Questions matter

The role of AI in educational assessment

What every assessment team should know for 2024

Alex Scharaschkin
14 February 2024

Copyright © AQA and its licensors. All rights reserved.

Outline

How AI differs from other ways of working with data

…and why has it suddenly become ubiquitous?

Benefits and risks of using AI in assessment contexts

…both in standardised summative assessments and in the classroom

How we are we applying and developing AI tools at AQA

Outlook for 2024

Q&A/discussion

Copyright © AQA and its licensors. All rights reserved.

What is AI?

A way computers can ‘learn’ from data, to perform certain tasks

- rather than human programmers having to figure out how to give the computer instructions to do the tasks

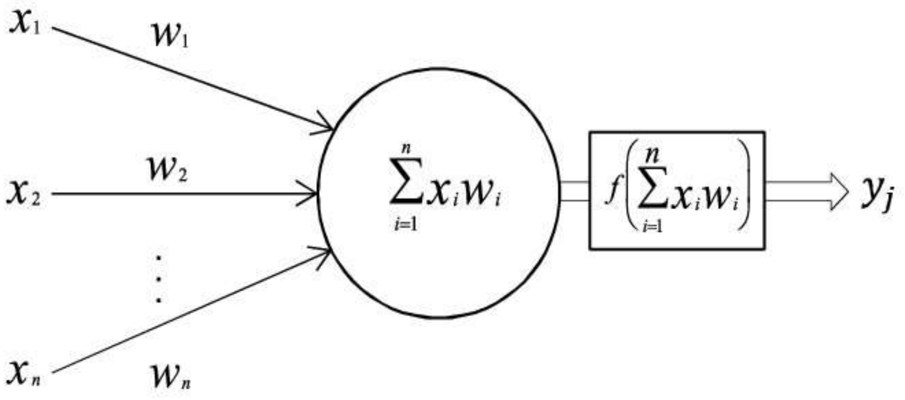
Tasks could be ‘make a decision/recommendation’

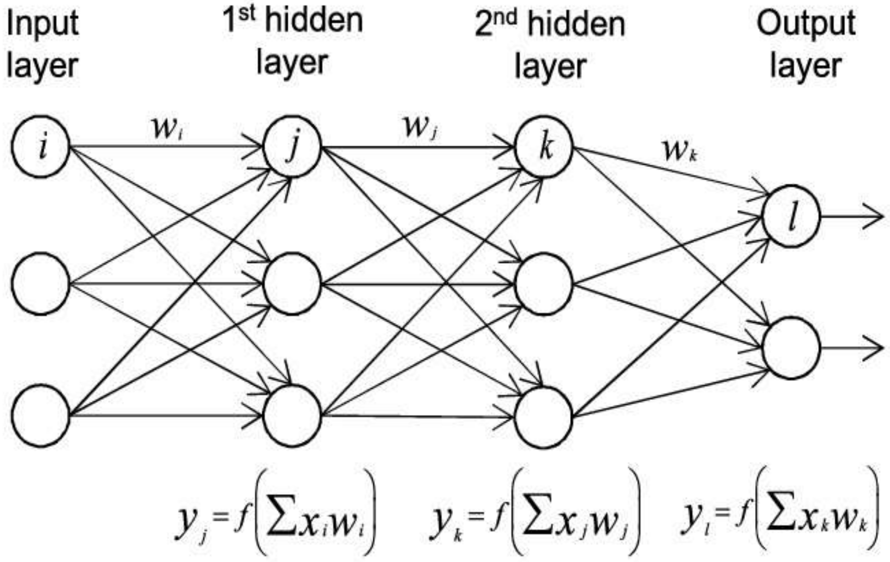
- facial recognition
- recommender engines

Or ‘produce some output’ (generative AI)

- ChatGPT
- DALL E2; Midjourney
- Jukebox

These tools are implemented using *neural nets*




Input layer
1st hidden layer
2nd hidden layer
Output layer
i
w_i
j
w_j
k
w_k
l
$y_j = f\left(\sum x_i w_i\right)$
$y_k = f\left(\sum x_j w_j\right)$
$y_l = f\left(\sum x_k w_k\right)$

Copyright © AQA and its licensors. All rights reserved.

Why has AI suddenly become ubiquitous?

Neural networks have been around since the 1940s

But three developments have now come together:

- **Computer processing speeds**

  Networks can be enormous: there are 175 billion connections in GPT3.5

  Environmental impacts of massive computing resources… (cf. mining bitcoin)

- **Network engineering**

  Large language models; transformer architecture; generative AI

- **Availability of training data – the internet**

  Ethical issues – copyright; intellectual property; tagging disturbing content;…

Copyright © AQA and its licensors. All rights reserved.

Benefits and risks of using AI in educational contexts

- **IMHO, the methods of artificial intelligence (so far) produce *simulacra* of intelligence. But tools such as ChatGPT aren't *intelligent*.**

  An actor giving a convincing performance as a doctor isn't a doctor.

- **Regulating activity aimed at producing 'artificial general intelligence' (AGI) is important…**

- **…but a more immediate risk, especially in educational settings, is *believing that a convincing-sounding piece of output from a chatbot is actually true*.**

- **Moreover, by using ChatGPT, for example, you're giving your data to OpenAI.**

- **However, clearly genAI is potentially transformative for many tasks that involve producing content**

  writing assessment tasks/items;
  generating feedback to learners;
  mapping assessment structures, or lesson plans, to a syllabus or specification, etc. …

- **Could be applied independently; in a human-in-the-loop model; or in a co-pilot model**

Copyright © AQA and its licensors. All rights reserved.

Appraising AI applications in educational contexts

- **It's easy, in principle, to stack together some open-source, or third-party-provided, apps, put on a front end, and market the result as an edtech product**

  E.g. a 'virtual teacher' that responds to students' questions, created by stacking

  (input) voice→text; text→ChatGPT; (output) text→voice; voice→animated avatar

- **Clearly, the quality, reliability, and usefulness of such applications can be highly variable**

  Biases; hallucinations; appropriateness of content; training corpus; educational/pedagogical validity…

- **At AQA, in line with our values as an independent education charity, we are determined to pursue the deployment and development of AI-enabled tools in an ethical, responsible, and trustworthy manner**

- **And as well as applying *AI to assessment*, we are also exploring the application of *assessment to AI* – how can we use assessment expertise to measure the quality (fairness, validity, reliability) of AI tools or products?**

Copyright © AQA and its licensors. All rights reserved.

How are we using AI in assessment at AQA?

- **Automated marking (relating to a mark scheme, with explanation)**

  As a form of **quality assurance** for high-stakes assessments (integration into next generation of our marking platform)

  (The current version already uses AI to assist image processing)

  High-stakes exams are **not** going to be marked by robots; but informed use of AI-enabled tools can help professionals do their jobs more easily, and focus on what's really important

  Trials are promising (both reading handwriting and marking for certain subjects/question-types)

- **Automatic item (task) generation**

- **Feedback to learners**

  Formative assessments, and teaching-and-learning products
  These must be **educationally valid** – based on an effective theory of learning

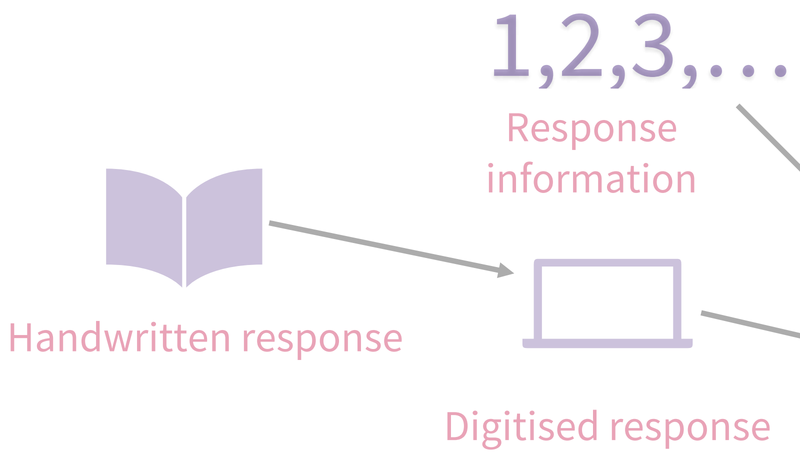
  ⇒ **Reducing teachers' workloads**

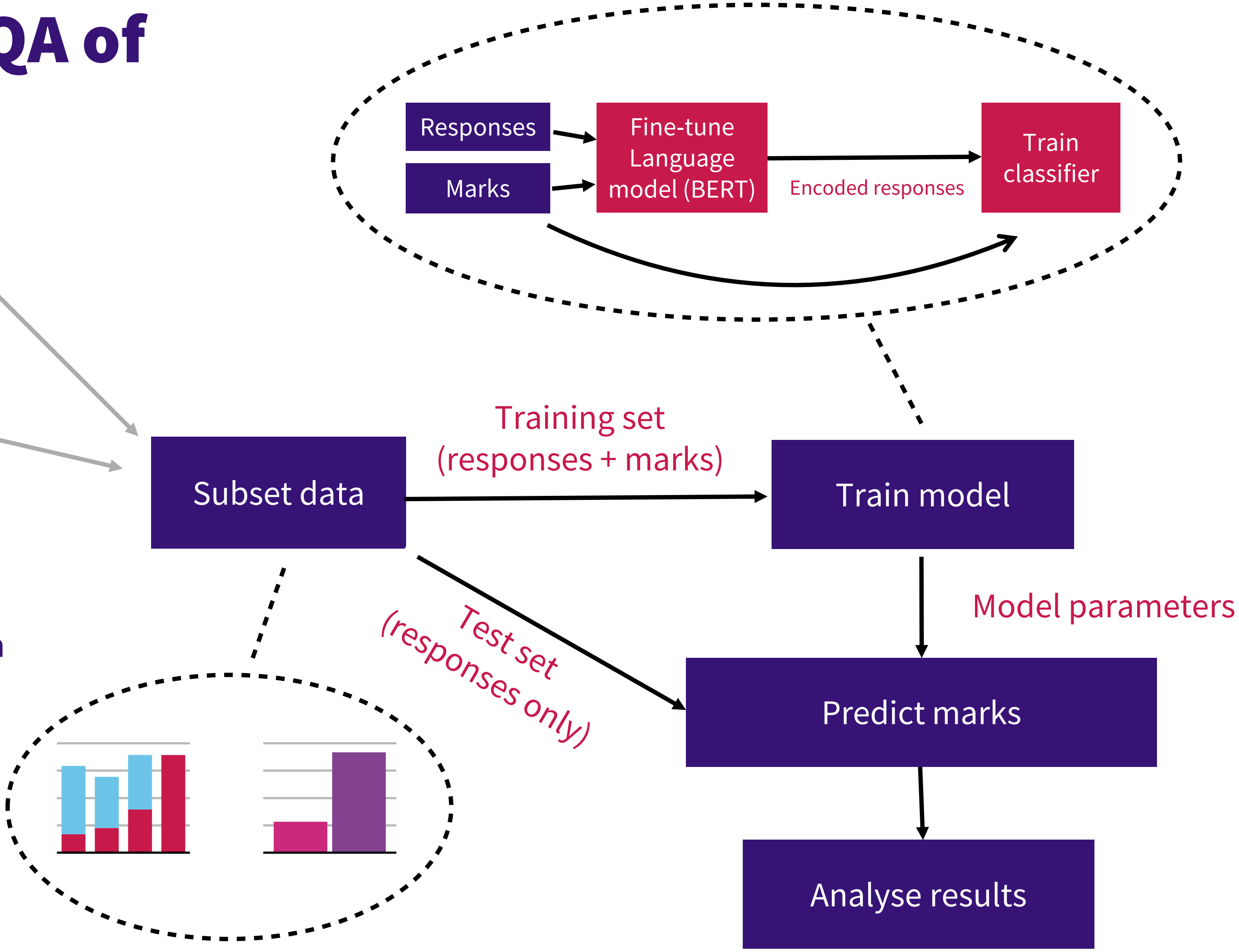
  Using professional expertise to greatest effect and highest impact - rather than being crowded out by routine work

- **A framework and toolkit for the ethical and responsible use of AI**

Copyright © AQA and its licensors. All rights reserved.

Example: using AI for QA of exam marking





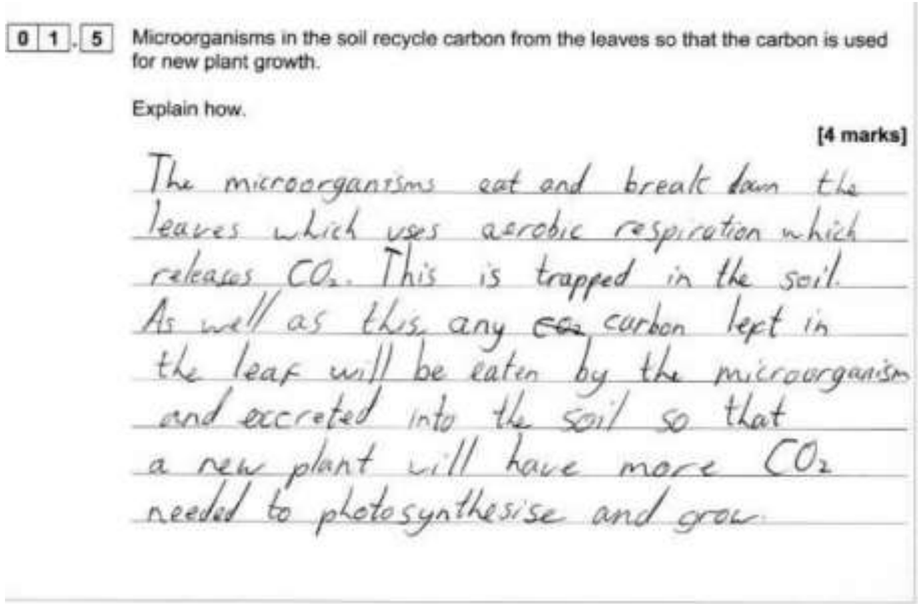
Dozens of experiments carried out for each item

Using different combinations of variables including:

- Data sub-setting (data imbalance, training sets)
- Using different mark classifiers
- Different amounts of fine-tuning (the process of customising the model for our data)

Copyright © 2023 AQA and its licensors. All rights reserved.

Example: using AI for QA of exam marking

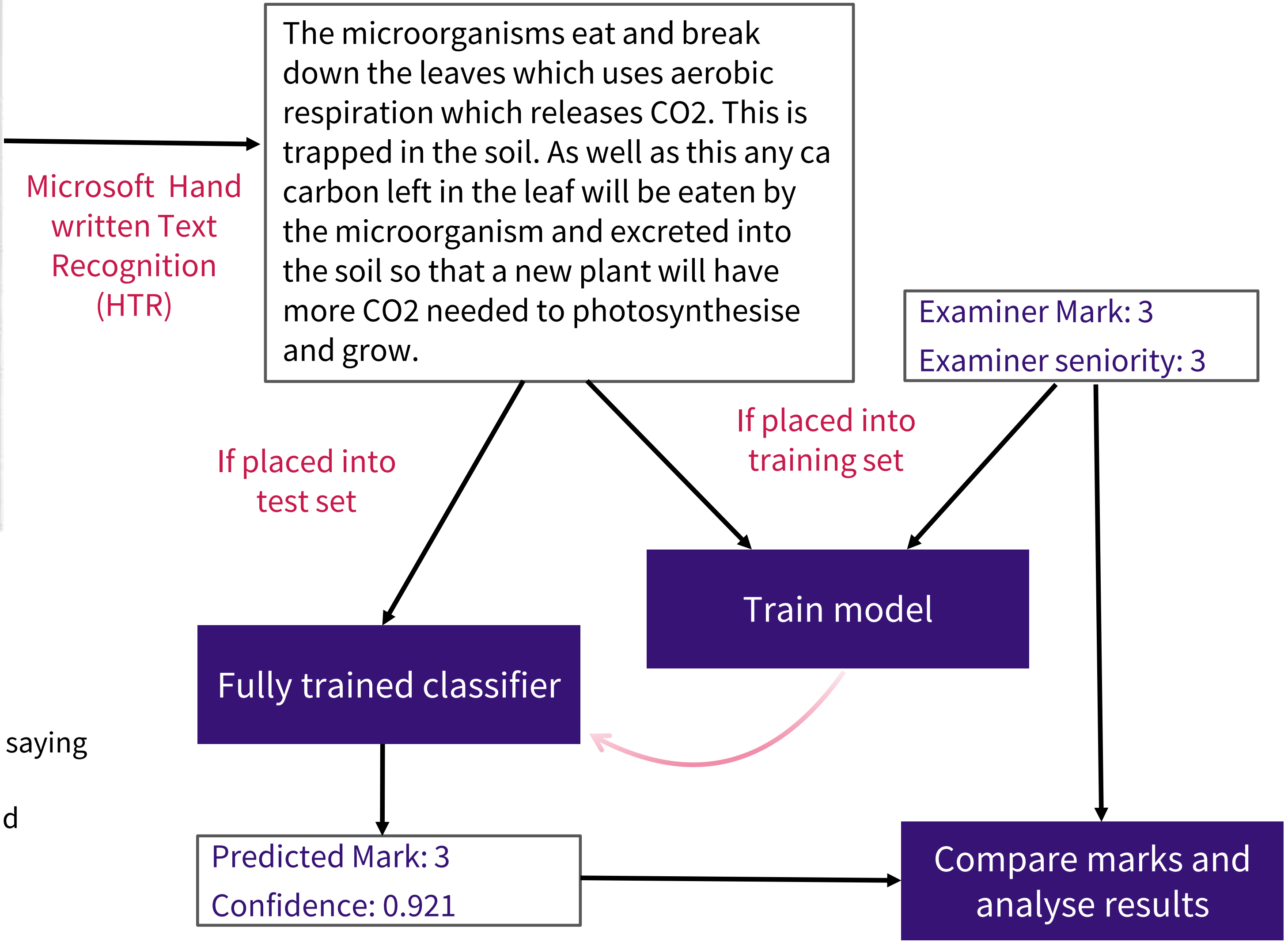


0 1 . 5 Microorganisms in the soil recycle carbon from the leaves so that the carbon is used for new plant growth.

Explain how.

[4 marks]

The microorganisms eat and break down the leaves which uses aerobic respiration which releases CO_2. This is trapped in the soil. As well as this any ~~CO_2~~ carbon left in the leaf will be eaten by the microorganism and excreted into the soil so that a new plant will have more CO_2 needed to photosynthesise and grow



Microsoft Hand written Text Recognition (HTR)

The microorganisms eat and break down the leaves which uses aerobic respiration which releases CO2. This is trapped in the soil. As well as this any ca carbon left in the leaf will be eaten by the microorganism and excreted into the soil so that a new plant will have more CO2 needed to photosynthesise and grow.

Examiner Mark: 3

Examiner seniority: 3

If placed into test set

If placed into training set

Train model

Fully trained classifier

Predicted Mark: 3

Confidence: 0.921

Compare marks and analyse results

The path taken by one response

- The script image is transformed into text
- It may be placed in the training dataset or test dataset
- Training data helps to produce the fully trained classifier by saying this text is worth 3 marks
- Test data is used to analyse our model’s overall accuracy and applicability

Copyright © 2023 AQA and its licensors. All rights reserved.

Results: QA of exam marking

- **We could predict senior examiner marks quite accurately for constructed-response style items**

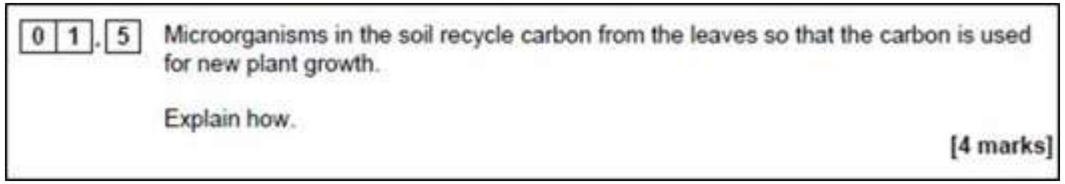
  ≥80% exact agreement with senior mark for 3- and 4-mark items; ≥65% agreement for 6-mark items

  100% agreement to within a 1-mark tolerance for 3- and 4-mark items; ≥80% agreement for 6-mark items

- **Further work ongoing to explore practicalities of using this approach at scale, to focus senior examiner interventions most effectively during the live marking period**
- **As well as looking at other functionality to enhance the exam marking experience**
- **Note the importance of keeping control of sensitive data – especially if working with external providers of large language models**
- **And, from a theoretical assessment perspective, exploring approaches to producing <u>valid explanations</u> of the mark awarded, grounded in construct-relevant features of the candidates' responses**

Copyright © AQA and its licensors. All rights reserved.

Example: using genAI to give feedback



0 1 . 5 Microorganisms in the soil recycle carbon from the leaves so that the carbon is used for new plant growth.

Explain how.

[4 marks]

Student Response:

The leaves die once they have fallen to the ground and small organisms such as insects as well as microorganisms break down the plant matter into its carbon compounds. They also release carbon compounds through respiration. As new plants grow in the same spot, they absorb the carbon by photosynthesis and through their roots to use to grow.

AI feedback:

The student's response explicitly mentions two out of the four key concepts in the mark scheme. They state that "small organisms such as insects as well as microorganisms break down the plant matter into its carbon compounds" which aligns with the mark scheme's answer of "decay / decomposition of leaves". Additionally, they mention that "microorganisms release carbon compounds through respiration" which corresponds to the mark scheme's answer of "respiration (by microorganisms / decomposers)". Therefore, the student is awarded 2 marks out of a possible 4.

Copyright © 2023 AQA and its licensors. All rights reserved.

Examiners’ reflections on AI-generated feedback

- Examiners felt the AI feedback was generally **a decent reflection of the mark scheme but was not always accurate**. An example given for one provider was marking for level of response, wherein the feedback suggested the number of ideas/concepts listed would gain marks (e.g. 3 ideas = 3 marks) which is misleading for students.
- However, the examiners thought **such feedback may not be useful for students**, as it mostly just restated the mark scheme and not the specifics of what the students hadn’t done or could be doing to improve their answer.
- Certain feedback looked impressive (e.g. created a narrative, listed page numbers from the textbook for concepts) but was **lacking in terms of providing genuine 'feed-forward'** - highlighting misconceptions and suggesting practical learning steps.
- For one of the examiners, reading the feedback and its lack of personalised detail reminded them “**how much value an expert teacher provides** for students in their learning experience”.

Copyright © 2023 AQA and its licensors. All rights reserved.

Outlook for 2024

- **A move from a ‘use ChatGPT to draft stuff and then check it’ approach…**

  …to ‘use genAI-enabled co-pilot functionality’ in many applications

- **Greater awareness of ethical, data protection, copyright, and safety issues**

  … to what extent will smaller providers be able to challenge big tech in applying genAI and LLMs?

  …and what should form part of the school curriculum on AI?

- **An ever-increasing number of edtech apps in the market, of varying educational quality**

  …with a need for trustworthy appraisals of their reliability and effectiveness

- **Scope for re-imagining formative assessment: digital tools that integrate assessment and pedagogy**
- **Re-thinking the purpose and design of summative non-examined assessment**

Copyright © AQA and its licensors. All rights reserved.

To conclude...

- **The combination of ubiquitous digital devices, and the rapidly evolving capabilities of AI, will transform what we teach, how we teach it, and how we assess young people**
- **It is essential that the professional assessment community has a leading voice in the application of AI to educational assessment**

  → to maximise the fairness, validity, reliability, and effectiveness of edtech products

  → while remaining open to new assessment paradigms, enabled by new ways of working with unstructured data that are unlocked by large language models

- **We plan to expand our work applying AI to educational assessment across several fronts. Watch this space!**

Copyright © AQA and its licensors. All rights reserved.

Thank you