

Markscheme

May 2025

Computer science

Higher level

Paper 3

© International Baccalaureate Organization 2025

All rights reserved. No part of this product may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without the prior written permission from the IB. Additionally, the license tied with this product prohibits use of any selected files or extracts from this product. Use by third parties, including but not limited to publishers, private teachers, tutoring or study services, preparatory schools, vendors operating curriculum mapping services or teacher resource digital platforms and app developers, whether fee-covered or not, is prohibited and is a criminal offense.

More information on how to request written permission in the form of a license can be obtained from <https://ibo.org/become-an-ib-school/ib-publishing/licensing/applying-for-a-license/>.

© Organisation du Baccalauréat International 2025

Tous droits réservés. Aucune partie de ce produit ne peut être reproduite sous quelque forme ni par quelque moyen que ce soit, électronique ou mécanique, y compris des systèmes de stockage et de récupération d'informations, sans l'autorisation écrite préalable de l'IB. De plus, la licence associée à ce produit interdit toute utilisation de tout fichier ou extrait sélectionné dans ce produit. L'utilisation par des tiers, y compris, sans toutefois s'y limiter, des éditeurs, des professeurs particuliers, des services de tutorat ou d'aide aux études, des établissements de préparation à l'enseignement supérieur, des fournisseurs de services de planification des programmes d'études, des gestionnaires de plateformes pédagogiques en ligne, et des développeurs d'applications, moyennant paiement ou non, est interdite et constitue une infraction pénale.

Pour plus d'informations sur la procédure à suivre pour obtenir une autorisation écrite sous la forme d'une licence, rendez-vous à l'adresse <https://ibo.org/become-an-ib-school/ib-publishing/licensing/applying-for-a-license/>.

© Organización del Bachillerato Internacional, 2025

Todos los derechos reservados. No se podrá reproducir ninguna parte de este producto de ninguna forma ni por ningún medio electrónico o mecánico, incluidos los sistemas de almacenamiento y recuperación de información, sin la previa autorización por escrito del IB. Además, la licencia vinculada a este producto prohíbe el uso de todo archivo o fragmento seleccionado de este producto. El uso por parte de terceros —lo que incluye, a título enunciativo, editoriales, profesores particulares, servicios de apoyo académico o ayuda para el estudio, colegios preparatorios, desarrolladores de aplicaciones y entidades que presten servicios de planificación curricular u ofrezcan recursos para docentes mediante plataformas digitales—, ya sea incluido en tasas o no, está prohibido y constituye un delito.

En este enlace encontrará más información sobre cómo solicitar una autorización por escrito en forma de licencia: <https://ibo.org/become-an-ib-school/ib-publishing/licensing/applying-for-a-license/>.

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	<ul style="list-style-type: none">• 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 each characteristic, up to [2] max

Hidden state;

Layered architecture (input, hidden, output);

Contains connections with trainable weights and biases;

Uses activation functions (e.g. tanh and sigmoid);

Learns temporal dependencies;

Handles sequences of varying lengths;

Operates in stateful and stateless modes;

Backpropagation through time (BPTT);

Uses hyperparameters (e.g., number of units, learning rate);

Supports selective memory

Supports sequence-to-sequence learning;

Maintains long-term memory via cell state;

Input gate (three-gate mechanism);

Forget gate;

Output gate;

Gate-based architecture for memory control.

Separates cell state from hidden state;

Improves long-term dependency learning;

(b) **Award [2] max**

Award [1] for definition, [1] for expansion, which can be an example of context

Pragmatic analysis aims to understanding meaning beyond syntax and semantics;

Delves into the context in which the sentence is used;

Pragmatic analysis looks at implied meaning/contextual meaning/behind the language;

Based on social or cultural norms;

Pragmatic analysis looks at the speakers' intention ("pass the salt" not "do you have the ability to pass the salt?");

Looks beyond the literal meaning of the sentence.

2. (a) **Award [4] max**
Accept any [4] from the following:
Award [1] for assigning scores, [1] for vectors, [1] for normalisation, [1] for weighted sum, [1] for related words, and [1] for applying the example, up to [4] max
Mechanism assigns attention scores to each word in the sentence (relative to other words);
Transforms each word into vectors (i.e. query, key, and value) through weight matrices;
Attention scores are calculated using the dot product of query and key vectors;
Attention scores are normalized through a softmax function to form a probability distribution;
Attention scores are used to create a weighted sum of all words;
Assign higher scores to important (contextually relevant) words;
Accept a suitable example. e.g. “chased” has a higher attention score than, “cat” and “mouse”
or “The” has a lower attention score than “cat”
- (b) **Award [4] max**
Award [1] for identifying reason, and [1] for elaboration
LLMs have more (trainable) parameters/more complex model;
Larger datasets are required but with insufficient data or poor regularisation;

Scale and architecture amplify sensitivity/pre-trained on large, diverse, and noisy datasets;
Then fine-tuned on smaller, task-specific datasets/increases the risk of narrow or biased content;

LLMs have increased task versatility/required to handle a range of NLP tasks;
This increases susceptibility to learning task-specific noise in the data;

LLMs require greater computational resources than RNNs;
Limited resources may constrain tuning (e.g., regularisation, early stopping)

Unlike RNNs, LLMs do not inherently model the sequence of the data;
This may reduce their ability to generalise across time-dependent patterns;

LLMs lack the inductive bias of RNNs (naturally focus on recent or local input patterns);
Without this bias, LLMs are more likely to model irrelevant or global noise;

Mark as [2] + [2].

3. *Award [6] max*
Award [2] max for advantages, [2] max for disadvantages, and [2] for a reasoned opinion/conclusion

Advantages

Internet provides a vast and varied dataset, spanning many topics, styles, and linguistic patterns; This diversity helps train LLMs to understand different cultures, languages, and contexts/with sufficient qualities to enable pre-processing to filter out unnecessary data;

Much of the internet is publicly accessible and unregulated, making it a low-cost source of data; Models benefit from the breadth of content without needing expensive licensing agreements;

Up-to-date information: Internet content is constantly updated, allowing models to learn from recent trends, social changes, or emerging terminology; This can improve chatbot relevance in rapidly evolving fields like law, insurance, or healthcare.

Disadvantages

The Internet is not a curated source/contains unreliable information; The model can inadvertently learn from this misinformation and spread it; Leading to negative consequences (e.g. customers receiving unreliable advice/ lawsuits/decline in customer satisfaction);

Internet data may contain cultural bias, harmful stereotypes, or offensive content; Using such data without careful filtering can reinforce prejudice in model outputs.

Many internet users do not give permission for their content to be used in model training; This can lead to public backlash, legal consequences, and reputational damage for companies deploying LLMs;

The internet contains vast amounts of irrelevant or off-topic material; Extensive pre-processing is required to extract domain-specific content, increasing time and computational cost.

Conclusion

Award [1] for weighing the benefits against the risks and [1] for suggesting a logical strategy

Example opinions worth [2] marks

Internet data may never be fully reliable or neutral, but its scale makes it indispensable. The key challenge isn't whether to use it, but how carefully it's selected and refined. For applications like chatbots in regulated domains, strategic filtering and validation must be built into the model pipeline from the start.

For an insurance chatbot, trust and accuracy matter more than scale. While internet data can help build a broad language foundation, relying on it without control risks reputational damage. A better long-term strategy is to pair that foundation with curated claims data and expert-reviewed language to align the model with industry standards and customer expectations.

Note: For both advantages and disadvantages, award [2] for one point with expansion or two separate points.

4. **Award [12] max**

Some of the following points may be discussed:

Increasing computational processing power

Hardware and Infrastructure:

- Upgrade hardware: use TPUs, GPUs, or more powerful CPUs.
- Use hardware acceleration (e.g., matrix multiplication optimisations).
- Deploy high-performance cloud infrastructure.

Parallelisation and Scalability:

- Implement parallel processing frameworks (e.g., batch processing or transformers' attention mechanisms).
- Use distributed computing or clustered servers for inference and training.
- Enable auto-scaling infrastructure to handle peak loads.

Software/Architectural Optimisations:

- Optimise APIs and backend code to reduce latency.
- Use microservices architecture to isolate and scale critical components (e.g., the NLP engine).
- Thread pooling or asynchronous processing to handle multiple conversations.

Response Time Improvements:

- Cache frequent responses (e.g., FAQs).
- Use load balancing to distribute incoming queries.
- Employ request prioritisation or throttling under heavy load.

Model-level optimisation and compression:

- Model distillation: Use smaller student models trained to mimic larger models for faster response times.
- Quantisation: Reduce model precision (e.g., float32 → int8) to reduce memory use and speed up inference.
- Early exiting and pruning: Skip unnecessary layers or connections during inference to improve efficiency.

Improving dataset

Data Collection and Curation:

- Use a domain-specific dataset (insurance-specific documents, FAQs, claim data).
- Include diverse linguistic patterns (from real customer queries).
- Balance data to represent all demographics.

Bias and Ethics Management:

- Remove culturally or demographically biased data.
- Conduct expert and ethical review of dataset inclusions.
- Monitor for fairness in prediction across user types.

Labelling and Structure:

- Improve annotation quality through clearer guidelines and trained annotators.
- Automatically label responses when appropriate, but review manually.

Pre-processing and Engineering:

- Clean noisy data (misspellings, irrelevant tokens).
- Use feature engineering to add useful context (e.g., intent classification).
- Augment data using paraphrasing, back-translation, or synthetic dialogue generation.

- Discussion of trade-offs between real and synthetic datasets is creditworthy, especially where it links to bias, authenticity, or model performance.

Validation and Testing:

- Perform stratified validation splits (training, testing, validation).
- Cross-validate to identify overfitting.
- Collect real-world user feedback for continuous improvement.
- Define performance metrics such as accuracy, latency, F1 score, or customer satisfaction ratings.
- Human-in-the-loop testing for manual review of chatbot outputs.
- Deploy A/B testing to measure improvement over previous models in real scenarios.

Conclusion

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

Please see markbands.

Marks	Level descriptor
No marks	<ul style="list-style-type: none"> • No knowledge or understanding of the relevant issues and concepts. • No use of appropriate terminology.
Basic 1–3 marks	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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.