

# 第 1 章 信息与计算机技术基础

步入 21 世纪，人类已生活在信息科技时代，信息成为经济与社会发展的主导因素，与物质、能源并称为支撑人类社会生存与发展的三大资源。“利用身外之物扩展自身能力”是人类文明进化的主要推动力，也是科学技术产生和发展的根本原因，因此对信息进行处理的手段与工具的更迭也伴随和引导着人类文明的发展与变革，其中，计算机的发明与普及极大地推动了这一进程，计算机以其高效、准确以及智能化的信息处理能力，成为当前在各行各业中渗透最深入的一种信息处理工具。了解以计算机为核心的信息技术基础知识和掌握信息技术的基本应用是当代大学生必须具备的基本素质。

本章主要内容包括信息概述、计算机概述、人工智能概述、计算机中数的表示和计算思维概述。

## 1.1 信息概述

“信息”是现代社会中广泛使用的一个概念，不同领域的科学家对“信息”的定义提出了不同角度和层次的阐释，至今关于信息的定义达 200 多种。信息论创始人，美国数学家香农基于在 1948 年创立的“通信的数学理论”，以通信系统为背景，提出：信息是能够用来消除不确定性的因素。控制论奠基人，美国数学家维纳认为：① 信息是组织程度的度量；② 信息是我们适应外部世界并使这种适应为外部世界所感知的过程中，同外部世界进行交换的内容的名称，这种信息不能脱离接收者的解释结构而存在。

我国著名信息科学家钟义信结合时代发展中信息概念边界的不断扩大，将原来各自独立研究的基于传感、通信、计算、智能、控制学科的信息定义高度统一为“全信息”理论，并以这一理论为基础，系统地研究和总结了全信息在整个信息处理过程的工作原理。本章将以钟义信提出的“全信息”理论为界定，阐述信息的明确定义、特征与价值。

### 1.1.1 信息的概念和特征

“全信息”理论中，对信息做了如下定义：作为主体（人和生物）所感知或所表述的“事物的运动状态及其变化方式”就是信息；信息不同于物质也不同于能量，信息具有巨大的价值。

信息具备如下主要特征。

#### 1. 载体的依附性

信息可以存在并负载于载体上，可以被无限地进行复制、传播。语言、数据、文字、



图像、声音、编码都是信息的载体。

### 2. 价值的相对性

拥有不同观察能力、不同理解能力和不同目的的观察者，从同一事物所获得的信息各不相同。

### 3. 价值的时效性

信息是事物的运动状态及其变化方式，当信息不再能反映事物变化后新的状态时，它的效用就会逐渐降低，直至完全失去。

## 1.1.2 信息科学

信息科学是以信息为研究对象、以信息运动规律为研究内容的一门学科，其研究目标是如何以科学的方法扩展人的智能、扩展人类认识世界并改造世界的能力。信息科学是在概括和综合多门学科的基础上形成的新兴横断科学，信息科学最基本的原理是：信息—知识—智能转换原理，如图 1-1 所示。

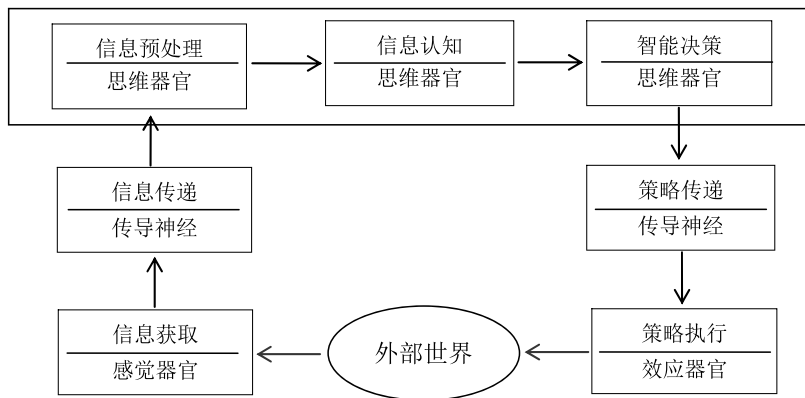


图 1-1 信息—知识—智能转换示意图

## 1.1.3 信息技术

技术是人类在认识和改造自然的过程中为了增强与扩展自身力量赢得更广阔的生存机会而产生与发展起来的，其中，信息技术是能够扩展人的信息器官功能的一类技术。而人的信息器官主要包括以下四类：

(1) 感觉器官。包括视觉器官、听觉器官、嗅觉器官、味觉器官、触觉器官和平衡感觉器官等。

(2) 传导神经网络。可以分为导入神经网络和导出神经网络等。

(3) 思维器官。包括记忆系统、联想系统、分析推理和决策系统等。

(4) 效应器官。包括操作器官（手）、行走器官（脚）和语言器官（口）等。



四类信息器官的主要功能如下：

(1) 感觉器官：获取信息——通过视觉、听觉、嗅觉、味觉和触觉感知（即获取）外部世界各种事物运动的状态和方式；平衡感觉器官则可以根据运动主体与客体之间的关系获取平衡信息。

(2) 传导神经网络：传递信息——通过导入神经网络把感觉器官获得的信息传递给思维器官；又通过导出神经网络把思维器官加工出来的信息传递给各种效应器官或某些内部器官。

(3) 思维器官：加工和再生信息——实际上它担负着存储信息、检索信息、加工信息和再生信息的复杂任务。通过存储、检索、加工信息得到对于外部事物运动规律的认识，通过再生信息表示主体对于外部事物进行改造的策略和意图。

(4) 效应器官：施用信息——通过操作器官和行走器官执行大脑发出的再生信息，或者通过语言器官表达再生信息，使这些信息产生实际的效果。

人类的这四类信息器官和它们的信息功能有机地联系在一起，使它们能够执行整体性的高级功能——用于认识世界和改造世界的过程中所需要的智力功能。图 1-2 为这四种信息器官互相联系形成智力功能的示意图。

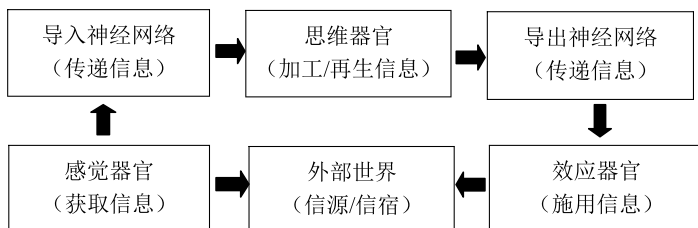


图 1-2 信息器官及其功能示意图

图 1-1 中，实现信息获取功能的技术对应“传感技术”，实现信息传递和策略传递的功能对应“通信技术”，实现信息预处理的功能对应“计算机技术和存储技术”，实现信息认知功能和智能决策的功能对应“人工智能技术”，实现策略执行的功能对应“控制技术”，实现反馈学习和策略优化的功能对应“信息系统自组织技术”，将以上过程进行概括归纳后得出信息科学的简化模型，如图 1-3 所示。

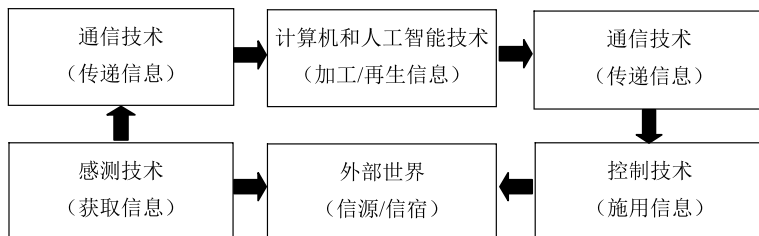


图 1-3 信息科学的简化模型图

从信息科学简化模型可知，扩展感觉器官功能的传感技术、扩展传导神经系统的通信技术、扩展思维器官预处理功能的计算机技术以及扩展效应器官功能的控制技术为外围技术，扩展思维器官认知功能和决策功能的人工智能技术为核心技术。



根据信息技术的定义和以上相应分析，得出信息技术的四项基本内容，即：

### 1. 感测技术

感测技术是感觉器官功能的延长。包括传感技术、测量技术、遥感技术、遥测技术等。它使人们能更好地从外部世界获得各种有用的信息。

### 2. 通信技术

通信技术是传导神经网络功能的延长。它的作用是传递、交换和分配信息，消除或克服空间上的限制，使人们能更有效地利用信息资源。

### 3. 计算机和智能技术

计算机和智能技术是思维器官功能的延长。包括计算机硬件技术、软件技术和人工智能技术，使人们能更好地加工和再生信息。

### 4. 控制技术

控制技术是效应器官功能的延长。其作用是根据输入的指令（决策信息）对外部事物的运动状态实施干预，即施用信息。

综上信息技术的定义如下：

- (1) 信息技术是指能够完成信息获取、传递、加工、再生和施用等功能的技术。
- (2) 信息技术是感测、通信、计算机和智能以及控制等技术有机结合的整体。

## 1.1.4 医学信息学

医学信息学又称卫生信息学或医学资讯学，是信息科学、医学和卫生保健学等学科的交叉学科。医学信息学的工具不仅包含计算机，还包括临床指导原则、医疗术语和通信系统。

医学信息学的覆盖领域包括临床信息学、临床研究信息学、护理信息学、口腔医学信息学、公共卫生信息学、医学图形信息学、健康信息学、生物信息学和药物信息学。

医学信息学的具体应用包括以下几个方面：

### 1. 医学信息标准的制定

医学领域的数据往往由不同的医疗、科研、管理机构和不同企业、个体产生，要实现跨部门、跨机构、跨场景的医学信息互联互通、数据共享和业务协同，就需要信息学专业人员联合相关机构制定相关医学信息标准，并由公认机构批准和发布，作为共同遵守、可重复使用的规范化准则和依据。例如，医疗信息编码标准制定了各种诊断、手术、药品、医疗器械等编码规范，为医疗信息数据的标识、查询、比对、统计和分析等提供了基础保障。目前，国际上流行的医疗信息数据标准主要有 ICD、SNOMED CT、DICOM、LOINC、HL7 等，国内也有自主研发的医疗信息数据标准，如中国医学科学院医学信息学术语（简称 CMA）、国家中医药管理局中医病证分类标准等。



## 2. 数据治理、管理与分析

在基础医学、临床医学、公共卫生、中医药学等领域的科学研究和业务活动中产生了大量原始性、基础性的医学数据。这就需要医学信息学专业人员和建立和维护数据管理机制，获取合理合法可用的健康医疗相关数据，进而用于数据分析挖掘以支持个体和人口健康并推动医学科技创新。

## 3. 促进卫生决策、进程和结果

医学信息学的主要目的是将所处理的数据、信息和知识有效地应用于临床诊疗实践、公共卫生决策、公众健康管理、医学科研与医药研发、医学科技评价与战略决策等方面，从而支持和促进临床医生、公共卫生专业人员、医学科研人员、患者和公众、管理者的决策，以优化流程、提高效率、促进创新。

# 1.2 计算机概述

计算机是一种能迅速而高效地自动完成信息处理的电子设备，它能按照程序对信息进行加工、处理和存储。在上一节中，我们知道计算机科学与技术是信息科学中重要的一个分支与支撑，本节将对计算机的产生、发展、特点、分类、应用和发展趋势等方面进行介绍。

## 1.2.1 计算机的发展历史

世界上第一台电子计算机由美国宾夕法尼亚大学、穆尔工学院和美国陆军火炮公司联合研制而成，于1946年2月15日正式投入运行，它叫ENIAC，是The Electronic Numerical Integrator and Calculator（电子数值积分计算机）的缩写。ENIAC使用了17468个真空电子管，耗电174千瓦、占地170平方米，重达30吨，每秒可进行5000次加法运算，3毫秒便可进行一次乘法运算。虽然它的功能还比不上今天最普通的一台微型计算机，但在当时它的运算速度、精确度和准确度也是历史上的计算工具无法比拟的，它将60秒射程的弹道计算时间由原来的20分钟缩短到30秒。ENIAC奠定了电子计算机的发展基础，开辟了一个计算机科学技术的新纪元。

ENIAC诞生后短短的几十年间计算机的发展突飞猛进。主要电子器件相继使用了真空电子管，晶体管，中、小规模集成电路和大规模、超大规模集成电路，带动计算机屡次更新换代。每一次更新换代都使计算机的体积和耗电量极大减小，功能极大增强，应用领域进一步拓宽。特别是体积小、价格低、功能强的微型计算机的出现使得计算机迅速进入了办公领域和家庭，在办公自动化和多媒体应用方面发挥了巨大的作用。目前计算机的应用已扩展到社会的各个领域。计算机的发展如下所述。

### 1. 第一代：电子管计算机（1946—1957年）

主要特点：

- (1) 采用电子管作为基本逻辑部件，体积大，耗电量大，寿命短，成本高。



(2) 采用电子射线管作为存储部件, 因容量太小, 后外存储器使用了磁鼓存储信息扩充容量。

(3) 输入输出装置落后, 主要使用穿孔卡片, 速度慢且容易出错, 使用十分不便。

(4) 没有系统软件, 只能用机器语言和汇编语言编程。

(5) 主要服务对象是军事应用, 包括导弹计算和与军事相关的空间的计算等。

## 2. 第二代: 晶体管计算机 (1958—1964 年)

随着半导体技术的发展, 20 世纪 50 年代中期晶体管取代了电子管, 晶体管计算机的体积缩小为电子管计算机的 1% 左右, 耗电量也只有电子管计算机的 1%, 但它的运算速度提高到每秒几万次。其主要特点如下:

(1) 采用晶体管作为基本逻辑元件, 体积减小, 重量减轻, 能耗降低, 成本下降, 计算机的可靠性和运算速度均得到提高。

(2) 普遍采用磁芯作为主存储器, 采用磁盘和磁鼓作为外存储器。

(3) 开始有了系统软件 (监控程序), 提出了操作系统概念, 并出现了高级语言。

## 3. 第三代: 集成电路计算机 (1965—1970 年)

1962 年, 世界上第一块集成电路诞生, 在一块只有 2.5 平方英寸<sup>①</sup>的硅片上集成了几十至几百个晶体管。计算机的体积进一步缩小, 运算速度可达每秒几百万次。其主要特点如下:

(1) 采用中小规模集成电路制作各种逻辑部件, 从而使计算机体积更小、更轻、耗电更少、寿命更长、成本更低, 运算速度得到大幅提高。

(2) 采用半导体存储器作为主存储器, 取代了原有的磁芯存储器, 使存储器的存取速度大幅提升。

(3) 系统软件取得长足发展, 出现了分时操作系统, 多用户可以共享计算机软硬件资源。

(4) 在程序设计方面采用了结构化程序设计, 为研制更加复杂的软件提供了技术支持。

## 4. 第四代: 大规模、超大规模集成电路计算机 (1971 年至今)

1971 年 Intel 公司的工程师们把计算机的算术与逻辑运算电路集成在一片微小的硅片上, 做成了世界上第一片微处理器 (Intel 4004), 该硅片上集成了 2250 只晶体管, 从此, 掀起信息革命浪潮的微型电子计算机 (简称“微型计算机”) 诞生。1983 年, 美国《时代周刊》第一期选出的 1982 年“时代风云人物”是 1 台微型计算机。微型计算机的体积小, 易于搬动, 运算速度达每秒上亿次, 正是我们目前正在普遍使用的一代计算机。其主要特点如下:

(1) 基本逻辑部件采用超大规模集成电路, 使计算机体积、重量、成本均大幅度降低, 出现了微型计算机。

(2) 作为主存的半导体存储器集成度越来越高, 容量越来越大; 外存储器除了广泛使

<sup>①</sup> 1 平方英寸=6.4516 平方厘米。



用软硬磁盘，还引进了光盘等移动存储器。

(3) 各种使用方便的输入/输出设备相继出现。

(4) 软件产业高度发达，各种实用软件层出不穷，极大地方便了用户应用。

微型计算机普及后，计算机技术与通信技术相结合形成的计算机网络把世界紧密联系在一起。随着多媒体技术崛起，计算机集图像、图形、声音和文字处理于一体，在信息处理领域掀起了一场革命，与之对应的信息高速公路在各国紧锣密鼓地筹划和实施。

## 1.2.2 计算机的主要特点

计算机的主要特点如下：

### 1. 运算速度快

计算机内部承担运算的部件叫作运算器，目前大多数个人计算机的 CPU 运算速度大约已达每秒 50 亿次，超级计算机可达每秒千万亿次。为保证运算结果的可用性，计算机控制导航要求“运算速度比飞机飞得还快”，气象预报需分析大量资料，运算速度必须跟上天气变化，否则便会失去“预报”的意义。

### 2. 计算精度高

计算机的计算精度是其他计算工具无法比拟的。利用计算机可以计算出精确到小数点后 200 万位的圆周率值。

### 3. 存储容量大

在计算机中承担存储职能的部件称为存储器。计算机可以将信息存储在存储器中，我国四大名著之一的《红楼梦》含标点共 87 万字，1GB 的空间可以存储约 671 部《红楼梦》。

### 4. 具有逻辑判断能力

计算机以逻辑运算进行逻辑判断，它能够分析命题是否成立，并且可以根据命题的成立与否采取相应的对策。逻辑运算是离散数学中重要的部分，因此离散数学也是深入了解计算机工作原理的基础。数学中有一个“四色问题”，该理论认为不论多么复杂的地图，如果想要使相邻区域颜色不同，那么最多只需 4 种颜色就够了。很久以前，不少数学家一直想去证明它或推翻它，但是一直没有成功，这个问题也就成了数学中著名的难题。直到 1976 年，两位美国数学家借助电子计算机进行了非常复杂的逻辑推理验证，从而使这个困扰了数学家们近一百年的问题终于被解决。

### 5. 可靠性高

计算机硬件故障通常是由元器件失效引起的，元器件越多，容错性就越好。计算机硬件中拥有上亿的元器件，为预防出现质量问题、提高系统稳定性与可靠性提供了巨大的保障。

### 6. 通用性强

现代计算机不仅可以进行科学计算也可用于数据处理、实时控制、辅助设计和辅助制



造、办公自动化等领域，具备渗透各行各业的通用性。

### 1.2.3 计算机的分类

计算机的高速度、高可靠性和高精度度，以及其所具有的海量存储信息的能力，在各领域得到了广泛应用。根据其用途不同，计算机可分为通用机和专用机两类。通用机能解决多种类型的问题，通用性强；而专用机则配有解决特定问题的软、硬件，功能单一，但能高速、可靠地解决特定问题。

通常，按照计算机的运算速度、字长、存储容量、软件配置及用途等多方面的综合性指标，将计算机分为巨型机、大型机、小型机、微型机、工作站、服务器。

#### 1. 巨型机

巨型机是计算机中运算速度最快，性能最高，实现技术最复杂的机型，主要用于尖端科学研究领域。反导弹武器、空间技术、大范围天气预报、石油勘探和地震预测等都要求计算机有很高的运算速度和巨大的存储容量，因此一些国家竞相投入巨资开发速度更快、性能更强的超级计算机。巨型机的研制水平、生产能力及其应用程度已成为衡量一个国家经济实力和科技水平的重要标志。

1983年12月26日，我国第一台命名为“银河-1”的亿次巨型计算机通过国家技术鉴定。它的成功研制标志着中国成为极少数能独立设计和制造巨型机的国家之一。

#### 2. 大型机

从核心而言，大型机是高性能的计算机，具有大量内存和处理器，能够实时处理以十亿计的计算和事务。大型机需要具备高度灵活性、安全性和及时响应能力以适用于商用数据库、事务服务器。

#### 3. 小型机

小型机功能和价格介于微型机和大型机之间。小型机用途广泛，既可用于科学计算、数据处理，也可用于生产过程中的自动控制和数据采集及分析处理。

#### 4. 微型机

以微处理器为中央处理单元而组成的个人计算机简称微型机或微机，英文简称为PC。20世纪70年代，微型计算机的出现引发了计算机软硬件领域的一场革命。它既可以用于日常信息处理，又可以用于科学研究。微型计算机的出现使得计算机真正面向全人类，成为大众化的信息处理工具。

#### 5. 工作站

工作站是一种高档微型计算机系统。它具有较高的运算速度，既具有大、中、小型机的多任务、多用户能力，又具有微型机的操作便利和良好的人机界面。工作站可连接多种输入/输出设备，而其最突出的特点是图形功能强，具有很强的图形交互与处理能力，因此



在工程领域，如计算机辅助设计 CAD 领域得到广泛应用。通常认为工作站是专为工程师设计的机型。

## 6. 服务器

服务器指在计算机网络中提供服务和资源的计算机，通常需要具有高速的 CPU 运算能力、长时间可靠运行以及强大的输入/输出能力。

## 1.2.4 计算机的应用

如今计算机的应用已广泛而深入地渗透到人类社会的各个领域。科研、生产、国防、文化、教育、医疗、生活都离不开计算机，其应用领域可归纳为以下六类。

### 1. 科学计算

强大的计算能力正是计算机的特长，在自然科学中，诸如数学、物理、化学、天文、地理等领域；在工程技术中，诸如航天、汽车、造船、建筑等领域；计算工作量巨大，均需要计算机工作提供算力支持。2015年5月，北京师范大学天文系教授张同杰带领的宇宙中微子数值模拟团队通过“天河二号”超级计算机系统完成了3万亿粒子数的宇宙中微子和暗物质数值模拟，宇宙大爆炸1600万年之后至今约137亿年的漫长演化进程得以被揭示。2020年，新型冠状病毒感染时，研究人员借助超级计算机的演算，发现其会让人体停止分解缓激肽，然后患者的肺部会被液体充满，进而引发炎症并带来呼吸困难。这一发现给疫情联防联控以及精准施策带来了帮助。

### 2. 数据处理

数据处理指用计算机收集、记录数据，经加工产生新的信息形式的技术，数据指数字、符号、字母和各种文字的集合。数据处理涉及的加工处理比一般的算术运算广泛得多，这类工作是当今计算机应用的主流。

### 3. 过程控制

过程控制又称实时控制，指用计算机及时采集检测数据，并在一定条件下迅速地对控制对象进行自动控制或自动调节，目前已在冶金、石油、化工、纺织、水电、机械和航天等领域得到广泛应用。微型计算机控制系统除了应用于工业生产，还广泛应用于交通、邮电、卫星通信等领域。

### 4. 多媒体技术

多媒体技术是应用计算机技术将文字、图像、图形和声音等信息以数字化的方式进行综合处理的技术。

### 5. 计算机通信

计算机通信是计算机技术和通信技术相结合的产物，计算机网络技术将不同地域的计



计算机通过通信线路连接起来，在网络软件的支持下达到资源共享的目的。多媒体技术的发展将计算机通信由简单的文字数据通信扩展到音频、视频图像的通信，使诸如远程会议、远程医疗、远程授课、电子商务等网上通信活动扩充和改变了人类进行交流的方式和途径。

### 6. 人工智能

人工智能的主要目的是用计算机模拟人类的智能，可用于机器人专家系统、推理证明、多轮对话等方面。

## 1.2.5 计算机在医学领域的应用

在医学领域，以计算机为核心构建的系统和应用如下：

### 1. 医院信息系统

医院信息系统是指利用计算机软硬件技术和网络通信技术等现代化手段，对医院及其所属各部门的人流、物流、财流进行综合管理，对在医疗活动各阶段产生的数据进行采集、存储、处理、提取、传输、汇总，并加工形成各种信息，从而为医院的整体运行提供全面的自动化管理及各种服务的信息系统。是否具备功能完善的医院信息系统已成为衡量医疗机构是否具备良好服务和先进水平的重要标准。

### 2. 医学图形图像处理

医学图形图像是反映人体内部结构的图像，是现代医疗诊断的主要依据之一。主要技术包括图像检测、图像分割、图像配准和图像融合技术，通过对图像进行分析和处理，实现对人体器官、软组织和病变体的位置检测、分割提取、三维重建和三维显示，进行定性分析与定量分析，从而提高临床诊断的效率、准确性和可靠性。临床广泛使用的医学成像种类主要有：X-射线成像(X-CT)、计算机断层扫描(CT)、正电子发射计算机断层显像(PET-CT)、核磁共振成像(MRI)、核医学成像(NMI)、超声波成像(UI)、显微镜下拍摄病理图像等。

### 3. 计算机辅助药物设计

新药研发是医药健康领域的重要部分，新药上市需经历临床前研究及临床 I、II、III 期研究后才能最终上市，低效率和高成本是新药研发的主要障碍。同时处理来自基因组学、蛋白质组学和临床试验中大量复杂的数据也为新药的研发带来了很大的挑战。依托人工智能技术的计算机辅助药物设计是近年来兴起的药物研发新手段，它借助计算机技术和人工智能算法对海量生物数据以及化学数据进行处理，以更高效地发现潜在的有效药物。

### 4. 智能诊断

智能诊断是当今健康医药的一个重要研究领域，目的是通过非人工的方式干预疾病诊断，提升患者的治疗效率。智能诊断主要涵盖临床虚拟助手、辅助诊断、疾病风险预测等方面，力求帮助提高医生的工作效率和减少人为失误。其中，临床虚拟助手的原理为将各



种病情的诊断标准、阈值判断、治疗处方、专家经验等以数据库的形式存储在计算机内，利用计算机强大的计算和逻辑推导能力模拟医生的工作过程，从而达到精确诊断与决策目的，可以在很大程度上解决临床医生知识的局限性问题，减少人为疏漏，提升医疗效率与药物的使用效率。

## 1.2.6 计算机的发展趋势

英特尔（Intel）创始人之一戈登·摩尔（Gordon Moore）1965年提出“摩尔定律”，其内容为当价格不变时，集成电路的硅片上可容纳的晶体管数目，约每隔18个月便会增加一倍，性能也将提升一倍。换言之，每一美元所能买到的电脑性能，将每隔18个月翻两倍以上。这一定律揭示了信息技术进步的速度。在摩尔定律发现后10年，1975年摩尔又做了一些修改，将翻番的时间调整为24个月。

硅基技术的进步都在遵循摩尔定律，每两年将处理速度和内存容量提高一倍，随着晶体管不断逼近其尺寸的极限、新材质技术的不断突破和新型社会需求的产生，计算机还向以下四个方面发展：

### 1. 生物计算机

生物计算机是以蛋白分子替代硅，生物芯片上集成的元件密度比人的神经密度还要高100万倍，传递信息的速度也比人脑思维速度快100万倍。生物计算机传递信息时阻抗小，能耗仅为现代计算机的十亿分之一，由于具备生物特点，可进行自我组织和自我修复。

### 2. 量子计算机

量子计算机是一类遵循量子力学规律进行高速数学和逻辑运算、存储及处理量子信息的物理装置。量子计算机由于具有不可克隆的量子原理，因此在信息存储和处理方面具备较高的安全性。在新药研发领域，量子计算机有着巨大的价值，它能描绘出万亿计的分子组成，并且可选择出其中最有可能的方法，这将提高人们发明新型药物的速度，并且能够更个性化地对药理进行分析。

2023年5月，中国科学院院士、中科院量子信息与量子科技创新研究院院长、中国科学技术大学常务副校长潘建伟发表视频演讲时表示，我国科学家已经实现了255个光子的九章3号计算原型机，它针对特定问题的求解能力已经比经典的超级计算机快1000万亿倍。

### 3. 光计算机

光计算机是由光代替电子或电流，实现高速处理大容量信息的计算机。它以不同波长的光代表不同的数据，以大量的透镜、棱镜和反射镜将数据从一个芯片传递到另一个芯片，运算速度在理论上可达每秒千亿次以上。当电子通过晶体管和其他传统集成电路元件时，会遇到阻力并产生热量，而光子芯片不存在电阻问题，因为由激光产生的光子能快速通过波导、调制器、反射器等原件阵列，因此能耗极低，只有同等级电子芯片的六分之一，光计算机逐渐被研发及应用于人工智能、自动驾驶汽车、量子计算等领域。



## 1.2.7 计算机历史上的女性

长久以来人们普遍认为以计算机为代表的理工类学科是男性的主场，而女性更多地从事于人文社科领域。事实上，女性科学家同样为计算机的发展做出了不可磨灭的贡献。本节主要介绍一些对计算机发展有重大影响的女性科学家，以期鼓励更多女性学习者勇敢地探索计算机世界。

### 1. 第一位程序员：艾达·洛夫莱斯

艾达·洛夫莱斯（Ada Lovelace，1815—1852 年）如图 1-4 所示。她十多岁时结识了后世誉为电脑之父的查尔斯·巴贝奇（Charles Babbage，1791—1871 年），后者提出了差分机和分析机的概念及原型，艾达发明了一种通过分析机计算伯努利数的算法，被认为是世界上第一个公开发表的计算机程序，她也因此被认为是世界上第一位程序员。并且，她富有远见地认为分析机不仅可以用于计算，还可以处理一切能加以抽象并用指令描述的任务，洞见了现代计算机的通用性。



图 1-4 艾达·洛夫莱斯的水彩肖像

为了纪念她，NVIDIA 以 Ada 命名了在 2022 年发布的图形处理器微架构，该架构用于打造出色的游戏创作、专业图形、AI 和计算性能，被 NVIDIA GeForce 40 系列显卡采用。

### 2. “不可思议的格蕾丝”：格蕾丝·霍珀

格蕾丝·霍珀（Grace Hopper，1906—1992 年）如图 1-5 所示。1949 年，格蕾丝·霍珀参与了开发通用自动计算机 1 号（UNIVAC I）的工作，这是世界上第一台商用电子计算机。当时人们用机器语言编写程序，而她认为可以用接近英文写作的方式编写程序，即高级编程语言。两年后，她在这款计算机上开发出了世界上第一个编译器。COBOL（通用商业语言，common business oriented language）是世界上第一个高级编程语言，它大量采用了霍珀的思想，因此霍珀也被称为“COBOL 之母”。在开发马克一号期间，为了节约存储空间，霍珀各用两位数来存储年、月、日，这个方法也延续到了 COBOL 语言中，也被多数编程语言、操作系统和应用软件采纳，从而导致了“千年虫”问题（只用两位数无法区分 2000 年和 1900 年）。20 世纪 70 年代，霍珀又提出了分布式网络。



图 1-5 格蕾丝·霍珀

霍珀曾给过年轻人一条寄语：“停在港口的船很安全，但这不是我们造船的目的。”



### 3. 中国计算机之母：夏培肃

夏培肃（1923—2014年）如图1-6所示。夏培肃一生强调自主创新在科研工作中的重要性，坚持做中国自己的计算机。1958年，她负责设计研制通用电子数字计算机107计算机，如图1-7所示，这是中国第一台自行研制的通用电子数字计算机。从20世纪60年代开始，夏培肃在高速计算机的研究和设计方面做出了系统的创造性的成果，解决了数字信号在大型高速计算机中传输的关键问题。她负责设计研制的高速阵列处理机使石油勘探中的常规地震资料处理速度提高了10倍以上，为中国石油勘探做出了重大贡献。她提出了研制功能分布式计算机的设想，领导团队成功研制了GF10系列计算机，还提出了最大时间差流水线设计原则，根据这个原则设计的向量处理机的运算速度比当时国内向量处理机快4倍。多年来，夏培肃还负责设计、研制成功多台不同类型的并行计算机。2002年，中科院计算所自主研发了一款通用CPU“龙芯一号”，这是我国首枚拥有自主知识产权的通用高性能微处理芯片，龙芯系列芯片后来被广泛应用于北斗系统，“龙芯一号”也被称为“夏50”，“夏”就是向夏培肃的致敬。



图1-6 夏培肃

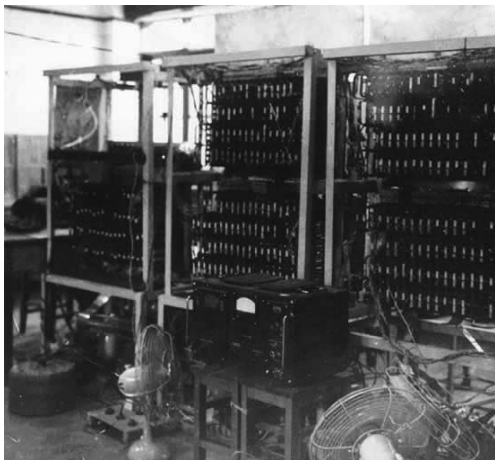


图1-7 107计算机

### 4. 中国龙芯之母：黄令仪

黄令仪（1936—2023年）如图1-8所示。1960年，黄令仪创建了国内首个半导体实验室并研发出了我国的半导体二极管，1965年，为了突破“两弹一星”中的瓶颈，国家成立了中国科学院计算技术研究所二部专攻微型计算机和三极管等项目，这是中国第一个芯片研究团队，黄令仪呕心沥血，带领团队成功研制出了半导体三极管。2001年，65岁的黄令仪已进入退休生活，此时，中科院胡伟武教授（夏培肃的学生）向全国发出了打造中国芯的集结令，黄令仪受邀毅然选择加入龙芯研发团队并成为项目负责人，2002年8月10日我国首款通用CPU“龙芯1号”研制成功，走出国产计算机无芯可用的历史困境。在黄令仪与一众芯片科学同行的不懈努力之下，“龙芯3号”（见图1-9）等一大批国产高性能芯片应运而生，使得众多重大领域（如复兴号高铁）实现了百分之百的国产化。



图 1-8 黄令仪

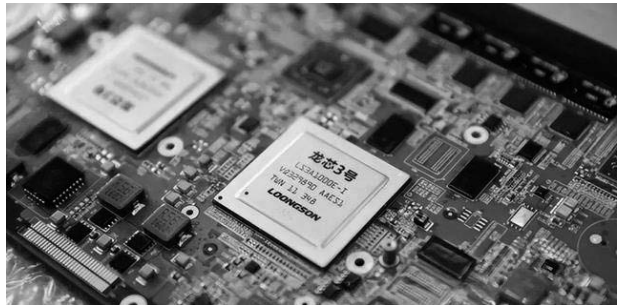


图 1-9 “龙芯 3 号” 芯片

## 1.3 人工智能概述

最初的人工智能研究是 20 世纪 30 年代末到 50 年代初的一系列科学发展交汇的产物，在 20 世纪 40 年代和 50 年代，来自数学、心理学、工程学、经济学和政治学等不同领域的科学家们开始探讨制造人工大脑的可能性。1956 年，麦卡锡、明斯基等科学家在美国达特茅斯学院开会研讨“如何用机器模拟人类的智能”，首次提出“人工智能”（artificial intelligence, AI）这一概念，标志着人工智能学科的诞生。

人工智能是研究和开发能够模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门新的技术科学，研究目的是促使智能机器会听（语音识别、机器翻译等）、会看（图像识别、文字识别等）、会说（语音合成、人机对话等）、会思考（人机对弈、定理证明等）、会学习（机器学习、知识表示等）、会行动（机器人、自动驾驶汽车等）。

2017 年我国印发了《新一代人工智能发展规划》，明确了我国新一代人工智能发展的战略目标：到 2020 年，人工智能总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点，人工智能技术应用成为改善民生的新途径；到 2025 年，人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为我国产业升级和经济转型的主要动力，智能社会建设取得积极进展；到 2030 年，人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心。

### 1.3.1 人工智能的分类

从可应用性看，人工智能可分为专用人工智能和通用人工智能。面向特定任务的人工智能称为专用人工智能。

从发展程度的角度看，人工智能可以分为弱人工智能、强人工智能和超人工智能。



### 1. 弱人工智能

弱人工智能（weak artificial intelligence, weak AI）。是实现部分思维的人工智能，仅针对某项特定任务，如：2016—2017年，AlphaGo在围棋比赛中战胜人类冠军；2020年英国《自然》杂志发表的一项研究报告了一款特殊的人工智能AI系统检测乳腺癌的能力甚至超过了人类放射科专家。但它的智能也只局限于特定方向，无法用于处理其他问题，因此称为弱人工智能。

### 2. 强人工智能

强人工智能（strong artificial intelligence, strong AI）。是指在各个方面都可以与人类智能相比肩的人工智能。

### 3. 超人工智能

超人工智能（artificial super intelligence, ASI）。其基本在所有领域都比人类智能更强，包括社交能力、科技创新、艺术设计能力等。

## 1.3.2 人工智能在医学中的应用

### 1. 计算机视觉

计算机视觉是指使用计算机及相关设备对生物视觉进行模拟，通过采集的图像识别物体信息，并根据此信息做相应的判断和决策。目前该技术被广泛应用于医疗影像诊断中，快速识别并标定异常结构，为医生提供参考，提高诊断效率。计算机视觉依赖于深度学习技术，而深度学习强大的识别能力则依赖于庞大的、可供训练的数据集。

### 2. 自然语言处理

自然语言处理也在生物医药领域得到了广泛的应用，自然语言处理是使用自然语言与计算机进行通信的技术，主要目标是使计算机能够理解人类的语言与符号，进一步通过语言与符号传递信息。目前，基于自然语言模型，利用在线问诊平台积累的大量问诊数据构建医学智能问答系统，能够解决常见患者的医学问题；通过对话摘要技术从对话中提取信息，能够辅助医生完成电子病历，从而减轻医生的工作负担。

### 3. 医疗机器人

人工智能和机器人技术的不断发展催生出了医疗机器人新兴产业。基于机器人的不同功能和应用领域，它们可以被划分为多个类别，如手术机器人、康复机器人、服务机器人等。其中，占比规模最大的是手术机器人，医生通过显示屏和内窥镜仔细观察患者体内的病灶情况，通过机器人手中的手术刀将病灶精确切除。但机器人手术若出现意外结果，则难以进行责任判定，这也困扰并限制了机器人在医疗领域的应用。



### 1.3.3 人工智能的伦理与争议

#### 1. 技术奇点与人工超级智能

技术奇点指的是一个特定的时间点，在该时间点上，技术的发展已经累积了足够的量变，并引发飞跃式的质变。对于人工智能而言，技术奇点将是人工超级智能的诞生。尽管目前人工智能的发展还相对有限，我们无法判断人工超级智能是否以及何时会诞生，但人们已经开始担忧，在人工智能加持的环境下，人类是否仍能始终保持主导权。这类担忧的极点则为人工智能叛变（AI takeover），即具备高度发达的人工智能程序、计算机或机器人产生了自我意识，向人类发起叛变，最终完全奴役或彻底灭绝人类。

#### 2. 与人类的竞争及技术性失业

从理论上来说，机器学习不可能做到百分之百准确，但在生产环境中，任何不确定和出错的风险都是难以忍受的，这将导致两个问题：一是人们是否能完全将控制权交给人工智能；二是当人工智能的判断与人类的判断相左时，究竟应以哪一方为准。特别是在机器人医疗、自动驾驶等与生命安全密切相关的行业，确权与认责仍旧存在巨大争议。

另一方面，人工智能在特定生产任务中已取得不错的成果，如生产线上的残次品鉴别。与人类相比，人工智能可以全年无休地工作、长期成本低，且不存在长期重复性工作导致的疲劳，诸如此类的劳动密集型岗位将最先面临人工智能导致的技术性失业冲击。在过去人们认为人工智能难以取代的创意行业中，伴随着大语言模型（large language model, LLM）的发展，人们发现人工智能生成内容（AI generated content, AIGC）也表现出令人惊艳的效果，甚至在某些程度上可以以假乱真，又激起了新一轮的热议。

#### 3. 滥用与违法犯罪

机器学习需要大量可供学习的数据作为训练模型的基础，然而，在相关法律法规尚未完全明晰且没有数据取得方式约束规则的情况下，许多包含个人隐私的数据也被用于训练模型，这些数据随时都有泄露风险。另一方面，训练集的数据分布对模型效果有决定性影响，其中可能包含种族等方面的隐性歧视：由于少数民族的数据占比非常低，在生成任务中很可能被忽略或无法正确处理；若提供的数据本身带有强烈的倾向性，那么训练出来的模型也很可能带有同样偏见。

同时，人工智能在语音迁移、人脸迁移（Deepfake）等任务上已经有不错的表现，这其中可能存在被用于违法犯罪行为的的风险，如电信诈骗。2017年，一些通过 Deepfake 制作的名人虚假视频和虚假新闻等开始在互联网上传播，这类恶意行为不仅对当事人造成了极大的名誉损害，而且在民间引起不必要的恐慌和误解。

#### 4. 能源消耗与环境破坏

以深度学习为代表的人工智能需要消耗大量的计算资源来完成训练和推理任务，其中的电力消耗非常巨大。阿姆斯特丹自由大学博士研究生亚历克斯·德·弗里斯（Alex de Vries）



预测，2027年新制造的人工智能备用电量将与荷兰整个国家的年用电量相当，甚至仅谷歌一家公司的用电量就能赶上爱尔兰全国的年用电量。由于许多地方难以使用清洁能源，仍然通过以煤为动力的发电方式来满足上述巨大的用电需求，人工智能的发展令碳中和的落实面临巨大考验。同时，这些设备在运行时也会产生巨大热量，需要大量的冷却水来保持它们处在安全的温度区间。据微软和谷歌披露，两者2022年的用水量分别为640万立方米和2120万立方米，且多为淡水。

## 1.4 计算机中数的表示

计算机内部是一个二进制的世界，在二进制系统中只有0和1两个数，不论是指令还是数据，在计算机中都采用二进制的编码形式。图形、声音等信息也必须转换成二进制编码形式，才能由计算机进行存储。因为在计算机内部，信息的表示依赖于机器硬件电路的状态，信息采用的表示形式直接影响到计算机的结构与性能。

### 1.4.1 计算机与二进制

采用二进制编码表示信息有以下几个优点：

(1) 用0, 1表示电磁状态的两对立面，容易在技术上实现。如晶体管的截止与导通、磁芯磁化的南极与北极、电位的低与高等。假如采用十进制，要制造具有10种稳定状态的物理电路，增加了物理元器件的复杂度。

(2) 运算简单。数学推导证明，对R进制数进行算术求和或求积运算，其运算规则各有 $R(R+1)/2$ 种。如采用十进制，则有55种求和与求积的运算规则；而对于二进制，则仅有3种，因而简化了运算器等物理器件的设计。

(3) 机器可靠性高。由于电压的高低、电流的有无等都是两种分明的状态，因此基2码的传递抗干扰能力强，鉴别信息的可靠性高。

(4) 通用性强。0和1不仅成功地运用于数值信息编码，而且适用于各种非数值信息的数字化编码。特别是0和1两个符号正好与逻辑命题的“真”与“假”两个值相对应，从而为计算机实现逻辑运算和逻辑判断提供了方便。

计算机存储器中存储的都是由“0”和“1”组成的数据编码，不同的编码分别代表不同的含义，可用于表示机器指令，有的表示二进制数据，有的表示英文字母，有的则表示汉字，还有的可能表示色彩与声音。存储在计算机中的信息可采用各自不同的编码方案，即使是同一类型的信息，也可以采用不同的编码形式。

### 1.4.2 数制的概念

数制是以表示数值所用的数字符号的个数来命名的，例如，十进制数由十个数字组成，即0, 1, 2, 3, 4, 5, 6, 7, 8, 9。在计算机内的数值信息用二进制数表示，基本符号只



有 0 和 1。

任何一种数制都包含两个基本要素：基数和位权。所谓某进位制的基数是指该进制中允许使用的基本数码的个数，例如，十进制的基数就是 10，逢十进一；二进制的基数就是 2，逢二进一。诸如此类，对于  $R$  进制数，有数字符号  $0, 1, 2, \dots, R-1$ ，共  $R$  个数码，基数是  $R$ ，运算时遵循逢  $R$  进一。

位权是指数制中每一固定位置对应的单位值，也就是一个数在不同位置上代表的数值大小。例如：个位上的“1”，就表示 1，而十位上的“1”，表示的确是 10 个 1。

例如：十进制数  $598 = 5 \times 10^2 + 9 \times 10^1 + 8 \times 10^0 = 500 + 90 + 8$ 。

### 1. 常用进位计数制

#### 1) 二进制

二进制由 0 和 1 两个数字组成，2 就是二进制的基数，逢二进一。二进制的位权是  $2^i$ ， $i$  为小数点前后的位序号。

#### 2) 八进制

八进制由八个数字组成，即 0,1,2,3,4,5,6,7 这八个数字，八进制的基数是八，逢八进一。八进制的位权是  $8^i$ ， $i$  为小数点前后的位序号。

#### 3) 十进制

十进制由十个数字组成，即 0,1,2,3,4,5,6,7,8,9 这十个数字，十进制的基数是十，逢十进一。十进制的位权是  $10^i$ ， $i$  为小数点前后的位序号。

#### 4) 十六进制

十六进制由十六个数字组成，即 0,1,2,3,4,5,6,7,8,9,a,b,c,d,e,f 这十六个数字，十六进制的基数是 16，逢十六进一，十六进制的位权是  $16^i$ ， $i$  为小数点前后的位序号。

### 2. 不同数制之间的转换

#### 1) 二、八、十六进制数（非十进制数）转换为十进制数

$R$  进制数转化成十进制数：数码乘以各自的权再累加。二进制数转换成十进制数的常用方法就是按权展开，然后按照十进制规则计算。

例：将二进制小数 1101.011B 转换成十进制数。采用按权展开法，过程如下：

$$\begin{aligned} 1101.011B &= (1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) D \\ &= (8 + 4 + 0 + 1 + 0 + 0.25 + 0.125) D \\ &= 13.375D \end{aligned}$$

例：将八进制数 5675 转换成十进制数。

$$\begin{aligned} (5675)_8 &= 5 \times 8^3 + 6 \times 8^2 + 7 \times 8^1 + 5 \times 8^0 = 2560 + 384 + 56 + 5 \\ &= (3005)_{10} \end{aligned}$$

例：将十六进制数 3B 转换成十进制数。

$$(3B)_{16} = 3 \times 16^1 + 11 \times 16^0 = 48 + 11 = (59)_{10}$$

#### 2) 十进制数转换为二、八、十六进制数（非十进制数）

十进制数转换成  $R$  进制数的方法：

整数部分：除以 R 取余数，直到商为 0，余数从右到左排列。

小数部分：乘以 R 取整数，整数从左到右排列。

(1) 十进制数转换成二进制数：数值由十进制转换成二进制时，要将整数部分和小数部分分别进行转换，再组合起来。

十进制整数部分转换成二进制数的最简便方法是“除 2 取余”法，也称为基数除法。

十进制小数部分转换成二进制数的常用方法是“乘 2 取整”法，也称为基数乘法。

例：(25.3125)<sub>10</sub> 转换为二进制数。

整数部分和小数部分的转换方法不同。

整数部分的转换（除 2 取余法）。

$$(25)_{10} = (11001)_2 \quad \text{先余为低，后余为高}$$

小数部分的转换（乘基 2 取整法）

$$(0.3125)_{10} = (0.0101)_2 \quad \text{先取整为高，后取整为低}$$

综上所述：

$$(25.3125)_{10} = (11001)_2 + (0.0101)_2 = (11001.0101)_2$$

(2) 十进制数转换为八、十六进制数以此类推。

3) 二进制和八进制、十六进制间的转换

由于 8 和 16 都是 2 的整数次幂，即  $8=2^3$ ， $16=2^4$ ，所以一位八进制数就相当于 3 位二进制数，而一位十六进制数就相当于 4 位二进制数。因此，八进制、十六进制同二进制之间的转换极为方便。

(1) 八进制和十六进制转换成二进制。由于每一个八进制数对应二进制的三位，每一个十六进制数对应二进制的四位，因此，八进制数转换成二进制数的方法是：用 3 位二进制数取代每一位八进制数；十六进制数转换成二进制数的方法是：用 4 位二进制数取代每一位十六进制数。

$$\text{如：} 2C1D(H) = 0010\ 1100\ 0001\ 1101(B)$$

$$\quad \quad \quad 2 \quad C \quad 1 \quad D$$

$$7123(O) = 111\ 001\ 010\ 011(B)$$

$$\quad \quad \quad 7 \quad 1 \quad 2 \quad 3$$

(2) 二进制转换成八进制和十六进制。整数部分：从右向左进行分组。小数部分：从左向右进行分组。转换成八进制三位一组；转换成十六进制四位一组。不足位时，整数部分在左边补零，小数部分在右边补零。

$$\text{如：} \underline{0011}\ \underline{0110}\ \underline{1110}.\underline{1101}\ \underline{0100}(B) = 36E.D4(H)$$

$$\quad \quad \quad 3 \quad 6 \quad E \quad D \quad 4$$

$$\underline{001}\ \underline{101}\ \underline{101}\ \underline{110}.\underline{110}\ \underline{101}(B) = 1556.65(O)$$

$$\quad \quad \quad 1 \quad 5 \quad 5 \quad 6 \quad 6 \quad 5$$

### 3. 二进制数的算术运算和逻辑运算

基本的算术运算有四种，即加、减、乘、除运算。