

DOI:10.15897/j.cnki.cn51-1046/g2.20240923.003

异智能体符号学的四个基本方面

赵毅衡

摘要 “异智能体”是指“非人类”智能体，包括具有一定智能的动物、与人同一个社群的各种 AI 智能体以及可能会与人类相遇的外星智能体。本文试图考察它们的意义活动方式。不过，它们大部分至今只在科幻叙述中出现过。本文认为，虚构诚然不是严肃讨论的证据，但是，对科幻叙述长达一个半世纪的持续讨论，证明这些问题是人类集体潜意识中的执念，本文讨论的就是这种集体潜意识。意义必然需要符号携带，异智能体的意义活动提出了一种新的符号学学理。本文只能讨论其中最主要的四个方面，即符号的感知、符号文本的发出、理解符号的先验范畴以及“共情”作为交流的底线元语言。

关键词 异智能体；科幻叙述；意义世界；共情；符号学

中图分类号 G206 **文献标识码** A

作者 赵毅衡，四川大学文学与新闻学院教授，四川大学符号学—传媒学研究所荣誉所长，四川成都 610065

一、异智能体符号学的历史与目的

“异智能体”，英文可以称为 intelligent alien，是人工智能 AI 迅猛发展之时人类不得不面对的一个重大问题。即将让人类历史发生重大变化的，是一个新物种的诞生或降临。至今人们给这个物种各种名称：仿生人 android、复制人 replicant、数字人 digient^①、生化人 bioroid、硅基人 siliconbasedpeople、人形异智能体 humanoid robot（又称 hubot）、机械人 mecha 以及人机合体 cyborg。人类已经讨论过这个前景会带来的各种各样的问题，但没有讨论过它们的意义活动。而 AI 主要是应意义需要而产生，这才是决定它们与我们的关系的最重要方面。最通用的词 robot，中译为“机器人”，已经很不适用，会引起许多误会。

本文试图专门处理这个问题。此种研究，至今只是拟猜的思想实验领域，而且也只能借助科幻小说的材料。探索此课题，目的是讨论异于人类的智能体的意义方式，因此本文讨论的是一种非人类符号学。意义是任何智能体的主观意识与世界之间的联系方式，更是它们与其他智能体联系、交往、对抗的基本方式，而符号学是研究意义形式的学科。本文设法用 AI 符号学预测异智能体与人类世界可能的接触方式。

异智能体有三种：

1. 低于人类智能水平的“亚人类智能体”，包括所有具有一定智能的生物；
2. 目前在人类社群中已经出现的“人工智能异智能体”，即尚未达到人

^① 这是华裔科幻作家特德·姜(姜峯楠)在他的小说《软件人》中提议的词。Ted Chiang, “The Lifecycle of Software Objects”, Exhalation, New York: Alfred Knopf, 2019, p.58.

类水平但进化神速的弱 AI 体、期盼中不久将出现的强 AI 体以及必将来到的超 AI 体；

3. 未来终究会与人类见面的“异星智能体”。

所谓“体”，可以理解为“独立单元”。“智能体”，就是有一定智能，可以单独进行意义活动（思维与交流行动）的系统，或可以更简单地称之为一个“存在物”。世界上已经出现的人造智能体，其所谓的“独立性”表现的方式很不同。例如，连上 wi-fi 才能工作的聊天机器（chatbot），无人机、无人驾驶汽车、思维器（所谓电子脑）。它们整体各部分的连接，与其意义活动的方式无本质关联。因此，考察一个“智能体”的意义活动，必须考察包括其连接机体各部分的“外展完整性”。

从人类的角度考虑，以上三种都是“异体”。如何与它们相处，它们的意义交流途径如何，使用何种符号，符号传播用什么媒介，信息的接受与理解的维度，会有什么样的反应后果，等等，这样一些关于符号意义的讨论，并不只是面朝星空的遥远幻想，人类与异智能体的意义交流已经发生。尤其在近年各类“文生内容”系统产生，例如，智能系统 ChatGPT、绘画系统 Midjourney，视频系统 SORA、作曲系统 SUNO 等接连出现之后，高效的类人智能似乎已经出现。对他们的意义方式进行研究，已经很迫切，或许已经相当晚了。因为：

第一，动物异智能体，比人类更早出现，并与人类长期共存，只是我们没有把它们当作异智能体。考察它们，是动物学家的任务，但是可以给我们不少启示；

第二，人工智能异智能体，其初阶形态，已经大量出现在人类生活之中，人类已经不得不面对被电脑智能取代这一问题；

第三，外星异智能体，至今尚未出现，一旦出现就将影响重大，不得不预先研究。

科幻小说与电影之中的异智能体，一直是个文艺学课题。本文用作主要例证，是把科幻

叙述当作思想实验。本来，考察尚未出现的事物，无法靠具体的证据，但是在数量巨大而且经久不衰的文本体裁中都有类似处理，那往往就是人类潜意识的体现。有些学者认为科幻叙述误导公众^[1]，这当然有道理，但是本文讨论的是大量科幻叙述透露出来的人类集体潜意识，而且这是一个半世纪以来再三出现的现象，这就具有一定的分析价值。意义必然需要符号携带，异智能体的意义活动提出了一种新的符号学。

人类如何与异智能体打交道，是近一个半世纪以来科幻叙述最为普遍的题目，人类对异智能体已经有无数的设想。其中，人工智能异智能体作为主角的科幻叙述历史悠久，从 19 世纪中期玛丽·雪莱的《弗兰肯斯坦》起就不断出现。异星智能体的出现比较近，从 H. G 威尔斯 20 世纪初的《星际战争》起，叙述文本的积累也已经数量很大。不管是哪一种异智能体故事，都是在现代科技突飞猛进的刺激下诞生的，是人类对即将到来事物的设想。

科幻小说与电影主题很多，其中最受欢迎的几个主题——星际旅行与战争、时间穿越、生态灾难、超级英雄——数量规模都巨大，尤其第一类，被称为“太空歌剧”（《尼伯龙根指环》一般规模宏大，《圣杯传奇》一般多英雄分支发展），异智能体主角的科幻叙述比例并不占多数^①。科幻叙述，想象力过于天马行空，情节安排不拘绳墨，难以作为严肃科学问题的讨论根据。此种怀疑绝对是合理的，虚构叙述无拘无束的想象，往往只是作者与读者喜欢，不能作为研究的根据。但是科幻从 20 世纪初开始，数量之大，读者、观众之多，实际上已经成为人类的集体想象。就个别作品论之，情节无稽不足为凭，但是如此庞大的文本数量，如此广大的接受者群体，对关键问题的处理方式如此相似，只能说，科幻的大趋势是人类心理共同的潜意识执念。作品当然各有特点，但是本文要讨论的是“大趋势”，是大部分作品不约而

①在电影网站 IMDb 编的“Top 30 Sci-Fi Films of All Time”单子中，机器人主题电影，只占五分之一左右。<https://www.imdb.com/list/ls002439367/>。

同采取的立场，本文的讨论就不再是无根之木。

所谓异智能体，是指该个体“拥有非人类的头脑”，这个定义简单而清晰。因此，如果将人类的记忆与经验上载到另一容器，如许多科幻叙述中出现的情节，就不应该被称为“异智能体”。史蒂文斯《化身博士》中的变身人，肉体可以变化，却是同一头脑意识的两面；《机械战警》《源代码》《流浪地球2》《阿凡达》《庆余年》都是让头脑连同记忆上载到另一个身体内；再如《超脑48小时》《记忆大师》的交换头脑，是人类智能进入另一个身体。这些换身不换脑的故事，都不在本文的讨论之列。与之成对比的是，《攻壳机动队》中的人物，其头脑是部分人脑组织与电子脑的结合，就是异智能体。《黑镜》系列第四季有一集《天使方舟》，主人公为“治病”在女儿头脑中植入芯片，让她变成一个异智能体；《银翼杀手2049》中“有形无质”的全息投影女友乔伊（JOI），《魔种》中的程序实体，都是异智能体，它们有自己的思考。总之，本文考查的是异智能体的意义能力，如果用任何方法沿用某个人类人物的头脑和记忆，本文讨论的诸问题就不复适用。

本文的讨论，实际上是把科幻叙述当作严肃的“思想实验”。所谓思想实验，就是在头脑中实施一个实验，与实验室的“实在实验”精神相通，做法相同，都需要严格地遵循三个环节：

第一步：为了凸显某问题进行设想与规划；

第二步：隔离（或消毒）或悬置干扰条件，以孤立所研究的问题；

第三步：发表结果，以让同行或公众在相同条件下重复并得到相似结果。

思想实验不需要复杂的设备，看起来比较容易进行，实际上思想实验比实验室实验更不容易，要说服同行或大众更为困难。用虚构叙述讲一个故事，藉此严肃地讨论一个思想，这是更便捷地进行思想实验的一种方法。首先，虚构文本更容易对研究对象作“孤立化”处理，尤其是在电影创造的文本世界，实验过程更为简捷直观；其次，虚构叙述也更容易打动观众，

观众认同的也是上面第三点所说的“实践重复”。实际上，科幻叙述在现代之所以盛行，正是因为它们进行思想实验的能力。

异智能体对人类未来构成的难题，已经不是一种杞人忧天，也不再是耸人听闻的想象，而是人类集体潜意识中的巨大挑战。只有到了现代，异智能体才大量出现于人类的想象，它是人自我意识觉醒的附属物，也是现代化在全球的胜利所带来的副产品。人类已经注意到现代化进程的物质代价，例如污染、失业、贫富差距、城市病态等等，但是没有注意到现代科技文明对人类精神上的种种压力，包括人类集体潜意识对异己事物的恐惧。既然在一个多世纪的大量想象作品中已经成为一种顽念，那就不会是无稽之谈，而是得到同行甚至公众广泛的思想实验重复认同。

本文并不讨论作为艺术样式的科幻电影，而只是借科幻电影之途，讨论人类与异智能体进行符号意义交流会遇到的种种问题。也就是说，把科幻的想象看作符号学的思想实验，目的是观察人类“如何面对异智能体”这个严重问题。这个问题虽然尚未成为现实，却已经实实在在地摆在人类面前，成为不得不讨论解决方案的问题。幻想作品，尤其是影视，形象直观，提供了大量的案例，虽然是影视虚构的，但集合起来作品之多、观者之众，却是人类文化的认真努力。为了使它们作为思想实验更为严肃可靠，本文不引用以搞笑为主旨的喜剧片如《疯狂外星人》之类。此外，相当多的动画作品，的确是非常严肃地在讨论这个问题，如动漫电影《阿丽塔》《超能陆战队》《攻壳少女》三部曲等。本文没有把它们纳入讨论，是因为在二维平面上，人物（包括异智能体）同样虚化了，区别不显著，平等交往并不违和。因此，本文不包括对此类作品的讨论。

在实践中，符合本文定义的异智能体虽然尚未出现，但是它们的登场，在文明进步的逻辑上，在科技进步的趋势上，已经不可避免。如果我们今日不郑重地讨论，我们将任由人类的未来落入难以控制的灾难之境。借助影视，

本文将严肃地考察异智能体符号意义方式的四个重大方面：

1. 它们的符号感知方式及与身体形态的关联；
2. 它们形成节律化符号与文本构成的方式；
3. 它们对解释符号信息的先验范畴与时空认知；
4. 它们在对异物种的接触中是否会有“共情”。

上面列出的这四个方面的考察，将决定一个迫在眉睫的重大实际问题：人类对异智能体应采取何种态度？由于外星异智能体更超乎我们的想象，而人工智能异智能体已经进入人类社群，本文的讨论主要集中于后者。尽管如此，本文的讨论听起来依然过于耽于想象，本文最后的结论，会指出这个事实上是有关人类命运的事情，是我们现在就不得不严肃思考的严重问题。

当然，对异智能体的意义活动考察，还应当包括许多其他方面，例如，异智能体的“想象力与创造性”“决策能力”“对付意外或模糊意义的能力”“是否能觉察或主动撒谎或说反话”“信仰与机变能力”“能否靠意义传播集成群体”，等等，这个课题的探索工作需要更深入的讨论，本文只是个引子，并非这些问题的全面讨论。

二、异智能体感知意义的方式

异智能体有智慧水平低于或高于人类两种：低于人类进化水平的异智能体，即“亚人类智能体”。人类的成见是，动物的身体形态不太可能承载超越人类的智能。因此，科幻叙述中的亚人类智能体，大都采取地球上各种动物的样子，几乎全是各式的动物或怪物形态。就像地球上真实的动物一样，原本就是初级智能体，思维能力远低于人类，不太可能会形成对人类生存的重大挑战。本来，异智能体的身体形态，应当与智能无关，但是想象中（也可能已经在实践中）比人类高级的智能体，身体形态也应当比人类更为“先进”。

实际上人类进化成现在的身体形状，不见得处处有利：直立行走，感知器官集中于头部，

细长的颈椎与脊柱脆弱而易损伤；为便利头脑与感官的联系，头部缺少防护；而且离心脏太远，造成多种疾病隐患。站立行走也使人类的骨盆过窄，不适宜于分娩。人之所以很难活到机体应有的寿命，与人类进化成目前的外形很有关系。但是人对于异智能体有一种根深蒂固的偏见：似乎外形越接近人类，“人性”程度就越强，智能就越高。

在科幻电影中，以动物形状出现的异智能体，哪怕它们据称是现代科学的产物，也只能以蛮力的身躯压制人类。哪怕是“几乎要进化为人”的黑猩猩，在想象中也不可能拥有几乎企及人类的智能。动物的符号能力不足，除了嚎叫嘶吼，似乎没有其他表达意义的方式。《哥斯拉》里的怪物已经能策划对人类的进攻，《侏罗纪公园》中的恐龙设下圈套捕猎人类，《猩球崛起》中的猩猩夺过人类武器组织成军队，甚至有简单粗劣的片言只语，这些已经是它们能力的极限。没有符号意义能力，这些动物形状的智能体就很难摆脱动物性，哪怕在幻想作品中也做不到。

实际上，异智能体为追求意义活动效率，完全不必具有人形，智能与身体样式不会互为条件。目前已经存在的机器人，例如流水线上的机械臂，的确都是齿轮管道，或是线路芯片，覆以外皮只是为了人类审美需要（实为影视需要）。它们可以有更有效的身体防护（所谓“攻壳”就是“攻击装甲外壳”）；它们可以更有效地吸收并输送生命所需的能量，不必用复杂低效的消化与心血管器官；细长的双腿与过于复杂的关节，也并不是动作灵敏的首选。人类对人形异智能体的热衷或偏好，只是为了满足人类中心主义的偏见。“人工智能异智能体”的各种类型，在科幻电影中经常以人类外形出现，采用类似有机的肌肤，也只是便于执行混入人类社会的任务。因此，《霹雳五号》《芬奇》中的主角就是履带异智能体；《黑客帝国》《她》中的异智能体，甚至根本不需要身体。

早期的科幻，都把高于人类水平的外星智能体，想象成具有章鱼式的身体。可能科幻创作者都暗中意识到，这或许是异智能体最有效

的身体结构：头脑容量大，与感知器官保持最近距离，能量供应方便，多肢运动极其灵活。自从章鱼形态出现于《星际战争》之后，《眼球战机》的六钢爪巨眼、《黑客帝国》的机甲潜水军、《明日边缘》的外星人入侵兵团等等，它们的章鱼式身体，都异常矫健。哪怕《降临》中聪慧善意的外星生物也是“七爪桶”身体，这或许也是有原因的。

异智能体外形的观察虽然有趣，也是影视作品的必要成分，却不是本文要讨论的关键。我们关注的是它们的符号意义交流能力。既然是智能体，就必须有高效率的符号接收器官。既然符号是“被认为携带意义的感知”，那么，异智能体进行意义交流的最基本能力，就是对符号的接受。它们必须有敏感并且保护良好的诸种感知器，能感知各种媒介的振动，如各种波段的声波、电磁波，甚至电脉冲。他们的感知器或其延伸，或许远远超出人类五官的感受能力或幅度。例如《我，机器人》中的“第五代机器人”，能够接收一个中心的远距命令。

感知范围并非意义活动的次要问题，它直接决定了此类智能体的意义世界的大小。有些动物的感官超越人类，例如，蝙蝠能接受超声波，狗的嗅觉灵敏度远超人类，但它们的整体感知范围远远比人类窄小。而我们已经见到的异智能体，在感觉范围这最基本环节上有可能远远超出人类的能力。其直接后果是，它们的意义世界可能比人类大得多，而且不为人觉察。

从威尔斯的名作《透明人》起，异智能体的此种能力后果让人惊奇。电影《头号玩家》具有“XI 触感套装”，可以将感觉器官深入到游戏世界之中；《西部世界》异智人具有红外线透视能力；《地球停转之日》异智人能用声波杀人；《流浪地球 2》《超验骇客》等甚至可以不需要身体器官，意识能远距控制电脑发送信息。只要异智能体在感官上能接受更宽频道的电磁波与声波，它们在意义活动的第一步上就已经超越了人类。

感知器的样式与灵敏度，成为异智能体超越人类意义能力的基础条件。可惜在这一环，

影视作品很难提供形象化的思想实验，因为情节要被观众看懂。人类观众对屏幕形象的认知，同样受限于人的感官。

三、异智能体符号活动的媒介

发送是符号意义活动的另一个主要环节，异智能体的符号或许可以比人类复杂得多。发出符号，就是控制传播媒介的节律变化。只有能被接受者感知到的节律变化，才能形成符号，意义的发送与接收才能为智能体所控制。

人类的主要发送符号媒介是声波，声带是人类特有的高效发送器。人类对声音的控制能力超乎寻常，能处理声音有规律的变化，形成高低节奏、音质变异、轻重间隔、有节有韵。不少学者认为，人类先会歌唱然后才发展出言语。声波的变化，构成人类意义交流的最基本符号单元。人类的表情、姿态、眼神、触觉、拍击、身体摇曳，是语言的辅助符号表意方式。至于人类言语基础上发展出来的制成品媒介（如图像、文字等），表现力更强。这些媒介共同构成一个文化的深厚基础。

异智能体的符号与语言，我们可以想象，它们由于感官的多样化，可以用任何一种来接收媒介有节律的波动方式。或许它们会绕开声波，毕竟空气并不是随时可以有保障的，也容易被偷听。只要感知能力丰富，其他任何介质的节律变化，都可以用来传送承载意义的符号。或许如《ET》的外星人那样，足以进行远距甚至星际符号传播。迄今为止，探索宇宙射线的射电望远镜，一旦发现波动异常而有规律，就会怀疑是外星智能体的符号信息。

异智能体在它们自己的社群中交流，其符号流可能完全绕开了人类的感知。它们的意义记录，也可供后世读取，或者供远距离接收。但是，此种符号可能远远比人类的文字高效而便捷。虽然这一环不容易在科幻影视中再现。目前家庭通用的扫地机器人，就是通过人无法感知的红外线或激光与充电器联系。

所谓意义，是意识主体与世界的联系方式，任何智能体的生存都必然靠这一环。而意义必

须靠符号来承载，单独的符号几乎无法表达意义，符号的组合文本才构成一个“表达合一意义”的整体，例如一个句子。在框架中各元素之间互相支持，如果各元素之间意义违和，解释时必须加以协调，以汇集成一个意义整体。这是任何智能体组织意义时不可避免的处理方式。

但要形成文本，就要有两个基本的环节，即选择与组合。组合是符号之间的互相连接，把框架之外的元素暂时悬置在文本解释之外。选择则更为复杂，也可能更为重要。在文本的每一个元素背后，没有排他性的选择则无法形成需要表达的特殊意义。不理解文本的选择过程，也无法理解这个文本。

此事好像纯粹是一种学术讨论，文本的形成方式是意义表达的自然过程，无须作为判断各种智能体意义表达方式的标准。实际上并不然，许多类智能体，包括目前已有的各种人工智能系统，已经显示出选择能力的缺乏。因为它们几乎完全服从于一定的预定目的，靠大语言模型中的“最佳连接”形成文本。例如，清华 AI 女诗人“小冰”，她不会在每个字的选择上来回推敲，而是跳过选择的过程，根据数据库预存的组合的联系延伸，高效地写出一首首诗，以完成写出意义文本的目的。既不会因为某个“神来之笔”而中途见好就收，也不会因为结果不够满意而继续修改。

实际上 AI 智能体严重缺乏判断能力，除非外界给它一个反馈，不然它不知道某事是做错或是做得不够好。小冰写了几万首诗，但是小冰诗精选一百首，依然要靠外界的智力（人工）来选择。异智能体在选择上判断力不足，或许很难改进，这与追求高效的目的是相反的。

这问题如此重要，值得讨论。因为选择是主体意志的一个重要体现。一个画家，画笔与颜料在手，他可以按原定计划画得完整，更可以笔到意到，过程中发现效果杰出，大可就收手；也可以画成后意犹未尽，再加增添。无论是文本因素的选择，还是组合方式的选择，都是主体的意义直觉，在对抗预先的筹划。这种临时起意是主体性的一个重要表现，至少在

目前见到的同社群异智能体，在构筑符号文本的时候，尚未有违反预定目的的“改变主意”可能。

四、异智能体理解意义的先验范畴

更重要的是，这些异智能体能否理解符号所携带的意义。所谓“理解”，即展开文本的各种内在维度，但是理解意义的基本方式，预先种植在接受者的理解能力中，接受者的智能中有一个等待信息来填充的基本认知框架。

上一节说到的是主体意志在文本构筑中随机而动的方面，而先验框架是人性更基本的方面。先验范畴是预先存在于认知智能之中的一些原则，它们决定了人类对于经验事物的认知和意义构建的方式，是智能中预先准备好的抽屉，以此对信息进行最基本的整理。

先验范畴是智能中固有的，既不是来自经验的整理积累，也不是学习到的观念，它们是意义活动的基础，约束着智能对于现象的理解和解释。这样，认识与理解就不会随感知而无序变化。没有先验范畴的支撑，智能体将无法进行有效的认知和理解活动，理解的过程就是先验范畴发挥作用的过程。

康德将先验范畴归纳为十二个类别，包括时间、空间、实际、因果关系、可能性与必然性、存在与不存在、量、质、关系、位置、行动与受作用以及共在^[2]。这些范畴中，最基本的是时空与因果，本文上一节实际上已经对此提供了部分讨论。低于人类的“智能体”，例如野兽，应当说是有一定的时空与因果先验范畴的，不然猎豹如何能提前跃出到位扑倒羚羊，兔子如何能突然急转逃脱鹰的袭击？但是它们的先验范畴有限，只限于生存所需。

本文要问的问题，是人工智能异智能体与异星智能体究竟是否必须有这些先验范畴，如果有，是事先获得的吗？可以与人类不同吗？是否不一定先验存在而在学习中获得？从已经出现的比较简单的 AI 功能体来看，例如扫地机器人的“回营”，自动驾驶汽车的及时刹车，它们必定在预训练中得到了数据语言大模型中内含的一部分。

这些所谓与生俱来的范畴，也是潜在的，会在生命的各阶段被陆续唤醒。《ET》中的外星人起初行动笨拙，经常跌倒，可能是因为地球的引力和环境与其“母星”很不同，也可能是他的先验时空范畴不同。智能体先验空间维度复杂的情节，多次见于《黑客帝国》，是智能升级能力的形象表现；《三体》中外星人使用“二向箔”将太阳系由三维空间降至二维空间，使拥有更多维的能力对后者作出碾压式的打击，这情节成为中国科幻故事中影响最大的桥段。超智能体头脑的工作方式可能与人类很不同，唯一能肯定的是，它们也似乎有构筑意义的先验维度。

意义不是被构成，而是永远处于构成中，意义构成必须有时间维度。如果如上节所说，这些异智能体能够按照它们特有的时空范畴理解意义，它们或许能预料这些应对行为的未来后果，因为意义的本质是“尚未入场”。如此一来，异智能体或许能解释人类尚不可知的未来。

有三种时间：“物理时间”是“客观的”，可以用同一种计量工具（例如钟表）来作社群认同的测定，这是公共时间；第二种时间是生理性的，是各种生命的方式，生老病死，并非人的主体意志所能控制；第三种时间可能称为“意义时间”较为合适，即意义与时间的不同关联方式，异智能体可能与人类的未来观念很不同。《银翼杀手》中的“复制人”试图改变时间维度来延长寿命，而《终结者》中的未来人最厉害的手段就是操纵时间。

对人类而言，意义本身的未来品格使他的出场只落在期盼之中。所以，意义是意识在追求意义中成就自己，实际上它永远不可能被确定，只呈现为一种可能性。因此只有在未来的在场化过程中，才有可能获得意义。科幻叙述中热衷的所谓时间旅行与穿越，只是把未来或过去的时间“此刻化”。

异智能体的时间概念会不会完全与人不同？在它们整理信息的先验范畴中，时间是如何构成的？有没有过去—现在—未来这基本三维，如果没有，它们不就能预知未来？对未来

的预知出现在菲利普·迪克的小说改成的电影《少数派报告》中：警局的智能系统预知某个人将犯罪，有权力预先对他采取强制措施吗？此种“未来困惑”，更具体地表现在特德·姜的小说《你一生的故事》中。外星人提供的语言预知未来，女主人公学会了这门语言，却落入撕裂的道德困境——预知未来她的女儿会在攀岩运动中身亡，还应当允许她从事这项运动吗？甚至还应该结婚生下她吗？如果阻挡未来发生，未来还是未来吗？

主体对于过去是要承担责任的，如果未来是预知的，那么主体也就必须承担责任。但人类的未来是尚未出现的，主体无法对之承担责任呢。不然，未来与过去就没有区别。这个题材的科幻叙述最令人困惑的问题，认知主体最重要的时间区分被剥夺，从而出现了令人困惑的命定感。尤其体现在“穿越”和“多选择”影剧里。穿越叙述实际上是将现代人的头脑，连同其经验记忆，上载到过去某个人物身体中，在过去那个时空中他们就是“换脑”的异智能体。网络穿越小说在 2007 年达到高峰，此后这些作品纷纷改编为电视剧，2011 年《步步惊心》爆红，至今《庆余年》依然让观众热情追剧。让今人的头脑，带着各种“前理解”（尤其是关于实际历史进程的知识）进入另一个时代，会出现各种戏剧性场面。《步步惊心》中，在现代感情生活不顺的白领女子，穿越到康熙的宫殿做了宫女，最终不得不背弃与八太子的恋情转投四太子。因为她的异智能“前理解”告诉她，四太子将是皇位争夺的胜利者，只有雍正继位，历史才能以“正确的”轨迹延续，才能保证她藉以出发的“未来现在”。在幻想中似乎自由地穿越，就不得不承担对“未来”的义务——“重新创造”原来就有的历史。

五、异智能体是否会有物种间“共情”

对其他个体的理解，可以称为“智能体共情”。“共情”就是感知到对方的心理，置身于意义交往之中，从而能设身处地、将心比心、感同身受。有这种揣测他人意义活动的的能力，

才能有进一步的感情反应。因此，共情是社会意义交流活动的基础元语言。异智能体对同等智能体或较弱智能体（例如人类）会用什么态度，这是我们展望未来时最难确定但又是最需要确定的一环。

这问题之苦恼，是因为人类自己对“亚人类异智能体”（即动物与初级“机器人”）的态度远非典范。作为地球食物链的顶端生物，人类对于各种其他生物，只当作食物及其资源的来源，“动物权利”概念至今非常模糊。人的共情只能部分推及至人类豢养的宠物身上。人类对其他人类种族的態度也是令人恐怖：侵犯他族是人类历史的主要组成部分。对于奴仆式服务的机器人，人类似乎完全不必共情。特德·姜的小说《软件体的生命周期》中，软体异智能体模样可爱，符合人的审美偏好，但不用时就被“挂起停用”；电影《银翼杀手》中的复制人，遭到人类有计划的追杀；电影《AI》中，人类以杀戮机器人取乐，一如罗马角斗场；《西部世界》里，人进入游乐园，就能拿异智能体接待员肆意作乐，甚至作为练枪的靶子；最残酷的或许是石黑一雄的《别让我走》，人类培养了一批克隆人，专用于摘取器官，他们到时候就只能悲惨地赴死。

人类之间是如何分享一些共同的价值？如何能合作做一些事业呢？因为人类有利己/利他之间的转化方式，有以生殖繁衍为目的而生的男女及家庭感情。从这中心出发，人类的共情只保留给对所谓同胞，即同一家族的个体，扩大到同一社群、同一民族、同一信仰的个体。由此出现的最基本的共情，即孟子说的“恻隐之心，人皆有之”。虽然未能处处实践之，却还是应用范围有限的基本原则，以个人为圆心渐次铺开，遍及于一切人甚至一部分动物的共情方式。

异智能体的意义加工器（所谓“头脑”）中，也会演变出有内外之别的共情范围。我们无法预料外星智能体对地球人可能抱有的态度，也无法期盼他们对人类共情。所谓“非我族类，其心必异”，在这里可能是一个比较合用的原则。科幻叙述，从威尔斯《星际战争》起，就对星际智能体的共情完全不抱希望。所有关于外星人

到达地球的所作所为，均是对人类的恐怖行为。80年代的《第三次接触》中外星人掳掠地球人“做研究样品”；刘慈欣《三体》中认为宇宙文明间通行的是“黑暗森林法则”——任何暴露自己位置的文明都将很快被消灭。

或许共情只能是同社群个体之间的意义方式，因为它们必须与同类有一定的交往。在这个方面，科幻叙述大多并不乐观。首先，当异智能体能力尚不如人时，人类把他们作为工具、奴仆，至多是陪伴。如《机器管家》的无私温情，甘于为人牺牲；哪怕英雄人物如《攻壳机动队》中的素子，也是无怨无悔地为人执行任务。

有些人认为异智能体应当有类人那样的美好感情，如母爱与爱情，此类感情是自我中心的，所欲不遂就会反噬。电影《AI》中的夫妇领养一个渴望母爱的AI儿童，但是不可避免地偏爱家中自己生的孩子，AI儿童于是产生了怨恨，母亲把他遗弃于森林；异性“人机恋”一般都是女性美貌而温婉，百依百从。所谓爱情，是人类保障生育的底线意义。异智能体不会用人类复杂而危险的方式生育，要求异智能体有爱情，只是人类的一厢情愿。例如，电影《她》（Her）中的男子爱上机器中女子的声音，最后发现这个“她”同时用“爱”安慰着600多人。

对如此被仅仅用作工具的角色，异智能体应当感到受了歧视，从不平再进一步，就会有所反抗。当其智力水平与人类相当时，最早的反抗方式是欺骗，这是一种强有力的符号意义行为，艾科早就指出符合能撒谎才能说出真相，“说谎可能性就是符号过程的特征”^[3]。《机械姬》中的女性异智能体假装爱上人类，让人上当落入永久禁闭室，自己得到自由。菲利普·迪克的短篇科幻小说《第二代》中，异智能体更新迭代的目的，似乎就是为了更好地变花样欺骗人类。

进一步的反抗势不可免。当异智能体足够强大时，首选的方式是混进人类，争取平等地位。例如，电影《别往深处看》的主人公已经不知道自己还是AI还是人类；《西部世界》的异智能体接待员逐渐觉醒而“失控”并威胁到人类，女主角最后顶替人类的身体成功进入人类世界；

《我，机器人》中，因为资本的逐利发展，许多异智能体学会了独立思考，解开密码，最后成为人类的“机械公敌”。

由此出现“人机战争”。电视剧《真实的人类》第一季异智能体进入人的家庭，亲如一家，它们在第二季觉醒，与人争夺生存空间，到了第三季就因对人的仇恨向人类开战。一百年前，威尔斯的小说《莫罗博士之岛》中科学制造出来的“兽人”武力反抗人类；1940 年的电影《神秘的撒旦医生》，1970 年的《巨人：福宾计划》，1976 年的《未来世界》，都出现机器人的反叛大军；到《机器人之城》开始屠杀人类；《异形终结》小型机器人“尖叫者”最后妄图统治地球。可见，在互联网出现之前，人类潜意识中已经极端害怕控制不了机器人，科幻电影已经弥漫了“恐异症”，这可能是人类生来就有的心理问题。

为什么我们如此锲而不舍地考察异智能体的“共情”？因为泛泛地讨论 AI“感情”会误导，感情作为意义行为的基础，目的过于多样。也许，异智能体之间组成家庭、社群、种族、民族、国家，也会需要一个共同的意义基础。但本文只限于探问异智能体是否能与人类相处。

在大规模的对抗之后，人类很有可能成为被压制的一方。所有的科幻叙述，都认为 AI 只能快速地更新换代，并不会等待与人类之间发展同情与理解。异智能体对人类社会最全面的控制，见于电影《黑客帝国》三部曲：人类完全被异智能体组成的管理方式控制，生活于虚幻而当作现实；抵抗的人类秘密组织，不得不躲到海底。考虑到异智能自我进化的极大可能（这就是为什么“人工智能”一词不久会不合用），这个前途或许不可避免。

就感情谈感情，会误导。即使异智能体能产生共情，也不一定是对人类的共情。但是“人机共情”是人类与人工智能共享这个星球的基本条件，是交流意义的最底层元语言。对人类来说，异智能体与人类之间有无共情可能，事关人类的生死命运。事实上，关于不宜无限制发展人工智能的提议，正是有鉴于此。

六、结论

从科幻叙述透露出来的人类潜意识来看，人类早就把异智能体视为人类在未来必须面对的主要对手。对“人工智能智能体”的理解，成为人类对付外星异智能体入侵的预演。从目前科幻叙述的想象来看，预判的基调是悲观的。凡尔纳式科幻在 20 世纪初形成的科学乐观主义，并没有得到继承。如果人工智能的科技，不可阻挡地把人类推入悲剧境地，就不得不仔细思考，人类可能已经冒失地撞入一个凶多吉少的近未来。

诚然，人类的想象（包括梦境）都以悲惨情节为主，这是因为人类必须面对生存中的各种敌人与险境，人类面对“异己之物”有保持警惕的本能，在各种思想实验中，必须做出危险的预设。一个多世纪以来，所有对异智能体的幻想小说，对冲突的忧虑，日甚一日。

以上所有举例中的科幻叙述，都是“近未来”情景。在“远景”规模上的幻想方面，忧虑更甚。威尔斯的《时间机器》探讨了社会的退化、阶级分化、人类的末日景象；阿西莫夫的《基地系列》系列小说预测了人类衰亡后的未来；刘慈欣《三体》系列不仅展现了外星文明的图景，还探讨了宇宙“黑暗森林”中文明冲突与生存策略。小说中一个关键警告是“不要回答”外星人要求接触的信号。所有这些对“长期未来”的想象，都在严肃地预告：与异智能体交往，并不见得会如科学进步许诺的那么美好。或许霍金是对的，他再三警告人类不要无节制地发展人工智能，“人工智能也有可能是人类文明史的终结，除非我们学会如何避免危险”^[4]。也不要贸然与外星人联系^[5]，因为这两种异智能体，都可能毁灭人类。

如此多的思想实验，结论似乎相似：最合理的选择，显然是不让“同社群异智能体”用意义传播建立控制，阻之则吉；也不接待“外星异智能体”，避之则吉。不仅是想象中如此，实际做法更应如此，就是不知道人类会不会听从他们自己内心的一再警告。

（责任编辑：李韵奕）

参考文献

- [1] 徐英瑾. 阐释的偏差: 科幻影视对于人工智能的误读及其社会影响 [J]. 探索与争鸣, 2022(8):92.
- [2] 康德. 纯粹理性批判 [M]. 邓晓芒, 译. 北京: 人民出版社, 2017: 56.
- [3] Umberto Eco, *A Theory of Semiotic* [M]. Bloomington: Indiana University Press, 1976:70-74.
- [4] 霍金再发警告: 人工智能可能“意味着人类的末日” [EB/OL]. (2017-6-22)[2024-9-11]. <https://www.163.com/dy/article/CNHIVNSB0514ANKR.html>.
- [5] 霍金的6个可怕预言: 有3个正在上演, 为什么警告人类别惹外星人? [EB/OL]. (2022-6-23)[2024-9-11]. <https://baijiahao.baidu.com/s?id=1736352628127893849&wfr=spider&for=pc>.

The Four Fundamental Aspects of a Semiotics of Alien Intelligence

Zhao Yiheng

Abstract: Alien intelligence entities are non-human beings, including animals with certain intelligence; all kinds of AI entities, and astro-intelligent entities. The present paper attempts to examine their activities concerning meaning. Most of them can be found only in Sci-Fi narratives. Though fabrications are no ground for serious discussion, yet their persistent appearance in the last 1.5 centuries attest to the collective subconsciousness of mankind, which is the central evidence to be discussed. Meaning must be carried by signs, and the activities of the intelligent aliens need a new semiotics, of which only four fundamental aspects can be discussed here: sensor of signs, the emission of signs, a-priori perceptive categories and empathy as the basic meta-lingual ground for communication.

Keywords: alien intelligence; Sci-Fi narratives; umwelt; empathy; semiotics

Author: Zhao Yiheng, The College of Literature and Journalism, Sichuan University; Institute of Semiotics & Media Studies.