

## 《自然语言处理》课程教学大纲

课程英文名	Natural Language Processing				
课程代码	B050165s	课程类别	专业课	课程性质	专业选修
学 分	2		总学时数	32	
开课学院	计算机学院		开课基层教学组织	智能信息处理课程组	
面向专业	计算机科学与技术、智能计算与数据科学(计算机科学与技术)、计算机科学英才班(计算机科学与技术)、软件工程		开课学期	第 5 学期	

注：课程类别是指学科基础课/专业课/实践环节/通识公共课/公共基础课/其他；

课程性质是指通识必修/通识选修/学科必修/专业必修/专业选修/实践必修/实践选修。

### 一、 课程目标

自然语言处理 (Natural language processing, NLP) 或计算语言学是信息时代最重要的技术之一。NLP 的应用无处不在，因为人们几乎用语言交流一切：网络搜索、广告、电子邮件、客户服务、语言翻译、虚拟代理、医疗报告等。本课程为自然语言处理的基础教程，从计算的角度来研究人类语言。本课程的主要重点是理解课程大纲中列出的各种 NLP 任务、有效解决这些问题的算法以及评估其表现的方法。将重点关注统计和神经网络学习算法，这些算法在文本语料库上进行训练，以自动获取执行任务所需的知识。课堂讲授将讨论一般问题以及目前的抽象算法，并提供一些算法的实现版本，作为课程项目的一部分进行扩展和实验。

课程的主要知识单元有：系统性介绍自然语言处理常用的理论知识，包括自然语言处理任务限制、技术范畴、语料库、中文自动分词、数据预处理、自然语言处理工具、数据可视化等及相关算法，目的是使学生学会自然语言处理相关的基础知识，以便为应用涉及的任务分析、技术选型奠定基础。同时介绍自然语言处理的命名实体识别、文本信息抽取、文本分类、问答与摘要、机器翻译、语义计算、知识图谱、情感分析等技术领域与重要应用，学习自然语言处理常见应用任务的内涵和解决这些任务的常用方法。

通过理论教学和实践活动，达到以下课程目标：

**课程目标 1：**能够理解自然语言处理的基本概念，包括研究内容和应用领域。能够阅读有关自然语言处理的专业论文，具备阅读专业论文的能力

**课程目标 2：**掌握基于概率统计模型的自然语言处理方法，能够利用现有自然语言处理工具进行自然语言相关论题的理论与实验研究

**课程目标 3:** 能够利用本课程及论文阅读所掌握的知识, 选定某个主题, 开发实现一个自然语言处理领域的实验系统

**课程目标 4:** 具备基本的科学素养和创新意识, 在课程实践的过程中, 及时了解自然语言处理的国内外新技术和发展趋势, 及时掌握国家相关方面的科技战略需求, 树立强烈的爱国主义使命感与责任心, 同时能客观分析人工智能, 特别是自然语言处理技术发展中涉及到的伦理和社会问题。

## 二、 课程目标与毕业要求对应关系

根据计算机科学与技术专业培养方案中毕业要求指标点的分解、课程与毕业要求的对应关系表, 《自然语言处理》支撑毕业要求 1 的指标点 1-3 和 1-4, 毕业要求 2 的指标点 2-2, 毕业要求 4 的指标点 4-3, 毕业要求 8 的指标点 8-1, 毕业要求 10 的指标点 10-1, 课程目标与毕业要求、指标点的对应关系如表 1 所示。

**表 1 课程目标与计算机科学与技术专业毕业要求对应关系**

毕业要求	指标点	课程目标及支撑权重
毕业要求 1: 工程知识: 掌握数学、自然科学、工程基础、计算机专业领域的知识, 并能应用于计算机领域复杂工程问题的解决方案中。	1-3 能够运用计算机专业知识, 对计算机领域复杂工程问题解决方案进行分析与优化。	目标 2: 0.4 目标 3: 0.6
	1-4 掌握计算机系统、物联网、人工智能、大数据、网络安全等某个专业领域的知识, 并用于解决计算机领域的复杂工程问题。	目标 2: 0.4 目标 3: 0.6
毕业要求 2: 问题分析: 能够应用数学、自然科学和工程科学的基本原理, 以及科学思维方法, 对计算机领域的复杂工程问题进行识别、表达和分析, 并通过文献查阅与研究获得有效结论。	2-2 具备文献检索能力, 能够对计算机相关领域的复杂工程问题进行调查、研究分析。	目标 1: 0.6 目标 3: 0.4
毕业要求 4: 研究: 能够基于包括计算学科在内的科学原理, 采用科学方法研究计算机领域的复杂工程问题, 包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4-3 能够收集、分析与解释数据, 并通过信息综合得到合理有效的结论。	目标 1: 0.2 目标 2: 0.3 目标 3: 0.5
毕业要求 8: 职业规范: 具有人文素养和社会责任感, 能够在计算机工程实践中理解伦理道德、遵守职业规范、履行社会责任。	8-1 具有人文社会科学素养、社会责任感。	目标 1: 0.2 目标 4: 0.8
毕业要求 10: 沟通: 具备一定的国际视野和跨文化沟通能力, 能够就专业领域的复杂工程问题与业界同行及社会公众有效沟通, 包括文字表达和语言交流。	10-1 能够就计算机复杂工程问题撰写报告和设计文稿、陈述发言、清晰表达或回应指令。	目标 2: 0.4 目标 3: 0.6

根据软件工程专业培养方案中毕业要求指标点的分解、课程与毕业要求的对应关系表, 《自然语言处理》支撑毕业要求 1 的指标点 1-3 和 1-4, 毕业要求 2 的指标点 2-2, 毕业要求 4 的指标点 4-

3, 毕业要求 8 的指标点 8-1, 毕业要求 10 的指标点 10-1, 课程目标与毕业要求、指标点的对应关系如表 2 所示。

表 2 课程目标与软件工程专业毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
毕业要求 1: 工程知识: 掌握数学、自然科学、工程基础、软件工程专业领域的知识, 并能应用于软件工程领域复杂工程问题的解决方案中。	1-3 能够运用软件工程专业知识, 对软件工程领域复杂工程问题解决方案进行分析与优化。	目标 2: 0.4 目标 3: 0.6
	1-4 掌握某个专业领域知识, 并用于解决软件工程领域复杂工程问题。	目标 2: 0.4 目标 3: 0.6
毕业要求 2: 问题分析: 能够应用数学、自然科学、工程科学的基本原理和科学思维方法, 对软件工程领域复杂工程问题进行识别、表达和分析, 并通过文献查阅与研究获得有效结论。	2-2 具备文献检索能力, 能够对软件工程领域复杂工程问题进行调查、研究分析。	目标 1: 0.6 目标 3: 0.4
毕业要求 4: 研究: 能够基于包括计算学科在内的科学原理, 采用科学方法研究软件工程领域复杂工程问题, 包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4-3 能够收集、分析与解释数据, 并通过信息综合得到合理有效的结论。	目标 1: 0.2 目标 2: 0.3 目标 3: 0.5
毕业要求 8: 职业规范: 具有人文素养和社会责任感, 能够在计算机工程实践中理解伦理道德、遵守职业规范、履行社会责任。	8-1 具有人文社会科学素养、社会责任感。	目标 1: 0.2 目标 4: 0.8
毕业要求 10: 沟通: 具备一定的国际视野和跨文化沟通能力, 能够就专业领域的复杂工程问题与业界同行及社会公众有效沟通, 包括文字表达和语言交流。	10-1 能够就软件工程领域复杂工程问题撰写报告和设计文稿、陈述发言、清晰表达或回应指令。	目标 2: 0.4 目标 3: 0.6

根据计算机科学英才班(计算机科学与技术)专业培养方案中毕业要求指标点的分解、课程与毕业要求的对应关系表,《自然语言处理》支撑毕业要求 1 的指标点 1-3 和 1-4, 毕业要求 2 的指标点 2-2, 毕业要求 4 的指标点 4-3, 毕业要求 8 的指标点 8-1, 毕业要求 10 的指标点 10-1, 课程目标与毕业要求、指标点的对应关系如表 3 所示。

表 3 课程目标与计算机科学英才班(计算机科学与技术)专业毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
毕业要求 1: 工程与科学知识: 能够掌握数学、自然科学、工程基础、计算机科学技术理论知识, 并应用在计算机相关领域的复杂工程问题和基础科学问题的解决方案中。	1-3 能够运用计算机专业知识, 对复杂工程问题解决方案进行分析与优化。	目标 2: 0.4 目标 3: 0.6
	1-4 掌握计算机系统、人工智能、大数据、网络安全等某个专业领域的知识, 并用于解决计算机相关领域的复杂工程问题。	目标 2: 0.4 目标 3: 0.6

毕业要求 2：问题分析：能够应用数学、自然科学和工程科学的基本原理，以及计算科学思维方法，对计算机相关领域的复杂工程问题进行抽象分析与识别、建模表达和形式化论证，并通过文献查阅与研究获得有效结论。	2-2 具备文献检索能力，能够对计算机相关领域的复杂工程问题进行调查、研究分析。	目标 1：0.6 目标 3：0.4
毕业要求 4：研究：具有基本的科学素养和研究意识，具备良好的科学思维能力，对未知事物有探索精神和研究兴趣。具有运用数学和自然科学方法解决复杂问题的能力，能够采用科学方法研究计算机相关领域的复杂问题，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4-3 能够收集、分析与解释数据，并通过信息综合得到合理有效的结论。	目标 1：0.2 目标 2：0.3 目标 3：0.5
毕业要求 8：职业规范：具有人文素养和社会责任感，能够在计算机工程实践中理解伦理道德、遵守职业规范、履行社会责任。	8-1 具有人文社会科学素养、社会责任感。	目标 1：0.2 目标 4：0.8
毕业要求 10：沟通：具备一定的国际视野和跨文化沟通能力，能够就专业领域的复杂工程问题与业界同行及社会公众有效沟通，包括文字表达和语言交流。	10-1 能够就计算机复杂工程问题撰写报告和设计文稿、陈述发言、清晰表达或回应指令。	目标 2：0.4 目标 3：0.6

根据智能计算与数据科学(计算机科学与技术)专业培养方案中毕业要求指标点的分解、课程与毕业要求的对应关系表，《自然语言处理》支撑毕业要求 1 的指标点 1-3 和 1-4，毕业要求 2 的指标点 2-2，毕业要求 4 的指标点 4-3，毕业要求 8 的指标点 8-1，毕业要求 10 的指标点 10-1，课程目标与毕业要求、指标点的对应关系如表 4 所示。

**表 4 课程目标与智能计算与数据科学(计算机科学与技术)专业毕业要求对应关系**

毕业要求	指标点	课程目标及支撑权重
毕业要求 1：工程知识：能够掌握数学、自然科学、工程基础、计算机软硬件知识、人工智能、智能计算和数据科学的基础理论及专业知识，并应用在人工智能、智能计算和大数据专业领域及其他相关交叉领域的复杂工程问题的解决方案中。	1-3 能够运用计算机专业知识，对计算机领域复杂工程问题解决方案进行分析与优化。	目标 2：0.4 目标 3：0.6
	1-4 掌握人工智能、智能计算、大数据等某个专业领域的知识，并用于解决计算机领域的复杂工程问题。	目标 2：0.4 目标 3：0.6
毕业要求 2：问题分析：能够应用数学、自然科学、工程科学、人工智能、智能计算和大数据的基本原理，对人工智能和大数据专业领域及其他相关交叉领域的复杂工程问题进行识别、表达、分析和抽象建模，并通过文献查阅与研究获得有效结论。	2-2 具备文献检索能力，能够对计算机领域的复杂工程问题进行调查、研究分析。	目标 1：0.6 目标 3：0.4
毕业要求 4：研究：具有基本的科学素养	4-3 能够收集、分析与解释数据，并	目标 1：0.2

和研究意识，能够采用科学方法研究人工智能、智能计算和大数据专业领域及其他相关交叉领域的复杂工程问题，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	通过信息综合得到合理有效的结论。	目标 2: 0.3 目标 3: 0.5
毕业要求 8: 职业规范：具有人文素养和社会责任感，能够在人工智能、智能计算和大数据工程实践中理解伦理道德、遵守职业规范、履行社会责任。	8-1 具有人文社会科学素养、社会责任感。	目标 1: 0.2 目标 4: 0.8
毕业要求 10: 沟通：具备一定的国际视野和跨文化沟通能力，能够就专业领域的复杂工程问题与业界同行及社会公众有效沟通，包括文字表达和语言交流。	10-1 能够就计算机领域的复杂工程问题撰写报告和设计文稿、陈述发言、清晰表达或回应指令。	目标 2: 0.4 目标 3: 0.6

### 三、 课程目标与教学内容和方法的对应关系

《自然语言处理》的课程目标与教学内容、教学方法的对应关系如表 5 所示。

表 5 课程目标与教学内容、教学方法的对应关系

教学内容	教学方法	课程目标			
		1	2	3	4
1. 自然语言处理概述	课堂讲授、自学	●			●
2. 语言模型	课堂讲授、课堂练习、文献查阅	●	●	●	
3. 概率图模型及序列化标注	课堂讲授、课堂练习、课后实践		●	●	
4. 文本分布式表示模型	课堂讲授、课堂练习、自学		●	●	
5. 基于统计学习的句法分析	课堂讲授、课后实践、文献查阅		●	●	
6. 统计及神经机器翻译	课堂讲授、自学、文献查阅	●	●	●	
7. 问答及对话系统	课堂讲授、自学、文献查阅		●	●	
8. 情感分析技术	课堂讲授、文献查阅			●	●
9. NLP 系统的伦理与社会问题	课堂讲授、文献查阅	●			●

课程教学的详细内容与要求如下：

#### 1. 自然语言处理概述

##### (1) 教学内容：

- 自然语言处理的基本术语
- 自然语言处理的基本概念

##### (2) 教学重点：介绍自然语言处理研究的主要对象和基本方法思想

##### (3) 教学难点：自然语言处理的学科定位

##### (4) 教学要求：了解自然语言处理的基本概念及其基本工具和方法，了解自然语言处理的基本技

术和方法。

**思政融合点 1:** 指导学生课后通过查阅文献,了解国内外自然语言处理领域的相关研究成果,特别是国内比较有影响的自然语言处理研究机构,撰写相关调查报告,引发学生的家国情怀。在后续课程内容学习中,结合学习内容逐步展开相关讨论。

## 2. 语言模型

### (1) 教学内容:

- N-gram 语言模型;
- 概率语言模型的定义
- 模型概率估计及平滑算法
- 模型评价
- 语言模型构造方法
- 神经语言模型

(2) **教学重点:** 语言模型的定义、概率估计及平滑算法。

(3) **教学难点:** 平滑算法,神经语言模型

(4) **教学要求:** 理解经典语言模型与神经网络语言模型,掌握常见的概率估计和平滑算法,了解语言模型的评价方法,掌握困惑度 perplexity 的定义,掌握神经语言模型的定义。

**思政融合点 2:** 根据最大似然估计方法中的各种平滑技术,引申出共同富裕的讨论。

## 3. 概率图模及序列化标注

### (1) 教学内容:

- 隐马尔可夫模型定义及参数估计
- 解码算法
- 隐马尔可夫模型在词性标注任务中的应用
- 最大熵马尔科夫模型和条件随机场模型介绍

(2) **教学重点:** 隐马尔可夫模型的定义、词性标注

(3) **教学难点:** 解码算法

(4) **教学要求:** 熟悉词性标注和序列标注的经典算法,理解词性标注任务的主要功能,掌握隐马尔可夫模型的定义及参数估计,了解解码算法原理

## 4. 文本分布式表示模型

### (1) 教学内容:

- 文本向量空间表示模型
- 词嵌入模型
- 文档分布式表示模型

(2) **教学重点:** 词嵌入模型

(3) **教学难点:** 词嵌入模型

(4) **教学要求:** 掌握文本向量空间表示模型的应用,理解词嵌入模型的基本原理和应用。

**思政融合点 3:** 结合文本向量空间表示方式和词嵌入模型的对比, 挖掘其中的科学思维要素, 培养学生的演绎和归纳思维能力, 提高学生的辩证思维和创新思维能力。

## **5. 基于统计的句法分析**

### **(1) 教学内容:**

- 句法分析的概念
- 基于短语结构的句法分析
- 基于依存结构的句法分析
- 形式语法的基本理论
- CKY 句法分析方法
- 概率上下文无关文法的句法分析技术
- 句法分析评价

### **(2) 教学重点: 句法分析的主要技术、CKY 句法分析方法、句法分析的概率统计模型**

### **(3) 教学难点: 概率统计句法分析中的模型参数估计和解码算法**

**(4) 教学要求:** 熟悉基于短语和依存结构的句法分析方法, 掌握 CKY 句法分析算法、概率统计句法分析方法。

## **6. 统计及神经机器翻译**

### **(1) 教学内容:**

- 机器翻译的基本思想
- 统计机器翻译模型框架
- IBM 翻译模型介绍
- 统计机器翻译系统的构建方法
- 神经机器翻译模型及进展介绍

### **(2) 教学重点: 统计机器翻译模型相关技术、神经机器翻译模型的基本结构**

### **(3) 教学难点: 统计机器翻译模型构建、神经机器翻译的结构**

**(4) 教学要求:** 掌握机器翻译的总体思想, 熟悉统计机器翻译的模型结构和构建方法, 了解基于神经网络的机器翻译模型, 能初步实现基于循环神经网络的机器翻译模型。

## **7. 问答及对话系统**

### **(1) 教学内容:**

- 问答系统概述
- 基于知识库的问答系统
- 基于文档检索的问答系统
- 对话系统的组成及构造方法

### **(2) 教学重点: 问答系统的总体结构、对话系统的构造。**

### **(3) 教学难点: 知识库的组织及其在问答系统中的应用、对话系统的设计与实现。**

**(4) 教学要求:** 掌握问答系统的基本构成、了解知识库在问答系统中的应用、熟悉对话系统的组

成和构造方法。

**思政融合点 4:** 通过智能客服系统的现状描述，引导学生探讨智能客服应用场景中可能存在的伦理问题，比如骚扰电话、个人隐私保护的话题，培养学生健康的技术观。

## **8. 情感分析技术**

### **(1) 教学内容:**

- 情感分析的基本概念
- 情感分析的基本性质
- 情感词典的定义和构建方法
- 情感分析的方法
- 细粒度情感分析的方法

**(2) 教学重点:** 情感分析的基本概念、情感词典的构建和应用、情感分析方法、细粒度情感分析方法

### **(3) 教学难点:** 细粒度情感分析方法

**(4) 教学要求:** 能够区分情感分析、情感分类等基本概念，掌握情感分析的主要方法，熟悉情感词典的构建方法，了解细粒度情感分析方法的关键技术。

## **9. NLP 系统的伦理与社会问题**

### **(1) 教学内容:**

- NLP 模型中的公平性问题
- NLP 模型中不公平和偏见现象产生的原因
- 不同类型的偏见和不公平现象
- NLP 模型去偏见的方法

**(2) 教学重点:** NLP 模型中的偏见现象、去偏见方法

### **(3) 教学难点:** 去偏见方法

**(4) 教学要求:** 掌握机器学习和 NLP 系统中的偏见和不公平现象的具体特征，了解偏见和不公平的具体形式，熟悉如何在模型应用中消除偏见和不公平问题

**思政融合点 5:** 此处融合点直接关联于人工智能伦理问题，着重探讨 NLP 系统中存在的各种伦理和社会问题，如性别歧视、民族歧视、不公平等问题，引发学生对于人工智能伦理问题的讨论，开展隐性的价值观教育。

## **四、 实践环节及要求**

本课程不安排课内实践课时，但会根据课程需要安排适当的课外实践任务，主要围绕以下三个主题：

### **1) 文本分类/聚类系统设计与实现**

#### **a) 文本向量空间的表达方法**



- b) 文本向量空间距离的计算模型
  - c) 基于朴素贝叶斯的分类方法
  - d) 基于 KNN 的聚类方法
- 2) 基于概率上下文无关文法的句法分析方法
- a) 概率上下文无关文法的表示方法
  - b) 基于短语结构的句法树
  - c) CYK 句法分析方法
  - d) 基于 PCFG 的句法分析方法
- 3) 基于 RNN 的机器翻译系统
- a) 双语语料库的处理方法
  - b) RNN 的参数选择
  - c) 机器翻译模型的评价方法

## 五、 与其它课程的联系

**先修课程：**程序设计基础，数据结构，概率论与数理统计

**后续课程：**无

## 六、 学时分配

本课程总学时 32 学时，其中讲课 32 学时。如表 6 所示。

表 6 学时分配表

教 学 内 容	讲课 时数	实验 时数	实践 学时	课内 上机 时数	课外 上机 时数	自学 时数	习题 课	讨论 时数
1. 自然语言处理概述	2					2		
2. 语言模型	2				4	2		
3. 概率图模型及序列化标注	6				4	2		
4. 文本分布式表示模型	4				2	2		
5. 基于统计学习的句法分析	6				2	2		
6. 统计及神经机器翻译	4				2	2		
7. 问答及对话系统	4					2		
8. 情感分析技术	2				2	2		
9. NLP 系统的伦理与社会问题	2							
合 计	32				16	16		
总 计	32 讲课学时 + 16 课外上机时数 + 16 自学时数							

备注：自学学时用于预习、复习、习题、自学、课堂拓展等学习活动等。

## 七、 课程目标达成途径及学生成绩评定方法

### 1.课程目标达成途径

课程目标的达成途径如表 7 所示。

表 7 课程目标与达成途径

课程目标	达成途径
课程目标 1：能够理解自然语言处理的基本概念，包括研究内容和应用领域。能够阅读有关自然语言处理的专业论文，具备阅读专业论文的能力	以引导式、启发式和总结式教学方法为主，通过重点/难点内容讲解、课后作业、布置学生文献查阅、进行随堂提问等模式，帮助学生充分理解自然语言处理的相关概念，让学生重点掌握自然语言处理的研究内容和主要技术。
课程目标 2：掌握基于概率统计模型的自然语言处理方法，能够利用现有自然语言处理工具进行自然语言相关论题的理论与实验研究	以引导式、启发式和总结式教学方法为主，通过重点/难点内容讲解、课后作业、布置学生文献查阅、进行随堂提问等模式，结合课外实践，帮助学生掌握基于概率统计方法的自然语言处理模型的构建和实现。
课程目标 3：能够利用本课程及论文阅读所掌握的知识，选定某个主题，开发实现一个自然语言处理领域的实验系统	通过布置课程大项目形式，引导学生开展课程实践活动，重点帮助学生掌握自然语言处理开发环境配置、研究课题选择等任务
课程目标 4：具备基本的科学素养和创新意识，在课程实践的过程中，及时了解自然语言处理的国内外新技术和发展趋势，及时掌握国家相关方面的科技战略需求，树立强烈的爱国主义使命感与责任心，同时能客观分析人工智能，特别是自然语言处理技术发展中涉及到的伦理和社会问题。	通过课堂讲授、课后自学、文献查阅、课堂讨论、分析对比、总结报告等各种方式，让学生对自然语言处理现状与发展趋势有所了解，建立终生学习的意识；同时，进一步了解目前国内相关先进技术与取得的成就，从而建立强烈的民族自豪感与家国情怀。引导学生了解人工智能，特别是机器学习和自然语言处理领域涉及到的伦理和社会问题。

## 2. 学生成绩评定方法

该课程为考查课程。该课程采用形成性评价与终结性评价相结合的评价方法，学期总评成绩由三部分构成：平时成绩，占比 40%；期末课程大作业，占比 60%。各部分的具体评价环节、关联课程目标、评价依据及方法和在总成绩中的占比，如表 8 所示

表 8 课程考核与成绩评定方法

考核项目	考核内容	考核关联的课程目标	考核依据与方法	占总评成绩的比重
平时成绩	课程思政实践	(4)	关于自然语言处理技术的发展历程，通过文献查阅，课后报告，课堂讨论等形式，考查学生对我国相关技术领域情况的了解程度及其科学思维能力情况。	5%
	作业及课堂表现	(1)(2)(3)	作业：以是否提交、提交是否及时、作业质量做评价依据；作业质量按 A、B、C、D、E 五级评分（可再细分），未及时提交的作业评估分降一档处理，期末根据评价等级转换为相应得分。 课堂表现：根据课堂提问和讨论情况给分	35%
期末	课程报告	(1)(2)(3)(4)	功能实现：以程序验收考核作为评价依据，通	60%

末成绩	告		过现场提问考核，重点考察原创性； 实验报告：依据是否按时提交，内容完整度、原创性等	
总评成绩		(1)(2)(3)(4)	=平时成绩*40%+期末成绩*60%	100%

表 9. 考核内容详细评分标准

考核内容	评分标准			
	90-100	75-89	60-74	<60
课程思政实践	报告条理清晰，文字流畅，相关性较强；内容完整且材料丰富，体现强烈的使命感、责任心与民族自豪感	报告条理清楚，字数≥3000，相关性较好；内容完整，材料不够丰富，能体现学生的使命感、责任心与民族自豪感	报告有一定条理，基本相关；内容基本完整但材料较少，能体现学生的使命感与民族自豪感	内容少，或有抄袭现象，体现不出学生的使命感与民族自豪感
课堂参与	学习通课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 15%	学习通课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 50%	学习通课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 85%	学习通课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班后 15%
实践作业	程序运行流畅，功能完善，性能好；代码独创性好；算法与数据结构设计或代码实现等有 2 处及以上创新点	程序运行正常，功能基本实现，但不够完善、优化等，代码自创率应高于 50%；算法与数据结构设计或代码实现等至少有一处创新点	程序运行基本正常，有少量 bug；功能实现至少达到 80%，代码自创率应高于 30%；基本没有创新性	程序运行 bug 多，功能实现低于要求的 80%
期末考查	报告条理清晰，文字流畅内容完整，字数≥3500，数据结构及算法设计合理且效率高，有 2 个以上创新点或改进，参考文献≥5 篇，查重率≤20%，汇报 PPT 图表清晰，设计美观，答辩过程脱稿讲解，分析条理清晰，问题回答准确，小组协作好	报告条理清楚，内容较完整，字数≥2500，数据结构及算法设计合理且性能较好，有 1 个以上创新点或改进，参考文献≥3 篇，查重率≤35%，汇报 PPT 设计较美观，答辩过程脱稿讲解，条理较清楚，问题回答基	报告内容基本完整，字数≥1500，数据结构及算法设计基本合理，参考文献≥2 篇，查重率≤50%，汇报 PPT 美观性及内容一般，答辩过程大部分内容脱稿讲解，大部分问题能基本正确回答，分工基本合理	报告内容不完整，字数少于 1500，数据结构及算法设计不能满足题目基本要求，查重率>60%，有抄袭现象，汇报 PPT 界面文字多，内容不完整，答辩过程基本念 PPT，只能回答少量问题，分工不够合理

		本正确，分工较合理		
--	--	-----------	--	--

## 八、 教学资源

表 10 课程的基本教学资源

资源类型	资源
教 材	本课程不设定固定教材
参考书籍或文献	宗成庆,《统计自然语言处理》,清华大学出版社,2016 Dan Jurafsky and James H. Martin. Speech and Language Processing Yoav Goldberg. A Primer on Neural Network Models for Natural Language Processing Christopher D. Manning and Hinrich Schutze, Foundations of statistical natural language processing, MIT Press, 1999
教学文档	课件、电子书、实验指导书等

## 九、 课程目标达成度定量评价

在课程结束后,需要对每一个课程目标(含思政课程目标)进行达成度的定量评价,用以实现课程的持续改进。

课程目标达成度的定量评价算法:

- 1、 使用教学活动(如课程思政实践、课后作业、课堂练习、实验验收、课堂讨论、互动、阅读报告等等)成绩或期末考试部分题目得分率作为评价项目,来对某个课程目标进行达成度的定量评价;
- 2、 为保证考核的全面性和可靠性,要求对每一个课程目标的评价项目选择超过两种;
- 3、 根据施教情况,评价项目可以由教师自行扩展,权重比例可以由教师自行设计;
- 4、 对某一个课程目标有支撑的各评价项目权重之和为1;
- 5、 使用所有学生(含不及格)的平均成绩计算。

表 11.课程目标达成度定量评价方法

课程目标	课程目标达成度评价方式
课程目标 1: 能够理解自然语言处理的基本概念,包括研究内容和应用领域。能够阅读有关自然语言处理的专业论文,具备阅读专业论文的能力	课堂表现: 0.3 思政实践: 0.2 实践作业: 0.3 期末报告: 0.2
课程目标 2: 掌握基于概率统计模型的自然语言处理方法,能够利用现有自然语言处理工具进行自然语言相关论题的理论与实验研究	课堂表现: 0.3 实践作业: 0.3 期末报告: 0.4
课程目标 3: 能够利用本课程及论文阅读所掌握的知识,选定某个主题,开发实现一个自然语言处理领域的实验系统	实践作业: 0.4 期末报告: 0.6
课程目标 4: 具备基本的科学素养和创新意识,在课程实践的过程中,及时了解自然语言处理的国内外新技术和发展趋势,及时掌握国家相关方面的科技战略需求,树立强烈的爱国主义使命感与责任心,同时能客观分析人工智能,特别是自然语言处	课堂表现: 0.3 思政实践: 0.5 期末报告: 0.2

理技术发展中涉及到的伦理和社会问题。	
--------------------	--

## 十、 说明

本课程大纲主要用于规范计算机科学与技术、软件工程、计算机科学英才班(计算机科学与技术)、智能计算与数据科学(计算机科学与技术)及相关专业的《自然语言处理》课程的教学目标、教学内容、教学方法、教学要求以及考核评价方法等，承担该课程的教师必须遵照本大纲安排授课计划、实施教学过程，完成学生各个阶段与各方面的学习成果考核与评价；在学期末，需对课程目标和课程支撑的毕业要求指标点进行达成度评价。

本课程大纲自 2021 级开始执行，生效之日原先版本均不再使用。

## 十一、 编制与审核

表 11 大纲编制与审核信息

工作内容	责任部门或机构	负责人	完成时间
编制	智能信息处理课程组	黄孝喜	2022.3
审核	智能信息处理课程组	彭勇	2022.3
审定	计算机学院教学大纲编委会	吴海虹	2022.5