8825 – 6406M

Markscheme

November 2025

Computer science

Higher level

Paper 3

11 pages

Subject details: **Computer science HL paper 3 markscheme**

Mark allocation

Candidates are required to answer **all** questions. Total 30 marks.

General

A markscheme often has more specific points worthy of a mark than the total allows. This is intentional. Do not award more than the maximum marks allowed for that part of a question.

When deciding upon alternative answers by candidates to those given in the markscheme, consider the following points:

- Each statement worth one point has a separate line and the end is signified by means of a semi-colon (;).
- An alternative answer or wording is indicated in the markscheme by a "/"; either wording can be accepted.
- Words in (…) in the markscheme are not necessary to gain the mark.
- If the candidate's answer has the same meaning or can be clearly interpreted as being the same as that in the markscheme then award the mark.
- Mark positively. Give candidates credit for what they have achieved and for what they have got correct, rather than penalizing them for what they have not achieved or what they have got wrong.
- Remember that many candidates are writing in a second language; be forgiving of minor linguistic slips. In this subject effective communication is more important than grammatical accuracy.
- Occasionally, a part of a question may require a calculation whose answer is required for subsequent parts. If an error is made in the first part then it should be penalized. However, if the incorrect answer is used correctly in subsequent parts then **follow through** marks should be awarded. Indicate this with "**FT**".
- Question 4 is marked against markbands. The markbands represent a single holistic criterion applied to the piece of work. Each markband level descriptor corresponds to a number of marks. When assessing with markbands, a "best fit" approach is used, with markers making a judgment about which particular mark to award from the possible range for each level descriptor, according to how well the candidate's work fits that descriptor.

General guidance

Issue	Guidance
Answering more than the quantity of responses prescribed in the questions	• In the case of an "identify" question read all answers and mark positively up to the maximum marks. Disregard incorrect answers. • In the case of a "describe" question, which asks for a certain number of facts *eg* "describe two kinds", mark the first two correct answers. This could include two descriptions, one description and one identification, or two identifications. • In the case of an "explain" question, which asks for a specified number of explanations *eg* "explain two reasons …", mark the first two correct answers. This could include two full explanations, one explanation, one partial explanation *etc*.

1. (a) *Award **[2]** max.*
*Award **[1]** for the role and **[1]** for effect (why they matter/how they change).*

Role [1]
Weights are numerical parameters on connections;
Weights set the strength/sign (positive/negative) of influence;
Each connection has a weight used when computing neuron outputs;
Weights combine current input with the previous hidden state to produce the new hidden state (and output);

Accept student phrasing: "each weight has a contribution to the output," "weights decide how much a node influences the next layer."

Effect [1]
Weights are adjusted during training to reduce error/loss (match desired output);
Gradients found by backpropagation through time (BPTT);
An optimizer like SGD/Adam updates the weights;
Weights capture patterns/dependencies in sequences;
Accurate weights produce better predictions;
Inaccurate weights can hurt learning (e.g., vanishing/exploding issues).

Accept student phrasing: "updated during backpropagation," "SGD updates the weights," "helps the network learn patterns."

(b) *Award **[2]** max.*
*Award **[1]** per point.*
Activation functions;
Backpropagation Through Time (BTT);
Batch processing;
Bias units;
Cost/loss function;
Deep learning;
Feedback connection/loop;
Gated Recurrent Units (GRUs);
Gradients;
Hidden states/storing previous inputs in memory;
Hyperparameters;
Layers (Input, Hidden, Output);
Learning rate;
Memory of past inputs;
Number of hidden layers;
Optimisation algorithms (e.g. SGD, Adam)
Parameter initialization importance;
Real-time processing;
Recurrence;
Regularization techniques (e.g. dropout, L2);
Sequence processing;
Vanishing and Exploding Gradients.

Accept other correct RNN-specific characteristics.

2. (a) *Award **[4]** max.*
*Award **[1]** for advantage and **[1]** for expansion.*
Reduced waiting time/chatbots provide immediate responses to customer inquiries;
Satisfying customer needs (questions about policies, coverage, claims, and premiums);

Chatbots are operational 24/7, offering round-the-clock service;
Unlike human agents who are available only during specific hours;

Fewer employees needed;
lead to significant savings/deployed elsewhere;

Chatbots can handle a large volume of queries simultaneously;
Reducing the need for a large customer service team;

Chatbots are immediately scalable;
Managing increasing volumes of customer interactions/handling peak periods;

Chatbots are consistent in their advice/uniform and reliable customer service;
Whereas humans are prone to variability in mood/fatigue;

Chatbots can offer personalised advice/tailor their responses to customer needs;
By accessing customer profiles and history;

Chatbots collected valuable data from customer interactions;
Which can be analysed to gain insights into customer needs, preferences, and behaviour patterns;

Chatbots can be programmed to adhere strictly to regulatory guidelines;
Ensuring all advice given is compliant with current insurance laws and regulations.

Mark as [2] + [2]

(b) *Award **[4]** max.*
*Award **[1]** for reason and **[1]** for expansion.*
Omitted information: Customers may not realise they are talking to an AI;
A human is less likely to omit key policy details, exclusions, or disclaimers;

Ensuing Accountability: When decision making is not transparent (i.e. how it was arrived at);
It becomes challenging to attribute responsibility for advice;

Decision making: Cannot easily understand how the chatbot arrived at a particular conclusion;
A chatbot giving advice that cannot be explained or justified leads to distrust;

Unidentified bias: Transparency is crucial to identify and rectify biases in training data;
Without it, discriminatory outcomes may result (e.g. certain groups charged higher insurance)/suboptimal advice for a certain demographic;

User Trust and Confidence: A lack of transparency can lead to diminished trust in chatbots;
Trust is fundamental to user adoption and the effective use of technology.

Regulatory and Compliance: In insurance, transparency is a legal requirement;
A lack of transparency can lead to compliance issues/violation of industry regulations;

Hindrance to Improvement and Optimisation: Transparency in a chatbot's decision-making process is crucial for ongoing improvement/optimisation;
Without insights developers will struggle to fix performance issues.

Mark [2] + [2]

3. *Award **[6]** max.*
Award [1] for a correct definition of syntactic analysis.
Award [1] for a correct definition of semantic analysis.
Award [1] for complementarity parts of NLP (used together).
Award [1] for dependency/order (syntactic first, then semantic).
Award [1] for techniques for syntactic analysis e.g. POS tagging.
Award [1] for techniques for semantic analysis e.g. context analysis.
Award [1] for a suitable example that differentiates between syntactic and semantic analysis.

Definition: Syntax examines sentence structure, focusing on word arrangement/grammatical rules;
Definition: Semantic focuses on the meaning and interpretation of words and sentences;
Complementarity: Both are used in tandem to ensure accurate and contextually relevant results;
Dependency: Syntax analysis lays the groundwork for structure, while semantic analysis adds meaning and context.
Techniques: Syntax analysis use parsing trees/context-free grammars/part-of-speech (POS) tagging;
Techniques: Semantic analysis employs context analysis/entity recognition/relationship extraction.
Example: "The customer filed a claim." Understands that "file a claim" means starting an insurance process, not physically placing a piece of paper into a folder.

4. *Award **[12]** max.*
Student may discuss some of the following points:

Impact of datasets on processing power requirements

- High quality datasets typically require less pre-processing, which can reduce computational load. This significantly decreases the time and resources needed for model training.
- High quality datasets lead to more efficient training and inference processes, ensuring that the model doesn't spend resources correcting biases, a process that may require additional layers or mechanisms to handle errors. The model can focus more on learning the core patterns rather than correcting for data-induced errors.
- A model trained on poor-quality data may require additional training iterations, complex architectures, or more sophisticated regularisation techniques, all of which increase processing power demands.
- With high-quality data, simpler models might achieve the same performance as complex models trained on lower-quality data.
- Accurate and representative data allows for more straightforward feature extraction, reducing the computational complexity of the model.
- High-quality, representative data improves the model's ability to generalize to new, unseen data, reducing the need for extensive retraining or fine-tuning.
- Overfitting is less likely with unbiased and comprehensive data, reducing the need for complex regularization techniques that consume additional processing power.
- Models trained on high-quality data are often more scalable and require less ongoing maintenance and retraining.
- Efficient data handling and processing reduce the need for constant hardware upgrades or cloud processing power, leading to more sustainable and cost-effective model deployment.

Creating high-quality dataset

- *Diverse Data Sources*: Collect data from a wide range of sources to ensure diversity and capture a broad spectrum of instances and scenarios, reducing the risk of biases associated with specific sources.
- *Inclusive Representation*: Ensure that the data set includes representation from all relevant groups, especially in terms of demographics like age, gender, ethnicity, and geography.
- *Identify and Mitigate Existing Biases*: Conduct a thorough analysis to identify any inherent biases in the data. Use techniques like stratified sampling to ensure that all subgroups are adequately represented.
- *Data Cleaning and Preprocessing*: Meticulously clean the data to remove errors, duplicates, and irrelevant information, which can introduce biases or inaccuracies. Normalize and standardize the data to ensure consistency across the dataset.
- *Feature Selection and Engineering*: Carefully choose features that are relevant and do not inadvertently introduce bias. Feature engineering should be done thoughtfully to avoid creating features that could lead to biased interpretations.
- *Expert Review and Consultation*: Involve domain experts in the process to identify potential sources of bias or areas where the dataset may not be representative. This can also help in understanding which data might be relevant or irrelevant to the problem at hand.
- *Regular Updates and Maintenance*: Regularly update the dataset with new data to ensure it remains relevant and representative of current conditions in the insurance industry.
- *Ethical Considerations and Privacy*: Ensure that the data collection and usage comply with ethical standards and privacy laws. Anonymise and secure sensitive data to protect individual privacy.
- *Validation and Testing*: Validate the dataset using statistical methods to ensure it is representative and unbiased. Employ cross-validation techniques to check for consistency and reliability.
- *Feedback loop*: Implement a feedback mechanism to identify and correct any biases or issues that might emerge over time.
- *Synthetic Data Generation:* Using techniques to generate synthetic examples can make the dataset more balanced.

Computational processing requirements

- *Massive Scale of parameters*: Large language models often have billions or even trillions of parameters. For example, GPT-3 has 175 billion parameters. Managing and updating this vast number of parameters during training demands significant computational resources.
- *Processors*: Training LLMs typically requires clusters of high-performance GPUs (Graphics Processing Units) or TPUs (Tensor Processing Units). These specialized processors are capable of handling the massive parallel computations needed for deep learning tasks.
- *Large datasets*: LLMs are trained on extensive datasets comprising a wide range of internet text. Processing and learning from these large datasets require substantial memory and storage, along with computational power.
- *Distributed processing*: Due to the enormous computational demands, training LLMs often involves distributed computing across multiple machines or even across data centres.
- High Bandwidth: High bandwidth is essential to feed data to the model at a rate that keeps the GPUs or TPUs fully utilised.
- *Sophisticated Cooling Systems*: The processing hardware requires advanced cooling systems in data centres to maintain optimal operating temperatures.
- *Training times*: Training a large language model for NLP can take days to weeks, depending on the model size and the specifics of the training regimen.
- *Optimization Techniques*: Mixed precision training, model parallelism, and data parallelism are often employed to manage memory usage and improve computational efficiency.

Conclusion

A final measured conclusion is included in which the candidate links together the various points in evaluating convergence.

Marks	Level descriptor
No marks	• No knowledge or understanding of the relevant issues and concepts. • No use of appropriate terminology.
Basic 1–3 marks	• Minimal knowledge and understanding of the relevant issues or concepts. • Minimal use of appropriate terminology. • The answer may be little more than a list. • No reference is made to the information in the case study or independent research.
Adequate 4–6 marks	• A descriptive response with limited knowledge and/or understanding of the relevant issues or concepts. • A limited use of appropriate terminology. • There is limited evidence of analysis. • There is evidence that limited research has been undertaken.
Competent 7–9 marks	• A response with knowledge and understanding of the related issues and/or concepts. • A response that uses terminology appropriately in places. • There is some evidence of analysis. • There is evidence that research has been undertaken.
Proficient 10–12 marks	• A response with a detailed knowledge and clear understanding of the computer science. • A response that uses terminology appropriately throughout. • There is competent and balanced analysis. • Conclusions are drawn that are linked to the analysis. • There is clear evidence that extensive research has been undertaken.

[12]