机器学习在以人为本方向的发展

机器学习是当代计算机科学中最重要也是最成功的技术之一。虽然它可以应用于许多人类感兴趣的问题，但机器学习的研究往往仅仅将算法应用于模型中，是以一种客观的方式进行的。然而，它隐藏了人在调整算法、收集数据、建模过程，以及利用机器学习的经验等方面所做的工作。

从以人为本的角度审视机器学习，包括明确地认识人的工作，根据人的工作经验来重构机器学习工作流程，以及探索人与智能系统互相适应方法。以人为本的机器学习不仅会带来更有用的机器学习工具，而且会带来新的理解机器学习的方法。

近年来，不同的应用领域和都已经探索了使用机器学习的环境，并展示了在人类环境中构建机器学习的不同方法。

利用机器学习研究社会科学

社会科学通常处理复杂的、大型的数据集，而机器学习具有利用这些数据来加快研究过程的潜力。然而，机器学习和社会科学对数据分析的理解不同。社会科学经常采用理论和假设驱动的方法来处理数据，而机器学习通常处理收集和分析的数据，而不考虑理论框架。

Chen把机器学习的编码过程描述为一种经过分离、分组和链接，并将数据按系统顺序排列，以便表达和解释的过程。并将模糊性识别作为编码过程中的一个关键挑战，从而来避免程序员在编码过程中对不确定性的问题产生疑惑，或者在不同程序员之间产生分歧的问题。

视觉决策中的用户信息预测

人类不确定性的信息有可能影响机器学习的结果，而且研究者很难获得关于不确定性的信息。虽然可以问被测人员自身有多自信，但信心的判断本身可能就是一项艰巨的任务，会需要大量的时间与精力。

Smith等人通过使用机器学习来估计用户在可视化决策任务上的信心，解决了训练数据提供者的不确定行为导致的训练数据不准确或不一致的问题。研究人员没有明确询问用户的信心，而是使用在用户提供数据时收集到的隐含信息，特别是用户在执行任务时的眼睛注视模式的信息。研究人员通过利用卷积神经网络，提出了一种关于眼睛注视模式与时间的二维表示方法。这些数据适合学习常用于图像数据的技术，可以成为以人为中心的机器学习工作流程的一个组成部分。

通过群体注释者标注医疗数据

Dumitrache等人通过研究使用群体注释者来标注医疗文本。这项工作的目的是确定预料中医疗术语之间的关系，如“发热导致头晕”这句话表明了“发热”和“眩晕”之间的关系是“原因”。这种类型的关系抽取是一个关键问题，就像许多机器学习任务一样需要大量的标注训练数据。如果让医疗专家对这些数据进行标注，可能需要大量的精力。

因此，利用群体注释者进行标注来对机器学习系统进行训练是一个很有吸引力的选择。但语料中的许多句子是含糊不清的，研究者利用模糊性原理，对群体注释者的标注进行加权。然后，作者将经过加权、群体注释者标注的数据与专家提供的数据进行了比较。结果表明，群体标记数据方法与专家的结果更具有竞争性。

多发性硬化症患者运动能力的测量

基于计算机的传感系统和机器学习可以改善人类对运动能力的评估，然而，要使这种系统在临床实践中发挥作用，就需要对应算法提供比简单机械的评估更多的信息。临床医生在复杂的决策过程中必须将算法评估的信息与他们自己的知识结合起来，并与他们的同事进协作评估。

在将人机交互方法应用到机器学习的系统设计中，通过与临床医生经过一系列的设计迭代，通过可视化来帮助机器学习系统来进行多发性硬化症评估的解释工作，从而来更好的了解这一领域的挑战以及用户的需求。另一方便，通过可视化也可以来表示算法决策的具体过程，包括关于数据质量的信息以及可由临床医生解释的临床信息。

通过交互进行手势-声音映射

Francoise将机器学习应用于手势交互设计，来寻找比基于传统鼠标键盘或触摸屏更丰富的新的交互形式。他们把手势看作是控制声音合成的一种手段，基于人类动作以及对应的声音来定义复杂的非线性映射。

传统人机交互使用的快速原型方法不太适合基于移动交互的挑战，这是由于移动交互专注于界面设计，而且其重点是人的动作而不是视觉显示。另外，手势交互式一种具体的技能，我们只知道如何通过动作来做手势，却不能明确地表示出具体的象征。这使得我们很难清晰的通过代码来定义手势。研究者通过开发一种通过交互进行手势与声音映射的方法，并将其用于训练机器学习模型，最终通过混合评估方法进行定性研究，使得用户可以通过手势与系统进行互动，使得手势成为了系统的一部分。

正如我们看到的，以人为本的机器学习不是一种单一的方法，而是各种各样的问题、方法、技术和理论的集合，这些都可以并且应该在今后继续进行探索。未来几年，随着机器学习从实验室转移到现实社会中，如何确保这些系统对人类更加有用的问题变得更加迫切。以上不同方面的发展给出了如何使机器学习的研究更加以人为本，在新的机器学习技术的开发与应用中更加突出人的目标和经验。