

人工智能与自然语言理解培训协议

甲方：
姓名：**汤浩浩**
联系电话：**16655171202**
个人邮箱：**thhemail@163.com**

乙方：人工智能与自然语言处理课程培训组
代理人：**胡迪**
联系电话：**18910538029**
通讯地址：**北京市朝阳区北四环中路27号盘古大观**
联系邮箱：**deeplearning.nlp.zh@gmail.com**

甲方委托乙方进行关于人工智能与自然语言理解的培训课程，为保证双方权益，保障课程质量，签订本协议。

一、相关术语定义：

本协议中出现的相关具体术语定义为：

1. 不可抗力：所谓不可抗力，是指合同订立时不能预见、不能避免并不能克服的客观情况。包括自然灾害、如台风、地震、洪水、冰雹；政府行为，如征收、征用；社会异常事件，如罢工、骚乱三方面；
2. 在线直播课程：是指利用互联网计算机在线进行视频、语言教学的方式；
3. 录播课程：是指利用互联网内容，利用以及录制好的课程进行学习；

二、培训事宜

乙方负责为甲方提供人工智能与自然语言理解的系列培训课程，甲方负责完成课程以及课程相关的作业，项目练习，论文阅读以及简历修改和模拟面试环节。

三、培训项目

乙方负责为甲方提供以下课程：

1. 《人工智能与自然语言理解》：该课程为在线课程，每周在线课程3小时，如果遇到国家法定节假日或不可抗力因素，则会调整课程，但是乙方调整课程前需提前至少3天通知甲方。
2. 《科学计算与数学建模》：该课程为在线课程，每次课程3小时，如果遇到国家法定节假日或不可抗力因素，则会调整课程，但是乙方调整课程前需提前至少3天通知甲方。

3. 《人工智能英语能力提升》：该课程为在线课程，每次课程3小时，如果遇到国家法定节假日或不可抗力因素，则会调整课程，但是乙方调整课程前需提前至少3天通知甲方。
4. 《Python高级程序设计与AI编程》：该课程为录播课程，使用的是Udacity的课程内容，乙方负责甲方的作业辅导和课程答疑。
5. 《算法分析与设计》：该课程为Stanford University的录播课程，乙方负责为甲方提供申请的渠道，负责课程作业辅导与课程答疑。
6. 其他：Tensorflow, Keras, Numpy, Pandas, matplotlib 的使用
7. 简历修改、辅导答疑与职业生涯规划

四、辅导内容：

乙方为甲方提供以下辅导内容：

- a). 仔细阅读10篇经典论文；
- b). 8次编程与AI理论练习；
 1. 实现西部世界对话智能系统；
 2. 北京市地铁自动换乘；
 3. 上海市外卖小哥送餐路线规划问题
 4. 中文拼写错误自动纠正
 5. 实现贝叶斯分类器，依据药物说明书进行药物适应症自动识别
 6. 手动实现一个神经网络模型
 7. 使用词向量发现新词汇
 8. 新闻抄袭分类智能模型
- c). 4次项目实战练习；
 1. 新闻人物言论自动提取或 PDF 重要信息自动标注系统
 2. 细粒度客户评论自动分类
 3. 文本自动摘要系统
 4. 面向服务的对话机器人

五、就业指导服务：

- a). 普通班、保过班提供2次简历修改服务；
- b). 普通版提供1次模拟面试服务，保过班提供2次模拟面试服务；

六、课程安排：

课程讲授为2019年3月31日 至 2019年8月14日

第一部分 基础篇 经典人工智能模型方法

第一周 人工智能导论与语法树，自动机理论

1.1 形式语言与语法树

- 1.1.1) 形式语言与正则表达式
- 1.1.2) 语法树与语言表示
- 1.1.3) 使用语法树自动生产语言
- 1.2 自动机理论**
 - 1.2.1) 有限自动机基础
 - 1.2.2) 语法树解析
 - 1.2.3) 语法树解析的实际例子
- 1.3 作业：基于Syntax Tree实现西部世界对话智能系统**
 - 1.3.1) 数据驱动的编程
 - 1.3.2) Python 实现句子生成与语法解析

第二周 智能搜索策略

- 2.1 智能搜索方法论**
 - 2.1.1) 搜索问题与决策问题
 - 2.1.2) 智能搜索典型问题分析(传教士过河、八皇后等问题)
- 2.2 智能搜索的实现**
 - 2.2.1) 深度搜索，广度搜索与最优搜索
 - 2.2.2) 搜索剪枝问题
- 2.3 作业：北京市地铁自动换乘**
 - 2.3.1) 数据获取爬虫的建立
 - 2.3.2) 编写智能搜索Agent

第三周 动态规划与线性优化：

- 3.1) 优化问题**
 - 3.1.1) 优化问题的背景
 - 3.1.2) 优化问题的现状与常用方法
- 3.2) 动态规划**
 - 3.2.1) 动态规划的原理
 - 3.2.2) 动态规划的典型实例
 - 3.2.3) Python 实现动态规划的最佳实践
- 3.3) 线性优化**
 - 3.3.1) 线性优化的原理
 - 3.3.2) 线性优化经典作业
 - 3.3.2) Python 线性优化最近实践
- 3.4) 作业：上海市外卖小哥送餐路线规划问题**
 - 3.4.1) 问题复杂度分析
 - 3.4.2) 使用动态规划解决实例
 - 3.4.3) 使用线性规划解决实例

第四周 自然语言理解初步

- 4.1 词向量**
 - 4.1.1) 文本表示初步
 - 4.1.2) 降维与 embedding 的原理；
 - 4.1.3) 词向量初步知识
 - 4.1.4) Python 词向量使用的最佳实践
- 4.2 关键词提取**
 - 4.2.1) 关键字提取的主要方法与挑战
 - 4.2.2) 基于频率的TFIDF
 - 4.2.3) 基于图关系的 Text-Rank
 - 4.2.4) 基于机器学习的方法
 - 4.2.5) 基于词向量与图网络的方法
 - 4.2.6) Python 关键词提取的最近实践
- 4.3 实体识别**
 - 4.3.1) 实体识别的原理与现状
 - 4.3.1) 实体识别的应用场景
 - 4.3.2) Python 实体识别的最佳实践
- 4.4 依存分析**
 - 4.4.1) 依存分析的原理与现状
 - 4.4.2) Python 依存分析的最佳实践

第五周：搜索引擎与文档检索：

5.1 自动检索系统

- 5.1.1) 搜索引擎与文档搜索的背景
- 5.1.2) 基于关键字的文本搜索
- 5.1.3) 布尔代数搜索

5.2 PageRank

- 5.2.1) PageRank 原理
- 5.2.2) PageRank 的其他应用场景
- 5.2.3) Python 实现大规模搜索系统的关键能力与算法实例

项目实训一：新闻人物言论自动提取 或 PDF 重点信息智能标准系统

- 数据获取，数据转换，数据标准化
- 词向量的构建
- 依存分析与实体识别，重要信息识别
- 搜索系统
- 综合实现

第二部分 机器学习与深度学习

第六周 统计概率模型：

6.1 语言模型

- 6.1.1) 语言模型的历史背景与意义
- 6.1.2) 语言模型的种类

6.2 统计概率模型

- 6.2.1) 条件概率
- 6.2.2) Python统计语言概率的实现
- 6.2.3) Good-Turing Estimation

6.3 编辑距离与文本相似度

- 6.3.1) 文本相似度的主要方法
- 6.3.2) 编辑距离的原理
- 6.3.3) 编辑距离的 python 实现

6.4 作业：中文拼写错误自动纠正

- 6.4.1) 语言模型的构建
- 6.4.2) 编辑距离的计算
- 6.4.3) 自动纠错算法的实现
- 6.4.4) Python源代码完整分析

第七周 经典机器学习一

7.1 机器学习的历史发展与原理

- 7.1.1) 机器学习的背景与原理
- 7.1.1) 机器学习的主要流派
- 7.1.2) 机器学习的现状分析

7.2 过拟合与欠拟合

- 7.2.1) Bias和 Variance
- 7.2.2) 模型能力的分析
- 7.2.3) 数据能力的分析
- 7.2.4) 过拟合与欠拟合的原理与策略

7.3 训练集，测试集，准确度

- 7.3.1) 数据对机器学习模型的影响
- 7.3.2) 训练集、测试集、准确度的关系

第八周：经典机器学习二：

8.1 经典机器学习模型

- 8.1.1) 回归和分类
- 8.1.2) 逻辑回归
- 8.1.3) 贝叶斯分类器
- 8.1.4) KNN模型，

- 8.1.5) SVM
- 8.1.6) 决策树
- 8.1.7 Python) 机器学习模型的最近实践

8.2 Ensemble 机器学习方法

- 8.2.1) Ensemble 机器学习的原理
- 8.2.2) Random Forest 随机森林
- 8.2.3) XGBOOST模型

第九周：经典机器学习三：

9.1 非监督/半监督学习与聚类模型：

- 9.1.1) K-Means算法与实例
- 9.1.2) 层次聚类与实例
- 9.1.3) 基于 embedding 的聚类机器实例

9.2 机器学习常见实践问题分析

- 9.2.1) 天气预测
- 9.2.2) 文本分类
- 9.2.3) 图像分类
- 9.2.4) 机器阅读理解
- 9.2.5) 博弈问题

9.3 作业：实现贝叶斯分类器，依据药物说明书进行药物适应症自动识别

第十周 深度学习初步

10.1 神经网络

- 10.1.1) Loss函数, Backpropagation
- 10.1.2) 梯度下降
- 10.1.3) softmax, cossentropy
- 10.1.4) Optimizer 优化器

10.2 神经网络的实践分析

- 10.2.1) 模型的稳定性
- 10.2.2) 模型的可解释性
- 10.2.3) 模型的运行分析

10.5 作业：手动从零实现一个神经网络模型

- 10.3.1) 实现神经元
- 10.3.2) 实现拓扑排序
- 10.3.3) 实现 Backpropagation
- 10.3.4) 实现神经元权重自动调整
- 10.3.5) 利用完成的神经网络模型进行真实机器学习任务

项目作业二：细粒度客户评论自动分类

- 数据预处理过程
- 数据分析与整理
- 模型的分析与搭建
- 模型的调整与分析
- 模型的部署与发布

第十一周 word2vec

11.1 word embedding与词向量

- 11.1.1) 词向量的原理
- 11.1.2) 哈夫曼树与 Negative Samples
- 11.1.3) GloVe, CoVe, ELMo 等高级词向量方式
- 11.1.4) Python 利用神经网络训练词向量的实例

11.2 句子向量

- 11.2.1) 句子向量的使用场景与背景
- 11.2.2) 句子向量的构建与评价标准
- 11.2.3) Python 构建句子向量的实例

11.3 词向量的高级用法

- 11.3.1) 利用词向量找到隐藏重要词汇

11.3.2) 利用词向量找到新词汇

11.4 作业：使用词向量自动整理同义词

第十二周 CNN卷积神经网络

12.1 卷积神经网络与 Spatial Invariant

12.1.1) 卷积神经网络的历史背景

12.1.2) 卷积神经网络空间平移不变形(Spatial Invariant)的原理

12.1.3) 卷积神经网络与 weights sharing

12.1.4) 卷积神经网络的原理及Python 实现

12.2 Pooling, Dropout 与 Batch Normalization

12.2.1) Pooling

12.2.2) Dropout

12.2.3) Batch Normalization

12.3 CNN 的可视化

12.4 经典 CNN 模型分析：

12.4.1) LeNet

12.4.2) AlexNet

12.4.3) GoogLeNet

12.4.4) VGG, ResNet

12.4.5) DenseNet

12.5 作业：进行萝莉和正太的分类

12.5.1) Python 深度学习环境的搭建

12.5.2) Keras, Tensorflow 的介绍与使用方法

12.5.2) 使用 Keras 搭建CNN 模型模型

12.5.3) 模型的调试与优化

12.5.4) 模型的发布

第十三周 RNN循环神经网络

13.1)序列模型

13.1.1) 时间序列问题的分析

13.1.2) 时间序列模型存在的问题挑战

13.2)RNN 循环模型

13.2.1) RNN 的原理

13.2.2) RNN 的相关问题

13.3)LSTM 与 GRU

13.3.1) LSTM的原理与实现

13.3.2) GRU 的原理

13.3.3) Python 进行 RNN 模型的最佳实践

13.4)RNN 训练的高级问题

13.5)Transfer Learning 迁移学习

13.5.1) 迁移学习的背景

13.5.2) 迁移学习的方法

13.5.3) Python 实现迁移学习的最佳实践

第十四周 深度学习高级问题

14.1 Seq2Seq, Transform, BERT

14.1.1) Seq2Seq 的原理

14.1.2) Seq2Seq 中的搜索方法

14.1.3) Attention注意力机制

14.1.4) Python Seq2Seq模型的最佳实践

14.1.4) Transform

14.1.5) BERT

14.2 强化学习

14.2.1) 强化学习的原理

14.2.2) 强化学习的常用方法与实例

14.2.3) 强化学习面临的挑战

14.3 自动对话机器人、文本自动摘要生成、文本自动阅读理解、自动驾驶等深度学习高级问题

第十五周：面向服务的智能客户机器人与新闻自动摘要生成

15.1 实训三： 文本自动摘要系统的构建讲解与导引

- 自动摘要的问题背景与挑战
- 中文文本摘要遇到的问题
- 使用 TextRank 进行文本自动摘要的实现
- 使用 Sentence Embedding 句子向量进行文本自动摘要的实现
- 完整的文本摘要系统所需要的技术能力分析

15.2 实训四： 面向服务的对话机器人的构建讲解与导引

- 对话机器人的历史背景
- 使用语法树进行对话的实现
- 意图分析与识别
- 文本相似度匹配
- 文本快速检索
- 对话机器人的整体架构分析

第十六周：目前人工智能与局限性、前沿 NLP 问题的现状及发展情况

16.1 学习能力迁移问题，样本迁移问题

16.2 机器学习的可解释性

16.3 非结构数据的处理的计算

16.4 经典 AI 模型的计算复杂性

16.5 AI 产业化面临的问题：

16.5.1) 数据标注与结构化数据

16.5.1) 问题定义与认识

16.5.2) 少量数据与机器学习的不可行性

16.6. 基于人类背景知识的常识推理与认知问题

七、 就业与职业生涯辅导

2019年7月14日 - 2019年9月28日

八、 乙方的责任与义务

1. 乙方负责为甲方按时提供课程， 按时完成作业辅导及课程辅导。 如果乙方因故需要调整课程内容或调整课程时间， 需要提前一周告知甲方并征得甲方同意；
2. 乙方负责批改及审核甲方提交的作业及项目；
3. 乙方在甲方未能及时完成相应工作时， 需要及时督促；
4. 乙方需为甲方的个人信息保密， 具体保密内容见第十条“保密义务”；
5. 乙方未完成以上1至4任一条款的， 甲方可随时要求乙方退还学费， 乙方若未能恢复服务， 乙方需在甲方提出退款要求7天内退还相应学费。

九、 甲方的责任与义务

1. 甲方需要按时完成乙方布置的作业， 按时完成课程；
2. 甲方只能在协议规定的服务时间联系乙方。 协议规定的时间是指所有学员的每周一次的在线课程， 通关班学员的每周集体答疑， 保过班学员每周一次的一对一答疑， 以及简历修改， 模拟面试。 其他时间段内乙方不需要服务甲方。
3. 甲方需要提供相关注册网站的注册信息， 报告用户名， 密码， 以方便乙方检查甲方的作业进度；

4. 甲方需要首先自己努力尝试解决问题，在努力无果的情况下，再请求乙方的支持；
5. 甲方需要按时参加答疑课程；
6. 甲方需要按时注册相关课程的账号进行学习；
7. 甲方需要为乙方的相关信息保密，具体保密内容见第十条“保密义务”；
8. 甲方若未能完成相应1-8的任一相应责任与义务的，乙方须单独提示甲方完成，若提醒3次甲方，甲方仍然不能履行其责任与义务的，乙方可单方面终止服务并保留学费。

十、 保密义务

1. 乙方不得泄露任何关于甲方的信息，信息包括但不限于甲方的姓名，电话，毕业院校，工作单位，地址，邮箱，年龄等任何与甲方有关的信息；
2. 甲方不得在未经允许的情况下，向任何人或组织泄露关于乙方的信息，信息包括但不限于：
 1. 培训时候如何使用互联网；
 2. 相关培训人员的姓名；
 3. 相关培训人员的工作单位；
 4. 相关培训的具体内容；
 5. 相关培训进度；
 6. 培训时候所使用的讲义、幻灯片、数据源、数据包；
 7. 培训时所布置的作业；

若乙方在甲方不允许的情况下泄露任何关于甲方的信息，在甲方提出声明后，需乙方需立即退还甲方 全部学费 * 2 的赔偿金。超过三天退款的以每天 全部学费*2*0.03%的利率 补缴滞纳金。

若甲方泄露关于乙方的信息，再有直接证据指明的情况下，需要在7天之内赔偿乙方人民币 80000.00 元， 大写 捌万元整。超过七天未付款的，每天以0.03%的利率缴纳的滞纳金。

十一、 合同金额及支付

甲方所报课程为 通关班，课程费用为人民币 4800.00 大写 肆仟捌佰元整 元。需在协议签署日期2019年3月29日之前将金额转入支付宝账号 18910538029（胡迪）中，否则取消此次培训协定。转账是请备注“学生姓名 + 通关班”。

十二、 关于预科期间付费

在甲方未进行编程测试，或编程测试未通过，而进行了预科课程的学习的情况下。若甲方希望在预科课程期间获得乙方的课程辅导，甲方可在预科期间提前付款，乙方负责辅导甲方完成预科课程。

十三、 退款协议

1. 开课十日无条件退款；在2019年3月31日 - 2019年 4 月9日 24点前，甲方可要求乙方退款，乙方须无条件3日内将全部学费退还至甲方账户。

十四、 保过班全额退款协议

保过班退款： 若保过班学员完成课程且投递简历大于20家公司，但未获得相应的人工智能算法工程师、机器学习/深度学习工程师、自然处理工程师、数据科学家、算法工程师等相关职位offer（包括实习offer），或签约薪资低于25万(限北京、上海、广州、深圳、杭州),则全额退款。

14.1 全额退款在以下条件均满足下成立：

1. 完成全部八次的Assignments的80%， 完成全部四次Projects；
2. 《人工智能与自然语言处理》，《科学计算与数学建模》课程出勤率大于80%；
3. 《算法设计与分析课程》的作业正确率 $\geq 80\%$ ；
4. 《Python高级程序设计》的课后作业及练习全部完成；
5. 完成满足要求的论文阅读任务的80%， 即， 完成至少8篇满足课程要求的论文阅读练习；
6. 主动投递简历超过20家公司；
7. 投递的20家公司的职位必须为：
 - a) 算法工程师（含算法专家， 算法研究员，以及其他相关职位）
 - b) 人工智能工程师（含人工智能专家， 人工智能研究员，以及其他相关职位）
 - c) 机器学习工程师（含机器学习专家， 机器学习研究员等相关职位， 人工智能工程师以及其他相关职位）
 - d) 数据分析师（含数据分析专家， 数据分析员， 数据分析研究员， 数据分析科学家， 以及其他相关职位）
 - e) 自然语言处理工程师（含NLP工程师， 自然语言（NLP）专家， 推荐算法工程师， 文本挖掘工程师， 文本挖掘专家，以及其他相关职位）
 - f) 数据科学家（含数据挖掘工程师， 数据挖掘专家， 研究科学家，以及其他相关职位）
 - g) 搜索算法工程师（含搜索算法专家）
8. 在课程开始一年后，即，2020年3月31日之前，未收到任何一家公司的offer；
9. 通过简历初选被公司邀请面试后， 参加面试的公司比例必须不低于80%的邀约公司(数字为小数时四舍五入)， 例如， 同学A发送24封简历后收到了12个公司的面试邀请， 该同学必须至少参加 $12 * 0.8 = 10$ 家 公司的面试；
10. 学员必须将每次的面试反馈反馈给课程组；（面试报告电话面试， 视频面试）
11. 学员通过面试并拿到offer后主动告知本课程组

14.2 以下条件下不满足全额退款条件：

1. 任意和14.1条款相违背的行为；
2. 学员主动拒绝所有offer；
3. 学员连续15天失去联系；
4. 学员连续2次没有参加课程；
5. 学员没有自己可以日常使用的笔记本电脑或台式机；
6. 学员没有设定合适的互联网访问设备；
7. 学员或课程组因为不可抗力而中断课程；
8. 学员因故电脑损坏或网络损坏而且没有及时修复的；
9. 学员因故转入其他非计算机算法行业的；
10. 学员因故没有完成学业的；

14.3 全额退款时间

全额退款时间为2020年3月31日，乙方必须为满足14.1全部条款且没有违反14.2任意一条条款的学员退还全部学费，退还学费以合同签署时所写数额为准。若乙方延迟退款，每超过1天，需要交纳1.0%学费的滞纳金。在超出应该退款时间7日之后，乙方需无条件退还全部学费以及全部滞纳金，否则甲方可向司法机关提起追责。

14.4 全额退款方式

1. 乙方必须以现金或支付宝/微信转账，银行卡汇款的形式进行。不能以支票、欠条、期货、股票、实物等进行退款；
2. 必须以人民币进行交易，不得以除人民币之外的任何货币进行交易；
3. 不得使用各种比特币进行交易。

十四、其他约定

1. 若本协议中出现任何与中华人民共和国法律法规或当地法律法规相违的条款，本条款作废；

甲方签字：

汤浩浩

2019年3月2日

乙方签字：

胡迪

2019年2月16日