

数据使用许可协议

许可方（甲方） : 数据堂（北京）科技股份有限公司

住所地 : 北京市海淀区中关村大街 18 号科贸大厦 B 座 18 层

法定代表人 : 齐红威

项目联系人 :

通讯地址 : 北京市海淀区中关村大街 18 号科贸大厦 B 座 18 层

电话 : 010-82538190

电子信箱 : services@datatang.com

被许可方（乙方） :

住所地 :

法定代表人 :

项目联系人 :

通讯地址 :

电话 :

电子信箱 :

数据堂（北京）科技股份有限公司，成立于 2011 年（股票代码：831428），专业的人工智能数据服务提供商，致力于为全球人工智能企业提供数据获取、处理及数据产品服务。为了更好的推动产业发展，加快人工智能模型及算法研究进展，数据堂决定针对高校、科研机构及重点合作伙伴免费开放 1505 小时中文普通话语音数据基础库的非商业用途使用权，借此回馈 AI 行业。

本合同甲方许可乙方在约定范围内无偿使用 1505 小时中文普通话语音数据基础库（以下简称“本数据”）。双方经过平等协商，在真实、充分地表达各自意愿的基础上，达成如下协议，并由双方共同恪守。

第一条 数据使用许可

乙方为高校、科研机构或甲方指定的合作伙伴。甲方许可乙方对“1505 小时的中文普通话语音数据基础库”享有限制性的、非独占的、不可转让，非商业用途的、禁止演绎使用权，在本合同有效期间内，乙方可以在本合同约定的范围内以科研为目的，对该数据具有储存、使用权利。该数据及数据衍生品（包括但不限于模型）不得以商业使用为目的进行使用。

双方确定，甲方保留本数据包括所有销售、交易、出租、复制、发布、转让等权利在内的全部权利和知识产权。

第二条 本数据描述

1. 格式：16kHz 16bit，wav，单声道
2. 录音环境：安静的室内；包含不影响语音辨识的背景噪音
3. 录音内容：30 万条口语化句子
4. 录音人：6408 人；男性 2999 人，女性 3301 人；≤20 岁 1481 人，21~30 岁 4412 人，31~40 岁 244 人，40 岁以上 163 人；录音人员分布于广东、福建、山东、江苏、北京、湖南、江西等 34 个省级行政区域
5. 设备：安卓:iOS=9:1
6. 语言：普通话；有口音的普通话
7. 应用场景：语音识别；机器翻译；声纹识别
8. 准确率：句准确率不低于 98%
9. 市场价值：361200 元人民币

第三条 双方的权利与义务

1. 在第一条数据试用许可规定的范围内，乙方无偿使用本数据。
2. 本协议期限内，乙方需允许甲方在进行开源数据推广及数据集使用效果跟进时提及乙方名称及 Logo，包括但不限于：
 - (1) 数据堂官网进行 Logo 展示
 - (2) 数据堂自媒体文章中提及乙方名称及 Logo
 - (3) 第三方媒体跟进数据集使用情况时，提及乙方名称等。
3. 当乙方公开发表展示（公开发表展示的形式包括但不限于参加学术会议、公开发表论文、申请专利等）利用本数据的全部或者部分获得的科研成果（科研成果包括但不限于会议论文、会议报告所用 PPT、学术著作、专利等）时，必须在所获科研成果中注明使用“数据堂中文语音数据库”或“Datatang Chinese Mandarin Speech Dataset”，并且同时注明出处

<https://www.datatang.com>。乙方应当在其得知科研成果取得或者公开发表展示后的 5 日内书面通知甲方，并向甲方说明注明科研成果的名称及出处。

4. **甲方仅许可乙方以科研为目的使用本数据。乙方不得以商业目的使用该数据，或者利用本数据开发和生产商业产品。**
5. 乙方有义务保护数据的完整性，妥善储存本数据，并对本数据及其技术细节进行严格保密。
6. **在未得到甲方事先许可的情况下，乙方不得公开发布本数据的任何内容，或者以销售、交易、出版、传播、输送、翻译或其他方式向第三方披露本数据的全部或者部分，除在符合本条第三条第 3 款规定的研究成果公开发表的情形外。数据的任何格式或者任何复制品视同原始数据。**
7. 乙方应当在法律法规规定范围内合法使用本数据，并确保本数据不会被用于违法目的，也不会损害社会公共利益。
8. 本公司不承诺对所提供的数据中可能存在的与现势的某种不一致或者缺陷进行修改。本公司不因该数据本身的瑕疵而对使用方所造成的任何后果承担任何责任。

第四条 法律责任

1. 乙方保证其机构具有独立法人资格；或者乙方经过法人合法授权且在权限内签订本合同。若乙方有以下情形之一，应当支付甲方本数据的全部市值，361200 元人民币。
 - (1) 乙方在获得数据后，决定将本数据用于商业开发。此时乙方应当在决定作出的 3 日内书面通知甲方，并在通知甲方后的 20 日内支付价款；
 - (2) 乙方超出科研目的利用本数据，包括但不限于将数据衍生品用于商业用途等。
2. 若乙方出现本协议第四条第 1 款 (2) 项情形，且造成甲方其他损失的，乙方应当向甲方赔偿其他损失。
3. 对乙方违反约定本协议第三条规定使用数据，未妥善保管数据造成数据泄漏，或者无论以何目的向第三方提供全部或者部分本数据的，乙方在赔偿甲方本数据的全部市值 361200 元人民币的基础上，还应当赔偿对甲方造成的损失。双方约定损失数额具体计算方法为以下两种之一：
 - (1) 若乙方向固定对象提供本数据，损失数额 = 乙方提供的第三方的数量 × 36.12 万元人民币
 - (2) 若乙方向不特定对象提供本数据，或者公开发布本数据的，应当赔偿甲方损失 100 万元人民币。

当乙方造成实际损失超过以上损失计算方法得出的数字时，应当按照实际损失计算数额。

若乙方在合同终止或解除后未履行第六条规定的删除义务，仍然存有全部或者部分数据（包括任何数据的复制品）的，应当按照数据集原价赔偿甲方损失。

4. 若乙方在其以本数据而获得的研究成果公开发表展示时未按照本协议第三条第 3 款规定对本数据进行注明，或者注明有错误的，应当首先采取措施改正。成果已经发表不能改正的应当赔偿甲方人民币 5 万元。
5. 对由于乙方利用本数据而给第三人造成的损失或者违反国家法律法规的，乙方单独承担责任，甲方概不负责。

第五条 合同的生效与终止

本数据使用许可协议的合同期限为自签约之日起 1 年。在合同有效期内，若乙方违反本合同项下任何约定，则甲方有权随时解除合同。

第六条 合同终止后的义务

乙方在本合同终止或者解除后 5 日内，应当销毁所有其根据本协议获得的储存在硬盘、移动设备或者其他存储介质上的数据（包括任何数据的复制品）。在数据销毁后，乙方应当向甲方提供一份由乙方法定代表人签字的书面声明，确认乙方已经按本条的要求对数据进行了销毁。

本协议终止后，乙方仍然受到本合同第三条第 3、4、5、6 款，第四条、第六条的制约。

第六条 纠纷解决

双方因履行本合同而发生的争议，应协商、调解解决。协商、调解不成的，确定按以下第 2 种方式处理：

1. 提交 _____ 仲裁委员会仲裁；
2. 依法向甲方所在地有管辖权的人民法院起诉。

第七条 本合同一式 2 份，具有同等法律效力。

第八条 本合同经双方签署后生效。

（以下无正文）

甲方：_____（盖章）

法定代表人 / 授权委托人：_____（签名）

年 月 日

乙方：_____（盖章）

法定代表人 / 授权委托人：_____（签名）

年 月 日