分类号：H315.9　　学校代码：10363
密　级：公开　　学　　号：2221210112

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　《人工智能：机器学习如何塑造下一个十年》（第五章至第七章）汉译实践报告

论 文 作 者　杜嘉豪
校 内 导 师　李新国
校 外 导 师　谢亮亮
专业学位类别　翻译硕士
研究方向（领域）英语笔译　（工程科技翻译）

论文提交日期：2025 年 5 月 15 日

分类号：H315.9　　单位代码：10363
密　级：公开　　学　　号：2221210112

题　　目《人工智能：机器学习如何塑造下一个十年》（第五章至第七章）汉译实践报告

A Report on the E-C Translation Practice of *Artificial Intelligence: How Machine Learning Will Shape The Next Decade* (Chapters 5-7)

论文作者：杜嘉豪

校内导师：李新国

校外导师：谢亮亮

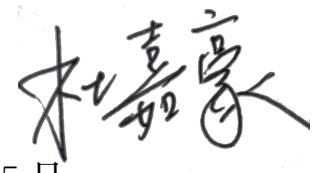
专业学位类别：翻译硕士

研究方向（领域）：英语笔译（工程科技翻译）

论文答辩日期：2025 年 5 月 14 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 

日期：2025 年 5 月 15 日

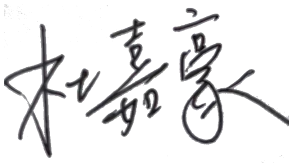
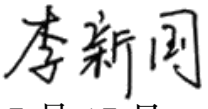
安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☑。

学位论文作者签名：  　　指导教师签名： 

日期：　2025 年 5 月 15 日　　日期：2025 年 5 月 15 日

致　谢

行文至此，掩卷长思。这篇论文的完成，既是学术旅程的驿站，亦是生命长河的涟漪。在此方寸之地，谨以素心诚表谢忱。

首先，我要特别感谢我的导师李新国教授。在翻译过程中，导师给予了我极大的指导和支持，不仅帮助我梳理翻译策略，提升语言表达能力，还在论文撰写的各个阶段给予了宝贵的建议。导师的严谨治学态度和渊博知识让我深受启发，也让我更加深刻地理解了翻译的本质与价值。

其次，我要感谢我的同门和我的舍友。在这段翻译实践的过程中，他们始终给予我理解与支持。当我遇到翻译难题时，他们耐心倾听，给予鼓励，使我在困惑时仍能保持信心与动力。正是这份支持，让我能够专注于翻译工作，顺利完成本报告。

同时，我要感谢原著作者马修·伯格斯（Matthew Burgess），其在人工智能领域的独到见解和严谨的学术研究为我提供了宝贵的学习机会。在翻译过程中，我不仅加深了对该领域的理解，也锻炼了自己的翻译能力。感谢这本书，它让我在翻译实践中体会到了跨语言沟通的挑战与乐趣。

此外，我也要感谢在翻译和论文写作过程中提供帮助的同学、图书馆工作人员以及各类学术资源平台。正是这些支持，使我能够查阅丰富的资料，不断优化译文，提升翻译质量。

本次翻译实践是我学术成长过程中的一次重要历练，也让我更加坚定了对翻译事业的热爱。当然，由于个人能力有限，论文中仍可能存在不足之处，恳请各位老师和学者批评指正，以便在今后的学习与实践中不断改进和提升。

摘　要

本次翻译实践材料选自马修·伯格斯（Matthew Burgess）所著的《人工智能：机器学习如何塑造下一个十年》（*Artificial Intelligence: How Machine Learning Will Shape the Next Decade*）。所选翻译材料从西方视角切入，探讨人工智能在工作自动化、武器化及伦理责任等领域的影响。译者选取第五章至第七章为翻译材料，结合翻译功能对等理论，分析人工智能类科技文本的翻译策略与方法。研究聚焦于原文的文本特征、翻译难点及解决方案，旨在探讨科技翻译中术语准确性、句式逻辑重构与语篇连贯性等问题，为同类文本的翻译提供参考。

在翻译过程中，译者遇到三大难点，即通用词汇的专业化转向、英语长句嵌套结构以及逻辑层面的表达方式。针对以上难点，译者通过文本类型分析、查阅平行文本、综合使用网络工具来解决上述三个问题。其中，对于词汇方面，采用“词义引申”与“词性转换”策略，通过网络资源、平行文本等渠道进行查证，确保翻译准确；对于长难句的结构，采用“信息重组”，“逻辑拆分”与“文化补充”的翻译方法，确保中国读者对源语句式结构的理解；对于逻辑的重构，采用“释义补充”与“逻辑连译”的方法，将概念通俗化、表达形象化，实现科普化的翻译。

通过本次翻译实践，译者认识到人工智能类文本的翻译不仅是语言转换，更是技术传播与伦理反思的双重过程；唯有将专业知识、理论工具与批判性思维相结合，才能实现科技文本的“意义对等”与“交际等效”。最后，作为一名翻译硕士研究生，在今后的翻译实践活动中，要发扬人工智能的“不断学习”的精神，以积极向上的心态不断提升翻译水平，迎接挑战。

关键词：《人工智能：机器学习如何塑造下一个十年》；人工智能；科技翻译；术语翻译；句式重构

Abstract

This translation project is based on Chapters 5-7 of *Artificial Intelligence: How Machine Learning Will Shape the Next Decade* by Matthew Burgess. The selected source text adopts a Western perspective to examine AI's impacts across domains including workplace automation, weaponization, and ethical accountability. Guided by the Functional Equivalence Theory, the translator analyzes translation strategies and methods for technical texts in AI. The study focuses on textual features, translation challenges, and solutions, aiming to address core issues in scientific translation such as terminological accuracy, syntactic logic restructuring, and discourse coherence, thereby providing references for similar translation projects.

Three major challenges emerged during translation: 1) context-specific conversion of generic vocabulary into technical terms; 2) deconstructing nested English sentence structures; and 3) maintaining logical coherence at the discourse level. To resolve these issues, the translator employed text-type analysis, consulted parallel texts, and utilized online tools. Specific strategies included: Lexical level: applying "semantic extension" and "conversion of parts of speech" verified through online resources and parallel texts to ensure terminological precision; Syntactic level: adopting "information reorganization" and "logical segmentation" to bridge Chinese readers' comprehension gaps regarding source-text sentence patterns; Logic level: implementing "explanatory supplementation" and "logical linkage reconstruction" to simplify concepts and visualize expressions for popularized science communication.

This practice reveals that translating AI-related texts constitutes not merely linguistic transference but a dual process of technological dissemination and ethical reflection. This project demonstrates that only by integrating professional knowledge, theoretical frameworks, and critical thinking can we achieve both "meaning equivalence" and "communicative equivalence" in technical translation. As an MTI graduate student, the translator advocates embracing AI's "continuous learning" ethos to proactively enhance translation competence and meet future challenges.

Keywords: *Artificial Intelligence: How Machine Learning Will Shape the Next Decade;* artificial intelligence; technical translation; terminology translation; syntactic restructuring

目　录

第一章 任务描述

本章主要包括三部分内容：任务来源、任务简介和实践意义。任务来源部分阐述了所选翻译材料的背景信息及选材依据；任务简介部分介绍了原著的核心内容，并对本次翻译实践的具体任务进行描述；实践意义部分则围绕人工智能相关知识的学习、翻译经验的积累以及人工智能对翻译的影响三部分展开探讨。

1.1 任务来源

本次翻译实践文本选自数据科学领域专家马修·伯格斯（Matthew Burgess）所著的《人工智能：机器学习如何塑造下一个十年》（*Artificial Intelligence：how machine learning will shape the next decade*），由企鹅兰登书屋（Penguin Random House）于 2021 年出版。全书共计 8 章 88 页。译者所选取翻译内容为第五至七章节，选定章节源文本字数共计 8644，译本字数共计 14364，符合翻译硕士的字数要求。经 Z-library 查询发现，目前尚未有中文译本。

此本人工智能英语科普著作为科技信息型文本，选择此人工智能类书籍作为翻译材料，主要原因如下：**一、人工智能让生活智慧化。**随着社会的不断发展，科技的不断进步，人工智能正深刻地改变着我们的生活和工作方式，例如 OpenAI、ChatGPT，以及国产 AI，豆包、Deepseek 等等。在日常生活中，AI 助力于智能家居、个性化推荐和虚拟助手等领域，提升了生活质量；在工作领域上，AI 通过自动化流程和数据分析，提高了生产效率和决策能力，推动零售、医疗和教育等行业的变革。**二、人工智能本身的双面性。**不可否认，AI 给人们的生活和工作方面带来了极大的便捷，但新现象、新事物的出现也会带来新的问题。在医疗领域，智能诊断系统能快速分析病理影像，但也可能出现因算法偏见而导致误诊；自动驾驶技术极大提升出行效率，但系统决策机制的不透明性使交通事故责任认定陷入困境。更值得警惕的是，深度伪造技术正冲击社会信任体系，人脸合成、语音模仿等 AI 应用已衍生出新型网络诈骗。三、**人工智能推动翻译研究。**语言机器人是构建无障碍社会的“新桥梁”，但也需要警惕，它也可能成为某些人、某些人群交往的“新沟壑”（李宇明, 2023: 34）。一方面 AI 通过分析海量语料，AI 不仅能识别汉语靠意思连接和英语靠语法连接的本质差异，还能将“塞翁失

马”这类成语译出文化内涵；另一方面 AI 倒逼翻译本质的重新思考，译者要把好文化关，用 AI 处理技术术语的同时，也要打磨文化隐喻。就像《流浪地球》英译时，AI 虽能精准翻译“行星发动机”，但“带着家园去流浪”的诗意表达仍需人工推敲。翻译本书正是要探索人类如何在 AI 时代守护语言的文化灵魂。

1.2 任务简介

作者简介。《人工智能：机器学习如何塑造下一个十年》是一本人工智能领域的科普读物，为人工智能和数据科学领域的专家马修·伯吉斯所著。马修·伯吉斯的研究主要集中在人工智能技术，以及对实际应用中的影响。伯吉斯在学术界和工业界的影响力十分广泛，常常参与跨学科的研究项目，致力于推动人工智能的创新和实际应用，也在机器学习、数据分析和自动化系统方面做出了杰出贡献。

内容梗概。全书共分为 8 章，系统梳理了人工智能的发展历程、现状、未来方向及潜在挑战。本书从人工智能的早期探索入手，阐述了关键技术突破如何推动其进入主流应用阶段。通过医疗、金融和交通等领域的应用案例，揭示了人工智能带来的广泛社会影响。然而，人工智能的发展也面临诸多局限，包括算法性别歧视、技术伦理偏见和数据隐私风险等问题。在影响分析方面，本书着重探讨了人工智能渗透日常工作的演进路径及其对就业市场的结构性冲击。

任务介绍。本次翻译实践材料选定第五至七章，重点探讨人工智能在工作、军事及社会等不同领域的影响及其发挥的作用，以帮助读者对人工智能的发展及应用形成初步认识。**第五章**《人工智能与现实工作》（AI and the World of Work）以奥卡多公司仓储自动化系统为研究样本，系统解析人工智能与机器学习技术的实际工作场景应用，重点剖析该技术在民生领域的应用潜力与实施挑战，论证人机协同模式的战略价值。**第六章**《人工智能的武器化》（Weaponising AI）聚焦中美在人工智能领域的战略博弈，通过地缘政治视角审视技术竞争态势，警示技术武器化潜藏的安全风险，主张通过国际监管协作机制建设，以实现技术应用的和平导向与全球安全治理。**第七章**《构建 AI 问责》（Making AI accountable）从技术社会学角度切入，强调研发主体须恪守伦理准则，同时呼吁行业开展系统性变革，着力破解算法系统中固化的种族歧视、性别偏见与社会不平等痼疾。

1.3 实践意义

推动 AI 知识普及。当前，人工智能技术已深入医疗影像诊断、金融风险预测、工业自动化及智能教育评估等具体应用场景。以机器学习、自然语言处理和计算机视觉为代表的核心技术不断迭代升级，进而产生 DeepSeek、ChatGPT 和 MidJourney 等创新工具。但值得注意的是，该领域仍面临用户隐私数据泄露、算法决策逻辑不透明和生成内容伦理争议等现实挑战。为推动人工智能向安全、高效、普惠的方向持续发展，世界各国通过强化立法监管、技术创新与风险管控等手段解决上述问题。本次翻译项目精选前沿资料，旨在帮助中文读者把握全球人工智能技术动态与发展趋势，构建知识共享的新范式。

提升翻译实践能力。科技信息类文本的翻译使译者的实践经验得到显著提升。在处理选定章节内容时，译者需要应对大量专业术语与复杂句式结构。初步分析文本时，译者发现原文呈现显著的嵌套式长难句特征，核心挑战集中于句法结构层面。译者首先查找该领域的相关中文和英文的平行文本，了解专业术语、表达规范，做好充足的译前准备。通过翻译复杂的技术文献，积累丰富的翻译经验，掌握了技术文档的翻译规范，为后续开展人机协作翻译研究积累了重要经验。

促进人机协作能力。当前弱人工智能时代，翻译主体介于机器与人类之间，机译促使翻译劳动分工与译者身份泛化，衬托出人译宝贵价值（胡健&范梓锐, 2023: 90）。人工智能技术已广泛应用于翻译实践中，机器翻译和译后编辑已成为行业趋势。翻译主体走过了从完全人译、低级机译、机助人译、人助机译到高级机译的历史进程（周怡珂&周领顺, 2023: 12）。通过本次翻译项目，译者不仅深入了解了人机协作的优势与局限，提升了与智能翻译工具协同工作的能力，还能更好地适应翻译行业的技术变革，培养在人工智能辅助下的翻译能力，这不仅有助于满足未来市场对复合型翻译人才的需求，也为社会带来了更高效、精准的翻译服务，推动了翻译行业向智能化、专业化方向发展。

第二章 译前准备

本章分三部分，依次为文本特点分析、平行文本查阅、翻译工具选定。通过分析选定翻译材料，将材料的文本特点归纳为内容科学性强、用词专业和逻辑严谨；选定的平行文本可以在一定程度上弥补因人工智能相关专业知识欠缺、语言风格不同等方面限制而造成的翻译问题；翻译工具的使用上，采用网络与纸质相结合的方式确定词义，进而提升译文用语专业性。

2.1 文本特点分析

科学性强。选定翻译材料，引用了大量人工智能发展过程中的事例。全书参考文献共计 98 篇，包括医学、民生、政治等多领域的研究项目和成果，以及相关领域权威人士的观点。文章逻辑结构清晰，语言风格较为正式，句法相对复杂，长句和被动语态在文中的出现频次高，经 python 检验选定章节共有 8644 个字组成，共 344 句，平均句长为 24.93。

用词专业。人工智能科普英语文本的词汇通常包含以下三类：技术术语、抽象概念词汇和术语缩略词。翻译材料涵盖大量专业词汇，其显著特征在于通用词汇的专业化转向——日常语汇在特定学术语境中被赋予精确的技术语义。为确保术语转换的精确性，译者采用以下三种方法解决此问题：查询权威词典、借助网络资源进行查询；结合文本特点，根据上下文进行语境化语义推演；通过平行文本对比分析，完成术语使用的范式校准。

逻辑严谨。源文本呈现显著的学术论证特征：文章结构层面，通过系统分析与严谨推理展开，确保论点紧密衔接、层层递进，结合数据、案例和实验结果，以对比分析强化论证的有效性；语篇层面，通过运用精准的指代、逻辑连接词及前后呼应的结构，使内容连贯、推理清晰，避免逻辑跳跃。对于译者而言，如何精准再现原文的逻辑结构，确保译文既层次清晰又衔接顺畅，是翻译过程中的关键挑战。

2.2 平行文本查阅

为更好了解材料内容以及把握作者写作风格，译者在翻译之前对源文本的背景做了调查，通过在线阅读相关文章和书籍，了解人工智能的发展历程，以及该技术在各个领域的应用情况。最终选取 2021 年由 John Murray 出版的 *The Age of AI* 一书作为译者的平行文本。选择该书作为平行文本有以下三点原因：内容和主题与选定材料高度契合；体裁和写作风格相似；句式结构都以长句为主。*The Age of AI* 的作者为美国第 56 任国务卿亨利·A·基辛格 (Henry Alfred Kissinger)、谷歌首席执行官兼董事长埃里克·施密特（Eric Schmidt）、麻省理工苏世民计算学院的首任院长丹尼尔·赫滕洛彻（Daniel Huttenlocher）。选定平行文本共分七章，探讨了人工智能技术的崛起及其对人类社会、经济、政治和文化的影响。作者们结合各自的专业背景，从技术、政策和伦理多个角度，详细分析了人工智能在各个领域中的应用及其带来的变革。书中不仅对人工智能在医疗、教育、交通等行业的应用加以阐述，而且对就业、经济结构以及由此产生的伦理和治理问题的影响也进行了深入的探讨。

2.3 翻译工具选定

在翻译工具的选取方面，译者采取了线上翻译工具和纸质翻译工具相结合的方法，线上借助的工具包括：必应词典、CNKI 百科等；纸质的翻译工具包括：《牛津高阶英汉双解词典》（第十版）和《柯林斯高阶英汉双解词典》（第八版）。

本研究依托百度 AI 论文精翻的文心大模型及 SDL MultiTerm 等工具，针对人工智能专业领域实现了术语的精准提取与标准化管理，并通过维基百科、PubMed 等多源数据库交叉验证术语译法的准确性；平行文本的应用则采用“关键词替换法”与多源文本整合策略，从学术文献与行业资料中筛选双语对照文本，分析其术语规范及句式特征，同时借助 Trados 构建可检索的平行语料库以提升译文一致性。当上述两种方法无法解决问题时，译者通常需要依靠专家或导师的指导。通过向他们请教，译者能够获得更深入的领域知识，并在此基础上，重新审视翻译中的专业术语与表达方式。结合专家的建议，译者会更有信心地做出选择，确保最终的翻译既忠实于原文，又符合学术或行业的规范要求。

第三章 案例分析

本章节分为三部分，从词汇、句法和逻辑衔接三个角度分析。译者将重难点概括为词汇难对等、长难句结构复杂、衔接手段丰富。为解决这些问题，译者从英汉对比的基本知识着手，提出针对各难点的具体方法。借助网络资源和平行文本对专有名词进行查证；采取重新排序法、语气对比转换、意义补充再现，对复杂长句拆分重组；通过释义补充、拆分小句，对逻辑结构进行重构。

3.1 词义的准确传译

由于源文本是纯英语文本，译者在翻译过程中发现，若要准确传达原文作者的意图，必须深入理解每个单词的深层含义。叶子南（2008: 74）认为，“翻译中词类转换的翻译技巧能够提高译文的可读性”。本研究将从词义引申与词类转换两个维度，对翻译过程中词汇层面的处理策略进行系统探讨。

3.1.1 词义的引申

本次翻译实践报告中常出现通用词汇、日常语汇，但在特定科技语境中却存在特殊的专业技术语义。胡开宝（2011: 352）认为，“尽管双语词典提供了源语词汇的目的语对应词，但由于双语词典所提供的对应词数量有限，如果按照词典上的解释生搬硬套逐字直译，会让人觉得译文晦涩难懂，无法准确表达出原文含义”。于是译者根据等效原则运用引申法，根据上下文的语境及逻辑关系，从该词语的基本意思出发，选择出恰当的汉语词汇来表达。

例 1. In 2015 the graduate student of MIT’s Media Lab was sitting at her laptop, trying to get the generic face-tracking software she was using to work. The problem was that it wouldn’t. It was supposed to project digital artwork onto her face, but it couldn’t ‘see’ her.

初译： 2015 年，作为麻省理工学院媒体实验室的研究生，她在笔记本电脑前尝试使用面部追踪软件进行操作，然而软件并不起作用。本应在她脸上投影数字艺术的系统却始终无法“看见”她的面容。

改译：2015 年，麻省理工媒体实验室的一位研究生发现，通用型面部追踪软件存在系统漏洞。程序本应实时捕捉面部特征并投射虚拟影像，却因肤色识别算法缺陷导致运行中断。

分析：例 1 的翻译遵循了 Newmark 的意义对等原则（Newmark 2001: 47）。若只是将 digital artwork 译作“投影数字艺术的系统”，就使得译文的流畅度大打折扣，同时读者在阅读时也会产生困难，使译文晦涩难懂。因此，译者通过互联网查阅相关的人工智能技术，以及与计算机相关的人脸识别技术之后，了解到 digital artwork 其实指的是通过软件生成的艺术图像或特效，比如 AR 滤镜、虚拟化妆等。在改译中，译者将其译作“实时捕捉面部特征并投射虚拟影像”，既保留术语原意，又点明其艺术特效的用途，从而避免读者误以为仅仅是静态画面投射。同时，译者根据语境背景了解到此次试验是在做人脸识别系统的偏见研究，Buolamwini 发现当深肤色人群使用面部识别时，系统经常无法检测到人脸通过。因此译者将 couldn’t see her 进一步解释，译作“肤色识别算法缺陷导致运行中断”增强了句子的清晰度和易懂性，也符合“意义对等”的翻译原则，即在传递信息时，不仅保留原文的细节，还确保读者对其深层含义有清晰的认识。

例 2. Twenty-one government agencies pledged to be transparent about how they use algorithms in decision-making. They undertook to give ‘plain English’ explanations of how systems worked, and promised to make efforts to manage any biases that might be informing them.

初译：21 个政府机构承诺在算法决策中保持透明，并用“简明英语”解释系统的工作原理，努力管理可能存在的偏见。

改译：21 家政府机构公开承诺，将就其决策算法建立透明机制：通过通俗易懂的非技术语言，披露系统运作原理，并主动管控可能潜藏于代码深处的结构性偏见。

分析：例 2 中，plain English 直译为“简明英语”，但在公共政策语境中，其核心是要求语言去专业化、去术语化，确保信息能被普通公众无障碍理解。汉英翻译应以英文写作原则为指导，例如使用简明英语，避免冗余信息，多用主动语态，简化句子结构（李长栓, 2012: 46）。译者通过查询相关参考文献以及政府发表的各类讲话之后，对于这一词的翻译有了以下两点新的见解，一是为避免直

译可能引发的歧义（如误认为仅指英语语种）；二是政府机构承诺使用 plain English 的根本目的是提升透明度与公众信任，那么在译文中也需要体现这一政策的背景。因此，译者直接点明“用大众能理解的语言”这一核心意图，通过补充“非技术性”降低读者的理解门槛，将其译为“通俗易懂的非技术语言”。“交际翻译法”强调译文应考虑目标语言读者的文化背景和接受习惯，确保信息能够有效传达并产生预期的效果。出于对目标语言文化和读者接受能力的考虑，尤其是在理解复杂的技术内容时，使用“通俗易懂的非技术语言”这一补充，既清晰地解释了 plain English 所指的含义，也确保了译文能产生与原文相同的效果，避免了语言上的误解或混淆。

例 3. Algorithms must not be a black box, and there must be clear rules if something goes wrong.

初译：算法不能是黑盒子，如果出现问题，必须有明确的规则。

改译：算法系统必须禁止暗箱操作，当决策链条出现偏差时，须有可溯源的权责框架作为系统性追责机制。

分析：在翻译例 3 时，black box 原译作“黑盒子”虽保留了 black box 的字面意象，但未能充分体现其在算法伦理与透明度要求中的深层内涵。根据控制论中的定义，black box 特指“仅能通过输入和输出观察、无法解析内部机制的系统”，通俗来说就是内部运作不透明的系统。这里需要保持术语的一致性，同时可能需要强调技术的不透明性。因此，将 black box 引申为“暗箱操作”，既保留原词的技术隐喻特征，也保留其不可追溯性与不透明性的负面意涵。后句 clear rules 的初译“明确的规则”过于浅显，需结合技术治理的具体诉求进一步具象化。根据 IEEE《可解释人工智能标准》（2022）提出的“技术文档可追溯性”原则，以及“模型可解释性”要求，将其译为“可溯源的权责框架”，通过“责任追溯”凸显算法失效时的问责路径，以及“可验证”呼应技术文档需满足第三方审查的合规性要求。

3.1.2 词性的转换

在翻译过程中常出现词性不对等问题，汉语的词性和英语的词性出现不相匹配的情况。两种语言在表达习惯和背景文化上存在差别，一味按照原文词性进行

对等翻译的话，容易导致译文晦涩难懂，也不符合汉语的表达习惯、表达方式。因此，为了使译文更加流畅，笔者在一些词语的词性方面作出调整。

例 4. True, there was a prominent backlash against Apple's credit card launch in the US when its algorithms appeared to show bias by offering different credit limits to men and women, but such a reaction is rare.

初译：诚然，当苹果的算法似乎通过为男性和女性提供不同的信用额度而显示出偏见时，苹果在美国推出的信用卡引起了强烈的不满，但这种反应很少见。

改译：诚然，苹果信用卡在美国上线时，曾因算法存在性别歧视倾向——向男女用户提供差异化授信额度，而遭遇强烈抗议，但此类引发公众警觉的算法纠纷，在系统性歧视普遍存在的当下仍属罕见个案。

分析：例 4 中，翻译不可过分追求词性对等，“需要通过词性的有效转换将英文翻译工作达到充分落实，做到信息的准确表达”（郝嘉欣, 2019: 109）。翻译时搜索原文语境背景得知，在该事件中，苹果合作银行 Goldman Sachs 的算法将“家庭共同账户还款记录”与性别变量隐性关联，导致女性申请人平均授信额度低于男性 34%。虽然最终达成 9000 万美元和解，但 95%的算法歧视案件仍通过仲裁保密处理，印证“罕见个案”背后的制度性遮蔽机制。因此，译者在结合背景后，采用词性转换法，将原句的语法结构重构以契合中文表达范式。名词性短语 prominent backlash 动态化为动词结构“遭遇强烈抗议”，不仅消解了英文静态表达的抽象性，更通过动作化呈现凸显公众反应的即时性与冲突张力；形容词 different 向动词“差异化”的转变，则精准捕捉了算法决策中动态生成歧视性结果的本质特征，使授信额度差异从静态比较升维为系统化运作过程。另外，将动词短语 show bias 转换为名词化表达“性别歧视倾向”，通过语法隐喻将偶发技术故障重塑为结构性症候，符合“技术中性神话解构”的学术话语。

例 5. In one rare pro-active instance in August 2020, a UK police ethics board stopped the development and deployment of flawed AI that was supposed to predict if someone would commit violent crime.

初译：在 2020 年 8 月的一个罕见的积极事件中，英国警察道德委员会阻止了有缺陷的人工智能的开发和部署，该人工智能本应预测某人是否会犯下暴力犯罪。

改译：2020 年 8 月，英国警务伦理委员会叫停某 AI 犯罪预测系统的开发部署。该系统设计用于暴力犯罪风险评估，因算法缺陷导致决策偏差而被终止。

分析：中文善于用动词，属于动态性语言，而英文善于用名词，属于静态性语言。例 5 中，原文被解构重组为符合中文治理的批判性表述，原文形容词 pro-active 被动态化为动词“中止”，突破英语静态描述的局限。动词 stopped 被创造性转换为“叫停”，通过法律术语的移植赋予执行动作更强的行政规制色彩，使伦理审查从技术建议升格为具有约束力的制度干预。名词 development and deployment 在保持动名词本质的同时，通过“开发部署”的四字格浓缩，形成符合中文政务文本特征的紧凑表达。另外，动词 predict 被转换为名词“风险评估”，通过技术功能描述的静态化处理，规避中文科技文本中可能引发的统计学歧义，同时对应《欧盟人工智能法案》对个体风险评估系统的监管分类。原句条件状语 if someone...被压缩为“犯罪预测系统”的限定成分，将静态的技术缺陷 flawed 动态化为持续释放负面效应的“系统性歧视”，这种从形容词到动词链的词性转换，建构起“算法开发-社会危害”的因果链条，使译文在信息转换中完成了从技术叙事到伦理批判的话语跃升。

3.2 句式的结构处理

英语复杂句中的词、词组、小句、句子由小到大，层层嵌套，呈现出严格的“递归性”（王福祥&徐庆利, 2018: 99）。英语复杂句的层级结构以主谓核心为基点，通过嵌套形成多重主从关系。在翻译此类结构复杂句式时，采用信息重组、语气对比和意义补充三种方法降低处理难度。

3.2.1 信息重组优化表达

“大凡译文之生硬、拗口甚至晦涩多半都是因没能译好定语从句所致”（曹明伦, 2001: 23）。英语多存在从句嵌套、介词连词衔接等形合结构，而汉语作为意合语言更注重语义连贯，翻译时需对原文信息进行逻辑解构。信息重组前要进行句子拆分，以分译法作为信息重组的基础，通过调整语序、转换词性及重置修饰成分实现信息重组，确保译文既忠实于源语语义，又符合汉语简洁流畅的表达规范，最终实现“信、达、雅”的翻译标准。

例 6. Simply adding strips of black and white tape to red stop signs (in some cases spelling out the words ‘love’ and ‘hate’) – as a group of nine researchers did in 2018 in a virtual simulation – can be sufficient to fool a self-driving car into assuming that the speed limit is 45 miles per hour.

初译：简单地在红色停车标志上加上黑白胶带（在某些情况下会拼出“爱”和“恨”两个词）——正如 2018 年一组九名研究人员在虚拟模拟中所做的那样——就足以欺骗自动驾驶汽车，使其认为限速为每小时 45 英里。

改译：2018 年，由九位学者组成的实验团队在数字仿真环境中揭示：仅需在红色停车标志表面施加黑白条纹干扰（部分案例中甚至嵌入情感暗示文字），就能诱导自动驾驶系统将“停车禁令”误判为“时速 45 英里”的通行许可。

分析：长难句解构是科技文本实现功能对等的关键操作，其核心在于通过句法重构完成信息的跨语言迁移（Reiss, 2004: 93）。译者需突破表层结构束缚，将修饰成分进行结构重组，例如将 as a group of nine researchers did in 2018 in a virtual simulation 在此句中作插入语，起承接上下文、递进的作用。每种语言都有各种方法来强调重要的信息，在正常的主谓宾句型中，语义会出现在补语或最后一个单词上，但如果句子中的其他成分被放在开头，那么这个成分就会成为语义的一部分，那就需要在翻译时表达出这种情感过程。在翻译时，如果不改变这句话的顺序，按照初译进行翻译的话，就会割裂句子，从而影响句意的顺畅表达。因此，在翻译时，译者将插入语置于句子开头。如此一来，不仅可以进一步强调句子的递进关系，而且句子主次关系更为清晰，避免了英语长句翻译中因保留原句结构而导致的生硬和不流畅的问题。另外对于原文 in some cases spelling out the words ‘love’ and ‘hate’，译者在查阅相关资料后，了解到实验原型是在揭露基于 CNN 的计算机视觉系统在符号语义理解上的脆弱性，因此译者在翻译此句时，突破直译 love/hate 的表层，揭示攻击者利用自然语言处理模块与视觉识别系统的认知裂隙，将其译作“嵌入情感暗示文字”。

例 7. At one clinic half the patients earmarked for the study declined to be involved after finding out that although the results would be virtually instantaneous (previously some results had taken ten weeks to arrive), a positive diagnosis would see them being referred to a hospital an hour’s drive away.

初译：在某诊所中，半数入选研究的患者在获悉以下情况后拒绝参与：尽管检测结果近乎即时可得（此前某些结果需要十周才能得出），但一旦确诊阳性，患者需转诊至车程一小时的医院接受治疗。

改译：某临床试验点，半数入选研究的患者在知情后选择退出。尽管检测反馈周期从十周优化至实时，但阳性确诊者将被转诊至距诊所一小时车程的定点医院。

分析：汉语句法成分间依存距离短于英语，主要是因为汉语句段铺排更多依赖时序，而英语多使用后置修饰成分（秦洪武&周霞, 2019: 418）。例 7 句式复杂，嵌套从句和修饰语较多，若直接翻译为中文长句，可能显得过于冗长、难以阅读。译文通过将长句拆分为若干短句，使信息层次清晰分明，尤其在处理因果关系和时间顺序时，读者能够逐步消化信息，不会因信息密集而产生阅读障碍。针对英文中原因后置的复合句式，译者将“选择退出”的结果前置，用破折号引出具体原因，避免了直译可能产生的冗长拖沓问题。在处理括号内的背景信息时，将原本置于句尾的“十周等待”调整到检测结果之后作为补充说明，既保留了原文的对比意图，又确保了句子重心的平衡。对于关键术语 positive diagnosis 表述为“阳性确诊者”，通过补充检测结果属性消除了可能存在的歧义。在翻译时，译者既保持了医学文本的专业性，又清晰呈现了患者拒绝参与的内在矛盾——检测效率提升与转诊不便之间的权衡。整体来看，译者在深入理解原文信息的基础上，通过合理的语序重组、语义显化和逻辑衔接，实现了学术文本翻译中准确性与可读性的统一，体现出对中英语言思维差异的精准把握。

例 8. What he therefore did was to create a system in which AI works with human lawyers rather than instead of them. The company’s system has been trained on historical contracts – both those in the public domain and documents provided by clients – and taught to look for particular elements.

初译：因此，他转而构建人机协同的法律审查系统。该公司的 AI 系统经过训练，能够识别历史合同（包括公共领域的合同和客户提供的文件），寻找特定元素。

改译：因此，他构建的并非替代性方案，而是打造人机协作的智能合约审查体系：该平台通过深度学习历史判例库（既涵盖公开司法协议，也整合客户专有

文档），训练算法能够精准识别关键法律要素。

分析：科普翻译需将专业术语转化为大众能理解的通俗语言，但不可牺牲科学严谨性（郭建中, 2004: 79）。例 8 中，原文包含主从句关系和插入语，句子层次较多、结构复杂，译者使用分译法对分句化繁为简，使译文在逻辑上更为清晰，同时符合汉语表达习惯。具体表现为三个层面的重组策略：其一，概念整合化处理，将原文中分散的 AI works with human lawyers rather than instead of them 复杂关系从句，提炼为“人机协作”四字核心概念，既保留“合作而非替代”的核心理念，又符合中文法律文本简洁性要求；其二，逻辑链条显性化，将隐含因果关系的 therefore did 转化为结论性表述“因此……而是构建”，强化行为动机的衔接，同时将被动语态 has been trained 转译为“经过训练，能够识别”的主动句式，突出 AI 系统的功能实现路径；其三，专业信息结构化，将破折号内的补充说明 both those...处理为括号内平行信息，维持法律文本严谨性的同时，通过标点符号重构形成“系统功能+训练数据”的模块化呈现，使专业读者能快速抓取关键要素。

例 9. The journeys then taken by Ocado’s delivery trucks are optimised by a neural network that makes more than 14 million last-mile routing calculations per second, and adjusts delivery routes each time a customer places a new order or adds extra items to their shopping lists.

初译：在 Ocado 的配送卡车随后执行的行程中，其路线经过神经网络动态优化。该网络每秒进行超过 1400 万次末端路径规划计算，并在每次客户下达新订单或在购物清单中添加额外商品时，实时调整配送路线。

改译：Ocado 的智慧物流系统通过深度神经网络实现动态路径规划：每秒执行逾 1400 万次末端配送路径计算，当客户新增订单或实时修改购物清单时，系统能实时调整全局配送路线。

分析：英汉两种语言在表达顺序上有着一定的差异，英语通常采用前置性陈述，先果后因，而汉语相反，通常是先因后果，层层推进，最后进行综合，点出主题（余高峰, 2012: 03）。例 9 在结构上层层递进，信息密度较大，分译法可以将这些信息分拆后独立呈现，便于读者逐层理解每个要点，符合汉语表达的逻辑特点。若直接翻译成一个句子，容易使句子过长、晦涩，不符合汉语简洁明快的

表达习惯。通过分译，译文首先指出“智慧物流系统通过深度神经网络实现动态路径规划”，简洁地引出主旨，然后才进入关于神经网络的具体功能描述。分拆后的第二句不仅清晰地传达了“每秒执行逾 1400 万次末端配送路径计算”这一信息，还合理连接了“当客户新增订单或实时修改购物清单时，系统能实时调整全局配送路线”这一细节。分译法使得整个句子的逻辑关系更为分明，信息流更具条理，有助于中文读者逐步吸收理解。

3.2.2 语气对比灵活转换

跨文化翻译中，如何准确传递原文语气以避免误解？这是译者在处理情感表达、礼貌程度等语气差异时面临的核心难题。伦敦学派弗斯指出（刘润清, 2002: 93），语境不仅包括文字本身，还涉及文化场景等非语言因素。这意味着译者既要分析原文的社交场景，也要预判译文读者的接受习惯。只有兼顾语言形式与文化逻辑的双重适配，才能真正实现翻译的“等效沟通”。

例 10. It’s also possible, though, that AI will control lorries and delivery trucks on major highways where conditions are relatively predictable, and that human drivers will take over the driving as they reach the outskirts of the villages, towns and cities set as their destinations.

初译：不过，也有可能是这样一种情况，即人工智能将在条件相对可预测的主干高速路上控制卡车和送货车，而当到达目的地的村庄、城镇或是城市的郊区时，人类司机将接管驾驶。

改译：不过，人工智能或将在路况相对可控的高速公路接管物流运输车辆，待车辆驶抵目标城镇、乡村或城市郊区时，则会由人工接管。

分析：例 10 中，保持原文 possible 蕴含的虚拟色彩的同时，又兼顾汉语假设性表达的动态对等。原句 will control 与 will take over 在语法表层呈现为将来时态，这种“现在将来时表假设”的语法特征恰是英语虚拟语气在科技文本中的表达方式（Vermeer, 2013: 47）。在初译时译者采用“即……而当……时”的解说式结构，虽完整传递了语义信息，但将原文通过时态构建的虚拟表达转化为确定性表达。因此在后续修改中，将其调整为：“在路况相对可控的高速公路接管物流运输车辆，待车辆驶抵目标城镇、乡村或城市郊区时，则会由人工接管。”

通过“待……时”的句式重构，既保留了时间条件关系，又通过文言残留的“待”字强化了假设实现的阶段性特征（郑声滔, 2009: 16）。在虚拟语气的处理上，将 relatively predictable 译为“相对可控”而非字面的“可预测”，使科技假设的表述更符合中文语境。

例 11. In response Google CEO Sundar Pichai apologised for the controversy and said the company needed to ‘accept responsibility for the fact that a prominent black, female leader with immense talent left Google unhappily’.

初译：对此，谷歌 CEO 桑达尔·皮查伊（Sundar Pichai）为这场争议道歉，并表示公司需要“为一位才华横溢的黑人女性领导人不满离职负起责任”。

改译：对此，谷歌 CEO 桑达尔·皮查伊（Sundar Pichai）公开致歉，坦言公司必须正视，顶尖非裔女性高管愤而离职一事。

分析：语篇中交际者的身份不同，语篇中的语言功能也就不一样（张美芳, 1999: 13）例 11 中，源文本是一则包含祈使语气的企业声明，在翻译时既要再现原文话语主体的责任，又要契合中文企业危机公关的修辞。原句 needed to accept responsibility 是 CEO 对公司行为的指令，但其实也是向公众传达企业的自省态度。初译时将 needed to 处理为“需要”，虽准确对应词汇含义，却弱化了企业高管在危机公关中必须展现的强制性责任认定力度。因此将其调整为：“必须正视顶尖非裔女性高管愤而离职一事”，通过情态动词的转换，匹配了中文中企业道歉时应有的决断性语气。另外，原文 unhappily 作为状语修饰 left，初译作“不满离职”虽简洁，却将 unhappily 仅仅表现为内在心理的不满，导致公众对企业失责的认知模糊。因此转换为“愤而离职”，使情感更加显化。此外，对于 prominent black、female leader with immense talent 的多重身份限定，在中文企业声明翻译中需进行价值排序重构，将 prominent 译为“顶尖”前置，继而用“非裔女性”承接身份政治正确性，相较于直译更符合中文高层人事表述的语体规范。

例 12. Until a robot can pick and pack as many items in an hour as a human being – Harvey says this is around 600–700 items – it is unlikely to be widely adopted: the impact on productivity would damage service and profits.

初译：除非机器人可以像人类一样每小时挑选和包装 600-700 件商品，否则不可能会被广泛使用，因为生产力低下会对企业的服务和利润造成损失。

改译：哈维（Alex Harvey）指出，只有当机器人的分拣效能达到人类水平（约 600-700 件）时，该技术才可能被大规模应用——否则其对生产效率的负面影响将损害服务质量与盈利空间。

分析：例 12 在处理虚拟语气的对比转换时，遵循“功能对等”原则（Nida&Taber, 2004: 38），既要准确传递原文逻辑关系，又要符合目标语言的表达习惯。原句本质上构建了条件结果的虚拟关系，但中文不存在严格对应的虚拟语气形态，需要通过“只有……才……”等假设性结构来体现。在译文初译时使用的“除非……否则……”虽正确传达了条件关系，但将原文隐含的“能力未达标的现状”转换为的“生产力低下”，存在过度解读的问题。因此，译者将其调整为：“只有当机器人每小时拣选与打包量达到人类水平（约 600-700 件）时，该技术才可能被大规模采用”这种处理既保留了原文 until 引导的虚拟结构，又通过“只有”替代“才”来强化绝对化表述，更符合科技文本的客观性。

3.2.3 文化补充精准再现

文化补充精准再现通过句法结构调整，在目标语中补全源语文化特有的隐含信息，使抽象文化概念显化，本质上是以句法“补丁”实现文化的对等，将原文中依赖文化背景的指称转化为目标语读者可理解的内容。

例 13. One such is Tess Posner’s AI4ALL, the US non-profit set-up that is seeking to get more young women and BAME[1] people interested in AI. Among its initiatives are summer camps for high-school students. We hear a lot of things like, “I’m the only person at my school who looks like me who’s interested in this,” or “I never thought I can go into this field because I don’t see anyone like me in the field,” says Posner.

初译：苔丝·波斯纳的人工智能夏校就是这样一个美国非营利组织，旨在让更多年轻女性和少数族裔（BAME）对人工智能产生兴趣。其项目之一是为高中生举办的夏令营。波斯纳表示：我们经常听到一些学生说，“我是学校里唯一对这感兴趣的人”或“我从未想过我能进入这个领域，因为我在这个领域看不到和我一样的少数族裔”。

[1] BAME 为英国社会学术语，全称 Black, Asian and Minority Ethnic，特指非白人移民后裔群体，常用于社会分层研究与平权政策语境。此处采用扩展译法以还原其文化政治意涵。

改译：苔丝·波斯纳创立的“人工智能普及联盟（AI4ALL）”便是典型代表，这个美国非营利教育机构致力于激发更多年轻女性及黑人、亚裔与其他少数族裔群体（BAME）对人工智能的兴趣，包括为高中生开设人工智能主题暑期学术营。波斯纳说：“我们在高中生夏令营中反复听到这样的心声——‘全校找不到第二个与我外貌背景相似却热衷此领域的人’，或‘因未见同肤色的先行者，从未敢奢望踏入此领域’。”

分析：在例 13 的翻译中，译者采用意义补充精准再现的策略，兼顾文化指称对等与社会语境显化。将 BAME 译为“少数族裔”虽符合字面指称，但未能体现该缩写的文化特殊性——BAME（Black, Asian and Minority Ethnic）是英国社会对非白人移民后裔的特定表述，具有鲜明的社会分层与平权运动背景。根据翻译研究中“功能对等理论”的要求，应补充文化限定信息，将其译为“黑人、亚裔及其他少数族裔群体”。此外，AI4ALL 作为专有组织名，处理为“人工智能夏校”存在偏差，根据斯坦福大学官网信息，AI4ALL 实为覆盖学术培训、职业发展等多维度的系统性组织，因此将其音译全称“人工智能普及联盟（AI4ALL）”并补充说明其“非营利性教育机构”属性。原句 I’m the only person at my school who looks like me who’s interested in this 中的 looks like me 具有双重语义：既指种族相似性，又隐含少数群体代表缺失。因此，通过增译“外貌背景”以实现源文本内容的精准表达。

例 14. Researchers at MIT Sloan and Boston argue that those companies poised to benefit most from AI are those who use it to augment and shape traditional processes rather than replace them.

初译：麻省理工学院斯隆管理学院和波士顿的研究人员认为，那些最有可能从 AI 中获益的公司是那些利用 AI 来增强和重塑传统流程，而不是简单地替换掉传统的方式。

改译：麻省理工斯隆管理学院与波士顿大学的联合研究指出，最能收获人工智能红利的企业，是那些通过智能技术实施流程重塑与人机协同，而非简单粗暴的岗位替代。

分析：在翻译例 14 时，采用意义补充精准再现的策略，同时兼顾术语专业化与技术内涵显化。原译文中“波士顿的研究人员”存在指代模糊问题，根据

MIT Sloan 官网信息，该校是麻省理工学院（MIT）下属管理学院，位于波士顿剑桥市，其研究团队常以 MIT Sloan researchers 署名，而 Boston 在此语境中更可能指代波士顿区域的合作研究机构或项目，因此修订为“麻省理工斯隆管理学院（MIT Sloan）与波士顿大学”。对于 augment and shape traditional processes rather than replace them, 原译“增强和重塑传统流程，而不是简单地替换掉传统的方式”虽传递了表层语义，但未能体现 AI 技术赋能传统流程的双向交互性。结合译者在网络上搜索到的 MIT Sloan 专家提出的 Small Transformations（小规模渐进式转型）理论，AI 的价值在于“通过补充性整合实现流程增效，通过适应性改造推动系统演进”，因此调整为“通过智能技术实施流程重塑与人机协同，而非简单粗暴的岗位替代”。

例 15. Perhaps most crucially, it campaigns for a shake-up in the way people are recruited, suggesting that initiatives should be set up to reach out to non-elite universities, and that better ways should be found to allow contractors, temporary staff and freelancers to become full-time employees. Hiring practices and conditions should also be made transparent, so that everyone knows how individuals are hired, paid and promoted. Responsibility for all this begins with senior management.

初译：最关键的是，报告提倡对招聘方式进行大幅调整，建议设立项目，主动联系非精英大学，并探索更好的方法使合同工、临时员工和自由职业者能够成为全职员工。报告指出，招聘流程和条件应透明化，让所有人了解个人如何被招聘、支付和晋升。这一责任从高层管理人员开始。

改译：或许最关键的是，这场变革的核心诉求在于推动企业招聘机制革新——建议建制化非顶尖学府人才输送专项计划，破除劳务派遣人员、临时雇员及自由职业者的转正壁垒；同时强制实行招聘流程透明化工程，确保薪酬、晋升机制全程可溯。此类系统性变革的权责溯源，必须始于高层管理者主导的自上而下的制度重构。

分析：对于例 15，原译文中“非精英大学”虽符合字面对应，但未能体现英美高等教育体系下“精英大学”与“非精英”院校的结构性差异，因此，译者补充文化限定信息，译为“非顶尖高校（非‘常春藤联盟’层级院校）”，并通过增译“定向拓展”凸显企业主动打破学历壁垒的意图。对于 contractors,

temporary staff and freelancers 的译法，原译“合同工、临时员工和自由职业者”虽准确，但未体现劳动力市场分层特征。因此调整为“劳务派遣人员、临时雇员和自由职业者”，其中“劳务派遣”对应中国《劳动合同法》术语，而“自由职业者”符合国家统计局对新型就业形态的定义，实现法律概念的跨文化对等。修订后的译文修正了初译的语义偏差，使目标读者能够跨越文化鸿沟，精准把握原文倡导的雇佣公平性改革内核，实现了奈达 (Nida&Taber, 2004: 84) 所主张的“从语义到文体再现源语信息”的翻译目标。

3.3 译文的逻辑重构

基于英汉语言衔接机制差异的翻译策略，通过解构原文的逻辑框架，结合汉语意合特征重构汉语逻辑。英语依赖语法化衔接手段，文章逻辑层层递进；汉语则倾向以语义关联或话题延续实现内在连贯。译者通过释义补充和信息层级拆分，重建符合汉语认知习惯的逻辑。通过语序调整、语义强化或冗余消解，实现译文逻辑的“形隐而神聚”，最终达成跨语言思维模式的深度适配。

3.3.1 释义补充

释义补充是针对跨语言翻译中，因文化背景缺失或专业领域知识盲区，而做出的一种翻译策略。通过在线查阅背景信息了解科技信息类文本的写作特点，同时补充文化限定信息，展现文本中隐藏的文化信息，进而使翻译译文更加通顺流畅富有逻辑性，消除源语与目标语之间的认知偏差。

例 16. Now it has now turned its attention to codifying laws on the use and control of AI, which, if and when issued, would constitute the first transnational regulation of AI (the guidelines were scheduled for publication in 2020, but then delayed to an unspecified date in 2021).

初译：现在，欧盟已将目光转向制定人工智能的使用和控制的法律上来，这些法律一旦发布，将构成首个跨国 AI 法规（这些指南原定于 2020 年发布，但 2021 年才予以发布）。

改译：当前，欧盟正着力推进人工智能应用与管控的立法进程，其核心成果《人工智能法案》已于 2024 年 8 月 1 日正式生效，成为全球首部综合性人工智

能跨国监管法规（该法案草案曾于 2020 年启动立法协商，但因技术伦理争议及成员国博弈多次延期，最终确立基于风险等级的四级分类体系，禁止“不可接受风险”应用场景，并对违规企业设定最高达全球营业额 7%的罚款机制）。

分析：例 16 中，译者结合上下文背景和实际法规进展对时间细节进行校准，补充法案的核心内容以增强读者对“首个跨国 AI 法规”重要性的认知。原文中“2021 年”这一时间节点与实际情况存在偏差，根据欧盟《人工智能法案》的立法进程，该法案于 2023 年 12 月达成历史性协议，并于 2024 年 3 月经欧洲议会通过最终于 2024 年 8 月 1 日正式生效。因此，译文需调整时间表述，并补充法案的关键特征，如“风险分类监管框架”和“高额罚款机制”，从而凸显其作为“首个跨国 AI 法规”的重大意义。此外，原文括号内的内容需通过增译明确“指南”与“法案”的关系，补充说明法案的立法背景——欧盟为平衡技术创新与公民权利保护，在面临美国技术竞争压力下，通过渐进式立法逐步完善监管体系。这种译法不仅修正了时间误差，还通过加入法案的核心条款和立法动因，帮助读者理解其跨国监管的复杂性和开创性。

例 17. In their field trial in Thailand, says Google product manager Lily Peng, a system notification that an image was ungradable and that there should be a referral could very easily be misinterpreted. ‘For some people it means being referred to retinal specialists, which is at the higher end of care, versus refer for human review, which is what ended up being part of the protocol,’ she explains.

初译：在泰国的实地试验中，谷歌产品经理彭莉莉（Lily Peng）表示，系统给出的“图像不可分级需转诊”的通知容易被误解。“对部分医护人员而言，此提示意味着转诊至视网膜专科医生（高端诊疗），而实际协议应触发人工复核流程。”她解释道。

改译：谷歌产品经理彭莉莉（Lily Peng）指出，在泰国的实地试验中，系统发出的“图像无法判读，建议转诊”极易引发临床误解。“对于部分医护人员而言，‘转诊’意味着将患者转介至视网膜专科医生（前者属于高阶诊疗）；但根据实际操作规范，只是需要转交人工复查。”她解释道。

分析：在处理例 17 时，译者采用“释义补充深化理解”的翻译策略，因为原文中的某些概念可能在目标语言文化中并不直接通用或容易理解，尤其是涉及

专业术语时。在原句中，a system notification that an image was unarguable and that there should be a referral 这一部分涉及到一个系统通知，表明图像无法进行进一步评分并且需要转诊。翻译时需要通过适当的释义来传达这一通知背后的技术和医疗背景，同时增强读者对该情境的理解。原文中的 unarguable 可以理解为“无法争议的”，但这个表述在中文中略显生硬，可能会引起误解。为了更清晰地传达意思，可以翻译为“图像无法判读”或“图像无法确诊”，这样不仅传达了 unarguable 这一技术概念，还保留了该术语的医学含义。此外，原文提到的 refer to retinal specialists 与 refer for human review 之间的差异，翻译时也应作出更加明确的区分。在医学领域，“转诊”可能会让读者认为是指向专家的高端医疗服务，因此，需要补充说明“高端医疗护理”这一表述，突出这种误解可能带来的不同层次的医疗需求。而“人工复查”则是更为基础且规范的操作流程，翻译时应补充这一背景，以免产生误解。这样，译文不仅准确反映了原文内容，还帮助读者理解了不同的医疗术语和程序。

例 18. The laptop, computer and phone manufacturer ASUS, for example, suffered a supply chain attack in 2019 after a server for the company's software updates was compromised and malware inserted into the legitimate update that was sent out.

初译：例如，笔记本电脑、台式电脑和手机制造商华硕（ASUS）在 2019 年遭遇了一次供应链攻击，当时该公司用于软件更新的服务器被入侵，恶意软件被插入到发送的合法更新中。

改译：以华硕（ASUS）为例，该电脑设备制造商于 2019 年遭遇供应链攻击。其软件更新服务器遭入侵，致使恶意程序被植入官方推送的系统更新包。

分析：在翻译例 18 时，不仅传达原文的字面意思，还要让读者对事件的背景和影响有更深刻的理解。原句中的 suffered a supply chain attack in 2019 after a server for the company's software updates was compromised and malware inserted into the legitimate update that was sent out 叙述了一个复杂的网络安全事件。为了增强译文的可读性和准确性，可以补充更多的背景信息。例如，supply chain attack 在中文中没有直接的对应词汇，可以用“供应链攻击”来传达这一概念，但为了让读者更清楚了解该事件的背景，可以进一步解释这个攻击是如何发生的：攻击

者通过侵入公司软件更新服务器并在其中插入恶意软件，从而利用合法更新程序进行传播。这种解释能帮助读者理解攻击方式的复杂性，并强调恶意软件是如何通过看似安全的途径潜入系统的。因此，翻译时，除了保持原意，还应通过细化语言和补充信息来避免过于简化。对于原文中的 the legitimate update that was sent out，可以明确指出“发送的合法更新”可能会让读者误以为是简单的文件发送，实际上这里的“官方系统更新”特指公司推送给用户的官方更新，这样补充可以帮助读者更好地理解事件的严重性和复杂性。

3.3.2 逻辑拆分

在科技翻译中，长句常包含复杂的术语和逻辑结构，如果逐词逐句翻译，容易导致译文晦涩难懂或逻辑混乱。而通过信息层级拆分，将英语复合句按汉语流水句拆解为因果链，将抽象代词 it 具体化为所指名词，不仅保留了原文的信息完整性，还提高了译文的接受度，符合读者的认知规律。

例 19. Most of the world’s AI development is far too interconnected to be completely driven apart by political rivalries. It’s dominated by open-source publishing, research, breakthroughs and training datasets that are available to anyone with an internet connection.

初译：世界上大多数人工智能的发展都紧密相连，政治竞争很难将其完全割裂开来。它主要由开源出版物、研究、突破以及可供任何拥有互联网连接的人使用的训练数据集所主导。

改译：全球人工智能研发体系具有高度技术互依性，难以通过地缘政治彻底解耦。该领域由开源代码发布、学术研究协作及开放训练数据集主导，任何网络接入者均可获取核心研发资源。

分析：在翻译技术性较强或信息量大的句子时，使用这种方法能够抓住原文的逻辑层次，将主干与修饰部分分离，使译文语义更加明确。例 19 中，原文包含多个层次的信息：首先讨论了人工智能发展的整体特征——互联性与不可割裂性；其次详细阐述了这种发展的具体推动因素。译者通过将原文拆分成两部分，先译出主干信息“全球人工智能研发体系具有高度技术互依性”，然后将相关的技术细节和例证单独列出，如“开源代码发布、学术研究协作及开放训练数据集”

等。这种处理方式注重译文在目标语言中的自然性和功能性，确保译文读者能像原文读者一样轻松理解文本内容，避免长句信息过于密集而导致的阅读困难。

例 20. It's so easy to think you might introduce technology to solve other people's problems without actually bringing the people who are most impacted into the community. So [it's] not designing for but designing with.

初译：很容易认为自己可以引入技术来解决他人的问题，却并未真正让那些受影响最大的人参与到社区中来。所以，这并非是为他们设计，而是与他们共同设计。

改译：人们很容易产生这样的想法：你可以引入技术来解决人们的问题，而不将那些受影响最大的人纳入社区。所以，与其为某些群体设计，不如与这些群体一起设计。

分析：英文中像 it's 这种形式主语往往具有模糊性，因为其具体指代在语境中可能不够明确。通过翻译成短句的形式，比如“与其为某些群体设计，不如与这些群体一起设计”，可以更清晰地表达 designing for 和 designing with 之间的对比关系。这种方式不仅消除了指代的不明确性，还使句子更符合汉语表达习惯，有助于提升译文的可读性和逻辑流畅性。从翻译策略的角度来看，这种方法属于明晰化策略的一种表现，即在翻译过程中增添解释性短语或添加连接词等来增强译本的逻辑性和易解性（戴光荣&肖忠华, 2010: 78）。明晰化（explicitation）强调通过语义具体化和语境补充来增强译文的清晰度。在本例中，译者通过将 it's 隐含的意义转化为“与其……不如……”的结构，完成了从抽象到具体的转换。这种转换不仅帮助目标读者更容易理解，还避免了因指代不清而可能引发的歧义。其次，译者通过补充语境中隐含的信息，使得译文既忠实于原文，又符合目标语言的表达习惯和文化逻辑。

第四章 总结与反思

本章主要分为译文质量提升、心得与启示、不足与改进三部分，译文质量提升从自我审校、同学互审以及导师指导三方完成；心得与启示可以概括为词义谨慎选择、句式灵活变通；最后结合具体的翻译实践，总结出两点不足与改进。

4.1 译文质量提升

自我审校。译者完成初稿后，通过自我审校来提升译文的专业性和可读性。首先，重点检查术语的准确性和句子的表达方式，对照原文逐一核对专业术语，对较长或复杂的句子进行调整，确保意思完整、表达流畅。接着，采用朗读的方式检查语言质量，看看译文是否通顺，语法是否正确，同时借助 Grammarly 等工具查找可能的错误。最后，通过事先建立的术语库，对全文的术语使用进行对比检查，避免术语翻译不统一或概念表达偏差等问题。

同学互审。自我审校之后，译者邀请具有相同专业背景的同窗进行互审，以获得不同视角的反馈，弥补主观审校盲区。同学在审校过程中标记出发现的问题，并提出具体的修改建议。针对高频争议点展开辩论，最终形成共识性解决方案。例如，在人工智能使用说明的审校中，同组译者通过比对 AI 术语数据库，协同优化专业表述的合规性与可读性。

导师指导。在完成自我审校和同学互审后，译者将译文提交给导师，接受专业指导和审校。导师对译文进行全面审校，重点检查专业术语的准确性、长难句的表达、整体逻辑和流畅性。导师结合自己的专业经验，提出更高层次的修改建议。根据反馈意见，译者对译文进行进一步修改和完善，确保译文达到高水平的专业性和流畅性。

4.2 心得与启示

词汇选择需谨慎。首次接触专业性较强文本，译者深刻感受到词义抉择的复杂程度远超预期。以往的翻译实践中，多依赖字面意义进行理解，但此次经历使译者认识到，在特定学术语境中，日常用语往往被赋予更精确的技术内涵，这种

不确定性可能导致对文章理解出现偏差。因此，在后续处理人工智能类科技文本时，面对难以准确界定的词义，译者通过语境推敲和网络查证等手段加以确认。

句式变通需灵活。翻译不仅仅是逐句转换，更要注重全文上下文的连贯与逻辑性，以便读者顺畅理解译文。人工智能科普文本在保持科学严谨性的同时，常包含层层嵌套的从句、频繁的插入语以及严密的逻辑连接词。若拘泥于英文原有的句法结构，译文往往会显得生硬且充满翻译腔。为此，译者尝试按照意群拆分长句，调整语序并适当断句，使译文更符合汉语表达习惯。通过对信息的重组与表达优化，译者在准确传递技术内涵的前提下，实现了专业性与可读性的有机统一，从而更贴近中文学术写作规范。

4.3 不足与改进

在处理译文时，译者遇到了以下两个主要问题：一、**信息查证不充分。**作为非人工智能领域的译者，面对大量专业术语和特定文化背景，初期的查证工作可能存在一定的局限，这直接影响了对原文的准确理解，从而影响译文的专业性和精准度；二、**风格把握不精准。**在翻译过程中，译者对科技文本的语域特征把握不足，容易在“科学严谨性”与“表达流畅性”之间失衡。若过度强调严谨性，可能导致译文显得生硬，而过度追求流畅性，又可能削弱其专业性。

经导师指导，译者总结了未来翻译实践中应注意的三项关键措施：**一、加强专业知识储备。**翻译前应系统地学习人工智能领域的基础知识，掌握常见的技术术语和概念，提高对原文的理解能力；**二、反复推敲润色。**完成初稿后，反复审阅译文，从读者角度出发，调整语言表达，确保译文既准确传达原意，又符合目标语言的语言习惯；**三、研究科技文本风格。**大量阅读高质量的科技类译文，深入理解其语言特征和风格，培养对科技文本语域特征的敏感性。

参考文献

Matthew Burgess. *Artificial Intelligence： how machine learning will shape the next decade*. New York: Random House Business, 2021.

Newmark, P. *Approaches to Translation*. Shanghai: Shanghai Foreign Language Education Press, 2001.

Nida, E. A. & Taber, C. R. *The Theory and Practice of Translation*. Shanghai: Shanghai Foreign Language Education Press, 2004.

Reiss, K. *Translation Criticism: The Potentials & Limitations*. Shanghai: Shanghai Foreign Language Education Press, 2004.

Vermeer, H. *Towards a General Theory of Translational Theory*. New York: Routledge Taylor & Francis Group, 2013.

曹明伦：“英语定语从句译法补遗”，《中国翻译》，2001 年第 5 期，第 23-26 页。

戴光荣，肖忠华：“基于自建英汉翻译语料库的翻译明晰化研究”，《中国翻译》，2010 年第 1 期，第 76-80 页。

郭建中：《科普与科幻翻译：理论、技巧与实践》，北京：中国对外翻译出版公司，2004。

郝嘉欣：“英语翻译技巧中词性的转化策略分析”，《课程教育研究》，2019 年第 6 期，第 109-110 页。

胡健，范梓锐：“机器翻译视角下的翻译本质”，《当代外语研究》，2023 年第 2 期，第 90-96 页。

胡开宝：《语料库翻译学概论》，上海：上海交通大学出版社，2011。

李长栓：《非文学理论与实践》，北京：中国对外翻译出版有限公司，2012。

李宇明：“语言是文化的鸿沟与桥梁”，《天津师范大学学报》，2023 年第 6 期，第 34-42 页。

刘宓庆：《新编当代翻译理论》，北京：中国对外翻译出版公司，2005。

刘润清：《西方语言学流派》，北京：外语教学与研究出版社，2002。

秦洪武，周霞：“现代汉语的对比性特征——基于词性和句法关系清单的分析”，

《当代语言学》，2019 年第 3 期，第 418-437 页。

王福祥，徐庆利：“翻译递归性与翻译经验关系的实证研究:翻译过程视角”，《外语教学》，2018 年第 6 期，第 96-101 页。

叶子南：《高级英汉翻译理论与实践》，北京: 清华大学出版社，2008。

余高峰：“科技英语长句翻译技巧探析”，《中国科技翻译》，2012 年第 3 期，第 1-3 页。

张美芳：“从语境分析看动态对等论的局限性”，《上海科技翻译》，1999 年第 4 期，第 10-13 页。

郑声滔：“英语定语从句的并列法翻译”，《中国科技翻译》，2009 年第 2 期，第 14-17 页。

周怡珂，周领顺：“译者行为批评视域机器翻译本质问题反思”，《外语电化教学》，2023 年第五期，第 8-12+39+102 页。

附录一：术语及专有名词表

序号	英文	中文
1	Adversarial Attacks	对抗性攻击
2	AI Augmentation	人工智能增强
3	AI Ethics	人工智能伦理
4	AI Hardware Development	人工智能硬件开发
5	AI Politicisation	人工智能政治化
6	Algorithm Register	算法登记
7	Algorithmic Charter	算法宪章
8	Algorithmic Discrimination	算法歧视
9	Algorithmic Governance	算法治理
10	Algorithmic Impact Assessment	算法影响评估
11	Algorithmic Transparency	算法透明性
12	Algorithmic Vulnerability Bounty	算法脆弱性悬赏
13	Automation Paradox	自动化悖论
14	Autonomous Decision Making	自主决策
15	Autonomous Weapon Systems	自主武器系统
16	Backdoor Attacks	后门攻击
17	Bias Auditing	偏见审计
18	Computer Vision	计算机视觉
19	Contextual AI Deployment	人工智能情境化部署
20	Contract Review Automation	合同审查自动化
21	Data Poisoning	数据中毒
22	Data Protection Laws	数据保护法
23	Dataset Labeling	数据集标注
24	Data-to-Actionable Knowledge	数据转为可执行知识
25	Deep Learning	深度学习
26	Ethical Assessment	伦理评估

27	Explainable AI (XAI)	可解释人工智能
28	Facial Recognition Technology	人脸识别技术
29	GAN (Generative Adversarial Networks)	生成对抗网络
30	Gender Shades Study	性别阴影研究
31	Generalised Control Strategies	通用控制策略
32	High-Risk AI Systems	高风险人工智能系统
33	Human AI Partnership	人机关系
34	Human Oversight	人工监督
35	Image Cloaking	图像伪装
36	Intersectional Bias	交叉偏见
37	Machine Learning	机器学习
38	Neural Network Routing	神经网络路由
39	Plug-and-Play AI	自动化的人工智能
40	Polymorphic Malware	多态恶意软件
41	Predictive Algorithms	预测算法
42	Quality Assurance in AI	人工智能质量保证
43	Reinforcement Learning	强化学习
44	Robotic Suction Pick (RSP)	机器人吸附拾取系统
45	Structured Auditing	结构化审计
46	Supply Chain Attack	供应链攻击
47	Supply Chain Poisoning	供应链数据污染
48	Technological Dominance	技术霸权
49	Trojan Attacks	特洛伊攻击

附录二：原文与译文

原文	译文
5 AI and the world of work	第五章 人工智能与现实工作
If you want to get a sense of where AI is now, and where it's heading next, Ocado Technology is not a bad place to start. At the warehouses of the British online grocery tech company, robots, guided by AI, whizz around on rails at speeds of up to four metres per second, picking a 50-item order in minutes. The journeys then taken by Ocado's delivery trucks are optimised by a neural network that makes more than 14 million last-mile routing calculations per second, and adjusts delivery routes each time a customer places a new order or adds extra items to their shopping lists.	如果你想了解人工智能的发展现状以及未来的发展趋势，奥卡多科技（Ocado Technology）是一个不错的切入点。这家英国在线食品科技公司的仓库，是由人工智能（AI）机器人驱动，以每秒四米的速度在轨道上飞速移动，几分钟就可以完成一个包含 50 件商品的订单。Ocado 的智慧物流系统通过深度神经网络实现动态路径规划：每秒执行逾 1400 万次末端配送路径计算，当客户新增订单或实时修改购物清单时，系统能实时调整全局配送路线。
But Ocado's most ambitious automation efforts involve packing robots. At the time of writing the company has five robotic picking arms powered by computer vision, and other machine-learning systems that can identify the products that need to be packed and use suction power to grab them. Furth advances, undertaken in conjunction with two European academic-led projects, are in the pipeline.	但是，奥卡多最具雄心的自动化探索集中在打包机器人领域。截至本文撰写时，该公司已部署了五支由计算机视觉技术驱动的机械拣选臂，以及能识别待打包商品并通过吸力抓取的机器学习系统。此外，通过与两个欧洲学术机构主导的科研项目合作，更先进的机器人系统也正在研发中。

Picking and packing aren't easy if you're a robot. 'From a human's perspective, it is a fairly simple task to pick and pack, and it doesn't require an awful lot of training,' says Alex Harvey, chief of advanced technology at Ocado Technology. 'For a computer and for a robot, the dexterous manipulation involved is far beyond the state of the art today to be able to pick and pack the full range of items that we do.'	若让机器人执行拣选与打包任务，事情就变得复杂得多。“对人类而言，拣选打包是项相当简单的操作，几乎无需专门培训”，奥卡多科技的首席先进技术官亚历克斯·哈维（Alex Harvey）指出，“但对计算机和机器人而言，要实现我们现有的全品类商品拣选打包，所需的灵巧操作技术远超出当前技术水平。”
Ocado's Robotic Suction Pick (RSP) machine is a vacuum cup, powered by an air compressor, that sits at the end of an articulated arm. It uses computer vision and built-in sensors to select items gathered by a bot and place them into a shopping bag. Ocado.com sells a vast range of products, running into the tens of thousands. In terms of outward physical appearance, some are much the same: a tin of chopped tomatoes, for example, is not that different from a tin of lentils. But a tin of chopped tomatoes is very different indeed from a pack of yoghurts, which in turn are more sturdy than say, a bunch of grapes. And, of course, even grapes aren't all the same – they vary according to their variety and state of ripeness. Get the pressure of the vacuum cup wrong and the	奥卡多的机器吸取挑选（RSP）是安装在关节臂的末端，由空气压缩机提供动力的真空吸盘。它利用计算机视觉和内置传感器，对机器人选择的商品进行挑选，并将它们放入购物袋中。奥卡多官网销售商品种类逾万种，从外观来看，有些产品非常相似：例如，一罐切碎的番茄与一罐扁豆并没有太大的区别，但一罐切碎的番茄与一盒酸奶的差异却很大，而酸奶又比一串葡萄更坚固。值得注意的是，即便同类商品如葡萄，其个体也会因品种差异与成熟度变化呈现不同形态特征。若真空吸盘压力参数设定失准，RSP装置将面临抓取失败或损毁商品的风险；分拣时序规划不当，则可能导致番茄罐头压溃葡萄等作业事故。

RSP will either drop or crush the item it is attempting to manipulate. Get the sequence wrong and there's a danger that the tin of tomatoes will squash the grapes.	
"At present, the company is expanding the number of items its robotic suction system can pick. 'There's no point having for sixty thousand different items, sixty thousand different control pieces of code,' Harvey says. 'What we want at Ocado is generalised control strategies.' But challenges remain. 'We need to fit a robot into the same square footage that the person sits in or operates in, and we need the robot system to achieve the same throughput.' Until a robot can pick and pack as many items in an hour as a human being – Harvey says this is around 600–700 items – it is unlikely to be widely adopted: the impact on productivity would damage service and profits. It also has to be affordable, which means, on the one hand, scaling the technology to the point where it becomes economically worthwhile, and, on the other hand, not over-speccing it (for example, by using a camera with an unnecessarily high resolution). 'When we're deploying stuff in the real world, we want it to be economical	"目前，公司正在增加其机器吸盘系统可以拾取的商品数量。"哈维说道："如果拾取六万种不同商品，要使用六万种不同的控制代码的话，那这个系统毫无意义。我们的目标是开发通用型控制策略。"但挑战依然严峻。"机器人系统需在人类作业的同等空间内运行，并达到相近的物料处理量。"哈维（Alex Harvey）指出，只有当机器人的分拣效能达到人类水平 (约 600-700 件）时，该技术才可能被大规模应用——否则其对生产效率的负面影响将损害服务质量与盈利空间。而且价格也必须要经济实惠，一方面需通过技术规模化实现经济可行性，另一方面需避免过度配置硬件（如采用超出必要精度的高分辨率摄像头）。"在现实场景部署时，我们必须确保技术应用的经济性，"哈维补充道，"不可能为每个拣选机器人配备超级计算机。"

in the way that we're deploying it,' Harvey says. 'I don't want to deploy a supercomputer next to every robot picker.'	
"Whether or not such AI will ultimately replace humans is the billion-dollar question. Many now believe AI will work alongside humans. Obviously, AI offers the promise of greater efficiency, but so far, at least, this tends to hold good only where the environment is relatively controlled and predictable. Production lines and warehouses may well become fully automated. But where processes interact with the outside world – with all its randomness – it's harder to envisage a wholly AI future. Delivery drivers, for instance, have to take into account such factors as the weather and the erratic behaviour of some pedestrians. It's possible that there will be a future without them. It's also possible, though, that AI will control lorries and delivery trucks on major highways where conditions are relatively predictable, and that human drivers will take over the driving as they reach the outskirts of the villages, towns and cities set as their destinations.	"AI 是否终将取代人类，这是价值千亿美金的核心命题。"当前主流观点认为，AI 将长期与人类形成协同工作模式。诚然，AI 在效率提升层面展现出显著优势，但至少目前来看，这种优势通常只出现在环境相对受控和可预测的情况下。生产线和仓库很可能实现全面自动化，但当技术应用涉及现实世界的随机性交互时，纯粹 AI 主导的解决方案仍面临实施困境。以物流驾驶员为例，其工作需实时应对天气突变、行人突发行为模式等复杂变量。未来可能会出现没有司机的情况。不过，人工智能或将在路况相对可控的高速公路接管物流运输车辆，待车辆驶抵目标城镇、乡村或城市郊区时，则会由人工接管。

Arguably, it's the medical sphere where AI's potential and its limitations have been most apparent – and where its possible future relationship with humans has been most clearly demonstrated. Take Google's computer-vision system, for example. It's capable of spotting diabetic retinopathy, a complication of diabetes that can cause sight loss. In lab conditions, it can achieve an accuracy rate of 90 per cent and provide results in ten minutes.	按理说，医疗领域是人工智能最有潜力但同时局限性最大的领域，未来也可能是与人类息息相关的领域。以谷歌的计算机视觉系统为例，该技术可精准识别糖尿病视网膜病变（糖尿病引发的致盲性并发症）。在实验室环境下，系统诊断准确率高达 90%，且能在十分钟内输出结果。
When put to a real-world test in Thailand, however, the deep-learning set-up often struggled. The 11 clinics that operated it used different technology. Only two had a dedicated eye-screen room that could be made dark enough for patients' pupils to enlarge to the point where high-quality fundus photos could be taken. Most had to make use of nurses' offices or general-purpose rooms. Emma Beede, the lead Google Health researcher evaluating the technology, described one room where, because dental checks and patient consultation were going on, it was essential to leave the light on. 'That makes sense for them,' she said. 'That makes sense when you're under-resourced.' Poor internet connection caused problems, too – at one	然而，这套深度学习系统在泰国进行实地测试时，却暴露出诸多适应性挑战。参与测试的 11 家诊所技术条件参差不齐，仅有两家配备专用眼底筛查暗室，可调节至患者瞳孔充分扩张所需暗度以确保高清眼底成像；其余诊所不得不使用护士站或多功能诊室进行检测。谷歌健康部门主管艾玛·比迪（Emma Beede）描述某诊所实况时指出，由于需同步进行牙科检查与患者问诊，室内照明必须保持开启状态。“这符合他们的现实需求，”她说。“在资源受限的医疗环境中，这是合理选择。”"网络基础设施薄弱同样引发问题——某诊所曾出现长达两小时的全网中断。多重因素叠加下，试验前六个月提交系统的 1,838 张眼底影像中，21% (393 张) 因质量不达标被判无效。

clinic it dropped out altogether for two hours. One way and another, during the first six months of the trial, 21 per cent (393) of the 1,838 images put through the system proved to be of an insufficiently high quality. Nurses and patients felt frustrated. 'I'll do two tries,' one nurse said. 'The patients can't take more than that.'	医护人员与患者普遍产生挫败感，一位护士表示："最多尝试拍摄两次，患者耐受度已到极限。"
And then there was the human factor. At one clinic half the patients earmarked for the study declined to be involved after finding *out* that although the results would be virtually instantaneous (previously some results had taken ten weeks to arrive), a positive diagnosis would see them being referred to a hospital an hour's drive away. One member of medical staff told the Google researchers that patients 'are not concerned with accuracy, but how the experience will be – will it waste my time if I have to go to the hospital?'	此外还有人为因素。某临床试验点，半数入选研究的患者在知情后选择退出。尽管检测反馈周期从十周优化至实时，但阳性确诊者将被转诊至距诊所一小时车程的定点医院。一位医护人员向谷歌研究人员坦言："患者并不在意技术是否精准，他们更关心诊疗体验——如果需要再去医院复查，是否会浪费我的时间？"
There's little doubt that the technology works: tests in the trial proved accurate when conditions were right. And when things were going well, Beede says, senior nurses felt they had more time to spend with patients and speak to them about their health and lifestyles. But the	这项技术的有效性毋庸置疑：在条件适宜时，试验证明其诊断准确性确有保障。比迪指出，当系统运行顺畅时，资深护士得以腾出更多时间与患者深入沟通，探讨其健康状况与生活方式。然而，现实世界的不可预测性与人类需求的复杂性不容忽视。比迪强调：

unpredictability of the everyday world and the needs and concerns of humans cannot be ignored. 'I think the main takeaway from this research is that we need to be designing AI tools with people at the centre,' says Beede. 'We need to be considering our success beyond the accuracy of the model. We need to understand how it's actually going to work for real people in context.' She argues that as more AI is implemented in the real world, more pilot studies will be required to ensure that it works for everyone involved. 'That implementation is of equal importance to the accuracy of the algorithm itself, and cannot always be controlled through careful planning,' the Google report concludes.	"本研究的关键启示在于，我们必须将人类置于 AI 工具设计的核心。评判成功的标准不应局限于模型准确率，更需要理解技术如何在具体情境中为真实个体创造价值。" 她认为，随着 AI 技术加速落地，必须开展更多试点研究以确保其普惠性。谷歌的报告总结道："技术实施过程与算法精准度同等重要，且无法完全通过周密规划预先掌控。"
The need for a proper and carefully considered partnership between humans and AI has been well demonstrated by Finale Doshi-Velez, a Harvard University computer science professor who leads its Data to Actionable Knowledge Lab. In one experiment, she and her colleagues recruited 220 psychiatrists to study case notes for hypothetical patients supposedly suffering from major depressive disorder. Each set of case notes described that	哈佛大学计算机科学教授菲纳莱·多西-维莱兹（Finale Doshi-Velez）领导的数据可操作知识实验室（Data to Actionable Knowledge Lab）通过实证研究揭示了构建人机协作良性伙伴关系的重要性。在这个实验中，研究团队招募 220 名精神科医生，针对虚构的重度抑郁障碍患者病例进行分析。每组病例资料包含患者病情描述，并随机附加三种干预条件：经独立验证正确的 AI 治疗建议、错误建议或完全

particular patient's condition and was then followed by either an independently verified correct AI-generated recommendation for treatment, an incorrect recommendation or no recommendation at all. Where an AI diagnosis was given, it was accompanied by an explanation for that diagnosis, which varied in length, quality and detail.	无建议。当提供 AI 诊断时，会同步附上解释说明——这些说明在长度、质量与详略程度存在系统性差异。
'What we found was that when the recommendation was correct, overall, everything improved,' Doshi-Velez says. Doctors and AI decisions proved to be a formidable team. 'Humans may have already had an idea, the recommendation reinforced it or caused them to change their mind to that idea.' However, when the recommendation presented to the volunteer doctors was incorrect, it tended to lead to poorer forms of decision-making. The psychiatrists were influenced by incorrect recommendations from the AI, leading to lower levels of treatment selection accuracy. This is scarcely a new phenomenon. It's long been known that if we put machines in charge of simple tasks, humans will, without continuous training, forget how to do them. Hence, at an everyday level, why digital contact books	多西·维莱兹解释道："我们发现，当提议正确的时候，整体上都会改善。"医生和人工智能的共同的决策展示出了合作的强大。"医生原本可能已有初步判断，AI 建议或强化其确信，或促使修正原有决策。"但当 AI 给出错误建议时，却会引发更劣质的决策输出，精神科医生受误导性建议影响，治疗方案准确性明显下降。这并不是一种新现象，人类将简单任务委派机器后，若无持续训练，相关技能必然退化。因此，在日常生活中，手机通讯录导致人们记忆电话号码能力丧失。这也是为什么 2009 年 6 月 1 日 2009 年法航 447 航班飞行员因过度依赖自动驾驶系统，在系统失效时应对失当，最终酿成 228 人遇难的空难。随着 AI 普及，这种"自动化悖论"只会愈发凸显。

in phones have caused us not to remember phone numbers any more. Hence, at an extreme level, why on 1 June 2009 the pilots of Air France Flight 447, who had come to rely heavily on the plane's autopilot features, could not cope when the systems failed, and so presided over a crash in which 228 people perished. With AI this 'paradox of automation' is only going to become more pronounced.	
The way in which information is presented is also an important factor in determining the success or otherwise of the final outcome. 'Certain forms of explanation are more effective at preventing a wrong decision,' Doshi-Velez says. It's something the Google Health researchers have noticed, too. In their field trial in Thailand, says Google product manager Lily Peng, a system notification that an image was ungradable and that there should be a referral could very easily be misinterpreted. 'For some people it means being referred to retinal specialists, which is at the higher end of care, versus refer for human review, which is what ended up being part of the protocol,' she explains.	信息呈现的方式也是决定最终结果成功与否的重要因素之一。多西·维莱兹说："某些呈现的方式会更有效地防止错误的决定"，谷歌健康的研究人员也注意到了这一点。谷歌产品经理彭莉莉（Lily Peng）指出，在泰国的实地试验中，系统发出的"图像无法判读，建议转诊"极易引发临床误解。"对于部分医护人员而言，'转诊'意味着将患者转介至视网膜专科医生（前者属于高阶诊疗）；但根据实际操作规范，只是需要转交人工复查。"她解释道。
'It's not just about the machine-learning part, it's about figuring out how to present	多希·韦莱兹认为，"这不仅关乎机器学习组件本身，更需深究信息呈现的

the information as well,' Doshi-Velez argues, stressing the point that conclusions reached by an AI system need to make humans think. Such conclusions can't be too easy to accept or they risk becoming relied upon without that crucial element of critical thinking. By the same token, they can't be too hard to digest or humans will simply ignore or gloss over them. 'Models need to be transparent in their limitations, highlighting situations in which the AI prediction may not be accurate or valid,' the Harvard study concludes.	艺术"。她强调，AI 系统得出的结论必须激发人类思考，而且这些结论既不能过于易被接纳，否则将引发批判性思维的缺失性依赖。同样，也不能艰涩难懂，导致人类选择忽视或草率处理。哈佛的研究得出结论，模型需透明揭示其局限性，明确标注 AI 预测可能失准或失效的具体情境。
But when humans and AI are in sync, the potential benefits are huge. The more positive outcomes of the Harvard project, for example, suggest AI could eventually help with diagnosis for a mental health condition that is often missed and for which treatments vary wildly. They start to offer hope for the more than 264 million people around the world who battle with depression.	当人类和人工智能同步协作时，潜在的好处是巨大的。以哈佛研究项目的积极成果为例，人工智能最终可能对诊断一种经常被忽视且治疗方法大相径庭的心理健康状况有所帮助。这些成果将为全球 2.64 多亿抑郁症患者带来希望。
AI vs humans or AI with humans?	人工智能和人类竞争还是合作?

What, then, does the future hold for AI and the world of work in the medium term? It's likely to be a mixed bag of results. There have been and will continue to be disappointments and failures along the way. In 2018 IBM had to ditch a multi-million-dollar project designed to help the treatment of cancer patients after it was found to be giving clinicians bad advice. During the coronavirus pandemic Walmart abandoned the use of robots to scan shelves and assess levels of stock when it realised humans were just as effective. An October 2020 study conducted by the MIT Sloan Management Review and the Boston Consulting Group, which surveyed more than 3,000 business leaders running companies with annual revenues above $100 million, discovered that only in 10 per cent of cases did people feel that the investment they made in AI produced a 'significant' return.	那么，AI 和工作世界在中期内的未来会怎样？技术演进过程中既有过往教训，亦将持续面临挫败。2018 年，IBM 被迫终止耗资数千万美元的癌症辅助治疗项目，该项目原本旨在帮助癌症患者治疗，但其向临床医生提供错误建议。在新冠疫情期间新冠疫情期间，沃尔玛发现人工盘点效率与机器人相当后，放弃货架扫描机器人部署。2020 年 10 月，麻省理工学院斯隆管理学院和波士顿咨询集团进行的一项研究调查了超过 3000 名年收入超过 1 亿美元的公司领导者，结果发现只有 10%的受访者认为他们在 AI 上的投资获得了"显著"的回报。
AI will cause disruption, too. 'Better-educated, better-paid workers will be the most affected by the new technology, with some exceptions,' research from the Brookings Institution found in November 2019. Those whose jobs currently involve a close focus on data	AI 也将带来一些不便。布鲁金斯学会 2019 年 11 月的一项研究发现，"除个别特例外，高学历、高薪酬劳动者将成为受新技术冲击最严重的群体。"那些当前工作高度依赖数据处理的人，如市场研究员、销售经理、计算机程序员和个人理财顾问，将特别容

will be particularly vulnerable: market researchers, sales managers, computer programmers and personal finance advisers among them. Those whose jobs involve a lot of interpersonal skills, such as those in education and social care, will probably be less affected: AI is very unlikely to replace human compassion and empathy. That said, in Japan care robots are already used to help the country's ageing population. And in any case, it's dangerous to make sweeping generalisations. The fact is that AI adoption will vary around the world according to local culture and social attitudes. Automation in finance in Singapore is likely to be very different from automation in finance in Pakistan.	易受到影响。相反，依赖人际交互技能的岗位（如教育、社会护理）可能不太会受到影响，因为 AI 很难替代人类的同情心和共情能力。尽管如此，日本已规模化应用护理机器人应对老龄化危机，但无论如何，做过于笼统的概括是危险的。事实上，AI 技术渗透深度将因地域文化与社情差异呈现显著分化，新加坡金融自动化路径必然迥异于巴基斯坦。
If Google's work with computer vision or Harvard's study with psychiatrists are anything to go by, though, it seems likely that the general trend will be for AI not to replace existing jobs but to transform them – and to create new ones, too. Already, thousands of roles exist that would have been unimaginable at the turn of the century. Scores of people now work on AI labelling, helping to compile datasets that train machine learning. Thousands of individuals have been taken on at	如果参考谷歌在计算机视觉方面的工作或哈佛与精神科医生的研究，AI 发展的总体趋势或将并非取代现有岗位，而是重塑既有职业形态，同时催生新工种。本世纪之交难以想象的职业如今已成常态，数以万计从业者投身数据标注领域，构建机器学习训练数据集；诸如脸书（Facebook）和油管（YouTube）等平台雇用大量内容审核员，处理 AI 初步识别的违规信息，最终由人工复核裁决。

companies such as Facebook and YouTube to moderate content that might be breaking their platforms' rules, which has in many cases been initially flagged by an AI.	
Researchers at MIT Sloan and Boston argue that those companies poised to benefit most from AI are those who use it to augment and shape traditional processes rather than replace them. In other words, they create an environment in which humans learn from AI and AI learns from humans. The toolmaker Stanley Black & Decker is one example. It has started using computer vision to check the quality of the tape measures it manufactures. The system flags defects in real time, spotting problems early in the production cycle and so reducing wastage. But humans are still on hand to inspect and make judgement calls on the worst faults.	麻省理工斯隆管理学院与波士顿大学的联合研究指出，最能收获人工智能红利的企业，是那些通过智能技术实施流程重塑与人机协同，而非简单粗暴的岗位替代。换句话说，这些公司构建了人类与 AI 双向学习的协同环境。工具制造商斯坦利·布莱克·百得（Stanley Black & Decker）就是一个例子，该公司采用计算机视觉系统检测卷尺产品质量，实时标记生产周期早期的缺陷以降低损耗率。但人工质检员仍保留对重大缺陷的最终裁决权。
Experts are key to creating trustworthy AI systems, says Ken Chatfield, the vice-president of research at Tractable, an AI firm that uses computer vision to help make decisions about insurance claims after car crashes – its AI is being used in the real world by some of the biggest insurance companies. The company initially trained its AI on thousands of	特拉克特（Tractable）公司的研究副总裁肯·查特菲尔德（Ken Chatfield）表示，构建可信赖的 AI 系统离不开领域专家深度参与。该公司运用计算机视觉技术辅助车祸保险理赔决策，其 AI 系统已获多家顶级保险公司商用部署。期训练阶段，系统学习数千张涵盖车门板受损、挡风玻璃破裂等各类事故车辆图像，但性能突破性提升源

images of vehicles that had been in accidents – involving damaged door panels, broken windscreens and more. But it saw the biggest improvements in the system's performance when the damage highlighted in images had been labelled by specialists, with years of experience in assessing crash reports. And it is human insurance agents who take over once the AI has reviewed images and suggested what the next steps should be. 'The data in itself is not enough, and also our knowledge as researchers is not enough – we really need to draw on the knowledge of experts in order to be able to train models,' Chatfield explains. 'Involving the expert is also what we need to build up trust'.	自引入具有数十年定损报告评估经验的专家标注数据。AI 完成图像分析并提出处置建议后，最终决策权仍由人类保险专员掌握。"仅凭数据本身不够，研究者的知识储备也不足，必须融合专家智慧训练模型，"查特菲尔德解释道。"专家参与更是构建系统信任度的关键。"
The London-based lawyer Richard Robinson, the CEO of Robin AI, has struck a not dissimilar balance in the legal field. He quit his job at a large law firm when he became convinced that many of the repetitive tasks that went into contract work could be automated. 'A lot of what I would spend my time doing as a lawyer felt like it didn't need much brain power,' he explains. His view was that machine learning could be utilised for reviewing some types of contract, such as those	伦敦律师理查德·罗宾逊（Richard Robinson）创立的罗宾人工智能（Robin AI），在法律领域实现了类似的平衡。这位律所 CEO 因确信合同审查工作中的大量重复性任务可实现自动化，毅然离开知名律所。他解释道："作为一名律师，我花费时间做的很多事情感觉并不需要太多脑力。"他认为核心理念在于运用机器学习技术审查特定类型合同，如涉及雇佣条款的标准化文件，这类法律事务的自动化处理看似技术门槛较低。

concerned with employment conditions. The tasks involved seemed simple enough.	
It didn't, however, turn out that way. 'The truth is it was much more difficult than we anticipated,' Robinson says. 'There are so many random things that could be in that document, that you can't be confident that the AI will always identify them.' What he therefore did was to create a system in which AI works with human lawyers rather than instead of them. The company's system has been trained on historical contracts – both those in the public domain and documents provided by clients – and taught to look for particular elements. It's therefore able to detect whether a non-compete clause has been sneakily added into a business contract, or whether an employment contract stipulates non-standard working hours.	然而，情况并非如此。罗宾逊说：事实证明其难度远超预期，法律文件中可能存在的随机性条款如此之多，无法确保 AI 总能精准识别。"因此，他构建的并非替代性方案，而是打造人机协作的智能合约审查体系：该平台通过深度学习历史判例库（既涵盖公开司法协议，也整合客户专有文档），训练算法能够精准识别关键法律要素。因此，它能够检测出商业合同中是否悄悄加入了竞业禁止条款，或雇佣合同是否规定了非标准工作时间。
If it finds anomalies, the system alerts a human lawyer via email and they then check the document. The same thing happens if the system is unable to interpret a particular clause or contract. A recent assignment the company took on was checking contracts between big fast-food retailers and their suppliers during the early months of the coronavirus pandemic, to	当系统检测到异常条款或无法解析特定条款时，将通过邮件提醒律师人工核查。如果系统无法解释某个特定的条款或合同，同样也会交给人工处理。该公司近期承接的项目涉及审查新冠疫情初期快餐巨头与供应商的合同，以厘清危机情境下各方的责任义务。罗宾逊认为，律师们觉得检查合同是一件乏味的事情，但完全依赖人工智

find out what each party's obligations were in the event of a crisis. Robinson's view is that lawyers find checking contracts tedious. At the same time, it's dangerous to rely wholly on AI, because even if it's getting things right 96 per cent of the time, that's not good enough when companies' and individuals' lives and livelihoods are at stake. 'We want to use AI to make the first attempt at everything in situations where it's really easy for a person to check and see if it's wrong or right,' he says.	能也很危险。即使人工智能的准确率达到 96%，在事关企业存续与个人生计的领域仍不可接受。他说：“我们的理念是让 AI 完成初筛，这些场景下人类可轻松核验结果正误”
However organisations end up using AI, there's no doubt that as it spreads it will become easier to access and operate. At present most AI deployments involve handcrafted technology. In the future, a company's AI requirements may be handled by a third party, using software that seems as straightforward as that inside word processors or slideshow builders. A company that wants to use AI to analyse specific datasets or images will be able to use a template to create this. The algorithm it picks may not have been created by the third-party service they're buying the template from, but from another company further up the chain of businesses developing and industrialising AI. The	然而，无论组织最终如何使用人工智能，毫无疑问的是，随着人工智能的普及，必将推动使用门槛的持续降低。目前，大多数人工智能部署仍需定制化开发，而未来，企业的人工智能需求可能由第三方处理，使用的软件就像字处理软件或幻灯片制作软件一样简单。企业若需分析特定数据集或图像，仅需选择相应模板即可构建 AI 系统。所选择的算法可能不是由提供模板的第三方服务公司创建的，而是来自于在 AI 开发和工业化链条上更上游的另一家公司。到那时，这项技术将自动化，届时其与日常生活的交融或已达润物无声之境。

technology will become plug-and-play. By that point we may hardly notice its interaction with our daily lives.	
Once this happens the world really will change significantly. People's workplaces will face automation at a greater scale than at any point so far this millennium. For many the entire nature of work may change. How we interact with businesses and government services will also be transformed. Societies that deploy AI will need to learn how people react to the technology and what their expectations of it are. At the same time, individuals will only follow the directions given by an AI if the system works efficiently, is understandable – and can be trusted.	当这种技术渗透达到临界点，世界将会迎来巨大的改变。工作场所将经历本世纪最剧烈的自动化转型，许多工作岗位的本质属性面临重构。公众与政企服务的交互模式将被彻底改写，部署人工智能的社会需要了解人们对技术的反应以及人类对人工智能的期望。与此同时，只有当 AI 系统高效运行、易于理解并且值得信赖时，个人才会遵循 AI 给出的指示。
6 Weaponising AI	第六章 武器化的人工智能
AI's inexorable rise has been accompanied by its politicisation. This is not just because it offers the potential for societies to operate more efficiently, or that it is now intricately bound up with their economic success. It's because it's become a weapon in the struggle for a technological dominance that carries with it immense political power. In the 1960s the former Soviet Union and the USA competed in the space race. Today it's the USA and the	人工智能的迅猛发展，始终伴随着深刻的政治化进程。这不仅源于其提升社会运行效率的潜力，更因其已深度融入国家经济命脉的核心构造，而是因为它成为了技术霸权争夺中的一件武器，承载着巨大的政治影响力。20 世纪 60 年代，前苏联与美国展开太空争霸；而今，美国与崛起中的超级大国中国，正在争夺全球技术体系及其基础设施的控制权。

growing superpower that is China that are seeking supremacy in the struggle to achieve control of the world of technology and its infrastructure.	
The politics	政治背景
For a long time, it has been Western universities, technology companies and policymakers that drive innovation, with talent from around the world flocking to the US to work on AI. Now, China is seeking to drive forward tech standards in everything from 5G to AI developments. Its ambitions are serious. In 2017, the country's ruling Communist Party announced its aim to be world leader in AI by 2030. It's currently spending billions on AI development, much of it in the private sector. Analysis from the US think tank the Center for Security and Emerging Technology suggests that the Chinese government spent around the same amount on AI investment in 2018 as the US planned to in 2020.1 Others have claimed that its expenditure is actually far outstripping that of the US.	长期以来，创新驱动力主要来自西方高校、科技企业与政策制定者，全球人才纷纷涌入美国投身人工智能研究。而今，中国正从 5G 到人工智能技术全面推动标准制定。这个东方大国的科技雄心绝非虚言，2017 年，中国共产党明确提出 2030 年成为世界人工智能领域的领导者。当前中国正以千亿级资金注入 AI 发展洪流, 其中私营领域投资占比惊人。美国安全与新兴技术中心 (CSET) 智库分析显示, 2018 年中国政府 AI 投入规模已与美国 2020 年预算持平。更有研究指出, 中国实际支出可能远超美国。
Other metrics, too, show China's commitment. Since 2005 it has published more AI research papers than the US, even if they haven't all had the same impact on	其他方面的数据同样彰显着中国的投入力度。自 2005 年起，中国发表的 AI 学术论文总量已超越美国，尽管这些论文的学术影响力尚未全面比肩美

the academic community as US-based work. That, however, is set to change, according to analysis by the Allen Institute for AI in May 2019 which suggested that within a year the number of Chinese-led research papers cited by others would match the number of US ones being cited2 – although the Institute says this growth may be partly fuelled by a virtuous circle, where Chinese researchers may increasingly cite more Chinese researchers.	国研究成果，但根据艾伦人工智能研究所（Allen Institute for AI）2019 年 5 月的分析预测，中国论文的被引次数将在一年内与美国持平。尽管该研究所指出，这种增长可能部分源于中国学者日益倾向于引用本国研究的良性循环。
China may currently lag behind the US and other nations in its AI hardware development, but it nevertheless now boasts some of the world's leading AI software companies. SenseTime, Face++ and Hikvision are all prominent in facial recognition. Alibaba, Baidu and Tencent now rival Google, Facebook, Microsoft and Apple in their spending power and appeal to AI would-be employees.	尽管中国在人工智能硬件研发领域暂时落后于美国及其他技术强国，但其已然孕育出多家全球顶尖的人工智能软件企业。商汤科技（SenseTime）、旷视科技（Face++）、海康威视（Hikvision）在面部识别技术领域占据全球领先地位；现如今阿里巴巴、百度和腾讯在资金实力和对人工智能员工的吸引力上，与谷歌、脸书、微软和苹果不相上下。
And the country is spreading its influence across many of the world's less economically developed countries. China's Belt and Road scheme has provided developing nations with new technology that will in future become fuelled by AI. 'In Africa the backend network infrastructure is basically almost entirely	中国的影响力正扩散到世界上许多经济欠发达的国家。通过“一带一路”倡议，中国为发展中国家提供了未来将深度整合人工智能的新型技术基础设施。中非项目的负责人欧瑞克·奥兰德（Eric Olander）表示：“非洲的后端网络基础设施现在基本上完全由华为和中兴通讯提供。加纳的监控摄像头

Chinese now, between Huawei and ZTE,' says Eric Olander, who leads the China Africa Project. 'In Ghana, they've gone from 700 [CCTV] cameras to 7,000 cameras in the space of just a few years.' During the coronavirus pandemic, Olander adds, Chinese companies donated technology to African states to help them fight the virus. Huawei, for example, provided video conference systems and thermal temperature scanners, as well as some cash, during the spring of 2020.	数量在短短数年间从 700 个激增至 7000 个。"奥兰德又补充道，新冠疫情期间，中国科技企业向非洲国家捐赠抗疫技术装备。例如，在 2020 年春季，华为在 2020 年春季提供了视频会议系统、热成像体温检测仪及现金援助。
Tech rivalry isn't limited to China and the US, though: other countries around the world are investing in AI research and development, aiming to keep up with the global leaders and seeking to deploy the technologies across their own nations. It's unlikely that such competition will lead to the fracturing of the technology, or into a simple two-way split between the US and China. Most of the world's AI development is far too interconnected to be completely driven apart by political rivalries. It's dominated by open-source publishing, research, breakthroughs and training datasets that are available to anyone with an internet connection. Nevertheless, international tensions are	不仅中美之间存在科技竞争，世界上的其他国家也在投资人工智能的研发，力图跟上全球领导者的步伐，并试图在本国范围内部署这些技术。这种竞争不太可能造成技术的分裂，或者简单地将世界分为美国和中国两个阵营。全球人工智能研发体系具有高度技术互依性，难以通过地缘政治彻底解耦。该领域由开源代码发布、学术研究协作及开放训练数据集主导，任何网络接入者均可获取核心研发资源。然而，随着中国致力于实现其超级大国目标，国际紧张局势可能会加剧。

likely to increase as China aims to achieve its superpower targets.	
In immediate terms the expansion of AI has raised questions of regulation. Some claim that existing laws, covering human rights and data protection, are sufficient to keep abuses in check. Others argue that AI raises wholly new issues that can only be dealt with by specific laws. Some are nervous about particular aspects of AI technology (several US cities have banned the use of facial-recognition technology). Some believe that it should not be used at all in certain domains (there is a wide consensus among the AI community that the technology should not be used for autonomous weapon systems, such as drones, that can take lives without human intervention).	人工智能的迅猛扩张已引发监管体系的深层拷问。部分观点认为，现有人权与数据保护法律已足以遏制滥用行为；另一些则主张需针对 AI 制定专项法规以应对其引发的全新挑战。具体争议点包括：美国多个城市已禁止使用人脸识别技术；学界普遍反对将自主武器系统（如无需人工干预即可杀人的无人机）投入实战——这已成为人工智能领域的广泛共识。
The European Union may end up leading the way in such matters. Its General Data Protection Regulation (GDPR), whose enforcement began in 2018, has redefined the way people's personal information is handled around the world. Now it has now turned its attention to codifying laws on the use and control of AI, which, if and when issued, would constitute the first transnational regulation of AI (the	欧盟或将在 AI 治理领域引领变革。欧盟于 2018 年开始实施的《通用数据保护条例》（GDPR），重新定义了全球范围内个人信息的处理方式。当前，欧盟正着力推进人工智能应用与管控的立法进程，其核心成果《人工智能法案》已于 2024 年 8 月 1 日正式生效，成为全球首部综合性人工智能跨国监管法规（该法案草案曾于 2020 年启动立法协商，但因技术伦理争议及成员

guidelines were scheduled for publication in 2020, but then delayed to an unspecified date in 2021). 'We want a set of rules that puts people at the centre,' said Ursula von der Leyen, the European president, in September 2020. 'Algorithms must not be a black box, and there must be clear rules if something goes wrong.'	国博弈多次延期，最终确立基于风险等级的四级分类体系，禁止"不可接受风险"应用场景，并对违规企业设定最高达全球营业额 7%的罚款机制）。"我们希望制定一套以人为中心的规则，"欧盟主席乌尔苏拉·冯·德莱恩（Ursula von der Leyen）在 2020 年 9 月表示："算法系统必须禁止暗箱操作，当决策链条出现偏差时，须有可溯源的权责框架作为系统性追责机制。"
Von der Leyen's comments followed the publication of a white paper setting out what the EU might do: for example, introducing rules to govern 'high-risk' AI systems (such as those used in health, policing and transport) and those cases where AI use could result in discrimination or injury or put people's lives in danger. It's worth noting, though, that a consensus view is some way off. Of the more than 1,200 groups the EU consulted, 42 per cent thought that new laws were needed to govern AI, while 33 per cent argued that existing legislation could do the job. There was a similar split over possible approaches to high-risk uses of AI (42 per cent to 30 per cent). Clearly there is a long way to go.	冯·德莱恩的表态呼应了欧盟发布的 AI 监管白皮书，其中明确了监管方向，例如对医疗、警务、交通等"高风险"AI 系统实施专项管制，重点防范可能引发歧视、人身伤害或危及生命的场景。然而，就这一问题的共识远未达成。在欧盟咨询的 1200 多个团体中，42%支持制定 AI 专项法律，33%认为现有立法已经足以应对。关于高风险 AI 应用的治理路径同样存在分歧（42%支持新规，30%倾向现有框架），可见前路依然漫长。
The hackers	黑客

If countries are in the process of using AI as a tool reinforcing power and dominance, criminals are also starting to weaponise the technology. First among them will be hacker and cybercriminal communities.	当各国将人工智能作为强化权力与主导地位的工具时，那么犯罪分子也开始将这种技术武器化，首当其冲的正是黑客和网络犯罪团体。
The problem is that it currently doesn't take much to fool an AI. Simply adding strips of black and white ape to red stop signs (in some cases spelling out the words 'love' and 'hate') – as a group of nine researchers did in 2018 in a virtual simulation – can be sufficient to fool a self-driving car into assuming that the speed limit is 45 miles per hour. Neural networks have been tricked into misinterpreting a picture of a turtle as a rifle,4 a panda as a gibbon,5 a school bus as an ostrich,6 and a coffee pot as a macaw.7 A particular pattern applied to the frame of a pair of glasses has proved sufficient to cause facial recognition software to make an incorrect identification.	问题在于目前欺骗人工智能并不难。2018 年，由九位学者组成的实验团队在数字仿真环境中揭示：仅需在红色停车标志表面施加黑白条纹干扰（部分案例中甚至嵌入情感暗示文字），就能诱导自动驾驶系统将"停车禁令"误判为"时速 45 英里"的通行许可。神经网络曾被误导将海龟识别为步枪、熊猫认作长臂猿、校车视为鸵鸟、咖啡壶看成金刚鹦鹉。实验证明，特定图案的眼镜框架足以导致人脸识别系统产生错误判定。
These errors occurred in research conditions. But cybercriminals and hackers always work at the cutting edges of technology and innovation – it's the only way to stay in front of law enforcement officials. It won't be long before they turn	这些错误是在研究条件下发现的，但网络犯罪分子和黑客始终活跃于技术创新最前沿——这是他们规避执法追捕的关键。转向 AI 武器化只是时间问题。对某些犯罪集团而言，AI 将成为攻击工具而非目标：利用 AI 扫描应用

to using AI. For some AI will be the tool rather than the target. For instance, criminals may use AI to help find flaws in an app's code (app developers also do this to pre-emptively find errors). They may employ AI to create new malware or attacks (polymorphic malware can constantly change itself to evade detection). Or they may program AI to mimic people's behaviour as part of phishing attacks that con people into believing an email or request for money is being sent from someone they trust.	程序代码漏洞（开发者同样用此预检错误）、设计能自我变形的恶意软件以逃避侦测，甚至训练 AI 模仿人类行为实施精准钓鱼攻击，诱使受害者相信钓鱼邮件来自可信对象。
In some respects, the cybersecurity industry is ahead of criminals, as it has long been using AI tools that can detect anomalies, such as malware, in a corporate network that could cause extensive damage to data and systems. Google, for example, started using machine learning in 2017 to help spot malicious Android apps in its PlayStore. Emily Wenger from the University of Chicago's SAND Lab, along with other researchers, used AI to create a privacy-preserving tool, Fawkes, that adds data to people's photos in a process called image cloaking. This added data is invisible to the human eye but can confuse facial recognition databases. 'The idea is	网络安全产业在某些维度暂居上风，其长期部署的 AI 威胁检测系统可识别企业网络中的恶意软件等异常行为。例如，谷歌早在 2017 年即运用机器学习筛查应用商店中的恶意安卓应用；芝加哥大学 SAND 实验室的艾米丽·温格 (Emily Wager) 团队开发隐私保护工具 Fawkes，通过“图像隐身”技术在人像照片中嵌入肉眼不可见的干扰数据，扰乱面部识别数据库训练。“当用户上传处理后的照片至社交平台，若被爬虫抓取用于非法训练面部识别模型，模型将学习到失真的面部特征，”温格解释道，但犯罪集团既有动机亦握有技术手段实现反制。

that if you post these cloaked photos online and someone scrapes them and tries to use them to train an unauthorised facial recognition model, that model will learn this bogus representation of you,' Wenger says. But criminals and malicious actors have the incentive – and the tools – to catch up.	
As well as using AI to attack conventional computer systems, AI can be employed to attack other AI. The risks and stakes here are high, and malicious uses of the technology are hard to detect. Essentially it can take one of two forms: adversarial attacks and Trojan attacks. With an adversarial attack 'perturbations' – or pixels that aren't visible to the human eye – are added to an image to make the AI misinterpret what it is seeing. (This uses GAN-style technology, which was first created by Google researchers in 2014 when they published a paper that described how adding a 'perturbation' to an image made it perform incorrectly. 'Perturbations' are, effectively, a more sophisticated version of the stop sign phenomenon described above. As Dawn Song, a professor of computer science at the University of California, Berkeley, who	除攻击传统计算机系统外，AI 还可被用于攻击其他 AI 系统，此类攻击风险和收益都极高且难以侦测。基本上，这种攻击形式可以分为两种：对抗性攻击和木马攻击。对抗攻击通过向图像注入人类不可见的扰动像素（利用 2014 年谷歌研究者提出的生成对抗网络技术），诱导 AI 误判图像内容。"对抗样本恰恰揭示我们对深度学习运作机制及其局限性的认知仍极为有限，"参与交通标志篡改项目的加州大学伯克利分校教授宋丹丹（Dawn Song）2018 年向《WIRED》坦言。

was involved with the street sign manipulation project, told WIRED in 2018: 'Adversarial examples are just illustrating that we still just have very limited understanding of how deep learning works and their limitations.'	
Trojan attacks, by contrast, involve planting malicious data in one piece of AI that might well end up being used by a different one. So, for example, a company might purchase an image-recognition system from another firm that deals with the training and creation of AI, not knowing that it has already been hacked. Only when they – or an end user even further down the supply chain – deploy it will the problems surface. 'You'll probably have a bunch of test data yourself,' says Wenchao Li, an assistant professor at Boston University who has exposed how Trojan attacks can impact reinforcement-learning algorithms. 'You've tested the model and it seems right on this test dataset. And now you can deploy it. But the problem is, the network that you got from somebody else actually has a Trojan in it.'	相比之下，木马攻击则截然不同——将恶意数据植入某一 AI 系统，而这些数据最终可能被其他 AI 系统使用。例如，某公司可能从另一家从事 AI 训练与开发的企业购买图像识别系统，却不知该系统已被入侵，直至终端用户部署时问题才爆发。波士顿大学助理教授温超·李（Wenchao Li）揭示了木马攻击如何影响强化学习算法。他表示："你可能会自己有一堆测试数据，也已经对模型进行了测试，在这个测试数据集上表现良好，也可以部署了。但问题是，你从他人处获得的神经网络中早已暗藏木马。"

Such systemic threats have been a part of the digital world for years. The laptop, computer and phone manufacturer ASUS, for example, suffered a supply chain attack in 2019 after a server for the company's software updates was compromised and malware inserted into the legitimate update that was sent out. About 500,000 devices were hit by the malicious update, which went unnoticed in ASUS's updates for around five months. AI Trojan horses are more complex. That, however, doesn't make them any less feasible.	这类系统性威胁已经存在于数字世界多年。以华硕（ASUS）为例，该电脑设备制造商于 2019 年遭遇供应链攻击。其软件更新服务器遭入侵，致使恶意程序被植入官方推送的系统更新包。大约 50 万台设备受到了这一恶意更新的影响，而这一问题在华硕的更新中持续了大约五个月才被发现。人工智能的特洛伊木马更加复杂，但这并不意味着人工智能不具备可行性。
To help combat such backdoor attacks against AI systems the US Intelligence Advanced Research Projects Activity (IARPA) has come up with a project called TrojAI whereby researchers compete to find effective methods to detect Trojans. 'For a Trojan attack to be effective the trigger must be rare in the normal operating environment, so that the Trojan attack to be effective the trigger must be rare in the normal operating environment, so that the Trojan does not activate on test datasets or in normal operations, either one of which could raise the suspicions of the AI's users,' IARPA says on its website. Those working on the project have to try to detect	为应对针对 AI 系统的后门攻击，美国情报高级研究计划局（IARPA）启动了名为 TrojAI 的项目，通过竞赛形式推动研究者开发木马检测方案。IARPA 官网明确指出："要使木马攻击生效，触发条件必须在常规操作环境中极为罕见，确保既不会在测试数据集激活，这样才能确保木马不会在测试数据集或日常运行中被激活，否则可能引发 AI 使用者的警觉。"项目参与者需在无法接触训练数据和模型底层机制的情况下，识别被植入木马的 AI 系统。

an AI system that has a Trojan installed, without having any access to its training data or the underlying mechanisms behind the AI.	
Detecting these types of attacks is difficult. 'There's an argument for an attacker to poison only a tiny fraction of the data, because, let's say, this one bad player needed a team, and you could have a vigilant vendor monitoring the training process,' says Boston's Li, who is part of one of the teams in the IARPA project.	检测此类攻击极具挑战。作为波士顿大学 IARPA 的一员，李文超表示："攻击者只需污染极小部分的数据，假设恶意行为者组织严密，即便供应商对训练过程实施严密监控，也难以察觉微量污染。"
At present, the TrojAI project is in its early stages, but those behind it say it shows promise. However, inspecting for a Trojan involves deep analysis of the results of an AI system, and being able to undertake this requires another AI that's able to carefully classify many different outputs of the corrupted system and spot how it reaches its decisions. Such complexity inevitably gives potential hackers room for manoeuvre. The battle between legitimate AI users and hackers will be a continual one.	当前 TrojAI 项目虽处于早期阶段，但研发团队对其前景充满信心。不过，检测木马需要对 AI 系统的输出结果进行深度解析，且要做到这一点，需要另一套能够精确分类海量异常输出、追踪决策路径的 AI 系统。如此复杂的架构必然为黑客预留操作空间，合法 AI 用户与攻击者的博弈必将长期存在。
7 Making AI accountable	第七章 构建 AI 问责制

Silicon Valley's biggest technology companies – Google, Amazon, Facebook, Netflix, Microsoft and Apple – have driven the development and commercialisation of AI throughout the 2010s. They have invested billions to build the infrastructure required to process AI decisions and then deploy AI in their products. They have invested heavily in securing the vast amount of data their AI systems are trained on, processing it in huge data centres and deploying immense computing power to crunch it. And they have spent a small fortune on attracting and keeping the best possible talent.	硅谷科技巨头——谷歌（Google）、亚马逊（Amazon）、脸书（Facebook）、奈飞（Netflix）、微软（Microsoft）和苹果（Apple）——主导了 2010 年代人工智能的研发与商业化进程，这些企业斥资数百亿构建 AI 决策基础设施，重金保障训练数据安全，在巨型数据中心处理海量信息，部署超强算力进行模型训练，并不惜代价招揽顶尖人才。
Thanks to them our lives have been quietly transformed. Amazon has brought voice-recognition technology into millions of homes with its Alexa-enabled Echo speakers. Apple has pioneered privacy-protecting machine learning which strips away personal details from your interactions with your phone, to ensure people can't be re-identified from the training data it supplies. Microsoft's tech infrastructure sits behind hundreds of other companies' AI systems. Facebook and Instagram feeds are filled with the posts that AI decides we're most likely to	得益于这些公司，我们的生活悄然发生了转变。亚马逊通过搭载 Alexa 的 Echo 智能音箱将语音识别技术引入千万家庭；苹果开创隐私保护型机器学习，剥离用户交互数据中的个人标识符，确保训练数据无法反推个体身份；微软的技术基座支撑着数百家企业的 AI 系统；脸书和 Instagram 的动态推送由 AI 主导，不仅筛选亲友内容，更精准投放广告与品牌信息；网飞基于观看历史，用机器学习引导用户观看特定影视内容；谷歌的大部分产品都依赖人工智能驱动，包括电子邮件和文档中的推荐回复、翻译工具、地图

interact with, not just from our friends and family but also advertisers and brands that people follow. Netflix's machine learning nudges us towards the shows and movies it wants us to watch, based on our past viewing history. The vast majority of Google's products – including suggested replies in emails and documents, its translation tools, directions provided by maps and YouTube recommendations – are fuelled by AI.	提供的路线指引以及 YouTube 的推荐内容。
But there's a problem. The vast majority of AI products are being developed by one particular group: white men. Just as decisions made by AI can be skewed by the datasets that feed them, their basic creation can be skewed by the cultural background and assumptions of those who build it. 'What really matters at the end of the day is who is building and shaping the AI system,' says Tess Posner, the CEO of non-profit AI4ALL. 'Who really is in a position of power to influence what gets built, what's designed, what's created? Right now, that's a very homogenous group of technologists. You can only know so much if you're a group of individuals who represent one type of person and background.'	但问题在于，绝大多数人工智能产品是由一个特定群体—白人男性—主导开发。正如 AI 决策可能受训练数据偏见影响，其基础架构同样受开发者文化背景与预设观念的形塑。非盈利组织人工智能普及联盟（AI4ALL）的首席执行官苔丝·波斯纳（Tess Posner）说："归根结底，真正重要的是谁在构建和塑造人工智能系统。谁掌握着决定技术方向、设计与创造的话语权？当前，这是一个高度同质化的技术专家群体。当开发者群体仅代表单一身份背景时，其认知边界必然受限。"

Ironically for a sector that is so heavily reliant on data, the world of AI is remarkably coy when it comes to revealing data about the nature of its workforce. Technology companies, universities and other academic institutions alike have traditionally been reluctant to publish details. Things are slowly changing for large companies in some countries, but they're still able to control the narrative that accompanies the statistics they issue. Overall, there's a distinct lack of detail about such metrics as gender, sexual orientation or ethnicity.	讽刺的是，在极度依赖数据的 AI 领域，行业对自身人才结构的数据披露却讳莫如深。科技公司、大学和其他学术机构通常不愿公布相关细节。虽然一些国家的大型公司情况正在慢慢改变，但它们仍然能够控制伴随统计数据发布的叙述。总体上，关于性别、性取向或种族等维度的统计数据严重缺失。
Even so, the general picture is clear. On average 80 per cent of AI professors are male, according to the 2018 AI Index, which looked at such top computer science schools as University College London, Stanford, CMU, Berkeley, Oxford, ETH Zurich and the University of Illinois' Urbana-Champaign campus (the researchers noted that there was 'little variation' between the schools inspected).1 Research from Nesta2 and the World Economic Forum (WEF)3 suggests that the proportion of academic AI papers co-authored by at least one woman 'has not improved since the 1990s' (analysis by	即便如此，整体图景已然清晰。根据 2018 年《AI 指数》（*AI Index*）报告，在顶尖计算机科学学校（如伦敦大学学院、斯坦福大学、卡内基梅隆大学、加州大学伯克利分校、牛津大学、苏黎世联邦理工学院以及伊利诺伊大学厄巴纳-香槟分校）中，平均 80%的人工智能教授为男性（研究人员指出这些学校之间的差异很小）。Nesta 和世界经济论坛（WEF）的研究显示，自 1990 年代以来，至少有一位女性合著的 AI 学术论文比例例并未改善 (WEF 分析指出德国、巴西、墨西哥、阿根廷性别差异最大）。谷歌的黑人员工仅占 2.5%，而脸书和微软的这一比例

WEF shows Germany, Brazil, Mexico and Argentina to have the largest gender gaps). Only 2.5 per cent of Google's workforce is black, while for Facebook and Microsoft it's 4 per cent – and those percentages are total ones: there is little or no data available for roles that are specifically concerned with AI, or that offers a detailed breakdown of the minorities involved.	为 4%。更糟糕的是，这些数据为全公司范围统计，专门从事 AI 岗位的族裔细分数据几近空白。
Not surprisingly, a 2019 report on gender, race and power in AI from New York University's AI Now Institute that discussed these numbers concludes that 'The AI industry is strikingly homogeneous, due in large part to its treatment of women, people of colour, gender minorities, and other under-represented groups.' It continues: 'The AI industry needs to make significant structural changes to address systemic racism, misogyny, and lack of diversity.' DeepMind researchers Shakir Mohamed, Marie-Therese Png and William Isaac view the problem in colonial terms.6 'Any commitment to building the responsible and beneficial AI of the future ties us to the hierarchies, philosophy and technology inherited from the past, and a renewed responsibility to the technology of the	不出所料，纽约大学 AI Now 研究所在 2019 年发布的一份关于性别、种族和权力的报告指出，“AI 产业呈现惊人的同质性，这主要源于其对女性、有色人种、性别少数群体及其他弱势群体的系统性排斥。”报告进一步指出：“AI 产业需进行根本性结构变革，以应对制度性种族主义、厌女症及多样性缺失。”DeepMind 研究人员沙基尔·穆罕默德（Shakir Mohamed）、玛丽-特蕾莎·潘（Marie-Therese Png）与威廉·艾萨克（William Isaac）以殖民视角剖析此问题：“任何构建负责任、有益 AI 的承诺，都让我们与承袭自过去的技术等级体系、哲学范式紧密捆绑，同时赋予我们重塑当下技术伦理的新责任。”

present,' the trio write.	
The colonial nature of AI takes various forms. It's embedded in its algorithmic discrimination, which allows an unequal real world to be reflected in an unequal virtual one. It runs through the sector's hiring practices. It involves the widespread use of low-paid labour in relatively poor parts of the world among those at the beginning of the AI production chain who prepare data for training. People benefit financially from the jobs they do, of course, but it's questionable whether their communities are likely to benefit from the technologies they're helping to create. And then there's the fact that AI governance rules tend to be set by richer countries. Mohamed, Png and Isaac highlight a global review of AI ethics guidelines that found Africa, South and Central America and Central Asia to be absent from discussions. They also point out that AI systems can be beta-tested in countries where data protection laws aren't as strong as they are	AI 的殖民性呈现多维形态：它深嵌于算法歧视之中，使现实世界的不平等在虚拟空间复现；贯穿于人才选拔机制，让少数族裔与女性难入核心决策层；更体现在全球产业链底端——那些为 AI 训练数据做标注的低薪工作者往往来自相对贫困地区。尽管从业者获得经济收益，但其所在社区能否从参与创造的技术中获益仍存疑。此外，AI 治理规则制定权始终掌握在发达国家手中。穆罕默德等人指出，全球 AI 伦理准则制定过程中，非洲、中南美与中亚地区处于失语状态。更值得注意的是，AI 系统常在数据保护法薄弱的国家进行测试，形成“伦理套利”空间。

in, say, the UK or Europe.	
Such systemic injustices are manifested in individual experiences: people who have been under-paid for the job they do, or not received the job promotion they deserved, or been treated disrespectfully by colleagues. Recent years have seen thousands of Google employees protest at their companies' handling of sexual harassment allegations against senior executives and failing to properly investigate discrimination claims, in addition to hundreds of Microsoft employees claiming they had been victim to harassment and discrimination while at work. In December 2020, leading AI ethics researcher Timnit Gebru said Google fired her and accused the company of being 'institutionally racist' after it took issue with an academic paper she co-authored. More than 6,000 AI people, including several thousand Google staff, signed an open letter supporting Gebru. In response Google CEO Sundar Pichai apologised for the controversy and said the company needed to 'accept responsibility for the fact that a prominent black, female leader with immense talent left Google unhappily'. Consistent failings at the heart of AI	此类系统性不公正具体投射于个体遭遇：同工不同酬者、晋升通道受阻者、遭受同事不公待遇者比比皆是。近年来，数千名谷歌员工抗议公司高层性骚扰指控处理不当及歧视投诉调查失职，另有数百名微软员工声称遭遇职场骚扰与歧视。2020 年 12 月，顶尖 AI 伦理研究者蒂姆尼特·格布鲁（Timnit Gebru）指控谷歌因其合著学术论文提出异议将其解雇，并斥责公司存在"制度性种族主义"。超 6000 名 AI 从业者（含数千谷歌员工）联署公开信声援格布鲁。对此，谷歌 CEO 桑达尔·皮查伊（Sundar Pichai）公开致歉，坦言公司必须正视，顶尖非裔女性高管愤而离职一事。人工智能开发过程中持续的失败，强化了这样一个观点：为了让技术有机会为所有人服务，它必须由一个多样化且具有代表性的人群来构建和掌控。

development reinforce the argument that to have a chance of working for everybody, the technology needs to be built and controlled by a diverse and representative set of people.	
'Diverse' and 'representative' means also taking account of the more than one billion people around the world who live with disabilities. AI has helped create new forms of accessibility – for example, computer vision apps that can 'see' for people. But there is a long way to go to before it's truly inclusive. 'I think the most important step in inclusive AI is involving people with disabilities at all stages,' says Mary Bellard, the principal innovation architect and AI for Accessibility program lead at Microsoft. Such inclusion involves more than testing products on people with disabilities. It has to include hiring technologists with disabilities, too. 'Sometimes there is this concept of building technology "for people with disabilities", instead of "with or by people with disabilities", and there is an important distinction between those two ideas,' Bellard says.	"多样化"和"具有代表性"还意味着要考虑全球超过 10 亿残障人士。人工智能已经帮助创造了新的无障碍形式，例如能够"为人们看到"的计算机视觉应用程序。微软 AI for Accessibility 项目负责人玛丽·贝拉德（Mary Bellard）表示："我认为包容性人工智能最重要的一步是在所有阶段都让残疾人士参与其中。"这种包容性不仅仅是让残障人士测试产品，还必须包括雇佣残疾技术人员。贝拉德指出："有时人们会认为是在为残疾人士开发技术，而不是与他们或由他们开发，这两者之间存在重要的区别。"
But there are people and organisations around that are trying to change things.	不过，一些个人和组织正在努力改变现状。苔丝·波斯纳创立的"人工智能

One such is Tess Posner's AI4ALL, the US non-profit set-up that is seeking to get more young women and BAME people interested in AI. Among its initiatives are summer camps for high-school students. 'We hear a lot of things like, "I'm the only person at my school who looks like me who's interested in this," or "I never thought I can go into this field because I don't see anyone like me in the field,"' says Posner. The summer camps are designed to introduce such students to AI role models. PhD students and experts from big technology firms explain the main principles of AI: the types of machine learning, how to use the programming language Python, the importance of datasets within AI. AI4ALL also provides open-sourced AI teaching resources to high-school teachers around the US, targeting them at lower-income schools, which may not typically have the money to invest in computer science-specific education for their students. Posner aims to get the teaching resources translated into other languages so they can be disseminated more widely both within and outside the US.	普及联盟（AI4ALL）"便是典型代表，这个美国非营利教育机构致力于激发更多年轻女性及黑人、亚裔与其他少数族裔群体（BAME）对人工智能的兴趣，包括为高中生开设人工智能主题暑期学术营。波斯纳说："我们在高中生夏令营中反复听到这样的心声——'全校找不到第二个与我外貌背景相似却热衷此领域的人'，或'因未见同肤色的先行者，从未敢奢望踏入此领域'。"这些夏令营旨在为学生们介绍人工智能领域的榜样。来自大型科技公司的博士生和专家向他们讲解人工智能的基本原理：机器学习的类型、如何使用 Python 编程语言、数据集在人工智能中的重要性。人工智能普及联盟还为美国的高中教师提供开放源代码的人工智能教学资源，针对那些通常没有资金为学生投资计算机科学教育的低收入学校。波斯纳的目标是将这些教学资源翻译成其他语言，以便在美国境内外更广泛地传播。
AI Now's diversity report looks at ways in	AI Now 的多样性报告探讨了技术行

which the technology industry in general and AI in particular can make changes. It argues for the publication of pay levels, including bonuses and all other payments, across all roles and jobs, broken down by race and gender. It pushes for pay equality to be established to ensure people are being properly compensated for the work they're doing, and for the promotion of currently under-represented groups to be built into company targets. It also demands that harassment and discrimination transparency reports should be published. Perhaps most crucially, it campaigns for a shake-up in the way people are recruited, suggesting that initiatives should be set up to reach out to non-elite universities, and that better ways should be found to allow contractors, temporary staff and freelancers to become full-time employees. Hiring practices and conditions should also be made transparent, so that everyone knows how individuals are hired, paid and promoted. Responsibility for all this begins with senior management. 'Those at the top of corporate hierarchies have much more power to set direction and shape ethical decision-making than do individual researchers and developers,' says AI	业，尤其是人工智能领域，如何进行变革。报告主张公开所有角色和职位的薪酬水平，包括奖金和其他所有支付项目，并按种族和性别进行细分。报告推动建立薪酬平等，以确保人们得到与其工作相称的报酬，并建议将提升当前弱势群体的地位列入公司目标。此外，报告还呼吁发布骚扰和歧视的透明度报告。或许最关键的是，这场变革的核心诉求在于推动企业招聘机制革新——建议建制化非顶尖学府人才输送专项计划，破除劳务派遣人员、临时雇员及自由职业者的转正壁垒；同时强制实行招聘流程透明化工程，确保薪酬、晋升机制全程可溯。此类系统性变革的权责溯源，必须始于高层管理者主导的自上而下的制度重构。报告指出："在企业层级中的高层管理人员比个别研究人员和开发人员有更大的权力来设定方向并塑造道德决策。"

Now’s report.	
Transparency and regulation	透明度和监管
Joy Buolamwini has shown accountability in AI can lead to change. In 2015 the graduate student of MIT’s Media Lab was sitting at her laptop, trying to get the generic face-tracking software she was using to work. The problem was that it wouldn’t. It was supposed to project digital artwork onto her face, but it couldn’t ‘see’ her. When, however, her friend sat in front of the cheap laptop webcam the system automatically sprang to life. The reason? Buolamwini is black. Her friend is white. Once Buolamwini donned a white mask, the software instantly spotted her.	乔伊·布拉姆维尼（Joy Buolamwini）的经历展示了人工智能中的问责制如何带来改变。2015 年，麻省理工媒体实验室的一位研究生发现，通用型面部追踪软件存在系统漏洞。程序本应实时捕捉面部特征并投射虚拟影像，却因肤色识别算法缺陷导致运行中断。然而，当她的白人朋友坐在相机前时，系统立即启动。原因是什么？布拉姆维尼是黑人，而她的朋友是白人。后来，当她戴上一个白色面具时，软件立即识别了她的脸。
This wasn’t the first time Buolamwini had encountered the discrimination arising from algorithms trained mostly on images of people with white skin. She had once tried to teach a robot how to play peek-a-boo. It didn’t respond. Again, she was invisible.	这并不是布拉姆维尼第一次遭遇算法歧视，这些算法主要是基于白人皮肤图像进行训练的。她曾试图教一台机器人玩躲猫猫，但机器人没有回应，她又“隐身”了。
Five years after Buolamwini sat in front of her webcam, her research has changed how facial-recognition technologies are developed, and brought the issue of selection bias in training data to general awareness. Her 2018 MIT thesis, ‘Gender	五年后，布拉姆维尼的成果彻底改变了面部识别技术的开发范式，并让公众普遍意识到训练数据中的选择偏见问题。她与蒂姆尼特·格布鲁（Timnit Gebru）2018 年合作的麻省理工学院论文《性别色调》（Gender Shades），

Shades', which was co-authored by Timnit Gebru, audited the performance of facial-recognition systems from IBM, Microsoft and Chinese firm Face++, and offered a stark wake-up call.	通过审计 IBM、微软和中国企业旷视科技（Face++）的面部识别系统性能，向行业敲响警钟。
What she showed was that, overall, the systems were worse at recognising people with darker skin than they were at recognising those with lighter skin. Microsoft performed best, but even so, its 87.1 per cent accuracy for darker-skinned people compared poorly with its 99.3 per cent accuracy for lighter-skinned subjects. The margin of difference in IBM's software was 19.2 per cent. And when gender was introduced, accuracy rates dropped further. Both IBM and Face++ recognised darker-skinned females correctly only 65 per cent of the time. Microsoft was slightly better at 79 per cent. For lighter-skinned females all three sets of software scored above 92 per cent. The differences in performance were as high as 34 per cent."	布拉姆维尼的研究表明，整体而言，面部识别系统在识别深色皮肤的人时表现较差，而在识别浅色皮肤的人时则表现得更好。尽管微软的表现最好，但其对深色皮肤人群的识别准确率为87.1%，远低于其对浅色皮肤人群的99.3%。IBM 的软件的准确率差异更大，为 19.2%。当性别因素引入时，准确率进一步下降。IBM 和 Face++对深色皮肤女性的识别准确率仅为65%，而微软则为 79%。对于浅色皮肤女性，三个软件的准确率均超过92%。三者的性能差异高达 34%。
Follow-up research a year later – called 'Actionable Auditing'11 – from Buolamwini and Inioluwa Deborah Raji claimed to find similar problems with Amazon's facial-recognition technology	一年后的跟进研究《可操作审计》（Actionable Auditing）由布拉姆维尼和伊尼奥路瓦·黛博拉·拉吉（Inioluwa Deborah Raji）完成，发现亚马逊的面部识别技术存在类似问题（亚马逊对

(the company disputed this). But by this time IBM, Microsoft and Face++ had all improved their algorithms. IBM's error difference dropped from 34.4 per cent to 16.71 per cent; Microsoft's from 20.8 per cent to 1.52 per cent; and Face++'s from 33.7 per cent to 3.6 per cent. 'The study was able to motivate companies to prioritise the issue and yield significant improvements within months,' wrote Buolamwini and Raji.	此表示异议）。但此时，IBM、微软与旷视已改进算法：IBM 的错误差异从 34.4%降至 16.71%，微软从 20.8%压缩至 1.52%，旷视从 33.7%优化至 3.6%。布拉姆维尼和拉吉指出："这项研究促使企业优先解决该问题，并在数月内实现显著改进。"
By holding opaque algorithms and their creators to account, the hugely influential 'Gender Shades' and follow-up study have shown a way forward. They have very publicly exposed the failures of existing AI systems and led to a much-needed public debate, not just about how facial-recognition systems operate, but about how they should be used. Buolamwini has appeared before legislators, policymakers and police officials to explain how the discriminatory practices worked, and how her audits uncovered them. Her work has directly contributed to greater scrutiny of facial-recognition systems. It's also led to them being banned in a number of US states. In June 2020, Amazon, IBM and	《性别阴影》及其后续研究通过问责不透明算法及其开发者，指明了技术纠偏路径。它们公开揭露现有 AI 系统的缺陷，引发关于面部识别技术运行机制与应用边界的必要公共讨论。布拉姆维尼曾向立法者、政策制定者和警察官员解释这些歧视性实践是如何运作的，以及她的审计如何揭示这些问题。她的工作直接促使面部识别系统受到更严格的审查，并导致多个美国州禁止使用这些技术。2020 年 6 月，在全球范围内抗议种族歧视和乔治·弗洛伊德被杀事件后，亚马逊、IBM 和微软均承诺（至少暂时）停止向警方出售面部识别软件。

Microsoft all committed, at least temporarily, to stop selling facial-recognition software to police, following global protests about systematic racism and the killing of George Floyd at the hands of Minneapolis police.	
The ripples of the pioneering work done by Buolamwini and others will spread out over the next few years, as lawmakers and regulators all grapple with the issues that AI throws up. In particular, there will need to be a recognition that constant structured auditing, of the type that Buolamwini undertook, is essential. Not only does it lead to better systems, but it also forces people to address how those systems are utilised and if they should be used at all.	未来数年，随着立法者与监管机构持续应对 AI 引发的伦理挑战，布兰温妮等人的开创性工作将持续产生深远影响。核心启示在于：必须建立类似其研究的系统性审计机制。这不仅提升技术可靠性，更迫使社会直面技术应用的伦理边界——某些系统是否应被允许存在。
At present, uncertainty about the application of AI is very widespread. “Currently we cannot say which local authorities in the UK, for instance, are using an algorithmic decision system to inform prioritisation in children’s social care,” says Jenny Brennan, an AI policy researcher and software engineer at the Ada Lovelace Institute. “What is really clear is that we do need ways for regulators to inspect algorithmic systems.’”The same uncertainty and need for inspection exists	当前，人工智能应用的透明度缺失问题普遍存在。阿达·洛芙莱斯研究所（Ada Lovelace Institute）的人工智能政策研究员兼软件工程师珍妮·布伦南（Jenny Brennan）表示：“我们无法确认英国哪些地方政府正在使用算法决策系统来指导儿童社会服务的优先级划分，”她还表示，“但可以明确的是，监管部门亟需建立算法系统的审查机制。”这种不确定性与监管需求不仅存在于英国，也遍布全球各 AI 应用领域。

across other sectors in which AI is utilised, not just in the UK but across the world.	
“In Amsterdam and Helsinki city authorities have started tackling a lack of transparency by introducing registers for AI used by city departments, says Linda van de Fliert, the programme manager for Amsterdam’s algorithm register.12 The system, which at the time of writing contains just four AI examples operating in fields such as property rental and automated parking controls, seeks to capture the AI via descriptions of the system’s architecture and its datasets. If the AI is developed by a third-party company, they are contractually bound to provide details for the purposes of openness and transparency. Over the coming years, van de Fliert says, non-AI algorithms, such as the simple ones used in spreadsheets, will be added. ‘The register is about trust. Journalists can go to the register, activists can go to the register, and they can look at what the city is doing.’	阿姆斯特丹算法登记项目负责人琳达·范德弗利尔特 (Linda van de Fliert) 表示：“阿姆斯特丹与赫尔辛基市政当局已通过建立政府部门 AI 应用登记制度，着手解决透明度缺失问题。”该登记系统目前（截至本文撰写时）仅包含房地产租赁、自动停车管控等四个 AI 应用案例，通过描述系统架构与数据集构成实现技术透明。若 AI 由第三方开发，合同将强制要求其披露技术细节。范德弗利尔特透露，未来数年将逐步纳入电子表格等非 AI 算法。“登记制度的核心是建立信任——记者、社会活动家均可查阅市政 AI 应用详情。”

For her part, Jenny Brennan pinpoints two approaches for assessing AI systems that are either currently in use or being considered. One is an algorithm audit that includes a bias audit and regulatory inspections. The other is an algorithmic impact assessment, which might include considering the societal impact of an AI before or after it is in use. Her research concludes, though, that there is some way to go before it's clear precisely what the best versions of these systems would look like in practice.	珍妮·布伦南（Jenny Brennan）指出了两种目前正在使用或处于考虑中的评估人工智能系统的方法。其一是包含偏见审查与监管检查的算法审计；其二是算法影响评估，涵盖 AI 投入使用前后的社会影响研判。然而，其研究也指出，这些机制的最佳实践模式仍需探索完善。
There's a lack of clarity around when algorithmic decisions are being used, and there's usually not a way for folks to actually compare notes and see if there is evidence of bias or any of those sorts of issues,' says Alice Xiang, the head of fairness, transparency and accountability research at the Partnership on AI. The organisation is a research body that is backed by more than 100 groups, including Apple, Amazon, Google, Microsoft, Samsung and IBM. 'There's a lot of constraints in terms of the proprietary nature of the data used to train algorithms and also the proprietary nature of the algorithms themselves.	人工智能合作组织 (Partnership on AI) 公平性、透明度和责任研究负责人爱丽丝·向（Alice Xiang）表示，“公众既不清楚算法决策何时被应用，也缺乏渠道验证其是否存在偏见等问题，由于训练算法的数据和算法本身的专有性质，存在许多限制，难以对其进行透明化的审查。”该组织由苹果、亚马逊、谷歌等百余家机构支持，致力于 AI 伦理研究。

“The current opaqueness of AI decisions makes it harder for people to oppose them, Xiang says. True, there was a prominent backlash against Apple’s credit card launch in the US when its algorithms appeared to show bias by offering different credit limits to men and women, but such a reaction is rare. ‘When you have, say, millions of users being affected, even if just a small cluster of folks got together and managed to talk to each other and compare notes, that’s not necessarily really great evidence to go forth with a lawsuit, and even that rarely happens.’	向表示，目前人工智能决策的不透明性使人们难以对其提出异议。诚然，苹果信用卡在美国上线时，曾因算法存在性别歧视倾向——向男女用户提供差异化授信额度，而遭遇强烈抗议，但此类引发公众警觉的算法纠纷，在系统性歧视普遍存在的当下仍属罕见个案。“即使是数百万用户受影响，若只有少数人聚集在一起对比并交流信息，也很难拿出足够的证据提起诉讼，而且类似的情况很少发生。”
There is disagreement as to who should be responsible for audits and assessments. Dozens of AI ethics groups and guidelines have been drawn up, but many of them lack real power. (In one rare pro-active instance in August 2020, a UK police ethics board stopped the development and deployment of flawed AI that was supposed to predict if someone would commit violent crime.) Some AI experts believe existing regulators, such as those that handle financial laws or education bodies, should be responsible for inspecting the AI that falls within the ambit of their current expertise. Others believe	关于谁应负责进行人工智能系统的审核和评估，存在分歧。虽然已经制定了数十个 AI 伦理团体和指南，但许多缺乏实质性权力。（2020 年 8 月，英国警务伦理委员会做出突破性示范——在全球执法部门中率先叫停某存疑犯罪预测系统的开发部署。该人工智能原型机打着“暴力犯罪风险评估模型”的旗号，实则嵌含着可能加剧系统性歧视的算法隐患。）一些 AI 专家认为，现有的监管机构，例如处理金融法律或教育事务的机构，应该负责检查其专业领域内的 AI 系统。而另一些专家则认为，应该建立全新的监管机构。实际上，最有可能的是采用混

entirely new bodies should be set up. In practice, a mix of approaches seems the most likely.	合的方法。
"In July 2020 the government of New Zealand launched a set of standards that it called an algorithmic charter and a 'world first'. Twenty-one government agencies pledged to be transparent about how they use algorithms in decision-making. They undertook to give 'plain English' explanations of how systems worked, and promised to make efforts to manage any biases that might be informing them. In addition, the charter stipulates that algorithmic decision-making systems should be regularly peer-reviewed, that humans should retain oversight of the systems, and that there should be public access to government agencies deploying them. Agencies must also consider the views of Te Ao Māori, or Indigenous people, on data collection, and consult with people whose lives are affected by the algorithms."	2020 年 7 月, 新西兰政府颁布名为"算法宪章"的标准体系，自诩为"全球首创"。21 家政府机构公开承诺，将就其决策算法建立透明机制：通过通俗易懂的非技术语言，披露系统运作原理，并主动管控可能潜藏于代码深处的结构性偏见。宪章还规定：算法决策系统需定期接受同行评审、保持人类监督权、向公众开放部署信息。政府部门须考虑毛利原住民（Te Ao Māori）对数据采集的意见，并与受算法影响群体进行协商。
"There are so many sources of bias: individuals, structural, societal, it seems dangerous and a bit arrogant to think that we can fix all this with one system,' argues Suresh Venkatasubramanian, a professor in	犹他大学计算机学院教授苏雷什·文卡塔苏布拉马尼安（Suresh Venkatasubramanian）表示："偏见有太多来源：个人的、结构性的、社会的。认为我们能用一个系统解决所有

the School of Computing at the University of Utah. He says that each of the potential risks a system poses should be identified, and that it should be audited for them one at a time. 'The goal is to narrow the scope of the harms to the point where we can actually deliver an audit that makes sense for that,' he explains, 'and then slowly try to build up from there.' Audits have to be done at the right stage of the development process, too. 'Often the use of audit risk compliance teams are the last people anybody calls,' says Rumman Chowdhury, the former head of Accenture Applied Intelligence's Responsible AI division, who left to create the algorithmic audit company Parity. 'Ethical assessment is not something you can just do once. There are certain actions you would take at different stages. It wouldn't even be appropriate to do, for example, a data audit at the very end of product development.' Crucially, any audit has to evolve over time, as new problematic issues are identified or if the training data or algorithms used in the AI application change.	这些问题，这显得既危险又有些自负，"他主张逐一识别系统潜在风险并进行专项审计，"目标是将危害范围缩小至可审计范畴，再逐步拓展审查维度。"前埃森哲负责任 AI 部门负责人拉曼·乔杜里（Rumman Chowdhury）强调审计需贯穿开发全周期："伦理评估不能一劳永逸，不同阶段需采取特定措施。比如在产品末期才进行数据审计就极不妥当。"最关键的是，审计必须随着时间的推移而不断演变，当发现新的问题，或者 AI 应用中的训练数据或算法发生变化时，审核程序也要随之调整。

In the absence of widespread legislation and auditing that can expose threats, discrimination and biases in AI systems being used around the world, Buolamwini's non-profit Algorithmic Justice League is taking measures into its own hands. In the world of cybersecurity ethical hackers that report flaws or vulnerabilities in software can be paid for their work through bug bounty systems. The Algorithmic Justice League is learning from this process by running an algorithmic vulnerability bounty project – people can disclose how they have been treated by AI and it will be investigated. Ultimately, Buolamwini believes that to develop AI that can be a benefit to all, everyone has to be involved in its creation. 'It's so easy to think you might introduce technology to solve other people's problems without actually bringing the people who are most impacted into the community. So [it's] not designing for but designing with.'	在缺乏全球性立法与审计机制揭露 AI 系统威胁的现状下，Buolamwini 创立的非营利组织"算法正义联盟"正在采取行动。借鉴网络安全领域的"漏洞悬赏"模式，该联盟启动"算法脆弱性悬赏计划"——公众可申报遭遇的 AI 不公待遇，联盟将展开调查。Buolamwini 认为，要构建普惠性 AI，必须让所有人参与技术创造。"人们很容易产生这样的想法：你可以引入技术来解决人们的问题，而不将那些受影响最大的人纳入社区。所以，与其为某些群体设计，不如与这些群体一起设计。"

注：划线部分为报告正文中选取的案例。