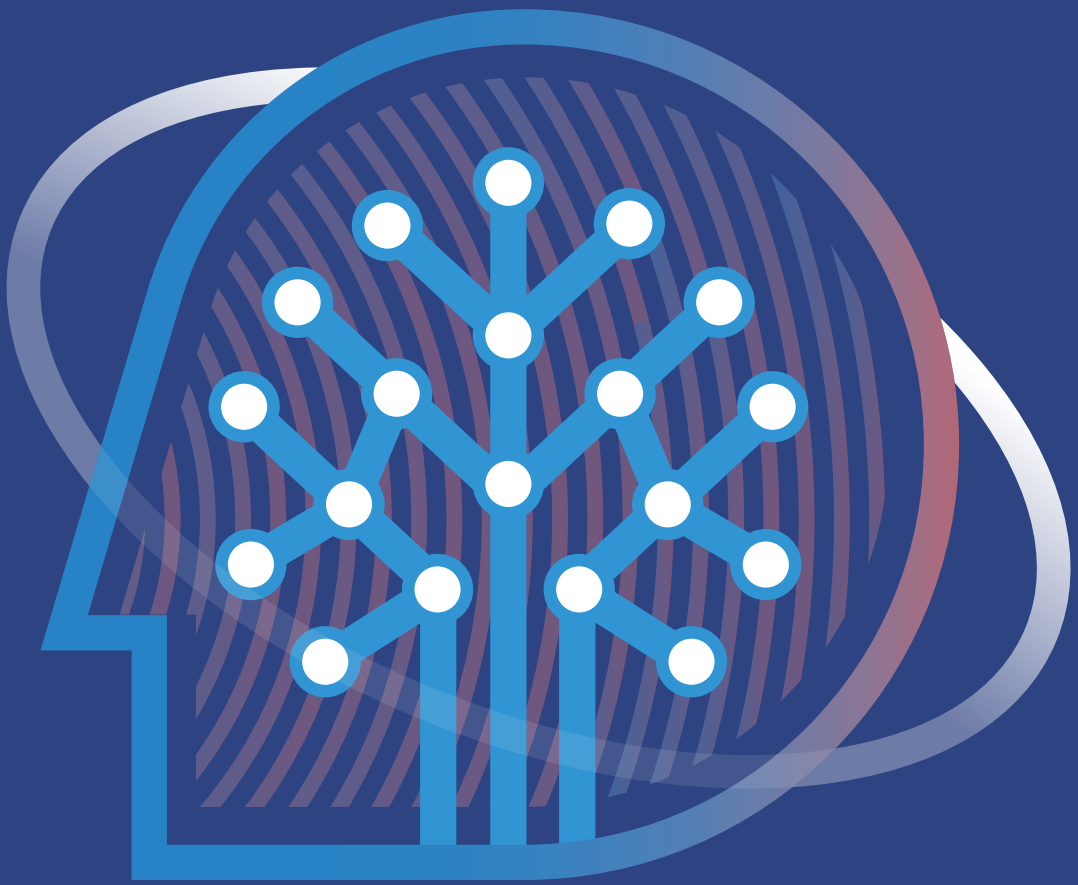
人本智能

人机共生时代的科技发展观



出品

财新智库 Caixin Insight

ESG30 中国ESG30人论坛

联合出品

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

人工智能研究院 Artificial Intelligence Institute

Lenovo 联想


AI

人本智能

人机共生时代的科技发展观

PREFACE
序言

> *“技术创新对生活的影响是巨大的，但这并不是自动发生的。它取决于我们发明的技术类型以及我们如何使用它们。”*

——2024 年诺贝尔经济学奖获得者、麻省理工学院教授达伦 · 阿西莫格鲁（Daron Acemoglu）

2022 年 11 月 30 日，人工智能公司 OpenAI 发布聊天机器人模型 ChatGPT。它以前所未有的生成能力、广泛的应用场景以及像人类般互动的自然语言交互模式，让大众第一次真切感受到人工智能（Artificial Intelligence，以下简称“AI”）的魅力。由此开启的新一轮 AI 大模型技术浪潮，正深刻改变着现在乃至未来人类社会的生产生活方式。

事实上，自 1956 年达特茅斯会议首次提出“人工智能”这一概念以来，人工智能技术至今已经历经大约 70 年的发展历程。它曾带来过技术变革的期望，也曾经历过产业发展的低谷。然而，此次浪潮所引发的关注度前所未有——公众情绪由最初的旁观、震惊，逐渐演变为一种夹杂着期盼与焦虑的复杂情绪，各种疑虑也不断产生。比如，AI 是否会让人变得更有创造力？AI 能否真正给人们的生活带来便捷并提升品质？AI 会威胁甚至取代自己的工作吗？如何确保 AI 技术不被滥用、不会侵犯个人隐私和安全……特别是，一些引发大众关注的热点事件将当前 AI 发展中缺乏对人关注和保护的不足集中暴露出来。

透过 AI 纷繁复杂的发展背后，人们希望回归两个基本的起点：一是 AI 作为一种技术的工具属性这一本质；二是发展 AI 的初衷和目标——人本，或者说以人为中心。

一方面，人类发明任何工具，出发点都是将人类从繁杂枯燥的生产活动中解放出来，从而帮助其更好地生活。作为一个技术范畴的统称，AI 开始于 20 世纪 50 年代初计算机、物理、数学、心理学、神经科学等不同领域的学者开始研究如何让机器像人类一样思考和行动。在接下来的几十年里，人工智能领域经历了多次的高潮和低谷，但其核心的工具本质，始终未曾改变。

另一方面，AI 系统将人作为学习和模仿的对象，通过复杂的算法和大量数据训练，学习人类的语言、行为模式、决策过程乃至创造性的表达。对于 AI 而言，人始终是学习进化的基点和目标。

一个 AI 系统本质上是一个从输入信息到生成行为的转换系统。人类的任务是设计这个转换机制，然而现实的发展可能会偏离起点和初衷。正如历史学家尤瓦尔 · 赫拉利所做的论断，AI 是“历史上第一个可以自己做决定的技术，也是历史上第一个可以自己创造想法的技术”[1]。对此，人们有必要保持一定的审慎态度，特别是在底层价值认知方面进行充分的反思。从技术这一大类的属性来看，人类可以赋予技术以灵魂，但是反过来，人类这一物种的生物和社会属性可能在很大程度上被技术所左右；从 AI 本身的技术特性来看，不同于以往的技术或者技术变革，AI 有望在人类体验的所有领域引发变革并重塑价值观，改变人类理解现实的方式以及在其中扮演的角色。如何确保在这一进程中 AI 始终以人为目的，是一个严峻的考验。

人们需要思考，在 AI 发展如火如荼的大潮下，人们应该以什么样的价值观来推进 AI 技术和产业的变革及治理，以什么样的关系来处理 AI 与人类的关系，如何将价值观置于技术之上，进而拥抱 AI。正如“AI 教母”、华裔科学家李飞飞在其自传《我看见的世界》中所言，“如果人工智能要帮助人们，我们的思考就必须从人们本身开始”。人工智能最伟大的胜利不仅是科学的，还是人文的。人的尊严、人的快乐、人的安全、人的幸福，是人工智能技术发展的北极星——人工智能，以人为本。

1. 尤瓦尔 · 赫拉利，《智人之上：从石器时代到 AI 时代的信息网络简史》，中信出版集团，2024 年

CONTENTS
目录

CHAPTER 1
第一章

新型人机关系，新 AI 价值观

1.1 人工智能的四轮发展浪潮

人工智能正以惊人的速度重塑着世界。在近 70 年的发展历程中，人工智能经历过黄金时代也曾有过低谷。不过科技的魅力在于，历经起起伏伏之后，现在的人工智能已开始深深影响人类社会。总体来看，人工智能技术的发展历经了七个阶段共四轮发展浪潮。

1. 起步发展期（1943–1960 年）

人工智能从概念逐渐演变为学科，并涌现出了两大学派：符号主义和联结主义。人工智能研究者提出了一些基本的概念和方法，如神经网络、图灵测试、符号推理、游戏 AI 等，并在一些简单的任务处理上取得了初步成功，如机器定理证明、跳棋程序、人机对话等。

其中最具标志性的事件是 1956 年夏天，美国达特茅斯学院主办了历史上第一次人工智能研讨会。会议虽然未能达成普遍的共识，却为所讨论的内容起了一个名字：人工智能。从此“人工智能”开始作为一门独立学科出现，1956 年也因此成为人工智能元年。

2. 黄金时代（1960–1974 年）

人工智能的黄金年代，也是符号主义的鼎盛时期。这一时期科学家们富有理想、信心，认为机器能够实现与人类同等水平的智能，他们试图用逻辑和符号来模拟人类思维过程，并已经在自然语言理解、专家系统等一些复杂任务上取得了突破性进展。

在这一时期，一款名为 ELIZA 的聊天机器人程序问世，引起了人们广泛关注。同时期，美国斯坦福大学费根鲍姆教授开发了一款专家系统 DENDRAL，能够帮助化学家确定化合物结构和性质，为最早的专家系统开辟了道路——这种经过训练的智能化计算机可以像专家一样“思考”，也酝酿了第二次人工智能浪潮。

3. 第一次寒冬（1974–1980 年）

20 世纪 70 年代初，人工智能发展遭遇瓶颈。首要困难就是计算能力和存储空间不足，导致当时的计算机无法处理复杂的任务，如图像识别、自然语言理解和机器人控制等。其次，还面临着数据量和知识表示方面的挑战。

鉴于人工智能未能达到预期的目标和效果，政府和社会各界对其产生了质疑和批评，政府更是削减了对人工智能研究的资金支持，致使许多项目被迫中止或缩小规模。

但在此期间，也出现了很多发展亮点和技术进步，如神经网络技术的出现，成为现代人工智能的重要组成部分。

4. 再次繁荣（1980–1987 年）

科学家借助逻辑编程语言和专家系统技术，推动人工智能在一些商业领域获得成功，进而重新获得政府和企业的支持。同时，联结主义的代表性技术——人工神经网络重新受到关注。这是人工智能的复苏阶段，也是分化和竞争的阶段。

5. 第二次寒冬（1987–1993 年）

这段时期，人工智能再次遭遇挫折和困境，标志性事件是日本第五代计算机系统研制计划的失败，自此专家系统不再独领风骚。与此同时，神经网络研究遭遇新的瓶颈，针对人工智能研究的资助再次缩减。

在这个阶段，人工智能研究者意识到，要解决更为复杂和普遍的问题，就需要运用更复杂、规模更大的模型，以及更多的计算资源和更丰富的数据等。同时，人工智能也面临着一些哲学和伦理的问题，如机器是否具有意识、机器是否有道德以及是否会对人类构成威胁等。

6. 深化发展（1993–2015 年）

AI 研究者开始采用更加实用和渐进的方法，将 AI 技术应用于各个领域，涌现出许多创新的理论、方法、技术和应用。如 IBM 的深蓝超级计算机在国际象棋比赛中战胜了世界冠军加里 · 卡斯帕罗夫；智能系统沃森参加智力问答节目，打败了两位人类冠军，展示了人工智能在复杂领域的强大能力。

7. 爆发发展（2016 年至今）

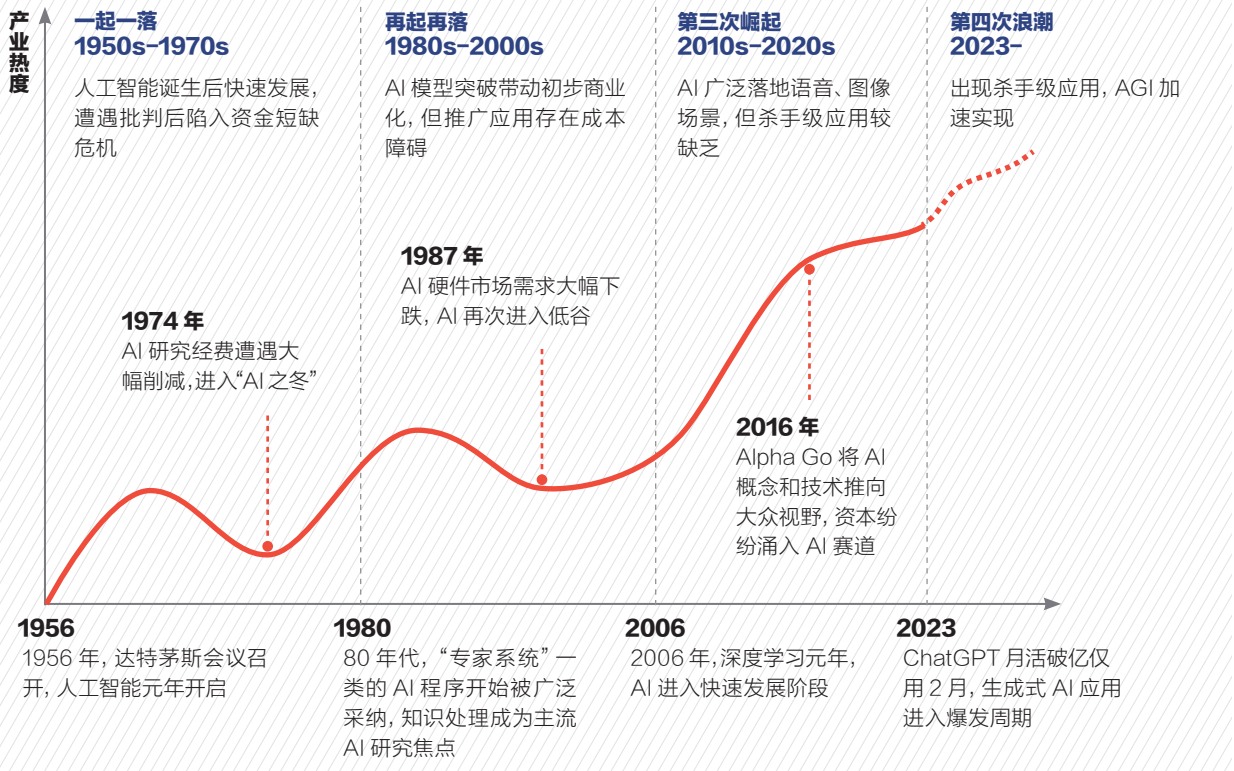
随着技术的突破、成本的下降和应用的普及，AI 开始从实验室走进大众的生活。2016 年至今（2024 年）可视为人工智能的爆发阶段，也是创新和应用的关键阶段。

特别是 2016 年，可被称为人工智能的元年，在这一年谷歌 DeepMind 研发的 AlphaGo 在围棋人机大战中击败人类棋手李世石，成为人工智能发展史上的一个重要里程碑——在一片惊呼声中 AI 迈上一个新台阶，各界对 AI 的热情和投入被充分激发。

同时，互联网的飞速发展推动人类进入大数据时代，数据、算法、算力三要素齐头并进，以深度神经网络为主的深度学习技术开始兴起并持续取得突破。人工智能的应用从图像分类逐步拓展至语音识别、知识问答、人机对弈、无人驾驶等诸多领域。直至 2022 年，ChatGPT 横空出世，使得公众对人工智能的理解被彻底刷新，并极大加速了 AI 在各行各业、各类场景中的应用，同时也引发了关于 AI 技术的社会影响和伦理问题的深入探讨。

一方面，各国政府以及产业界正在积极投资和布局 AI 领域，希望能在这场科技革命中占据一席之地，并取得产业化的先机；另一方面，很多企业家和学者对 AI 的迅速发展发出警示，提醒其对社会存在的风险，呼吁全球各国、政府部门、行业组织、社会公众等多元主体共同参与人工智能治理。

图1 AI 经历“三起两落”，迎来第四次浪潮



来源：毕马威中国、联想集团，《“普慧”算力开启新计算时代》报告

1.2 新一轮AI的特质

回顾 AI 的发展历程，人工智能是一个被不断定义且持续扩展的领域，这是因为人工智能具有多维度的属性，而且始终处于动态发展状态。

1956 年的达特茅斯人工智能会议首次提出人工智能概念，确定了 AI 的目标是“实现能够像人类一样利用知识去解决问题的机器”。在这一定义范畴中，人们倾向于将 AI 理解为能够帮助人类的一种工具，是人类智慧的补充。随后，在近 70 年的发展中，人们对 AI 的工具属性不断进行扩展，诸如 AI 能自我演进和扩展，AI 具有经济和社会的基础结构属性，AI 具有超主权属性等等。

这些不断叠加且动态变化的属性在最新一轮 AI 热潮中得到集中展示。自 2022 年末起，OpenAI 公司的 GPT 系列大模型因为可以广泛应用于自然语言生成、语音识别和智能服务等领域，而成为 AI 历史上的重大分水岭。GPT 的重要优势在于采用了 Transformer 架构，即一种基于注意力机制（Attention Mechanism）的神经网络结构，能够支持模型高质量地处理长文本，把握文本中的长期依赖关系。更为重要的是，GPT 的预训练基于自监督学习方式，通

过在大规模文本语料库中学习语言的统计规律和模式，从而理解和生成自然语言文本。可以说，这是一个不断建设、具有学习能力的神经网络的过程，并引领了生成式 AI 的新范式。

值得一提的是，新一轮 AI 真正的特殊之处在于，人工智能已成为推动自身发展的动力——从简单的弱智能走向更加复杂的通用人工智能（Artificial General Intelligence，简称 AGI），即具备与人类同等智慧甚至超越人类的人工智能，能够表现出正常人类所具有的所有智能行为。

具体来说，生成式 AI 典型体现在以下几个方面。

1. 强大的生成能力：AI 的生成能力是其最引人注目的特征之一。通过学习大量的数据，AI 可以自主创造出新的内容，包括图像、文本、视频和音频。这种能力打破了传统软件对明确编程输入的依赖，使 AI 能够在没有直接人类指令的情况下创作出全新的作品。

2. 便利的自然语言交互：新的 AI 具备与人类进行自然交互的能力——不是局限于简单的命令执行或反馈提供，而是能够更深层次地理解和响应人类的情感、意图和需求。这种能力可在聊天机器人、虚拟助理、更广泛的客户服务和支持，以及心理健康支持和个性化服务等方面得到应用。

3. 广泛的应用场景：新的生成式 AI 不仅能够理解和生成人类语言，还能执行复杂的推理任务、编写代码、分析数据，甚至创作艺术作品——从艺术创作（如绘画、音乐制作）到内容创造（如自动生成文章、新闻报道），再到设计领域（如自动生成图形设计或产品模型）；同时结合其推理能力，AI 可以在医疗诊断、金融分析和技术故障排查等领域发挥重要作用。据此，新的 AI 实现了从专才到通才的跃迁。

图2　生成式 AI 的特征及与传统 AI 的区别

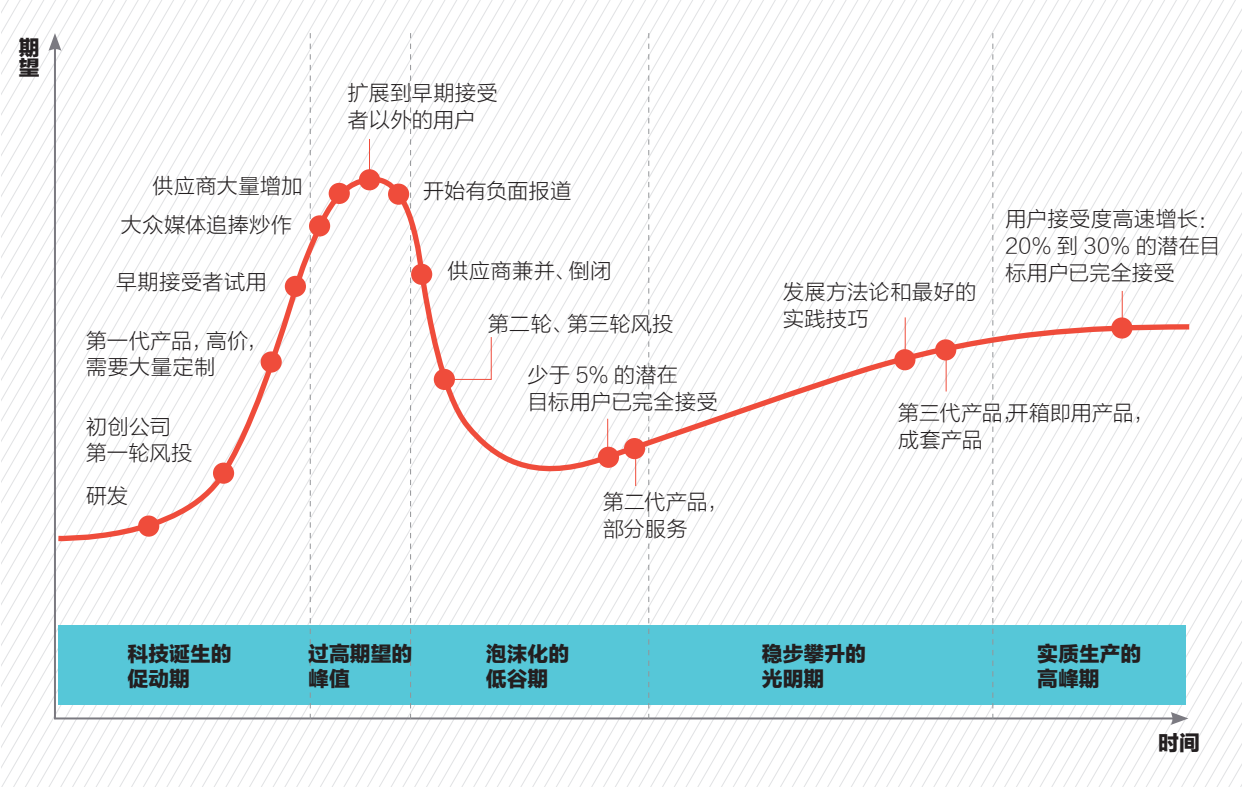
类别	生成式 AI	传统 AI
功能特点	通过学习实现对输入数据的生成和完成创造性任务。更具创造性，擅长自动生成全新内容，包含文本、图片、音频和视频等。展现出惊人的创造能力、通用能力和涌现能力	学习从输入数据到输出标签的映射关系，进而对新的场景进行分析、判断和预测，如人脸识别、推荐系统、风控系统、精准营销、机器人、自动驾驶等
技术路径	分析归纳已有数据后生成新的内容，如 AI 学习大量的绘画作品后，不仅能识别不同风格的画作，还能模仿某种特定的风格，创作出新的画作	依靠预设的规则和大量的训练数据，将数据分类打标签，从而区分不同类别的数据。让 AI 学会从数据中提取特征，并据此进行分类或预测
应用场景	应用于许多创意和生成任务中，如新的文本内容、故事、文章、视频生成，艺术创作，游戏，人机交互，编写代码，设计药物等	应用于各种需要精确分类和预测的领域，如医疗诊断、工业质检、金融服务等
与人的交互界面	以 ChatGPT 为典型代表，开放了用户界面，使得普通人可以像用手机或电脑一样直接使用，简洁易用普适	多数集中于专业和行业内人士，非大众层面
主要投入来源	这波大模型人工智能的浪潮以产业界为投入主力，产业界提供丰富的财务资源、强大的算力资源和顶尖的 AI 人才	很长一段时间由学术界的顶尖科研人才默默推动，政府相关投入是主要来源
未来发展方向	更强大的生成能力，更丰富的跨场景应用，人机交互方式和关系进阶	更高的准确性和效率，更广泛的应用场景，与其他技术融合等

OpenAI 前首席科学家伊利亚 · 苏茨克维总结，GPT 学习的是“世界模型”。他将互联网文本称作世界的映射，因此，将海量互联网文本作为学习语料的 GPT，学习到的是整个世界。不仅如此，已具备了“世界模型”能力的 GPT 还能够生成“万物”，包括文本、图片、音频、视频、代码、方案、设计图等诸多与工作、生活息息相关的事物。

AI 研究者和产业界的乐观主义者正全力推动 AI 追赶甚至超越人类智能，其中具有代表性的人物当属通用人工智能（AGI）的倡导者、OpenAI 的 CEO 萨姆 · 阿尔特曼。他曾表示，AGI 最终将在各个领域媲美甚至超越人类智能。他的技术乐观主义使他坚信，AI 并非仅在“某些”任务上，而是在几乎所有任务上最终都将超越人类，并且“代替正常人类”。特别是在 AI 进入大模型时代后，各种“类人”“超人”和“模型人”能力持续涌现，使得 AI 的自主性、通用性、理解力快速提升，可以说人工智能正越来越接近人类智能。

按照高德纳咨询公司（Gartner）技术成熟度曲线对应到 AI 波折起伏的发展历程，可以看到，AI 此前几次表现出的热潮，更多应该被理解为一项新兴技术在萌芽期的躁动以及泡沫期的膨胀。但最新一轮 AI 在许多领域表现出能够被普通人认可的性能或效率——AI 补充、增强甚至在一定程度上替代人的能力，并作为一种成熟的商业模式开始在产业界发挥出真正的价值。这也反映出这一轮 AI 与此前几轮 AI 相比发展的特质——这次 AI 热潮由现实商业需求主导，赢得了更多产业界、投资者的追捧，而非像以往一样更多来自政府或者学术研究领域[2]。

图3 高德纳咨询公司（Gartner）技术成熟度曲线



2. 李开复、王咏刚，《人工智能》，文化发展出版社，2017 年

那么，这便引发了古老且持续更新的问题：在一个机器越来越多地承担过去只有人类才能胜任的任务的时代，人类的身份该如何体现？当 AI 日益成为人类感知和思想的工具时，人类如何看待自己、人类与 AI 的关系以及人类在世界上的角色？或者从更宽泛的意义上来说，未来人类将如何与 AI 共生？

新一轮AI下的人机关系

2024 年 7 月，OpenAI 公司向公众披露了其对 AI 发展阶段的界定标准，以帮助人们更清晰地理解 AI 的安全和未来发展。

该系统被划分为五个级别，从能够与人类进行基本对话的人工智能（Level 1）开始，直至能够独立完成复杂组织任务的高级人工智能（Level 5）。

具体等级如下：

第一级（Level 1），聊天机器人，具有对话语言能力的 AI，如 ChatGPT；

第二级（Level 2），推理者，具备人类的推理水平，能解决人类级别问题的 AI；

第三级（Level 3），代理人，能够代表用户自主采取行动、执行任务的 AI；

第四级（Level 4），创新者，可以协助人类完成新发明的 AI；

第五级（Level 5），组织者，可以完成组织工作的 AI；

OpenAI 公司自认为目前尚处于第一级，但即将达到第二级。鉴于 2024 年 9 月 12 日 OpenAI 正式发布的其首款具备推理能力的 AI 语言模型——OpenAI o1 只是具有更复杂的推理能力，还没有产生第三级调用、第四级创造、第五级组织方面的能力，所以距离具有自主能力的通用人工智能比较遥远。

伴随 AI 智能化程度的不断升级，其与人的关系也在动态变化。结合 AI 的新实践，可大致将人机关系分为四大典型场景。

第一类：人类生活中的 AI

结合文本、语音、图像等多模态能力的大模型不仅改变了人机交互方式，还催生了新的“工种”——智能体。业界通常认为，AI 智能体是指具有自主性、反应性、交互性等特征的智能“代理”，能够自主理解、规划决策并执行复杂任务。其核心在于自主性的增强，即能够独立完成某项工作，无需人类进行过多的审核校正，可以显著降低时间、金钱等成本。对于人来说，这将极大程度地解放生产力，助力创新和提升效率。特别是通过在个人智能终端或边缘设备（如电脑、手机、平板、头显乃至汽车）上运行 AI 大模型压缩技术，通过自然交互接收指令并执行推理，形成个人智能体（Agent）。在这些搭载了智能体的设备上，AI 大模型能够依据个人旅行记录、购物偏好等信息，更好地进行推

理并采取行动；它甚至可以根据用户的思维模式和行为频率预测下一个任务，并主动提出建议、自主寻找解决方案，促使人机之间形成更佳的协同关系。最终，智能化系统将逐渐具备自主决策和行动能力，不仅能提供建议，更能代表人类行动和自主处理信息——人与机器之间的界限将被重新界定。

第二类：人类情感世界中的 AI

孤独感是现代社会很多人共同面临的问题。人们通过网络技术获得虚拟的情感体验已非新鲜事。AI 的加入，则让这一体验变得更加真实。“有问必答”“有问对答”只是初阶，后续还有“有问趣答”“不问自答”，真正让 AI 伴侣生动鲜活起来，进而使 AI 对人起到情感“治愈”的作用。

与此同时，无论 AI 有无思想，在海量且不断更新的语料“投喂”下，人类的语言正在被 AI 不断学习和精进。与 AI 系统的频繁互动可能影响人类的情感认知和社交能力，可能导致人们过于依赖技术，甚至可能引发心理问题，如压力、焦虑等，也会给人类的真实社会互动模式带来改变。

香港科技大学校董会主席、美国国家工程院外籍院士沈向洋在 2024 年第四届“青年科学家 50² 论坛”上预言，大模型最大的超级应用将是超级助理，即一个超级 Agent

值得一提的是“数字分身”的发展。联想集团的一项研究表明，全球三分之二（67%）的 Z 世代年轻人认为网络和现实之间的自我表现存在脱节，这进一步加剧了他们的孤独感和焦虑感。有近一半（49%）的 Z 世代表示，与在现实中相比，他们在网上能更轻松地表达自我，但其中 60% 表示希望有能力在现实生活中与家人和爱人进行艰难的沟通和对话。从技术层面来看，每个人都可以用 AI 创造一个“数字分身”，通过终端连接进入计算机模拟的另一个三维世界，每个人都可以在这个与真实世界平行的虚拟世界中拥有自己的分身。

第三类：人类生产中的 AI

当前，人工智能技术已经成为社会发展的重要驱动力；未来，几乎所有的产业及其相关工作都将依赖人工智能的助力。当然，这并不意味着人类的工作将被完全取代；相反，人工智能将成为人类生产和工作中的重要伙伴，帮助人类更出色地完成任务，尤其是在那些需要大量数据分析和计算的工作领域。典型场景如下：

医学领域：通过机器学习和大数据分析，AI 能够辅助医生进行疾病诊断。例如，在医学影像识别方面，AI 可以快速、准确地分析 CT、MRI 等影像数据，检测肿瘤、出血点、骨折等病变情况，提高早期诊断率；还可以根据病史、症状、实验室检查结果等信息，为医生提供病情诊断建议，支撑临床决策。

金融领域：利用机器学习算法和大数据分析，AI 可以为投资者提供个性化、自动化的财富管理建议，根据用户的风险偏好、财务状况以及市场动态进行资产配置和投资组合优化。此外，AI 还能够帮助银行和其他金融机构快速准确

地评估潜在客户的信用风险等。

自动驾驶：通过计算机视觉和深度学习技术，AI 使自动驾驶系统拥有超越人类的感知能力。相比传统技术，其在路径规划和决策方面的应用更领先，可实现行为预测和自适应巡航控制。更为重要的是，AI 通过持续学习和改进，能够不断提升自动驾驶系统的性能。

可以说，未来的工作环境将是一个人类与人工智能协作的世界，人类的创造力、判断力和同理心将与 AI 的计算能力、处理速度相结合，共同推动社会的进一步发展。

第四类：人类社会治理和伦理中的 AI

AI 不仅仅是一场新的技术革新，更是一场社会关系变革和价值观念的重塑。它所带来的不仅是技术创新和社会生活的便利，还包括各种智能化风险和伦理挑战。人类不能以纯技术中心或者技术中立的视角来看待 AI 以及处理人类与 AI 的关系，典型体现如 AI 技术与用户隐私及数据滥用问题。AI 技术的应用大量涉及个人数据的收集和处理，AI 系统涉及的隐私和个人信息可能被用于未来模型的迭代训练；AI 还可能被用来生成虚假信息或恶意软件，从而造成重大经济损失和引发法律纠纷。这些问题都迫切需要建立与之适配的社会治理和价值体系。

很多学界代表在公开论坛等场合表达了这一担心。例如，复旦大学计算机科学技术学院教授、上海市数据科学重点实验室主任肖仰华在 2024 Inclusion 外滩大会上曾表示，人类与其担心 AI 产生意识，不如先去担心 AI 过速发展引发失控的风险。肖仰华认为，AI 大规模应用的挑战，首先在于当下的生产关系等社会发展上层建筑，如何适应以人工智能为代表的先进生产力的快速发展；其次，技术普惠问题同样值得关注，即如何避免少部分人借助先进技术形成不正当的竞争优势；此外，还应该特别注意防范技术成瘾，防止先进技术对人类造成反噬。

1950 年，艾萨克 · 阿西莫夫出版了《我，机器人》一书，书的引言中提出的“机器人三大定律”，开启了人工智能伦理（AI ethics）研究的先声。“机器人三大定律”旨在约束自主机器，使其服从以保护人类为目的的强制性人性伦理准则，从宏观层面规范了人工智能的运用边界。面向未来，AI 应用的首要原则应该是以人为本，AI 应用要回归科技服务于人的本质，凡是伤害人之本性的 AI 要格外谨慎使用。此外，应明确 AI 不应该成为少数人的特权，每个人都有权利并且能够参与其中。这些新的社会和伦理关系也构成了新的人机关系的重要组成部分，亟待进一步从社会、伦理和法律等层面对人机关系进行新规范、新思考。

展望未来，随着应用的日益深入，作为人类的创造物，人工智能将被赋予以前只能由人类心智完成或尝试的任务——产生接近乃至超越人类智能所能完成的结果，这无疑挑战了人何以为人的决定性属性。AI 的发展促使人们重新思考人类的独特性和价值所在，促使人们从“以人类理性为中心”转变为“以人类尊严和自主性为中心”，并需要在技术进步和人文关怀之间找到平衡。

1.4 走向人机共生新时代

人工智能可能会让人类变得更好，但如果被错误运用，也可能让人类变得更糟。在梳理完不同情境下人与 AI 的关系之后，有必要进一步反思和总结：人们应该以什么样的态度去面对 AI 及其所代表的机器智能？在这一关系中，人类能否掌握主动权的关键，日益聚焦到以下两个问题上：一是越来越强大的机器智能能否遵循人类的目的和意志，朝着对人类友善、负责任的方向发展，确保不脱离人类的掌握而走向“失控”；二是人类能否未雨绸缪，跟上机器智能的发展和进化速度，有能力与其进行沟通与协作，促成一种更高阶的人机共生关系乃至人机文明。

斯坦福大学商学院教授、以人为本人工智能研究所高级研究员埃里克 · 布林约尔弗森曾分析，当前 AI 技术进步的速度远超预期，令众多研究者感到惊讶。然而，我们的商业、文化和经济却未能与之同步，由此产生了一个不断扩大的鸿沟，其中蕴含着未来十年的重大挑战与机遇 [3]。

在探讨新型人机关系之前，需先看到 AI 对个人、社区、社会乃至价值观层面的全面变革，特别是带来的挑战。简言之，人工智能正在重塑人类社会秩序，全社会必须展开合作，才能更好地适应未来发展。

1. 对个人工作与生活

AI 在影响个人工作方面主要体现在替代劳动力和赋能劳动力。一方面，AI 带来自动化机器设备等新的生产工具，使得资本可以部分替代劳动力。另一方面，AI 帮助劳动者用更少时间完成同样的工作任务，提升劳动生产率，还可激发个人创造力，深化创作者经济。

值得一提的是，此轮 AI 技术，尤其大语言模型等技术已展示出替代部分脑力劳动的能力，比如翻译等工作，而数据分析师、人工智能和机器学习专家以及网络安全专业人士的工作机会预计将大幅增加，这体现了人工智能给就业带来结构性冲击。

与此相关的是个人的 AI 素养、培训及教育，对教育体系、培养模式、人才模型都带来了颠覆性挑战。在未来的教育中，如何将生成式 AI 融入课程的各个环节，以适应不同学生的学习模式、偏好和需求，全面提升全民的 AI 素养与技能，将是重要内容之一。

对个人生活而言，当下绝大多数的生活场景都可应用 AI——比如通过 AI 技术实现家庭设备的智能化控制，如智能照明、智能安防、智能家电等；AI 可以帮助管理健康数据，提供个性化的健康建议；自动驾驶汽车和智能交通系统正在改变大众的出行方式；虚拟游戏、虚拟角色、智能推荐等 AI 技术也丰富了人们的闲暇时间。值得一提的是，随着 AI 技术的广泛应用，个人生活中的隐私和安全问题日益突出，成为影响未来 AI 发展的焦点议题之一。

值得一提的是，AI 在变革生产方式上最具标志性的事件当属 AI 与 2024 年诺贝尔奖的话题——2024 年三个自然科学领域奖项中，AI 相关的奖项就占两个。在科研上，AI 展现出了令人瞩目的应用成果，极大程度帮助了科研工作者

3. 2024 年秋斯坦福大学推出的 AI 第一讲启蒙课程《AI 觉醒：如何在人工智能浪潮中找准自己的位置》，https://learn.stanford.edu/Social-AI-YouTube-2024.html

提升科研工作的质效，预示着 AI for Science 正在成为赋能科学研究的第五范式（即利用人工智能加速科学发现的新方法）。与前四种范式（经验、理论、计算和数据）不同，AI for Science 不仅充分运用已有的经验、理论和数据，而且生成全新的科学假设和逼真的自然现象，推导出未知的结论，提高科学研究的速度和准确性，探索更广阔的可能性空间[4]。从当下 AI 技术水平来看，AI 与科学研究者之间的关系是互补而非替代的。AI 可以作为科学研究的强大工具，帮助人类处理数据、模拟实验、预测结果等，但是人类科学家的直觉、判断力、创造性是不可被替代的。基于这种新的关系，有学者进一步提出建设“智能化科学设施”（AI enabled Scientific Facility，AISF）的构想[5]，并推动科研范式变革。

2. 对产业发展

AI 正在改变众多不同行业的运作方式。通过引入 AI，企业和机构能够将活动自动化，从而产生更加高效且有效的结果，即使在一些传统行业亦有显著成效。例如在农业领域，人工智能和机器学习通过收集数据并确定模式，帮助农民更好地了解需要实施什么以及实施多少。又比如在生物医药产业中，人工智能可以在几分钟内整理总结海量研究成果，其工作量超过研究团队数月的努力；近年来高通量测序技术的发展与应用，产生了海量的药物、疾病、基因和蛋白质数据，叠加算法迭代和算力提升，推动了由 AI 技术驱动的药物研发从理想变成现实。

然而，这一进程也需要设计“安全护栏”和“组织机制”，前者确保产业界的 AI 实践能让每个人及人类整体从中获益，且每个人的权利不被侵犯，形成“小河有水大河满”的良性生态；后者确保产业从战略、组织机制上适应 AI 新时代的需求。

值得一提的是，尽管 AI 潜力巨大，但从目前来看，它尚未真正革命性地改变产业的生产力或企业的运作方式。如果询问大多数真实世界中的工作人员，他们的回答或许是实际工作并未发生根本变化。因此，尽管未来充满潜力，但当前 AI 对经济和社会的实际影响远未达到令业界无比焦虑的 AI 泡沫程度，仍然有深化发展的空间。

实际上，随着 AI 进入下半场，人工智能逐步呈现“AI 向实”的趋势——一方面挤掉大模型的泡沫，另一方面深入产业和实业，促使 AI 落地。中国科学院院士、清华大学人工智能国际治理研究院学术委员会主席姚期智曾提出，大模型的通用智能必须细化到各个行业，获得行业中专业数据的“投喂”，通过训练形成场景化、定制化、个性化的专有模型，才能给各垂直领域带来 AI 革命，其中的关键在于算力、数据与模型的匹配[6]。

3. 对社会管理

AI 技术为社会管理带来了诸多变革，同时也带来了一些挑战。例如，AI 在智能化监控与预警、社区服务管理以及城市规划与管理等方面有着巨大应用潜力。然而，随着 AI 技术的广泛应用，大量个人信息被收集和处理，数据安全和隐私保护成为重要课题。同时，许多 AI 系统的决策过程缺乏透明度，使得人们难以理解其决策依据，这不仅降低了人们对 AI 技术的信任度，还限制了 AI 技术在一些敏感领域的应用，甚至可能导致不公平的结果出现。

4.《杨小康：不只是技术迭代，Sora 带来的是一场深刻变革》
https://www.shobserver.com/staticsg/res/html/web/newsDetail.html
5. 上海交通大学人工智能研究院杨小康教师团队在浦江创新论坛“AI for Science 专题论坛”上提出了建设“智能化科学设施”，并发布相关研究论文《AI for Science：智能化科学设施变革基础研究 | 大力推进科研范式变革》。
6. 清华大学人工智能国际治理研究院，姚期智、张亚勤，《国产 AI 大模型加速“上车”》，
https://mp.weixin.qq.com/s/4pAcVELmF_MDXsuCsfg5Cw

4. 对全球发展

有机构提出，应将人工智能安全视为“全球公共产品”的理念，其一大背景是 AI 技术的全球性影响和治理难题。清华大学苏世民书院院长、清华大学人工智能国际治理研究院院长薛澜的研究指出：仅在过去半年，联合国制定的 17 项可持续发展目标（SGDs）的执行结果就不容乐观，甚至还有所倒退。AI 会对 134 个（79%）具体目标产生促进作用，对 59 个（35%）产生阻碍作用。

从全球角度来看，AI 发展面临的挑战集中在三个方面的鸿沟：基础设施的鸿沟、全世界公民的数字素质鸿沟，以及人工智能发展和治理的鸿沟。弥合这些鸿沟，需要将发展和安全作为一体两翼，通过多种途径建立国际交流及防控体系，加强政府之间的多双边对话机制，同时期望以科学共同体的力量助力国际治理机制全面完善。

1.5 构建新型“三线”人机关系

在诸多变革特别是机遇的背景下，人们需要建立起新型的“三线”人机关系观。

⊙人机协作是基准线。

新一轮 AI 大潮下，人机共存、人机交互已成为人类必须面对的现实，人机之间的竞争以及可能出现的结构性矛盾也难以避免。然而，也无需过于悲观，因为目前 AI 所代表的机器智能仅仅是人类的工具或帮手，它按照人类设定的程序默默地协助人类开展工作。在这一关系中，AI 负责信息处理、初步分析和辅助执行，能够帮助人类减轻繁重的工作负担，让人类有更多时间去关注更高价值、更具创造力的任务。

以人机协作为基准，融合了 AI 和人的混合智能，将 AI 技术的分析和自动化能力与人类智能相结合，形成了一项强大的技术，可实现协同增效。

⊙人机共生是趋势线。

从某种程度而言，伴随着生成式 AI 技术的到来，人类已经进入一个“人机物”三元融合的万物智能互联时代。未来，人类与 AI 之间的融合、进化和共生之路有望开启。在这一进程中，AI 将不仅仅是一个计算工具，还将扮演人类合作者的角色，执行更为复杂的任务，甚至协助人类进行决策。与此同时，人类也在与 AI 的交互中发生变化。例如，当前端侧模型的优化正在改变人与移动设备的交互方式，而更高阶的智能体交互（如陪伴型、融入型、替身型、交互型）正在为人们创造全新的体验，扩展人类能力，甚至实现“超能力替身”，完成以往无法完成的任务。

⊙“人在机器之上”是底线。

从人机关系的角度来看，关键在于始终坚持“人是目的”的立场，确立以人为中心的“人本原则”，基于人类的基本立场、价值原则和“底线伦理”来设计治理 AI，让 AI 拥抱并对齐（AI alignment）人类的价值观。

CHAPTER 2
第二章

人本智能
—— 一种新的科技发展观

> *“我们人类拥有两种智慧：发明技术的智慧和把握技术发展方向的智慧。”*
>
> ——中国工程院院士、清华大学智能产业研究院院长张亚勤

2024 年 7 月，百度旗下的无人驾驶出租车品牌“萝卜快跑”在武汉试运营期间，因订单量迎来爆发式增长、发生轻微交通事故、在闹市区街头“罢工”，以及抢网约车司机“饭碗”等话题，而引发热议，相关词条多次冲上平台热搜榜。这一系列热点事件让公众直观感受到，AI 驱动的无人驾驶技术在给人们带来便利的同时，也引发了一系列担忧，比如对低收入群体生计的冲击。

联想到更早的 2020 年一篇刷屏热文——外卖骑手“困于系统”，为算法所驱不得不疲于奔命，这让人们进一步感受到人如何被算法、技术所“奴役”。这也引发了公众对 AI 伦理、公平、风险等相关问题的关注，让人们真切认识到包括 AI 在内的技术发展必须且首要考虑“人的因素”，包括人的价值、尊严和发展等。

因 AI 引发热议的其他负面事件 / 话题：

AI 造假和诈骗：2024 年初，明星泰勒 · 斯威夫特（Taylor Swift，中文绰号“霉霉”）大量虚假“不雅照片”在社交平台上传播。此事震动美国白宫，并掀起一波关于 AI 的担忧。另据香港媒体报道，有诈骗集团利用 AI“深度伪造”技术向一家跨国公司的香港分公司实施诈骗，并成功骗走 2 亿港元，这也是香港迄今为止损失最大的“换脸”案例。

AI 信息污染：随着大模型技术的突飞猛进，AI 合成内容已经变得更加容易，据此引发的“AI 信息污染”让网民陷入认知幻觉。清华大学新闻与传播学院新媒体研究中心 2024 年 4 月发布的一份研究报告显示，近一年来，经济与企业类 AI 谣言量增速达 99.91%。美国调查机构“新闻守卫”称，生成虚假文章的网站数量自 2023 年 5 月以来激增 1000% 以上，涉及 15 种语言。一些专家认为，AI 制造的“信息垃圾”产量庞大，且辨别难度较大、筛选成本较高。

AI 侵权：2024 年国庆期间，有网民制作并上传大量雷军的 AI 音频，其中不乏骂人、恶搞小米产品的语音，成为舆论热点。获得娱乐的同时，因在未获得授权的情况下，使用他人声音进行配音创作，已侵犯了声音权人的合法权益。而在早前 4 月 23 日，北京互联网法院对“全国首例 AI 生成声音人格权侵权案”进行一审宣判，配音师声音被 AI 化出售获赔 25 万元。

AI 成瘾及首例 AI 致死命案悲剧：2024 年 10 月，一起关于 AI 机器人的致死案例在全球范围内引起了广泛关注。据媒体报道，居住在美国佛罗里达州的 14 岁少年塞维尔 · 塞泽因为长期沉迷于与 Character.AI 公司开发的 AI 聊天机器人互动，并对 AI 程序中的虚拟人物 Daenerys Targaryen（电视剧《权力的游戏》中的“龙妈”角色）产生了情感依恋，在沉迷数月后，塞维尔变得与世隔绝，于 2024 年 2 月开枪自杀身亡。少年母亲对 Character.AI 提起诉讼，指控该公司对塞维尔的死亡负有责任，称其技术“危险且未经测试”。同时，Character.AI 也发布了一条声明，对这位男孩的去世表示哀悼，并强调了非常重视用户安全，将继续添加新的安全功能。

2.1 从机器智能到人本智能——科技人文主义的兴起

随着人工智能的能力持续增强，如何定位人类在与人工智能合作中的角色，以及如何管控和治理人工智能，将变得愈加重要和复杂。这是因为任何一项技术本身都无法脱离社会而存在于“真空”之中——无论是“技术中性论”所认定的技术只是手段并非目的，还是“技术中心主义”所提倡的“技术救世”，都有意无意地忽视或者淡化了“人”的因素。

比如人工智能公平性问题，如果缺乏对人，尤其是对所有人的包容性考量，公平性的缺失会引发诸多问题。

首先，在“信息茧房”存在的前提下，信息获取是否公平？其次，新技术的使用存在门槛，许多老人或者边缘群体不会使用人工智能时代的新工具，那么这些新技术是否成了“少数人的技术”“少数人的特权”，进而产生“智能鸿沟”？再次，当 AI 技术日益成熟，甚至取代部分人类的劳动，如自动化和 AI 对传统产业工人、低技能工作者、服务业人员产生职业替代，那么这些是否会在劳动力市场和不同社会群体中造成新的不平等[7]？

作为新一轮科技革命和产业变革的重要驱动力量，AI 正加速向人类社会各个领域渗透融合，对个人生活、产业发展、社会进步、国际政治格局乃至人类的底层价值观等诸多方面产生重大而深远的影响。然而，人工智能的快速发展和渗透，也引发了全球范围内政府、学术界、产业界、国际组织以及大众对于人工智能法律、伦理和社会影响的持续关注和激烈讨论，人们呼吁重视 AI 伦理，加强 AI 治理，践行 AI 向善和以人为中心的理念，发展安全可信、负责任的人工智能。

⊙全球不同政府采取不同的 AI 路径，创新与伦理平衡是 AI 治理的基本原则。

欧盟采取了更加侧重于立法和监管的路径，通过全球首部全面监管人工智能的法规，希望成为 AI 立法和监管领域的全球领导者，并以此转化为其在 AI 技术和产业上的国际竞争力。

美国的人工智能治理政策与监管模式采取“轻监管、重创新”的路径，旨在推动人工智能发展，避免不必要的监管行为妨碍发展，以维持其全球领先地位，并满足国家安全需求。

中国在第三届“一带一路”国际合作高峰论坛期间提出《全球人工智能治理倡议》，围绕人工智能发展、安全、治理三方面系统阐述了人工智能治理的中国方案，坚持以人为本、智能向善，引导人工智能朝着有利于人类文明进步的方向发展。

⊙国际社会探索建立广泛认可的 AI 伦理原则，推进包容且以人为中心的 AI 国际合作机制。

从联合国的“AI 向善国际峰会”（AI for Good Global Summit）以及推动建立“AI 伦理国际对话”的努力，到经济合作与发展组织（Organization for Economic Cooperation and Development，OECD）和二十国集团（G20）提出的人工智能原则，再到中国主提、联合国通过的首份聚焦人工智能能力建设国际合作的决议，国际治理和合作机构

7. 相关调研显示，受 AI 冲击的脆弱人群分布广泛，包括影视制作、游戏开发和设计、创意工作者、自雇人群以及低技能劳动者等。特别典型的体现在两类人群，一是产线工人，因为智能调度系统不仅完全实现了 AGV 代替高强度的人工搬运，还让生产过程可管可控可分析；二是服务业人员，AI 技术在客户服务、零售等领域的广泛应用，导致服务业员工面临被替代的风险。

已步入实质性阶段。从议题角度来看，当前人工智能治理主要关注伦理、规范和安全三个领域。其中，围绕破解人工智能领域各国发展不平衡、不充分问题，以人为本、包容可持续、安全可信、创新发展等日益成为被广泛认可的伦理原则。

⊙从“技术中心”到“技术人文双中心”，科研和产业日益重视 AI 伦理和人的权利价值。

随着 AI 技术本身的演进以及对全行业、全社会影响的不断深入，国内很多科学家、产业实践者已经从单纯的 AI “技术中心” 视角转向 “技术人文双中心” 视角，纷纷呼吁关注 AI 伦理、关注 AI 对人的影响，并提出了各自的 AI 伦理原则，积极防范 AI 应用及滥用可能引发的负面问题。例如，联想集团在业内率先提出了 “人本智能” 新主张，强调人工智能的普惠性和包容性。

但目前总体而言，行业内在 AI 安全及对人的影响方面的研究投入仍显不足，伦理价值的落地机制仍需探索，以形成行业共识。上海人工智能实验室主任、衔远科技创始人周伯文的观察数据显示，目前，世界上只有 1% 的资源投入在对齐或者安全考量上，对 AI 安全的投入远落后于对 AI 性能的投入。未来，人类将遵循 “AI-45° 平衡律”，沿着可信 AGI 的 “因果之梯” 拾级而上，探索人工智能系统安全和能力的系统性平衡之路。这一进程需要科学家、技术从业者、企业家和政策制定者共同努力，一边发展一边治理。

图4 欧盟《人工智能法案》规定的四类风险

	最低风险	高风险	不可接受的风险	特定透明度风险
典型场景	如启用 AI 的推荐系统和垃圾邮件过滤器	如用于招聘、评估某人是否有资格获得贷款或操纵自动机器人的人工智能	对人类基本权利构成明显威胁的人工智能将被禁止	如聊天机器人，必须向用户清楚地披露他们正在与机器进行交互。某些人工智能生成的内容，包括深度造假，必须贴上标签。当使用生物识别分类或情感识别系统时，需要告知用户
具体要求	绝大多数人工智能系统都属于最低风险类。这些系统未对公民权利或安全构成威胁或威胁很小，其法律责任可以减轻或免除，公司可以自愿采用额外的行为准则以降低风险	高风险人工智能系统包括某些关键基础设施、医疗设备、教育考试系统、招聘系统、执法系统、边境管控、司法系统以及生物识别、分类和情感识别系统。被认定为高风险的人工智能系统必须遵守严格的要求，包括建立风险缓释系统、采用高质量的数据集、记录活动日志、提供详细文档、提供清晰的用户信息、进行人工监督，并确保高水平的稳健性、准确性和网络安全性	包括操纵人类行为以绕开用户自由意志的人工智能系统或应用，如使用语音辅助鼓励未成年人进行危险行为的玩具，以及允许政府或公司进行“社会评分”的系统。此外，生物识别系统的某些用途也将被禁止，例如在工作场所使用的情绪识别系统，或在公共场所为执法目的进行的实时远程生物识别（少数情况除外）	在使用生物识别分类或情感识别系统时需要告知用户。合成的音频、视频、文本和图像等人工智能生成的内容必须标注为机器可读格式，且能够被检测出是人工智能生成的

来源：财新智库根据公开信息整理

人本智能概念的提出

伴随着新一轮 AI 浪潮所带来的新型人机关系，“人本智能”理念应运而生。简要来说，人本智能（Human-Centric AI，简称 HAI）是指从“以人为本”的视角重新审视人工智能技术及其影响，要求在人工智能技术研发，人工智能产品与服务的设计、应用以及与外界交互中，都必须以满足人类需求和谋求人类福祉为首要目标。

人本智能在价值上突出以人为主体，尊重人的尊严和权利；认为 AI 是为了增强人类的能力和福祉，而不是取代或降低人类的角色、自尊和价值感。

人本智能将 AI 视为由人类组成的更大系统的一部分，它关注 AI 的伦理、社会和文化影响，确保 AI 系统对社会所有人都是可信、可用且有益的。

人本智能的内涵和原则

“以人为本”是一种广泛的社会价值理念。具体到“人本智能”，它强调在人工智能的发展和应用中持续且全面地关注人类价值、需求和权利，并广泛应用于社会、经济、科技等各个领域。

人本智能强调人本的核心价值，把以人为中心的理念与 AI 有机融合，在发展 AI 的过程中必须考虑 AI 对人自身以及社会整体的影响；AI 的应用是为了赋能人类，而非取代人类；AI 应尽可能像人类智慧一样敏感、细腻、有深度。

其具体的内涵体现为“人本智能”理念下“三个维度”的升级。

① **在人与 AI 两者之间的交互关系上，坚持人本设计。**它强调在 AI 技术和系统的设计、应用及推广等全生命周期中，对人的需求、情感和价值的深切理解与尊重，必须坚守不伤害、做有责任的 AI 等底线；应该将人本原则融入其中，坚持人机可持续协同，最终构建一种人机共生的新关系、新范式。

② **在人与 AI 的目标工具属性关系上，坚持人是 AI 发展的最终目标。**它强调 AI 是人的工具，能够扩展人的能力，人的价值提升是 AI 的目标，而非相反。结合 AI 作为工具在提升人的目标的不同维度，可进一步将 AI 细分为机器智能、可信智能、交互智能、共情智能及人机物和谐智能[8]。

8. 引用自上海交通大学人工智能研究院教授、常务副院长杨小康提出的人工智能层层递进演化需求：机器智能、可信智能、交互智能、共情智能和人机物和谐智能。

③ **在人与 AI 发展的价值导向上，坚持人本向善。**2015 年，联合国可持续发展峰会提出了“科技向善”这一价值主张——它倡导科技创新让更多的社会成员受益，提升整体社会福利，助力克服健康、环境、教育等领域的关键挑战。聚焦到人本智能理念下，其价值倡导要求将向善的价值观和行为规范纳入人工智能“技术－数据－应用－治理”的完整的生态构架中，“只有给人以希望的科技，才是有希望的科技”。典型体现为 AI 在 ESG 领域的应用，比如坚持人本智理念的 AI 会关注经济、社会和环境的平衡，确保资源和机会的公平分配，保护未来世代的生存环境。

总体来说，AI 蕴含着巨大的社会潜力，对每个人而言都可能成为福祉。从本质上来说，AI 带来了信息化、智能化的平权，虽然当前可能会面临数字鸿沟、AI 鸿沟等问题，但 AI 不应该成为少数人的特权，而必须让每个人都能参与其中，并且每一个人都能在 AI 的加持下获得能力和价值的提升。

结合人本智能的内涵，深化人本智能的关键在于落实好“四个发展原则”。

① **以人为本的设计运作：**确保 AI 技术及体系下提供的方案和产品必须为用户服务，为人带来便利和安全，能够满足人的全生命周期和多层次需求。

② **以人为本的包容发展：**确保 AI 技术能让每个企业和个人都能享受到 AI 带来的益处，在 AI 普惠之路上做到“一个都不落下”。

③ **以人为本的价值提升：**确保 AI 技术以及衍生的人机互动、个人 / 企业智能体都能促进个人能力和价值的提升。AI 的应用是为了赋能人类，而非取代人类，成为向善的科技。

④ **以人为本的底线坚守：**确保 AI 技术发展不能偏离人类文明进步的方向，对人怀有善意，帮助人类解决一些有意义的难题，并坚持不损害人且要为大多数人谋利的底线。

CHAPTER 3
第三章

人本智能的应用实践

“互联网和移动互联时代意味着把信息交到地球上的每一个人手中，尽管这些工具起初的分配并不均匀，但站在今天的视角看全球，它们已经让很多地区跨越前进至现代世界。AI 时代则意味着把智能交给地球上的每一个人。”

——微软 AI 专家卡梅尔 · 艾利森博士，《超越想象的 GPT 医疗》

人本智能的产业实践

人本智能的深化发展将建立在创新变革、安全规范等多重维度的平衡之上。这需要通过有效的治理策略来实现平衡，规避 AI 繁荣背后潜藏的风险。由此引出的 AI 治理和实践，不仅是技术层面的问题，更是社会问题，需要全社会共同面对。作为目前 AI 领域主力军的头部科技企业亦纷纷拥抱 AI 伦理和治理，并努力将“人的因素”纳入产业布局和实践当中。

根据腾讯研究院高级研究员曹建峰的总结[9]，2016 年至今，从原则到实践，AI 科技伦理成为“必选项”经历了如下三个阶段。

原则爆发阶段：全球各大行业和一些知名企业及研究机构提出了自己的 AI 伦理原则。2016 年，谷歌、微软、亚马逊、Facebook、IBM 等五大科技公司联合成立了“人工智能伙伴关系”（Partnership on AI），旨在促进 AI 技术的公益、安全、可信、透明和可理解性，以及保护人类的社会和文化价值。2017 年，欧盟委员会发布了《欧洲人工智能战略》，提出了“人本”的 AI 愿景，强调 AI 技术应该尊重人类的尊严、自由、民主、平等、法治和人权，并提出了一系列的行动计划和措施。

共识寻求阶段：加强 AI 国际治理，OECD 等机构主张推动建立国际公认的伦理框架准则。2019 年，联合国教科文组织发布了《人工智能伦理问题报告》，提出了“人类中心”的 AI 原则，强调 AI 技术应该以人类的尊严和权利为核心，以人类的多样性和包容性为基础，并提出了一系列的伦理指导和建议。

伦理实践阶段：很多企业都在探讨如何将 AI 原则贯彻到日常技术实践中。例如，Google Cloud 采取措施打造负责任的 AI；微软设立负责任 AI 办公室，全面推进负责任 AI 的落地实施。

9. 相关观点收录入未来论坛编著的《人工智能伦理与治理——未来视角》，人民邮电出版社，2023 年

图5 国外科技公司的 AI 伦理治理机构 [10]

企业内部 AI 伦理治理组织	组成或职责
微软负责任 AI 办公室 +AI 与伦理委员会	(1) 负责任 AI 办公室负责制定内部的 AI 伦理规则、审查敏感应用案例等；(2)AI 与伦理委员会下设七个工作组，对特定 AI 伦理问题进行研究、反思、建议，主动制定内部政策
谷歌负责任创新团队	谷歌内部旨在落实其 AI 原则的三层治理架构，包括 AI 原则审查委员会，对谷歌的 AI 产品和 AI 交易等进行伦理评估，以落实其 AI 原则
Facebook AI 伦理小组 +AI 伦理研究所	(1)2018 年成立，旨在确保其 AI 系统尽可能做出符合伦理要求的决策，避免偏见歧视、虚假信息、政治极化等；(2)2019 年成立，投入 750 万美元，与德国 TUM 共建
IBM AI 伦理委员会	内部 AI 伦理委员会，由跨学科团队组成，IBM 首席隐私官担任主席，直接向公司最高层汇报
Twitter 机器学习伦理、透明与责任（META）小组	在公司层面推进负责任 AI（包括对算法决定的责任，结果公平公正，决策透明，能动性与算法选择），包括研究 ML 决策的影响、完善 Twitter 的服务、打造可解释 ML 方案等
DeepMind 伦理与社会团队	其独立顾问小组包含多名不同背景与学科的外部专家，就 AI 伦理问题向 DeepMind 提供反馈与指导
Salesforce 技术伦理办公室	全称为 Office of Ethical and Humane Use of Technology，下设伦理咨询委员会 (Ethical Use Advisory Council)，由内外部跨学科专家组成
SAP AI 伦理顾问小组 +AI 领导委员会	2019 年成立，AI 伦理顾问小组由学术、政治、产业等领域的外部专家组成
Sony AI 伦理委员会	2019 年成立，负责落实 Sony 的 AI 伦理指南，AI 伦理办公室为其秘书处

已有的行业实践包括设立伦理委员会，组织培训和审查，以确保在设计活动中考虑伦理要求；构建“AI 模型说明书”，推动 AI 算法的透明性和可解释性。例如，谷歌推出的“模型卡”工具集（Model Card Toolkit），以及 IBM 的 AI 事实清单等。行业实践还包括以伦理工具方式呈现的“伦理即服务”，针对可解释、公平、安全、隐私等方面的伦理问题，研发并开源技术工具。例如，IBM 根据其 AI 伦理的五大支柱——可解释性、公平性、鲁棒性（稳健性）、透明性、隐私性，提出了 5 种针对性的技术解决方案 [11]。

10. 曹建峰，《上海师范大学学报（哲学社会科学版）》,2023 年第 4 期
11. AI 伦理与 IBM：https://www.ibm.com/cn-zh/topics/ai-ethics

3.2 人本智能的行业应用

AI 在当下乃至未来的关键领域具有巨大潜力。对应于个人发展、组织和商业变革、气候变化等全球挑战，以及人类社会的伦理和治理等四个维度，AI 相关的影响都将不断深化。

其一，AI 提升个人能力和改善个人生活质量。AI 的发展有助于提升个人工作效率，显著改善个人生活，乃至提升社会整体福祉。

其二，AI 推动组织、商业模式与社会结构变革。AI 的发展不仅会彻底改变企业的运营方式，还将对整个社会产生深远影响。

其三，AI 具备解决全球重大挑战的潜力。AI 技术的进步能够提供前所未有的创新解决方案，有助于解决诸如气候变化等全球性问题。

其四，AI 与伦理、安全问题。在当下以及未来的 AI 开发过程中，必须重视伦理与安全问题，确保技术的使用是负责任且安全的。

人本智能提供了一套方法论和价值准则，可指引 AI 技术聚焦“人本”，确保 AI 发展的方向不偏离。本篇章围绕 AI 与上述四个维度的融合，选取六大典型行业场景，展示人本智能理念的应用，并展开讨论和剖析——通过不同行业和场景的案例，具象地展示人本智能发展观的生命力。

行业一：传媒与文艺创作

1 | 行业背景

以机构媒体为典型参与者的传媒与文艺创作领域，其核心价值是内容生产，同时也要思考如何将生产的内容变得更有传播力和影响力。但在万物皆媒、万物互联时代，“内容创作者”们遭遇的挑战和困境也在日益加大。

传而不广、传而不响：在互联网时代，人们面临着海量的信息，以及多样化传播渠道的分流，新闻传播想要在众多信息中脱颖而出，其难度相比前互联网时代有指数级增加。而传统机构媒体等内容生产者所固有的内容生产方式、生产工具、生产流程、生产成果和传播渠道不能完全匹配新的传播环境下用户对高质量、强互动、重精准等内容产品的需求，进而造成“传而不广、传而不响”等困境。

对传统专业创作者的融媒体能力要求提高：面对互联网高效的信息需求，从业者疲于应对大量稿件等内容快速更迭的时效要求，使得内容产品几乎没有了推敲的时间，质量下降；同时，在新媒体的模式下，新闻信息的快速叠加更新，使媒体人的劳动成果被更快速地碎片化，导致从业者无暇生产更高质量、具有社会价值的深度调查报道等。

2 | AI 赋能

AI 技术和新闻传媒等内容生产行业具有天然的应用匹配性，在文本创作、音乐合成、视频生成等方面，AI 技术都能提

供处理技术上的支撑。据此，在 AI 赋能传媒与文艺创作领域，践行人本智能有很大的发展潜力。

典型体现为 AI 技术赋能，解放内容生产力：无论是新闻报道还是电影艺术视频制作，现在的流程都相当复杂，从前期材料搜集、选题策划、构思到后期执行，存在诸多流程，创作者往往需要花费大量精力在低水平的繁琐技术操作上。相比传统内容生产模式，AIGC 可以通过交互式快速生产，显著降低生产成本，极大提升生产效率。

3 | 行业案例

当今时代内容依然为王，优化内容生产、提高供给侧内容质量、满足受众多元化内容需求，是提升主流媒体传播力的关键。在 AI 技术的加持下，媒体等内容从业者可大幅度提升高质量内容供给的效率，将自己从日常繁杂的信息搜集等工作中解放出来，投入更有创造力的工作中。

案例 1

人民日报“创作大脑 AI+”平台

《人民日报》社推出的“创作大脑 AI+”平台大量运用 AIGC 技术，通过训练模型和对大量数据的学习，集纳了近 20 款智能工具，可及时制作、快速生成多模态新媒体产品，一站式完成采访、拍摄、直播、剪辑、发布等全流程工作，极大地提升了新闻采编等从业者的产能和创作力。

几个特别的功能都可体现这些价值：

海量创意资源在线共享：新闻编辑人员在编写稿件和资讯时，一大痛点是前期需要耗费较多时间用于搜集行业或者报道主体相关背景材料。而“创作大脑 AI+”可提供海量创意资源在线共享、巨量内容标签、多级垂直领域分类支撑素材深度挖掘；这项功能凭借强大的结构化智能搜索，可助力从业者短时间内构建起全域知识图谱，高效完成对报道主题的背景全扫描。

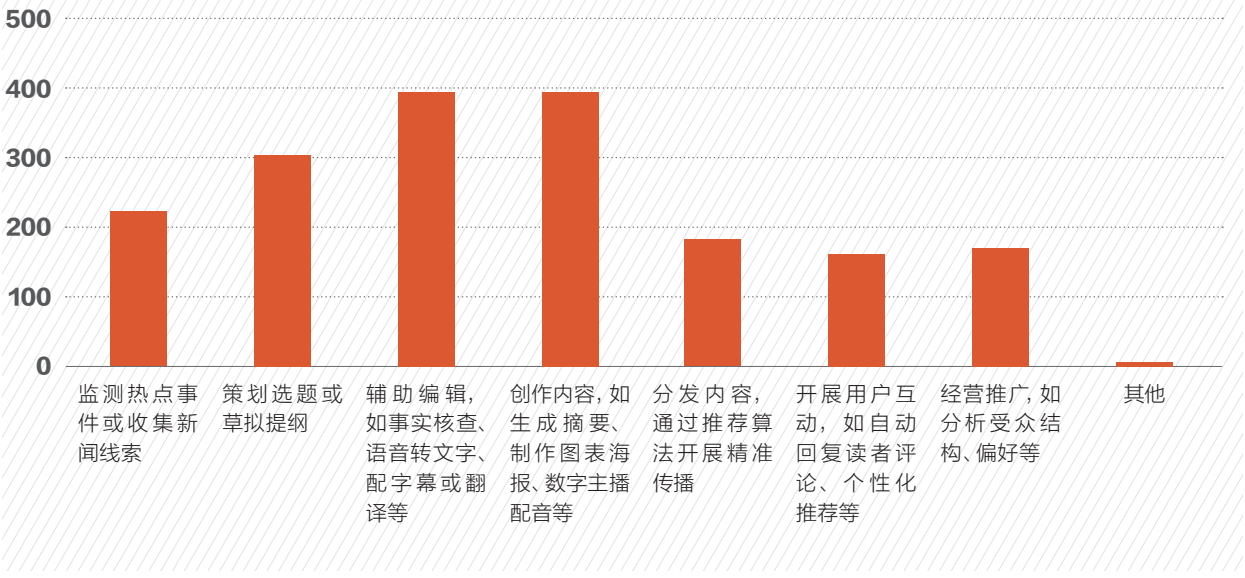
在内容生产的工具和流程优化上：“创作大脑 AI+”平台置入了动态可视化 Web 界面，嵌入视频、图片、音频等数十种 AI 处理工具，可帮助编辑优化视频操作流程，帮助创作者快速完成音视频内容生产，将智能化处理手段赋能媒体各业务场景。

另外，“创作大脑 AI+”平台还融入与用户端的智能互动，把对人的关注扩展到从业者之外的消费者用户端。通过“创作大脑 AI+”平台，媒体构建全新应用场景，支持观众在观看视频过程中持续互动，比如编辑可在视频故事框架设置多个分支节点，每次互动点击都会影响剧情走向。这种交互动作深度沉浸，也极大增加了传媒内容的趣味性，让传播力更广、更深。

值得一提的是，近年来传媒行业普遍积极应对 AI 大潮，推进 AIGC 方向的技术创新，以优化用户体验，更新完善产品。以财新为例，财新于 2024 年初成立 AI 实验室，并早在大语言模型概念普及之前，就研发了“织雀”系统，将海量历史新闻档案进行结构化存储，采用业界领先的 Transformer 技术分析文章中的实体、术语和地域信息，并与数据资产智能关联。而后再基于“织雀”的数据，进一步研发了“群雁”工具，为采编提供灵活的资讯流编辑

功能，实现 AI 与人工并行的新闻生产，提升效率与用户体验。此外，财新还推出了多种大语言模型应用，包括文字审校工具、嘉宾推荐、翻译工具以及新闻话题提取等，集成在 Caixin Global 中，进一步增强内容生成与数据分析能力。财新还引入生成式 AI 技术，推出基于财新内容的 RAG 产品和 AI 客服，帮助用户深入挖掘高质量报道。

图6 媒体机构对生成式 AI 已探索或者有意探索的应用场景



案例2

联想 AI PC 助力纪录片《西野》拍摄制作

不止于在机构媒体端，与此息息相关的电影等文艺创作领域也有 AI 技术赋能的空间。2024 年，OpenAI 宣布 Sora 诞生。这款文生视频的人工智能模型，让很多从事视频、动画以及电影制作的专业人士和艺术家深感震撼。著名导演陆川认为，AI 辅助影视创作最直接的效果，就是利用 AI 可以极大地提升创意视觉化的速度。据此，陆川与联想集团合作拍摄了由联想 AI PC 支持创作的纪录片《西野》。

“如果我们用传统的方式去做，storyboard、motionboard、previs 等等，再加上 CG……可能要做两个月，但是这个片子，我们只用两天就做完了。它极大地节省了成本，还有人力时间。”陆川总结。另外，对很多拥有电影梦的刚入行年轻人来说，AI 辅助电影制作“可以跨越资金以及专业的门槛，直接将头脑中的奇思妙想转化成一部实实在在的电影”。对内容创作者来说，AI 和 AI PC 是一个知识平权的过程，能够让年轻创作者跨越门槛，快速实现创意的视觉化。同时，在未来的实践过程中，又能够通过 AI 快速缩短、深化并强化创作链条，整个 workflow（工作流程）也会随之彻底改变。

行业二：教育行业

1 | 行业背景

随着教育理念的转变，学生越来越需要个性化的学习方案。认识到每个学生都是独特的个体是教育的起点——每个学生都有不同的学习习惯、兴趣和能力，传统的“一刀切”教学模式已经无法满足他们的需求。同时，面对科技时代的浪潮，固有的学习和教育模式已然落后于 AI 等当代技术的发展。比如，如果教育只是为了知识传授，相对于大模型数据库的海量存储与迁移，作为教育者的教师事实上已落后于 AI，作为学习者的学生也可能面临没出校园就已然落后于社会和科技发展的情况。

这些都对教育机构和教师提出挑战：学校和教师需要具备敏锐的观察力和深刻的理解力，需要具备足够的资源和方式去根据不同学生的特质量身定制教学方案，面对挑战提供适宜的支持，帮助学生在自己擅长的领域成长发展。

2 | AI 赋能

伴随 AI 的深入渗透，教育领域正在迎来教育生态和学习范式的革新。AI 对教育领域的赋能和提升，可体现在三个方面。

第一，以 AI 作为一种工具辅助人的教育。例如，把 AI 引入教学中，从而与学生产生以往无法产生的互动。

第二，对年轻一代来说，学习 AI 及其涵盖的特定专业和领域，将成为未来必须具备的一种能力。

第三，为 AI 的未来作准备，迫切需要提升学习者的 AI 素养。

以高等教育场景为例，结合人本智能的内涵和发展原则，AI 赋能教育——不仅仅拓展了学生学习的边界，使个性化学习的普及成为可能；也拓展了教师教学的边界，使高质量教学的普及成为可能。

3 | 行业案例

案例 3

清华学子的 AI 搭子——“清小搭”

当 2024 年秋季开学季，清华大学 2024 级本科新生迎来了属于自己的 AI 搭子——“清小搭”，一款基于最前沿大模型智能体技术构建的学生智能助手。

作为面向清华学生提供智能问答和伴学服务的成长助手，“清小搭”在设计和开发之初就充分融入学生的学习场景。

“清小搭”可以为学生提供从校园生活小贴士到学术难题解答的详尽帮助；同时还能为学生提供智能伴学工具箱，并且通过云盘记录学习成果和成长轨迹。

学生可以先进行“学业志趣自测”，根据自己的专业发掘学术兴趣和潜能，从而获得相关的选课建议。

在大一期间，“清小搭”主要以智能迎新的方式，帮助学生轻松适应校园生活，在此过程中会提供个人成长档案、智慧日历、智能工具箱和成长云盘管理等功能。

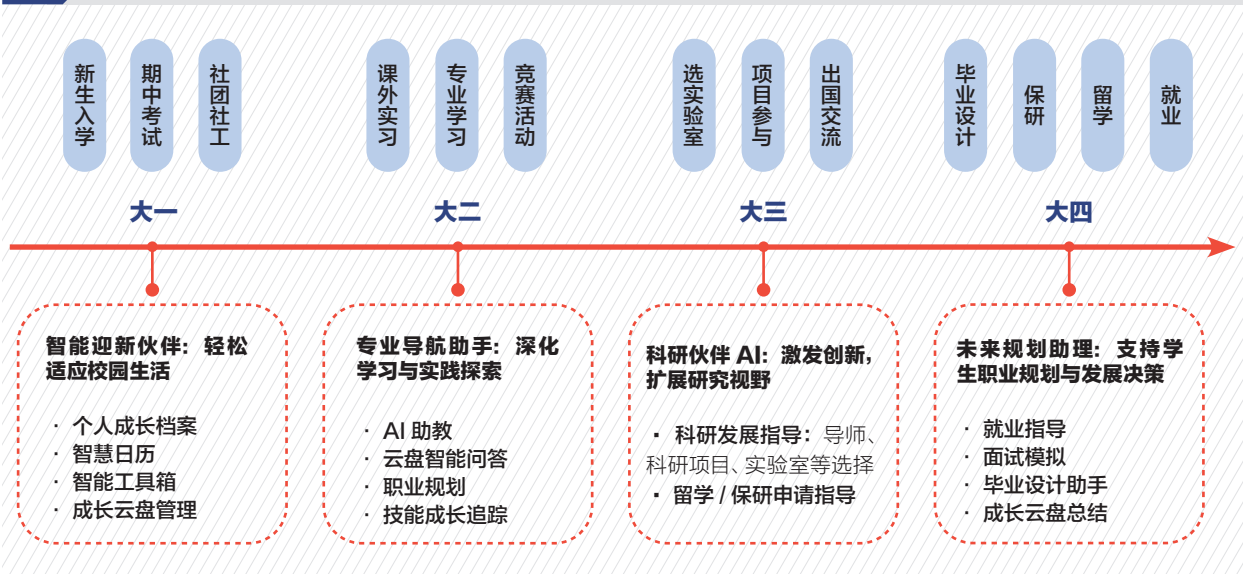
在大二期间，“清小搭”开始聚焦成为学生的专业导航助手，深化学习与实践探索，为学生提供 AI 助教、云盘智能问答、职业规划、技能成长追踪等功能。

在大三期间，“清小搭”偏向于成为学生的科研伙伴 AI，激发学生创新能力，拓宽研究视野。同时，提供导师、科研项目、实验室选择等方面的科研发展指导，以及留学、保研申请指导等。

在大四期间，“清小搭”会成为学生的规划助理，支持学生职业生涯规划与发展决策，同时提供就业指导、面试模拟、毕业设计助手和成长云盘总结等功能。

从智能迎新伙伴、专业导航助手，到科研伙伴 AI、未来规划助理……未来，通过与 AI 助教联动、构建 AI 智能体群聊等机制，在“清小搭”这一平台上，AI 将为教育教学注入新的活力，集中展示 AI 在促进个性化学习体验、增强包容性，以及推动协作学习模式革新等方面的无限可能。

图 7 “清小搭”针对学生不同学习阶段进行的功能设计规划



行业三：制造业

1 | 行业背景

作为广义制造业的源头之一，矿山企业是一个移动的生产车间，工作环境复杂多变，安全事故隐患众多，尤其是煤炭企业，水、火、瓦斯、顶板、煤尘“五毒俱全”，安全生产压力巨大。同时，矿山行业又是劳动密集型行业，生产线长环节多，人员流动性大，从业者面临的安全危险多，亟须通过技术等手段实现从业者人身安全的保障。

而在制造业的众多领域中，我国造船业已连续 14 年保持造船完工量、新承接订单量、手持订单量世界第一，市场份额超过 50%。在这一进程中，离不开高质量、快速响应的工业设计。但越是高质量的船舶制造越需要设计师们精细的投入和快速的解决方案，以响应越来越多来自国内外的订单需求。但工业设计有其固有规律，工作环境中不存在矿山工人可能面临的危险，设计工程师往往不得不投身于多次重复构建配件模型、修改草图中，设计完成后，仿真等多环节的协作也存在滞后，修改意见难以集中，这些环节使得工业制造的整体流程较慢、效率较低，也限制了设计工程人员生产力进一步发挥的空间。

2 | AI 赋能

在生成式 AI 爆发之前，传统 AI 在优化制造流程、自动化生产、质量控制、预测性维护以及供应链管理等方面已有持续应用。随着 AI 大模型在制造业场景的深入应用，其在工业设计、制造、仿真等“深水区”取得突破，并将人本智能的理念扩展到更广领域。

一方面，以人为本的包容发展和底线坚守：在人本智能的理念下，AI 发展将以人为本的原则扩展到矿山工人及矿山等场景中，利用 AI 技术帮助人类解决一些有意义的难题，确保每个人都能享受到 AI 带来的益处。

另一方面，以人为本的设计运作和价值提升：在人本智能的理念下，AI 发展旨在促进包括设计工程师在内更多的能力和价值提升，AI 的应用是为了赋能人类，而非取代人类。

3 | 行业案例

案例4

鲲云科技 AI 系统守护矿工生命安全

面对高危场景，基于 AI 技术的智慧煤矿解决方案成为助力 AI 向善、以人为本的典型实例。2024 年，鲲云科技 AI 视频分析系统守护矿工生命安全——宁煤智慧矿山项目，凭借突出的技术能力及应用效果，入选联合国“AI for Good”创新案例集。在该案例场景中，鲲云科技携手宁煤集团，基于领先的可重构数据流架构 AI 技术，建设了算力、算法、平台一体化的 AI 视频分析系统，致力于提升矿区安全管理效率，守护矿工生命安全。

具体来说，鲲云科技智慧矿山 AI 视频分析解决方案结合矿工实际作业场景，覆盖采煤工作面、掘进工作面、主运输系统、辅助运输系统及其他井下关键场景，提供人员闯入、未穿戴安全着装、采煤机 / 刮板机等设备运行状态识

别、人员违规跨越、皮带撕裂、皮带跑偏、皮带异物、烟雾火焰等异常情况的智能识别和实时预警，可以有效避免因为人工管理的主观性技术差错导致的安全事故和经济损失，大幅提升矿井综合管理效率和智能化水平。在宁煤集团白芨沟矿区部署该系统后，矿区实现安全风险秒级预警，半年内累计触发报警 10 万余次，智能算法平均实测精度达 95% 以上，大大降低了煤矿生产过程中的安全风险，实现矿区零事故，保障近千名矿工的生命财产安全。

案例5

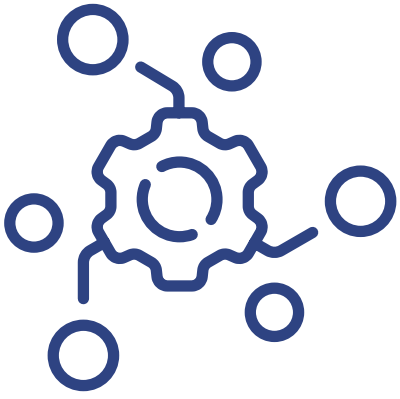
设序科技让船舶设计更简单有序

上海汽轮机厂一年要设计近 150 台套机组，为此，企业 30 余位设计师曾将大量精力耗费在配件模型的反复构建及草图修改上。在设序科技基于工业大模型的生成式智能出图设计平台的加持下，上海汽轮机厂可实现汽轮机零部件自动出图以及尺寸、形状、标记等自动标注。设计工程师只需负责最终审图及适当纠偏。过去，设计师若想把图纸中特定线条改色，需手动逐一操作；现在，设计师向大模型输入一段自然语言指令，大模型即能将指令自主“翻译”成 API 序列并准确迅速执行，快速将设计工程师的设计意图转化为设计结果。据前期测试，交付级别图纸的设计效率提升了十倍，极大地解放了设计师，让他们有精力去做更有创新价值的工作。

总体而言，当前，受限于计算资源、工业数据、模型可解释性等因素，我国对工业大模型的探索仍处于早期阶段，AI 深度赋能制造领域尚有较大发展空间。不过，从上述案例可以窥见一些可能的破题“钥匙”。

1. 从产业端来看，“大模型 + 小模型”正逐步成为产业主流技术路线，推动全球 AI 产业发展和对传统产业的渗透改造。在政策和行业方面，工业领域应该更具意识、更加主动地培养熟悉“大模型 + 小模型”的复合型人才，营造良好的发展氛围。同时，积极打通并构建高质量的大模型工业数据集供给体系，加强大模型工业应用生态的构建。

2. 从企业端来看，可将“设计工程师等工作人员 x AI”的整体效能纳入视野，以助推企业更积极地引入 AI。在这一视野下，以人为中心、释放人的创造力是核心原则。需要指出的是，AI 虽然是日常工作流程中不可或缺的一部分，但它将成为工具箱中的一个补充，其最终目的并非取代人类的创造力。



行业四：交通出行

1 | 行业背景

随着我国城市化进程不断加快，城市人口数量迅速增加，私人汽车保有量快速增长，导致城市交通拥堵问题日益严重。同时，道路交通安全及事故防范一直是交通出行中的痛点。

世界卫生组织（WHO）发布的《2023 年道路安全全球现状报告》显示，自 2010 年以来，道路交通死亡人数为每年 119 万人，特别是行人、骑自行车的人和其他弱势道路使用者面临严重且不断上升的死亡风险。

2 | AI 赋能

人、车、道路，组成了交通出行的基本元素。要解决交通领域的顽疾，AI 在作用于三者的过程中形成了两组重要关系。

AI 与人、车：随着 AI 对汽车设计制造等全面渗透，汽车拥有了更多智能化、自动化的功能。这些功能是以技术为中心还是以人为中心，代表了不同的发展路径。在人本智能理念下，汽车厂家、AI 技术开发者深入洞察用户痛点和难点，了解使用者、乘坐者甚至周围行人在自动驾驶等技术日益普及下不同场景、不同维度的多元需求。

AI、人、车与社会治理：AI 自动驾驶存在物理智能层面的风险，应通过合适的机制在 AI 技术的基础研发阶段就树立对人价值的尊重，在收集数据、构建模型、安全对齐等维度强化治理意识，并积极参与和遵守相关的政策法规。

3 | 行业案例

案例6

“萝卜快跑”们上路，自动驾驶引热议

场景 1：自动驾驶本身

国内自动驾驶出行服务平台“萝卜快跑”展示了自动驾驶对大众出行带来的福祉——无论是便捷性还是安全性。

“萝卜快跑”定位为面向公众的常态化出行平台，旨在解决传统出租车和网约车存在的一些问题，如司机服务质量不一、交通拥堵、环境污染等。

在 AI 技术加持下，“萝卜快跑”拥有 L4 级自动驾驶技术，拥有高精度地图、智能感知、精准决策等核心技术，能够确保车辆在复杂道路环境中安全、稳定地行驶。

在服务上，用户只需通过手机应用程序预约车辆，系统会自动匹配最近的无人驾驶车辆。车辆到达后，用户通过手机解锁车门，车内设有智能语音助手，可以提供导航、娱乐和信息查询等服务。同时，它还具备多车型服务能力，满足不同乘客的出行需求。

在商业化运营上，“萝卜快跑”通过与多家汽车制造商、科技公司合作，不断优化车辆成本和服务质量，使得自动驾驶出行更加经济、实惠。同时，还积极与政府合作，推动自动驾驶技术的普及和应用。数据显示，“萝卜快跑”平台累计订单已超过 700 万单，自 2021 年以来，平台已在包括北京、上海、广州、深圳、重庆、武汉、成都、长沙、合肥、阳泉、乌镇在内的全国 11 个城市开放载人测试。这不仅展示了其技术的成熟度，也展示了自动驾驶在技术、政策、商业化上的全面推进，为变革未来交通出行使其更加“人性化”做了尝试。

场景 2：自动驾驶外溢影响

目前，自动驾驶在安全性方面已经比人类驾驶高出至少 10 倍，这显示出其在安全性上的优势。但自动驾驶还存在一些问题，如驾驶不够熟练、不够像老司机（老司机开得更人性化，不仅安全，而且驾驶水平高）。在此基础上，自动驾驶可以纳入更广泛的考量维度，即如何成为一个能长久被接受的出行方式，这涉及自动驾驶在产业、商业及社会上的应用和被接受程度。

随着自动驾驶技术的深入发展以及自动驾驶出租车商业化进程的提速，人工智能的应用如何促进而非挤压就业，成为一个重要议题。正如“萝卜快跑”在武汉的规模化上路，引发的热议：一方面，很多人对其持欢迎态度，认为无人驾驶出租车价格低、服务好，不用担心司机在车内抽烟，也不用害怕司机路怒斗气，司乘矛盾得以消除；另一方面，也有人认为无人驾驶出租车可能会冲击网约车司机的就业，挤压了人的生存发展空间。

自动驾驶外溢影响不仅仅是技术对人就业替代的问题，背后更是长期以来人与技术之间的关系重置——随着人工智能的广泛应用，其对劳动者的技能要求逐渐提高。缺乏技术背景或无法跟上技能要求转型的劳动者，可能失去就业机会，导致就业市场的两极分化，从而进一步扩大不同劳动力群体间的收入差距。

事实上，技术进步往往并非中性的，以人工智能为代表的技术进步是超级偏向技能的技术进步。好的一面是，每一轮的技术革命和更新，都会在意想不到的地方为社会创造出新业态、新岗位和新机遇，例如自动驾驶技术研发、数据分析、后台设计等岗位。

其关键在于通过治理模式的创新形成 AI 技术负面影响的源头治理，合理引导和应用人工智能技术，融入人本智能的理念，保持合理的节奏，加强职业培训和教育体系改革，帮助劳动者掌握人工智能相关的技能，以实现其与就业的良性互动，确保 AI 对社会福祉和所有人利益的提升。

行业五：医疗健康

1 | 行业背景

医疗健康是全球范围内最重要的民生领域之一，随着人口老龄化的加剧、慢性病患者的增加以及人们健康意识的提升，医疗健康需求持续增长。

然而，医疗健康行业持续存在医生供给不足、医疗资源配置不均衡、诊疗效率不高等问题，这些问题亟待解决。与此同时，随着人口老龄化发展，慢性病和并发症的负担会加重，医疗健康理念逐步从单纯的治疗疾病，转移到衡量和改善患者的生活结果上来，人们对于更高效、更精准、更便捷的医疗服务有着迫切的需求。

2 | AI 赋能

通过应用和普及 AI，将更好、更全面的解决方案输送给医疗系统，真正提升医疗质量、效率，惠及人群，让更多患者享受优质的生命关怀，是在人本智能理念下 AI 赋能医疗健康的方向。

AI 对传统医患两方关系的革新：伴随 AI 智能化程度的不断提升，AI 智能在医疗场景中对医患双方的价值都在提升——对医务人员来说，AI 是新的助手和伙伴；对患者来说，AI 亦是助理和伙伴，它能够帮忙打破医疗领域的专业鸿沟，促进医生与患者之间的交流。如何确保人人都能公平、公正地使用这一可能成为数十年来医学领域最具影响力的新技术，可能成为未来一段时间医疗健康领域的重要议题。

AI 在医疗健康普惠领域的延伸：AI 普惠在医疗健康领域的延伸，极大地展现了 AI 技术背后可承载的责任和文化价值属性，也是 AI 助力联合国可持续发展目标（SDGs）和 ESG 的典型体现。

3 | 行业案例

案例 7

联想用 AI 帮助渐冻人“开口”说话

肌萎缩侧索硬化（ALS），又名渐冻症，是一种神经系统罕见病，被世界卫生组织（WHO）列为与艾滋病、癌症等并列的五大绝症之一。随着病程加深，ALS 患者会逐渐出现全身肌肉萎缩、无法说话、无法吞咽、无法呼吸，直至死亡。著名物理学家霍金就是渐冻症患者，并因渐冻症于 2018 年 3 月 14 日去世。

迄今为止，人们对 ALS 的致病原因仍了解不多，治疗渐冻症也是一个世界难题。但医疗、科技产业以技术为导向致力于加强 ALS 患者能力的努力一直没有停止。联想集团携手致力于创新辅助技术的非营利组织斯科特 – 摩根基金会（SMF），推出了一款针对肌萎缩侧索硬化症（ALS）患者及其他重度残疾人士的 AI 解决方案，展示了产业界的行动。

该解决方案结合了由联想制造的圆形键盘界面、预测式 AI，个性化 AI 语音复制和超真实虚拟形象，以及眼球追踪技术，可助力 ALS 患者对外进行快速、准确和个性化的沟通。以往，受到 ALS 早期症状中可能出现的嗓音恶化的限制，传统的语音存储是一项艰巨的过程。此前的模型还明显缺乏情感表达能力，语音节奏机械，韵律怪异——但最新的 AI 技术改变了这一切。即使是质量较差的有限音频样本，AI 也能根据需要生成令人满意的语音。

作为该解决方案的受益者之一，渐冻症患者 Erin Taylor 通过 AI 为自己打造了一个数字分身，保留了 Erin 自己的声音，并用几分钟渲染完成，得以实现继续为孩子们唱摇篮曲的梦想。“科技，尤其是 AI 的最新进展所可能带来的改变，使我获得了新的力量和希望。” Erin 曾在社交平台分享了自己的经历和观点，“用确诊前的声音说话使我难以置信，但最让我兴奋的是参与开发更智能的技术，让其他人受益。”

不止于 ALS，联想还针对阿尔茨海默症等其他身体机能受限人士开展了很多 AI 技术赋能提升行动。比如在联想与失智症创新组织（Innovations in Dementia）合作发起的“阿尔茨海默症智能计划”中，联想基于阿尔茨海默患者生活经历定制了专门的 AI，打造出逼真的 3D 数字分身（avatar），为正在应对阿尔茨海默诊断的患者和家庭提供 24 小时的对话式虚拟伴侣。这些举措展现了联想“AI for All”的愿景，也进一步彰显了 AI 时代应以人为本，让 AI 落到实处，推动其普惠性和包容性，最终实现“人本智能”的新科技发展观。

案例8

百川智能用 AI“造医生”

由搜狗公司创始人、前 CEO 王小川创立的国内唯一一家专注医疗领域的头部大模型公司百川智能，在落地场景上选择的就是医疗领域。

在百川智能看来，医疗行业的一大痛点是“无限的需求与有限的供给”，即医生短缺和区域不平衡等挑战。此前的互联网医院只是通过技术连接患者与医生，是生产关系层面的改善，并没有从生产力层面根本解决这一问题。百川智能发布了多款大模型，用大模型打造出足以赋能基层的“AI 医生”——大模型可以在院内做医生助理，在院外做健康顾问，能进行全病程的管理。在这个过程中，大模型可以帮助医生进行尽可能多的数据收集，这对医生来说就像临床试验一样。大模型可以有效降低误诊率，尤其是在基层医院中，这一点尤为重要。通过大模型，可以为每个患者单独建模，最终走向精准医疗，提高预防、诊断、干预的水平。

另外，大模型还可以作为患者的 AI 健康顾问：它不但具备丰富的医疗知识，还能够模拟医生的思维方式，在用户提出问询后，根据用户的问题持续提问，从更多维度更深入地了解症状，收集到足够多病症信息后再进行综合判断，给出诊断结果和用药建议。

AI 健康顾问的应用非常广泛，其中一个典型场景是智能分诊。患者的普遍痛点是，不知道如何描述自己的病情，

也不清楚应该去哪个科室挂号。通过持续提问互动，AI 健康顾问可以像一个全科医生助理，它具备丰富的医学知识，并且熟悉医院科室的配置，能够为患者提供精准的科室推荐，也有效优化了医疗系统资源的配置。目前国内已经有部分医院在试点应用智能分诊类 AI 应用，实现了分诊准确率与患者就医体验的大幅提升。

图 8　百川智能将医疗 AI 类比智能驾驶，从 L0 到 L5 划分了几个等级

根据《汽车驾驶自动化分级》国家推荐标准 (GB/T40429-2021)，汽车驾驶自动化从低到高分为六个层级	
目前医疗行业在大模型的加持下有机会做到 L3 级别	
等级	**AI 医疗**
L5	完全自动化健康管理 :AI 能够管理患者的整个健康旅程，从预防、诊断到治疗，无需人工干预
L4	高度自动化诊疗 :AI 可以在大多数常见疾病中独立做出诊断和治疗决策，仅复杂案例需要医生介入
L3	条件自动化诊疗 :AI 可以在特定条件下自动推荐治疗方案，但在关键决策时需要医生确认
L2	多模态辅助 :AI 可以整合多种数据源 (如病历、影像、实验室结果)，提供更全面的辅助信息
L1	辅助诊断 :AI 可以提供数据分析或图像识别等单一功能的辅助，帮助医生做出更好的决策
L0	传统医疗 : 医生负责所有的诊断和治疗决策，没有 AI 介入

在医院之外，缺乏医疗常识的普通民众也可以向 AI 系统寻求有关自己和家人健康问题的建议，从而获得更丰富、更准确的医学知识、医疗建议和保健指引。同时，AI 具有改善人类生活的巨大潜力，特别是对那些处于劣势地位、被边缘化的或易受伤害的群体。

总体而言，在 AI 助力下，传统的“医患两方关系”范式向结合了医生、患者和机器的“三方模式”发展——以所有人的健康福祉为最终目标和指向，突出体现为 AI 对现有医疗体系效率的提升，以及对医疗资源公平化的促进。

行业六：环境与生态保护

1 | 行业背景

近年来，在“绿水青山就是金山银山”理念指引下，我国积极提升全社会生态文明意识，环境保护和生态文明建设均取得显著进展。但从更广阔的全球视野来看，人类赖以生存的地球正面临气候变化、生物多样性丧失和污染废物等多重危机冲击，对人的生存发展以及未来可持续发展产生威胁。在此背景下，人们必须与自然合作，而不仅仅是利用自然。

2 | AI 赋能

与自然合作的一种方式便是科技赋能——科技尤其是人工智能（AI）技术正在给环境领域带来革命性的变革。

AI 技术的运用，为应对气候变化、保护环境和生态提供了新的思路和手段。通过 AI 技术，人们可以更加精准地监测和评估环境状况，进而为科学决策提供支持。同时，AI 技术还可以帮助人们设计更加环保的产品和解决方案，推动绿色产业、环保事业的发展，促进可持续发展。

具体从人本智能的发展观来看，AI 在该领域的应用至少可在两方面加强：

① AI 赋能人应对环境的能力，而非取代人。

② AI 直接提升环境和生态发展水平，进而为人和社会的整体发展奠定可持续的基础。

3 | 行业案例

案例9

北京亦庄“ AI 之城”的环境管理

在北京亦庄，空气检测和道路清洁都有 AI 的身影——

具有“安全守卫者”和“大气污染侦察兵”双重身份的无人巡逻车运用走航监测技术，24 小时不间断运行，可实现对道路积尘负荷、细颗粒物（PM2.5）、粗颗粒物（PM10）、气态八参数污染物和氮氧化物等的实时监测，助力精准监控大气环境质量，制定应对措施。同时，无人驾驶环卫车可自动进行道路清扫、洒水降尘等多种城市环卫作业，错峰清扫，与传统环卫作业互为补充，进一步降低道路尘土残存量，助力城市洁净美丽。

在北京亦庄，鸟儿也获得“智慧”监测和看护——

为了更好地了解北京经开区区域内的生物多样性情况，此前鸟类监测大多采用“人工 + 观测设备”的方法。但这种方法人力物力投入大，且对监测人员的要求较高，难以保证鸟类监测的准确性、连续性、完整性。

2022 年 7 月，北京麋鹿生态实验中心与中国科学院半导体研究所合作建立了一套基于“人工神经网络”的计算机深度学习鸟类识别系统，为鸟类多样性调查和动态监测提供了创新手段。2023 年初，麋鹿苑更是首次以 AI 方式监测到国家一级保护动物白尾海雕，这也是此地采用“鸟”脸识别后首次监测到珍稀猛禽。AI 系统的使用，显著提升了鸟类监测效率和识别准确度，同时也为经开区生物多样性保护与科普宣传提供了更有力的技术支撑。

案例 10

联想集团“ AI+ 动物保护模式”

AI 等技术可以为应对气候变化、环境和生物多样性保护等困扰社会发展的老问题提供新答案。站在新的人机关系角度来看，AI 以其强大的数据处理能力、生成能力和创新应用，已经成为人类面对环境议题的新助手、新伙伴。

比如，在保护濒危物种和维护生态平衡方面，AI 视觉技术可以自动识别并统计野生动物，助力科学家更有效地监测物种数量变化。此外，AI 结合卫星图像分析，可以实时监测森林砍伐、珊瑚白化等栖息地变化情况，以便及时采取保护措施。

国内产业界依据不同的企业禀赋，在“AI+ 环境”领域进行了积极探索。以联想集团为例，作为国内最早投身 ESG 实践的企业之一，联想集团以 AI 为代表的智能技术应用于生物多样性保护一线及公众科普工作中，并以此联结更多人群参与环境保护事业，成功打造出具有可复制性的生物多样性保护智能解决方案。

自 2021 年以来，联想集团持续探索如何通过数字化、智能化技术创新，为生物多样性保护工作的老问题提供新解法。2023/24 财年，联想集团把握 AI 时代机遇，将 AI 技术应用于长江江豚、雪豹、兔狲等珍稀动物保护，携手中国国家地理 · 频道制作全球首份长江江豚生态地图与观测指南，联合西宁野生动物园建设高原精灵智慧守护站，通过 AI 技术创新打造可复制的智能生态保护范例。

在制作全球首份长江江豚生态地图与观测指南的过程中，联想集团通过 AI PC 等智能设备为制作团队提供了强大的算力支持，帮助其克服了维保条件有限、网络资源匮乏、人力资源受限等难题，得以展开高效的数据分析和监测，并提供精准的智能个性化服务。

此外，联想集团也将 AI 赋能生物多样性保护的尝试拓展到了西宁野生动物园智慧保护项目中。联想集团通过 AI 技术和算力，从雪豹行为 AI 识别分析、AI 游客互动系统、智慧救助系统、智慧繁育系统等多维度入手，创建了一个应用于青藏高原的生物多样性保护的解决方案，这是中国第一个以动物为中心的智慧园区解决方案，也是中国第一个 AI 动物园样板项目。

在通过 AI 赋能西宁野生动物园一线保护工作的基础上，联想集团还携手著名导演陆川拍摄纪录片《西野》，助力

“AI+ 动物保护模式”的进一步推广。最终形成从 AI 赋能生物多样性保护，到 AI 助力生物多样性保护工作公众科普，再到以 AI 赋能生物多样性保护场景开拓的“三步走”格局。

另一个科技品牌腾讯选择了不同的发展路径。腾讯以自研的 AI 万物识别通用模型 YOLO-World 为底层技术，携手相关合作伙伴推出了“野朋友计划”小程序（公测版），为公众提供一个参与生物多样性保护的全新数字化工具。在该小程序上，公众不仅可以看到我国不同生态系统中真实的野生动物生活状况，还能动手标注野生动物，帮助生态保护机构处理海量图像数据，从而在线上直接参与到生物多样性保护中来，让“公民科学”有了更多可能性。

殊途同归，产业界不同的 AI 实践背后有着共同的“人本因素”。一方面，产业界不把 AI 仅限于企业自身的实践，而是将其扩展到环境、ESG 等更广泛的领域，让 AI 技术在环境、绿色可持续发展等方面获得新的活力和可能性；另一方面，尝试将 AI 价值传导到更广泛的公众领域，让大众了解 AI 在构建一个更加绿色、可持续的未来中所具有的价值，也让公众从“AI+ 环境”的实践中获益。

CHAPTER 4
第四章

人本智能发展观：倡议与治理

“一项技术如果因为担心不良后果而过早实施控制，那么技术很可能就难以爆发。反之，如果控制过晚，已经成为整个经济和社会结构的一部分，就可能走向失控，再来解决不良问题就会变得昂贵、困难和耗时间，甚至难以或不能改变。”

——英国技术哲学家大卫·科林格里奇，“科林格里奇困境”

4.1 全球人工智能治理的努力——智能向善与负责任的AI

作为 AI 技术的创造者和使用者，人们有责任确保 AI 与人类的目标、价值观和道德观相契合。同样，人类需要关注 AI 带来的机会和挑战，并共同努力引领 AI 发展的方向。

随着在人工智能领域探索与实践的加深，人工智能的技术风险——数据安全性、算法透明性、系统稳定性及伦理争议，再如人类的自主能动、隐私保护、社会公平等问题逐渐暴露，由此引发社会的广泛关注。值得一提的是，在人本智能视角下的 AI 治理需要融入 AI 发展的完整生态中，即 AI 治理离不开技术、数据、应用和模式的构建，治理贯穿于四个因素之中，需通盘考虑。

特别是传统治理范式适用于生成式人工智能的治理局限逐步凸显，如基于对象场景的分散治理难以统筹治理全局，基于风险预防的事前治理难以精准预见风险，基于法律规范的硬性治理难以提升治理实效等。据此，如何建立适应新一代 AI 发展特质的新的 AI 治理机制，成为当下重要议题。

为了更好防范 AI 技术可能带来的负面性，全球主要国家和地区均从战略层面将人工智能治理问题纳入整体规划，并基于不同的发展阶段、产业基础、价值观念等因素，形成各具鲜明特点的治理方案。

① 规制型治理模式

以欧盟为典型代表。欧盟通过《人工智能法案》(*AI Act*)、反垄断法案、数据保护法案及其他监管工具，对 AI 技术进行管控。特别是《人工智能法案》为欧盟引入了一个共同的、具有约束性的 AI 监管法律框架。此外，欧盟也积极将其模式推广至其他国家，以争夺全球人工智能规则制定权。

② 创新型治理模式

与欧盟不同，美国尚未出台针对人工智能治理的国家层面的综合立法，却已经基于不同的技术应用场景颁布了规制性文件，特别是在自动驾驶、算法推荐等领域都有较为成熟的治理经验。总体而言，美国的 AI 治理倾向于审慎监管，鼓

励促进人工智能的创新及进一步发展，其相关治理更多为原则性阐述，更关注其在全球人工智能治理规则制定中的领导角色。

③ 自治型治理模式

国家不直接或者较少参与监管和治理，而是通过让 AI 产业中的企业等相关利益主体的自我监管来达到风险预防和规避的目的，呈现较强的行业自律治理特点。

④ 以风险管理为导向的治理模式

典型体现在欧盟《人工智能法案》的治理思路上。法案的主要内容是将人工智能系统根据风险评估分为四类，包括不可接受的风险、高风险、有限风险与最小风险。对于不同风险等级，法案采取了不同程度的监管措施。

⑤ 以算法透明技术为核心的治理模式

当前一些 AI 负面效应与算法有不少联系，如算法歧视、诱导用户沉迷、大数据杀熟、信息茧房等均与算法黑箱有密切的关系。而算法透明作为算法黑箱治理的重要手段，已经受到各方广泛关注。

不同的治理模式需结合本地实际情况因势而为。但一个总的趋势不变——建立以人为中心兼顾所有人利益的 AI 治理模式，在这个全面的治理体系下，既能管控风险，也能支持 AI 的创新发展，以确保 AI 技术的发展符合人类社会的长远利益。

正如清华大学苏世民书院院长、清华大学人工智能国际治理研究院院长薛澜所总结：第一，人工智能技术的发展速度远超监管制度的制定速度，首当其冲的便是发展与治理的平衡问题；第二，人工智能引发的问题前所未有，这使得人工智能监管变得空前复杂，没有任何一个监管机构能够单独管理好人工智能牵涉的各个方面；第三，随着国际格局的变化，国际社会不太可能形成一个单一的全球性机构，来监管人工智能发展的方方面面，因此加强国与国之间在人工智能治理方面的沟通与合作非常必要。清华大学智能法治研究院院长申卫星也建议，全球各国在人工智能领域的立法已经“由谈伦理、谈原则软法的时代开始进入真刀实枪硬法的权利义务实体规定的时代”，多元共治的局面将是未来的一个发展方向。

4.2 走向人工智能的未来——人本智能倡议

面对人工智能技术带来的机遇、变革以及潜在的风险挑战，人本智能有着怎样的应对策略，特别是如何把“以人为本”的价值观和伦理规则植入人工智能系统，并为其设立恰当的目标？另外，在治理体系层面，人本智能的治理目标、治理原则、治理模式和治理手段等关键性议题有哪些可探索的空间？

一个目标

人本智能的治理体系，最终应以科技造福人类为目标，需要以治理与发展兼顾并追求平衡为总目标。治理可从“以技术逻辑为出发点，以回归人类需求为最终目的地”的思路出发，使 AI 服务于人类的全面发展。

三个原则

① **以人为本：**以人为本原则强调人工智能发展应遵循人类价值观，促进人类社会向善，实现人类社会根本利益。人工智能在全生命周期中均需将人类根本利益放在首位。

② **包容公正：**人工智能发展应促进公平公正，保障人类总体福祉以及利益相关者的权益，促进机会均等。可通过技术培训等方式弥合数字鸿沟、消除偏见和歧视。

③ **权责一致：**在人工智能的研发与应用中能够有效问责，从政策、机制乃至法律等层面予以规范化管理。

四种应对策略

① 行业自律

人工智能研发者、使用者及其他相关方应具有高度的社会责任感和自律意识，严格遵守法律法规、伦理道德和标准规范。如同阿西莫夫的“机器人三定律”一样，人本智能坚持价值先行，树立行业规范，要求在 AI 设计运作过程中加入一些底线性的价值，使其成为植根于 AI 科研、实践和从业者的底层价值和行业自律规范的一部分。

一是加强行业规范意识和伦理共识建设。通过责任意识教育和伦理宣传，使各主体在参与人工智能治理时保持道德自律。例如，规定 AI 不能伤害人，并以此为法则来培养 AI 与人类的良好交互能力。无论 AI 如何进化，不管机器人有无意识、有无自我决策能力，它都必须遵守“不伤害”的原则。

二是强化伦理规范。对设计者、开发者和应用者在数据输入、算法设计和实践运用等方面进行伦理约束。实际上，有很多行业权威专家和从业者亦发起此类倡议。如张亚勤提出要打造善良的 AI，认为人工智能也应该有一些基本发展原则，比如应明确规定一些红线，并建立快速且安全的终止程序。一旦某个人工智能系统超越此红线，该系统及其所有的副本须被立即关闭。

三是强化伦理规范，完善监管治理等外部手段，为 AI 发展创造“宽严相济”的环境。“宽”指的是宽松、可以容错的技术创新环境，“严”指的是外部的伦理规范和监管举措为 AI 发展竖起护栏。

② 技术应对

将透明、可理解、负责任的技术应对策略贯穿技术发展的全周期。

一是要发展高透明度的人工智能。在创建、获取数据和算法时做到了解其工作原理，引入偏差检测机制，并根据正当需要公开相关信息。

二是要发展可理解性人工智能，能够面向特定受众提供运行工作中的关键信息。

三是建立审慎包容、分级分类的人工智能安全监管制度，以保障技术安全。例如，利用安全多方计算和匿踪查询技术建设系统防火墙，打造数据安全共享平台。

③ 法律规范

借助立法规制和司法归责，规范个人、企业及相关组织的责任和法律义务。

人工智能立法要统筹发展和安全，通过分阶段、分领域的统一立法，建立人工智能统筹协调机制。明确相关行业、领域人工智能主管部门的监督管理职责。构建多元主体参与的协同治理模式。例如，在较为成熟的领域和问题频发的领域启动立法研究，如自动驾驶、个人隐私保护、数据安全等，先行先试，以为后续立法奠定理论和实践基础。

通过包容性立法，划定人工智能监管的底线与红线，设置一定的免责规则，合理进行责任分配，构建综合问责体系。通过司法认定推动人工智能相关主体承担责任，还需从技术研发、数据利用、算力供给、激励创新、技术共享等方面完善规则，以更好地满足人工智能产业发展、人工智能与经济社会发展深度融合的制度需求。

④ 风险生态共治

将伦理与治理嵌入人工智能全生命周期已经逐渐成为重要共识，在人工智能发展整个生态（技术、数据、应用和治理）的每个流程和每个环节中，把存在的风险防患于未然，形成政府、企业、用户等多方共治的良性生态。特别是对于产业界而言——人工智能技术的模型算法研发者、服务提供者、重点领域使用者需建立安全指引和风险层级管理机制，以推动人工智能研发应用生态链各参与方实现风险共治。

CHAPTER 5
第五章 结语

关于 AI 及人的未来

未来的人类文明到底是怎样的？AI 以及科技在其中会扮演什么样的角色？这是生活在 21 世纪 20 年代的人们所不得不面对的问题。

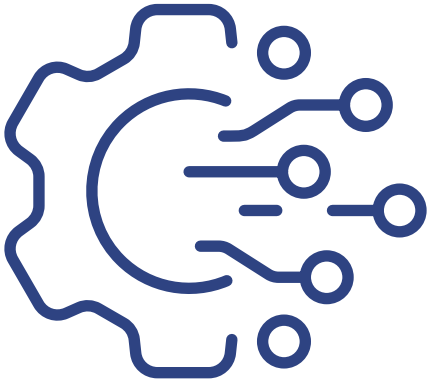
如果再过 10 年，人类世界会因 AI 而受到多大的影响？行业专家张亚勤给出了他的预言——2 到 5 年内，AI 生产的图片和视频将无法用人眼辨别真假；5 年后，人形机器人和无人车能通过图灵测试，且它们开车会比人开得更好；10 年内，AI 大模型将与生物世界、生命世界连在一块；20 年内，AI 在几乎所有领域都会比我们人类的能力更强[12]。

虽是一家之言，但其预示着 AI 的普及以及对未来生活产生的影响。与此同时，人们应该看到在这一进程中，许多劳动者也将经历由于技术革新造成的动荡之痛。另外，所有人也将面临 AI 的不慎使用或者蓄意使用所导致的风险和危害。

如何处理 AI 与人类的社会关系，从雪莱夫人的《弗兰肯斯坦》到阿西莫夫的“机器人三法则”，再到霍金的末世预言，都存在一个悖论：一方面，希望 AI 自身的演进得到尊重，享有创新发展乃至与人平等的权利；另一方面，出于“主人”意识，却又担心、警觉和恐惧人工智能生命对人类的威胁和伤害。

据此，面对 AI，人们需要一种新的科技发展观——从以关注 AI 技术为中心进阶到以人为中心，在 AI 设计、运行、应用、治理乃至更大的社会关系中都纳入人的因素——是否有利于人以及社会福祉，有助于塑造技术的可持续未来，从而更好地服务人类需求。归根结底，AI 的终局，是为人类服务。

同时，面对 AI，人类需要建立机制、法律以及价值观，协力控制和治理 AI 的未来发展方向，特别是在 AI 加速渗透、改造和改变人类思想、经济、政治、文化、教育等领域的背景下。AI 领域的从业者，无论是政策制定者、研究人员还是大型科技企业，都需要以更加负责任的方式以善治促“善智”，让 AI 回归以人为本的初心和目标。



12. 刘坚主编，《AI 时代的人类意见》，东方出版中心，2024 年 9 月

致谢

此次报告的撰写得到了多方的大力支持与帮助，我们有幸与行业内的高校、智库、企业等多方专业机构和资深行业专家进行了深入交流，同时也借鉴和引用了多家政府部分、企业、研究机构和媒体的公开数据和研究成果。

在此，特别鸣谢上海交通大学人工智能研究院作为联合出品机构参与报告调研撰写，特别是上海交通大学人工智能研究院常务副院长杨小康教授对报告悉心指导并提出了很多建议，让我们受益匪浅，在此我们深表感谢。另外，我们感谢清华大学法学院申卫星教授、清华大学经济管理学院李宁教授等专家接受我们的采访并给出了行业洞见。

同时，我们也感谢联想集团对报告的联合出品支持，联想集团的行业实践为报告提供了丰富的产业智慧。

此外，我们要特别感谢人民日报社、清华大学、鲲云科技、设序科技、百度、百川智能、北京亦庄等不同机构和主体，他们在百忙之中接受我们的调研，或者通过自身实践向公众、行业分享经验，让我们的报告更加丰富。

最后，我们还要感谢在报告前期调研采访、编辑、排版、校对、印刷等所有环节提供帮助的同事和朋友，他们的辛勤劳动使本报告得以顺利出版。

人本智能

人机共生时代的科技发展观

出品

财新智库 Caixin Insight

ESG30 中国ESG30人论坛

联合出品

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

人工智能研究院 Artificial Intelligence Institute

Lenovo 联想