》中缺失的片段。

苏美尔泥碑文本中通常包含个人姓名和日期,根据这些线索,可以追溯了解某人在当时社会中承担的角色。例如,形成于公元前2300年左右,被称为“Mama-ummi档案”的80块泥碑文本描述了一位名叫Mama ummi的女主管,她负责一个由180名织工组成的团队。这份资料表明,当时的女性有很多的工作机会,这一点非常出人意料。

面对浩如烟海的楔形文字行政信息,加拿大多伦多大学的希瑟·贝克于2017年建立了完美的楔形文字机器翻译和自动分析系统。在不久前的测试中,