



人文学者对话大模型设计、研究者：如果通用人工智能必然来临，那么——

作为工具或朋友的AI，如何长出善良

圆桌对话

刘梁剑主持：感谢林咏华总工程师极为专业而又前瞻的讲解。这几年华东师大哲学系希望对话方式推动一些新的研究增长点。包括哲学内部

范式改变：成本升降带来AI应用的巨变

刘梁剑：贺樑老师，您来自计算机学院，可否先谈谈对林工主旨演讲的感想？

贺樑：两点印象尤其深刻。第一，范式带来的变化。范式一词内涵丰富，其中也涉及未来人工智能的研究方法。从更通俗的“成本”角度来阐述，ChatGPT等大模型出来后，使用计算机和人工智能技术来赋能行业的成本大幅降低，几乎被打到地板上去了，这可能砸了一些计算机从业者的“饭碗”。一旦某一个地方变成成本洼地，很多事的运转逻辑就会很不一样。

第二，林工把我们这个行业内部不太被人理解但又关键的环节——基础模型、专业基础模型和基础模型的微调，用比喻方式说得很形象——通识教育阶段到读研究生、博士生，最后是上岗培训。最核心的是，服务成本大幅降低，使人工智能的能力真的变得让各行各业可用了，大众的生活由此会产生大变化。

刘梁剑：使用成本降低，但是训练基础模型的成本很高。就像国家的基础设施一样，是不是应该由国家来做这个投入？

林咏华：确实是高成本。一旦不

同学科之间的对话；外部与其他学科如教育、艺术、科技的对话；还有跟现实进行对话。最近在建设应用伦理专业硕士点，让伦理与其他学科之间的对话更有制度保障。今天与科技的对话就是跨界对话的尝试。

开源或者开源之后不能商用，势必导致每一个公司都只能自己去做大模型。除高成本外，背后的基础设施也要靠能源推动。大模型训练需要巨大的算力，同时也是巨大的电力消耗。

所以，第一，产业期待开源，以智源为开始，希望看到有更多的研发团队或公司可以加入开源队伍。

第二，如果要做出高质量的基础大模型，除了前期投入，关键还在于未来要有持续迭代。可持续发展的产业实践是，基础大模型至少三个月或半年有一个新的版本，才能满足更多新知识融入、新技术融入。费用由谁出？在科研界、产业界，现在还没有定论。

贺樑：我认为，大模型的投入应该是市场行为和政府行为相结合。目前整个大模型领域还是美国优于我国，这也是中美在未来科技竞争中的制高点之争。这一轮大模型的竞争是因为Open AI、GPT-3.5没有开源，导致大家有一点恐慌。因为没有开源，未来市场就会被Open AI的生态全部抢占。因此很多企业纷纷在商言商去参与研发。我相信政府会做好引导，提供有组织的科研、有组织的市场竞争。

工具或朋友：人工智能与人类智能关系？

刘梁剑：我想问一个“杞人忧天”的问题，人工智能的发展会不会取代和消灭人类智能？

人类或许只是文化进化链上的一个起点

郦全民：6月16日在巴黎举行的科技创新展上，马斯克说，人工智能技术可能是有史以来最具颠覆性的技术。我觉得从进化角度可去掉“可能”两字。

其实，人是两种进化的主体，首先是生物进化。人在与环境作用的过程中，将心智生成的知识或者经验外化，通过口语，后来又发明文字实现外化，使得人从动物界提升出来。这个飞跃让人成为文化世界或者更广义地讲信息世界的主体，从发明文字算起已有几千年了，因此，人也是文化进化的主体。到了20世纪中叶，发生了最具颠覆性的变化，标志是电子计算机的问世，从此，人类开始把心智运作的力量外化，而不仅仅是外化结果。近十几年来，随着机器学习的发展，这种外化能力突然提速了，以至有了AI科学家、AI工程师。

原来，人类是文化世界的唯一主体，现在出现了人工的智能体，也能生成文化产品，比如文本、图画，为文化进化作出贡献。如何为它们定位呢？是看作新主体还是只看作工具？假如这些AI科学家、AI作者越来越聪明，甚至有自主的意识，又该怎么看？

人工智能的主要创始人明斯基有一个观点：一旦出现这样的智能体，那它就是人类新的后代，是“智能孩子”。其实，如果承认进化论，承认人是进化而来，很难说人类就是进化链上的最后一环，很可能是智能进化或者文化进化的起点。如有朝一日，“智能孩子”会不会把人类像动物一样关起来？我说，这种担忧是人类中心主义的表现。

也请教后面两位同行。

不可控风险：碳基文明被硅基文明所超越

潘斌：人类常常持一种极为理想化的设想，即人工智能仅仅只是人类智能的工具、延展、增强。随着技术迭代的爆发式增长，这种想法可能只是

训练时会设伦理框架，但注入善良依然是挑战

付长珍：从伦理学角度看，这些年很多人都意识到——社会越发展，伦理越重要；科技要发展，伦理需先行。国家也多次出台相关文件，倡导科技伦理要贯穿科技活动全过程，各个环节都要有伦理在场。原有的伦理学是一种近距离伦理，主要侧重后果评估。人工智能时代的伦理学，应该阐发一种面向人类未来的新型责任伦理意识。伦理应该成为科技发展的方向盘，需要前瞻性地预测可能存在的伦理风险。伦理与技术同向而行，确保科技向善。

特别想请教林工，大模型作为底座进行研发和预训练时，对于价值观、安全性、伦理风险等，是否也有一些具体的考量和举措？

刘梁剑：我想这里有一种矛盾，一方面用伦理规范技术，人是目的；另一方面又认为人不可能永远是目的。

林咏华：从偏理工的脑袋想来，这是必须要考虑的。我们至少在几个环节去解决：第一，对大模型的预训练数

据做安全过滤，包括去除伦理偏见等。目前未必所有的企业都这么做，但是智源作为开源基础大模型的提供者，为了减少后续使用者的顾虑，我们从源头开始规范；第二，微调训练的数据也要做这样的处理；第三，今天没展开讲的强化学习，它速成跟人类的价值观对齐。技术上至少是三个环节的流程化处理。

用偏人类的脑袋，我也经常在思考，怎么把人类的真善美、人类的榜样

数字&人文圆桌对谈

嘉宾：林咏华、贺樑、郦全民、付长珍、潘斌、刘梁剑

“数字强国”系列
第163-1期 / 2023年第1期



林咏华



贺樑



郦全民



付长珍



潘斌



刘梁剑



6月20日，163期文汇讲堂“数字强国”系列启动，首期《AIGC驱动生产力跃升与良好未来塑造》在华东师大举办。北京智源人工智能研究院副院长兼总工程师林咏华应邀作主讲，华东师大哲学系杨国荣致辞《科技发展与人类生活》，计算机学院贺樑、哲学系郦全民、付长珍、潘斌、刘梁剑参与数字&人文圆桌对谈，13位听友现场互动。校内外7600余人观看华东师大

微信号直播。现场发放了50枚NFT数字徽章。

本次讲座由文汇报社、上海树图区块链研究院、华东师大中国现代思想与文化研究所、华东师大哲学系伦理与智慧研究中心联合主办。

上图为：圆桌对话围绕四大问题展开充分讨论，并呈现观点的多元。

本版整理 李念 摄影 周文强 版式 李洁

据做安全过滤，包括去除伦理偏见等。

目前未必所有的企业都这么做，但是智源作为开源基础大模型的提供者，为了减少后续使用者的顾虑，我们从源头开始规范；第二，微调训练的数据也要做这样的处理；第三，今天没展开讲的强化学习，它速成跟人类的价值观对齐。技术上至少是三个环节的流程化处理。

用偏人类的脑袋，我也经常在思考，怎么把人类的真善美、人类的榜样

注入到大模型？俗语“三岁见老”，当一个人是善良、有爱心的，没有必要告诉他很多的条条框框，实际上哪怕会犯错，也不会犯大错。但是今天做大模型机器学习时，不得不给出很多的条条框框。坦白说，做安全对齐、伦理对齐时，维度也很多。所以这是一个开放的问题。是否有机会把善良与爱注入到机器呢？如果不能，或许这就是机器和人很重要的区别。

解放或焦虑：大模型带来自由还是失业？

刘梁剑：即便不想有偏见，但怎么没有偏见呢？类似这样的难题就出来了。这里也涉及到伦理风险。马克思畅想的“在共产主义社会，人们上午打猎，下午捕鱼，傍晚开始畜牧，晚饭之后从事哲学批判”，自由社会会因为ChatGPT等大模型出世而提前到来吗？

依据惯性，人类未必都会善用自由时间

潘斌：能出现这样自由状态有一个根本性的前提，即物质产品的极大丰富。通用人工智能的出现会使人类得到更多的劳动解放，拥有更多的自由时间。但依据个体惯性以及数字技术的强大控制力，很多人可能还会沉溺于数字世界的娱乐与游戏之中，在一定意义上，技术的进步不一定必然带来自由人状态。

相反，我们更需要去重视与思考的问题是，它会不会导致很多人失业？技术进步与就业机遇之间如何保持一种适当而有机的协调？

进入老龄社会，人类需要高效的机器人辅助

林咏华：短期可能会使现有职业岗位的数量发生明显变化，但未来将促进职业种类的变化变迁。人类社会有很多文明的进步、技术的进步，带来工作岗位性质的改变，这是正常现象。人类

总是很聪明地找到更多其他需要人类去做但未做的更良好事物。我们相信职业社会有很强的自适应调整能力。另外从人口结构的变化来看，目前100个人需要奉养30位老年人，到了2045至2050年，100个人甚至要奉养50—60位老年人，此时我们需要有更高效率的工具包括智能机器人，来辅助我们更少的人完成更多的事情。

郦全民：今天讨论的是生产力的跃升。生产力有三个要素：劳动者、生产资料和生产对象，其中工具属于生产资料。以往的技术基本上是生产工具，人是惟一的劳动者，但AI不一样，它可以充当劳动者，所以，对生产力的跃升必将是革命性的。

劳动是阶段性需求，激发人的潜能是目标

刘梁剑：请教付老师，为什么失业是人类进化的一个标志？

付长珍：当我们被解放出来获得更多自由时间的时候，我觉得问题不是担心我们失业了，而是该思考如何来安排我们的闲暇并使之不无聊。

人工智能技术的涌现正是出于人的需要，促进人更好的存在。劳动只是生产力不够发达时的需要，因此应该扩展劳动的定义。更有价值的劳动，在于激发出人的更大潜能，提升人的内在本质，塑造高尚人格和美好生活，使人更好地成为人。

来自人类后天的定义，主观性较强。只要人还是世界主体，就有望创造出超越机器能力的职业。但前提是，要规避的标准是由AI来确立和判断，那样会沦为当下的下围棋一样。这些充满创意的环节一旦都变成了机械的确定，就容易索然无味。

为保护隐私，可部署到私有化网络中

通信行业柴忠於：通过ChatGPT询问公司业务，而公司又禁用ChatGPT，具体落地时如何保护隐私？

通用人工智能：何时出现，拥有自主意识？

刘梁剑：形而上的讨论总能脑洞大开，走得较远，这是哲学的特性。那么从技术的角度看，通用人工智能是否有可能出现？更进一步，超强人工智能是否有可能出现？这里的关键是出现自主意识。贺樑：虽然大模型当下依然处在初级阶段，大部分能力还是由数据喂出来的，但“涌现”能力让人吃惊。这种“涌现”能力是一个复杂系统从量变到质变的过程。原先我认为，量变到质变只是在物理世界里存在，在数字世界凡事都可掌控。但这次“涌现”指出了一条人工智能进阶的新途径——复杂到一定程度，它自然会产质变。

林咏华：现在ChatGPT还是一个困在数字域的囚徒。本来以为这一轮最早被替代的是蓝领，实际最早被替代的是白领。AI+X、X的那些学科的发展并没有像人工智能那么快，这也导致机器人有一个很聪明敏捷的脑袋，但是四肢的

准确性、灵活性还没有匹配上。比如，在分拣上的速度，机械臂满足不了10毫秒以内的灵活性抓取。所以离大家期望的老有所依、为我们做风吹日晒下的事情等设想还比较远。

是否有自主意识？这次大模型引领的AI是否到了一个“奇点”？我个人分成三层来看：第一也是最低的，AI可以自主去应对人类的需求，调用有限掌握的几个工具完成人类的任务；第二它是否有能力自主地去学习还没有学过的东西；第三它是否能够自主地去创新，去突破一些我们没有让它突破的难题。比如说怎么让飞机净重减少一半，或者会自主地去看到海洋里因为污染而濒临灭绝的生物，并去设想拯救方案。

现在看到，目前可以初步达到第一个层次，未来10年可以达到第二个层次，而第三个层次将会是很难达到的目标。



13位听友现场互动，其中优秀提问奖获得者与嘉宾合影。

现场互动

大模型走向更智能，尚在集体摸索中

AI行业 Convertlab 公司主管高鹏：今年2月起，我们一直折腾于在大模型上做自动化应用，发现它没有想象中那么聪明。请问技术天花板在哪里？

林咏华：从从业者的角度来看，因为国内大模型大多数情况都还是要依赖人去调教。坦白来说，目前我们还没有摸到天花

众筹降低开发成本，需更智慧的机制

商业协会平渊海：有没有想过发动相同利益的企业众筹开发行业大模型？

林咏华：这个设想从今年上半年开始就一直就有。每一个行业都可以至少找到八到十家共同性质的企业一起众筹，训练行业大模型，做出来后在企业内

审美创造岗位前提：人不依赖AI判断

华东师大哲学系丁洪然：审美是否不可能被AI替代，由此创造出更多的职业选择和生活方式？

贺樑：人类会被动或主动地想办法设计出适合做的事，审美是其中一个可能。它一是来自生物自然的感受；二是

林咏华：关于数据隐私，基于不同的大模型部署方式，通常有两种考虑。一是类似于ChatGPT的某些大模型，在云上回答用户上传问题，理论上，运营商都需要获得授权才能访问用户上传的数据。

二是对于要求信息保密的行业，包括金融机构、医疗机构等，可支持大模型的私有化部署，部署到企业的内部网，以避免商业信息的泄露。

机器如替代人，符合“宇宙的自然选择”

华理博士生胡杨：机器是不是必将

替代人类或者替代人类的创造力？

郦全民：技术发展了会出现超级智能，我觉得这是好事。人类假如要离开地球、离开这个生物系统，进入另外一个状态，就需要新的智能物种超出生物的生态系统，实现进化新的方式。目前，科学界就有一个“宇宙的自然选择”学说，主张普适达尔文主义。人工智能体所处的电子生态系统与人所处的生物生态系统不同，有更广阔的进化空间，所以，我相信AI会超越人类。