

智能社会研究

Journal of Intelligent Society

中华人民共和国工业和信息化部主管

哈尔滨工程大学主办

Journal of Intelligent Society

JIS

第3卷
2024
第6期

智能社会研究

中华人民共和国工业和信息化部主管

ZHINENG SHEHUI YANJIU



杂志公众号二维码
官网网址 www.jis.ac.cn



定价：45.00 元

智能社会研究

(双月刊)

2024年
第6期

2024年11月10日出版

2022年11月10日创刊

总第13期

目 次

论文

数字素养对低龄老年人幸福感的影响研究

- 基于 CFPS 2020 年数据的实证分析 宋思瑶 贾开阳(1)
通用人工智能著作权问题的法律因应 罗艺 曹子贤(26)
人工智能生成技术方案可专利性界定路径 宋红松 王瑞新(48)

研究报告

上海数字适老 2.0 探索

- 构建新旧媒介并行的微粒社会 张林(64)

译文

人工智能中的人的形象

- 与约瑟夫·魏岑鲍姆对话 伯恩哈德·珀克森 著 王立秋 译(88)

计算的限度

——约瑟夫·魏岑鲍姆与聊天机器人 ELIZA

..... 大卫·贝里 著 王立秋 译(107)

从约瑟夫·魏岑鲍姆到 ChatGPT

——与令人倾倒的 AI 技术的批判交锋

..... 克里斯蒂安娜·弗洛伊德 著 王立秋 译(137)

书评

数字平台帝国化的制度演进

——评莱顿维塔的《云端帝国：数字平台如何重塑经济与世界》

..... 牛一帆(169)

技术对社会的再结构

——从卡斯特的网络社会到智能社会 叶莹菲(188)

CONTENTS

THESES

A Study of the Impact of Digital Literacy on the Well-Being of the Under-Aged Older Adults: Empirical Analysis Based on CFPS 2020 Data ... Song Siyao, Jia Kaiyang(1)

Legal Cause of the Copyright Issue of General Artificial Intelligence Luo Yi, Cao Zixian(26)

Pathway to Defining Patentability for AI-Assisted Inventions Song Hongsong, Wang Ruixin(48)

RESEARCH REPORT

Exploring Digital Aging 2.0 in Shanghai: Constructing a Granular Society with Parallel
New and Old Media Zhang Lin(64)

TRANSLATED TEXTS

The Image of Man in Artificial Intelligence: A Conversation with Joseph Weizenbaum
..... written by B. Pörksen; trans. by Wang Liqiu(88)

The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot
..... written by D. Berry; trans. by Wang Liqiu(107)

From Joseph Weizenbaum to ChatGPT
..... written by C. Floyd; trans. by Wang Liqiu(137)

BOOK REVIEWS

- The Institutional Evolution of the Digital Platform Empire: A Review of Vili Lehdonvirta's
Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control Niu Yifan(169)
- The Restructuring of Society by Technology: From Manuel Castells' Network Society to the Intelligent Society Ye Yingfei(188)

计算的限度^{*}

——约瑟夫·魏岑鲍姆与聊天机器人 ELIZA

大卫·贝里 著^{**}

王立秋 译^{***}

摘要:约瑟夫·魏岑鲍姆于 20 世纪 60 年代开发的 ELIZA,堪称有史以来最具影响力的计算机程序之一。ELIZA,尤其是其最著名的人格 DOCTOR,至今仍是程序员的灵感源泉。它引发了关于人工智能(AI)更广泛的讨论,催生了诸多对其的模仿。作为所有对话界面和聊天机器人的鼻祖,ELIZA 尤其令研究 AI 和计算领域的工程师、历史学家和哲学家着迷。ELIZA 能凭借相对少量的计算机代码生成类人回应,这一能力为大量类似程序铺平了道路。这些程序以对话智能体和其他人机界面的形式出现,在计算机科学领域催生出数个全新的研究领域。本文将探讨魏岑鲍姆对 AI 的贡献,并结合生成式 AI(如 ChatGPT)的当代发展状况,思考其更具批判性的著作。此外,本文还将通过研究有关 ELIZA 的讨论,为当前围绕机器学习和 AI 的争论提供一些思路。

关键词:魏岑鲍姆 ELIZA 人工智能 机器学习 聊天机器人 ChatGPT

一、引言

人工智能(AI)承诺采集并重塑人类最强大的能力——语言、创造力、

* 本文原题为“*The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot*”,原刊于 *Weizenbaum Journal of the Digital Society* 3(3),由作者以 CC-BY 协议许可翻译发表。

** 大卫·贝里(D. Berry),萨塞克斯大学媒体、艺术与人文学院。

*** 王立秋,哈尔滨工程大学人文社会科学学院。

推理和智力的本质。进步正飞速发生,在过去短短十年间,AI 就产生了实际影响,自 2019 年以来,这种发展更是显著加速。2022 年,像 OpenAI 的 ChatGPT 和 Stability AI 的 Stable Diffusion 这样的突破,被认为以前所未有的方式提升了人类在语言和艺术方面的创造力与生产力。在本文中,我想联系 ELIZA 来思考这一领域近来的进展。在许多人看来,ELIZA 是如今所谓生成式 AI 的前身,它是一个相对简单的聊天机器人(对话智能体)程序,使计算机内基于对话的界面成为可能。

ELIZA 由约瑟夫·魏岑鲍姆于 20 世纪 60 年代开发。可以说,它是有史以来最具影响力的计算机程序之一。ELIZA,尤其是其最著名的人格 DOCTOR,仍然吸引着程序员,并催生了各种对它的模仿。虽然 ELIZA 影响了计算机科学乃至更广泛的文化,但它的初始源代码从未被公布或广泛传播。^① 不过,它是所有聊天机器人的前身、对话式人机互动的鼻祖,也是大众对计算机可能形象的想象(如 1968 年电影《2001:太空漫游》中的 HAL9000 和 2013 年电影《她》中的操作系统)的灵感来源。这个所有对话界面和聊天机器人的始祖仍然令研究人工智能和计算的工程师、历史学家和哲学家尤为着迷。值得注意的是,ELIZA 在对计算历史产生实质性影响的同时,也使其永远陷入与性别和阶级的问题性假设的纠葛之中(见下文)。ELIZA 能够利用相对少量的计算机代码生成类人回应,凭借这一能力为大量类似程序铺平了道路。这些程序以对话智能体和其他人机界面的形式出现,在计算机科学内部激发了几个全新的研究领域(Boden, 1977)。

值得注意的是,OpenAI 在 2022 年发布 ChatGPT,为复杂大语言模型(LLM)系统 GPT-3 设计合适界面时,决定采用聊天机器人的形式。虽然

^① 作为一个 2021 年形成的国际研究团队的一部分,我们已经在 MIT 档案馆中重新发现了 ELIZA 丢失的用 MAD/SLIP 语言编写的代码。关于这一进展的细节已经发布于 ELIZA 档案网站,参见 <https://sites.google.com/view/elizagen-org/the-original-eliza>。

该公司尚未充分解释使用这种基于文本的模型背后的逻辑,但使用该系统本身就表明,很可能是创造人格的能力促使开发者这样设计系统。为进一步探究这一点,本文认为,回顾过去对理解像 ChatGPT 和类似产品这样的 AI 产品的特殊性或许有用。这就意味着要思考以前的系统,尤其是 ELIZA。我尤其想要考察约瑟夫·魏岑鲍姆(Weizenbaum, 1976)在《计算机的力量和人的理性》中呈现的反思。在书中,他联系人类思想不可简化(不能把人的思想简化为逻辑函数,尤其是简化为 ELIZA)这一特性来构建自己的研究框架。^① 这本书提出了合理性与逻辑——尤其是计算机程序和数学公式中能实现的逻辑——之间关系的问题。正如魏岑鲍姆(Weizenbaum, 1976: 11)所论证的:“把计算机引入我们已然是‘技术社会’的社会……只是强化和放大了先前的那些压力,这些压力迫使人对自己的社会采取一种日益高度合理性的看法,并为自己塑造一个越来越机械的形象。”魏岑鲍姆认为,应该对给计算机分配的任务设限。这意味着要为计算的部署建立规范限制,因为计算机会影响人在世界上寻找自身定位的欲望。他担心计算机严格遵循工具编程逻辑时形成对同情或人际尊重的伪装,认为这种欺骗将会无意间给双方带来改变的交流性相遇变成一种造成疏离的接触。他(Weizenbaum, 1972: 316)也预见到了对计算机决定提出异议的困难,他写道:

人不再为“机器说的话”负责。因此,就不存在对错,不存在正义与否的问题,不存在关于同意或不同意什么的理论。最终,也不存在让人能够质疑“机器说的话”的基础了。

^① 在这里,我们可以指出计算系统(尤其是聊天机器人,不管是 ELIZA 还是当代如 Siri 那样的界面)的性别化(更完整的讨论可参看 Marino, 2006)。

重要的是,我们要开始学着从“后端”操作和“前端”界面这两个角度来理解这些新的 AI 系统,因为作为计算系统,它们可能会变得越来越普遍。就像瓦特 1776 年发明的蒸汽机一样,AI 常被视为具有多种潜在应用的通用技术。因此,我们可以预计,AI 的投入使用将对全球经济产生变革性影响。人们已经在日常语境中使用 AI 了,如电子邮件的垃圾邮件过滤功能、媒体的推送系统、导航应用以及支付交易的验证与核实。虽然 AI 显然有对社会和个人产生巨大影响的潜力,但描绘这些影响是一项艰巨的任务,特别是考虑到 AI 领域经常宣扬自己的发现和潜力。^① 从监管的角度来看,AI 系统的自主性引发了一系列独特的问题,涉及责任、保险、公平性、风险与安全,甚至创意内容的所有权。这些问题要求我们深入思考透明性与偏见、基础设施决策,以及创造、使用和评判它们所需的各种新数字技能。

我们的社会愈发依赖数字技术,这些技术包含了计算合理性(*computational rationalities*),进而也包含了算计的合理性(*calculative rationalities*),并且围绕维持社会学意义上的人类推理能力这一事情提出了诸多挑战(Berry, 2011)。我们对小型软件应用日益依赖,很快就会出现问题,因为它们是自动化、网络化的,并且被接入了更大的软件平台和服务,这使其运作更加复杂。虽然这些系统中有许多最初是为支持或辅助人们进行大量活动、分析和决策的判断而设计的,但它们的运行早已超出用户的理解范围,可同时它们又不可或缺。这样,这些设备通过使人类理性的认知过程卷积短路,通过给予某些逻辑进程工具关系特权,改变了人类理性能力。于是,才有了“我们社会中,个体普遍存在的无力感……人们彼此之间、对自己工

^① 在我看来,就技术的变革潜能来说,当前的机器学习和 AI 系统能够把人在当代数字基础设施、数字应用和数字过程中不得不做的“无聊之事”自动化,这一事实比说它们有“创造力”或者能使文化自动化更具说服力。

作的普遍疏离……以及普通人那种生活在庞大系统缝隙中的感觉”(Rosenberg, 1980: 49)。魏岑鲍姆提出的这些说服、欺骗和中介概念,与后来 AI 发展轨迹中普遍存在的问题非常契合。

值得注意的是,发达资本主义社会的经济无政府状态,与理性化和技术相互交织,这种交织倾向于阻碍反思劳动。在此条件下,工具理性的价值被赋予了特权地位,毕竟其体现于合理性概念本身。将算计和理性思考混为一谈,意味着不能被简化为数的事物都是幻想或形而上学的。结果,社会对批判思想的敏感性可能下降,这也体现为个体化潜能的削弱。在当代计算社会,人们有十足的动力采用理性化的行动与思考方式,也就是插入算法的方式。而转向使用 AI,更确切地说是使用机器学习的自动化系统,会进一步推动这个带来疏离感的过程。这表明,能够批判这些系统、避免再次陷入缺乏对系统运行充分依据解释的抽象行为的解释性理解极为重要(Berry, 2021)。正如魏岑鲍姆所说:“即便大型计算机系统,也不一定不可理解。它们可理解性的秘密在于,这些系统是非常稳健的理论模型。”(Rosenberg, 1980: 45)

我们对合理性的构想在多大程度上被重新表述为一种将人的思想理解为算计而非判断的形式,这正是汉娜·阿伦特(Arendt, 1972: 11)联系政治决策和政策所研究的内容。阿伦特的研究显然影响了魏岑鲍姆的思考。^①他引用阿伦特的话说:“他们(政策制定者和执行者)不仅聪明,还为(自己)‘合理’……而不下判断感到自豪;他们算计……一种对现实可算计性的从根本上非理性的自信(变成了)决策的主旋律。”(Weizenbaum,

^① 魏岑鲍姆提及刘易斯·芒福德、汉娜·阿伦特、雅克·埃吕尔、西奥多·罗斯扎克、肯尼斯·博尔丁,称这些人表达了“对科学技术肆意发展所造就状况的深切关注”(Weizenbaum, 1976: 11)。魏岑鲍姆在《计算机的力量和人的理性》里指出,出版前刘易斯·芒福德“通读了全书”(Weizenbaum, 1976: x)。

1976: 14, 强调部分为原文所有)。在同一文本的稍后部分,他又阐述道:

计算机能够做出司法判决,也能够做出精神疾病诊断。它们抛硬币的方式能比最有耐心的人所采用的方式复杂得多。关键在于,不应该让计算机承担这类任务。在某些情况下,它们甚至能得出“正确的”决定,但这种决定总是且必然建立在没有人愿意接受的基础之上。
(Weizenbaum, 1976: 227)

要联系 ELIZA 来研究这些问题,思考魏岑鲍姆在设计和编写该系统时所关注的语境和一些争论是有帮助的。接下来,我首先要考察魏岑鲍姆开发 ELIZA 的环境,从而理解他是如何构思界面的可供性(affordances)的;其次,思考 ELIZA 与现代用生成式 AI 开发的聊天机器人之间的异同;最后,在结论部分提出关于这种联系的一些反思,并对魏岑鲍姆提出的警告加以引申。

二、ELIZA 和早期的 AI

约瑟夫·魏岑鲍姆于 1923 年 1 月 8 日出生在柏林的一个犹太家庭。他的父亲是皮草技术员,母亲叫亨利埃特。^① 13 岁时,他随父母逃离纳粹德国,移民到美国并在底特律生活。他后来说自己“是在恰逢合理性的终极流产的成长岁月进入世界的”(Dembart, 1977: 1)。虽然他最初在韦恩大学学习数学,但 1941 年“他因战争中断了学习,其间在空军气象部门服役”(Sack, 2018)。1946 年,他返回大学,并分别于 1948 年和 1950 年获得学士

^① 魏岑鲍姆 2008 年 3 月 5 日去世于德国。

和硕士学位。^① 此后，“他在底特律的韦恩大学参与设计和建造了一台数字计算机”(MIT, 2008)。1955 年，他搬到西岸，加入通用电气公司，与 SRI 合作开发了一个名为电子记录机会计 (electronic recording machine accounting, ERMA) 的银行自动化系统。^② ERMA 是为帮助美国银行管理此前一直依靠人工处理的海量纸质支票而开发的。在 20 世纪 50 年代初，银行业在本质上处于危机边缘：

1943—1952 年，美国的支票使用量翻了一番，从每年签发 40 亿张增长到 80 亿张。银行家估计，到 1955 年，支票数量每年将增加近 10 亿张；到 1960 年，每年签发的支票数量将高达 140 亿张。(Fisher & McKenney, 1993: 44)

ERMA 系统让人们能够利用打印在支票底部的磁编码字体，借助磁墨水字符识别技术实现支票处理自动化。在向媒体做最终演示时，ERMA 还无法完全自动运作，需要一名程序员躲在后面的屋子里帮忙掩盖系统故障。这种用人来替代当前尚未运行或未完成的算法的技巧，被称为“奥兹国的巫师”编程。费舍尔、麦肯尼认为：

SRI 工程师……以竖大拇指的形式表示系统运行稳定，可继续演示。在设计周期的早期，这种样机形式被视为有益，因其提供了研究和了解用户预期与要求的途径。莫兹利等人认为，这种方式尤其适合探

^① 魏岑鲍姆解释说：“促使我投身数学的是……在所有可学的东西里，数学看起来是最容易的。数学就像一场游戏，完全是抽象的。在认为数学最容易这一认识背后，隐藏着与之对应的一种认识，即真实生活是最难的。我从孩童时期起就这么认为了。”(Dembart, 1977)

^② SRI 国际是斯坦福大学于 1946 年创立的一家非营利公司，原名为斯坦福研究中心。1970 年，SRI 脱离大学独立，于 1974 年更名为 SRI 国际，且越来越聚焦于国际事务 (SRI International, 2016)。

索待投入使用系统的设计可能性。ERMA 的展示很顺利,媒体从未质疑(计算机系统)不够理想。(Fisher & McKenney, 1993: 55)

莱茵戈尔德(Rheingold, 1985: 164)认为,“最终将这个过程成功计算机化”是“世界银行系统计算机化的一个里程碑”。不过,在计算过程中,用欺骗代替算法的引入和使用,使其看起来平稳,以及用人力劳动取代计算机这种做法,预见了魏岑鲍姆后来提出的关于自动化的担忧。^①

在通用电气任职期间,魏岑鲍姆着手开发一个名为对称列表处理器(symmetric list processor, SLIP)的编程系统,并于1963年完成。SLIP是一组函数,这些函数通过调用例行程序(routines)来执行列表处理,这些例行程序最初由机器代码编写,后来用Fortan编程语言编写。魏岑鲍姆本人将其称为“FORTEAN/SLIP程序”(Weizenbaum, 1963: 524)。在离开通用电气之前,魏岑鲍姆写过一篇名为《怎样使计算机看起来智能》(“How to Make a Computer Appear Intelligent”)的揭秘文章。文章引用马文·明斯基的观点来论证:一种以特定观察者无法理解的方式产生结果的活动,在该观察者看来,在某种程度上是智能的,或者至少有智能的迹象(Weizenbaum, 1962: 24)。在那篇文章中,魏岑鲍姆还描述了一个“五子棋”游戏,该游戏看似智能,但其底层编程实际上并未编入这种智能:

根据上述推理,“人工智能”程序的作者能在一段时间内愚弄部分观察者,其成功程度可通过被愚弄的观察者所占百分比乘以他们未察觉的时长来衡量。(Weizenbaum, 1962: 24)

^① 考虑到魏岑鲍姆于1955年或1956年加入这家公司,他是否出席了ERMA的发布并不明确。魏岑鲍姆不在费舍尔和麦肯尼所列的主要工程师或贡献者名单上,但他仍有可能听说过那种骗术,即通过在计算机和演示者之间安排一名工程师来假装直接操作计算机。

1993 年,在反思自己工作时,他承认这让他有了“骗子或老千”的名声 (Crevier, 1993: 133)。^① 他觉得:“从某种程度上说,那是我后来开发 ELIZA 的前身,也让我有了骗子或老千的名声。但另一方面,我毫无保留地说了出来。我的想法是创造‘计算机是智能的’这种强烈错觉。在那篇文章里,我花了很大力气解释,幕后没多少东西,机器并没有在思考。我把策略解释得很清楚,任何人都能写出那个程序,我写 ELIZA 时也是如此。”(Crevier, 1993: 133)

基于 SLIP 方面的工作,1963 年,他获得了麻省理工学院提供的电子工程助理教授职位。四年间,他就拿到了终身教职,于 1970 年升任计算机科学与工程正教授。^② 在那里,他用密歇根算法解码器 (Michigan algorithm decoder, MAD, 1959 年为 IBM7090 和相关计算机开发的一种编程语言与编译器)重写了 SLIP。MAD 基于 ALGOL 语言,但并非 ALGOL 语言。MAD-SLIP 这一术语常被用于指代二者的结合,可实际上,SLIP 和 MAD 始终是分开的,它代表一组可纳入 MAD 程序的列表处理函数。这与长期存在的误解相反,即 ELIZA 最初是用 LISP 而非 MAD-SLIP 编写的。值得注意的是,魏岑鲍姆后来证实,ELIZA 是从他 1962 年编写的那个游戏发展而来

① 我们不禁好奇,魏岑鲍姆是否读过大卫·毛雷尔(D. Maurer)1940 年出版的《大老千》(*The Big Con*)。毛雷尔后来写到,如今,骗子不会再背着一大堆事先写好、涵盖所有可能情况发展的脚本到处跑了,他们什么都不写。不过,他们凭经验就能知道可能出现的情况,也知道他们的台词应对这些情况是完美的。所以,如果内鬼提供了一间办公室,那么就有很多常备的应变方式。骗局的所有参与者都牢记这些应变方式,并且能根据给定的信号采用相应的应变方式(Maurer, 1974: 89; Weil, 2015)。很难想象有比这更好的对 ELIZA 的描述了。有趣的是,魏岑鲍姆也在《计算机的力量和人的理性》(Weizenbaum, 1976: 121)中收录了一小段关于“骗子”的离题讨论。

② 魏岑鲍姆后来先后在哈佛大学教育研究生院、斯坦福大学、柏林工业大学以及德国汉堡大学任职。他还是美国科学促进会会员、纽约科学院和欧洲科学院成员(MIT, 2008)。

的。^① 克勒维耶认为：

20世纪60年代中期，魏岑鲍姆在MIT时试图让计算机用英语与人交谈。他指出，像STUDENT这样的现有程序应用领域有限，而且描述这些领域的知识与程序结构本身密不可分。魏岑鲍姆还探究了程序通过向用户提问来获取更多信息的有限能力：“我很清楚，我们在接下来的几个星期里解决不了这个问题。”魏岑鲍姆告诉我：“所以我开始思考替代方案……我把我所有的办法整合起来，做起了ELIZA这个项目。”(Crevier, 1993: 133–134)

ELIZA试图以一种相对简单的方式模拟人与心理治疗医师可能的对话。它有一个互动界面，用户可通过打字来回答软件所生成的问题。众所周知，很多人对它介入对话的能力印象深刻，觉得这个程序具有比实际更多的理解力与智能。写完ELIZA后，魏岑鲍姆承认：“我没意识到，极短时间接触一个相对简单的计算机程序，也会让相当正常的人产生强烈的妄想。”(Weizenbaum, 1967: 7)正如他所解释的：“选择这种对话模式，是因为精神科晤谈是为数不多的分类二元自然语言交流(categorized dyadic natural language)的例子之一。在这种交流中，参与双方的其中一方可以自由地摆出对真实世界一无所知的姿态。”(Weizenbaum, 1966: 42; Boden, 1977: 108)^②通过使用简单的模式匹配技巧，魏岑鲍姆让ELIZA程序能够把输入

^① 魏岑鲍姆(Weizenbaum, 1962)所描述的游戏其实是五子棋，而非跳棋。萨缪尔在谈及“机器学习”时暗示了跳棋，但未直接引用(Samuel, 1959)。不过，魏岑鲍姆确实参考过香农关于象棋的作品(Shannon, 1950; McCarthy, Minsky & Rochester et al., 2006)，而且很可能受到萨缪尔1959年那篇文章的影响。

^② 有趣的是，魏岑鲍姆在一次访谈中称，他“一开始构想ELIZA为一名酒保，但后来觉得心理治疗师更有趣”(Crevier, 1993: 136)。

的英文句子转变为对用户而言看起来有意义的输出。麦考黛克认为：

当时，问答机器在 MIT 极为火爆。博布洛(Bobrow)在 MIT 研究 STUDENT；拉斐尔(Raphaelb)在做后来的 SIR；“棒球”程序则是剑桥地区的成果。魏岑鲍姆趁势推波助澜，他连续多日早起，与开发过 CO-MIT 语言的邻居维克托·英格维(V. Yngve)一起研究模式匹配。(McCorduck, 2004: 292)

魏岑鲍姆在 ELIZA 程序里探究了许多此类问题。ELIZA 这个名字源于伊莱莎·杜立德，她是萧伯纳 1912 年戏剧《皮格马利翁》(也叫《卖花女》)中的一个工人阶级角色。事实上，魏岑鲍姆后来表示，他是偶然观看了由该戏剧改编、1956 年首演且 1964 年推出电影版的《窈窕淑女》才知晓这个角色的。他为那个后来与 ELIZA 相混淆的人格脚本所起的名字是 DOCTOR。虽然后来这个名字可能一直被视为代表心理治疗师的无性名字，但 ELIZA 延续了这个名字的内涵，将阶级、性别乃至种族与族群的社会结构融入程序中——因为 ELIZA 这个名字包含着萧伯纳戏剧以及 20 世纪 60 年代改编的音乐剧中对女性和社会阶级的再现所蕴含的假设。

ELIZA 会以打字的方式对所有英文输入做出回应，其脚本 DOCTOR 旨在模拟或“戏仿”一位罗杰斯学派心理治疗师。罗杰斯学派心理治疗由心理学家卡尔·罗杰斯于 20 世纪 40 年代创立，采用一种鼓励患者多说话的非指导性语言方法。魏岑鲍姆将这种方法用作构建 ELIZA 回应的模型，ELIZA 的回应大多是重复用户先前对话文本输入中的关键问题。魏岑鲍姆(Weizenbaum, 1967: 474)将 ELIZA 描述为“后来被称作 DOCTOR 的程序家族中的第一个程序……一个特殊成员。这些程序的家族姓氏也是 ELIZA”。这组程序是在麻省理工学院的一台 IBM 计算机上用 MAD-SLIP 创

建的,发布时只有一个一般性描述(Weizenbaum, 1966; Weizenbaum, 1967)。在描述程序如何解析句子时,魏岑鲍姆写道:

从左至右扫描输入的句子,在关键词词典里查找各个词。若某个词被识别为关键词(除关键词优先级问题外),则只需尝试包含该关键词的分解规则。(Weizenbaum, 1966: 38)

他接着详细阐述了特定关键词为何需要特殊处理:通常,当一个人使用像“每个人”“总是”或“没有人”这样的全称术语讲话时,实际上指的是某个非常具体的事件或人物。赋予“每个人”比“我”更高的优先级,就能生成“具体说说你想到的是谁”这样的回应(Weizenbaum, 1966: 39)。

魏岑鲍姆还提供了一个用于说明关键词检测的基本流程图,并描述了一些衡量关键词的具体算法:

魏岑鲍姆决定把领域知识放入一个与处理对话的程序模块相分离的模块。他推断,如果用不同的知识模块(他称之为“脚本”)来描述不同种类的知识,程序就能谈论各种各样的话题。今天想谈高级时装而非棒球?只要加载高级时装的软件模块即可!由于程序会像萧伯纳《皮格马利翁》中的伊莱莎·杜立德那样,学着把话说得越来越好,因此魏岑鲍姆将程序命名为《伊莱莎·杜立德》。(魏岑鲍姆后来指出,另一个原因是,“和杜立德小姐一样,它永远不清楚自己是否变得更聪明了”。)(Crevier, 1993: 134)

三、ELIZA 的社会意义

通过开发 ELIZA 程序和 DOCTOR 脚本,魏岑鲍姆揭示了一点:就计算与人的行为和理性的联系而言,它会带来深刻的影响。魏岑鲍姆试图对这些影响进行探索、提出异议并加以限制:

我惊讶地看到,与 DOCTOR 对话的人迅速对计算机产生了深厚的感情,并且毫不犹豫地将其拟人化……这表明,人们以这种方式与计算机对话,就好像它是一个能够与之用亲密的语言恰当、有效地交谈的人。(Weizenbaum, 1976: 6)

事实上,人们倾向于对计算机投入私人情感,这让魏岑鲍姆感到困惑与担忧。他担心机器可能取代人们内心的现实,还担心计算机也许会让“相当正常的人产生强烈的妄想”,并且强化那些将合理性与算计联系起来、认为人就是机器的想法。这便是后来所谓的“ELIZA 效应”,也就是人认为计算机系统有理解力和智能的倾向。侯世达(Hofstadter, 1995: 167)在 1995 年将其描述为“人的这样一种易感性:人们很容易从计算机串联的字符串(尤其是文字串)中读出远超应有的理解”,这个极具说服力的描述也能精确地描述现代的生成式 AI 系统(如 ChatGPT):

在这里,我们要记住,魏岑鲍姆版的 DOCTOR 不是在屏幕上一闪而过,而是或多或少被永久地印在纸上。我们可能会把该系统理解为受到文学和心理治疗影响的系统,留下打印记录(如日记)的治疗过程,给人的体验或许和在屏幕上的短暂接触有所不同。(Montfort, 2004)

的确,如今我们更熟悉出现在屏幕上(无论是台式电脑、笔记本电脑、智能手机还是平板电脑)的 ELIZA。但在 20 世纪 60 年代,计算机系统几乎总是使用电传打字机将输出打印在纸上。这种“屏幕的缺失”,是理解当时计算机的物质特性以及使用系统体验的关键。显然,ELIZA 捕捉到了一种思维模式,以及我们与计算机关系中的一种焦虑。魏岑鲍姆后来也通过对自身工作的哲学、社会和政治基础(我们也是基于此理解他的工作的)进行批判性考察,探索了这种思维模式和焦虑。还需指出的是,软件是一种动态且不断变化的媒介,软件程序很少只有一个版本。实际上,在编程过程中常常会出现多个相互矛盾的副本:

值得注意的是,ELIZA 存在多个版本。魏岑鲍姆承认自己一直在对 ELIZA 进行改进,因为“最初的 ELIZA 除了生成似是而非的回应外,做其他任何事都很困难”。他在 1967 年写到,后来的版本与旧版本主要有两方面的不同:其一,它包含一个求值器,这个求值器能够接受无限复杂的表达式(程序)并对其求值(执行),当然也能够将求值结果储存起来,以供后续检索和使用;其二,脚本的概念得到了扩展,现在程序可以同时包含三个不同的脚本,能够从磁盘存储单元中无限量供应的脚本里获取新脚本,并且现有的脚本之间还可以相互通信。(Weizenbaum, 1976: 478)

遗憾的是,这些后续的改进缺乏充分的记录,许多源代码仍然处于丢失状态。不过,如今人们已经从 MIT 档案中恢复了 ELIZA 的源代码列表以及

1966 年的一些脚本。^①

作为计划的一部分,我也参与了重建 ELIZA 代码的工作。我们已经确定,ELIZA 实际上至少有五个主要版本。其中三个版本已被遗忘(并且可能已经丢失),还有一个最终版本只在其他档案的文字资料中被提及。按照年代顺序,我们暂时识别出以下版本:

1. ELIZA 1965a: 仅以“.”和“,”分隔句子(证据源于 MIT 档案流程图)。
2. ELIZA 1965b: 用“.”“,”和“但是”来分隔句子,缺乏 NEWKEY 函数,包含未被记录过的 CHANGE 函数以及硬编码信息(这是从 MIT 档案中恢复的版本)。
3. ELIZA 1966 CACM: 包括 NEWKEY 函数、关键词堆栈和“but”分隔符。
4. ELIZA 1967: 添加了复杂的脚本处理,证据不仅来自魏岑鲍姆和泰勒的描述,也见于档案中现存的 ARITHM、F29、FIGURE 脚本。
5. ELIZA 1968+: 对这个版本的计划描述见于泰勒。不过,脚本 SPACKS、INTRVW 和 FVP1 也证明,在更为复杂的编程中,通过瓦尔特·丹尼尔(W. Daniel)开发的脚本文本,使用过这个版本。

^① 除 DOCTOR 外,MIT 档案中的脚本如下:ARITHM(可进行对话中的数学计算,求值并返回结果)、F29(似乎是 FIGURE 数学定义脚本的早期版本)、FIGURE(像是 F29 加上更多定义和回应后的拓展版本)、GIRL(简单演示脚本)、NEWENG(用于讨论新英格兰各州)。有脚本输出但无脚本本身的有:ELEVTR(讨论电梯物理学)、POLETA(讨论“竿与谷仓悖论”)、SPACKS(讨论 MIT 人文系引用的贝利·斯帕克斯的一句诗)、SYNCTA(讨论时间同步脚本的名称)。档案中还提及一些既无脚本也无脚本输出的脚本:ANTIPR、INTRVW(据描述是为四向量研究做准备的初步采访)、FVP1、FVQUIZ、CAN-VEC、FRANCE、FORVEC、MIT、ORTH1、PHOTON、RLPOLN、STATES、QMPROB、WATSNU、XYPOLN。相关重建工作可参看 <https://wg.criticalcodestudies.com/index.php?p=/discussion/108/the-original-eliza-in-mad-slip-2022-code-critique>。

魏岑鲍姆认为,ELIZA 开启了人类依赖和信任计算机的先河,而未来这样的例子还会越来越多。然而,这只是更大问题的一个元素而已,这个问题就是:

科学向人类许诺了权力。但就像在人类被权力的许诺诱惑时经常发生的那样,他事先就开始并且在整个过程中都在付出代价,而实际付出的代价,是奴役和无能。权力如果不是选择的权力,那它就什么也不是。工具理性可以做决定,但决定和选择完全不一样。(Weizenbaum, 1976: 259)

当时,人们围绕 ELIZA 和 PARRY 展开了激烈的争论——PARRY 是由肯尼斯·科尔比(K. Colby)于 1972 年设计的一个竞品系统(关于二者的结合,参看 Cerf, 1973)。科尔比试图创造一个具有偏执型精神分裂症特征的聊天机器人,这在当时引发了巨大的争议,并引起了人们对这类软件意义的广泛关注,这些争论也继续影响着当代围绕聊天机器人和 AI 的讨论。

ELIZA 提出的问题在大众文化中得到了更广泛的响应。例如,“火象剧场”(The Firesign Theater)在 1971 年的专辑《我认为我们都是这辆巴士上的笨蛋》(*I Think We're All Bozos on This Bus*)中虚构了它的各种可能性。在这张专辑中,主人公克莱姆去“未来博览会”参观,在那里,他与一台用模糊、听起来积极且与提出的问题几乎无关的回复来回答的计算机交谈。当克莱姆(他的名字被博览会的人格化系统错听成“阿克莱姆”)有机会提问时,他决定黑掉它,通过说“这是工作人员在说话。你好……”进入系统的维护模式;那台名叫 DOCTOR MEMORY 的计算机用“系统状态:运行时间”和运行时长来响应他。接着,克莱姆试图通过提出无法理解,有时甚至

无法解析的问题来使系统混乱,从而使系统崩溃。最后,克莱姆问:“为什么粥鸟要在空中下蛋?”这使得系统停止服务并关闭。这张专辑中出现了一个类似于 ELIZA 的系统,这让许多懂行的人好奇,这样内行的技术行话怎么会出现在一张喜剧专辑上。这一直是一个谜,直到 2015 年,“火象剧场”最初的成员之一菲利普·普洛克特在 Quora 上用一条评论解释说:“我(那张专辑里)所有的计算机语言,都是从我在洛杉矶工作博览会上找到的一张人们与 ELIZA 心理治疗师程序互动的打印资料上学来的。”(Proctor, 2015)

苹果的 iPhone 的 AI 聊天机器人 Siri 后来也提到了这个“大众文化时刻”,它会用“你好阿克莱姆,我能为你执行什么功能?”来回答“这是工作人员在说话。你好……”^①至于这个段子是怎么被放进 Siri 的,根据普洛克特的说法,是“史蒂夫·乔布斯在皮克斯一部普洛克特担任配音演员的电影放映会上遇到他,告诉他自己是‘火象的忠实粉丝’,并要求把这个桥段加入 Siri”(Tannenbaum, 2015)。^②如这里所展示的,在这些例子中,ELIZA 的影响和它提出的问题,从技术领域转移到了文化领域,并加剧了智能机器出现给公众带来的不安。^③

在魏岑鲍姆 1966 年的文章发表后不久,伯尼·科赛尔(B. Cosell)就用 LISP 编写了一个 ELIZA 程序。这个程序以魏岑鲍姆公布的算法为基础,不过重要的是,科赛尔并未见过实际的 MAD-SLIP 源代码。LISP 很快成为 AI 的主要语言,而魏岑鲍姆的 DOCTOR 脚本格式正是 LISP 的符号表达

^① 可惜的是,这个彩蛋在过去几年间的某个时候被移除了。

^② 最近的对话界面(也被称为可听界面)复制了 ELIZA 的诸多特征。这一情况在约翰·凯莱(J. Caylay)的《聆听者》(The Listeners)中也有体现。它是在扬声器或通过扬声器访问的用户与 Amazon Echo 之间进行的语言表演、装置,是由亚马逊分发的第三方应用。

^③ 最近,ELIZA 又出现在阿德里安·柴可夫斯基(A. Tchaikowsky)2015 年的科幻小说《时间之子》(Children of Time)里了。

方式。此外,ELIZA 的原始版本仅在学术圈小范围流传,这些因素很可能加深了那个长期存在的误解,即 ELIZA 一开始就是用 LISP 编写的。需要注意的是,科赛尔的代码也未曾广泛公布,而且该代码直到 2013 年才被恢复。科赛尔最著名的版本是 1966 年的那一版,GNU 的 EMACS 版和杰夫·施拉格(J. Shrager)的 BASIC 版都是基于此版本编写的。后者最初发布于 1973 年,1977 年由《创意计算杂志》(*Creative Computing magazine*)再版。^① 施拉格的 ELIZA 很短,能够快速手动重新键入,并且可以在任何支持 BASIC 语言(当时这种情况很常见)的个人计算机上运行(North, 1977)。^② 如今,多个系统上仍然存在类似的程序,包括手机版和能在网页浏览器上运行的 Java 版。这些版本就运行的程序语言而言稍有不同,并且仅在长度上存在差异——科赛尔的 LISP 版包含 2500 行代码,施拉格的 BASIC 程序仅有 250 行,其他版本则介于两者之间。^③

魏岑鲍姆后来在《计算机的力量和人的理性》中继续阐述操纵和误认的危险。在该书中,他认为计算机会给人类用户带来无法预见的负面影响。魏岑鲍姆早期有将人的工作自动化(如用计算机器取代银行职员)的经验,这影响了他日后对计算机给社会带来挑战的理解。魏岑鲍姆尤其担心人的理性会被有效的程序所取代。对他而言,有效的程序是一种操作,这种操作能让我们在形式数学里定义一个问题域,使该问题域能被机器进行后续计算。而这种思想的形式化又为将人类生活的各方面贬低为机器程序(如把

^① 理查德·斯托曼(R. Stallman)一直被视作 GNU 的 EMACS 版 ELIZA 的作者,但斯托曼告诉我,他并未编写过这个程序(个人通信,2022 年 11 月 13 日)。

^② DOCTOR 是伯尼·科赛尔依照魏岑鲍姆在期刊文章中的描述编写的 LISP 版 ELIZA。伯尼·科赛尔是 BBN(Bolt Beranek and Newman)科技公司程序员团队成员之一。1972 年,该团队制造出两台用作 ARPANET 通信路由的专用计算机,名为接口信息处理器(Interface Message Processor, IMP)。1972 年 9 月,文顿·瑟夫(V. Cerf)录制了 DOCTOR 和 PARRY 的一次对话,并于 1973 年将结果记录为征求意见稿(Cerf, 1973)。

^③ 我要感谢 2016 年与 2022 年的批判代码研究工作团队(Critical Code Studies Working Group),他们对 ELIZA 展开的丰富且详尽的讨论,极大地启发了我对 ELIZA 和 DOCTOR 的思考。

计算当作思考)创造了条件,引出大多数人的这一问题:

根本不理解计算机。所以,除非他们有很强的怀疑能力……否则就只能运用自己能想到的唯一一个类比,即按照自己的思考能力为模型,来解释计算机的智力壮举。(Weizenbaum, 1976: 10)

也就是说,存在一个共同假设:计算机对行为的抽象有助于解释机器的运作。在当代计算社会中,人们有充分的动力采用理性化、插入算法的行动和思维方式,自魏岑鲍姆写下相关文字以来,这种动力有增无减。例如,引入可测量的性能指标和输出标准,以及以计算为前提的监控和监视。我们还发现,在资本主义社会中,相较于互动(对话),常规工作反而增多了。同样严重的是算法系统缺乏合法性,部分原因在于,这些系统在加剧与民主权威相关的结构问题(包括它们在通信和媒体系统中日益广泛的部署)的同时,自身仍然不透明,它们不断增强的影响力表明,它们有可能产生系统性危害,造成危险的系统故障(Berry, 2014)。也许,只要稍微回顾一下2008—2011年的美国次贷危机,就能发现与理性化和计算相结合的“算计式理性”(calculative reason)与以利润为导向的公司和个人会形成多么令人头疼的组合。魏岑鲍姆敏锐地察觉到了这些问题,并试图阐明其中的重要议题,他评论道:“我们的社会对计算机系统(最初,我们只想让它们‘帮助’人们进行分析和决策)的依赖日益加深,早已超出用户的理解范围,并且对用户来说变得不可或缺(……这)是一个非常严重的发展。”(Weizenbaum, 1976: 236)

在这里,魏岑鲍姆所指的是算法化带来的物化与形式化,以及我们构建系统的方式。在创建系统时,人们不再确切知晓或理解系统的运作及决策方式。这就产生了一种危险:因害怕未知的后果,我们不愿修正系统。这进

而引出了增强技术所体现和支持的人的主体间历史概念的创造与维持问题。魏岑鲍姆 1976 年的评论就触及了这一点,他说:“《纽约时报》已经开始建立一个实时数据库……这离让系统来判定什么算事实,进而直接宣布其他所有知识、所有记忆不合法还有多远呢?”(Weizenbaum, 1976: 238)在复杂系统、庞大的数据集合与快速流动的数据流之间进行协调的能力,为社会将“规模”作为一个新的智识视域来运用提供了条件。在业界和知识界的研究中可以看到,人们越来越多地运用计算系统对复杂的“大数据”现象进行抽象、简化和可视化处理,并且倾向于使用过于简单化的因果和统计模型来理解复杂的社会现象,“社会物理学”概念就是一个例证(Pentland, 2015)。魏岑鲍姆可能会评论说,这种“技术、政治和社会的不可避免性”的混合,“对良知来说是一种强大的镇静剂”,它不但赋予计算决定的权力,还赋予其行动的权力(Weizenbaum, 1976: 241)。

魏岑鲍姆依据他和 ELIZA 及其他系统打交道的经验论证称,我们必须对人工智能及其技术发展持批判态度,如此才能理解当下的处境,并且防止其失控:

我不太清楚计算机科学及其分支学科(AI)是否尤其偏爱委婉。我们如此夸张又欣然地谈论能理解、能观看、能决定、能下判断等的计算机系统,却不承认自己在这些概念上的浅薄与极度的天真。而且,在这样的谈论过程中,我们麻痹了评估自身工作质量的能力,更重要的是,麻痹了识别和认知其最终用途的能力。(Weizenbaum, 1987: 44)

四、现代的迭代

魏岑鲍姆指出，人的思想和劳动自动化存在危险。这要求我们以一种拒绝忽视和平息现象层面的矛盾及矛盾主张、努力把握主体动态时刻(*dynamic moment*)的方式来解释。这种解释方式要始终为批判反思理解算法的历史和传统留出空间，接受传统意义结构的重要性，同时力求避免将其理想化。也就是说，在这些计算社会中，传统可能继续表现为基于欺骗和扭曲的互动，而这种互动往往会被不加反思地转换为算法形式，用魏岑鲍姆的话来说：

我们所讨论的各种系统的程序都具备一些非常重要的特征：从某种意义上说，它们都很简单；它们都会扭曲和滥用语言；在否认规范内容的同时，它们也都支持一种基于专业知识的威权主义。（Weizenbaum, 1976: 248）

现在，为进一步强调这点，我想简要考察一个例子，它将说明 ELIZA 的界面可能性与我们对计算机功能的预期为何可能以极其反人类的方式互动。在这个例子里，社会冲突内嵌于算法机器，劳动通过界面被转化为一种商品。这个例子就是用于隐藏社会劳动，使工人隐藏在网页表单、聊天机器人和应用程序接口（这些能让雇主将自己视为创新技术的建造者，而非不关心工作条件的老板）背后的计算（Irani & Silberman, 2013: 613）。在此语境下，算法造就的“自由”和“创造”可能实际上只是掩盖了控制和管理系统而已。

为理解社会冲突如何隐藏于算法形式之中，研究在新计算技术条件下

才成为可能的现代生成式 AI 系统(或聊天机器人)是有益的。从某种意义上讲,这是“大脑之间的合作”。斯蒂格勒(Stiegler, 2010: 47)指出,这种合作“通过语法化系统产生,该系统使神经系统活动最高水平上的所有质性任务的无产阶级化成为可能”。关键是,这是对劳动力的实时抽象,将劳动力抽象为一种潜能,一种像电力或水一样按需持续供应的“劳动力流”。经由聊天机器人和应用的调节,这种抽象造就了一种高度异化的劳动力。从某种意义上说,这种劳动力实现了控制论系统,该系统试图在紧密联系的反馈回路中让机器与人紧密耦合。这也描绘了魏岑鲍姆在《计算机的力量和人的理性》中所警告的威胁。

这方面的范例当属 OpenAI 的 ChatGPT。该公司首席执行官称 ChatGPT 能改变资本主义:“技术的改进及其对社会的积极影响将持续指数级增长。”(Konrad & Cai, 2023)ChatGPT 是基于生成式预训练 Transformer 模型 [大语言模型(LLM)的一种] 的聊天机器人,LMM 是基于真实对话、文本以及像 Reddit 那样的平台上的文章训练而成的。它让当代观察者感到惊奇,因为通过输入和问题来提示系统,它似乎将书面文本的生产机械化了。同时,系统的力量实际上隐藏在聊天机器人的“界面”之后。这种既通过软件汇集人类文化生产,又将其作为计算系统储备资源的想法,体现了魏岑鲍姆有先见之明地警示过的计算社会中流行的控制论思想。从很多方面来看,这使人类活动变得离散。不过,这也是人通过用于协调更普遍文化生产的计算层的非人化。它还表明界面如何将表面之下进行的社会劳动物化,从而使机器能够真正将数百万类写作和对话的痕迹据为己有,而这一切都是人们没有意识到且未曾同意的。

这也开启了按需求进行管理、控制、监控、分解与再集合的自动化文化生产的可能性。资本主义公司将其操作化后,就有造成社会休克的危险,因为新的技术实践往往能避开劳动法与劳动保护的监管,还能带来巨大利润。

这可能扭曲更广泛的经济,这是缺乏政府监管和工会等劳动组织监督的结果。或许我们还会发现,未来的社会冲突与这种自动化系统的扩大使用有关。实际上,它们营造出一种算法拜物教的感觉,劳动被算法掩盖,这意味着通过面板界面和编程代码互动的程序员或用户无须承认劳动的社会性。对界面另一边的工人来说,对能动性、体力劳动、情感控制和自我规训的要求可能造成严重的心理紧张(在失范与宿命论之间切换)。消费者或客户的一条差评就可能立即导致失业,而且资方几乎不会给出解雇原因。这是否会为必要性政治和社会学反身性创造条件还有待观察,但它肯定是传统工人阶级之外未来劳动祛魅的一个可能来源。我们在当前流行的基于文本界面的聊天机器人中也能看到类似的例子。和 ELIZA 一样,这些聊天机器人也是部分自动化、部分由人来管理客户的询问和问题(Marino, 2006)。不用说,ELIZA 启发创造认知工厂的计算系统这一想法会让魏岑鲍姆感到恐惧。

通过使用这种新范式,运用生成式 AI 的系统指出了新一类知识—权力化界面的各种可能性。在此,机器被界面模糊,它可能是人类劳动、聊天机器人或者其他形式的 AI。从界面用户的视角看,没有任何差别。这个程序将关系物化,在用户与底层过程之间创造出一种命令—执行关系。对手段的模糊(不论它可能编码何种剥削关系)以及在计算系统中对真实人类进行理性和抽象所带来的后果,也让人联想到魏岑鲍姆 1976 年的评论:

当个体不被当作完整的人来对待时,就被非人化了。各种形式的人类和社会工程……在绕过人的语境,尤其是绕开那些赋予人的语言以真实意义的语境时,就在做这样的事。(Weizenbaum, 1976: 266)

五、关于生成式 AI 意义的反思

通过考察先前有关早期计算的论证和讨论，我们找到了一种反思当下困境与问题的有用方式，这些困境与问题出现在我们急于将社会合理化和计算机化之时。身为计算科学家，魏岑鲍姆乐于反思自己的实践及其工作的意义：

这应该教会我们什么呢？特别是就至少保住甚至增加人在人类事务中的选择这一问题而言。当然是，与其说构建可靠的计算机软件依靠计算机科学的研究成果，不如说是对人类状况更深入的理论理解。
(Rosenberg, 1980: 103)

虽然 ELIZA 放在今天或许显得原始，但它仍反映出诸多设计决策，这些决策在我们如今使用的系统中依旧有所体现，比如人机如何互动，以及应在多大程度上允许计算管理我们的生活与心智等。它还表明，早期软件的发展脉络对于理解我们在何种程度上必须始终对过去的假设以及技术可能且应该如何发展的想法保持质疑非常关键。虽然如今我们不会把 ELIZA 视为 AI（即便过去我们曾有过这种错误认知），但我们也应认识到，类似的工具合理性仍然内置于机器学习和当今的 AI 系统之中。

今天的计算机科学环境与魏岑鲍姆当时的环境有实质性不同。他在学院环境下工作，虽得到美国国防部 MAC 计划的项目资助，却从实验研究视角探索技术（如 ELIZA）。如今，AI 日益被资本主义需求所涵盖（Berry, 2021; McQuillan, 2023）。多数 AI 的发展趋向计算的历史规范，包括强调技术层面将控制与执行分开的能力。计算的批判史显示，这种方法常被用

于与资本主义相符的进程，其中隐含一套假设，如在构建社会关系时对市场优先性的先验假设。其次——在我看来与之不无关联——当前理解 AI 的方式倾向于促使人们对其运作进行形而上学或形式主义的辩护与解释。这可能使思想的数学化成为一种价值，在这种价值观念下，思想的形式化不仅是思考 AI 的一种方式，而且是唯一可效仿的方式。这可能催生一种倾向于理念论，却不关注谁拥有和控制新“认知手段”的计算理论，就如布莱兹·帕斯卡尔在 17 世纪评论的那样：

数学的心智 (*esprit de géométrie*, 几何学精神) 和直觉的心智 (*esprit de finesse*, 敏感性精神) 是存在差别的。数学家不讲直觉，原因在于他们看不到眼前的事物。他们习惯了明确、显而易见的数学原理，习惯在正确理解和处理原理之前就得出结论，所以在需要直觉的问题中就会迷失。在这类问题里，原理不能像之前那样被处理……这些事物极为微妙且众多，需要一种非常微妙和精确的性情去感受它们，然后依据这种感知做出准确、正确的判断。最常见的情况是，我们无法像在数学中那样从逻辑上对其进行证明，因为我们认识原理的方式不同，那样做会是一项无尽的任务。“真”必须直接去看，一眼便知，至少在一定程度上不需要通过推理过程。所以，数学家很少按直觉行事，凭直觉行事的人也很少是数学家。因为数学家想用数学的方式处理直觉方面的事物，他们荒谬地想从定义开始，再按照原理……倒不是说心智不会这么做，而是说，心智会沉默、自然而简单地这样做。（Pascal, 1995: 151，译文有改动）

在不否认人文学科或社会科学对 AI 或机器学习的研究方法尚不成熟的情况下，我想魏岑鲍姆会认同，AI 思考中的严重问题确切地说正是这样

一个假设：数学化或形式化是理解的前提。毕竟，魏岑鲍姆本人解释过：

他们一开始可能只是相信，自己采用的运算不过是在描述所处理现象时一种便捷的速记方式。但随着他们用从外部语境借来的基本要素构建出越来越庞大的概念框架，为这些框架命名，并且像操控更复杂思想体系的元素一样操控它们，这些框架就不再仅仅是描述方式了，而是像马斯洛的锤子一样，成为他们世界观的规定。（Weizenbaum, 1976: 102）

随着以 ChatGPT 这类系统为代表的机器学习不断进步，人类思维能力被毁灭这一危险（可能会被误解为只是 AI 偶然的副作用）已随之而来了。不过，对于任何技术而言，在创造技术、应用技术以及围绕技术在社会中的使用进行限制和管制方面，人类总有选择。魏岑鲍姆富有先见之明地认识到，无知、无意义的机器可能会对我们的理解能力、主体间交流能力以及基本人性构成挑战。他发现人类对彼此存在自然的同理心，也明白这种同理心可能会被错误地导向计算机，还认识到对于技术主义者来说，利用这种同理心误导或欺骗人们以谋取利益、煽动或说服他们，乃至破坏民主本身具有巨大的诱惑。我们正生活在 ELIZA 所指向，而魏岑鲍姆试图警示的未来。现在，听从他的警告，明确可计算事物的限度是我们的责任。^①

首先，魏岑鲍姆开发 ELIZA 的初衷是探索计算机与人类交互的可能性。在当时的技术环境下，这是一项极具开创性的工作。他的贡献不仅在于创造了一个可模拟对话的程序，更在于引发了人们对人工智能本质的深

^① 在这里，我对“不可计算的东西”的理解，是从哪类问题适合计算的批判意义出发的，而非依据关于什么能或不能被算法再现的更具技术性的定义。在此语境下，不可计算的东西就是被社会定义为不适合计算的那些问题。

入思考。

在生成式 AI(如 ChatGPT)蓬勃发展的当下,魏岑鲍姆批判性更强的著作显得格外重要。他曾警示过过度依赖人工智能可能产生的风险,像人机关系失衡以及对人类思维方式的潜在影响等。ELIZA 虽是较为初级的程序,却已展现出计算机模拟人类对话的潜力,当代的机器学习和 AI 技术更是将这种潜力发挥到了新高度。

ELIZA 相关讨论为当下机器学习与 AI 的争论提供了独特视角。一方面,ELIZA 使人们认识到计算机在一定程度上能模仿人类语言行为,这为机器学习奠定了早期理论与实践基础;另一方面,人们也意识到仅追求语言相似性并不代表计算机具备人类智能。比如,ELIZA 的回应常基于预设模式和算法,缺乏真正的理解与情感。

当代,生成式 AI 面临诸多挑战与争议,如数据隐私、算法偏见、对就业市场的冲击等。魏岑鲍姆的思想能引导我们更谨慎对待这些发展。我们不能只陶醉于 AI 带来的便利和高效中,还得关注其背后可能潜藏的风险。同时,对于 AI 的发展方向,我们应更注重让其与人类社会更好地融合,而非单纯追求技术超越。

总之,考察魏岑鲍姆对 AI 的贡献并讨论 ELIZA,有助于我们更好地理解当前机器学习和 AI 发展中的关键问题,从而在这个快速发展的领域保持清醒头脑,引导 AI 朝着对人类有益的方向发展。

参考文献

- Arendt, H. 1972, *Crises of the Republic*, [S. l.] : Harcourt Brace Jovanovich Publishing.
- Berry, D. 2011, *The Philosophy of Software: Code and Mediation in the Digital Age*, [S. l.] : Palgrave.
- Berry, D. 2014, *Critical Theory and the Digital*, [S. l.] : Bloomsbury.

- Berry, D. 2021, "Explanatory Publics: Explainability and Democratic Thought." in B. Balasas & C. Rito eds., *Fabricating Publics: The Dissemination of Culture in the Post-Truth Era*, [S. l.] : Open Humanities Press.
- Boden, M. 1977, *Artificial Intelligence and Natural Man*, Cambridge, MA. : MIT Press.
- Cerf, V. 1973, "RFC 439-PARRY Encounters the DOCTOR." Faqs. org. <http://www.faqs.org/rfcs/rfc439.html>.
- Crevier, D. 1993, *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books.
- Dembart, L. 1977, "Experts Argue Whether Computers Could Reason, and If They Should." <https://www.nytimes.com/1977/05/08/archives/experts-argue-whether-computers-could-reason-and-if-they-should.html>.
- Fisher, A. & J. McKenney 1993, "The Development of the ERMA Banking System: Lessons from History." *IEEE Annals of the History of Computing* 15(1).
- Hofstadter, D. 1995, *Fluid Concepts and Creative Analogies*, [S. l.] : Perseus Books.
- Irani, L. & M. Silberman 2013, "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." <https://doi.org/10.1145/2470654.2470742>.
- Konrad, A. & K. Cai 2023, "Exclusive Interview: OpenAI's Sam Altman Talks ChatGPT and How Artificial General Intelligence Can 'Break Capitalism'." <https://www.forbes.com/sites/alexkonrad/2023/02/03/exclusive-openai-sam-altman-chatgpt-agi-google-search>.
- Marino, M. 2006, *I, Chatbot: The Gender and Race Performativity of Conversational Agents*, Ph. D thesis of University of California.
- Maurer, D. 1974, *The American Confidence Man*, [S. l.] : Charles C. Thomas Publisher.
- McCarthy, J. , M. Minsky & N. Rochester et al. 2006, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." <https://doi.org/10.1609/aimag.v27i4.1904>.
- McCorduck, P. 2004, *Machines Who Think*, [S. l.] : A. K. Peters.
- McQuillan, D. 2023, "Predicted Benefits, Proven Harms: How AI's Algorithmic Violence

- Emerged from Our Own Social Matrix.” <https://doi.org/10.51428/tsr.ekpj9730>.
- MIT 2008, “Joseph Weizenbaum, Professor Emeritus of Computer Science, 85.” <https://news.mit.edu/2008/obit-weizenbaum-0310>.
- Montfort, N. 2004, “Continuous Paper: The Early Materiality and Workings of Electronic Literature.” http://nickm.com/writing/essays/continuous_paper_mla.html.
- North, S. 1977, “Eliza.” *Creative Computing* [s.l.] .
- Pascal, B. 1995, *Pensées and Other Writings*, Oxford: Oxford University Press.
- Pentland, A. 2015, *Social Physics: How Social Networks Can Make Us Smarter*, London: Penguin.
- Proctor, P. 2015, “What Are the Best Siri Replies?” <https://www.quora.com/log/revision/82373474>.
- Rheingold, H. 1985, *Tools for Thought: The History and Future of Mind-Expanding Technology*, Cambridge, MA.: MIT Press.
- Rosenberg, R. 1980, *Incomprehensible Computer Systems: Knowledge Without Wisdom*, Ph. D thesis of MIT.
- Sack, H. 2018, “Joseph Weizenbaum and His Famous Eliza.” <http://scihi.org/joseph-weizenbaum-eliza>.
- Samuel, A. 1959, “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3(3).
- Shannon, C. 1950, “XXII. Programming a Computer for Playing Chess.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41(314).
- SRI International 2016, “About Us.” <https://www.sri.com/about>.
- Stiegler, B. 2010, *For a New Critique of Political Economy*, London: Polity.
- Tannenbaum, S. 2015, “With One of Its Easter Eggs, SIRI Evokes the Firesign Theatre, Hacker Culture, a 1960s Chatbot, and Steve Jobs.” <https://medium.com/@stannenb/with-one-of-its-easter-eggs-siri-evokes-the-firesign-theater-and-steve-jobs-86ea5b4874d3#.rn9smlfgm>.
- Weil, P. 2015, “Seriously Writing SIRI.” doi:10.20415/hyp/011.e05.

- Weizenbaum, J. 1962, "How to Make a Computer Appear Intelligent." *Datamation* 8(2).
- Weizenbaum, J. 1963, "Symmetric List Processor." *Communications of the ACM* 6(9).
- Weizenbaum, J. 1966, "ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine." *Communications of the ACM* 9.
- Weizenbaum, J. 1967, "Contextual Understanding by Computers." *Communications of the ACM* 10(8).
- Weizenbaum, J. 1972, "On the Impact of the Computer on Society." *Science* 176.
- Weizenbaum, J. 1976, *Computer Power and Human Reason: From Judgement to Calculation*, [S. l.]: W. H. Freeman and Company.
- Weizenbaum, J. 1987, "Not Without Us." *ETC: A Review of General Semantics* 44(1).

编委会主任：高 岩
编委会副主任：夏桂华 赵玉新
吕 鹏（中国社会科学院）
编 委：尹 航 冯仕政 冯全普
(按姓氏笔画排序) 吕 鹏（中南大学） 吕冬诗
朱齐丹 汝 鹏 苏 竣
李正风 来有为 肖黎明
邱泽奇 何晓斌 宋士吉
陈云松 陈华珊 郑 莉
孟小峰 孟天广 赵万里
赵延东 胡安宁 袁 岳
黄 萍 梁玉成 董 波
曾志刚 蔡成涛

青 年 编 委：丁奎元 王 磊 叶瀚璋
(按姓氏笔画排序) 邢麟舟 向 维 刘灿辉
刘松吟 刘春成 刘晓波
安 博 许馨月 孙宇凡
李子信 李天朗 李晓天
吴雨晴 何 丽 邹冠男
张咏雪 张承蒙 陈 苗
陈典涵 林子皓 周雪健
周骥腾 郑 李 胡万亨
茹文俊 贺久恒 贾雨心
郭媛媛 黄 可 梁 轩
曾 晨

编辑团队
主 编：郑 莉
编辑部主任：吴肃然
编辑部成员：林召霞 王立秋
李昕茹 李天朗
岳 凤

主管单位：中华人民共和国
工业和信息化部
主办单位：哈尔滨工程大学
出版单位：哈尔滨工程大学
出版社
地址：哈尔滨市南岗区
南通大街 145 号

国际标准连续出版物号：
ISSN 2097-2091
国内统一连续出版物号：
CN 23-1615/C

印刷单位：哈尔滨理想印刷有限公司
创刊年份：2022 年
出版日期：2024 年 11 月 10 日
发行单位：哈尔滨市邮局
订 阅 处：全国各地邮电局
邮发代号：14-375
发行范围：公开发行
定 价：45.00 元

投稿指南

本刊面向海内外学者征稿，欢迎社会科学及交叉学科的专家学者惠赐稿件。请在来稿首页写明文章标题、作者简介（姓名、工作单位全称、联系电话、详细通信地址、电邮地址等）。文稿需完整，包括标题（中英文）、作者姓名、作者单位、摘要（300 字左右）、关键词（3—5 个）、正文、参考文献等。所投稿件如受基金资助，请在标题上加脚注说明，包括项目全称和项目批准号。来稿请以中文撰写。

稿件采用他人成说的，须在文中以括注方式说明出处，并在篇末列出参考文献；作者自己的注释均作为当页脚注。中外文参考文献分开列出，中文文献在前，外文文献在后，并按音序排列。中文文献参照中文社会学权威期刊格式，外文文献参照 APA 格式。来稿中的图表要清晰，符合出版质量要求，必要时可单独提供图表压缩包文件。

稿件格式请参考杂志官网 (<http://www.jis.ac.cn>) “下载中心” 中的稿件模板。

投稿方式：请登录杂志官网投稿系统 (<http://www.jis.ac.cn>) 进行投稿。

编辑部联系方式

地 址：黑龙江省哈尔滨市南岗区南通大街 145 号哈尔滨工程大学主楼
北楼 N301 室，《智能社会研究》编辑部
邮 编：150001
电 话：0451-82588881
E-mail：mailto:jis@163.com

著作权使用说明

本刊已许可中国知网等网络知识服务平台以数字化方式复制、汇编、发行、信息网络传播本刊全文。本刊支付的稿酬已包含网络知识服务平台的著作权使用费，所有署名作者向本刊提交文章发表之行为视为同意上述声明。如有异议，请在投稿时说明，本刊将按作者说明处理。