

目 次

论文

平台企业的灵活积累机制

——一个合约理论的视角 滕 飞(1)

信息逆差:代码权力膨胀与公众权力让渡 施颖婕(21)

大语言模型能用来给主题模型中的主题编码吗?

——兼论大语言模型在社会科学领域应用的前景

..... 李 代 张博伦 周浥莽(33)

数字时代的科技公司与国家治理

——一个综合性的分析框架 钟瑞雪(60)

新闻大数据视角下测度区域一体化水平的新路径

——以长三角地区为例 李 军 董方杰(84)

解密数字社会中的幸福密码

——不同互联网使用程度下社会信任感和社会公平感对主观幸福感的

影响机制研究 郭媛媛 张 腾(103)

研究报告

- 数字时代安全科技价值报告 …… 陆菲菲 秦 昊 张友红 吕 鹏(133)
- 问题触发的算法模型响应机制探索 …… 许正军 袁 岳(153)

译文

数字健康的社会技术伦理

——生命伦理学进路的批判和延伸

…… 詹姆斯·肖 约瑟夫·多尼娅 著 何 丽 译(166)

书评

数字资本主义和智能技术的当代批判

——评《过度智能》 …… 周银知(186)

访谈

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼 …… 露西·萨奇曼 胡万亨(198)

CONTENTS

THESIS

The Flexible Accumulation Mechanism of Platform Enterprises: A Contract Theory Perspective	Teng Fei(1)
“Information Deficit”: Code Power Inflation and Public Power Cession ...	Shi Yingjie(21)
Can LLM Label Topics from A Topic Model: The Future of LLM in Social Science	Li Dai, Zhang Bolun, Zhou Yimang(33)
Technology Companies and National Governance in the Digital Age: A Comprehensive Analytical Framework	Zhong Ruixue(60)
A New Approach to Measuring Regional Integration from the Perspective of Big Data in News: A Case Study of the Yangtze River Delta Region ...	Li Jun, Dong Fangjie(84)
Decrypting the Happiness Code in the Digital Society: A Study on the Influence Mechanism of Social Trust and Fairness Perception on Subjective Well-Being under Different Internet Usage Levels	Guo Yuanyuan, Zhang Teng(103)

RESEARCH REPORTS

Report on the Value of Security Technology in the Digital Age	Lu Feifei, Qin Hao, Zhang Youhong, Lyu Peng(133)
Exploration of Algorithmic Model Response Mechanisms for Problem Triggering	Xu Zhengjun, Yuan Yue(153)

TRANSLATED TEXT

The Sociotechnical Ethics of Digital Health: A Critique and Extension of Approaches from

Bioethics written by J. Shaw, J. Donia; trans. by He Li(166)

BOOK REVIEW

A Contemporary Critique of Digital Capitalism and Smart Technologies: Review of *Too*

Smart Zhou Yinzhi(186)

INTERVIEW

Thinking Machine Intelligence Through Human Actions: An Interview with Anthropologist

L. Suchman L. Suchman, Hu Wanheng(198)

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼

露西·萨奇曼 胡万亨*

随着机器学习技术的快速发展和广泛应用,人工智能已成为当今人文社会科学领域的热点议题之一。实际上,早在 20 世纪 70 年代,以专家系统(expert system)为主流的早期人工智能研究就吸引了一批杰出的人文社科学者的关注,他们从哲学、社会学和人类学等角度做了深入的分析,对于我们理解人工智能在当下的发展仍有重要的借鉴意义。其中,著名人类学家露西·萨奇曼(L. Suchman)在 80 年代对美国施乐公司(Xerox)帕洛·阿尔托研究中心(Palo Alto Research Center, PARC)一款基于专家系统的交互式界面打印机的研发过程进行的详尽的民族志研究(Suchman, 1987; Suchman, 2006),是这方面的经典作品之一。萨奇曼目前为英国兰卡斯特大学社会学系荣休教授,近年来持续关注和研究人工智能技术的进展。她是人机交互(human-computer interaction, HCI)领域的重要开拓者之一,曾获美国语言学会“计算机与认知科学本杰明·富兰克林奖章”和美国计算机协会人机交互特别兴趣组“终身研究奖”。同时,萨奇曼长期活跃在科学技术学(science and technology studies, STS)领域,是科学的社会研究学会(Society for Social Studies of Science, 4S)“贝尔纳奖”(Bernal Prize)的获得者——这是该学会的最高奖项,历届获奖者包括 STS 领域的奠基者以及为增进对科学技术之社会维度的理解做出杰出贡献的学者。在本次访谈中,萨奇曼分享了自己作为一名人类学家在施乐公司 20 余年的学术研究和工作经历,对包括专家系统和机器学习在内的人工智能技术的长期思考,围绕当代自主武器系统(autonomous weapon systems)的批判性研究和社会活动,以及对人文社科学者如何研究及影响技术实践等话题的看法。

* 露西·萨奇曼(L. Suchman),英国兰卡斯特大学社会学系。胡万亨,美国康奈尔大学科学技术学系。

一、科技行业的人类学家

胡万亨:您在加州大学伯克利分校期间,从本科到博士接受的都是人类学训练。您是如何开始对技术,尤其是对机器和人工智能产生兴趣的呢?

萨奇曼:这其实挺偶然的,可能对所有的博士生来说都是如此。因为我刚开始读博时,根本就没有考虑过技术。20 世纪 70 年代,美国人类学界掀起了一股叫“向上研究”(studying up)的浪潮。这个说法出自人类学家劳拉·纳德(L. Nader)70 年代发表在《重塑人类学》(*Reinventing Anthropology*)这本书中的一章(Nader, 1974),她是我在加州大学伯克利分校第一年的导师。这本书收录了一系列文章,以反思人类学领域,尤其是美国人类学的政治性。由于人类学有着殖民主义的历史渊源,这本书试图思考如何转变这一领域,使之服务于其他不同的政治价值。

纳德的论点是,人类学应该将注意力转向权力机构、权力中心。我深受启发。我那时也受到了互动论(interactionism)方法的很大影响,这跟当时人类学的主流取向有所龃龉。我在伯克利最后选了杰拉尔德·贝里曼(G. Berreman)做导师,他是一位互动论人类学家,从互动论视角对印度的种姓问题做了非常出色的研究。他对我非常支持。我当时有个想法,即用一个在我现在看来属于后结构主义的方法来看待权力的日常运作机制,更具体地说是企业中权力的运作。

不过,作为加州大学伯克利分校的学生和 20 世纪 60 年代成长起来的孩子,我跟任何大公司的联系都为零,我没有任何在大公司的人脉。但我发现,我其实还是可以通过学校的渠道联系到大公司的——伯克利分校人类学系的语言实验室跟施乐公司的帕洛·阿尔托研究中心有合作。于是,我们人类学系的博士生就可以通过这层关系进入 PARC,我之前的文章写到过这一点(Suchman, 2013)。

由于伯克利分校跟施乐公司的关系,我就直接去了 PARC。我对 PARC 可以说是一无所知,而且那时我连计算机都没有见过呢——在我那个年代,这可想而知,对吧?那个时候计算机还不普及。我来 PARC 根本就不是出于对技术的兴趣,只是后来才发现这里的技术真的太值得研究了。可以这么说,我在某种程

度上偏离了最初的目标。尽管我在那里学到了很多关于公司的知识,但我还是有些“分心”了,我被 PARC 当时在交互式计算(interactive computing)领域所进行的工作分散了很多注意力。

那个时候正是交互式计算发展的早期阶段。1978 年我初到 PARC 时,人机交互的想法还非常新,我立马就对这种将计算机界面作为互动场所的理念产生了兴趣。我开始觉得,对人与计算机接触的过程中发生的事情进行一些互动论分析,应该会非常有趣。与此同时,我所在的实验室里也有认知科学和人工智能方面的研究项目。那时还是规划模型(planning model)在人工智能和专家系统中占主导地位的时代。

慢慢地,我越来越多地参与到这个项目中来。它要打造一个能够指导用户使用人机合作型复印机的专家系统原型,这在当时听起来还是挺疯狂的。就这样,我对研究技术产生了浓厚的兴趣。与此同时,我接触到了 STS,它给我的研究建立了一个参考框架,也让我加入了一个由其他社会科学家组成的圈子,他们以非常有趣的方式思考技术。

胡万亨:您最初感兴趣的是用互动论视角去考察权力运作,为了研究企业中的权力而来到施乐公司。当您看到人类与机器或复印机之间的互动时,发现用互动论来思考也很合适,这样说起来确实很自然。

萨奇曼:没错,我立马就意识到了这一点。并且,我对人机互动过程中所发生的事情,在多大程度上类似于人人互动是持怀疑态度的。这就是我的研究的起点。这个“类似”在哪些方面可以成立,它的限度在什么地方,又在哪里无法成立?这就是我的博士论文要回答的问题。

胡万亨:您对人机交互产生兴趣时,还在一个人类学项目里学习。在 20 世纪 80 年代,作为人类学家去研究技术是种什么感觉?我的印象是,这在当时应该不是传统的路数。人们对这类研究的反应如何?以及,那时的技术人类学领域大致是个什么情况?

萨奇曼:确实不太传统。当然,那时候已经有一些重要的学者做了相关研究。比如,莎伦·特拉维克(S. Traweek)基于对一个粒子物理实验室的民族志

研究,出版了《物理与人理:对高能物理学家社区的人类学考察》(Traweek, 1988)。但这本书更多是对科学实践的民族志,它的田野地点虽然不寻常,但就方法而言,我觉得它比我的研究更为传统。幸运的是,我的导师很有同理心,他的思想非常开放,对我也很支持。尽管我的博士论文离他所做的研究越来越远,我们仍然有互动论的共同研究取向,而且我们在关心的事情上有着相近的政治立场。他调解了我与系里的关系,帮助我渡过了博士生所要经历的各种关卡。但不得不说,与 STS 的联系对我来说越来越重要——那时不断发展的 STS 圈子,在研究取向上跟我正在研究的东西要一致得多。

还有一点,我的博士论文委员会里还包括赫伯特·布鲁默(H. Blumer)——一位出色的符号互动论学者,以及哲学系的休伯特·德雷福斯(H. Dreyfus)。所以,我的博士论文委员会里有着更了解我论文的实质主题的老师——不用说,尤其是德雷福斯。

胡万亨:原来如此!我们来谈谈您在施乐公司实习的经历吧!您在参与观察和民族志研究时,肯定也会有一些具体的工作职责吧。您在施乐公司具体做什么呢?

萨奇曼:这是一个非常有趣的过程,也是我融入 PARC 这个过程。跟我找到这份实习工作一样,我的工作内容也是有迹可循的。团队成员之前也接触过人类学家,因为在我来之前,埃莉诺·韦恩(E. Wynn)就曾在 PARC 待过,她的博士论文是关于办公信息系统和办公自动化的。我最开始实习时,加入的是“办公室研究小组”,那个时候还是办公自动化的早期阶段。我的同事对工作实践(work practices)很感兴趣,这在某种程度上是一种基础性的研究工作。但我到 PARC 后不久,“办公室研究小组”就解散了——企业界就是这样,上头下来个什么指令,一切就风云突变了。所以,我那时基本上是被另一个小组收留了,也就是约翰·布朗(J. Brown)领导的“认知和指导科学小组”。

我刚到时,基本上就是凭感觉干活。那是个暑期实习,我是埃莉诺的研究助理。她刚刚完成博士学位论文,正在做一项研究——施乐公司那时正在研发 Alto 计算机,并将其视为办公信息系统。他们决定进行一系列所谓的“探索”(probes),也就是将 Alto 引入工作场所。一开始,这些基本上都是在施乐的行

政办公室进行的,属于某种内部探索或实验。埃莉诺在施乐位于湾区的一个行政销售办公室做了一系列预访谈,然后给他们装备上 Alto 计算机。一段时间后她再回访,看看会发生什么。我最初是作为暑期实习生来转录这些访谈的,我在一台 Alto 上做了转录,用的是一个名为 Bravo 的早期文字处理包。

这份转录工作是一个学习实际使用 Alto 的机会——它还蛮好用的——也是一个在暑假期间结识同事的机会。在那段时间,我一直在留心,想弄清楚我自己在那里实际上可以做什么。暑期实习结束后,我获得了一个研究实习的机会,来进行我的博士学位论文研究。

所以,我经历了几次不同的转变。从一开始更多关注工作和自动化,到后来更多关注人机交互、人工智能和认知科学领域。PARC 是一个非常适合探索和思考的地方:一方面,我有很大的自由去关注有趣的事情;但另一方面,我必须不断努力让同事能够理解我,并且思考如何把我关心的问题变成一个博士论文题目,让我的博士论文委员会也可以理解,觉得有趣,同时还要结合 PARC 这边正在进行的实际工作。也就因此,这是一个共同打造研究项目并找到进入田野空间的方法的过程。我很幸运能够真正被施乐雇用,也同时攻读博士学位,很幸运能够来到这个非凡的地方,拥有如此多十分有趣的可能性。

话说回来,你也得时刻准备着,找到一些方法让自己变得容易理解,让周围的人对你的工作感兴趣。这些人基本上都是计算机科学家,还有一些是认知心理学家。所以,你得做大量的“转译”工作,而且尽量不要单纯地去迎合他们的想法,把自己完全“转译”成他们眼中“我可能是谁”“我可能在做什么”的样子。

胡万亨:您毕业后继续留在施乐公司工作,并且度过了 20 年的职业生涯。您能谈谈决定留在科技行业尤其是施乐的考虑吗?

萨奇曼:当然。我完成博士学位论文后,拿到了施乐的一个实际上相当于博士后的职位,得以暂时延续我在那里的工作。那时我对施乐已经足够熟悉,而且施乐也有足够多的潜在研究方向,我也有能力推进这些方向。所以,继续留下来就是一个自然而然的选择,这是一个继续推进我手头工作的绝佳机会。

我的博士后差不多快结束时,我跟当时的上司约翰·布朗说:“我在想博士后做完后的安排。”他大意说:“哦,我们给你提供了一个长期职位。我还没跟你

说吗?”所以,他其实已经直接把我调到了长期岗位上。这个情况之所以发生,是因为当时 PARC 还处于早期发展阶段,还没有以任何方式规范他们的招聘流程。我觉得当时应该没有人反对,但这个决定确实是在十分非正式的基础上进行的。那时,我已经深深地融入了公司,并且可以看到很多有趣的方向,所以完全没有问题。我对能够待在那里非常非常开心。正如你提到的,我在那里度过了接下来的 20 年。因此,从我拿到博士学位过渡到人生中的下一个重要阶段,顺利得出乎意料。

胡万亨:您在 2000 年加入兰卡斯特大学,回到了学术界。您是如何决定重返学术界的呢?

萨奇曼:我会说有推力——有拉力,也有推力。对我和我的社会科学同事来说,维持与 PARC 和施乐公司的关系变得越来越困难,我们越来越多地被要求扮演服务性的角色。就像我之前说的,这类角色对今天工业界的社会科学家来说是最常见的。彼时,我们在 PARC 成立了一个名为“工作实践与技术”的研究小组,旨在探索如何与相关工作场所中的用户合作开发一些信息技术原型机。

在此期间,我们与施乐最感兴趣的一些组织开展了一系列项目,与施乐产品项目部门以及全公司出色的界面设计师密切合作。在这样的合作过程中,我们小组成员之间变得密不可分。我们有一些自己的长期项目,而且我们之间的紧密联系也是非常必要的,因为我们想要跟加州交通部的土木工程师或硅谷一家大型律师事务所的律师建立合作,我们决心开展几个长期项目。与此同时,PARC 的风向也在发生变化。PARC 的研究议程经常变换,这会儿说“每个人都应该去那里”,过会儿又说“每个人都应该朝另一个方向去”。我们对此持抵制态度,我们坚持遵循此前积累下来并且认可的研究程序。

就我个人而言,一个决定性时刻是我跟当时的经理交流意见时,他告诉我,他认为我抗拒改变!他怎么能说我抗拒改变呢?但我意识到,当他这么说时,他的意思是我抗拒改变以适应新的议程和新的工作方式。我想,我和我的研究小组那时都感到,20 世纪 90 年代末的 PARC 的发展方向跟我们越来越背道而驰了。

最后,我们集体辞职了。我们也想过要一起去做一些事情,因为我们喜欢一

起工作。不过,失去了 PARC 这个平台,我们终究还是各奔东西了。就我个人而言,我一直很清楚,如果离开 PARC 的话,我还是想回到学术界。所以,在 PARC 工作期间,我也非常努力地维系回归学术界的桥梁。我之所以知道兰卡斯特大学,是因为他们社会学系的一些人跟“计算机支持的协同工作”(Computer Supported Cooperative Work, CSCW)会议联系密切。他们做了一些非常出色的工作,比如约翰·休斯(J. Hughes)等人对空中交通管制员的技术变迁的研究。

兰卡斯特大学还有一个非常强的 STS 项目,有约翰·劳(J. Law)与许多其他鼓舞人心的 STS 研究者,以及性别和女性研究方面的杰出同事。因此,当兰卡斯特大学发布了两个职位的招聘公告时,我就想我肯定会申请。这就是我最终来到兰卡斯特大学社会学系的原因。这是一个非常棒的系,而且那些真正对跨学科思考感兴趣的人遍布整个大学。我立即感到宾至如归。与此同时,我周围的每个人身上都学到了很多。所以我在那里度过了接下来的 20 年!

二、人类行动和机器智能

胡万亨:现在我们来谈谈您的具体研究吧!先聊聊您的经典作品《人机重构》(*Human-Machine Reconfigurations*)——最早的版本标题叫《规划和情境化行动》(*Plans and Situated Actions*)。书中对用“规划模型”(planning model)去理解人类行动,由此开发智能机器的想法提出了批评。您能简单解释一下书中的核心论点吗?

萨奇曼:规划模型的基本论点跟有关交互性(interactivity)的看法密切相关。当时在人工智能领域占主导地位的就是规划模型,它对人类智能和相互可理解性的概念有着非常强烈的认知主义色彩。它的基本想法是,所有真正要紧的事情都发生在我们的头脑中,在我们的大脑或心灵中,人的智能行动是目标导向的,是按照制定好的规划有效执行的。行动的要点就在这些规划之中,随后被执行出来。这种认知主义的观点能很好地映射到计算机技术的能力范围内,因为规划可以对应到程序里,计算机程序是对行动要求的具体化。人们也认识到,在规划执行过程中可能会出现障碍,可能会遇到必须适应的环境条件,但这些都是

行动者执行计划时需要解决的问题。具体到沟通行为上,它意味着我们相互理解的方式是通过认识彼此的规划。

我当时参与研发的系统就是基于规划模型的。回到复印机的例子,当我们进行复印时,我们有一个规划,有一个目标,有一种我们想要执行的具体规范。我们制定了实现这一目标的规划。因此,我们就想,专家系统需要能够识别用户的规划,然后为他们提供必要的指令,以使用这台机器执行该规划。我的分析就是基于人们在实际理解操作界面并试图使其变得易于理解时遇到的困难而展开的。

我当时对常人方法学和互动论分析很感兴趣,所以我录制了一些人们尝试使用复印机的视频来研究这个问题——最开始是比较原始的复印机,后来的复印机加入了专家系统。让我不可思议的是,我坐下来观看这些录像带,而且如果我足够仔细地观察,我就能弄清楚这些人想要做什么。然后,我发现他们实际上并没有按照设计者的意图来理解机器,机器的很多“举动”用户也并不理解。所以,我可以看到所有这些问题。我开始思考,我为什么能看到所有这些麻烦,但机器似乎完全没有注意到,专家系统似乎也完全没有,或者错误地识别了用户的行为。

我开始更仔细地观察这些互动,我在书中用一个四列表格列出了这些观察。我录制的视频里,人们要么在谈论他们正在做的事情,要么是两个人试图一起做点什么,所以在互相交谈。我可以听到他们彼此交谈的所有事情,看到他们围绕机器所做的所有事情,我将它们转录到表格的第一列中。然后,我可以看到其中那一小部分机器可以在现实中做到的事情,我将其转录到第二列中。当然,因为它只知道是什么(指令)会改变它的状态——即便是我们现在使用的更加复杂的计算界面,也还是如此。在我们人的世界中,唯一能跟机器交界的,就是那些能够实际影响机器状态的方面,比方说敲击键盘、点击触摸板或是其他之类的操作。所以,我开始认识到,这个系统只有一个小得令人难以置信的小孔,这是它能够了解正在发生的事情的唯一通道。

接着,我在表格的第三列指出了用户哪些操作得到了机器系统的响应。最后,在第四列,我列出了这些操作背后的设计原理,这些原理都是基于把用户的操作理解为对他们头脑中规划的执行,进而给系统设计相应的响应。如此,我能

够看到用户和系统之间的所有不一致(misalignment)。我由此得出的基本论点聚焦在人与机器之间存在的极大的不对称性(asymmetry)上:当我们参与行动、理解彼此以及世界时,我们有很多可以使用的资源,而计算机则少之又少。

基于对这种不对称性的观察,我对于规划与情境化行动的论点在很大程度上得益于哈罗德·加芬克尔(H. Garfinkel)的常人方法学中一个非常基本的观点:任何关于行动的规定——无论是说明、法律、规则还是规范——都必然以找到规定与当前具体情况之间关系的能力为前提。说明书上的指示预设了用户的这种能力。这些说明没有,也永远不可能,完备地具体指明如何找到这种关系。因为,不论是遵照什么样的指令,用户面对的都是一个不可简化的开放世界。自此之后,我所有的研究都贯穿着这个前提,即计算系统在某些重要方面需要在封闭的世界里运行,而计算机与用户交互的界面,本质上就是对开放世界、对我们在世界上的行为和活动的转译,使之能被计算机读取。

胡万亨:说到这本书本身,我很喜欢您在书中开篇举的例子,您谈到欧洲人和特鲁克人(the Trukese)如何以不同的方式在海上航行。

萨奇曼:是的。这是人类学里的一个经典,托马斯·格拉德温(T. Gladwin)和其他人描述了南太平洋的航海家如何在没有地图和仪器的情况下,不可思议地从一个岛屿准确到达另一个岛屿。

简单来说,这个故事说的是他们的参照点是周围的所有事物,如海浪、星星、风、太阳等等。这个例子的要点并不是说世界上有两种人,一种人总是按规划行动,另一种人总是看情况行动;相反,这个例子比喻的是我们(作为西欧人)思考目标导向之行动的方式,以及思考它的不同方式。在一些重要的方面,我想,我们所有的行动都涉及调动不同类型的资源,根据具体的情况去寻找资源。这始终是一个将我们可用的工具应用于所处的实际情况的问题。

胡万亨:其他一些学者也对人工智能提出了一些著名的批评,如您所知,比如德雷福斯、哈里·柯林斯(H. Collins)和戴安娜·福赛斯(D. Forsythe)。对于他们的看法,您如何定位自己的观点?

萨奇曼:我当时也读到了他们的作品,深受他们的影响,从中汲取了很多灵

感,尤其是从德雷福斯那里。我想,我们可以先从他的《计算机不能做什么》(Dreyfus, 1978)及其续集《计算机仍然不能做什么》(Dreyfus, 1992)说起。他带来的现象学视角,真正阐明了认知科学和人工智能是如何消除具身性和情境性的。我认为,他以一种非常有力的方式指出了计算的局限性,他的论点到今天仍然站得住。随后,他对联结主义(connectionism)的出现给他的论点带来的影响和挑战感到相当不安。我认为,他对此多虑了,他的论点仍然是成立的。

然后是柯林斯和福赛斯。我们每个人的出发点都略有不同,但殊途同归。柯林斯是一位知识社会学家,带来了知识社会学的视角(Collins, 1990)。在研究人工智能之前,他的《改变秩序》(Collins, 1985)一书是关于科学实验的复现的,但同样也是关于具身性和情境性的。那本书对我来说真的很有启发。福赛斯出身于更传统的文化人类学领域,她对当时的人工智能有很深的研究(Forsythe, 2001)——那时候还以知识工程(knowledge engineering)和专家系统为主。因此,她把对知识构成的人类学理解带了进来。这同样可以追溯到德雷福斯那里,并且与我提出的论点也非常有共鸣。

所以,我们提到的这些书,无论是哪一本,我读起来都觉得非常有共鸣,有很多共通之处。我们对人工智能的批评只是在取向上略有不同,主要是我们所借鉴的知识传统不一样。让我受益匪浅的另一本书是艾莉森·亚当(A. Adam)的《人工认知:性别与思维机器》(Adam, 1998),这本书非常精彩。艾莉森是一名计算机科学家、软件工程师,具有非常强烈的女性主义意识——她阅读了大量女性主义学术著作。

胡万亨:德雷福斯在他的《计算机仍然不能做什么》一书中谈到了机器学习。柯林斯于2019年出版了他第二本关于人工智能的书《人工虚拟智能》(Collins, 2018),主要讨论的就是机器学习。遗憾的是,福赛斯似乎并没有对此说太多,她去世得太早了。我想知道:您如何评价机器学习?如果您今天要再写一本关于人工智能的书,您之前的论点如何适用或不适用于机器学习?我想听听您对这个领域的现状有何看法。

萨奇曼:这是一个非常有趣的问题!因为从专家系统、知识工程和符号主义人工智能时代,到现在这种数据密集型、高算力模式的人工智能的爆炸式发展,

很多方面的情况都已经发生了翻天覆地的变化。但我的看法还是一以贯之的，还是可以追溯到封闭世界和开放世界的问题。

如果我们看看大语言模型，看看它的自我参照性 (self-referentiality)，不同之处也就在于，现在这些机器学习系统的能力不是依赖基于规则的操作，而是可用数据的规模以及能够从中识别相关性的计算统计，然后进行参数化和迭代以产生可理解的输出。这就是我对它的核心特征的描述。

我们仍然可以将这些模型视为一种封闭的世界系统、一种自我参照的系统。现在这些模型所谓的“脆性” (brittleness)，就是在它们遇到自我参照性的局限时显现出来的。事实上，大语言模型基本上是在以数字方式捕获的文本和图像上运行的，而完全不是在它们所指涉的真实世界中运行的。

因此，正如我此前研究的专家系统只能依照规划模型读取指令从而在不同的状态间切换而实现运作一样，当前的系统只能根据它们运行的数据集和数据模型中存在的相关性进行响应。它们其实什么都不知道，对吧？我的意思是，它们对所有这些数据所指涉的实质内容其实一无所知。有趣的是，现在已经有一些人在呼吁回归到混合模型上来，比如加里·马库斯 (G. Marcus) 这样的关键人物 (Marcus, 2018)。数据驱动的人工智能之路正在遇到瓶颈，在寻找下一步对策时，该领域的人们现在正回头转向符号人工智能——“好吧，现在我们需要回到逻辑”——不论如何，它仍然还是会处在一个逻辑关系或统计相关性的封闭世界中。

我觉得观察这一切、追踪这些巨大的变化真的非常有趣——这些变化同时也是我们如何生活和穿梭在开放世界与封闭世界中的方式的变迁——当然，我认为有些东西也没怎么变，有一些非常根本的方面并没有改变。还有一点没有变的是，人工智能界还是不愿意承认这一点，他们对类人智能的渴望和决心如此强烈。不知何故，在他们看来，如果承认这些局限性，就会贬损这项奇妙的技术——我认为它真的非常了不起。这些局限性在我看来，只不过是计算系统和作为有机体的动物之间永远都会存在的差异而已——我说的“动物”不仅仅是指人类，还包括人类之外的动物，甚至其他更多非人类世界的存在。

三、跨越学科边界,连接学术与社会

胡万亨:我们讨论了您早期对符号主义人工智能的批评,也探讨了今天的联结主义人工智能。当时的工程师如何回应您的批评呢?这对当今的人工智能研究者有怎样的启发?

萨奇曼:我认为“因人而异”总是非常重要的——人们会问:施乐对这项工作有何看法,或者 PARC 有何看法?但实际上,我们说的都是特定的一些人。你会遇到一些非常关键的搞技术的人——现在回想起来,我尤其感谢菲尔·阿格雷(P. Agre),他在我的思考过程中非常重要。我们都还在念博士的时候就认识了,我从他那里得到了很多灵感,也学到了很多東西,他那时正在麻省理工学院攻读人工智能方面的博士学位,而我在加州大学伯克利分校攻读人类学博士学位。在像麻省理工学院或 PARC 这样的地方,有计算机科学家,有非常投入的工程师,但他们也会有疑问,对吧?他们也会觉得在工作中找到问题和改进的机会,以及思考自己工作的新方式更有趣、更令人兴奋。

我在 PARC 的任务主要就是找到这些人。一路走来,我发现他们中的许多人确实对重新界定他们正在解决的问题或以不同的方式思考解决方案很感兴趣。我得说,那些人真的就是我的盟友。即便是理查德·菲克斯(R. Fikes)也对我的批评非常感兴趣,而他是开发我所研究的专家系统的、秉持人工智能的规划模型的人,他也是一个思想非常开放的人。

所以我认为,与其将人工智能社区视为一个同质的整体,不如与那些有兴趣批判性地反思自己实践的技术从业者建立联系。而且,他们也可以给你讲这些工作的进展,以便你能够充分了解田野的情况并提高自己的技术素养。刚去 PARC 的时候,我甚至觉得自己得成为一名会 LISP 语言的程序员,如此才能真正了解这里发生的事情。

后来,我意识到不用成为一名 LISP 程序员,因为我自己就是一个人类学家而不是计算机科学家啊!我当时觉得——也许有点事后诸葛亮了——哪怕我自己对这些技术只掌握了一点皮毛也没有关系,因为我身边都是世界上最好的 LISP 程序员,对吧?因此,我要做的就是去合作。于是我有了像兰迪·特里格

(R. Trigg)这样的合作者,我和他一起工作了很多年。我们共同撰写了“作为手艺的人工智能”(“AI as Craftwork”)一章(Suchman & Trigg, 1993)。兰迪是一个可以和我一起思考的人,他有这个知识和能力。我觉得这种合作令人兴奋且富有成果。不得不说,司马贺(H. Simon)实际上也觉得这种合作是必要的,这也许是我成功的标志之一——他和艾尔松索·维拉(A. Vera)写了一篇题为《情境化行动:一种符号主义的方法》的论文(Vera & Simon, 1993)。他们的基本观点是,我和其他人所提出的关于情境化行动的论点可以毫无问题地整合到符号主义人工智能里并为之所用。我认为,这是一种防御性的举措,好比说我们必须得“中和”掉这种批评一样。有些人还是会坚持己见,但会试图接受任何批评并使其中性化,这样它就不会以任何方式干扰或破坏他们的进展。这是一种跟搞技术的人打交道的模式。但对我来说,更具创造性的模式是与像菲尔和兰迪这样的人合作,他们对自己的领域有着自己的反思,有兴趣以各种各样的方式重新思考它,而且是跟我们这样的人一道。

胡万亨:我很喜欢您关于“中和”批评的说法。如今,科技公司似乎越来越流行聘请一些社会科学家,让他们做用户研究、界面体验等等。您如何看待人类学家、社会学家或其他社会科学家在科技行业中的角色?您在施乐的情况在20世纪80年代很常见吗?

萨奇曼:我想说,我的例子以及施乐少数那几个像我一样隶属于PARC的研究人员的情况是相当特殊的。虽然PARC是一个由施乐资助的研究中心,但它的运作非常独立。那个时期施乐公司有大量资金,而当我2000年离开PARC时,情况完全不是这样——施乐公司几乎破产了。但早在1970年,PARC刚成立时,有大量资金用于未来投资。PARC非常独立,那里的每个人都有博士学位,简直就是个学术研究组织的延伸。我们都得在学术上有所成就,所以我们需要发表论文、参加会议,这意味着我们有很多机会保持与学术界的联系。这原本就是一个不同寻常的情况,在此后的20年间也确实改变了。当我离职时,PARC更像是一家大公司内部的典型研发组织,非常营利导向。

我在PARC的职业生涯末期,公司希望它增加专利和许可证以及其他可以让研究变现的方式。正如之前所说,我加入时,它更像是一个学术研究机构。我

认为社会科学家融入工业界——我特地使用“融入”(incorporation)这个词——在很大程度上是营销方式的一种延伸,尤其是在设计方面,是“用户体验”行业成长的一部分。这个行业为拥有社会科学学位的人提供了就业机会。

一方面,我认为这很好,因为人们确实需要工作,社会科学给工业界和设计领域带来的意识观念很重要;但另一方面,我认为对社会科学家的投资比媒体所暗示的要少得多,那些被聘用的社会科学家是用户体验、界面设计业务的一部分,与工业界的其他活动一样,这些业务的时间框架非常短。所以,研究人员鲜有机会真正长期参与田野点的活动,也很少能生产出进入学术圈的知识,无论是人类学还是社会学。

因此,可以说,这样一些工作几乎没法给社会科学家提供充分的施展机会。我认为这并非幸事,但这就是业内的普遍状况,尤其是在公司利润增长放缓的情况下。做研究的空间变得越来越小,社会科学家通常都是第一批被裁的人,而且越来越多地需要为产品计划、品牌和营销提供有用的故事。

胡万亨:我们来聊聊您最近对自主武器系统的研究吧。您能简单介绍一下您在这个方向的工作吗?您正在解决的关键问题是什么?是什么原因促使您关注这个主题的呢?

萨奇曼:之所以做这个题目,可能是因为我在20世纪60年代的美国长大,是反战运动的孩子。但即使在80年代初的施乐公司,也有一些人在PARC从事这方面的研究。其中最著名的是计算机科学实验室的塞韦罗·奥恩斯坦(S. OrNSTein),他曾从事过地面环境下半自动化武器系统(semi-automatic ground environment, SAGE)的研发,这是美国第一个计算机化导弹防御系统。80年代初,塞韦罗越来越担心所谓的“战略性防御计划”,该计划将计算机控制引入了核武器系统。

因此,我与塞韦罗以及PARC和斯坦福大学的其他同道一起参与创立了计算机专业人员社会责任联盟(Computer Professionals for Social Responsibility, CPSR)。当年有一些计算机科学家讨论了用所谓的“预警发射”来控制核弹的自动响应系统的诸多不可靠之处。多年来我一直积极参与CPSR,这些工作演变成了一系列追踪武器指挥和控制系统的自动化,特别是计算机化的项目,自主

武器系统是相关工作在当下的题中应有之义。对我来说,这是对自动化——尤其是自动化瞄准——以及越来越多的远程控制武器系统的出现的更广泛担忧的一部分,它有可能导致自动瞄准能力被一发不可收拾地用到武器系统中来。

我最近加入了机器人军备控制国际委员会(International Committee for Robot Arms Control)。这是一个由计算机科学家、物理学家和社会科学家组成的很棒的团体,其试图推动围绕自主武器系统建立具有法律约束力的条约——自主武器系统在目标识别瞄准和启动发射等关键功能上进行了完全自动化。

我们所呼吁的禁令,针对的是那些在关键功能上去除了所谓“有意义的人类控制”的武器。我深信这项运动的意义,它真的非常重要。与此同时,我对“战争背景下的有意义的人类控制”这个想法也有一些疑问。因此,我的研究往大的方面来说,涉及分析有哪些各式各样的投资、话语和技术在支持和维系“现代战争可以是一种受控制的理性事业”这一理念。我认为事实并非如此。

在这项工作中,我主要参考和对话的是保罗·爱德华兹(P. Edwards)的《封闭的世界》(Edwards, 1996),还有约瑟夫·马斯科(J. Masco)及其他 STS 学者的工作。这些研究追溯了这种控制战争的想法在二战期间和战后,以及从冷战到如今的发展。

我最近在写一篇文章,是关于军队中封闭世界思维的当代化身的。用我们 STS 的话来说,就是对一个完全仪器化、军队掌握了完备信息的世界的“想象”(imaginary)(Suchman, 2020; Suchman, 2023),我认为这仍然是一种幻想。我认为追溯这种思维的逻辑以及它为对新武器系统的投资、政策制定和国际关系提供理由,并揭示这种思维方式的谬误之处都非常重要。这种思想确实是站不住脚的。这些谬误之处为我们开辟了空间,去思考一些替代方案来阻止军事主义以及处理国际关系。

胡万亨:我们来谈谈一些更广泛的问题吧!您的研究对 STS 和 HCI 等跨学科领域做出了重要贡献,也获得了美国计算机协会和 4S 奖项的认可。在您看来,定义 STS 和 HCI 领域的核心问题分别是什么,它们有何不同?您是如何参与这些领域的对话的呢?

萨奇曼:这两个领域对我来说都非常重要。HCI 领域的大会 CHI(computer

human interaction)规模庞大,而且非常技术中心主义(techno-centric)。但是,CHI内部有越来越多的人在致力于把人机交互问题放到一些更广阔的问题中去研究,比如系统开发的条件、用户是谁、用户所处的情境以及用户是如何被设定的。

HCI领域有很多非常有趣、更富批判性的工作从STS中汲取了灵感。而且,跟以往相比,我觉得越来越多的STS学者也在寻求参加HCI领域的会议,以便更接近那些实际参与研发工作的人员。同样,从事设计的人员也在从STS中寻找资源,以不同的方式去思考他们正在做的事情,扩展他们的工作框架。在PARC期间,我主要活跃在HCI和CSCW领域。但我一直坚持把人类学和STS作为我智识上的灵感来源,所以我在这两个世界之间来回穿梭。这确实让我收获颇丰。如果没有从STS获得新的思考来源,我想我不大可能在HCI领域做出这些贡献。自从我离开PARC以来,我更多地活跃在STS学界。

所以,如今我已经不太参加CHI或者CSCW会议了。但随着信息学院的兴起,比如康奈尔大学的计算机与信息学院,现在已经有越来越多的交叉对话和研究。信息学院对于那些没法完全归入某个学科类别、处在类别边界的人来说,真的是非常理想的归宿。同时,这类学者在那些以往技术性更强的会议中的影响力也越来越大。

胡万亨:作为一名人文社科学者,要研究计算机科学和智能技术的话,您觉得最重要的技能、资质是什么,或者应该有哪些准备呢?

萨奇曼:我会说是谦虚。我的意思是,你得意识到自己在其他相关领域的知识掌握程度和局限性,并始终努力加深你的理解。同样重要的是,要学会寻找合作者。在STS领域里,有的学者技术素养非常高,对研究主题的技术细节有着深刻的理解。但我认为,很多情况下,我们中的许多人无法在短短一生中在技术领域掌握跟社会科学相关领域同等的能力。因此,我们要不断努力去加深理解,同时寻求富有成效的合作。我现在非常开心能与一些像你这样的年轻学者进行持续的对话。跟你一样,他们也身处跨学科的交叉领域,也在真正深入机器学习的技术细节中,同时充分保持和发挥STS的敏锐意识。这样就能够把STS中的一些思想概念引进来,从而强有力地、有创造性地去思考技术实践。

所以,我认为持续学习和建立联系可能是最重要的事情,也是你能拥有的最重要的技能。

胡万亨:的确,这里的关键词是“协作”。您对年轻学者有什么特别的建议吗?

萨奇曼:寻找志同道合的学术同辈,不是吗? 这太重要了。如今的就业市场真的非常非常难。从我早先对自己在工业界工作经历的评论,你可以看到,我其实还是在学术界感觉更亲切。但我认为在工业界工作是一种非常有价值的经历,最起码也是宝贵的民族志。所以我想说,如果你确实很在意维持在学术方面的发展,你可以想办法在业界工作的同时跟学术界保持联络。

去参加 STS 会议并且报告论文,这就是我保持学术联系的方式,我就是确保自己能够出席并参与我想保持联系的学术活动。所以,我觉得,还是有很多办法可以找到工作、找到雇主,并保持你所关心的学术联系的。这确实有点挑战性,但也是可行的。

参考文献

- Adam, A. 1998, *Artificial Knowing: Gender and the Thinking Machine*, London: Routledge.
- Collins, H. 1985, *Changing Order: Replication and Induction in Scientific Practice*, London: Sage Publications.
- Collins, H. 1990, *Artificial Experts: Social Knowledge and Intelligent Machines*, Cambridge, MA.: MIT press.
- Collins, H. 2018, *Artificial Intelligence: Against Humanity's Surrender to Computers*, Cambridge: Polity.
- Dreyfus, H. 1978, *What Computers Can't Do: The Limits of Artificial Intelligence*, New York: HarperCollins.
- Dreyfus, H. 1992, *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA.: MIT Press.
- Edwards, P. 1996, *The Closed World: Computers and the Politics of Discourse in Cold War America*, Cambridge, MA.: MIT Press.
- Forsythe, D. 2001, *Studying Those Who Study Us: An Anthropologist in the World of Artificial*

- Intelligence*, Stanford: Stanford University Press.
- Marcus, G. 2018, *Deep Learning: A Critical Appraisal*, <https://doi.org/10.48550/arXiv.1801.00631>.
- Nader, L. 1974, "Up the Anthropologist: Perspectives Gained from Studying up." in D. Hymes ed., *Reinventing Anthropology*, New York: Vintage Books.
- Suchman, L. & R. Trigg 1993, "Artificial Intelligence as Craftwork." in S. Chaiklin & J. Lave eds., *Understanding Practice*, New York: Cambridge University Press.
- Suchman, L. 1987, *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge: Cambridge University Press.
- Suchman, L. 2006, *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge: Cambridge University Press.
- Suchman, L. 2013, "Consuming Anthropology." in A. Barry & G. Born eds., *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*, London: Routledge.
- Suchman, L. 2020, "Algorithmic Warfare and the Reinvention of Accuracy." *Critical Studies on Security* 8(2).
- Suchman, L. 2023, "Imaginariness of Omniscience: Automating Intelligence in the US Department of Defense." *Social Studies of Science* 53(5).
- Traweek, S. 1988, *Beamtimes and Lifetimes: The World of High Energy Physics*, Cambridge: Harvard University Press.
- Vera, A. & H. Simon 1993, "Situated Action: A Symbolic Interpretation." *Cognitive Science* 17(1).