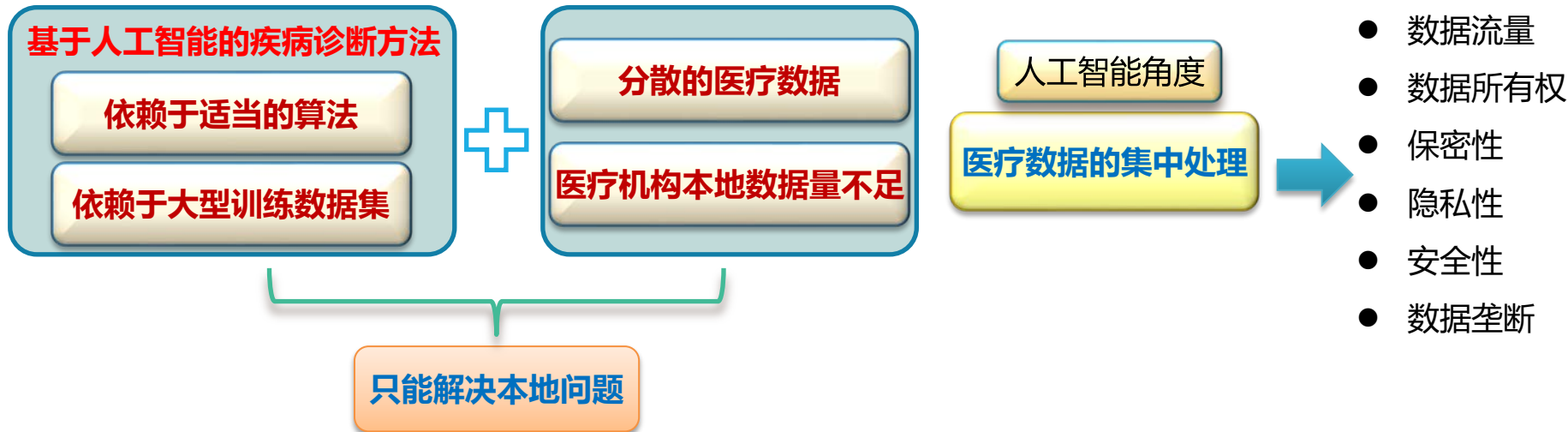


集群学习 (Swarm Learning)

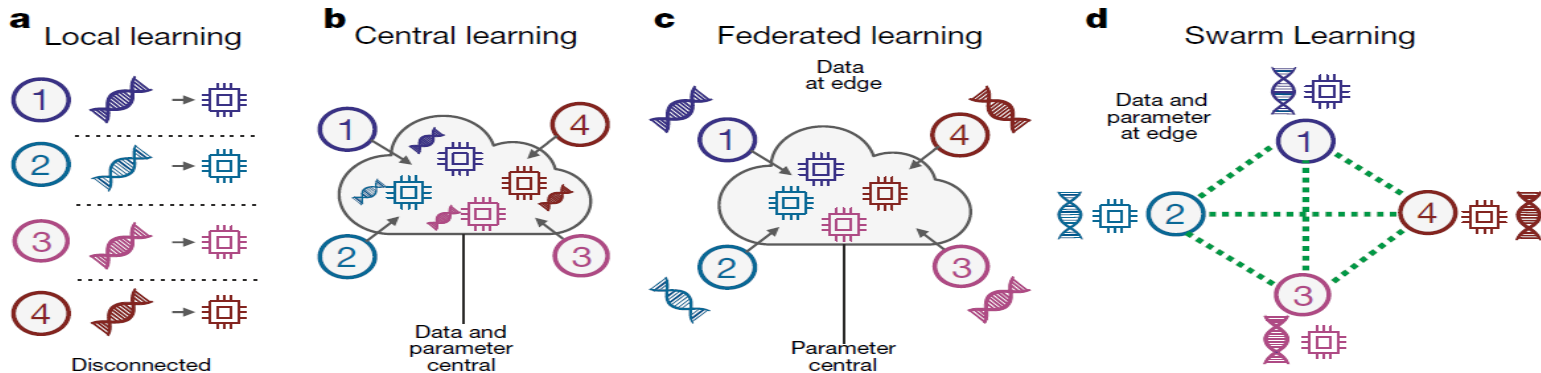
精准医疗的目标是能快速准确地检测出患有严重疾病和异质性疾病的患者，而机器学习有助于实现这一目标，例如根据病人的血液转录组数据来识别是否患有白血病，然而，应用到实际还存在很多问题。

集群学习是一种去中心化的机器学习方法，将边缘计算和基于区块链的对等网络结合起来，用于不同医疗机构之间医疗数据的整合。



集群学习 (Swarm Learning)

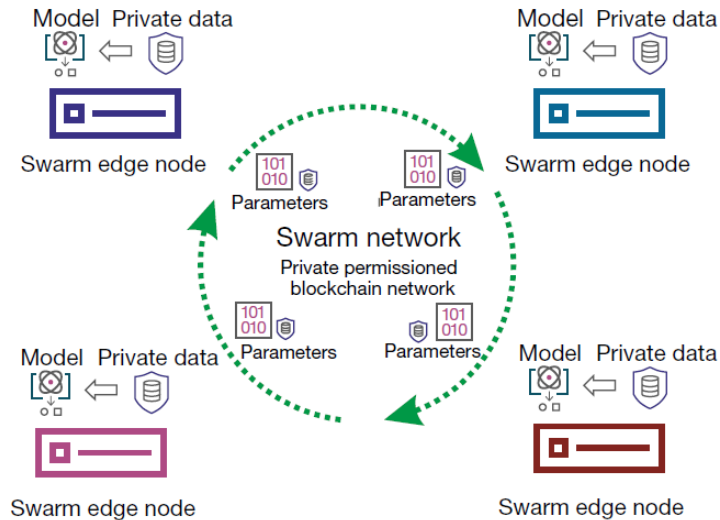
- Local learning: 数据和计算是在不同的、不相关的地点
- Central learning: 数据和参数是中心化的, 是基于云的机器学习, 存在数据重复、数据流量增加以及数据隐私、数据安全等方面的问题。
- Federated learning: 数据被保存在数据所有者当地, 计算在数据所有者当地的存储中心进行, 但参数设置由中央协调员协调, 使用专用参数服务器负责聚合和分发, 其他中央结构仍被保留。
- Swarm learning: 数据和参数在边缘 (边缘计算), 不需要中央协调员, 省去专用服务器, 通过Swarm网络共享参数, 并且在各个节点的私有数据上独立构建模型。



集群学习 (Swarm Learning)

集群学习的优势:

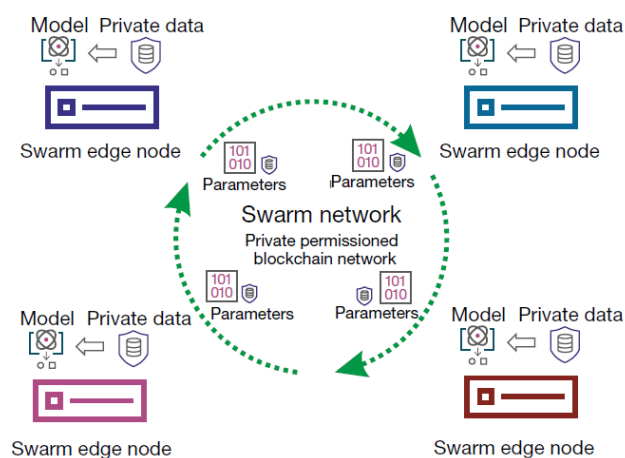
- 将大量医疗数据保存至数据所有者本地;
- 不需要交换原始数据, 从而减少数据流量;
- 提供高级别的数据安全保障;
- 能够保证网络中成员的安全、透明和公平加入, 不再需要中央托管员;
- 允许参数合并, 实现所有成员权力均等;
- 可以保护机器学习模型免受攻击。



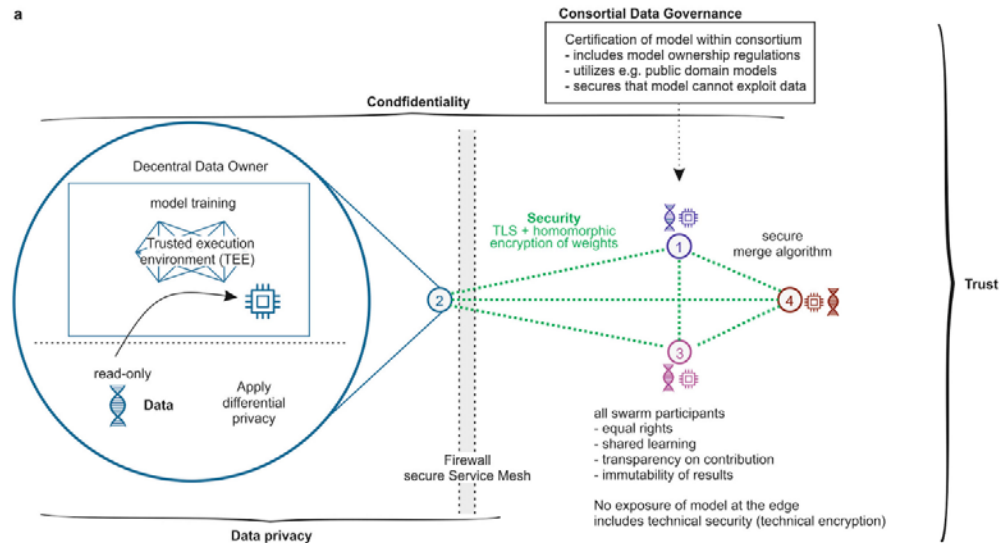
Swarm network

集群学习 (Swarm Learning)

集群学习是通过Swarm网络共享参数，并在各个节点的私人数据上独立构建模型。集群学习提供安全措施以支持数据主权、安全和保密性（这由私人许可的区块链技术实现）。



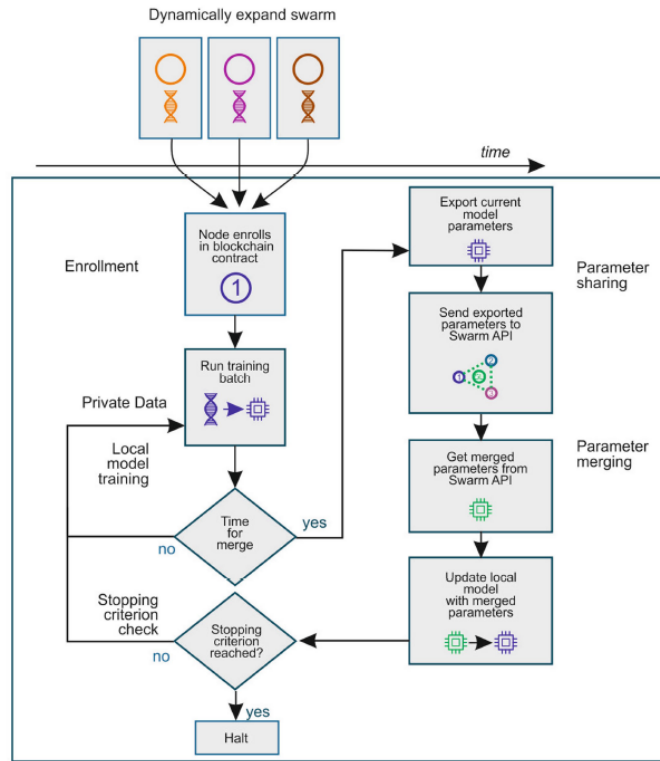
Swarm network



集群学习 (Swarm Learning)

集群学习的工作流程:

- 新节点动态加入
- 通过区块链智能合约进行注册
- 获得模型，进行本地的模型训练，直到满足定义的同步条件
- 将新节点参数输出到Swarm API中
- 通过Swarm API合并模型参数
- 在开始新一轮训练之前，使用合并后的参数更新模型



Enrollment, parameter sharing and merging within the swarm network

工作流程

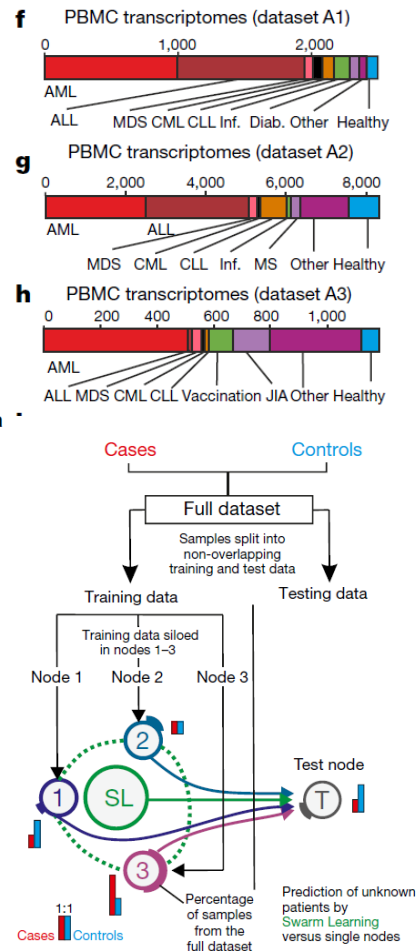
集群学习 (Swarm Learning)

案例： 集群学习预测白血病

研究使用来自12000多个患者的外周血单核细胞（PBMC）转录组数据，使用连续的深度神经网络（DNN）算法。

训练数据集以不同的分布方式被 “隔离 ”在每个Swarm节点上，模拟临床上的相关场景。每个节点可以代表一个产生医疗数据的医疗中心、一个国家或任何其他独立的组织。测试节点T使用独立的测试数据集。

使用来自急性骨髓性白血病（AML）患者的样本作为实例（cases），所有其他样本作为对照（controls）



集群学习 (Swarm Learning)

案例：集群学习预测白血病

在四种场景下进行研究，发现，在所有的情况下（改变节点的样本分布均匀度、cases/controls比例等），集群学习的预测性能均优于单个节点的预测性能，并且这种表现与数据收集方式或数据生成技术无关。

