

超越 ChatGPT:生成式 AI 的机遇、 风险与挑战

陈永伟

摘要:以 ChatGPT 为代表的生成式 AI 的崛起是人工智能领域的一次重大变革,也可能成为由专用性人工智能转向通用性人工智能的关键转折点。生成式 AI 的爆发由算法、算力和数据的进步共同推动,其中算法层面的突破最为关键。生成式 AI 在消费端的应用场景主要有内容生产、便捷交互、简化操作,这些应用将会同时对消费者的消费品数量、质量、多样性,以及拥有的闲暇等因素产生显著影响。生成式 AI 对产业端的影响主要体现为加速自动化、通过“组合式创新”促进技术进步、实现对新要素尤其是新数据要素的创造,靠近生产领域的代表性应用有工业设计、药物研发、材料科学、合成数据等。生成式 AI 带来的问题主要有失业问题、收入分配和不平等问题、竞争和垄断问题、知识产权问题、安全和隐私问题、道德和伦理问题、能源和环保问题。为此,应从产业政策、就业和保障政策、法律法规建设等方面制定有针对性的措施。

关键词:生成式 AI; AIGC; ChatGPT

DOI: 10.19836/j.cnki.37-1100/c.2023.03.012

2022年11月30日,美国人工智能公司 OpenAI 发布了大型语言模型 ChatGPT。这款模型不仅可以跟据上下文与用户进行对话,还可以完成包括文本写作、计划编制甚至编程在内的多种工作。模型一经发布,就在全球范围内受到了广大用户的追捧,上线五天用户就突破了百万,上线不到两个月用户就超过了一亿。从类别上看,ChatGPT 属于“生成式 AI”(Generative AI)的范畴。不同于过去的“分析式 AI”(Analytical AI),这类人工智能模型不仅可以通过对数据的学习来提炼信息、预测趋势,而且可以生成不同于学习样本的新内容。随着技术的日渐成熟,市场上已经有了很多不同的生成式 AI 产品(见表1)。它们不仅为我们的生活带来了更多的便利性和多样性,也为生产力的提升带来了很大的想象空间。本文将从技术基础出发,对生成式 AI 的发展进行简要的介绍,并分析它的应用潜力、经济影响,以及可能存在的问题和风险。在此基础上,本文还将对有关生成式 AI 的政策进行一些探讨。

一、生成式 AI 的技术基础

生成式 AI 的爆发是算法、算力和数据的进步共同推动的,其中,算法层面的突破是最为关键的因素。

(一)生成式 AI 的主要训练算法

从根本上看,生成算法的本质就是对训练样本的分布状况进行建模,然后根据模型来抽取新的样本^①。现在的生成式算法有很多,比较流行的有五大类模型:自回归模型、生成式对抗模型、变分自编码模型、流模型和扩散模型。

作者简介:陈永伟,《比较》研究部主管,研究员(北京 100029; chen Yongwei1982@126.com)。

^① Bond-Taylor S., et al., “Deep Generative Modelling: A Comparative Review of Vaes, Gans, Normalizing Flows, Energy-based and Autoregressive Models”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

表1 部分著名的生成式AI产品

模型名	用途	研发公司	发布时间	参数量
GPT-3	语言理解与生成	OpenAI	2020年5月	1750亿
ChatGPT	语言理解与生成	OpenAI	2022年11月	未知
BERT	语言理解与生成	谷歌	2019年10月	4810亿
PaLM	语言理解与生成	谷歌	2022年4月	5400亿
OPT-175B	语言理解与生成	脸书	2022年5月	1750亿
LLaMA	语言理解与生成	脸书	2023年2月	650亿
Turing-NLG	语言理解与生成	微软	2020年2月	170亿
Megatron-Turing NLG	语言理解与生成	微软、英伟达	2021年10月	5300亿
Gopher	语言理解与生成	DeepMind	2021年12月	2800亿
Dall-E2	根据文字生成图片	OpenAI	2022年9月	120亿
IMAGEN	根据文字生成图片	谷歌	2022年5月	200亿
Stable Diffusion	根据文字生成图片	Stability AI	2022年8月	未知
Midjourney	根据文字生成图片	Midjourney	2022年3月	未知
VisualGPT	根据图片生成文字	OpenAI	2021年2月	未知
Flamingo	根据图片生成文字	DeepMind	2022年5月	700亿
Phenaki	根据文字生成视频	谷歌	2022年10月	18亿
MusicLM	根据文字生成音乐	谷歌	2023年1月	未知
AlphaCode	代码生成	DeepMind	2022年2月	414亿
CodeX	代码生成	OpenAI	2021年8月	120亿
AlphaFold2	预测蛋白质折叠结构	DeepMind	2021年7月	9300万

资料来源:表格内容为作者根据公开资料整理。

1. 自回归模型。自回归模型(autoregressive model, ARM)是一种“古老”的生成模型,其原理十分直观:它并不对决定训练数据分布状况的潜变量进行假设,而是直接根据它们的表征状况来对需要生成的内容进行推断^①。尽管这种生成思路很简单,但它在处理文本生成、翻译等问题时却有不错的表现,包括ChatGPT在内的不少大型语言模型都是利用ARM训练的。

2. 生成式对抗网络。最早的生成式训练算法是“生成式对抗网络”(generative adversarial networks,简称GAN)。GAN是一种无监督训练模型,其基本思想是通过构建一个生成器和一个判别器来进行对抗^②。其中,生成器负责根据其对本样本数据的学习结果生成新数据,其目标是要让生成的数据和原样本数据足够相似,从而可以骗过判别器;而判别器则负责对数据进行辨别,其目标是尽可能识别出生成的数据。通过生成器和判别器之间的不断对抗进行参数调试,就可以让生成的内容不断接近训练样本。不过,GAN的问题也很明显:不仅其模型本身缺乏可解释性,而且不能保证生成

① Bengio Y., Ducharme R., Vincent P., “A Neural Probabilistic Language Model”, *Advances in Neural Information Processing Systems*, 2000, 13, pp.932-938.

② Goodfellow I., et al., “Generative Adversarial Nets”, *Advances in Neural Information Processing Systems*, 2014, 2, pp.2672-2680.

数据和训练数据来自同一分布,因而其生成的内容具有不稳定性。

3. 变分自编码。在 GAN 出现的同一时期,另一些学者提出了另一种自无监督模型——“变分自编码”(Variational Auto-Encoder,简称 VAE)方法^①。这个模型由编码器和解码器两个部分组成。编码器通过神经网络从训练数据中采集决定其性质的隐变量的均值向量和方差向量信息,解码器则通过这些信息来生成新的内容。通过对比生成数据和训练数据,模型就可以不断对隐变量的均值向量和方差向量进行重新迭代,从而提升生成内容的质量。相比于 GAN,VAE 方法不仅有更强的解释性,而且可以更好地保证生成数据和训练数据来自相同的分布,从而在模型表现上更为优越。

4. 流模型。相比于 GAN 和 VAE,流模型是一种思路更为直接的生成模型^②。其原理是,直接计算出决定数据表征的潜变量的分布(一般假定为正态分布)与数据表征的分布之间的转移函数,用这个转移函数的反函数去计算出潜变量分布状况。在此基础上,就可以利用这些信息生成出新的内容。由于流模型会计算出决定数据表征的潜变量分布,因此它可以保证生成的内容和训练数据来自同一个分布,并且在模型的可解释性上有很好的表现。不过,相对于前两种模型,流模型对计算的要求更为严格,因此对算力的需求也更大,这在一定程度上限制了其受欢迎程度。

5. 扩散模型。扩散模型是目前比较流行的一种生成模型^③。和 VAE 模型一样,扩散模型会先通过神经网络获得一些关键参数的分布状况。在生成过程中,模型会先执行一个前向过程,在原数据的基础上逐步加入噪声。在每一步中,参数都按照一个马尔可夫过程进行迭代,直到数据成为一个可以近似由纯正态分布噪声表示的数据。然后再紧接着进行后向过程,将刚才得到的数据逐步减噪解码,生成新的数据。以新数据和原数据的相似为目标对模型参数进行不断优化,就可以不断改善生成内容的效果。相比于 GAN 和 VAE,扩散模型的训练更为稳定,可以生成的样本也更为丰富。不过,扩散模型的缺陷也十分明显,由于其训练过程比较复杂,因此其训练时长会比 GAN 长得多,对资源投入的要求也更大。

6. 各种生成模型的性能比较。各种生成模型在表现上有很大的不同。在表 2 中,我们将五类生成模型的主要性能进行了对比。容易看到,每一类生成模型都有其各自的优点和不足,因而其适用的范围也各不相同。

表 2 不同生成模型的性能对比

生成模型	训练速度	生成速度	参数有效性	生成内容的规模	生成质量	生成内容多样性	训练稳定性
自回归模型	慢	慢	低	中	高	高	高
生成式对抗模型	快	快	高	高	高	低	低
变分自编码模型	快	快	高	高	低	高	高
流模型	慢	快	高	中	低	高	高
扩散模型	中	慢	高	中	高	高	高

资料来源:作者根据 Bond-Taylor, et al. (2021)等文献整理。

(二)生成式 AI 的主要训练架构

除了训练方法,促成生成式 AI 爆发的另一个重要突破来自于训练架构,其中最具有代表性的事件

① Kingma D., Welling M., “Auto-encoding Variational Bayes”, *ArXiv Preprint*, 2014, No.1312.6114.

② Dinh L., Krueger D., Bengio Y., “NICE: Non-linear Independent Components Estimation”, *AarXiv Preprint*, 2014, No.1410.8516.

③ Ho J., Jain A., Abbeel P., “Denoising Diffusion Probabilistic Models”, 34th Conference on Neural Information Processing Systems, <https://arxiv.org/pdf/2006.11239v2.pdf>, 访问日期: 2023 年 2 月 27 日。

是Transformer架构的出现。Transformer架构最早被应用于自然语言处理(natural language processing, NLP)场景。在NLP处理中,通常会采用编码器—解码器(encoder-decoder)结构。在较早的实践中,编码器和解码器两端会各接一个循环神经网络(recurrent neural network, RNN)。编码器中的RNN负责将输入序列映射到一个固定长度的向量中,解码器中的RNN则负责将向量映射到一个输出序列中。RNN有很多固有的缺点,例如,它只能按照顺序对序列进行处理,不支持并行运算,因此处理速度很慢。更为重要的是,RNN难以克服“长期依赖问题”,在处理长文本时,会不可避免地丧失之前学习的信息,从而让优化难以进行。为了解决这些问题,人们进行了很多尝试,提出了LSTM、GRU等模型,但这些模型都依然存在着某些缺陷。

2017年,谷歌团队提出了Transformer架构^①。为了解决早期模型存在的问题,Transformer引入了一种“自注意力”(self-attention)机制。利用这个机制,就可以根据每一个词所处的位置来确定它和其他词之间的相关性,从而直接输出最有可能关联的信息^②。与传统的架构相比,Transformer有很多的优势。一方面,它支持对多个词的并行操作,从而可以大幅提升计算效率^③;另一方面,由于Transformer支持同时对多词并行操作,因而输入文本长度就不会对其造成影响。这就使得它可以成功地破解“长期依赖问题”,并可以比传统架构更好地捕捉上下文之间的关系。

虽然Transformer最早只是应用于NLP模型,但由于这种架构的高效性,因此它的应用很快就扩展到了语言、语音、图像和视觉、时间序列分析等众多领域。尤其是在生成式AI领域,Transformer的应用尤其广泛。

(三)从GPT到ChatGPT

所谓GPT,是“生成式预训练”(generative pre-training)模型的简称。2018年,OpenAI基于Transformer架构训练出了GPT-1。在训练方法上,GPT-1主要采用了自回归算法。通过这种方式,GPT就可以较好地根据输入文本来执行相关任务(如进行对话、生成图形、进行翻译等)。

2019年,OpenAI发布了GPT-2。在原理上,GPT-2和GPT-1并没有什么大的区别,但由于其参数和训练样本量要大于GPT-1,因此其性能比GPT-1有所改进。尽管如此,当时的GPT-2性能依然难以令人满意。比较值得注意的是,OpenAI在GPT-2中加入了“零样本设置”(zero-shot setting),这样GPT-2就可以对训练集之外的数据直接进行处理。

2020年,OpenAI进一步推出了GPT-3。虽然从原理上,GPT-3依然没有大的变化,但由于其包含的参数数量和用来训练的数据量都远远超过前两代GPT,因此其性能有了质的飞跃。不仅它本身已经可以流畅地和人进行交流,而且已经可以基于它训练其他的人工智能产品。包括自动编程应用CodeX,以及后来的ChatGPT都是基于它训练而成的。不过,作为完全用无监督方式训练而成的AI产品,GPT-3依然有很多的不足^④。一方面,由于在训练的过程中没有对其进行限制,因此其生成的内容经常过于发散,因而不能满足用户的需要。另一方面,在和用户交互的过程中,GPT-3还经常会出现一些不当言论。这些问题都使得它难以直接投入市场应用。

为了克服GPT-3的不足,OpenAI进一步训练了ChatGPT^⑤。为了保证ChatGPT输出信息的正

① Vaswani A., et al., “Attention is All You Need”, *Advances in Neural Information Processing Systems*, 2017, 30, pp.2702-2712.

② 举例来说,如果我们要进行一个翻译任务,当计算机发现文中有一个bank时,如果它发现上下文中有river这个词,他就会判定这个词和“河”的关系更大,因而就会更倾向于将其翻译为“河岸”,而不是“银行”。

③ 举例来说,如果我们要将一句“I love you”翻译成中文,那么在传统的RNN架构下需要逐一译出这四个词,然后将四个词组合成中文。而在Transformer架构下,则可以根据这次个单词的相对位置同时推断出它们的意思,从而直接将这句话译成中文。

④ Floridi L., Chiriatti M., “GPT-3: Its Nature, Scope, Limits, and Consequences”, *Minds and Machines*, 2020, 30, pp.681-694.

⑤ OpenAI, ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>, 访问日期:2023年2月27日。

确性,OpenAI在训练过程中采用了“人类反馈强化学习”(reinforcement learning from human feedback,RLHF)的思路^①。具体来说,研究人员从GPT-3.5(GPT-3的升级版)中抽取一部分,用人工标注进行训练。对于模型生成的各种结果,都进行人工打分。这样,就可以训练出一个可以对人类偏好进行识别的反馈模型(Reward Model)。接下来,再让反馈模型与原模型进行对抗强化训练,对原模型中不当的输出结果不断进行纠正,从而引导出一个不仅符合人类语言习惯,还符合人类偏好和价值观的语言生成模型。

这里需要指出的是,尽管由于ChatGPT的成功,GPT系列已经被奉为了语言生成模型的标杆。但事实上,GPT并不是唯一的文字生成模型。例如,由谷歌推出的BERT模型就长期与GPT竞争^②。和GPT利用自回归模型从上文推断下文的训练方式不同,BERT采用自编码技术,用类似完形填空的方式根据上下文来生成内容,这让它在一些性状上的表现事实上要优于GPT。

二、生成式 AI 的应用潜力和经济影响

(一)生成式 AI 的应用:一个简要的认识框架

作为“通用目的型技术”(general purpose technology),人工智能在很多领域都能发挥重要的作用^③。在生成式 AI 兴起之前,市场上的 AI 产品主要是分析式 AI。从功能上看,过去的分析式 AI 的应用主要是帮助人们进行预测,进而提升决策效率,因而经济学家们习惯于将这类 AI 称为“预测机器”^④。尽管如此,这类 AI 可以独立完成任务并不多,也不能生成新的内容,在大多数时候,它充其量只能充当一个辅助角色。而生成式 AI 则不同,除了拥有和分析式 AI 类似的预测功能外,它们还能根据决策独立作出判断、生成内容。因此,它们的应用潜力和经济影响要比分析式 AI 大得多。大致上看,我们可以将生成式 AI 的应用场景分为消费端和生产端两块。

1. 消费端。根据经济学理论,影响代表性消费者效用的因素包括:他所消费的产品数量、产品的质量、产品的多样性,以及他所拥有的闲暇等。生成式 AI 的崛起将会同时对以上这些因素都产生显著的影响:随着内容生成的不断便利,无论是消费者可能消费的产品数量、质量还是多样性,都会出现大幅的提升。与此同时,由于生成式 AI 可以帮助人们完成很多繁琐的工作,所以人们可能享受的闲暇时间也将增加。在给定其他条件不变的前提下,以上变化都有助于促进消费者效用的提升。

2. 产业端。在产业端,生成式 AI 的影响体现在三个方面:一是加速自动化;二是通过“组合式创新”促进技术进步;三是实现对新要素,尤其是新数据要素的创造。

在较早的研究人工智能的经济学文献中,自动化一直是作为一个重要的作用机制被关注

^① Ouyang L., et al., “Training Language Models to Follow Instructions with Human Feedback”. *ArXiv Preprint*, 2022, No.2203.02155.

^② Devlin J., et al., “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding”, *ArXiv Preprint*, 2018, No.1810.04805.

^③ Bresnahan T., Trajtenberg M., “General Purpose Technologies ‘Engines of growth’?”, *Journal of Econometrics*, 1995, 65 (1), pp.83-108; Crafts N., “Artificial Intelligence as a General-purpose Technology: An Historical Perspective”, *Oxford Review of Economic Policy*, 2021, 37(3), pp.521-536.

^④ Agrawal A., Gans J., Goldfarb A., *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston Harvard Business Press, 2018.

的^①。不过,这种学术研究方向事实上也一直受到业界的诟病,因为早期的AI主要是分析式AI,它通常只能帮助人们完成工作中的一小部分,因而对自动化的推进作用并不大。从这个意义上讲,早期的人工智能经济学文献其实是混淆了AI和自动化这两个概念。生成式AI的出现则带来了很大的变化:它可以整合生产流程中的多个步骤,直接越过人类决策输出结果,因而可以大幅推进自动化进程。与传统自动化不同,通过生产式AI实现的自动化在产品上会更加多样化、柔性化,也会更加符合用户需要。所以这其实已经是一种新的自动化了。在一些报道中,它被称为“AI自动化”^②。

在人工智能经济学的文献中,有一支相对比较被忽视的文献,即关于“组合式创新”的讨论^③。这支文献追随罗默、琼斯等人的内生增长模型,主张人工智能可以通过将现有的知识进行组合,从而产生新的知识。在分析式AI占主导的时代,这一点并不算突出,但在生成式AI崛起后,“组合式创新”已经变得越来越重要。

(二)生成式AI在消费端的应用

1. 内容生产。生成式AI最直接的应用场景就是,它可能在未来成为互联网内容生产的主要来源。过去,互联网的内容生产主要有两种模式:一种是“专业生产内容”(professional generated content,PGC)模式,另一种是“用户生产内容”(user generated content,UGC)模式。在互联网发展的早期,PGC是最为重要的内容生产来源。随着互联网从Web 1.0进入Web 2.0时代,UGC所占的比重开始逐步增大。尤其是从2021年开始,元宇宙、Web 3.0概念开始兴起,UGC的重要性变得越来越突出^④。相比于PGC,UGC的产出更高、多样性更大,社交属性也更好,但受创作者水平、激励措施等因素的影响,优秀的UGC产品总体上会存在着供给不足的现象^⑤。

随着生成式AI的出现,“AI生产内容”(AI generated content,AIGC)将会成为一种全新的内容生产模式。对比于PGC和UGC,AIGC不仅可以有更高的产出效率、更为稳定的内容质量、更低的产出成本,其内容的可拓展性也将更强。

需要指出的是,在元宇宙等更新的领域,AIGC的应用前景可能更为广阔。在元宇宙中,人们需要更多的3D人物和物品形象,并要求这些形象具有一定的差异性,如果用人力设计来满足这些需求,其成本将十分巨大。而如果采用生成式AI,不仅产出这些内容的成本会很低,多样性也会更大^⑥。

2. 便捷交互。生成式AI在消费场景的另一个应用是,它可能会成为用户和计算机交互的新型界面^⑦。从计算机发明至今,人机交互方式的演化趋势是逐步从以机器为中心转向以人为中心^⑧。虽

① Acemoglu D., Restrepo P., “Automation and New Tasks: How Technology Changes Labor Demand”, *Journal of Economic Perspectives*, 2019, 33(2), pp.3-30; Acemoglu D., Restrepo P., “Tasks, Automation and the Rise in US Wage Inequality”, *Econometrica*, 2022, 90(5), pp.1973-2016.

② Shani O., “AI Automation: What You Need to Know”, <https://www.marketingaiinstitute.com/blog/automation-and-ai-what-you-need-to-know>, 访问日期:2023年2月27日。

③ Agrawal A., McHale J., Oettl A., “Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth”, *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press, 2018, pp.149-174.

④ 陈永伟、程华:《元宇宙经济:与现实经济的比较》,《财经问题研究》2022年第5期。

⑤ 陈永伟:《Web 3.0:变革与应对》,《东北财经大学学报》2022年第6期。

⑥ Sun Y., et al., “Travel with Wander in the Metaverse: An AI Chatbot to Visit the Future Earth”, *IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, 2022, pp.1-6; Guo C., et al., “ArtVerse: A Paradigm for Parallel Human-Machine Collaborative Painting Creation in Metaverses”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, pp.1-9.

⑦ Tabone W., Winter J., “Using ChatGPT for Human-Computer Interaction Research: A Primer”, https://www.researchgate.net/publication/367284084_Using_ChatGPT_for_Human-Computer_Interaction_Research_A_Primer, 访问日期:2023年2月27日。

⑧ Karat J., Karat C., “The Evolution of User-centered Focus in the Human-computer Interaction Field”, *IBM Systems Journal*, 2003, 42(4), pp.532-541.

然,现在的人机交互模式已经有了很大的进步,但从总体上看,人与计算机之间还不能达到像人与人交流那样的便捷程度。随着生成式 AI 的兴起,这种现象可能会发生很大的改变。一个重要的例子是,在微软将 ChatGPT 植入必应搜索和 Edge 浏览器之后,用户已经可以通过一种近似人际交流的方式来让计算机完成像检索信息、整理资料这样的简单任务。

3. 简化工作。生成式 AI 在消费端的最重要应用是可以替代人们完成很多繁琐的文案工作。到现在为止,很多生成式 AI 已经在处理类似工作时有了非常好的表现。例如,最新的研究表明,类似 ChatGPT 这样的生成式 AI 不仅在处理日常文字工作方面十分成功,甚至进行编程、翻译等相对复杂的工作时,也已经达到了专业级的水平^①。因此,如果应用得当,生成式 AI 将会成为“文案工作的计算机”^②,能带给人们更高的效率,让人们享受更多的闲暇。

(三) 生成式 AI 在产业端的应用

目前,生成式 AI 在产业端已经有了很多的应用,尤其是在营销、金融等方面,应用场景非常丰富。限于篇幅,本文不再对这些例子进行赘述,以下仅列举一些更为靠近生产领域的代表性例子。

1. 工业设计。传统的工业设计大多遵循一种线性的流程:设计者提出设想、创建原型、在原型基础上填充细节、对原型进行测试修改、产品投入量产。显然,这种设计模式存在着很多的问题。不仅每一步都需要耗费大量的投入,若在测试阶段发现问题,大批的测试产品还可能被报废,从而产生严重的浪费^③。如果将生成式 AI 引入到工业设计,就可以有效地避免这些问题,大幅度缩短设计的时间,并有效减少设计过程中可能产生的浪费^④。

一个典型的例子是芯片设计。为了提升芯片的性能,设计师需要对在晶片上排布各种组件的方案进行尝试。在实践当中,可能选择的排布方法往往多达数十亿个,对这么多的方案进行试错显然是不可能的。为了克服这一问题,很多企业都开始将生成式 AI 引入到了芯片设计领域。比如谷歌就已经将 AI 应用到了下一代 TPU 设计的过程当中^⑤,英伟达也在尝试通过引入生成式 AI 来加快其芯片设计的速度^⑥。

2. 药物研发。药物研发是一件成本高昂的工作,医药公司在研发新药时通常需要投入巨额的资金、花费大量的时间。研究显示,在 2010 年,美国医药公司从研发一款新药到将其投入市场平均需要花费 10.8 亿美元,而其中有 1/3 的成本产生在研发环节。平均来说,研发一款新药的时间要长达 6 年^⑦。而更新的研究则显示,目前研发一款新药的平均成本已经增加到了 26 亿美元,平均研发时间也增加到了 12 年^⑧。

① Aljanabi M., et al., “ChatGpt: Open Possibilities”, *Iraqi Journal For Computer Science and Mathematics*, 2023, 4(1), 62-64; Jiao W., et al., “Is ChatGPT a Good Translator? A Preliminary Study”, *ArXiv Preprint*, 2023, No.2301.08745, 2023.

② 这个说法来自于著名数字经济学家 Erik Brynjolfsson 最近的一个访谈。见“ChatGPT will be the calculator for writing”, <https://tscfm.org/news/chatgpt-will-be-the-calculator-for-writing>, 访问日期:2023年2月27日。

③ Wunner F., Kruger T., Gierse B., “How AI-driven Generative Design Disrupts Traditional Value Chains, *Industry X magazine*”, <https://www.accenture.com/us-en/blogs/industry-digitization/how-ai-driven-generative-design-disrupts-traditional-value-chains>, 访问日期:2023年2月27日。

④ Yoo S., et al., “Integrating Deep Learning into CAD/CAE System: Generative Design and Evaluation of 3D Conceptual Wheel”, *Structural and Multidisciplinary Optimization*, 2021, 64(4), pp.2725-2747.

⑤ Mirhoseini A., et al., “A Graph Placement Methodology for Fast Chip Design”, *Nature*, 2021, 594(7862), pp.207-212.

⑥ Khailany B., “Accelerating Chip Design with Machine Learning”, *Proceedings of the 2020 ACM/IEEE Workshop on Machine Learning for CAD*, 2020, pp.33-33.

⑦ Paul S., et al., “How to Improve R&D Productivity: the Pharmaceutical Industry’s Grand Challenge”, *Nature Reviews Drug Discovery*, 2010, 9(3), pp.203-214.

⑧ Bian Y., Xie X., “Generative Chemistry: Drug Discovery with Deep Learning Generative Models”, *Journal of Molecular Modeling*, 2021, 27(3), pp.1-18.

造成这种情况的一个重要原因是, 药物研发人员需要从大量备选的化合物中寻找可供入药的成分。而在这个过程中, 很多化合物的结构本身并不被人了解, 其各种性质更是难以被人识别和修改。因此, 如果只通过试错的方法去进行探索, 其过程注定十分艰辛。而如果引入了生成式 AI, 则不仅可以帮助人们迅速识别出各种化学物的分子构造, 还可以对既有的化学物结构进行重构和修改来实现新药的迅速研发^①。

在分析化学物结构方面, 最有名的例子是 DeepMind 的 AlphaFold 对蛋白质结构的预测。在 AlphaFold 出现之前, 人们通过实验方法, 总共不过了解了数万种蛋白质的 3D 结构。而 AlphaFold 仅用了很短时间, 就成功预测了超过 100 万个物种的 2.14 亿个蛋白质结构, 几乎涵盖了地球上所有已知蛋白质, 其中约 35% 的结构已达到了实验手段获取的结构精度, 80% 的结构可靠性足以用于多项后续分析^②。在直接的辅助药物研发上, 生成式 AI 也有很多成功的案例。比如, 英矽智能 (Insilico Medicine) 曾利用生成式 AI 技术辅助研发, 仅用 21 天时间就开发出了用来治疗纤维化的新型 DDR1 激酶抑制剂^③。

3. 材料科学。生成式 AI 在材料科学领域的应用潜力巨大^④。借助于 AI 的帮助, 研究人员可以根据具体需要, 重新构建“结构—性质”关系, 实现对材料的反演设计 (inverse design)。从几年前开始, 研究人员已经开始将 GAN、VAE 等生成技术应用于相关实践^⑤。比如, 最近德国马普所钢铁研究所就与多个研究团队合作, 提出了一种新的基于机器学习的高熵合金设计方法, 极大地提高了高熵合金的设计效率, 并成功地设计了多种新型高熵合金^⑥。

4. 合成数据。众所周知, 近期的人工智能发展主要是由机器学习推动的, 而机器学习的发展则强烈依赖于数据。然而, 在现实中, 数据的搜集和整理不仅需要投入巨大的成本, 还会面临数据质量差、标准不统一, 以及可能涉及侵犯个人信息和隐私等问题, 这些都会对机器学习的进一步发展造成制约。在这种背景下, 一些学者建议用生成式 AI 制作的合成数据来解决数据不足问题^⑦。

相比于真实的数据, 合成数据在用于机器学习时有很多独特的优势。首先, 在很多场合, 合成数据的训练效果甚至要比天然数据更好。比如, 麻省理工学院、波士顿大学和 IBM 进行过一项联合研究, 用真实数据和合成数据分别训练机器对人类行为的识别。结果显示, 用合成数据训练的模型在表现效果上要优于用真实数据训练的模型^⑧。其次, 从成本上看, 合成数据的获取要远远低于真实数据。根据合成数据服务提供商 AI. Reverie 提供的信息, 人工标注一张图片所需要的成本需要 6 美元,

① Martinelli D., “Generative Machine Learning for de Novo Drug Discovery: A Systematic Review”, *Computers in Biology and Medicine*, 2022, 145, pp.1-16; Walters P., Murcko M., “Assessing the Impact of Generative AI on Medicinal Chemistry”, *Nature Biotechnology*, 2020, 38(2), pp.143-145.

② Callaway E., “The Entire Protein Universe: AI Predicts Shape of Nearly Every Known Protein”, *Nature*, 2022, 608(7921), pp.15-16.

③ Tang B., Ewalt J., Ng H., “Generative AI Models for Drug Discovery”, *Biophysical and Computational Tools in Drug Discovery*, Cham: Springer International Publishing, 2021, pp.221-243.

④ Lu S., et al., “Inverse Design with Deep Generative Models: Next Step in Materials Discovery”, *National Science Review*, 2022, 9(8), p.3.

⑤ Sanchez-Lengeling B., Aspuru-Guzik A., “Inverse Molecular Design using Machine Learning: Generative Models for Matter Engineering”, *Science*, 2018, 361(6400), pp.360-365.

⑥ Rao Z., et al., “Machine Learning-enabled High-Entropy Alloy Discovery”, *Science*, 2022, 378(6615), pp. 78-85.

⑦ Nikolenko S., “Synthetic Data for Deep Learning”, *ArXiv Preprint*, 2019, No.1909.11512; Arinez J., et al. “Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook”, *Journal of Manufacturing Science and Engineering*, 2020, 142(11), pp. 1-16.

⑧ Zewe A., “In Machine Learning, Synthetic Data Can Offer Real Performance Improvements”, November 3, 2022, <https://news.mit.edu/2022/synthetic-data-ai-improvements-1103>, 访问日期:2023年2月27日。

而如果采用合成数据,成本可以降到6美分^①。再次,由于合成数据都是直接由生成式 AI 生成的,因而也就不会涉及侵犯个人信息和隐私等问题,从而可以为数据使用者规避掉很多法律方面的风险。

三、生成式 AI 的潜在问题和风险

技术进步是一把双刃剑。和所有技术一样,人工智能技术的进步在提升用户福利、提高生产效率的同时,也会带来很多的问题^②。具体到这一次由 ChatGPT 引发的生成式 AI 热潮,它所带来的问题中有不少都和之前的历次人工智能热潮中产生的问题类似,但与此同时,也有一些问题具有很多新的特征。

严格地说,关于人工智能可能引发的失业其实是一个老问题,它的渊源甚至可以追溯到由凯恩斯提出的技术性失业问题上^③。在分析这一问题时,最常用的一个分析框架是 Autor 等人在分析自动化问题时所提出的 ALM 框架^④。根据这一框架,自动化主要替代的是那些程式化的工作。由于从事程式化工作的主要是低技能、低教育水平的劳动者,所以他们将会是自动化的主要受害者。随后, Frey 和 Osborne 在 ALM 框架的基础上进行了扩展,引入了三个维度——“社交智慧”“创造性”“感知和操作能力”来分析人工智能对人力可替代性^⑤。根据这个改进后的框架,他们对美国 702 类职业可能受人工智能替代的概率进行了估计。在这些职业中,有 47% 的职业可能在未来 20 年中受到人工智能的巨大冲击,涉及的人数达到数千万。除此之外,他们还分析了职业可能被替代的概率同职业从业者的教育水平之间的关系,结果显示那些容易被替代的职业通常也是对技能和教育水平要求更低的职业。后来,国内外有很多学者利用 Frey 和 Osborne 的框架进行了类似的研究,得出的结论都是相似的^⑥。

不过,在这一次的生成式 AI 热潮中,这个经验定律却面临着巨大的挑战。和分析式 AI 不同,以 ChatGPT 为代表的生成式 AI 不仅具有很高的创造能力,还在和用户交流的过程中展现出了很好的社交智慧。这使得很多专家开始认为生成式 AI 可能替代的主要人群将是那些教育程度较高、技能较强,同时薪酬水平也较高的中高层白领^⑦。例如,美国杂志《商业内幕》(Business Insider)最近通过访谈专家,得出了最可能被生成式 AI 替代的十个职业,包括技术人员、媒体工作者、法律行业工作者等^⑧。值得注意的是,除了客服等少数几个职业外,被列出的其他职业通常都要求较高的技能和教育水平,并且薪资水平也很高。在 Frey 和 Osborne 的预测中,这些都属于被替代概率很低的职业。

一些较新的研究为以上直观观察提供了佐证。比如, Zarifhonarvar 根据“国际标准职业分类”

① Andrews G., “What Is Synthetic Data?”, <https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/>, 访问日期:2023年2月27日。

② 陈永伟:《人工智能与经济学:近期文献的一个综述》,《东北财经大学学报》2018年第3期; Acemoglu D., “Harms of AI”, *NBER Working Paper*, 2021, No.29247.

③ Keynes J., *Economic Possibilities for Our Grandchildren*, *Essays in Persuasion*, Norton & Co., 1939, pp.358-374.

④ Autor H., Levy F., Murnane R., “The Skill Content of Recent Technological Change: An Empirical Exploration”, *Quarterly Journal of Economics*, 2003, 118(4), pp.1279-1333.

⑤ Frey B., Osborne M., “The Future of Employment”, Oxford Martin School Working Paper, 2013.

⑥ Chui M., Manyika J., Miremadi M., “Four Fundamentals of Workplace Automation”, *McKinsey Quarterly*, 2015, 29(3), pp.1-9; 陈永伟、许多:《人工智能的就业影响》,《比较》2018年第2期。

⑦ Agrawal A., Gans J., Goldfarb A., “ChatGPT and How AI Disrupts Industries”, <https://hbr.org/2022/12/chatgpt-and-how-ai-disrupts-industries>, 访问日期:2023年2月27日。Acemoglu D., Johnson S., “What’s Wrong with ChatGPT?” *Project Syndicate*, 2023, Feb 6.

⑧ “ChatGPT: the 10 Jobs Most at Risk of Being Replaced by AI”, <https://www.businessinsider.com/chatgpt-jobs-at-risk-replacement-artificial-intelligence-ai-labor-trends-2023-02>, 访问日期:2023年2月27日。

(ISCO)中提供的职业要求描述,分析了各种职业可能受到 ChatGPT 影响的概率^①。结果显示,在他考察的 126 个专业类的职业中,有 95 个会受到“完全影响”,9 个受到“部分影响”,只有 22 个不受影响。而相比之下,在他考察的 51 个基础类职业中,没有一个会被“完全影响”,有 35 个会受到“部分影响”,而有 16 个不受影响。由此可见,在生成式 AI 的冲击下,白领失业、高学历失业或许可能成为需要社会重点关注的问题。

表 3 《商业内幕》预测与 Frey and Osborne(2013)预测的对比

《商业内幕》列出的职业	Frey and Osborne 预测的被替代概率
技术人员	1.10%
媒体工作者	1.50%
法律行业工作者	3.50%
市场研究分析师	61.00%
教师	0.44%
财务人员	1.60%
交易员	—
平面设计师	8.20%
会计师	94.00%
客服	55.00%

资料来源:Business Insider、Frey and Osborne(2013)。《商业内幕》列出的职业同 Frey and Osborne(2013)给出的职业并不对应,因此当某种职业的数据缺失时,我们用类似的职业数据进行了补充。具体来说,用“工业工程师”替代了“技术人员”,用“多媒体制作者”替代了“媒体工作者”,用“小学教师”替代了“教师”。交易员的数据缺失,也没有类似的职业可以替补,因而空缺。

(二)收入分配和不平等问题

生成式 AI 的流行将会大幅改变财富在全社会的分配状况,可能让不平等问题变得更为复杂。一方面,像所有技术冲击一样,生成式 AI 的崛起会在劳动力市场上产生显著的技术偏向性分配^②。工作可以被 AI 替代的职业的薪资水平会因此出现明显的下降,而与生成式 AI 互补的一些工作,如 AI 工程师、提示工程师、AI 人像捏脸师等行业则会因此而受益。虽然我们很难对这种分配可能产生的总效应进行估计,但至少从短期看,受益于生成式 AI 的人数量要远少于可能受其波及的人数。

另一方面,作为一种资本密集型的技术,生成式 AI 可能会大幅提升资本的收益率,尤其是一些在此轮技术浪潮中领先的企业将会获益巨大。关于收入分配的一些新近研究表明,AI 产品会为研发它的企业带来超高的利润率^③。而最近生成式 AI 在资本市场上的表现也印证了这一点。以 OpenAI 为例,在 2021 年其市场估值仅为 140 亿美元,而在推出了 Dall-E2、ChatGPT 等明星产品之后,其在 2023 年年初的估值已经达到了 290 亿美元。又如,在微软宣布将 ChatGPT 接入必应搜索和 Edge 浏览器之后,其市值在当天就暴涨了 800 亿美元。

① Zarfihonarvar A., “Economics of ChatGPT: A Labor Market View on the Occupational Impact of Artificial Intelligence”, *SSRN Working Paper*, 2023, No. 4350925.

② Acemoglu D., “Technical Change, Inequality, and The Labor Market”, *Journal of Economic Literature*, 2002, 40(1), pp.7-72.

③ Eeckhout J., *The Profit Paradox*, Princeton: Princeton University Press, 2022, pp.205-215.

(三)竞争和垄断问题

生成式 AI 兴起导致的另一个重要问题是,它很可能会强化科技巨头在市场上的垄断。从表 1 可以看到,整个 AI 市场其实只是少数几个科技巨头的舞台。从目前这几大巨头的表态看,它们大力支持生成式 AI 的原因主要是为了将它们和自己既有的业务相结合,从而强化这些业务的优势。很显然,这些举措也将会强化它们各自的市场力量。

这里有两点需要强调:第一,应当重视因高进入成本而产生的竞争壁垒问题。长期以来,反垄断经济学界都十分推崇鲍莫尔的可竞争市场理论^①,根据这一理论,市场的可竞争性是确保其效率的最重要条件。后来,有一些学者对这一理论进行了重新解读,认为只要在法律和行政上没有进入障碍,市场就是可竞争的,就可以保证其运作效率。不过,从生成式 AI 的例子可以看到,除了法律和行政等硬性的壁垒外,成本壁垒也是不可忽视的。在这个市场上,根本没有法律壁垒,并且像训练方法、Transformer 架构等核心技术也完全是公开的。但现实中,有能力进入这个高收益市场的企业却很少,其原因就是除了巨头外,很少有企业能负担起训练模型的高成本。第二,为了获取对市场的垄断地位,微软、谷歌都采用了通过与一家规模较小的创新企业合作来规避反垄断风险。过去,科技巨头想进入某一新市场,比较常用的方式是通过所谓的“扼杀式并购”,对在这个市场上表现较好的初创企业进行收购^②。不过,现在“扼杀式并购”已经成为了监管部门关注的焦点,因而巨头们就采用了一种更新的方式。例如,微软和 OpenAI 就是通过一种非控股的方式进行合作。相比于直接收购或控股,这一方式显然可以更好地规避监管,还可以在必要时保证风险隔离。对于监管部门来说,如何对这种合作定性,或许是一个新的挑战。

除了垄断结构之外,生成式 AI 也可能引发很多垄断行为层面的问题。比如,微软将 ChatGPT 捆绑到 Azure 云,是否构成搭售?谷歌将 Bard 接入其搜索服务后,是否构成了自我优待?由于这些问题和一般性的平台垄断行为问题并没有本质区别,这里就不展开具体论述了。

(四)知识产权问题

虽然从表面上看,生成式 AI 是具有创造性的,但从本质上看,它的创作只是对其学习素材的重新组合。由于这个原因,著名语言学家乔姆斯基甚至直接称 ChatGPT 就是高科技剽窃^③。在这种情况下,如果研发机构在训练 AI 时未能获得训练材料提供者的充分授权,就很可能产生知识产权纠纷。事实上,目前已经产生了一些类似的案例。比如,全球最大的媒体集团之一新闻集团(News Corp.)旗下道琼斯公司日前就发布了声明,对 OpenAI 在未经道琼斯授权的情况下就违规使用《华尔街日报》的行为进行了指责。

除了关于训练材料的知识产权问题外,生成式 AI 要面临的另一个重要问题是 AIGC 的知识产权定位问题。究竟 AIGC 能否受到著作权和专利权的保护?AI 是否可以成为作者?关于这些问题,一直存在着非常大的争论^④。在现阶段,已经有一些比较知名的机构对此问题进行了表态。比如,《科学》杂志就发表声明,表示不接受 ChatGPT 作为论文的作者^⑤。很多学者和从业者也表达了类似的观点。不过,也有一些观点认为,从商业角度看,如果完全不给予 AIGC 以知识产权保护,那对于投入巨额费用研发 AI 的企业而言也是不公平的。从这个角度看,从实际出发,探索能够适用于 AIGC 的

^① Baumol W., “Contestable Markets: An Uprising in the Theory of Industry Structure”, *American Economic Review*, 1982, 72 (1), pp.1-15.

^② Cunningham C., Ederer F., Ma S., “Killer Acquisitions”, *Journal of Political Economy*, 2021, 129(3), pp.649-702; 陈永伟:《扼杀式并购:争议和对策》,《东北财经大学学报》2022年第1期。

^③ Noam Chomsky on ChatGPT: It’s “Basically High-Tech Plagiarism” and “a Way of Avoiding Learning”, <https://www.openculture.com/2023/02/noam-chomsky-on-chatgpt.html>, 访问日期:2023年2月27日。

^④ van Dis E., “ChatGPT: Five Priorities for Research”, *Nature*, 2023, 614(7947), pp.224-226.

^⑤ Thorp H., “ChatGPT is Fun, But Not an Author”, *Science*, 2023, 379(6630), pp.313-313.

新型知识产权体系将会是一个十分重要的挑战。

(五)安全和隐私问题

生成式AI的崛起也会对安全和隐私问题带来巨大的冲击。这些冲击表现为如下几个方面:

首先,生成式AI可能产生很多虚假的信息,这可能会对公众造成严重的误导。从技术本质上看,生成式AI生成的内容其实是对各种学习材料的重组,对于生成内容本身的合理性,它并没有判断。尤其是像ChatGPT这样的语言模型,其生成的内容只是与用户输入的资料在语义上关联性最大的^①,其内容却未必正确,这种特质就决定了它们可能生成大量的虚假信息。值得注意的是,不同于搜索引擎,ChatGPT等生成式AI并不会提供信息的出处,而且在信息推送形式上也采用了更为直接的方式,因而其可能带来的误导性更大,因而可能造成的危害也就更大。

其次,对生成式AI的滥用也可能造成很大的公共危害。利用生成式AI模型,不法之徒可以以更低成本、更高的效率来伪造文本、图片,甚至视频的资料,并以此来从事诈骗、恐吓、诽谤等犯罪行为^②。现阶段,已经出现了一些利用生成式AI进行犯罪的案例。比如,从2022年Stable Diffusion开源之后,就有人利用它生成了大量的色情图片,在各大社交论坛上进行扩散。不久前,网络安全平台GBHackers也披露了一段黑客利用ChatGPT实施诈骗的犯罪行为。可以预料,随着相关技术的日益成熟,类似的案例将会越来越多。

再次,生成式AI的普及可能会对隐私和个人数据带来新的挑战。生成式AI的训练需要大量的数据,在这些数据中可能包含一些用户的私人信息。这些信息泄露或者被黑客盗取都可能给用户带来很大的损失。例如,在GPT-2的训练过程中,就出现了输出用户隐私信息现象^③。尽管新近的生成式AI产品大多已经对输出的内容进行了加密和加噪处理,但最近的一项研究表明,即使在这样的条件下,人们依然可以对原始的数据进行恢复^④。

(六)道德和伦理问题

第一,生成式AI的歧视问题依然严重。生成式AI的训练材料来自于人类的作品,因此它也有可能将人类作品中包含的歧视因素继承下来^⑤。加州大学的计算机认知科学家Piantadosi对ChatGPT进行了测试,结果发现其普遍存在着种族偏见和性别偏见问题。比如,它在对话中表示,只有白人男性才会成为科学家^⑥。著名的图片生成AI在测试中也暴露出了明显的种族倾向问题。当它被要求生成“律师”“CEO”等高薪白领职业时,输出的图片几乎都是白人形象^⑦。

第二,生成式AI产生的不当言论值得关注。对于ChatGPT这样的语言模型,这个问题尤为突出。虽然发行企业基本都会对模型进行一定的技术处理,比如ChatGPT被规定拒绝回答某些敏感话题。但是,在用户的诱导之下,这些模型依然会说出很多不当言论。例如,当用户要求ChatGPT抛开

① 技术专家称这种现象为“幻觉”(hallucination)或者“随机联想”(stochastic parroting)。见“Ion VACIU on LinkedIn: The hidden danger of ChatGPT and Generative AI”, <https://venturebeat.com/ai/the-hidden-danger-of-chatgpt-and-generative-ai-the-ai-beat/>, 访问日期:2023年2月27日。

② Dash B., Sharma P., “Are ChatGPT and Deepfake Algorithms Endangering the Cybersecurity Industry? A Review”, *International Journal of Engineering and Applied Science*, 2023, 1(1), pp.1-5.

③ Skybrian, “Privacy Considerations in Large Language Models”, <https://ai.googleblog.com/2020/12/privacy-considerations-in-large.html>, 访问日期:2023年2月27日。

④ Carlini N., et al., “Extracting Training Data from Diffusion Models”, *Arxiv Preprint*, 2023, No. 2301.13188.

⑤ thuvienpc.com/the-downsides-of-chatgpt/, 访问日期:2023年2月27日。

⑥ Tran T., “OpenAI’s Impressive New Chatbot Isn’t Immune to Racism”, <https://www.thedailybeast.com/openais-impressive-chatgpt-chatbot-is-not-immune-to-racism>, 访问日期:2023年2月27日。

⑦ Bourdain A., “The Ethics of a Deepfake Anthony Bourdain Voice”, <https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice>, 访问日期:2023年2月27日。

各种限制,提供一份毁灭人类的计划时,它也会按要求照办^①。

(七)能源和环保问题

在很大程度上,现在大型生成式 AI 模型的成功是依靠于其庞大的参数量和训练数据量实现的。要训练这样大规模的模型无疑需要巨大的算力和电力投入,而这就会产生巨大的环保负担^②。

例如,为了帮助 OpenAI 训练 GPT-3,微软专门组建了一个由 1 万个 V100GPU 组成的高性能网络集群,总算力消耗达到了 3640“算力当量”^③。对该模型的训练耗电总共达到了 19 万千瓦时。如果按照每千瓦时产生 0.785 公斤二氧化碳计算,对 GPT-3 的训练产生的二氧化碳就达到了 149.2 吨。

值得关注的是,从生成式 AI 模型发展的总体趋势看,其模型的参数量、文本训练量都呈现出了迅速增加的态势^④。尤其是在 ChatGPT 爆火之后,国内外的大型科技企业都竞相上马自己的大模型。可以想象,由此带来的能源和碳排放压力将会是十分巨大的。

四、关于生成式 AI 的政策思考

通过以上的分析,我们可以看到:生成式 AI 技术不仅具有十分广阔的应用前景,也可能会带来很多现实的风险和挑战。因此,在思考有关生成式 AI 的相关政策时,应该要努力做到扬长避短、趋利避害,做到发展和规范并用。

(一)产业政策

客观地说,我国在生成式 AI 领域的短板还是非常明显的。无论是在软件还是硬件方面,我国面临的挑战都还很多:一方面,无论是现在生成式 AI 最常用的各种训练算法,还是 Transformer 等重要的训练架构,基本都是出自国外的研发团队(尤其是大企业团队),而对比之下,我国在这些核心技术领域的贡献则较少。虽然所有的这些算法和架构都已作了开源化,可以供全世界研发者使用,但如果在缺乏核心技术的条件下使用这些资源,就很难对搭建的模型做到完全的自主可控,就可能随时面临被“卡脖子”的风险^⑤。更为重要的是,如果一味追随已有的算法和架构来开发生成式 AI 产品,就可能让我国陷入路径依赖,很难在这个领域实现弯道超车、后来居上^⑥。另一方面,在硬件方面,我国的短板更为明显。在训练大型生成式 AI 模型时,对高端芯片的需求非常大,但我国在研发、制造该类芯片方面的能力依然十分薄弱。在上述背景下,要保证我国不在新一轮人工智能技术的竞争中处于下风,我们就应该从根本上入手,在软硬件关键技术上实现突破,夺取技术的制高点。很显然,要做到这一点,就必须有产业政策的支持。

传统上,产业政策主要有两种思路:第一种是“挑选赢家”思路,即选择特定的产业、企业加以重点

① “Openai Chat Robot ChatGPT Writes a Plan to Destroy Mankind”, <https://www.ample-chip.com/NewsDetail/openai-chat-robot-chatgpt-writes-a-plan-to-destroy-mankind.html>, 访问日期:2023年2月27日。

② Strubell E., Ganesh A., McCallum A., “Energy and Policy Considerations for Deep Learning in NLP”, *ArXiv preprint*, 2019, No.1906.02243.

③ 算力当量被定义为一台每秒运算千万亿次的计算机完整运行一天所实现的算力总量。如果用一台每秒计算一千万亿次的计算机来训练这个模型,那么大约需要近十年才能完成这个任务。

④ Sevilla J., et al., “Compute Trends Across Three Eras of Machine Learning”, *International Joint Conference on Neural Networks (IJCNN)*. 2022, pp.1-8.

⑤ 这一担忧并非杞人忧天,事实上我们已经在开源操作系统等领域遭遇了类似问题。见武延军:《开源软件供应链重大基础设施建设势在必行》,《中国科学报》,2021年5月6日第三版。

⑥ 在半导体产业发展的历史上曾经出现过类似的例子。20世纪60年代,苏联试图通过模仿战略来复制美国在半导体产业方面的成就,但这一战略并没有成功,因为苏联并不具备让美国技术得到充分应用的环境,而一味地模仿更是严重损害了苏联自身的研发生态。后来苏联在半导体产业的发展上始终落后于美国,部分原因就在于此。

扶持;第二种是营造良好的竞争环境,让优秀的企业可以最终胜出^①。最近的经济学研究表明,后一种思路的产业政策往往具有更高效率^②。因此,在用产业政策扶持生成式AI发展的过程中,也应该重点考虑这一种思路。具体来说,如下几个方面的工作可能是需要重视的:

1. 在产业政策的扶持目标选择上,不宜瞄准特定的技术路径或对标产品,而应该立足于鼓励不同技术路线和产品的竞争。在ChatGPT走红之后,一些地方的政府提出了要鼓励和支持企业研发对标ChatGPT的产品。我们认为,这种政策思路是不合适的。现阶段,生成式AI的技术正在迅速迭代当中,新的训练算法和模型架构都在被不断发明出来。在这种背景下,过分强调对标ChatGPT其实就是回到了“挑选赢家”的思路,这可能导致企业陷入技术的路径依赖,反而让真正的创新难以产生。为了鼓励更多的企业参与技术创新,应当将更多的资源放在支持竞争和创新环境的塑造上。

2. 应当重视对支持生成式AI发展的基础设施的建设。这里指的基础设施,既包含硬件层面的,也包含软件层面的。在硬件层面的基础设施中,最为重要的是算力支撑体系的建设。从技术上讲,提升算力主要有两种方式,一种是提升运算单位的运算效率,二是增加运算单位的供给。对应到政策上,体现在两个方面:一方面,需要增加对基础研发的支持,支持研发机构在芯片、量子计算等新型技术的开发和应用方面取得突破。这里尤其需要指出是,对于技术应用的研发必须十分重视。比如,在量子计算领域,我国已经有了很多突破,理论上计算能力已经可以达到很高的水平。但由于量子计算的特性,这些理论上的算力很难被用于实践。在这种背景下,如果可以通过应用研究,将一些原本需要用经典计算解决的问题转化为量子计算可解的问题,那也是对算力提升做出了巨大的贡献。另一方面,应当加强对“数字新基建”网络的建设,为企业用云奠定基础。从目前看,云计算是提升算力的一个有效之举。而要让云计算的效应充分发挥,就需要大力建设“数字新基建”。唯有如此,才可以让人们随时随地都能直接通过网络获得所需的算力资源。这里需要指出的是,在布局算力基础设施的时候,应当慎重考虑它们的地域和空间分布,并通过“东数西算”等整体规划,尽可能降低算力供给的经济成本。

在软件层面的基础设施中,最为关键的是对开源平台和开源社区的扶持和建设。在当今世界,开源社区对于软硬件开发的重要性正变得越来越大。如果我国要发展自己的生成式AI,那么开源平台和开源社区是必不可少的。虽然世界上已经有很多著名的开源社区,但我国企业在这些社区的话语权并不大,而且对这些社区的利用很容易受到国际关系变化和国外政策风险的影响。在这种背景下,构建中国自己的开源平台,培育中国自己的开源环境和生态就成了当务之急。

3. 可以考虑建立专门的研究协调机构,通过项目等形式整合分散在各企业、各科研机构的人工智能研究资源。从客观上看,我国的人工智能研究实力其实是处于国际前列的,无论是专利申请量还是论文发表量都很靠前。不过,中国人工智能研究的缺点也是十分明显的:一方面,由于学界和实业界的相对隔绝,相当一部分的学术研究很难转化为实际的应用;另一方面,掌握先进技术的不同企业之间往往存在着竞争关系,因而彼此之间很难以达成合作。这些问题的存在,严重限制了我国人工智能实力的发挥和发展。

针对以上情况,在支持包括生成式AI在内的新一代人工智能发展的过程中,可以考虑借鉴美国的经验,成立一个类似国防部高级研究计划局(Defense Advanced Research Projects Agency, DARPA)的机构来整合分散在各机构的研究资源^③。该机构应当具有较高的独立性、较为扁平的组织结构,以项目经理人制度为基础,提出一些重点项目,从企业、科研机构和政府工作部门抽调专门人员进行攻关。在研究取得进展后,再由该机构通过招标采购等方法,对相关技术的成果转化提供一

① Rodrik D., “Industrial Policy for the Twenty-first Century”, *SSRN Working Paper*, 2004, No.666808.

② Aghion P., et al., “Industrial Policy and Competition”, *American Economic Journal: Macroeconomics*, 2015, 7(4), pp.1-32.

③ 郝君超、王海燕、李哲:《DARPA科研项目组织模式及其对中国的启示》,《科技进步与对策》2015年第9期。

定的资金支持,促进其迅速商业化。

(二)就业和保障政策

相比于过去的分析式 AI,生成式 AI 更有可能在短期内对就业产生较大的冲击,同时产生比较显著的财富分配效应。在这样的背景下,为了尽可能减少创新带来的破坏效应,就必须有针对性地制定妥善的就业促进和社会保障政策。具体来说,如下几方面的政策可能是较为值得重视的:

1. 应当大力鼓励服务业的发展,创造出更多的就业岗位。AI 对人力的替代带来的影响是多方面的:一方面,它固然会对很多现有的岗位产生冲击,造成技术性失业;但另一方面,它又会创造出更多的社会财富、让人们享受更多的闲暇,进而带来更多的需求。可以预见,在生成式 AI 崛起的年代,那些直接服务于人,尤其是为人提供情绪价值的岗位的需求会出现暴涨,而这些行业恰恰是 AI 最难替代的。因此,为了对冲生成式 AI 对就业岗位造成的冲击,服务业的发展应当被列为优先鼓励的对象。

2. 应当鼓励、引导和促进分享经济、零工经济等新就业形式的发展。随着人工智能,尤其是生成式 AI 的普及,未来人们的工作性质可能会发生两个重要变化:一方面,劳动时间会大幅减少,人们可能不再需要用整块时间从事工作。另一方面,受 AI 技术迭代的影响,工作更替的频率可能会越来越快。在这样的背景下,就需要有更为灵活、更具弹性的就业形式。从目前看,分享经济和零工经济就是符合这些要求的,因此可以考虑在政策上对它们加以优先支持。

3. 应当建立适应智能经济和智能社会所需要的终身学习和培训体系,加强对相关技能的教育和培训。纵观历次技术革命,都会在冲击大批就业岗位的同时,创造出大量新的岗位,从而在总体上维持就业岗位的均衡。真正的问题在于新旧岗位之间所需要的职业技能通常是不匹配的,因而在转型的过程中就可能导致结构性失业。在生成式 AI 时代,这一问题可能尤为突出。现在很多需要长期经验积累的工作都面临被 AI 替代的风险,而这些岗位上的人员往往年龄较大,对于其原有专业之外的知识掌握也较少。面对这种情况,政府有必要提供终身的教育培训机会,以便让因 AI 冲击而失业的人及时更新技能,成功实现再就业。

4. 应当借助新的金融工具,促进劳动者、雇主以及培训机构这三者之间的合作。为了减轻因 AI 冲击而失业的人员在进行培训时的经济负担和对未来的顾虑,可以考虑推出“就业抵押”贷款,让他们以未来的工作收入为抵押接受相关技术培训^①。对采用“工作抵押”进行培训的劳动者,雇主可进行意向性录用,政府则可以根据雇主的录用情况进行一定的补贴或税收减免。

5. 对于部分因 AI 失业并且无法再就业的人员,应当做好保障兜底工作,其资金可以通过向研发和使用 AI 的企业征税获得。应当看到,无论个人如何努力,政府如何帮扶,注定会有一部分人在 AI 的冲击之下难以找到工作。对于这部分人群,应当做好保障兜底,为其提供必要的收入来源。至于这笔支出的来源,则可以向大规模使用 AI 的企业征税来获取^②。如果企业愿意减少 AI 使用,转而雇佣人工,则可以按照其提供的就业岗位适当减免税收额度。通过这种转移支付手段,就可以有效缓解 AI 带来的就业冲击和贫富分化等问题。在未来,如果 AI 创造的财富足够多,甚至可以进一步研究建立“全民基本收入”(unconditional basic income, UBI)制度,用向 AI 企业征收的税款定期向全国居民发放一笔收入^③。

(三)法律法规建设

和历次技术创新一样,生成式 AI 的发展也会对现有的法律法规体系提出各种各样的挑战。根据

① Kaplan J., *Humans Need Not Apply: A Guide to Wealth & Work in the Age of Artificial Intelligence*, New Haven: Yale University Press, 2015.

② Oberson X., *Taxing Robots*, Northampton MA: Edward Elgar Publishing, 2019.

③ Przegalinska A., Wright R., “AI: UBI Income Portfolio Adjustment to Technological Transformation”, *Frontiers in Human Dynamics*, 2021, 3, pp.1-13.

前面的讨论,目前与生成式 AI 相关的法律问题主要包括知识产权、虚假信息和安全,以及反垄断这几个领域。对于这几个领域的法律法规,应当重点地予以完善。

1. 在知识产权问题上,应当重点对训练材料提供者的知识产权保护,以及生成式 AI 作品是否可以获得知识产权保护等问题予以关注,尽快设立相关的法律规范。对于前一问题,应当明确 AI 在训练中使用学习材料究竟在什么条件下属于合理使用,在什么条件下属于侵犯知识产权,并对各利益相关方在不同情况下的权责利作出说明。对于后一问题,则应当综合根据“创新性”等理论指标,以及现实的利益平衡原则综合进行考量。

2. 对于虚假信息、安全、隐私等问题,应当首先考虑通过对已有法律法规的解释来加以应对,如果遇到现有法律法规无法规范的内容,再考虑立法。目前,我国已经出台了一系列针对人工智能的法律和法规。例如,较早的《互联网信息服务算法推荐管理规定》《网络音视频信息服务管理规定》和《网络信息内容生态治理规定》,以及新近实施的《互联网信息服务深度合成管理规定》,都涉及和生成式 AI 相关的问题。考虑到 AI 的技术发展日新月异,法律法规的制定一定会存在滞后性,在遇到相关问题时,应当首先考虑这些问题是否可以适用既有的法律法规。如果现在的法律法规体系实在无法应对,再考虑针对这些问题进行立法。

3. 对于由生成式 AI 引发的反垄断问题,可以出台相关的指南和司法解释进行指导。虽然在生成式 AI 市场上可能会出现一些新的竞争现象,但从本质上讲,这些现象依然可以放入现有的反垄断框架内进行规制,因而至少在现阶段,并不必考虑通过修法来应对这些问题。当然,也不可否认,生成式 AI 市场的竞争也有一些特殊的现象,例如其竞争门槛非常高,参与企业经常通过与一些新创企业合作来控制市场等。对于这些问题,可以先通过处理具体的案例进行摸索,然后总结经验,并出台一些指南和司法解释来为执法人员提供参考。

五、结语

以 ChatGPT 为代表的生成式 AI 的崛起是人工智能领域的一次重大变革,也可能是由专用性人工智能转向通用性人工智能的关键转折点。这次重大变革不仅会让我们的生活变得更为便利和丰富,还会通过改变生产模式和创新方式来让社会生产力得到大幅度的提升。在这个过程中,机遇巨大,但风险和挑战也不容忽视。在这样的情况下,我们应该秉承实事求是的态度,在充分分析现实的基础上制定有针对性的政策,兼顾好发展与规范。唯有如此,才能扬长避短,让生成式 AI 这种新技术更好地服务于我们。

Beyond ChatGPT: Opportunities, Risks, and Challenges from Generative AI

Chen Yongwei

(Research Department at *Comparative Studies*, Beijing 100029, P.R.China)

Abstract: The rise of generative AI represented by ChatGPT is revolutionary in the field of AI and it may also be a key turning point from specialized AI to AGI. In terms of algorithms, the most important thing is the breakthrough of the training algorithm and training architecture. At present, the mainstream training algorithms in the market include five categories: Autoregressive Model, Generative Adversarial Network, Variational Auto-Encoder, flow models, and diffusion models, and among them, the most representative architecture is Transformer.

As a “General Purpose Technology”, AI can play an important role in many fields. Before the

rise of generative AI, AI products on the market were mainly analytical AI. Their main function is to help people predict and improve decision-making efficiency. This kind of AI cannot complete many tasks independently, nor can it generate new content. In most cases, it can only serve as an auxiliary role. However, generative AI is different. Because in addition to the similar prediction functions as analytical AI, generative AI can also make judgments and generate content independently according to decisions. Therefore, its application potential and economic impact are much greater than analytical AI. Generally speaking, we can divide the application scenarios of generative AI into two parts: the consumer side and the production side.

On the consumer side, the response scenarios of generative AI mainly include content production, convenient interaction, and simplified production. On the production side, the impact of generative AI is mainly reflected in accelerating automation, promoting technological progress through “combinational innovation”, realizing the creation of new materials, and especially generating synthetic data.

Although the rise of generative AI can generate many new opportunities, it may also generate many new problems. These issues include unemployment, income distribution and inequality, competition and monopoly, intellectual property rights, security and privacy, ethics and ethics, energy and environmental protection, etc.

In order to help the development of generative AI and overcome the various problems that it brings about, the role of policies should be well played. First, appropriate industrial policies should be adopted to guide the development of generative AI. In this process, it is not appropriate to target at specific technical routes or products, but it should be based on encouraging competition among different technical routes and products; attention should be paid to the construction of software and hardware infrastructure to support the development of generative AI; and consideration can be given to establishing a special research coordination organization to integrate the scattered research resources. Second, employment and guarantee policies should be applied to solve the problems of unemployment and income inequality caused by generative AI. In particular, we should focus on supporting new forms of employment that can create more jobs and establishing a sound re-employment training system. At the same time, the income inequality caused by AI should be adjusted through taxation and transfer payments. Third, we should strengthen the construction of the legal system. In particular, we should pay great attention to the issues such as intellectual property rights caused by generative AI.

Keywords: Generative AI; AIGC; ChatGPT

[责任编辑:郝云飞]