1. (a) What are the four main headings under which the British Computer Society Code of Conduct is described?

    What are the principal duties imposed by the 1984 Data Protection Act on those who hold permanent data about individuals?

    (5 marks)

   (b) Distinguish between *Knowledge Discovery in Databases* (KDD) and *data mining*, and explain briefly the current motivation for KDD in industry.

    (5 marks)

   (c) Name two "ground breaking" expert systems and indicate the domain in which they operate.

    In the context of expert system development what is the "knowledge acquisition bottleneck"?

    (5 marks)

   (d) Explain the meaning of the expression "parallel speed-up".

    (5 marks)

   (e) The time complexity of algorithm A is O(1), and algorithm B executes n times slower than A. Explain what is meant by the statement: "the time complexity of algorithm A is O(1)". What is the time complexity of B.

    In this context, what characteristics of a data set determine its size?

    (5 marks)

   (f) Explain what is meant by the term *deterministic model*, and briefly explain why it is necessary to run a stochastic simulation model several times in order to obtain a good estimate of the performance of the system being simulated.

    In connection with simulation models, explain the meaning of the terms *state* and *event*.

    (5 marks)

   (g) Explain the distinction between software verification and software validation, and give the quotation by Boehm that is directly relevant in this context.

    Explain the difference between black-box testing and white-box testing, stating which is most suited to verification and which to validation.

    (5 marks)

   (h) With the help of a diagram, briefly outline the essential features of Boehm's spiral model of software development and explain how it can accommodate

both the classic 'waterfall' model and a more evolutionary model. **NOTE** that whilst you need only indicate the general features of the spiral model, you should certainly include the meaning of each sector in the diagram.

(5 marks)

2.   (a)   One area to which data mining is commonly applied is that of "basket analysis". Explain what this is, and suggest how knowledge may be extracted from "basket analysis" databases.

(8 marks)

(b)   To what extent is parallel computing used to solve problems more quickly? Illustrate your answer using "expression evaluation".

(8 marks)

(c)   Is there anything to be gained through the application of parallel computing techniques to data mining.

(4 marks)

3.   Consider the following scenario: You are a programmer/analyst working in a team which has been commissioned by a government agency to conduct an investigation into the causes and treatment of a recently-discovered and potentially fatal disease. The data available for the study includes medical records of patients, recording a history of the progression of the disease and the treatments that have been attempted. Also included is other information about the patients, such as occupations, place of residence, etc, which might be relevant to understanding how they contracted the disease.

(a)   Discuss the role that modelling and simulation might play in conducting this investigation. What kind of simulation model would be most appropriate?

(6 marks)

(b)   What ethical issues are relevant to your involvement in this scenario?

(6 marks)

(c)   Suppose that, at an early stage in the investigation, your simulation appears to show that people living near nuclear power stations are more likely to contract the disease. Discuss your approach to this result as an ethical computer professional.

(8 marks)

4. (a) With the aid of a diagram, or otherwise, define the main stages in the *waterfall model* of the software life cycle, and for each stage, give two activities that are associated with it.

(6 marks)

(b) The following passage, which addresses the area of algorithmic complexity, contains a number of technical inconsistencies. Identify and briefly explain three of the inconsistencies.

P and *Q* are two decision problems of which *P* is undecidable and *Q* is intractable. *Q* is a member of the class *NP*, and the solution to any instance of *Q* is an integer in the range 1, …, 6. *A* and *B* are deterministic polynomial-time algorithms that solve *P* and *Q* respectively. When implemented and run on an IBM microcomputer *B* has been confirmed as having time complexity $O(n^3)$, and as having time complexity $O(n)$ when the same implementation is run on a Hewlett-Packard microcomputer.

(6 marks)

(c) When undertaking the activities associated with the various stages of the waterfall model of the software life cycle to produce a piece of software, many problems are encountered. Some of these problems are poorly-defined, some are intractable, and some are undecidable. Identify and briefly discuss **four** such problems, and in each case, indicate at which stage of the waterfall model you would expect to meet the problem. **NOTE** Your answer must include consideration of at least one each of: a poorly-defined, an intractable, and in undecidable problem.

(8 marks)

5. (a) Describe the stages appropriate for testing large software systems and then discuss the factors that an organisation would need to consider in deciding on the overall strategy for the verification and validation of large software systems that it develops.

(10 marks)

(b) Describe the characteristic features of expert systems and discuss the impact these features have on the verification and validation of such systems.

(10 marks)