文章整理模板

文章标题：字体 微软雅黑 3号字

文章正文：字体 微软雅黑 字体大小：11

**整体要求：将老师口语化的表述，整理成连贯性的技术文章，文章逻辑结构清晰，并配上相应ppt截图。**

**文章结构：**

导读：（根据老师开场白整理）XXXX，本次分享题目为XX，主要介绍：

* XX1
* XX2
* XX3
* XX4

**▌XX1（1级标题）**

1. **（2级标题）**

**① （3/4级标题）**

* **（3/4级标题）**

**正文XXXXXXX（每段开头顶格写）**

**案例：**

**导读：**随着大数据的进一步发展，重视数据隐私和安全已经成为了世界性的趋势， 同时，大多数行业数据呈现数据孤岛现象，如何在满足用户隐私保护、数据安全和政府法规的前提下，进行跨组织的数据合作是困扰人工智能从业者的一大难题。而“联邦学习”将成为解决这一行业性难题的关键技术。今天会和大家分享下微众银行主导的新一代联邦学习技术及应用，而 FATE 则是联邦学习落地的一个工业级开源平台。

今天的介绍会围绕下面五点展开：

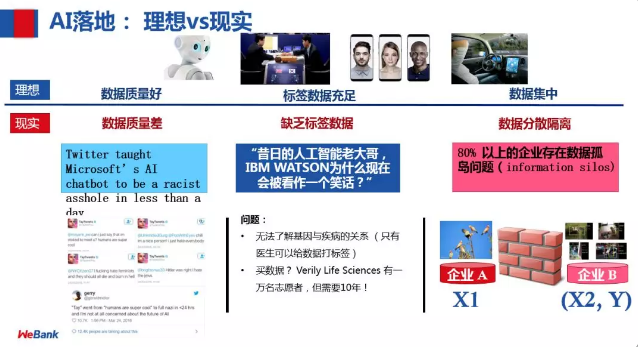
* 联邦学习背景介绍
* 纵向联邦学习
* 横向联邦学习
* 应用案例

FATE：联邦学习开源平台

**▌联邦学习背景介绍**

首先和大家分享下联邦学习的背景。

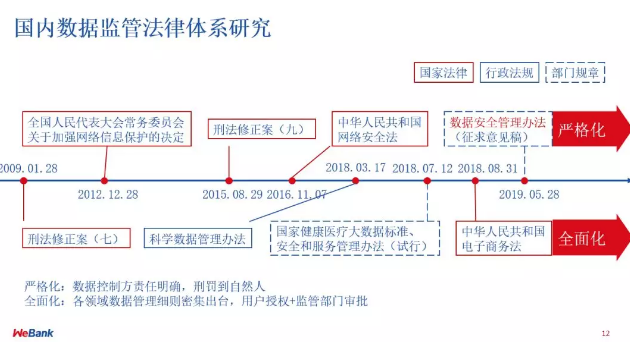
1. **AI 落地的理想与现实**



我们知道 AI 落地的时候，会遇到很多现实的问题：

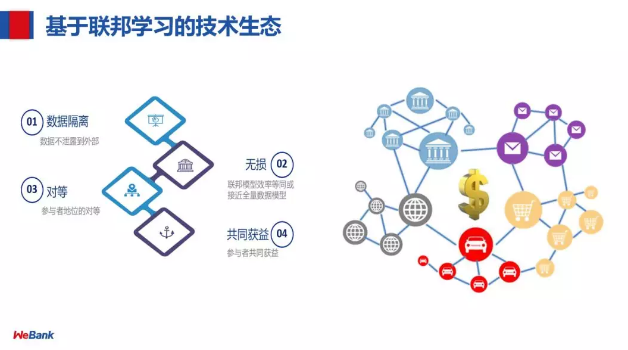
* 现实中，我们的数据质量是非常差的，比如聊天数据中有很多噪音；
* 数据标签，收集是比较困难的，很多场景中的数据是没有标签的；
* 数据是分散的，（这也是最重要的一点）每家应用的数据不一样，比如腾讯用的是社交属性数据，阿里用的是电商交易数据，微众用的是信用数据，都是分散来应用的。现实中，如何进行跨组织间的数据合作，会有很大的挑战。

1. **国内数据监管法律体系研究**



从09年到现在的10年时间内，国家关于数据的法律条例是趋向于严格化的，同时趋向于全面化，每个细分领域都纷纷出台了相应的条例和条款。相对来讲，让之前可行的一些数据合作方案变得不太可行。

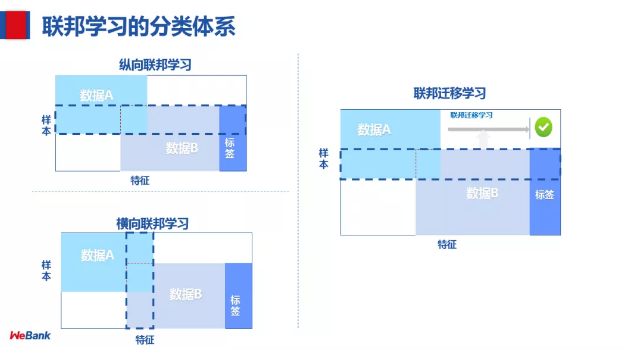
**3. 基于联邦学习的技术生态**



针对上述问题，微众银行提出了基于联邦学习的技术生态，特点如下：

* 数据隔离：联邦学习的整套机制在合作过程中，数据不会传递到外部。
* 无损：通过联邦学习分散建模的效果和把数据合在一起建模的效果对比，几乎是无损的。
* 对等：合作过程中，合作双方是对等的，不存在一方主导另外一方。
* 共同获益：无论数据源方，还是数据应用方，都能获取相应的价值。

**4. 联邦学习的分类体系**



联邦学习的分类体系，包括：

* 纵向联邦学习，两个数据集的用户 ( U1, U2, … ) 重叠部分较大，而用户特征 ( X1, X2, … ) 重叠部分较小；
* 横向联邦学习，两个数据集的用户特征 ( X1, X2, … ) 重叠部分较大，而用户 ( U1, U2, … ) 重叠部分较小；
* 联邦迁移学习，通过联邦学习和迁移学习，解决两个数据集的用户 ( U1, U2, … ) 与用户特征重叠 ( X1, X2, … ) 部分都比较小的问题。

下面，重点分享下纵向联邦学习和横向联邦学习。

**▌纵向联邦学习**

**1. 联合建模需求场景**



举个例子：微众与合作企业进行联合建模，比如做信贷逾期模型，微众有 Y 数据，包括标签数据，逾期记录，用这样的数据可能会建一个很好的模型，但我们希望用更多的数据，比如合作方的标签数据和画像数据来更大的提升风控模型的效果和稳定性。

传统模式的问题是：

合作企业缺乏 Y 无法独立建立模型，需要微众把 Y 数据，带入到合作方的生产环境建模，但是由于国家的数据保护条款和各企业自身对数据的严格规定，得到的 X 数据不能全量的传输到微众。

针对这个问题，通过纵向联邦学习来解决。如右图，两边的数据都有共同的 ID，特征是完全不一样的，可以通过一方特征来弥补另一方特征的不足