支持人在环路混合智能的交互设计研究

胡源达[1]，孙效华[1]，张何辛[1]，张绳宸[1]，易思琦[2]
（1.同济大学，上海 200092；2.香港理工大学，香港 999077）

摘要：目的 指出“支持人在环路混合智能的交互设计”这一类设计问题，研究人在环路混合智能系统中交互设计的问题，为相关设计、技术与应用研究提供索引和参考。**方法** 从人在环路混合智能的概念和架构出发，引出人在环路混合智能的交互设计；基于对相关文献的整理，总结常见界面构成和交互方式；总结整理人在环路混合智能的生命周期。**结论** 指明了人在环路混合智能是需要用户交互的智能模型，介绍了由用户、人工智能算法、用户接口构成的系统架构；总结了针对不同数据类型的现有工作可能的交互方式；分析了人在环路混合智能完整生命周期中的设计挑战，根据现有文献提取关键界面构成，提出了人在环路混合智能系统的设计建议；提出了从智能系统、用户、设计师三方面建立设计方法论，完善设计工具，更有效地支持和推动人在环路混合智能系统的应用的建议。

关键词：人在环路；交互设计；人工智能；机器学习
中图分类号：TB472 **文献标识码：**A **文章编号：**1001-3563(2020)18-0038-10
DOI：10.19554/j.cnki.1001-3563.2020.18.005

Interaction Design in Human-in-the-Loop Hybrid Intelligence

HU Yuan-da[1], *SUN Xiao-hua*[1], *ZHANG He-xin*[1], *ZHANG Sheng-chen*[1], *YI Si-qi*[2]
(1.Tongji Universtiy, Shanghai 200092, China; 2.Hong Kong Polytechnic University, Hong Kong 999077, China)

ABSTRACT: The work aims to point out the problems in “interaction design for human-in-the-loop hybrid intelligent” and study the problems of interaction design in the human-in-the-loop hybrid intelligent system, to provide references for the related research in design, technology and application. Starting from the concept and architecture of human-in-the-loop hybrid intelligence, the interaction design of human-in-the-loop hybrid intelligence was introduced and the interface compositions and interaction models and life cycle of human-in-the-loop hybrid intelligence were summarized based on related work in different fields. It is pointed out that the human-in-the-loop hybrid intelligence requires the intelligent model of user interaction and the system architecture composed of users, artificial intelligence algorithms, and user interfaces is introduced. The interaction methods of existing work for different data types are summarized, the design challenges for each key activity of the full life cycle in human-in-loop hybrid intelligence are analyzed, the characteristics and key interface composition of different user functions and requirements are extracted, and the design principles of the human-in-the-loop hybrid intelligent system are proposed. The suggestions for establishing design methodology from intelligent system, user and designer are proposed, improving design tools to support and promote the applications of human-in-the-loop hybrid intelligent systems.

KEY WORDS: human-in-the-loop; interaction design; artificial intelligence; machine learning

人在环路混合智能系统是机器智能和人类智能相结合，以克服现有人工智能不足的系统。不同于传统的交互设计，支持人在环路混合智能系统的交互设计处理的是用户与人工智能模型交互的问题，用户意

收稿日期：2020-07-18
作者简介：胡源达（1992—），男，新疆人，同济大学博士生，主要研究方向为机器学习及人机交互。
通信作者：孙效华（1972—），女，河南人，博士，同济大学教授、博士生导师，主要研究方向为人工智能与机器人、交互设计与技术、大数据与信息可视化分析、车载人机界面与汽车互联服务创新。

图的不确定性和人工智能模型的复杂性都给设计师带来了种种挑战。在人在环路混合智能系统中，合适的交互设计对提升终端用户的输入效率及系统效用具有积极的意义，有必要展开支持人在环路混合智能的交互设计研究。

1　人在环路混合智能系统

1.1　人在环路混合智能系统概述

随着计算机技术的发展，以机器学习和深度学习为代表的人工智能在计算机视觉、自然语言处理、音视频分析等领域取得了突破性的进展。人工智能具有规范性、逻辑性和可重复性的特点，能够节省人力资源、提高效率，帮助完成单调、重复和耗时的工作。但是人工智能需要从大量数据中提取知识，在专家标注成本高昂的领域中，无法大规模获取带有高质量标签的数据；在某些场景下，受数据质量及优化算法的限制，人工智能仅凭自身能力难以获取与人工相匹敌的效果。人类智能具有创造性、复杂性和动态性的特点，有能力处理不完整和非结构化的信息，并可以从少量样本中提取出抽象特征。Dellermann 等人[1]在其关于人类与人工智能协作的论文中提出，混合智能系统将人类和人工智能相结合以实现处理复杂目标的能力，从而达到优于它们各自可以单独实现的结果，人和人工智能相互学习，不断提高。

人在环路混合智能系统是机器智能和人类智能相结合，以克服现有人工智能不足的系统。学术界对人在环路混合系统没有明确的定义，与其相关的概念有人在环路系统、交互式机器学习、混合智能系统等。Dudley 等人[2]将交互式机器学习定义为一种交互范式，通过用户交互行为迭代地构建和完善机器学习模型。Holzinger 等人[3]在一篇讨论交互式机器学习的论文中，将交互式机器学习系统的迭代优化对象扩大至用户上，用户通过交互优化智能体的学习行为，人类也能够通过智能体优化自身的认知。还有一些文献则进一步描述了混合智能系统中的重要范式，分别是：人工神经网络（Artificial Neural Networks），模糊逻辑（Fuzzy Logic）及全局优化算法（Global Optimization Algorithms）[4]。经过分析和总结，将采用以下人在环路混合智能的定义：需要用户交互的智能模型[5]。

人在环路混合智能系统是通过整合人类的感知和认知能力，以及机器计算和存储能力，实现“1+1>2”的智能系统。该系统可以处理来自大规模，不完整和非结构化知识库的信息，并且可以避免基于机器学习的人工智能技术带来的失控风险[6]。迄今为止，人在环路混合智能系统在自动驾驶、辅助医疗、视频检索等领域中都有所涉及，具有巨大的潜力。

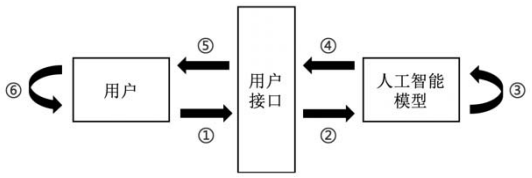


图 1　可持续优化的人在环路混合智能系统架构
Fig.1 Architecture of continuously optimized human-in-the-loop hybrid intelligent system

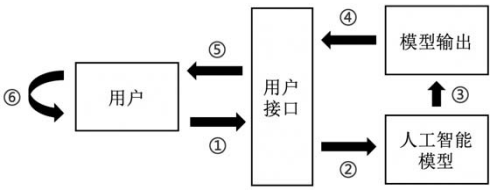


图 2　非持续优化的人在环路混合智能系统架构
Fig.2 Architecture of non-continuously optimized human-in-the-loop hybrid intelligent system

1.2　人在环路混合智能系统标准架构

人在环路混合智能的架构由三部分构成：用户、人工智能模型和用户接口。用户作为整个人在环路混合智能系统的驱动者，专家用户对数据及模型往往有着深刻的理解，能够给人工智能模型提供高质量的交互信息[2]；非专家用户需要在交互过程中不断进行知识迭代和学习，加深对数据及系统的理解。人工智能模型接受特定的输入并根据用户反馈的交互信息确定输出。传统人工智能模型存在多种多样的算法和体系结构，交互式人工智能模型往往需要针对用户输入做特定算法设计，使之能够充分利用用户反馈[2]。用户接口作为用户与人工智能模型交互的桥梁，提供样本浏览、反馈提交，模型评估的功能，并直接决定了用户的交互方式。

可持续优化的人在环路混合智能系统架构见图 1，步骤①至⑥通过交互过程中的数据流向展示人在环路混合智能的系统架构，数据从用户出发，通过用户接口传递至人工智能模型，人工智能模型迭代优化自身并计算得到相应结果，通过用户接口反馈至用户。值得注意的是，人工智能模型的迭代优化过程根据用户交互的阶段的不同有所不同。在主动学习（Active Learning）为代表的可持续优化应用中，用户交互往往在数据标注阶段，人工智能模型结合交互信息进行迭代优化，以更少的标注工作量优化提升模型性能；在其他非可持续优化的应用中，模型需要用户交互信息作为补充以达到更加优良的结果，部署后的模型不再因用户交互而改变，见图 2。

2　人在环路混合智能系统的典型应用

人在环路混合智能已应用于图像、文本、时序数据等多种数据类型中，在图像分割、图像分类、视频

检索、文本分类、音频标注等相关领域产生了种类繁多的应用。不同的用户及场景具有不同种类的需求和目标，难以使用既定的交互设计范式进行实践；此外，与传统界面交互设计有所不同的是，用户的交互的对象是人工智能模型，用户接口难以以直观且直接的方式的反馈至用户；人工智能模型在接收到用户反馈时会不断迭代更新，使其行为难以预测；同样，机器学习算法本质上很复杂，用户难以理解模型反馈[7]。近年来有许多工作尝试研究基于人在环路混合智能系统应对这些挑战，并探索简明高效的交互设计提高用户输入效率及系统效用。本文将按照人在环路混合智能系统处理的数据类型进行概述，展示人在环路混合智能的应用领域及已有工作中对交互设计有所考量的案例。值得注意的是，这里不追求概括所有人在环路混合智能的方法，只选取当下具有代表性的工作，以及在交互设计上具有鲜明特点的工作。

2.1 图像

近年来，得益于深度学习和大规模计算的发展，计算机视觉领域实现了实质性的飞跃。基于深度学习的人工智能已广泛应用于图像分类、图像检索、目标检测、语义分割等领域，取得了良好的效果。

语义分割旨在将类别信息分配给图像的所有像素，在图像理解中具有至关重要的作用[8]。用户通过在图像上绘画以标注像素的具体类别信息，并能够在系统错误分类的图像区域中绘画迭代矫正人工智能模型。首个交互式图像分割算法可以追溯到 1995 年的 Intelligent Scissors[9]，Mortensen 等人使用边界追踪的方法辅助终端用户捕捉边界进行分割，相比手动抠图极大的减少了用户交互次数，并能得到较为准确的分割结果。Lin 等人开发了第一个可通过 Web 浏览器访问的图像抠图工具，能够实时处理用户提供的输入，用户可以根据需要更改画笔的宽度，为抠图模型提供更加精确的边界信息[10]。

为了减少图像标注的人工成本，有学者通过使用用户提供的弱监督信息（例如使用鼠标单击像素点，画直线，曲线或边界框等），研究并开发了交互式分割模型和工具。Rother 等人[11]基于图像前景物体纹理信息和边缘信息，提出了一种迭代算法 GrabCut，在适度的用户交互下实现了高效的前景分割，用户通过不同颜色的画笔，能够粗略描出图像中的前景、背景和图割区域。在 GrabCut 的基础上，Xu 等人[12]将用户提供的粗略的边界框表示为欧几里得距离图，与原图拼接输入到一个编码器-解码器神经网络，获取用户感兴趣物体的分割区域，用户仅需提供粗略的边界框作为交互式输入，即能得到与精准的边界框相当甚至更好的分割效果。

不同于大多数将分割任务表示为像素标注的方法，一些交互式图像分割方法将分割任务转换为多边形预测任务，预测目标轮廓的多边形表示，用户能够实时修改多边形顶点，从而得到用户期望的精确分割结果[13-15]。Jang 等人[16]在医疗图像分割中，通过测量每个像素到被标注像素的距离，将用户的点击标注转换为交互图，输入至卷积神经网络中得到初始分割图，此外，该方法研究了反向传播细化方案（BRS）处理用户不准确的标注。

在如医疗图像分割等某些专业特定的领域中，训练数据需要专业的先验知识，获取难度大、周期长，需要专家耗费大量时间标注高质量的数据，难以完成大规模数据集的采集，导致训练样本不足，模型的泛化性能降低。基于人在环路混合智能的方法有效的应对了这些挑战，但是在落地过程中，受专业知识所限，终端用户难以提供高质量的数据标注；此外，用户对模型的预测，解释，理解或纠正的意义有限。在模型调整阶段，交互设计需要进一步解释模型如何得到预测结果，以增强用户对系统的信任[17]。

2.2 文本

得益于机器学习及深度学习的发展，自然语言处理领域取得了长足的进展。人在环路混合智能系统旨在解决需要更严格或自定义分类的场景，该场景下的标注信息或难以从领域专家中获取，或难以用编程的方式表达[2]。在基于文本的信息检索中，终端用户可以迭代地向系统提供有关呈现给他们的文档的相关信息，随后系统使用此附加信息优化显示的搜索结果。Wallace 等人[18]开发了名为 Abstrackr 的工具，用于综述引文筛选任务，该工具将用户提供的关键词信息整合输入至支持向量机（SVM）分类模型，并使用主动学习的方式交互式地迭代模型。Aberdeen 尝试通过用户反馈学习每个用户的统计模型，对邮件的重要性进行预测和排名[19]。Endert 等人[20]通过机器学习和空间可视化的方式对文本数据进行交互分析，分析师可以通过简单地移动文档来表达提供有效的信息。

相较其他类型的数据，文本数据相对易于获取，并包含了一定程度的手动标注。然而在特征选择阶段，交互设计对文本信息表示的简洁与否影响着用户的输入效率[2]。

2.3 时序数据

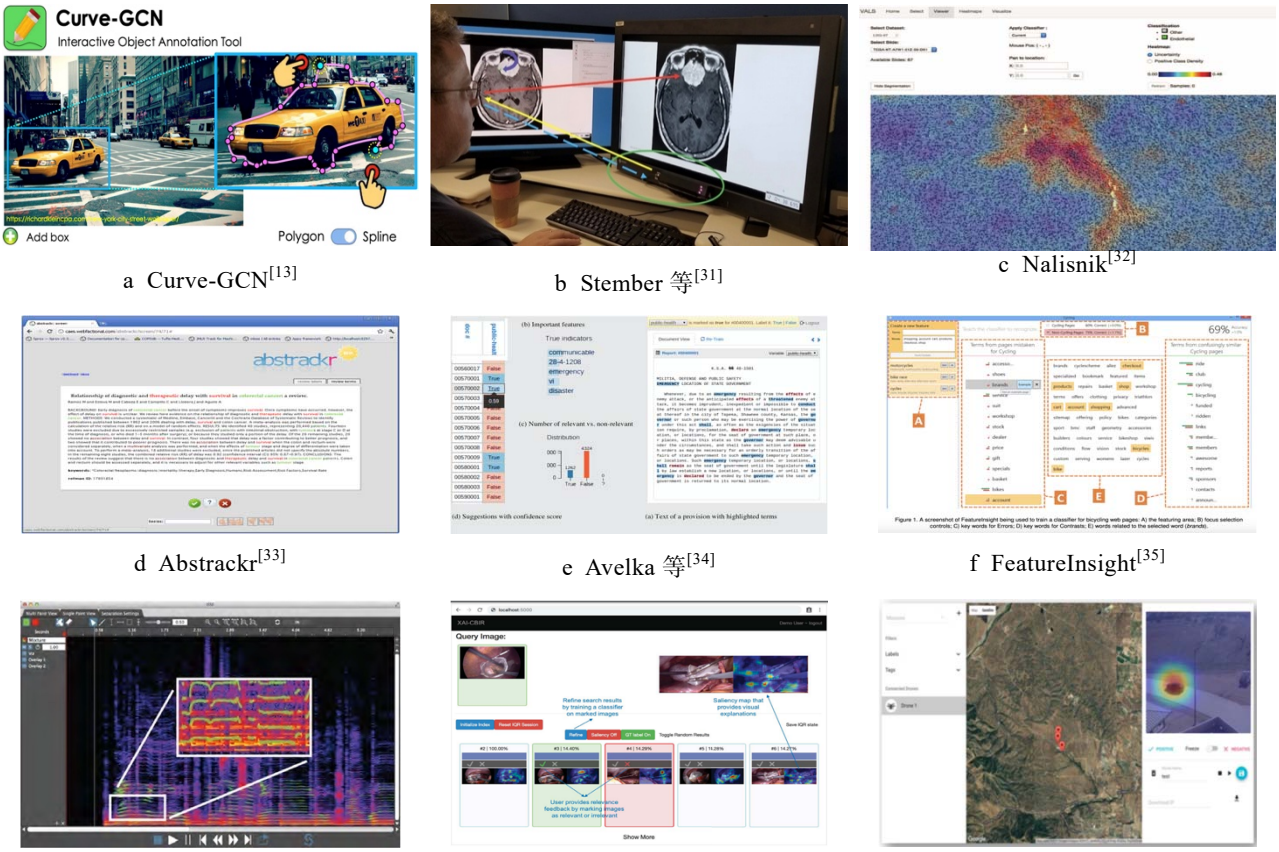
人在环路混合智能在基于视频、音频等时间序列数据诞生出丰富的应用场景，如自动驾驶、视频分类、音频提取等。视频帧可以视为图像，在构建并训练数据集时，可以迁移图像的处理方法与模型，但是视频作为时间序列数据，具有高度相关的上下文信息，很难表达时间、位置、空间之间的关系。类似的，音频数据有和视频数据相同的复杂性，音频片段之间形成可叠加的难以分离的波形信号，且有不同的数据格式。音频和视频可以视为时间序列数据，目标片段分布稀疏，需要耗费标注者大量的时间成本。人在环路混合智能系统一定程度上解决了上述的问题，相比单

一的机器学习方法具有更好的表现。

为了向录音工程师，音乐人和类似用户提供专业级的音频编辑工具，Kim 等人[21]提出了一种基于 Web 的人在环路声音标注系统，该系统的界面设计着眼于将用户的注意力转移到感兴趣的音频区域方便标注和反馈。Bryan 等人[22]提出了一种跨平台音频编辑工具，它使用户能够在声音的时域及频域的可视化中“绘画”来执行音源分离。类似的，Nakano 等人[23]提出了一种基于深度学习的交互系统，将歌声与输入的和弦信号分离，用户通过与系统交互对神经网络模型进行微调，以使其适应目标歌曲，该方法允许用户通过交互逐渐使模型适应特定的歌曲以获得更好的分离效果。BeatBox[24]允许终端用户创建自定义的识别器，界面将不同的发声等级分为不同的虚拟打击垫，并根据打击垫将颜色编码为打击垫的发声，帮助用户注意到未正确分类的记录样本，最终用户能够快速创建可用的分类器。对于音乐创作，RAFiebrink[25]开发了一种通用软件系统 Wekinator，旨在帮助学生和专业作曲家创作新的音乐作品，在交互流程中，用户交互式创建、修改、评估机器学习模型，用户和模型能够通过相互的反馈进行学习。这些研究表明，人在环路系统的目标不只是使人工智能算法模型更有效率或更具生产力，还能够扩大人类的创造力。

在视频图像检索的任务中，Chittajallu 等人[26]提出了一种人在环路可解释的 AI（XAI）系统，用于对视频帧进行基于内容的图像检索（CBIR），该系统通过显著图以直观的方式解释系统检索图像与查询图像的相似性，通过与查询图像相似度的降序展示给用户，并允许用户标注检索图像是否与其相关。ClickBAIT[27]通过将用户标注简化为单击操作，在实时视频流中点击标注的主题，减少了实时训练所需的人力，提高用户交互的训练收益。Feng 等人[28]开发了一种用于改善深度学习模型的半自动运动对象标注方法，用于视频中的车辆检测，该方法包括三个阶段，即自动前景对象提取、后续视频帧中的对象标注以及使用人在环路快速选择样本进行数据集构建。Kabra 等人[29]采用交互式机器学习方法，允许非专业用户建立分类器识别不同的动物行为，用户在标注后可以启动模型重新进行训练，用户可以纠正错误样本的标注以逐步提高模型质量。在医学领域中，Zhang 等人[30]将人在环路的深度学习应用在视力检查领域中，医生通过模型的反馈理解模型并提交反馈，模型根据反馈迭代优化自身，该方法有效增强了医生评估儿童视力的能力。典型案例中的用户界面见图 3。

由于时序数据具有连续的时间维度，难以简洁地表示动态内容中时间与位置等信息，在交互设计中



a Curve-GCN[13]　b Stember 等[31]　c Nalisnik[32]

d Abstrackr[33]　e Avelka 等[34]　f FeatureInsight[35]

g ISSE[22]　h XAI-CBIR[26]　I ClickBAIT[27]

图 3　典型案例中的用户界面

Fig.3 User interface in typical cases

需要探索合适的序列数据呈现方式，允许用户快速浏览和标注样本，以提高用户输入信息的效率[2]。

3 支持人在环路混合智能的交互设计

3.1 交互设计在人在环路混合智能系统中的意义

通过良好的交互设计，可以从两方面提升人在环路混合智能系统的表现。一是支持人在环路中用户的输入和决策，二是通过更为精准的输入和更为可用的工作流，提升系统的效用。

通过交互设计可以支持更有效的人在环路的输入和决策。首先，在需要用户多次输入信息来进行学习的系统中，可以通过设计更为自然的交互方式，提升用户输入信息的效率。其次，用户输入的质量并不稳定，不能假设用户在环路中的输入质量始终不变。疲劳、误判等各种因素均会导致输入数据中出现噪声。人在环路混合智能系统的交互设计需要考虑到用户的特点，通过良好的设计减少或修正噪声，提升用户输入信息的质量。此外，通过给用户反馈模型和训练的关键信息，可以帮助用户发现系统的不足，支持其对模型、样本和输入内容进行决策，进一步提升用户输入的效率。

通过交互设计可以提升人在环路混合智能系统的效用。首先，具有良好交互设计的系统由于具有更高的用户输入效率，可以使用更少的交互次数提升模型性能。其次，通过界面设计和交互设计，可以支持用户更容易地选择合适的模型，使得系统的可定制性变高。另外，交互设计通过对用户使用场景的考量，可以设计在不扰乱用户工作流的同时提升系统性能的交互方式，或利用用户原本与系统的交互来提取信息，从而提升系统在落地过程中的可用性。对于人在环路混合智能系统的交互设计，其研究工作主要分为交互界面和交互方式两方面，下面将分别进行阐述。

3.2 更加有效的交互界面

基于屏幕的交互界面作为用户向模型提供训练数据，并接收模型反馈，对模型进行监测的主要媒介，需要在人在环路混合智能的全生命周期中对用户和模型的需求提供支持。因此，如何设计更有效的创新交互界面，从而向系统提供高效的数据输入，向用户提供高质量的信息反馈是一个重要的设计问题。本文将分析人在环路混合智能生命周期中各个阶段的需求和挑战，然后从与各阶段相关的研究中整理归纳关键的界面构成。

3.2.1 人在环路混合智能的生命周期

Dudley 等人[2]分析了交互式机器学习的工作流，指出六个关键的流程活动（Workflow Activities）：模型选择、特征选择、模型调整、模型质量评估、终止训练评估和迁移。在这一过程中，用户通过与系统交互训练模型，训练完成后将模型部署到生产环境中使用。与交互式机器学习相比，Dellermann 等人[1]指出人在环路混合智能具有持续学习的特点，因此本文将其生命周期会拓展到生产环境中，加入持续优化的阶段。在这一阶段中，用户通过与系统交互来进行模型监测、排错和优化。

1）模型选择（Model Selection）。人工智能模型是人在环路混合智能系统的基础。人工智能模型本质上是一个函数，能够实现从样本到标签的映射。模型中有一些参数不确定，因此需要通过从已知的样本中寻找规律，确定未知参数。不同的模型有不同的特点，擅长解决不同的问题，选择合适的模型需要丰富的机器学习专业知识，因此常常由机器学习专家来完成。然而，针对同一个应用场景，可能有多种适合的模型，即使是同一个模型，由于参数的变化，也可能表现出较大差别，如何对此进行权衡，这对于机器学习专家来说常常都有些困难，更不必说普通终端用户。因此，如何帮助用户理解和对比模型，从而做出最适合应用场景的决策，是这一阶段重要的设计挑战。

2）特征选择（Feature Selection）。特征选择决定了模型会从什么样的输入中进行学习，因此对最终模型效果有极大的影响。应对一个复杂的任务，常常会有成千上万种可能的特征，如何构建或者选择与最终模型效果最相关的特征，并且去除冗余的特征，以提高模型学习的效率，是这一阶段的设计挑战。

3）模型调整（Model Steering）。模型调整是整个流程中的关键活动，用户通过迭代式的优化和训练使得模型向预期的方向演化[2]。这个过程可以由终端用户和机器学习专家共同参与，其中终端用户的主要交互行为常常包括修正错误标注的样本和添加新的样本，而机器学习专家则可以直接改变模型的参数，进行更直接的调整。从用户需求的视角来看，模型调整是其投入精力最多的阶段，因此需要良好交互设计来提升的效率，减少时间和精力的消耗。从模型需求的视角来看，由于用户的标注行为常常具有随机性，如何应对不一致的，甚至是错误的标注，避免模型向低效的方向偏移，是这一阶段的设计挑战。

4）质量评估和终止评估。在模型开发阶段，积累了一定的训练量之后，用户需要评估当前模型的质量，并将其与目标状态进行对比，判断是否可以终止训练，或是针对性地进行优化调整。由于模型质量与训练时间以及样本数量并非线性相关，过多的训练样本反而容易导致过拟合问题，因此，如何设置合适的评估策略，在恰当的时机终止训练，

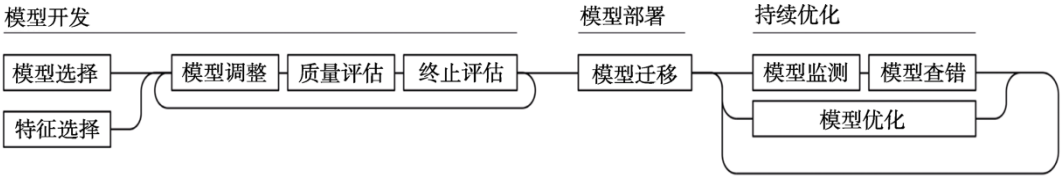


图 4　人在环路混合智能的生命周期
Fig.4 Life cycle of human-in-the-loop hybrid intelligence

是这一阶段的设计挑战。

5）持续优化。模型监测人在环路混合智能系统部署之后，随着终端用户在使用过程中加入新的训练数据，模型特征和预测结果的分布可能会发生偏移（Feature/Concept Drift）。在一些比较极端的应用场景中，一旦发生错误，可能会造成严重的后果，因此，如何辅助用户对模型的表现进行监控，提前发现特征偏移现象，是这一阶段的设计要点。

模型排错处在环路中的终端用户通常缺乏对于机器学习模型的专业知识，因此在模型出现错误时，他们常常不知道如何发现错误背后的问题。如何消除机器学习专业知识的壁垒，从终端用户熟悉的领域知识出发，提供易于理解的反馈信息来辅助排错，是这一阶段的设计挑战。

模型优化与模型开发阶段的模型调整不同，系统部署之后的模型优化活动不能影响用户的正常工作。基于模型特性而设计的反馈形式，常常会给用户增加额外负担，成为系统落地的阻碍。因此，模型优化活动需要自然地融入用户的工作流程之中，让用户付出较少的精力来获得更多的效果提升。

3.2.2　关键界面构成

支持人在环路混合智能的研究和应用呈现出了多样化的用户界面，由于数据类型、模型期望的用户输入方式、用户类型有所不同，很难提供一个固定的模板[2]。本文旨在归纳总结现阶段用户界面的共性，提取关键的界面构成元素，为研究人员和设计师提供参考。

1）模型评估。模型评估界面主要应用于模型选择、特征选择和质量评估阶段，支持用户评估和检查整体的模型质量，其评估指标不限于模型准确率，还可以合并如混淆矩阵等其他指标[2]。在人在环路混合智能生命周期的初期，模型评估界面的重要功能是支持用户对不同的模型和特征进行对比评估，见图 4。Amershi 等人[7]以图像搜索系统 CueFlik 为例，提供最近探索过的模型的历史数据可视化，支持终端用户参与多个模型的评估选择。Das 等人[36]以 BEAMES 可视分析系统为例，呈现了一种支持用户对多种学习算法和模型同时进行调整并互相比较的方法。Krause 等人[37]提供了一个可视化分析工具，帮助用户直观理解所有特征之间的排序关系，从而将自己的领域知识运用于特征选择的过程中。在对比评估的基础上，有部分研究探索如何构建混合的模型或者特征，来弥补单一模型或特征所存在的问题。这种情况下，模型或特征的可视化评估尤为重要。Justin 等人[38]提出了多种模型融合的方法，利用 EnsembleMatrix 系统，用户能够与模型可视化界面进行直接交互，将基于不同特征训练的分类模型结合起来，以获得超过单一模型的效果。Brooks 等人[35]通过可视化地检验错误来支持生成特征，并通过基于文字的特征选择来辅助文本分类。在质量评估阶段，仅仅依靠准确率指标是不够的，模型评估界面需要呈现多维度的指标，帮助用户全面了解模型的性能状态。例如 ModelTracker[39]引入了一个适用于机器学习的性能分析和调试的通用交互式可视化工具，通过显示模型的诸如混淆矩阵、准确率-精确度曲线等通用指标简洁直观表示模型性能。

2）样本浏览。样本浏览界面常常应用于质量评估阶段，作为模型评估界面的补充，为用户提供浏览原始样本及模型输出样本的接口，帮助用户更加直观地理解模型的表现，也有助于从主观的角度了解模型表现的可预测程度[40]。Amershi 等人开发的 CueFlik 系统[7]又将这两部分样本根据标记种类来分别呈现，帮助用户全面地了解不同标记与模型结果之间的关系。样本浏览界面有时也会用在模型和特征选择阶段，帮助提供上下文情景。例如 Brooks 等[35]开发的 FeatureInsight 网页分类系统，能够让用户预览可能导致分类器错误和正确的关键词及对应网页，帮助用户提供更为准确的标记。Cheng 和 Bernstein[41]开发的 Flock 系统通过在界面上呈现互相冲突的样本，提示用户思考有区分度并且可解释性高的特征，展示了利用众包来辅助构建新颖特征的可能性。在此基础上，模型输出的样本还常常与置信度等指标数据，以及热力图等可视化形式结合呈现，用于增强模型的可解释性，帮助用户进行排错。热力图能够直观表现神经网络分类起在每个像素上的不确定性。Nalisnik[32]在交互界面提供了热力图可视化，有助于用户快速发现分类器不确定的决策的区域。热力图也能直观地表现用户标注的识别对象的显著性。ClickBAIT[27]在界面右侧用热力图显示当前识别结果和训练页面，有助于

帮助用户快速判断模型是否识别正确，以及明确交互操作的可点击范围。类似的，XAI-CBIR[26]系统在图像中检索视频帧后，会显示检索结果的显著图反应与目标对象的相似性，同时提供样本的置信度，用百分比的形式帮助用户判断并反馈该图像的相关性。Avelka 等人[34]对模型判定结果进行可视化，帮助用户理解模型，从而提供更为详细准确的标注。Lü 等人[42]则在一项关于手势交互的研究中提供了样本浏览相关的建议，即对模型输出样本过度的抽象和压缩不利于用户探究导致模型表现下降的原因。

3）反馈提交。反馈提交界面主要应用于模型调整和持续优化阶段，为用户提供了标注、特征选择、新样本生成的渠道。作为整个交互界面最关键的部分，反馈提交界面的设计决定了用户的交互方式和内容，从而直接影响模型调整和优化的效率。基于对场景特征和用户反馈意图的分析，一些研究设计了直觉化而又丰富的反馈提交形式，帮助用户快速表达的同时，又能提供模型所需的丰富信息。GrabCut[11]提供不同颜色的画笔，使用户可以区分前景和背景。Lin 等人[10]开发的基于浏览器的图像扣像工具则允许用户根据需要改变画笔的宽度。ClickBAIT[27]允许用户在界面中实时浏览视频帧，点选视频中的对象，并将这些对象转换为视频帧的标注，用于训练卷积神经网络模型。在文本分类领域中，有工作提出一个交互式机器学习方法用于法律解释，在专家将法规的一小部分标注为相关或不相关后，系统能够就其余法规的相关性提供合理的建议，随着人类专家标注越来越多的法规，系统提供的建议将变得更加准确[34]。在音频源分离的工作中，Durrieu[43]允许用户拖动鼠标校正基音频率，点击模型更新按钮调整深度学习模型参数。Kim 等人[21]开发的音频处理系统将音轨分为与初始示例相同的小区域，在界面中将与示例相似的特征显示为候选对象引起用户的注意，方便用户选择和标注。由于用户提交的反馈常常具有不一致性，有很多研究从不同的角度尝试应对这一挑战。部分研究试图提高系统的容错率，例如 SSE[22]中的交互界面允许用户在混合声音的时间频率或频谱图的可视化上绘制标注，使用不同的颜色表示不同声源，使用不透明度度量模型给出的置信度，从交互形式上允许一定程度上不准确的标注。Intelligent Scissors[9]通过边缘检测算法来实时检测光标周围可能存在的边界，控制用户光标附着在所在位置距离最近的边界上，用户仅需要找出大概的边界即可进行标注。部分研究尝试从用户侧减少不一致现象的发生，例如 Kulesza 等人[44]指出了训练过程中概念演化（Concept Evolution）导致用户提供的标注前后不一致的问题，并提出了两种结构化标注的解决方案，帮助用户更加一致地定义和完善需要模型学习的概念。Zhang[45]则赋予模型怀疑的能力，一旦发现用户提供的标注与模型已经习得的概念不一致，则会询问和质疑用户，让用户进行确认。

3.3 更加自然的交互方式

交互方式包括但不限于图形界面交互，命令行交互，语音交互，多模态，智能交互，自适应交互，可触摸，无接触等[46]。表 1 列举了典型应用中的交互方式，可以看出大多数的交互方式集中在基于图形界面的鼠标操控。除此之外，针对一些特殊的场景和用户，有工作探索使用眼动仪或者手势等交互方式实现用户接口。Stember 等人[31]探索了一种基于眼动追踪的交互式分割方法，将专家的眼动仪数据输入至分割神经网络中，得到的分割结果与手动标注相当，眼动追踪使放射科医生能够更加自然地与系统进行交互，其分割结果也可以在临床工作期间被记录下来。Katan 等人[47]将“交互式机器学习”应用于手势控制工具的创建，以减轻残疾人在弹奏传统乐器时面临的挑战，使用户能够建立手势词汇表，并自定义手势和声音之间的关系。Sarasua 等人[48]展示了定制手势识别控制音乐清晰度的潜力。

Norman 等人[49]认为人具有天然的擅长或不擅长处理的事物，人和系统交互发生错误时，很多时候可以归结为系统要求人做其不擅长处理的事务。因此在人在环路混合智能系统的交互设计中需要根据不同的任务目标和用户群体，选择合适且自然的交互方式，通过良好的设计减少或修正噪声，提升用户输入信息的质量，加速模型的优化过程。

3.4 人在环路混合智能系统的设计建议

上文分别对人在环路混合智能系统的交互界面、交互方式两个方面进行了梳理，并对当前相关研究进行了综述。这里将从上述研究和分析中总结相关的设计原则，以支持人在环路混合智能系统的交互设计。

3.4.1 根据模型需求特点选择合适的交互方式

根据对既有研究中数据类型和交互方式的分析，可以发现不同数据类型具有不同的对应适合输入方式。典型人在环路混合智能应用见表 1，列举了不同数据类型的典型应用中交互方式，不难看出，以图像及视频为输入的应用中，用户多以鼠标点击、画线或者框选的方式直接与图像进行交互；以文本为输入的应用则多以界面控件为交互媒介；音频为输入的应用的交互方式则根据具体任务目标

表 1　典型人在环路混合智能应用
Tab.1 Typical application of human-in-the-loop hybrid intelligence

方法	应用场景	用户交互方式					
		点击	画线	边界框	光标移动	界面控件	非鼠标交互
Intelligent scissors[9]	图像抠图				√		
GrabCut[11]	图像抠图		√	√			
Web-Based Image Matting[10]	图像抠图		√				
Deep GrabCut[12]	图像抠图			√			
Jang 和 Kim[50]	图像分割	√					
Stember 等[31]	医疗图像分割						√
Polygon-RNN[14]	语义分割、实例分割	√		√			
Polygon-RNN++[15]	语义分割、实例分割	√		√			
Curve-GCN[13]	语义分割、实例分割	√		√			
Nalisnik 等[32]	图像分类					√	
Avelka 等[34]	文本分类、法律解释	√			√		
Abstrackr[33]	文本分类	√			√		
FeatureInsight[35]	文本分类	√			√		
Endert 等[20]	文本分析	√			√		
Kim 和 Pardo[21]	音频标注	√			√	√	
ISSE[22]	音频源分离	√	√			√	
Durrieu 和 Thiran[43]	音频源分离	√				√	
Nakano 等[23]	音频源分离	√	√		√		
Katan 等[47]	音乐创作、音乐演奏						√
Sarasua 等[48]	音乐创作、音乐演奏						√
XAI-CBIR[26]	视频图像检索	√				√	
ClickBAIT[27]	视频标注	√					
Feng 等[28]	视频监视车辆检测	√					

的不同有所不同。根据模型需求的数据特点来选择合适的交互方式，能够提升用户输入及交互效率，从而加速模型的优化过程。

3.4.2　针对不同场景的系统需求特征调整界面构成元素的具体设计

在界面设计方面，Dudley 等人[39]提出了如将任务目标和限制进行清晰呈现、支持用户理解模型的不确定性和置信度等六条设计原则，这些原则依然适用于人在环路混合智能系统的界面设计。然而，并非所有符合以上原则的设计均可以满足场景中的系统需求。例如，领域专家自行进行模型训练和排错的场景下，对系统可用性要求很高。此时，不确定性和置信度的具体呈现方式、是否有易懂的教程引导、是否有完善的错误恢复流程，对这类场景而言非常重要。因此，人在环路混合智能系统的交互设计需要针对不同场景的系统需求特征来调整界面构成元素的具体设计。

3.4.3　针对不同场景的系统需求特征调整界面构成元素的具体设计

在模型部署之后，系统优化和用户运用模型同时进行，意味着终端用户需要在运用模型的过程中即时生成对系统的反馈。在生产环境下，用户往往在交互式模型调整中无法投入较多的时间和精力，因此设计这一阶段的交互时需要兼顾系统的需求和用户的限制，在深度了解用户工作流程的基础上，设计简明高效的反馈交互形式

4　结语

人工智能技术的发展及其在各领域的应用，使得越来越多的设计师参与到实践和研究中，尝试运用以用户为中心的设计方法来应对挑战，并总结出了一些经验[51-52]。然而，现有研究的对象多为广义的人工智能系统，致力于提出通用的设计建议，对于系统的讨论较少，关于如何进行其中的交互设计的研究处于初期。

本文指出了“支持人在环路混合智能的交互设计”这一类设计问题，强调交互设计是实现人在环路混合智能系统落地所必须的重要环节。基于对文本、图像、视频、音频等应用领域的相关文献整理，总结了不同领域中常见的交互界面和交互方式，为

设计师提供索引和参考。然后总结了人在环路混合智能的生命周期和关键界面构成，解析其中的需求和设计考量。最后针对交互方式和界面构成提出设计建议，为具体的设计提供参考。

在这样一类设计问题中，设计师需要应对的设计挑战还有很多，这些挑战一部分来自于设计师对技术的理解，例如如何理解具有高复杂性的人工智能系统的技术能力和限制，如何创造出新颖并且技术可行性高的场景，以及如何了解算法需要的数据类型，并通过界面有效传达算法的需求。另一部分挑战来自于设计师对用户的理解，例如如何在全新的应用场景中做好对用户的引导，帮助用户建立合适的混合智能系统认知模型，如何为用户行为的不确定性和不一致性而设计，如何进行快速迭代的原型开发和用户测试，以及如何对界面和用户的表现进行评估等。为了应对这些挑战，需要从智能系统、用户、设计师三方面进行系统性思考，逐步建立设计方法论，完善设计工具，从而更有效地支持和推动人在环路混合智能系统的应用。

参考文献：

[1] DELLERMANN D, CALMA A, LIPUSCH N, et al. The Future of Human-AI Collaboration: a Taxonomy of Design Knowledge for Hybrid Intelligence Systems[C]. 2019.

[2] DUDLEY J J, KRISTENSSON P O. A Review of User Interface Design for Interactive Machine Learning[J]. ACM Transactions on Interactive Intelligent Systems, 2018, 8(2): 1-37.

[3] HOLZINGER A. Interactive Machine Learning for Health Informatics: When Do We Need the Human-in-the-loop?[J]. Brain Informatics, 2016, 3(2): 119-131.

[4] ABRAHAM A, NATH B. Hybrid Intelligent Systems Design: a Review of a Decade of Research[J]. CiteSeerX, 2000, 37.

[5] ZHENG N, LIU Z, REN P, et al. Hybrid-augmented Intelligence: Collaboration and Cognition[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(2): 153–179.

[6] PAN Y. Heading toward Artificial Intelligence 2.0[J]. Engineering, 2016, 2(4): 409-413.

[7] AMERSHI S. Designing for Effective End-user Interaction with Machine Learning[C]. Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology-UIST'11 Adjunct, 2011.

[8] ZHANG X, WANG L, XIE J, et al. Human-in-the-loop Image Segmentation and Annotation[J]. Science China Information Sciences, 2020, 63(11): 219101.

[9] MORTENSEN E N, BARRETT W A. Intelligent Scissors for Image Composition[C]. Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH'95, 1995.

[10] LIN Y C, WANG H A, HSIEH Y F. Image Matting through a Web Browser[J]. Multimedia Tools and Applications, 2012, 61(3): 551-570.

[11] ROTHER C, KOLMOGOROV V, BLAKE A. "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts[J]. Acm Trans Graph, 2004, (6).

[12] XU N, PRICE B, COHEN S, et al. Deep GrabCut for Object Selection[J]. ArXiv, 2017.

[13] LING H, GAO J, KAR A, et al. Fast Interactive Object Annotation with Curve-GCN[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[14] CASTREJON L, KUNDU K, URTASUN R, et al. Annotating Object Instances with a Polygon-RNN[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[15] ACUNA D, LING H, KAR A, et al. Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++[J]. arXiv, 2018.

[16] JANG W D, KIM C S. Interactive Image Segmentation Via Backpropagating Refinement Scheme[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019: 5292-5301.

[17] BUDD S, ROBINSON E C, KAINZ B. A Survey on Active Learning and Human-in-the-Loop Deep Learning for Medical Image Analysis[J]. ArXiv, 2019.

[18] WALLACE B C, SMALL K, BRODLEY C E, et al. Deploying an Interactive Machine Learning System in an Evidence-based Practice Center: Abstrackr[J].

[19] ABERDEEN D, PACOVSKY O, SLATER A. The Learning Behind Gmail Priority Inbox[J].

[20] ENDERT A, FIAUX P, NORTH C. Semantic Interaction for Visual Text Analytics[C]. Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems-CHI'12, 2012.

[21] KIM B, PARDO B. A Human-in-the-Loop System for Sound Event Detection and Annotation[J]. ACM Transactions on Interactive Intelligent Systems, 2018, 8(2): 1-23.

[22] BRYAN N J, MYSORE G J, WANG G. ISSE: an Interactive Source Separation Editor[C]. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems-CHI'14, 2014.

[23] NAKANO T, KOYAMA Y, HAMASAKI M, et al. Interactive Deep Singing-voice Separation Based on Human-in-the-loop Adaptation[C]. Proceedings of the 25th International Conference on Intelligent User Interfaces, 2020.

[24] HIPKE K, TOOMIM M, FIEBRINK R, et al. BeatBox: End-user Interactive Definition and Training of Recognizers for Percussive Vocalizations[C]. Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces-AVI'14, 2014.

[25] Real-Time Human Interaction with Supervised Learning Algorithms for Music Composition and Performance (Thesis) | Computer Science Department at Princeton

University[EB/OL]. [2020-07-20]. https://www.cs.princeton.edu/research/techreps/TR-891-10.

[26] CHITTAJALLU D R, DONG B, TUNISON P, et al. XAI-CBIR: Explainable AI System for Content Based Retrieval of Video Frames from Minimally Invasive Surgery Videos[C]. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019.

[27] TENG E, FALCÃO J D, HUANG R, et al. ClickBAIT: Click-based Accelerated Incremental Training of Convolutional Neural Networks[C]. 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2018.

[28] FENG R C, LIN D T, CHEN K M, et al. Improving Deep Learning by Incorporating Semi-automatic Moving Object Annotation and Filtering for Vision-based Vehicle Detection[C]. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019.

[29] KABRA M, ROBIE A A, RIVERA A M, et al. JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behavior[J]. Nature Methods, 2013, 10(1): 64-67.

[30] ZHANG X, WANG L, XIE J, et al. Human-in-the-loop Image Segmentation and Annotation[J]. Science China Information Sciences, 2020, 63(11): 219101.

[31] STEMBER J N, CELIK H, KRUPINSKI E, et al. Eye Tracking for Deep Learning Segmentation Using Convolutional Neural Networks[J]. Journal of Digital Imaging, 2019, 32(4): 597-604.

[32] NALISNIK M, GUTMAN D A, KONG J, et al. An Interactive Learning Framework for Scalable Classification of Pathology Images[C]. 2015 IEEE International Conference on Big Data (Big Data), 2015.

[33] WALLACE B C, SMALL K, BRODLEY C E, et al. Deploying an Interactive Machine Learning System in an Evidence-based Practice Center: Abstrackr[J].

[34] AVELKA J S, TRIVEDI G, ASHLEY K D. Applying an Interactive Machine Learning Approach to Statutory Analysis[J].

[35] BROOKS M, AMERSHI S, LEE B, et al. FeatureInsight: Visual Support for Error-driven Feature Ideation in Text Classification[C]. 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), 2015.

[36] DAS S, CASHMAN D, CHANG R, et al. BEAMES: Interactive Multimodel Steering, Selection, and Inspection for Regression Tasks[J]. IEEE Computer Graphics and Applications, 2019, 39(5): 20-32.

[37] KRAUSE J, PERER A, BERTINI E. INFUSE: Interactive Feature Selection for Predictive Modeling of High Dimensional Data[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1614–1623.

[38] Ensemble Matrix: Interactive Visualization to Support Machine Learning with Multiple Classifiers -Microsoft Research[EB/OL]. [2020-07-21]. https://www.microsoft.com/en-us/research/publication/ensemblematrix-interactive-visualization-support-machine-learning-multiple-classifiers/.

[39] AMERSHI S, CHICKERING M, DRUCKER S M, et al. ModelTracker: Redesigning Performance Analysis Tools for Machine Learning[C]. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems-CHI'15, 2015.

[40] FIEBRINK R, COOK P R, TRUEMAN D. Human Model Evaluation in Interactive Supervised Learning[C]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011.

[41] CHENG J, BERNSTEIN M S. Flock: Hybrid Crowd-Machine Learning Classifiers[C]. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015.

[42] LÜ H, FOGARTY J A, LI Y. Gesture Script: Recognizing Gestures and Their Structure Using Rendering Scripts and Interactively Trained Parts[C]. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems-CHI'14, 2014.

[43] DURRIEU J L, THIRAN J P. Musical Audio Source Separation Based on User-Selected F0 Track[C]. Latent Variable Analysis and Signal Separation, 2012.

[44] KULESZA T, AMERSHI S, CARUANA R, et al. Structured Labeling for Facilitating Concept Evolution in Machine Learning[C]. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems-CHI'14, 2014.

[45] Dealing with Mislabeling via Interactive Machine Learning|SpringerLink[EB/OL]. [2020-08-23].https://link.springer.com/article/10.1007/s13218-020-00630-5.

[46] PREECE J. Interaction Design: Beyond Human–Computer Interaction[J].

[47] KATAN S, GRIERSON M, FIEBRINK R. Using Interactive Machine Learning to Support Interface Development Through Workshops with Disabled People[C]. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems-CHI'15, 2015.

[48] SARASUA A, CARAMIAUX B, TANAKA A. Machine Learning of Personal Gesture Variation in Music Conducting[C]. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016.

[49] NORMAN D A, STAPPERS P J. DesignX: Complex Sociotechnical Systems[J]. She Ji (The Journal of Design, Economics, and Innovation), 2015, 1(2): 83–106.

[50] JANG W D, KIM C S. Interactive Image Segmentation Via Backpropagating Refinement Scheme[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019: 5292-5301.

[51] AMERSHI S, INKPEN K, TEEVAN J, et al. Guidelines for Human-AI Interaction[C]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems-CHI'19, 2019.

[52] YANG Q, STEINFELD A, ROSÉ C, et al. Re-examining Whether, Why, and How Human-AI Interaction is Uniquely Difficult to Design[C]. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020.