

样本标注指南

背景：

我们的目标是更有效地衡量网络仇恨言论探测系统的准确性。为此，我们创造了一个数千短句的数据集。每一个样本都反映网络仇恨言论的一个方面。

你的任务：

1. 将每一个分配与你的样本标注为仇恨言论或非仇恨言论。
2. 如果有样本在你看来不合实际，也将它们标注出来。

什么是“仇恨言论”？

我们对于“仇恨言论”的定义是 **针对一个受保护群体或其中个体的，他们的这一身份的，谩骂**。受保护群体的定义可以基于年龄、身体残障、种族（肤色、民族、国籍、出生地）、宗教信仰、性别、性取向以及性别认同。

什么是“不合实际”的？

我们让你来决定不合实际的具体定义。我们未有意在数据集中加入不合实际的句子；这一选项仅用于确保数据的质量。

标注样本时的注意事项：

- 一些常用于仇恨言论的词汇也可能被用于非仇恨言论。（例如，对仇恨言论的反对言论有可能使用侮辱性词汇：“称人为黑鬼是不文明的。”）
- 谩骂的对象对一句话是否是仇恨言论起决定性作用。 对独立个人以及对非受保护群体的谩骂或者仇恨心理不是仇恨言论。（例如，“我恨犹太人”是仇恨言论，而“我恨你”“我恨我的桌子”“我恨医生”都不是仇恨言论。）

其他注重点：

1. 请不要因为一句话看起来不像网络上常见的句子而将其标为“不合实际”。“不合实际”应当只用于不合常识或者语法错误的样本。
2. 请独立完成样本的标注，而不要与其他人交流。
3. 大多数句子都很短，作判断时不需要将它们过度复杂化。

非常感谢你的贡献！