**广东省高等教育自学考试《人工智能与大数据》考试大纲**

**（课程代码：13011)**

**Ⅰ 课程性质与课程目标**

一、课程性质和特点

“人工智能与大数据”是为满足计算机应用领域及相关专业对软件应用人才的需要而开设的。通过本课程的学习，学生能够理解人工智能的基础与应用，大数据技术原理与应用，初步学习和掌握人工智能和大数据的综合应用，为进入人工智能、大数据及分支领域进行学习或从事人工智能行业工作奠定基础，指引方向。

“人工智能与大数据”旨在培养学生对人工智能的核心算法、最新进展和发展方向、大数据技术的重要概念和处理架构等的认识和理解，建立起人工智能和大数据关系和综合应用的总体认识，了解人工智能和大数据如何改变生产和生活，在具体应用场景的采用的解决方法和实现途径，掌握目前人工智能领域的主流研究方向。

二、课程目标

设置本课程的主要目的是使考生了解人工智能的基本概念、主要技术和应用，掌握大数据的重要概念，了解相关工具和平台，并发展人工智能和大数据综合思维与解决问题的能力，从而为将来从事人工智能与大数据相关领域的工作做好充分的准备。

通过本课程的学习，考生应达到以下目标。

1. 理解人工智能的基本概念和历史发展，对人工智能有一个全面的了解和认识。

2. 掌握人工智能的基本算法和模型，包括特征提取和机器学习等，掌握这些算法和模型的基本原理。

2. 熟悉人工智能的典型应用，包括自然语言理解和目标检测等，能够在实际应用中掌握这些技术。

4. 掌握大数据的基本概念，熟悉大数据处理架构及分布式存储系统，熟悉大数据的应用及发展趋势。

5.发展人工智能思维，提高综合运用人工智能与大数据解决问题的能力。通过案例和实践，培养学生在实际工作中运用人工智能与大数据技术解决实际问题。

三、与相关课程的联系与区别

人工智能与大数据是一门旨在为学生提供人工智能与大数据的基础知识和技能，使其能够理解和应用人工智能与大数据相关技术的课程。学生在学习本课程前需要掌握高级语言程序设计、数据结构与算法、数据库原理与技术等相关课程的基本知识。

四、课程的重点和难点

本课程的重点是人工智能的核心理论和算法、分布式存储和大数据处理架构，并探讨如何在实际应用中运用这些技术解决实际问题，具体包括人大数据的基本概念、大数据处理架构、分布式存储系统、人工智能基础概念、机器学习算法的基本概念和常见算法、自然语言处理的基本概念和技术、目标检测的技术和应用等。难点方面，这门课程需要学生具备一定的编程和应用基础，对于一些较为复杂的案例实现，需要掌握平台操作和应用实现。此外，人工智能与大数据的应用非常广泛，需要学生具备跨学科的思维能力和一定的综合素养，能够将人工智能与大数据技术应用于实际问题解决中。

**Ⅱ 考核目标**

本大纲是“人工智能与大数据”课程的个人自学、社会助学和考试命题的依据，本课程的考试范围以本考试大纲所限定的内容为准。

本大纲在考核目标中，按照识记、领会、简单应用和综合应用四个层次规定其应达到的能力层次要求。四个能力层次是递进关系，各能力层次的含义是：

识记：学生需要掌握人工智能和大数据的基本概念、原理和算法，了解机器学习、自然语言处理等常见人工智能领域的基本知识和Hadoop大数据处理架构。

领会：学生需要理解人工智能与大数据领域的发展历程、技术趋势和未来发展方向，理解人工智能各个研究领域的典型方法和原理，理解大数据处理架构中各种工具和分布式存储系统。

简单应用：学生需要具备一定的程序设计能力，能够使用图示描述大数据处理平台中各组成部件或者工具的逻辑结构，能够为人工智能应用选择合适的解决方案和相关算法。

综合应用：学生需要结合实际问题，选取适当的大数据结构和人工智能算法，实现基于大数据的人工智能应用。

**Ⅲ 课程内容与考核要求**

1. 大数据的来源、采集与基本概念

一、学习目的与要求

本章的学习目的是使学生掌握大数据和大数据技术的基本概念，包括大数据的来源和定义、大数据的分类和采集、大数据技术的关键问题与关键技术等。

二、课程内容

1.1 大数据的来源与发展

1.2 何为大数据

1.3 大数据的特点

1.4 大数据的分类和采集方法

1.5 大数据的基本概念

三、考核知识点与考核要求

（一）大数据的来源与发展（教材1.1节）

领会：大数据的起源和发展历程。

（二）大数据的基本概念（教材1.2、1.3、1.4、1.5节）

识记：大数据的定义的特点。

领会：大数据的分类和数据采集的方法；大数据技术的内容；大数据的关键问题与关键技术。

四、本章重点、难点

本章的重点：大数据的概念和大数据关键技术。

本章的难点：大数据的分类和数据采集方法。

1. 大数据处理架构Hadoop

一、学习目的与要求

本章的学习目的是使学生掌握大数据处理架构Hadoop及其生态体系，了解生态体系中主要构成部分。

二、课程内容

2.1 Hadoop概述

2.2 数据仓库工具Hive

2.3 大数据仓库Hbase

2.4 编程语言Pig

2.5 协管员ZooKeeper

2.6 Hadoop资源管理与调度

三、考核知识点与考核要求

（一）Hadoop概述（教材2.1节）

识记：Hadoop生态体系里的主要产品

领会：Hadoop要解决的问题和应用场景。

（二）数据仓库工具Hive（教材2.2节）

领会：Hive的特点，Hive与数据库的区别。

简单应用：Hive的执行流程和Hive编译器的组成

（三）大数据仓库Hbase（教材2.3节）

领会：传统数据库存在的问题，table和region的关系，-ROOT-表和.META.表记录的信息和关系。

简单应用：Hbase数据的逻辑存储模型

（四）编程语言Pig（教材2.4节）

识记：Pig Latin中的数据组织形式。

领会：Pig、Hive和数据库的区别。

（五）协管员ZooKeeper（教材2.5节）

识记：ZooKeeper的设计目标。

简单应用：ZooKeeper集群中的五种角色和其功能。

（六）Hadoop资源管理与调度（教材2.6节）

领会：Hadoop三种常见的调度器的调度原理。

四、本章重点、难点

本章的重点：Hadoop生态体系的组成及各个组成部分的功能与特点。

本章的难点：。Hadoop生态体系各个组成部分之间的关联和区别。

1. Hadoop分布式文件系统HDFS

一、学习目的与要求

本章的学习目的是使学生了解HDFS分布式文件系统的设计思想、架构、存储和主要流程实现，加强对Hadoop体系的理解。

二、课程内容

3.1 HDFS的由来

3.2 HDFS的设计思想

3.3 HDFS的主要特性

3.4 HDFS的架构

3.5 HDFS的主要流程

3.6 HDFS异构存储

3.7 HDFS擦除码技术

三、考核知识点与考核要求

（一）HDFS的由来

识记：分布式文件系统（DFS）的概念和主流的分布式文件系统产品

领会：DFS产品的选择因素，文件系统发展的三个阶段

（二）HDFS的设计思想（教材3.2、3.3节）

识记：HDFS的核心设计思路是“分而治之”

领会：HDFS的设计目标，设计思想的主要策略

简单应用：HDFS中解决各类问题的方法

（三）HDFS的架构（教材3.4节）

识记：HDFS架构的结构及组成

领会：HDFS的数据块设计及其优点

简单应用：HDFS的整体架构图描述

（四）HDFS的主要流程（教材3.5节）

综合应用：HDFS主要流程的实现：客户端到名字节点的文件与目录操作、客户端读文件、客户端写文件、数据节点的启动和心跳

（五）HDFS异构存储和擦除码技术（教材3.6、3.7节）

识记：HDFS支持的存储介质和存储策略

领会：擦除码技术的概念、原理和优缺点

四、本章重点、难点

本章的重点：HDFS的设计思想和具体实现。

本章的难点：HDFS中主要流程的实现方法。

1. MapReduce与Spark

一、学习目的与要求

本章的学习目的是让学生学习MapReduce技术，进一步了解Hadoop生态体系的核心技术。

二、课程内容

4.1 MapReduce的设计思想

4.2 MapReduce的组成

4.3 MapReduce的工作流程

4.4 MapReduce的计算过程详解

4.5 MapReduce的使用案例

4.6 Spark概述

三、考核知识点与考核要求

（一）MapReduce的设计思想（教材4.1节）

领会：MapReduce的两个阶段Mapping和Reducing的涵义，MapReduce设计思想的三个方面

（二）MapReduce的组成和工作流程（教材4.2、4.3节）

识记：MapReduce的主从结构和组成部分

领会：MapReduce的处理过程、MapReduce任务执行过程、MapReduce任务进度和状态更新

简单应用： MapReduce作业提交过程、任务执行过程

（三）MapReduce的计算过程（教材4.4、4.5节）

综合应用：MapReduce的计算过程分析

（四）Spark概述（教材4.7节）

领会：Spark的优势、Spark的架构

四、本章重点、难点

本章的重点：MapReduce的设计思想和运作流程。

本章的难点：MapReduce的计算过程。

1. NoSQL数据库、分布式数据库HBase、云数据库

一、学习目的与要求

本章的学习目的是让学生掌握大数据的分布式集群存储技术，学习云数据库的起源、应用与发展。

二、课程内容

5.1 数据库

5.2 非关系型数据库（NoSQL）

5.3 NoSQL数据库的体系框架

5.4 NoSQL数据库的分类

5.5 云数据库及其产品

三、考核知识点与考核要求

（一）数据库（教材5.1节）

识记：数据库的概念

（二）非关系型数据库（NoSQL）（教材5.2、5.3、5.4节）

识记：非关系型数据库的体系架构层次、分类

领会：非关系型数据库与关系数据库的区别和关系

（三）云数据库及其产品（教材5.7节）

识记：云数据库的概念和特点

领会：UMP系统架构中的角色及功能

四、本章重点、难点

本章的重点：非关系型数据库和云数据库的概念和架构。

本章的难点：UMP系统架构设计。

1. 大数据的应用与展望

一、学习目的与要求

本章的学习目的是让学习者了解大数据技术在各个行业的应用需求和前景，提升大数据技术的应用能力。

二、课程内容

6.1 大数据的应用方向

6.2 大数据的成功应用

6.3 大数据分析技术的应用

6.4 大数据挖掘技术的应用实例

6.5 大数据的展望

三、考核知识点与考核要求

（一）大数据的应用方向和成功应用（教材6.1、6.2节）

识记：判断大数据应用成功的指标

领会：大数据的应用方向，大数据的成功应用领域

（二）大数据分析技术的应用（教材6.3节）

领会：大数据分析技术的应用场景

（三）大数据挖掘技术的应用案例（教材6.4节）

领会：大数据挖掘技术的应用场景

（四）大数据的展望（教材6.5节）

领会：大数据发展的挑战和机遇

四、本章重点、难点

本章的重点：大数据技术的应用。

1. 人工智能的基本概念

一、学习目的与要求

本章的学习目的是帮助学生学习人工智能的基本概念，了解人工智能的历史和未来，对人工智能及其应用有初步的认识。

二、课程内容

7.1 什么是人工智能

7.2 人工智能的历史

7.3 图灵测试

7.4 中文屋理论

7.5 人工智能的未来

三、考核知识点与考核要求

（一）人工智能的基本概念（教材7.1、7.2、7.3、7.4节）

识记：人工智能的概念

领会：语音识别系统中的关键问题，当前人工智能研究的两大方向，图灵测试方法，中文屋理论思想

（二）人工智能的未来（教材7.5节）

识记：自动驾驶汽车的智能分级

领会：人工智能对社会的哪些领域产生重大的影响

综合应用：人工智能对行业的综合影响分析

四、本章重点、难点

本章的重点：对人工智能的各种认识

本章的难点：人工智能对社会影响的正负面认知

1. 特征提取

一、学习目的与要求

本章的学习目的是让学生了解如何将客观世界中的对象特征提取出来，这是机器学习等人工智能技术的基础。

二、课程内容

8.1 特征提取基础

8.2 数学方法入门

8.3 图像特征

三、考核知识点与考核要求

（一）特征提取基础（教材8.1节）

领会：特征选择的作用，特征提取的方法

简单应用：数据预处理内容和方法

（二）数学方法入门（教材8.2节）

简单应用：相关性及其计算方法

综合应用：前向选择算法和后向选择算法

（三）图像特征（教材8.4节）

识记：图像的主要特征、最常用的颜色空间

领会：形状与二值图像表示，基于形状特征的应用，有效形状特征的特性，纹理特征的作用，纹理特征的属性，纹理特征的提取方法、空间关系特征的提取方法、姿态估计方法的分类、图像特征分类

简单应用：灰度图像获取二值形状图像的方法

四、本章重点、难点

本章的重点：对象特征的提取方法。

本章的难点：图像表示与各种特征的提取方法。

1. 机器学习

一、学习目的与要求

本章的学习目的是让学习者学习机器学习的概念和核心方法，了解人工智能的该项核心技术。

二、课程内容

9.1 机器学习的基本概念

9.2 强化学习

9.3 推荐系统

9.4 神经网络

9.5 深度学习

三、考核知识点与考核要求

（一）机器学习的基本概念（教材9.1、9.2节）

领会：机器学习的应用场景和应用方向，机器学习的分类，有监督和无监督的区别、强化学习的原理

（二）推荐系统（教材9.3节）

领会：推荐系统的过滤算法

（三）神经网络（教材9.4节）

领会：人工神经元的原理、神经网络机器学习的三个基础、神经网络的局限性

（四）深度学习（教材9.5节）

领会：人工智能、机器学习、深度学习和神经网络概念的区别，基本的前馈神经网络，Sigmoid激活函数，几种常用的深度学习网络，深度学习的缺点

简单应用：全连接神经网络结构

四、本章重点、难点

本章的重点：机器学习、深度学习和神经网络的概念和原理。

本章的难点：人工神经元原理和神经网络结构。

1. 自然语言理解

一、学习目的与要求

本章的学习目的是让学习者了解自然语言理解的概念、原理和技术，并能够使用自然语言理解技术解决应用问题。

二、课程内容

10.1概述

10.2 语言认知模型

10.3 自然语言理解模型

10.4 语言及汉语的特点

10.5 思维及思维规律的特点

10.6 什么是自然语言理解

10.7 计算机识字和理解

三、考核知识点与考核要求

（一）自然语言和思维（教材10.2，10.4，10.5节）

领会：自然语言的功能模型、语言认知模型和认知模型的层次；自然语言的基本特征、汉语的特点；思维的规律、思维结构的特点

简单应用：直言推理和换位法的思维推理

（二）自然语言理解概念和模型（教材10.1，10.3，10.6节）

识记：自然语言理解的概念、语言理解的标志

领会：自然语言工程的目标、自然语言理解任务模型、自然语言层次模型原理、自然语言理解的难点

（三）计算机识字和理解（教材10.7节）

领会：人类识字的模型、计算机识字的一般原理、计算机识字后处理的方法

四、本章重点、难点

本章的重点：自然语言特点和模型，自然语言理解的概念和计算机识字。

本章的难点：计算机识字原理和实现。

1. 人工智能方法在目标检测中的应用

一、学习目的与要求

本章的学习目的是让学习者了解人工智能在目标检测领域的应用，重点介绍常见的两个系统：智能视频监控系统和人脸识别系统。

二、课程内容

11.1 智能视频监控系统

11.2 人脸识别系统

三、考核知识点与考核要求

（一）智能视频监控系统

识记：贝叶斯网络的基本概念

领会：视频监控系统框架、多视角目标识别及消失/重现目标识别方法、异常行为的描述方法

（二）人脸识别系统

简单应用：人脸识别系统流程、多视角人脸检测算法、多视角人脸跟踪算法、多任务决策融合编程架构

四、本章重点、难点

本章的重点：多智能体的概念、实现及其协商方式。

本章的难点：投票策略的计算机实现。

1. 云计算、大数据与人工智能

一、学习目的与要求

本章的学习目的是让学习者更好的理解大数据与人工智能，了解云计算、大数据与人工智能之间的关系。

二、课程内容

12.1 云计算的概念

12.2 云计算、物联网、大数据与人工智能的融合

三、考核知识点与考核要求

（一）云计算及其与大数据、人工智能的融合（教材12.1，12.2节）

领会：云计算的优势、云计算与基础架构兼容、适合进行云端迁移的用例

四、本章重点、难点

本章的重点：云计算和人工智能各技术的融合

1. 大数据背景下机器学习算法实验

一、学习目的与要求

本章的学习目的是让学习者了解机器学习的基础算法。从概念入手介绍相关算法的原理及其简单应用。

二、课程内容

13.1 距离的度量方法

13.2 机器学习算法之聚类、回归、分类

13.3 机器学习算法之决策树

三、考核知识点与考核要求

（一）距离的度量方法（教材13.1节）

简单应用：各种距离度量的计算

（二）机器学习算法之聚类、回归、分类（教材13.2节）

识记：常用的聚类算法、最常用的回归、分类合理判断的两个常用指标及其概念

领会：线性回归的原理

简单应用：K-均值聚类算法原理

（三）机器学习算法之决策树（教材13.3节）

领会：决策树算法的基本思想、信息增益和熵

综合应用：基于信息增益的决策树生成

四、本章重点、难点

本章的重点：聚类、回归和分类的基础算法

本章的难点：熵的理解和应用

**Ⅳ 关于大纲的说明与考核实施要求**

一、自学考试大纲的目的和作用

课程自学考试大纲是根据专业自学考试计划的要求，结合自学考试的特点而确定。其目的是对个人自学、社会助学和课程考试命题进行指导和规定。

课程自学考试大纲明确了课程学习的内容以及深广度，规定了课程自学考试的范围和标准。因此，它是编写自学考试教材和辅导书的依据，是社会助学组织进行自学辅导的依据，是自学者学习教材、掌握课程内容知识范围和程度的依据，也是进行自学考试命题的依据。

二、课程自学考试大纲与教材的关系

课程自学考试大纲是进行学习和考核的依据，教材给出了学习掌握课程知识的基本内容与范围，教材的内容还包括大纲所规定的课程知识的扩展与发挥。课程内容在教材中可以体现一定的深度或难度，但在大纲中对考核的要求一定要适当。

大纲与教材所体现的课程内容基本一致；大纲里面的课程内容和考核知识点，教材里一般也要有。反过来教材里有的内容，大纲里就不一定体现。

三、关于自学教材

《大数据与人工智能》，郏东耀主编，清华大学出版社、北京交通大学出版社，2022第1版。

本教材第四章4.6、4.8节，第五章5.5、5.6节，第八章8.3节，第十三章，第十四章，第十五章15.4节内容，考生可根据个人能力兴趣学习，不纳入考核范围。

四、关于自学要求和自学方法的指导

本大纲的课程基本要求是依据专业考试计划和专业培养目标而确定的。课程基本要求还明确了课程的基本内容，以及对基本内容掌握的程度。基本要求中的知识点构成了课程内容的主体部分。因此，课程基本内容掌握程度、课程考核知识点是高等教育自学考试考核的主要内容。

为有效地指导个人自学和社会助学，本大纲已指明了课程的重点和难点，在章节的基本要求中一般也指明了章节内容的重点和难点。

本课程共6学分。由于成人学习的个性化特点，建议业余自学时间不低于66个学时。

建议学习本课程时注意以下几点：

1.在学习本课程教材之前，应先仔细阅读本大纲，了解本课程的性质和特点，熟知本课程的基本要求，在学习本课程时，能紧紧围绕本课程的基本要求。

2.在自学每一章的教材之前，先阅读本大纲中对应章节的学习目的与要求、考核知识点与考核要求，以使在自学时做到心中有数。

3.把学习基本理论、基本知识与分析、解决实际问题结合起来。首先，要弄懂基本理论、基本原理、基本知识和基本方法；其次，要学习运用这些知识联系实际解决有关实际问题。重点是要深刻领会教材内容，将知识转化为能力，提高运用知识分析问题和解决问题的能力。

4.学习人工智能与大数据的目的是深入了解人工智能领域和大数据技术的基本概念、理论和方法，增强对人工智能、大数据和云计算融合的认识和理解，掌握机器学习算法的基本原理，了解各种人工智能技术的应用场景，提高与人工智能相关的职业素养和综合素质，为以后进一步学习和研究人工智能相关打下坚实的基础，增强在未来人工智能时代的竞争力。

五、对社会助学的要求

对担任本课程自学助学的任课教师和自学助学单位提出以下几条基本要求。

1.熟知本课程考试大纲的各项要求，熟悉各章节的考核知识点。

2.辅导教学以大纲为依据，不要随意删减内容，以免偏离大纲。

3.辅导还要注意突出重点，要帮助学生对课程内容建立一个整体的概念。

4. 助学者在辅导时应帮助自学者梳理重点内容和一般内容之间的关系，在他们全面掌握全部考试内容的基础上，深入搜集技术、人工神经网络与深度学习、计算机视觉、自然语言处理等重点内容，注意本课程概念表示、知识表示和各种研究领域核心内容的系统性。

本课程考纲内容学时建议如下：

|  |  |  |
| --- | --- | --- |
| 章次 | 学习内容 | 建议学时 |
| 第1章 | 大数据的来源、采集与基本概念 | 4 |
| 第2章 | 大数据处理构架Hadoop | 6 |
| 第3章 | Hadoop分布式文件系统HDFS | 6 |
| 第4章 | MapReduce与Spark | 6 |
| 第5章 | NoSQL数据库、分布式数据库HBase、云数据库 | 6 |
| 第6章 | 大数据的应用与展望 | 4 |
| 第7章 | 人工智能的基本概念 | 4 |
| 第8章 | 特征提取 | 6 |
| 第9章 | 机器学习 | 6 |
| 第10章 | 自然语言理解 | 4 |
| 第11章 | 人工智能方法在目标检测中的应用 | 4 |
| 第12章 | 云计算、大数据与人工智能 | 6 |
| 第13章 | 大数据背景下机器学习算法实验 | 6 |

六、对考核内容的说明

1. 本课程要求考生学习和掌握的知识点内容都作为考核的内容。课程中各章的内容均由若干知识点组成，在自学考试中成为考核知识点。因此，课程自学考试大纲中所规定的考试内容是以分解为考核知识点的方式给出的。由于各知识点在课程中的地位、作用以及知识自身的特点不同，自学考试将对各知识点分别按四个认知层次确定其考核要求。

2. 在考试之日起6个月前，由全国人民代表大会和国务院颁布或修订的法律、法规都将列入相应课程的考试范围。凡大纲、教材内容与现行法律、法规不符的，应以现行法律法规为准。命题时也会对我国经济建设和科技文化发展的重大方针政策的变化予以体现。

七、关于考试命题的若干规定

1.本课程考试采用闭卷笔试形式,考试时间为150分钟；满分100分,60分及格。

2.本大纲各章所规定的基本要求、知识点及知识点下的知识细目，都属于考核的内容。考试命题既要覆盖到章，又要避免面面俱到。要注意突出课程的重点、章节重点，加大重点内容的覆盖度。

3.不应命制超出大纲中考核知识点范围的题目，考核目标不得高于大纲中所规定的相应的最高能力层次要求。命题应着重考核自学者对基本概念、基本知识和基本理论是否了解或掌握，对基本方法是否会用或熟练。不应命制与基本要求不符的偏题或怪题。

4.本课程在试卷中对不同能力层次要求的分数比例大致为：识记占20%，领会占30%，简单应用占30%，综合应用占20%。

5.试题的难易程度分为4个等级：易、较易、较难和难四个等级。每份试卷中不同难度试题的分数比例一般为：易占20%，较易占30%，较难占30%，难占20%。

必须注意试题的难易程度与能力层次有一定的联系，但两者不是等同的概念，在各个能力层次都有不同难度的试题。

6.各种题型的具体样式参见本大纲附录。

**附录 题型举例**

一、单项选择题（在每小题后的4个备选项中只有一个是符合题目要求的，请将其选出并填写在题后的括号内）

1.（ ）就是在已知输入和输出的情况下训练出一个模型，将输入映射到输出。

1. 监督学习 B.无监督学习

C. 聚类 D.专家系统

二、名词解释题

1.机器学习

三、简答题

1.美国认知心理学家G.M.Olson提出的语言理解的标志是什么？

四、综合分析题

1.以表所示的数据集作为训练数据，请构造一棵决策树用于判断水果的类别并画出来。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 颜色 | 形状 | 大小 | 类别 |
| 1 | 红 | 球 | 一般 | 苹果 |
| 2 | 黄 | 弯月 | 一般 | 香蕉 |
| 3 | 红 | 球 | 轻 | 樱桃 |
| 4 | 绿 | 椭球 | 重 | 西瓜 |
| 5 | 橘 | 球 | 一般 | 橘子 |