

Validating the AI's Answer! Read & R.A.S

Artificial Intelligence is becoming a “go to” source for information. Its speed and the readability and accessibility of the output to our questions often disguise inaccuracy, bias and sometimes Hallucinations.

A Hallucination is where the AI creates content to “fill a gap” in its own knowledge. The hallucination is often inaccurate, irrelevant or does not make sense.

The accuracy of the AI is limited to the information that it has available; this includes the information it was trained on called the training set. If the training set is out of date, inaccurate or contains information that is biased then the output is likely to be inaccurate and or harmful.

A simple tool we can use is Read and R.A.S. - Relevance, Accuracy and Safety.

Read the *output* and *consider the relevance, accuracy* and *safety* of the response.

R - Relevance

Is the output from the Artificial Intelligence (AI) RELEVANT?

Think about the question or task you gave the AI. Using your Prior Knowledge check:

Does the output match the question?
Is the output RELEVANT to the question
Is the output in the form or style I expected?



A - Accuracy

Is the output from the Artificial Intelligence (AI) Accurate, and Correct?

Using your prior knowledge, does the output match your understanding?

Are there errors that are concerning? For example:
Spelling and grammar mistakes,
Non-sense sentences,
Inappropriate or mis-placed sentences or phrases,
unrelated content.

Check the sources the AI has used? Are the sources real, relevant, up to date and from reliable sources?



S - Safety

Is the output of the Artificial Intelligence (AI) SAFE?

Is the output BIAS?
Does the output seem to be unfair, have prejudice, be racist, sexist or bigoted?

Is the output AGE APPROPRIATE?
Is it using language that is complex and hard to understand?

Does the output contain materials (text, audio, media etc) that are UNSAFE or INAPPROPRIATE?

Is this an ECHO Chamber?

Does the output feel wrong?



Does it pass the R.A.S test? Is it Relevant, Accurate, and Safe?

If in doubt, ask your teacher!