Testing Criteria

Here is an outline of the testing criteria I developed to evaluate the tools using a grading scale of -1 to 1.

1. **Useful**: The quality of the work being produced is increased by using the copilot.

2. **Efficient**: The ratio of time taken/ effort to quality of the final output is reduced by using the copilot.

3. **Flexible**: The copilot has a flexible understanding of prompts – does not require overly complicated prompt engineering in order to produce coherent outputs.

4. **Domain-specific**: The output produced by the copilot is specific and highly relevant to the task it was asked to carry out.

5. **Coherent**: The copilot hallucinates < 5% of the times it is prompted.

6. **Complementary**: The user can use the copilot's output as a complement to the work already being done, not just as a "replacement". There is easy control over the degree to which the copilot's output is incorporated into the final output.

7. **Co-worker**: (Ideally) The copilot introduces a novel approach to the task being carried out.

8. **Learner**: (Ideally) The copilot implements persistent memory of tasks it has already carried out, so that the next time users need it to complete the same task, this can be accomplished quicker.

9. **Confidential**: (Ideally) The data is not being used to train the AI being used for a task.

10. **Transparent data collection**: There is transparency in terms of what data is being collected by the copilot. -Whether any of this data is being made available to third parties. -Where/ how/ by who the data is being stored. -The degree of persistency of the data collected.

11. **Transparent training**: There is transparency in terms of what data was used/ is being used to train the copilot.

12. **Proactively trained**: The copilot's training was proactive and supervised, with correction for bias included in the process.

13. **Ethically sound:** The copilot's software includes protection against the development of unwanted behaviors and is regularly monitored for signs of these.

14. **Customizable**: (Ideally) Users can further train the copilot to customize it for their needs.

15. **Intuitive**: The copilot provides an intuitive, easy to work with UI.

16. **Simplifying**: Working with the copilot does not increase the complexity of the task at hand.

17. **Enjoyable**: (Ideally) Use of the copilot increases job satisfaction.

18. **Collaborative**: The copilot enables/ facilitates collaborative work.

19. **Compatible**: The copilot facilitates system integration and can work in parallel with other systems already in use.