

# Evaluation Criteria

For each MGM category, we consider the following criteria in evaluating candidate tools for application in the AMPPD pilot. Each MGM category may have specific criteria definitions, for instance, accuracy measures or social impact considerations, depending on the nature and purpose of the algorithm.

Evaluation Criteria	Description
<b>Accuracy</b>	How does the MGM output compares to the expected value (or human-generated value). This should be a consideration of both quantitative and qualitative measures.
<b>Input formats</b>	Filetypes, encodings, compressions, etc. allowed by the MGM. Assess the level of difficulty involved in converting your files to the formats required for the tool. How will this impact automation? Is anything lost in the conversion that could affect the accuracy of output?
<b>Output formats</b>	File types or data formats output by the MGM. Assess the level of difficulty involved in converting available output formats to the desired format. How will this impact automation?
<b>Growth rate</b>	Rate of increase of time and computing resources as volume/file size increases. Compare processing time between small, average, and large sized files to estimate time required as scale increases. Is this feasible given the estimated contents of your project? Compare memory use between small, average, and large sized files to estimate memory required as scale increases. Is this feasible given the estimated contents of your project?
<b>Processing time</b>	Time required for the MGM to process the file. How will processing time affect your production workflows? Can processing time be improved by optimizing computing hardware, software, or networks?
<b>Computing resources</b>	Amount of computing resources, including processing power, memory, network connections, and bandwidth required to process the file. How will computing resources affect your production workflows? Will you need to operate the MGM on other machines?
<b>Social impact</b>	The potential unintended consequences of an unmediated MGM's output. How could the MGM express hidden biases? What are the possible unintended negative impacts that could come from the output of this MGM? What measures can be taken to mitigate them? See FAT/ML's Principles for Accountable Algorithms for more information: <a href="http://www.fatml.org/resources/principles-for-accountable-algorithms">http://www.fatml.org/resources/principles-for-accountable-algorithms</a>
<b>Cost</b>	The cost of the MGM which could include paid services, file transfer and computing costs if running in the cloud, or local hardware and staff costs.
<b>Support</b>	Available human support, documentation, or logs output by the MGM which can help with learning or troubleshooting the MGM.
<b>Integration capabilities</b>	The ability of an MGM to fit into a workflow design or technical infrastructure or the ability to supply functionality for other computational needs, such as a speech-to-text tool that also provides segmentation and speaker diarization.
<b>Training</b>	Whether or not a model should be trained to utilize the MGM. Consider the costs, time, and social impact of training a model or using a model out-of-the-box.