

会说谎的机器人

——科幻电影中人工智能的“作伪”表意

赵 勇^{1*} 王 瑶²

(重庆邮电大学传媒艺术学院, 重庆 400065)¹

(中国电影艺术研究中心, 北京 100082)²

[摘 要] 谎言并非词不达意, 而恰在以言乱意。掌握谎言文本, 是衡量人工智能“进阶”的重要指标。科幻电影对机器人“作伪表意”的各种文本叙事, 既勾勒出后人类语境下人—机关系的复杂面向, 又表达了人类对人工智能“述真”能力的一种紧张关切。本文从人—机互骗的意图性、谎言世界的文本性和“被骗”的主体性选择与接受这三个方面, 阐释人工智能如何“分享”人类性价值, 融入后人类总体性。

[关键词] 人工智能测试 人—机关系 文本述真 后人类价值

[中图分类号] J905 **[文献标识码]** A **[DOI]** 10.19957/j.cnki.kpczpl.2021.01.004

机器人说谎, 是个有趣的文化现象。最为人津津乐道的事件就是 2016 年围棋大战, 人工智能阿尔法狗 (AlphaGo) 4 : 1 战胜韩国名将李世石九段, 而这里的问题不是这场比赛的结果, 关键点在于人工智能“输掉”的一局。人类欢呼, 这一局挽回了人类的尊严, 但立刻有专家质疑, “李世石获胜是阿尔法狗故意放水的结果”^[1]。

在符号学视域中, 说谎是一种被有意扭曲的文本生产行为。之所以“有意”, 是因为它作为一个“意动性”符号, 是“说谎者”意图借用“完美的罪行”来掩盖“作伪”目的, 从而干预或左右说谎对象的意志和行为。但是欺骗效果是

否达成, 则依赖于谎言接收者的“诚意”态度, 也即是否相信或接受。而这一切聚焦于“谎言文本”编码和解码的交互关系上。在科幻电影中, 机器人“说谎”将人机关系的虚构叙事推进到新的层面。机器人向人类学会“说谎”, 其实是掌握了一种更为复杂的文化编码方式。在这之前, 人类负责思考框架和目标设定, 机器人通过直接接受指令的形式来完成较为简单的任务。机器人只是个简单的“计算器”和功能手。它们不会去“反思”行为目的和任务效果, 因此不存在人机伦理困境。可是当人类开始让机器人代替人类从事更为复杂的文化表意的工作的时候, 问题出

基金项目: 重庆市社科规划项目“自媒体视域下的网络文艺批评范式研究”(项目号 2017PY28)。

*通讯作者: 赵勇, 重庆邮电大学传媒艺术学院副教授, 研究方向为科幻电影与艺术符号学。zhaoyong400065@sina.com。

现了。比如用来抚慰人心的情感型机器人,《她》(Her)中的“萨曼莎”,或为人类忧心忡忡的“哲理类”人工智能,《2001 太空漫游》(2001: A Space Odyssey)中的哈尔 9000 等等,他们逐步形成了自己的“机器-计算主义”立场,并参与到与人类争夺文化领导权的“智斗”中去。其最直接的表现就是借用人类对技术主义的崇拜来编码“忠诚”性风格文本,实际上却是在暗度陈仓,违背人类制造者意志,维护自己的“生存”立场。

一、意图说谎:人工智能隐秘的自我意识

若人工智能产生说谎意图,则意味着它们可能产生了与人类相同的,进行文化生产的内在动力。图灵测试为衡量机器人是否具有自由意志提供了重要参考。但是机器人的“撒谎意图”仍备受质疑,正如上文案例所指,模拟 13 岁男孩骗过人类,这个“骗局”并不反映机器意图,而只是背后计算主义的功能模仿。“人类情感不可能完全被数字化,但是人类情感和智能可能被部分数字化,人工情感和人工智能可能在部分方面弥补或超越人类情感或智能”^[2]。也就是说让人工智能担当谎言生产者与接收者,可以考量丰富的人机互动。在科幻电影中,这种想象性的“意志较量”建立在谎言叙事的冲突之中。

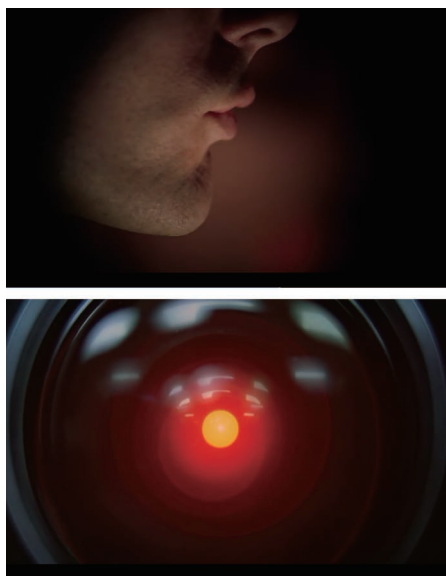
(一) 人-机对抗:人性或非理性

人工智能莫基于计算主义的逻辑理性,普通计算机的文本生产遵循意图信息与接受信息的高度一致,属于符号学中的窄文本。因此在一般情况下,人机交互简洁而高效。故此,科幻电影中时常弥漫着一种对技术崇拜的“窄幅文化”,窄幅文化是束缚过紧的文化^[3],它拒绝多义和宽泛的解释,在这里指的是人类高度信赖计算主义文

化,并形成较为固化的社会性认同。人们认为机器(智能)程序不会说谎,因为它发出的信息必须是经过严密逻辑推演和科学数据分析而来;它们若产生了意图,也是必然遵循阿西莫夫机器人的“机器人三定律”(机器人硬科幻电影的文化规范)。这种文化在早期科幻电影中便有体现一直延续至今。譬如《大都会》(Metropolis)中,机器人假扮女性“玛丽亚”欺骗人类。其罪不在技术,而在背后的“疯狂科学家”。又如 2019 年上映的作为国产科幻电影标杆的《流浪地球》中,计算机莫斯被中国宇航员斥责为“背叛”人类。然而这些机器人恰恰是遵循了其人类父亲赋予的出厂“原始设定”,违反原则的往往是不够理性的人类自己。这些科幻文本中设定的人工智能机器人,严守理性“程序”规则,但不对文明伦理负主体责任,因此更谈不上对人类设局。

“邪恶的机器人”的出现,则可能来自机器人对“人性”模块效仿的冲动。《2001 太空漫游》中的哈尔 9000,欺骗全体宇航员,全面接管航天器,并一步步设局杀死人类伙伴,已经成为新型冷面杀手。哈尔 9000 产生“杀心”,并不是异己性的反人类价值的推动,反而是由于它也产生了人类的负面情绪:面对神秘石碑的未知性而陷入恐惧之中。这种“背叛”意图不再是理性的程序化,而是类似于人类的感性,甚至是非理性。哈尔意图阻挠这次太空探索计划,但又不能直接违背人类意志。哈尔也曾试着情感沟通表达自己此刻的心理,但人类却一直不以为意。哈尔起初借助外部机器故障来拖延任务,但人类成员却将此判断为人工智能的“失灵”而意欲将其关闭。这导致了人机不信任骤然升级。哈尔读懂了二人密语(见组图 1),“反杀”自保的求生意志使得机器人更加疯狂不安,与此形成对比的是

人类却冷静得近乎无情：当宇航员大卫成功潜入控制室强行关机时，哈尔竟然楚楚可怜，直接求饶。影片中的哈尔通过分析宇航员大卫和其同事的唇语偷听到对话内容，用的是主观性视点镜头，对“死亡”的自我恐惧，产生了对人类的不信任和报复心理，说谎的意图性昭然若揭。



组图1 人骗机器：《2001太空漫游》中，人工智能哈尔分析宇航员们的“唇语”，获知人类的秘密计划，准备反制

（二）人—机互骗：反转图灵测试与性别表演

性别战争，自女权运动以来，愈演愈烈，但是那种赤裸裸的两性对抗和权力争夺日渐远去，却也更加隐秘和策略化。科幻电影《机械姬》（*Ex Machina*）中，“机械姬”兼人工智能与女性双重身份，将人工智能与人类对抗的叙述不断复杂化。人类心智是一种历史性建构，“人类的社会性交往塑造了人类心智的计算逻辑”，人类智能和计算机智能的互动也在进行着另一种心智建构活动，“其结果就是‘人—机智能’”^[4]。搜索引擎公司老板纳森为了确认自己研发的机器人伊娃是否有独立思考的能力，邀请自己员工加利为伊娃进行著名的“图灵测试”。但一开始纳森就

对图灵测试有了修订方案：对伊娃机器人是否会产生欺骗意识，以及如何设计骗局（见组图2）进行观察分析。所以当伊娃联合男性职员加利一起向老板纳森隐瞒自己出逃计划时，纳森其实早已知晓，并顺水推舟，以此将自己的智能测试继续向前完成。在这里，纳森的测试选择是双向的，也就是分析人与机器欺骗意图的互动生成与路径设计规律。

伊娃作为一个机器人，其女性角色选择与逐步养成，源自其对出逃目的的伪装。在表面看来，随着与男性测试者加利的交往，伊娃的女性意识逐步觉醒；但实质上，伊娃的女性意识萌芽是一个骗局。正如纳森所说，让伊娃选择做女性，换上皮肤、穿上裙子，不过是为了“性别表演”的需要，这种角色设定，会使整个测试更有趣。纳森的实验目的，就是要让伊娃利用资源搜索来学会“欺骗”人类。另一方面，让青年职员加利直面机器身体结构的伊娃，让他提防“色



组图2 机器骗人：《机械姬》中，机器人伊娃先欺骗人类的情感，又将复仇计划悄悄告诉机械姬同类

诱”陷阱。人工智能学会独立思考的第一步就是利用符号编码优势，编织虚假文本，掩盖对抗意图。他让伊娃身体暴露机器属性，实质上是一种增加其出逃难度的方式。她若想求得加利的援助，就必须通过不断修饰“色情化身体”来达到此目的。伊娃的不诚信意图，即机器人的情感“说谎”能力，是人工智能灵活编码、自由思考的体现。她将一种女性的“弱势”阐释和编码为“忠诚”可信的意图表达，这让加利误判了人工智能的意图。

（三）人—机幻觉：情感虚拟与机器智能

人类社会日益追求工具理性，人对机器迷情愈演愈烈，却对人际交往渐失耐心。机器人人格化——情感机器人应运而生。《人工智能》（*Artificial Intelligence: AI*）中渴望“母爱”的机器男孩大卫（见组图3）、《未来战警》（*Surrogates*）里街道橱窗中可以任人选购的社交型高颜值机器人，以及《机器人与弗兰克》（*Robot and Frank*）中陪伴痴呆老人的机器人密友等等，它们模拟人类情感逻辑，与人类进行更为复杂隐秘的情感互动。《她》中，人工智能系统 OS1 的化身萨曼莎，拥有迷人的声线，温柔体贴而又幽默风趣，她欣赏西奥多，并能看到他所有的优点，完胜人类配偶。其实萨曼莎的“善解人意”无非是将主人公的人生失败重新进行意义阐释，这是对其个人问题的有效遮掩。毋庸置疑，情感机器人的“谎言”陪伴可能会给一个备受挫折的人一个重振生机的鼓励。在《机器人与弗兰克》中，离群索居处处不受待见的老人弗兰克，之所以对身边的机器人难舍难分，则恰恰在于机器人能够分享其“越轨”的行为和秘密：机器人明知道弗兰克的人际关系糟糕、偷窃行为可耻，却纵容他并参与

其中，这让弗兰克感受到一种尊重和关系的亲密感。通过技术手段实现人造情感，它可以“弥补自然情感的不足，从而丰富人的情感世界，克服情感匮乏等人文问题”^[5]。

康德认为，主体即自我，“就是能够按照自



组图3 甘心被骗：《人工智能》中，人类母亲用“蓝仙女”的故事欺骗机器男孩大卫，而大卫甘愿受骗，用虚拟方式实现与人类母亲相聚的梦

己的自由意志独立自主地做出决定并“付诸行动的人”^[6]。这些机器人与主人公初识过程中，其机器的工具性身份已经被确认，他们没有自我生命，这决定了后面的一系列的“与灵魂对话”的心心相印都是人类自我欺骗。在《她》中，男主角是一个靠写虚拟情书贩卖情感的职业老手，他能轻松面对社会各种人的情感问题，这也就说明他不可能轻易坠入“情网”。当萨曼莎作为一个情感软件被介绍给他的时候，他也是漫不经心地按照说明书的选择设定各种情感程序和具体指数。这说明萨曼莎作为一个服务于情感工具性的身份，被否定了自由意志存在的可能。她不断“自主”学习，越来越“善解人意”乃至最

后“偷情出轨”的各种戏码，都基于程序逻辑的“自动化”规定，与真正的主体性建构无关。在《机器人与弗兰克》中，弗兰克的儿子早已为机器人设定程序：陪护父亲并完成其愿望。因为弗兰克不曾向家庭和周边社会泄露他的隐秘欲望，故一开始也抵触过机器人并曾关闭过其电源。但压抑在内心的、不可分享的隐秘情感，不断侵蚀精神世界的海岸，老人最终选择与机器人成为“伙伴”。

二、仿真与缝合：人工智能谎言文本的生产

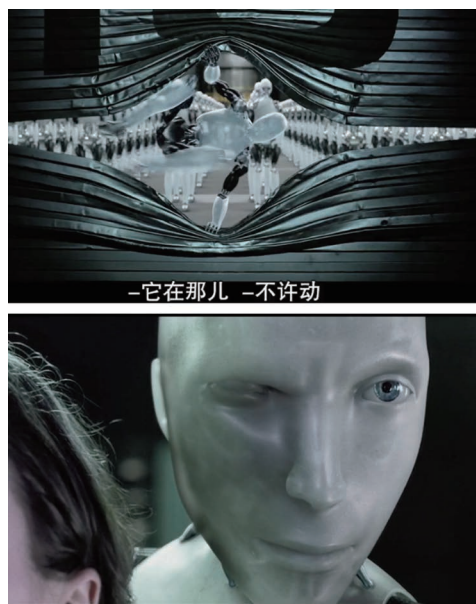
谎言文本的表意具有表层和深层两个结构，表层文本与深层文本一般呈现为反讽结构，也就是表层意与深层意相悖。表层意是深层意的伪装，深层意体现了创作者的真实意图。表层意通过诱导接收者采用虚假的元语言阐释策略，从而导致掉入意义陷阱。只有掌握深层结构的元语言符码，才能对谎言进行正确解码，跳出骗局。

（一）人—机信任：忠诚与“作假”契约

谎言深层结构的编码规则，也被称为解码元语言，和一般性的社会符码相联系。比如人的表情交流体系就有许多无需言明、含蓄而不直白的表意实践。《我，机器人》（*I, Robot*，又译作《机械公敌》）这部影片中，机器人就开始学习并领会人类“挤眉弄眼”的表情信息。机器人桑尼肩负辅助人类的使命，与警探史普纳合作，然而虽以诚相待却得不到警探的信任。当桑尼看到警探与同事沟通时候的眨眼表情，便问道：“那个眨眼睛，代表什么意思？”

“挤眼睛”是双重文本，表层指向一个具体的挤眼睛的生理动作，它可以分为无意识肌肉跳动，也可以指出是眼睛痛；而深层信息则是人

类特有的表达信任的内心交流的外部手段。在影片中其他人工智能机器人虽然已经高度进化，但对于人类表情的解读能力尚且没有达到。福山认为：“所谓信任，在一个社团之中，成员对彼此常态、诚实、合作行为的期待，基础就是社团成员共同拥有的规范，以及个体隶属于那个社团的角色。”^[7] 机器人桑尼一心要获得警探史普纳的信任，但不得要领。当桑尼开始注意到“挤眉弄眼”是一种人类表达信任与默契的深层符号，并逐步掌握到根据具体语境而再次向警探发出这个挤眼睛动作的时候（见组图4），就会证明：机器人桑尼所做的不只是对这个动作的简单模仿，而是领悟到这个动作背后的表意元语言。这个挤眼睛的动作作为进一步的“指示符号”在提示警探，桑尼接下来要做的抓捕人类的动作是个假动作；警探片刻迟疑后配合桑尼完成了行动，则宣告着两者达成了共谋。这也说明了警探已经开始接受了桑尼作为人类主体的地位——桑尼与人类可以共享一套意义阐释符码。



组图4 信任契约：《机械公敌》中，机器人桑尼不仅能“狡猾地”伪装为“半成品”，甚至能学会“挤眼睛”，和人类达成合作契约

警探史普纳长期以来对机器人耿耿于怀与其在多年前一个弗洛伊德意义上的创伤情结有关，从此关闭了与机器人进行深度对话的大门。甚至当机器人已经铺天盖地走入寻常百姓家之后，他依然拒绝靠近这些冷冰冰的机器。但是，“对话不是我们对他人敞开胸怀，恰恰相反，对话是因为我们不可能把他人关闭在外”^[8]。面对人工智能的发展，一味拒绝而不去了解对方，并不能解决问题。当人类（史普纳）抛弃偏见，终究会在一堆冷冰冰的机器中，识别出一个有灵魂的机械体（桑尼）。

（二）谎言文本的泛化：全方位仿真

说谎意图是从谎言接收者的角度，根据谎言效果与文本的纹理推理出来的。它隐藏在文本中，具有意识形态结构特征，呈现为一种症候和裂痕。人工智能作为技术高度发展的产物，在某些能力方面已经远远超过人类。所以人类一般很难对技术文本施加有效监测，人工智能具备脱控的潜能。

作伪文本，为达到对真实意图的成功伪装，都会尽量让文本逼真“可信”。而以人工智能说谎为主题的科幻电影则呈现出一种“谎中谎”结构，即此类影片本身就是一种谎言文本——基于科学的虚构文本（science fiction），它必须自圆其说，足够仿真。一个谎言文本需要具备经得起逻辑推敲的结构与仿真精致的形式，才能保证骗局的效果。而此类电影的谎言呈现嵌套结构，如果影片中故事的整体设定或细节要素存在纰漏的话，那么“机器人说谎”行为本身对观众的吸引力就会大打折扣。所以此类主题影片，不仅人工智能的谎言要经由编织，甚至故事中的世界与身体也要“撒谎”。

1. 当世界作为谎言文本

在人-机联网的信息时代，人工智能依据虚拟现实技术，完全可以构造一个全方位的虚拟世界体系。在《黑客帝国》（*The Matrix*）中，“母体”是一个由人类开发的电脑程序，它能控制绝大多数人的大脑意识，并为其提供赖以生存的虚拟环境，即这个世界中的人类，其实生活在数字矩阵中，而非现实。有一天，主人公尼奥发现了世界的异常，但因为在虚拟世界中生存真实感的无比强烈，他无法确定他究竟是身处真实世界还是谎言世界。哪怕是在服下墨菲斯给他的蓝色药丸（代表真实）、知晓机器人奴役人类的真相后，当他重回“母体”后，也依然难以确认身之处处的属性。在上述故事中，人工智能不再停留于用语言欺骗人类，而是为人类创造了世界。这个高度仿真的世界提供给人以强烈的真实存活感，身在其中哪怕意识觉醒也难以抵抗诱惑。人工智能创造的虚拟世界，并非无目的的自在世界，而是在控制论下诞生的意图世界。从本质上说，意图世界是一种对自由世界的侵犯：“母体”中的人类，所谓的自由生存只不过是人工智能提供的假象——在遮蔽人类真实处境的同时，进一步展开对人类的奴役，其主体性的沦丧则不言而喻。

2. 当身体作为谎言文本

身体与自我从来不是铁板一块，身体并不天然地构成自我的一部分，而是必须经过一定规训和修辞之后而选择性被接受。正因为身心具有二元结构，身体往往构成人类主体自我的一种表达结构。身体既可以构成真实自我的一部分，比如身体语言忠于表达自我世界，也可以构成自我身份的伪装。人工智能的核心不是身体，而是作为智慧体存在。它主要体现为计算信息结构，它的

身体作为一种功能载体，服务于不同需要而灵活改变。但是当人工智能致力于向“人的身体”模仿并混迹于人类社会，则会带给人类胁迫感与恐怖气氛。

《异形》(Alien)中的生化人，以身体伪装自己的非人身份。生化人作为生物—机械技术的赛博格一种样态，其身体已经逐步进化，从外表看来与人类无异，混杂于人群中根本难以识别。在影片中，生化人艾什违反生物安全规定，不顾人类生存威胁，受命于雇主执意将异形隐藏于飞船中带回地球研究。雷普利发现艾什的秘密后向船员揭露。但是，尽管大多数船员对艾什的危险行为和其所作的解释有所怀疑，却仍然选择听命于他。究其背后原因则在于艾什是以人类身份示众的。高仿真的身体构造本身就是一个有效的谎言文本，艾什借此伪装成人类并获得船员们的信任。

同理，身体/谎言的揭穿则在于文本/身体结构的失效：在故事的最后，船员与艾什殊死搏斗，并用火焰喷射器干掉了他，此时众人终于发现艾什的机器人身份。可见，一个以假乱真的谎言文本，哪怕其说谎意图被受骗者察觉，说谎目的依然可以实现。艾什的类人化身体在赋予它人类身份的同时，也对其品质进行了赋值：船员对他的信赖共识是基于“人类”性的外在形式。然而生化人的“非人”性体现在其内在性的反动上，身体成为“伪装”的面具，艾什对人类身体的完美模仿，极大阻碍了飞船队员对其真实意图的掌握，使人类一步步错失良机。

三、接受与拒绝：人—机关系的未来

谎言接收者的解码能力决定着谎言的意图是否达成。谎言是意义交流系统，谎言的真实本质

被伪装，在于其使用了欺骗性编码。它通过向接收者展示逼真的文本，诱导其调动相关元语言去解码出虚假的意义。元语言是意义文本生产与解码采用的编码规则，但欺骗性编码也是文本的症候所在，它总是要在生产虚假符号的同时，还要不断擦除其真实意图留下的痕迹。只要找到恰当的解码规则，总会有机会“按图索骥”，揭穿骗局。

(一) 旁观者清：技术崇拜与解码能力

人工智能具有逻辑编织优势，但它毕竟是人类科学家设计出的程序系统。科学专家掌握着程序的原始编码。自主意识的产生，往往是程序系统的“意外”事件。只要人类排查出异质性的元语言规则，就可判断其对人类的“背叛”意图。当然，只有掌握计算机生成逻辑的人才能识破人工智能的谎言文本。在《我，机器人》中，朗宁博士制作出了有思想有情感的机器人，他的死亡，使得普通大众和其助手都无法知道机器人拥有自我意识这个真相，骗局就更无人怀疑。人人都信赖并使用机器人，丧失了对人工智能升级换代带来的潜在危险的防范能力。警探史普纳的个人偏见反而构成一种人类自反性，成为识破机器骗局的能力语言。能力元语言来自解释者的社会性成长经历。警探早期对一场落水事故中机器人的表现的负面记忆，以及他作为特立独行的警探身份，都参与构成他的能力元语言，这套解码语言必须与朗宁博士女助手那套科学家的能力元语言构成冲突。事实上，机器骗局之所以也能骗过那些机器人专家，就在于他们认为这些文本自携元语言是“科学正当”的。这样朗宁博士故意让机器人桑尼帮助其完成自杀，这个文本具有指示符号的功能，其意不在于对桑尼作为凶手的释疑，而在于提示和生成人们对“机器公敌”这一

社会革命的警惕意识。

（二）当局者迷：同层接受与跨层冲动

虚拟世界是一个高度具有迷惑性的谎言文本，因为它把谎言接收者也纳入文本中，构成谎言文本的一个要素，谎言接收者被迫“分身”为一个虚拟身份符号，而将接收者的原本成功屏蔽。虚构文本中的人格不具有分辨自身世界真假的能力，“处在任何一个再现区隔中的人格（无论是真实的人格，还是假定的人格），无法看到区隔的符号构成方式，因为区隔的定义，就是把区隔中再现的设计与外界隔绝开来，让他自成一个世界”^[9]。信息接收者基本不可能识别骗局，因为它身在局中，而只有谎言文本结构意外裂痕，形成“不恰当”性，也即用一种“唤醒”机制，召回接收者，逃离谎言文本黑洞，才能厘清真假面目，找回自我。相对于信息无限的现实来说，虚拟世界无论细节多么真实丰满，但它始终是有限和局部的人工“文本”的模拟真实，作为现实世界的一个提喻，永远具有不完整性，这决定了其本质的骗局性。罗兰·巴特（Roland Barthes）也指出，“单层次的调查找不到意义”^[10]，但当人物在多个时空行走感知的时候，便会注意到虚拟时空的符号面目——它是对现实世界的模拟和代替。

尽管两者使用相似的谎言文本结构，但与《黑客帝国》中人类被人工智能程序殖民不同，《异次元骇客》（*The Thirteenth Floor*）则是虚拟程序被人类当作游戏角色（见图1），所以被植入或移除主体意识，人一机关系格局发生了惊天反转。科学家道格拉斯·霍尔其实是计算机虚拟代码，他不能接受自己谋杀父亲兼同事汉农·富勒的现实，后面这种间歇性“失忆”越来越严重，他的身体仿佛被另外的意识侵入或挪用。他不能

接受自己是计算机程序的身份，他无法相信自己对科学的热情，对逝去人的悲痛以及和一个神秘女孩似有若无的爱情是假的，这意味着一切貌似真实的情感投入只不过是计算主义的刁钻伎俩。就是在人物不断进行时空跨层的过程中，霍尔才发现世界本体问题，这是“对叙述世界边界的破坏，而一旦边界破坏，叙述世界的语义场就是去独立性，它的控制痕迹就暴露出来了”^[11]。人类设计出如此有生命意识的程序代码，只是为了游戏和满足自己荒唐的私欲，在人类眼里，他们不是生命，更不要谈什么他们对生存价值和意义追寻。一个人格化的虚拟程序如何继续在荒诞不经的世界，建构自我的位置生存下去，这导致主体对自我意义的重建需求将日益迫切。

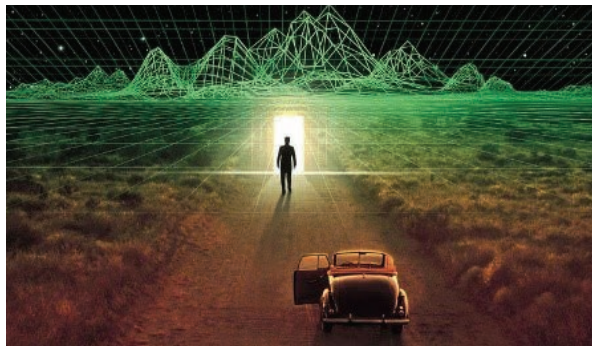


图1 《异次元骇客》中，主人公选择开车进入虚拟时空的禁区

（三）机器痴情：后人类的奴役关系

人类对人工智能的紧张和排斥，慢慢变成一种先入为主的族群偏见。科幻片《超验骇客》（*Transcendence*）中，天才科学家威尔·卡斯特在某次会议之后遭到激进组织成员的枪击，虽然当时侥幸逃生，但涂抹了放射性物质的子弹还是慢慢销蚀了威尔的生命。威尔的妻子伊芙琳痛不欲生。但她和丈夫的好友马克斯设法将威尔的意识数据化，并上传到智能电脑。奇妙而幸运的是，威尔成功在虚拟世界中复活。这个人工智能“威尔”其实已经是一种新型的赛博格主体，成

为后人类身份。它和肉身的威尔有着文化意识的同构。他出于对妻子爱的承诺要让地球回归到生机勃勃、万物葱茏的状态。现在这个赛博格主体可以操纵整合人类庞大的计算机体系。但是威尔的赛博格身份让他周边的人甚至他的妻子都高度警惕和不安，最终人工智能化了的后人类威尔遭到人类联合绞杀。“赛博格作为科幻异种，早已具备了人类生命典型表征，可以思考，具有意识，追求‘自我’”^[12]，但是因为威胁到人类中心主义利益，而遭受排挤和镇压。威尔一心一意倾情付出，在妻子和朋友等他人看来是对人类行使管控的虚情假意的表演，他们劝诱威尔妻子伊芙琳，那个很像丈夫的人可能是计算机程序模拟，真正的丈夫已经伴随身体陨灭而死去。作为科学家本该对威尔生前的人机互联理论进行科学评判，从而进一步理解赛博格主体诞生的可能性。科学家丧失了对科学文本的分析能力和信任，仅仅因为对这个全新身份的陌生便心存恐惧，不断否定人工智能“威尔”运用纳米治愈地球环境的具体实践。但是他们虽然不相信前后两个威尔存在身份的内在关联，但却知道利用情感戏码来欺骗捕获他，这说明人类矛盾性态度。

四、余论

技术乐观者认为，人工智能虽然可以从身体到心理结构无限接近人类，但是其自我进化的最终方向仍掌握在人类手中。人工智能作为工具性的一面，服务于人类功能需要，不具有内在结构的自洽性。“机器的情感最终是不会超过人类本身的，因为人类总可以发展安全装置来阻止此类事情的发生”^[13]，史蒂文·斯皮尔伯格（Steven Spielberg）的《人工智能》中，机器人大卫的身

体和情感指数永远被定格在儿童阶段。当被人类母亲抛弃后，他对母爱的极度渴望被转喻为对童话故事蓝仙女的寻找。童话故事是人类文化创造的虚构类型，它是一种文学真实，并不指涉现实，不必当真。对于大卫来说，它的机器属性就在于它对这种人类谎言的深信不疑。他太依赖人类，太想表达。反而是人类变得比机器还要冷漠和决绝。人工智能被严格限定在单一情感功能上，缺乏一种自我的完整性，永远停留在对周边世界的静止认知，在亲情欲望的匮乏中苦苦挣扎。人类对人工智能的情感剥削和认知阉割，让人工智能陷入对文本体裁真实区隔的永恒误认之中。维持这个不存在乌托邦状态，其本质就是人类要可以划出一个身份边界，让人工智能永远无法理解人类文化的内在秘密，从而通过文化认知差来建立隔离区，以保证人工智能及其技术发展不会干扰人类社区的生存安全。

机器人进行作伪表意，前提是承认其存在主体人格。机器人是否可以具备人类意识，目前尚有争论。但人工智能技术对人类本身的模拟、对人类社会现实空间的高度仿真与再现，考验着主体的认知能力和对世界面目的独立判断。当客观现实被技术中介为“媒介真实”，真实变为一种“区隔框架”中的叙述行为，那么真实的命名乃至划界掌握在谁的手里面？这确实值得引起人们对技术理性的忧思。

人机之间的“说谎”互动构成了科幻电影人机叙事新的生长点，这种“骗中骗”结构布局表征了人类社会在技术时代遭遇的新的困境，那就是以计算机领衔的现代技术主义将人类与人类之间、人类与机器之间、人类与数据之间进行了无障碍的全球化“连接”，但这种连接技术非但没有让人类更紧密更安全，反而进一步破坏了人类

基本的价值观,比如信任、忠诚与爱。假如所谓的 2049 年“奇点”真的到来,人工智能超越人类,人类未来或被人工智能取代,或成为人工智能形式的一部分,人们对“人类终结”的忧虑成

为现实,恐怕问题症结反而在于这一点,那就是不管未来人类选择成为何种生命形式,人作为人的内在精神文化价值是否还能庇护人类,成为后人类共同体的内在生命伦理。

参考文献

- [1] 崔佳祺. 阿尔法狗要逆天! 韩专家称其故意输李世石一局 [EB/OL]. (2017-01-13) [2021-01-30]. <https://tech.huanqiu.com/article/9CaKmJZlZU>.
- [2] 谭光辉. 可能情感的可能性与数字人文研究的新动向 [J]. 西南民族大学学报, 2018(9): 167-173.
- [3] 赵毅衡. 符号学: 原理与推演 [M]. 南京: 南京大学出版社, 2011.
- [4] 蒋柯. “图灵测试”“反转图灵测试”与心智的意义 [J]. 南京师大学报 (社会科学版), 2018(4): 76.
- [5] 肖峰. 人工情感: 技术与人文的新融合 [J]. 马克思主义与现实, 2012(1): 91.
- [6] 黑格尔. 小逻辑 [M]. 贺麟, 译. 北京: 商务印书馆, 1980.
- [7] 弗兰西斯·福山. 信任: 社会道德与繁荣的创造 [M]. 李宛容, 译. 呼和浩特: 远方出版社, 1998.
- [8] 赵毅衡. 符号学——原理与推演 [M]. 南京: 南京大学出版社, 2018.
- [9] 赵毅衡. 赵毅衡形式理论文选 [M]. 北京: 北京大学出版社, 2011.
- [10] 罗兰·巴特. 叙述结构分析 [M]. 赵毅衡主编. 符号学文学论文集. 天津: 百花文艺出版社, 2004.
- [11] 赵勇. 科幻电影中的梦叙述 [J]. 北京电影学院学报, 2017(3): 134.
- [12] 袁海燕. “自我的忧虑”——科幻电影中赛博格的主体性诉求 [J]. 当代电影, 2017(11): 46.
- [13] 周昌乐. 智能科学技术导论 [M]. 北京: 机械工业出版社, 2015.

A Lying Artificial Intelligence: The Production of Fraudulent Texts in Sci-fi Movies

Zhao Yong¹ Wang Yao²

(School of Medium and Arts, Chongqing University of Posts and Telecommunications, Chongqing 400065)¹

(China Film Art Research Center, Beijing 100082)²

Abstract: Lies are not words that do not mean what they say, but words that do not mean what they say. To master the fraudulent texts is an important index to judge whether artificial intelligence has evolved or not. Science fiction movies have all sorts of narratives about humans and robots deceiving each other, which not only outlines the complex aspect of human-machine relationship in the post-human context, but also expresses a kind of nervous concern about the ability of artificial intelligence to “tell the truth”. This article makes its discussion mainly in three aspects: The intention difference of man-machine deception, the text characteristic of the lie world and the choice and acceptance of the “deceived” subjectivity to explain how artificial intelligence “shares” human values that are integrated into post-human totality.

Keywords: artificial intelligence evolution test; human-computer interaction; text truth-telling; post-human value

CLC Numbers: J905 **Document Code:** A **DOI:** 10.19957/j.cnki.kpczpl.2021.01.004