# 1226

## 1.中国工程院院士孙凝晖给正国级、副国级讲课的万字长稿《人工智能与智能计算的发展》 【1】



委员长、各位副委员长、秘书长、各位委员：

人工智能领域近年来正在迎来一场由生成式人工智能大模型引领的爆发式发展。

2022年11月30日，OpenAI公司推出一款人工智能对话聊天机器人ChatGPT，

其出色的自然语言生成能力引起了全世界范围的广泛关注，2个月突破1亿用户，

国内外随即掀起了一场大模型浪潮，Gemini、文心一言、Copilot、LLaMA、SAM、SORA等各种大模型如雨后春笋般涌现，

2022年也被誉为大模型元年。

当前信息时代正加快进入智能计算的发展阶段，

人工智能技术上的突破层出不穷，逐渐深入地赋能千行百业，推动人工智能与数据要素成为新质生产力的典型代表。

习近平总书记指出，

把新一代人工智能作为推动科技跨越发展、产业优化升级、生产力整体跃升的驱动力量，

努力实现高质量发展。

党的十八大以来，以习近平同志为核心的党中央高度重视智能经济发展，

促进人工智能和实体经济深度融合，为高质量发展注入强劲动力。

### ****一、计算技术发展简介****

计算技术的发展历史大致可分为四个阶段，

算盘的出现标志着人类进入

第一代——机械计算时代，

第二代——电子计算的标志是出现电子器件与电子计算机，

互联网的出现使我们进入

第三代——网络计算，

当前人类社会正在进入

第四阶段——智能计算。

早期的计算装置是手动辅助计算装置和半自动计算装置，

人类计算工具的历史是从公元1200年的中国算盘开始，

随后出现了纳皮尔筹（1612年）和滚轮式加法器（1642年），

到1672年第一台自动完成四则运算的计算装置——步进计算器诞生了。

机械计算时期已经出现了现代计算机的一些基本概念。

查尔斯∙巴贝奇（Charles Babbage）提出了差分机（1822年）与分析机（1834年）的设计构想，

支持自动机械计算。

这一时期，编程与程序的概念基本形成，

编程的概念起源于雅卡尔提花机，通过打孔卡片控制印花图案，

最终演变为通过计算指令的形式来存储所有数学计算步骤；

人类历史的第一个程序员是诗人拜伦之女艾达（Ada），

她为巴贝奇差分机编写了一组求解伯努利数列的计算指令，

这套指令也是人类历史上第一套计算机算法程序，

它将硬件和软件分离，第一次出现程序的概念。

直到在二十世纪上半叶，出现了布尔代数(数学)、图灵机(计算模型) 、冯诺依曼体系结构(架构) 、晶体管(器件)

这四个现代计算技术的科学基础。

其中，布尔代数用来描述程序和硬件如CPU的底层逻辑；

图灵机是一种通用的计算模型，将复杂任务转化为自动计算、不需人工干预的自动化过程；

冯诺依曼体系结构提出了构造计算机的三个基本原则：

采用二进制逻辑、程序存储执行、以及计算机由运算器、控制器、存储器、输入设备、输出设备这五个基本单元组成；

晶体管是构成基本的逻辑电路和存储电路的半导体器件，是建造现代计算机之塔的“砖块”。

基于以上科学基础，计算技术得以高速发展，形成规模庞大的产业。

从1946年世界上第一台电子计算机ENIAC诞生到二十一世纪的今天，已经形成了五类成功的平台型计算系统。

当前各领域各种类型的应用，都可以由这五类平台型计算装置支撑。

第一类是高性能计算平台，解决了国家核心部门的科学与工程计算问题；

第二类是企业计算平台，又称服务器，用于企业级的数据管理、事务处理，

当前像百度、阿里和腾讯这些互联网公司的计算平台都属于这一类；

第三类是个人电脑平台，以桌面应用的形式出现，人们通过桌面应用与个人电脑交互；

第四类是智能手机，主要特点是移动便携，手机通过网络连接数据中心，以互联网应用为主，

它们分布式地部署在数据中心和手机终端；

第五类是嵌入式计算机，嵌入到工业装备和军事设备，通过实时的控制，保障在确定时间内完成特定任务。

这五类装置几乎覆盖了我们信息社会的方方面面，

长期以来人们追求的以智能计算应用为中心的第六类平台型计算系统尚未形成。

现代计算技术的发展大致可以划分为三个时代。

IT1.0又称电子计算时代（1950-1970），基本特征是以“机”为中心。

计算技术的基本架构形成，随着集成电路工艺的进步，基本计算单元的尺度快速微缩，

晶体管密度、计算性能和可靠性不断提升，计算机在科学工程计算、企业数据处理中得到了广泛应用。

IT2.0又称网络计算时代（1980-2020），以“人”为中心。

互联网将人使用的终端与后台的数据中心连接，互联网应用通过智能终端与人进行交互。

以亚马逊等为代表的互联网公司提出了云计算的思想，

将后台的算力封装成一个公共服务租借给第三方用户，

形成了云计算与大数据产业。

IT3.0又称智能计算时代，始于2020年，与IT2.0相比增加了“物”的概念，

即物理世界的各种端侧设备，被数字化、网络化和智能化，实现“人-机-物”三元融合。

智能计算时代，除了互联网以外，还有数据基础设施，支撑各类终端通过端边云实现万物互联，

终端、物端、边缘、云都嵌入AI，提供与ChatGPT类似的大模型智能服务，

最终实现有计算的地方就有AI智能。

智能计算带来了巨量的数据、人工智能算法的突破和对算力的爆发性需求。

### ****二、智能计算发展简介****

智能计算包括人工智能技术与它的计算载体，大致历经了四个阶段，

分别为通用计算装置、逻辑推理专家系统、深度学习计算系统、大模型计算系统。

智能计算的起点是通用自动计算装置（1946年）。

艾伦·图灵（Alan Turing）和冯·诺依曼（John von Neumann）等科学家，

一开始都希望能够模拟人脑处理知识的过程，

发明像人脑一样思考的机器，虽未能实现，但却解决了计算的自动化问题。

通用自动计算装置的出现，也推动了1956年人工智能（AI）概念的诞生，

此后所有人工智能技术的发展都是建立在新一代计算设备与更强的计算能力之上的。

智能计算发展的第二阶段是逻辑推理专家系统（1990年）。

1. A.费根鲍姆（Edward Albert Feigenbaum）等符号智能学派的科学家以逻辑和推理能力自动化为主要目标，

提出了能够将知识符号进行逻辑推理的专家系统。

人的先验知识以知识符号的形式进入计算机，

使计算机能够在特定领域辅助人类进行一定的逻辑判断和决策，

但专家系统严重依赖于手工生成的知识库或规则库。

这类专家系统的典型代表是日本的五代机和我国863计划支持的306智能计算机主题，

日本在逻辑专家系统中采取专用计算平台和Prolog这样的知识推理语言完成应用级推理任务；

我国采取了与日本不同的技术路线，以通用计算平台为基础，

将智能任务变成人工智能算法，将硬件和系统软件都接入通用计算平台，

并催生了曙光、汉王、科大讯飞等一批骨干企业。

符号计算系统的局限性在于其爆炸的计算时空复杂度，

即符号计算系统只能解决线性增长问题，

对于高维复杂空间问题是无法求解的，

从而限制了能够处理问题的大小。

同时因为符号计算系统是基于知识规则建立的，

我们又无法对所有的常识用穷举法来进行枚举，

它的应用范围就受到了很大的限制。

随着第二次AI寒冬的到来，第一代智能计算机逐渐退出历史舞台。

直到2014年左右，智能计算进阶到第三阶段——深度学习计算系统。

以杰弗里·辛顿（Geoffrey Hinton）等为代表的连接智能学派，以学习能力自动化为目标，发明了深度学习等新AI算法。

通过深度神经元网络的自动学习，大幅提升了模型统计归纳的能力，

在模式识别①等应用效果上取得了巨大突破，某些场景的识别精度甚至超越了人类。

以人脸识别为例，整个神经网络的训练过程相当于一个网络参数调整的过程，

将大量的经过标注的人脸图片数据输入神经网络，然后进行网络间参数调整，

让神经网络输出的结果的概率无限逼近真实结果。

神经网络输出真实情况的概率越大，参数就越大，

从而将知识和规则编码到网络参数中，这样只要数据足够多，

就可以对各种大量的常识进行学习，通用性得到极大的提升。

连接智能的应用更加广泛，包括语音识别、人脸识别、自动驾驶等。

在计算载体方面，中国科学院计算技术研究所2013年提出了国际首个深度学习处理器架构，

国际知名的硬件厂商英伟达（NVIDIA）持续发布了多款性能领先的通用GPU芯片，

都是深度学习计算系统的典型代表。

智能计算发展的第四阶段是大模型计算系统（2020年）。

在人工智能大模型技术的推动下，智能计算迈向新的高度。

2020年，AI从“小模型+判别式”转向“大模型+生成式”，

从传统的人脸识别、目标检测、文本分类，

升级到如今的文本生成、3D数字人生成、图像生成、语音生成、视频生成。

大语言模型在对话系统领域的一个典型应用是OpenAI公司的ChatGPT，

它采用预训练基座大语言模型GPT-3，引入3000亿单词的训练语料，相当于互联网上所有英语文字的总和。

其基本原理是：通过给它一个输入，让它预测下一个单词来训练模型，

通过大量训练提升预测精确度，最终达到向它询问一个问题，大模型产生一个答案，与人即时对话。

在基座大模型的基础上，再给它一些提示词进行有监督的指令微调，

通过人类的<指令，回复>对逐渐让模型学会如何与人进行多轮对话；

最后，通过人为设计和自动生成的奖励函数来进行强化学习迭代，逐步实现大模型与人类价值观的对齐。

大模型的特点是以“大”取胜，其中有三层含义，

1. 参数大，GPT-3就有1700亿个参数；
2. 训练数据大，ChatGPT大约用了3000亿个单词，570GB训练数据；
3. 算力需求大，GPT-3大约用了上万块V100 GPU进行训练。

为满足大模型对智能算力爆炸式增加的需求，国内外都在大规模建设耗资巨大的新型智算中心，

英伟达公司也推出了采用256个H100芯片，

150TB海量GPU内存等构成的大模型智能计算系统。

大模型的出现带来了三个变革。

一是技术上的规模定律（Scaling Law），

即很多AI模型的精度在参数规模超过某个阈值后模型能力快速提升，

其原因在科学界还不是非常清楚，有很大的争议。

AI模型的性能与模型参数规模、数据集大小、算力总量三个变量成“对数线性关系”，

因此可以通过增大模型的规模来不断提高模型的性能。

目前最前沿的大模型GPT-4参数量已经达到了万亿到十万亿量级，并且仍在不断增长中；

二是产业上算力需求爆炸式增长，千亿参数规模大模型的训练通常需要在数千乃至数万GPU卡上训练2-3个月时间，

急剧增加的算力需求带动相关算力企业超高速发展，

英伟达的市值接近两万亿美元，对于芯片企业以前从来没有发生过；

三是社会上冲击劳动力市场，北京大学国家发展研究院与智联招聘联合发布的

《AI大模型对我国劳动力市场潜在影响研究》报告指出，【2】

受影响最大的20个职业中财会、销售、文书位于前列，

需要与人打交道并提供服务的体力劳动型工作，

如人力资源、行政、后勤等反而相对更安全。

人工智能的技术前沿将朝着以下四个方向发展。

**第一个前沿方向为多模态大模型。**

从人类视角出发，人类智能是天然多模态的，人拥有眼、耳、鼻、舌、身、嘴(语言)，

从AI视角出发，视觉，听觉等也都可以建模为token②的序列，

可采取与大语言模型相同的方法进行学习，

并进一步与语言中的语义进行对齐，实现多模态对齐的智能能力。

**第二个前沿方向为视频生成大模型。**

OpenAI于2024年2月15日发布文生视频模型SORA，

将视频生成时长从几秒钟大幅提升到一分钟，

且在分辨率、画面真实度、时序一致性等方面都有显著提升。

SORA的最大意义是它具备了世界模型的基本特征，

即人类观察世界并进一步预测世界的能力。

世界模型是建立在理解世界的基本物理常识（如，水往低处流等）之上，

然后观察并预测下一秒将要发生什么事件。

虽然SORA要成为世界模型仍然存在很多问题，

但可以认为SORA学会了画面想象力和分钟级未来预测能力，这是世界模型的基础特征。

**第三个前沿方向为具身智能。**

具身智能指有身体并支持与物理世界进行交互的智能体，如机器人、无人车等，

通过多模态大模型处理多种传感数据输入，

由大模型生成运动指令对智能体进行驱动，

替代传统基于规则或者数学公式的运动驱动方式，

实现虚拟和现实的深度融合。

因此，具有具身智能的机器人，

可以聚集人工智能的三大流派：

以神经网络为代表的连接主义，

以知识工程为代表的符号主义

和控制论相关的行为主义，

三大流派可以同时作用在一个智能体，

这预期会带来新的技术突破。

**第四个前沿方向是AI4R(AI for Research)成为科学发现与技术发明的主要范式。**

当前科学发现主要依赖于实验和人脑智慧，

由人类进行大胆猜想、小心求证，

信息技术无论是计算和数据，都只是起到一些辅助和验证的作用。

相较于人类，人工智能在记忆力、高维复杂、全视野、推理深度、猜想等方面具有较大优势，

是否能以AI为主进行一些科学发现和技术发明，大幅提升人类科学发现的效率，

比如主动发现物理学规律、预测蛋白质结构、设计高性能芯片、高效合成新药等。

因为人工智能大模型具有全量数据，具备上帝视角，通过深度学习的能力，

可以比人向前看更多步数，如能实现从推断(inference)到推理(reasoning)的跃升，

人工智能模型就有潜力具备爱因斯坦一样的想象力和科学猜想能力，

极大提升人类科学发现的效率，打破人类的认知边界。这才是真正的颠覆所在。

**最后，通用人工智能③（Artificial General Intelligence，简称AGI）是一个极具挑战的话题，极具争论性。**

曾经有一个哲学家和一个神经科学家打赌：

25年后（即2023年）科研人员是否能够揭示大脑如何实现意识？

当时关于意识有两个流派，一个叫集成信息理论，一个叫全局网络工作空间理论，

前者认为意识是由大脑中特定类型神经元连接形成的“结构”，

后者指出意识是当信息通过互连网络传播到大脑区域时产生的。

2023年，人们通过六个独立实验室进行了对抗性实验，结果与两种理论均不完全匹配，哲学家赢了，神经科学家输了。

通过这一场赌约，可以看出人们总是希望人工智能能够了解人类的认知和大脑的奥秘。

从物理学的视角看，物理学是对宏观世界有了透彻理解后，

从量子物理起步开启了对微观世界的理解。

智能世界与物理世界一样，都是具有巨大复杂度的研究对象，

AI大模型仍然是通过数据驱动等研究宏观世界的方法，

提高机器的智能水平，对智能宏观世界理解并不够，

直接到神经系统微观世界寻找答案是困难的。

人工智能自诞生以来，一直承载着人类关于智能与意识的种种梦想与幻想，也激励着人们不断探索。

### ****三、人工智能的安全风险****

人工智能的发展促进了当今世界科技进步的同时，也带来了很多安全风险，要从技术与法规两方面加以应对。

首先是互联网虚假信息泛滥。这里列举若干场景：

**一是数字分身。**

AI Yoon是首个使用 DeepFake 技术合成的官方“候选人”，

这个数字人以韩国国民力量党候选人尹锡悦（Yoon Suk-yeol）为原型，

借助尹锡悦 20 小时的音频和视频片段、

以及其专门为研究人员录制的 3000 多个句子，

由当地一家 DeepFake 技术公司创建了虚拟形象 AI Yoon，

并在网络上迅速走红。

实际上 AI Yoon 表达的内容是由竞选团队撰写的，而不是候选人本人。

**二是伪造视频，**

尤其是伪造领导人视频引起国际争端，扰乱选举秩序，或引起突发舆情事件，

如伪造尼克松宣布第一次登月失败，

伪造乌克兰总统泽连斯基宣布“投降”的信息，

这些行为导致新闻媒体行业的社会信任衰退。

**三是伪造新闻，**

主要通过虚假新闻自动生成牟取非法利益，

使用ChatGPT生成热点新闻，赚取流量，

截至2023年6月30日全球生成伪造新闻网站已达277个，严重扰乱社会秩序。

**四是换脸变声，**

用于诈骗。如由于AI语音模仿了企业高管的声音，

一家香港国际企业因此被骗3500万美元。

**五是生成不雅图片，**

特别是针对公众人物。

如影视明星的色情视频制作，造成不良社会影响。

因此，迫切需要发展互联网虚假信息的伪造检测技术。

其次，AI大模型面临严重可信问题。

这些问题包括：

1. “一本正经胡说八道”的事实性错误；
2. 以西方价值观叙事，输出政治偏见和错误言论；
3. 易被诱导，输出错误知识和有害内容；
4. 数据安全问题加重，大模型成为重要敏感数据的诱捕器，

ChatGPT将用户输入纳入训练数据库，用于改善ChatGPT，

美方能够利用大模型获得公开渠道覆盖不到的中文语料，掌握我们自己都可能不掌握的“中国知识”。

因此，迫切需要发展大模型安全监管技术与自己的可信大模型。

除了技术手段外，人工智能安全保障需要相关立法工作。

2021年科技部发布《新一代人工智能伦理规范》，

2022年8月，全国信息安全标准化技术委员会发布《信息安全技术 机器学习算法安全评估规范》，

2022-2023年，中央网信办先后发布《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理办法》等。

欧美国家也先后出台法规，2018年5月25日，欧盟出台《通用数据保护条例》，

2022年10月4日，美国发布《人工智能权利法案蓝图》，

2024年3月13日，欧洲议会通过了欧盟《人工智能法案》。

我国应加快推进《人工智能法》出台，构建人工智能治理体系，

确保人工智能的发展和应用遵循人类共同价值观，促进人机和谐友好；

创造有利于人工智能技术研究、开发、应用的政策环境；

建立合理披露机制和审计评估机制，理解人工智能机制原理和决策过程；

明确人工智能系统的安全责任和问责机制，可追溯责任主体并补救；

推动形成公平合理、开放包容的国际人工智能治理规则。

### ****四、中国智能计算发展困境****

人工智能技术与智能计算产业处于中美科技竞争的焦点，

我国在过去几年虽然取得了很大的成绩，但依然面临诸多发展困境，

特别是由美国的科技打压政策带来的困难。

**困境一为美国在AI核心能力上长期处于领先地位，中国处于跟踪模式。**

中国在AI高端人才数量、

AI基础算法创新、

AI底座大模型能力（大语言模型、文生图模型、文生视频模型）、

底座大模型训练数据、

底座大模型训练算力等，

都与美国存在一定的差距，并且这种差距还将持续很长一段时间。

**困境二为高端算力产品禁售，高端芯片工艺长期被卡。**

A100，H100，B200等高端智算芯片对华禁售。

华为、龙芯、寒武纪、曙光、海光等企业都进入实体清单，

它们芯片制造的先进工艺④受限，

国内可满足规模量产的工艺节点落后国际先进水平2-3代，

核心算力芯片的性能落后国际先进水平2-3代。

**困境三为国内智能计算生态孱弱，AI开发框架渗透率不足。**

英伟达CUDA⑤(Compute Unified Device Architecture, 通用计算设备架构)生态完备，已形成了事实上的垄断。

国内生态孱弱，具体表现在：

一是研发人员不足，英伟达CUDA生态有近2万人开发，是国内所有智能芯片公司人员总和的20倍；

二是开发工具不足，CUDA有550个SDK(Software Development Kit, 软件开发工具包)，是国内相关企业的上百倍；

三是资金投入不足，英伟达每年投入50亿美元，是国内相关公司的几十倍；

四是AI开发框架TensorFlow占据工业类市场，PyTorch占据研究类市场，

百度飞桨等国产AI开发框架的开发人员只有国外框架的1/10。

更为严重的是国内企业之间山头林立，无法形成合力，

从智能应用、开发框架、系统软件、智能芯片，

虽然每层都有相关产品，但各层之间没有深度适配，无法形成一个有竞争力的技术体系。

**困境四为AI应用于行业时成本、门槛居高不下。**

当前我国AI应用主要集中在互联网行业和一些国防领域。

AI技术推广应用于各行各业时，特别是从互联网行业迁移到非互联网行业，

需要进行大量的定制工作，

迁移难度大，单次使用成本高。

最后，我国在AI领域的人才数量与实际需求相比也明显不足。

### ****五、中国如何发展智能计算的道路选择****

人工智能发展的道路选择对我国至关重要，

关系到发展的可持续性与最终的国际竞争格局。

当前人工智能的使用成本十分高昂，

微软Copilot套件要支付每月10美元的使用费用，

ChatGPT每天消耗50万千瓦时的电力，

英伟达B200芯片价格高达3万美元以上。

总体来说，我国应发展用得起、安全可信的人工智能技术，

消除我国信息贫困人口、并造福“一带一路”国家；

低门槛地赋能各行各业，让我国的优势产业保持竞争力，让相对落后的产业能够大幅地缩小差距。

**选择一：统一技术体系走闭源封闭，还是开源开放的道路？**

支撑智能计算产业的是一个相互紧耦合的技术体系，

即由一系列技术标准和知识产权将材料、器件、工艺、芯片、整机、系统软件、应用软件等密切联系在一起的技术整体。

我国发展智能计算技术体系存在三条道路：

**一是追赶兼容美国主导的A体系。**

我国大多数互联网企业走的是GPGPU/CUDA兼容道路，

很多芯片领域的创业企业在生态构建上也是尽量与CUDA兼容，这条道路较为现实。

由于在算力方面美国对我国工艺和芯片带宽的限制，

在算法方面国内生态林立很难形成统一，

生态成熟度严重受限，

在数据方面中文高质量数据匮乏，

这些因素会使得追赶者与领先者的差距很难缩小，

一些时候还会进一步拉大。

**二是构建专用封闭的B体系。**

在军事、气象、司法等专用领域构建企业封闭生态，

基于国产成熟工艺生产芯片，

相对于底座大模型更加关注特定领域垂直类大模型，

训练大模型更多采用领域专有高质量数据等。

这条道路易于形成完整可控的技术体系与生态，

我国一些大型骨干企业走的是这条道路，它的缺点是封闭，无法凝聚国内大多数力量，也很难实现全球化。

**三是全球共建开源开放的C体系。**

用开源打破生态垄断，降低企业拥有核心技术的门槛，

让每个企业都能低成本地做自己的芯片，形成智能芯片的汪洋大海，满足无处不在的智能需求。

用开放形成统一的技术体系，

我国企业与全球化力量联合起来共建基于国际标准的统一智能计算软件栈。

形成企业竞争前共享机制，共享高质量数据库，共享开源通用底座大模型。

对于全球开源生态，我国企业在互联网时代收益良多，

我国更多的是使用者，是参与者，

在智能时代我国企业在RISC-V⑥+AI开源技术体系上应更多地成为主力贡献者，

成为全球化开放共享的主导力量。

**选择二：拼算法模型，还是拼新型基础设施？**

人工智能技术要赋能各行各业，具有典型的长尾效应⑦。

我国80%的中小微企业，需要的是低门槛、低价格的智能服务。

因此，我国智能计算产业必须建立在新的数据空间基础设施之上，

其中关键是我国应率先实现智能要素即数据、算力、算法的全面基础设施化。

这项工作可比肩二十世纪初美国信息高速公路计划（即信息基础设施建设）对互联网产业的历史作用。

信息社会最核心的生产力是网络空间(Cyberspace)。

网络空间的演进过程是：

从机器一元连接构成的计算空间，

演进到人机信息二元连接构成的信息空间，

再演进到人机物数据三元连接构成的数据空间。

从数据空间看，人工智能的本质是数据的百炼成钢，

大模型就是对互联网全量数据进行深度加工后的产物。

在数字化时代，在互联网上传输的是信息流，

是算力对数据进行粗加工后的结构化抽象；

在智能时代，在互联网上传输的是智能流，

是算力对数据进行深度加工与精炼后的模型化抽象。

智能计算的一个核心特征就是用数值计算、数据分析、人工智能等算法，

在算力池中加工海量数据件，得到智能模型，再嵌入到信息世界、物理世界的各个过程中。

我国政府已经前瞻性地提前布局了新型基础设施，

在世界各国竞争中抢占了先机。

**首先，数据已成为国家战略信息资源。**

数据具有资源要素与价值加工两重属性，

数据的资源要素属性包括生产、获取、传输、汇聚、流通、交易、权属、资产、安全等各个环节，

我国应继续加大力度建设国家数据枢纽与数据流通基础设施。

**其次，AI大模型就是数据空间的一类算法基础设施。**

以通用大模型为基座，

构建大模型研发与应用的基础设施，

支撑广大企业研发领域专用大模型，

服务于机器人、无人驾驶、可穿戴设备、智能家居、智能安防等行业，覆盖长尾应用。

**最后，全国一体化算力网建设在推动算力的基础设施化上发挥了先导作用。**

算力基础设施化的中国方案，应在大幅度降低算力使用成本和使用门槛的同时，

为最广范围覆盖人群提供高通量、高品质的智能服务。

算力基础设施的中国方案需要具备“两低一高”，

即在供给侧，

大幅度降低算力器件、算力设备、

网络连接、数据获取、算法模型调用、

电力消耗、运营维护、开发部署的总成本，

让广大中小企业都消费得起高品质的算力服务，

有积极性开发算力网应用；

在消费侧，大幅度降低广大用户的算力使用门槛，

面向大众的公共服务必须做到易获取、易使用，像水电一样即开即用，

像编写网页一样轻松定制算力服务，开发算力网应用。

在服务效率侧，中国的算力服务要实现低熵高通量，

其中高通量是指在实现高并发⑧度服务的同时，

端到端服务的响应时间可满足率高；

低熵是指在高并发负载中出现资源无序竞争的情况下，

保障系统通量不急剧下降。保障“算得多”对中国尤其重要。

**选择三：AI+着重赋能虚拟经济，还是发力实体经济？**

“AI+”的成效是人工智能价值的试金石。

次贷危机后，美国制造业增加值占GDP的比重从1950年的28%降低为2021年的11%，

美国制造业在全行业就业人数占比从1979年的35%降低为2022年的8%，

可见美国更倾向于回报率更高的虚拟经济，

轻视投资成本高且经济回报率低的实体经济。

中国倾向于实体经济与虚拟经济同步发展，

更加重视发展装备制造、新能源汽车、光伏发电、锂电池、高铁、5G等实体经济。

相应地美国AI主要应用于虚拟经济和IT基础工具，AI技术也是“脱实向虚”，

自2007年以来硅谷不断炒作虚拟现实（Virtual Reality，VR）、元宇宙、区块链、Web3.0、深度学习、AI大模型等，

是这个趋势的反映。

我国的优势在实体经济，制造业全球产业门类最齐全，体系最完整，特点是场景多、私有数据多。

我国应精选若干行业加大投入，

形成可低门槛全行业推广的范式，

如选择装备制造业作为延续优势代表性行业，

选择医药业作为快速缩短差距的代表性行业。

赋能实体经济的技术难点是AI算法与物理机理的融合。

人工智能技术成功的关键是能否让一个行业或一个产品的成本大幅下降，

从而将用户数与产业规模扩大10倍，

产生类似于蒸汽机对于纺织业，

智能手机对于互联网业的变革效果。

我国应走出适合自己的人工智能赋能实体经济的高质量发展道路。

**主讲人系中国工程院院士，中国科学院计算技术研究所研究员、学术委员会主任**

文章来源：抗衰老产业联盟、中国信息界

免责声明：本文为网络摘录或转载，版权归原作者所有，内容为原作者个人观点，并不代表本公众号赞同其观点和对其真实性负责。如涉及作品版权问题，请与我们联系，我们将在第一时间删除内容！

## Refer

1. 中国工程院院士孙凝晖给正国级、副国级讲课的万字长稿《人工智能与智能计算的发展》https://mp.weixin.qq.com/s/4a\_u\_9kiO7QUwq6PVwYoLg
2. 《AI大模型对我国劳动力市场潜在影响研究》2024北大国发院 https://news.qq.com/rain/a/20241008A08B4O00