July 1, 2019

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
1	Dean Alderucci	Minor	Page 9, line 188, at the end of this bullet point	I recommend referencing another type of Data Standard: Document Standards to Facilitate AI Processing. It is important to be aware of such data standards, especially because NIST and government agencies are uniquely positioned not only to provide guidance but also to receive substantial benefits from such standards, as explained below.

Pages 8 – 10 of the Plan clearly describes the need for data standards, including guidance and requirements for big data analytics; data exchange; data quality; and data privacy. Another of type of data standard, "Document Standards to Facilitate AI Processing", is important but is not mentioned in the Plan.

Document Standards to Facilitate AI Processing are used where certain types of documents contain information we would like to extract using AI / Natural Language Processing (NLP), but current AI / NLP capabilities are insufficient to extract the information from those documents, forcing those documents to be processed manually. These documents include those created for submission to a government agency, and those created by a government agency. The information to be extracted from the documents can be anything that human readers want to glean from the documents, especially information that must be pieced together from multiple parts of a documents and that requires higher-level cognitive processing or 'common-sense' reasoning.

A specific example of an agency, type of document, and type of information in the document will be used to illustrate the principles described, but the principles apply to many more documents across different government agencies.

Assume that a patent application is drafted and submitted to the US Patent & Trademark Office (PTO) for processing. A PTO Examiner must read the application to determine whether the application | Page 9, line 188, at the end of this bullet point insert text such as the following:

"Since data also includes documents processed by and created for government agencies, data standards also include standards to facilitate AI processing of those documents. Such standards specify how documents should be created so information of interest can be more accurately identified and extracted by AI. This could deliver the benefits of AI processing in a shorter time and at a lower cost than by waiting for AI capabilities to advance to the point of being capable of processing documents in their current form." |

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
				satisfies various legal requirements. One such requirement is whether the claims portion of the application meets the "definiteness" requirement, i.e., whether a person of 'ordinary skill' in the relevant field of technology would find the claims to be ambiguous, vague, or otherwise unclear. Whether a claim meets the definiteness requirement depends on what is contained in other portions of the patent application, such as whether and how terminology in the claims are defined and generally how the technology recited in the claims is purported to operate. Unfortunately, state-of-the-art NLP is incapable of rendering this type of judgment, as it depends on data that is difficult to extract from the document (e.g., the specific meaning of technological terms) as well as higher-level cognitive inferences to be drawn from this data (e.g., how exactly a person of ordinary skill in a particular field of technology would understand the workings of technology recited in the claims). Other types of decisions the PTO Examiner would like to make about the patent application are likewise beyond state-of-the-art NLP capabilities. When AI is incapable of extracting a desired type of information from a particular type of document, there are two approaches to resolving this limitation. The first is perhaps more obvious: improve AI technology so that it can extract this type of information from the document. This is clearly a laudable goal. However, when the information to be extracted is extremely complex, it could take many years and significant research resources to develop AI that can perform the task. In some circumstances, a second approach can be quicker and less resource intensive: change the way the document is created in order to facilitate its processing by AI. That is, specify that the document should be created in a manner that highlights or otherwise makes	

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
				certain types of information more easily identified by AI / NLP software. In situations where the second approach imposes little additional costs in document creation but delivers significant benefits, the second approach is worthy of careful consideration. A notable benefit of improved AI document processing could be, e.g., automating tedious manual tasks and locating information in minutes rather than weeks. Where the document is to be processed many times over several years the benefits would be even more pronounced. Document Standards to Facilitate AI Processing would likely require consideration of factors such as the nature of the document being created, how it is created, the types of information that are desirable to extract from the document, use cases for such information, the benefits of AI processing of such information, how many times the document is to be processed by the agency or the public, the relative value of extracting such information, why the state-of-the-art and forthcoming AI techniques are unlikely to be capable of extracting the desired information, and the modifications to the document that could enable AI processing immediately or in the short-term. Document Standards to Facilitate AI Processing are intimately related to knowledge of state-of-the-art AI techniques and their respective capabilities. Only if we understand what current and forthcoming AI technology is capable of can we attempt to enable AI processing through an appropriate design of the documents. NIST and government agencies should have a strong interest in Document Standards to Facilitate AI Processing. Agencies frequently process documents submitted to the agency as well as produce documents for public consumption. AI processing of such documents would help in both situations. If the agency has the ability to influence or dictate the form of documents that are submitted, then	

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
				AI processing of agency documents could be accelerated even if the state-of-the-art is incapable of processing the documents in their current form. Finally, Document Standards to Facilitate AI Processing could be designed to promote the creation of training data for future AI systems. For example, part of the standard could require the document creator to apply tags or annotations to particular portions of the document. In conclusion, AI processing of documents could benefit government agencies and the public who interact with such agencies. Through appropriate Document Standards to Facilitate AI Processing, statutes, regulations, administrative decisions, documents to be submitted to agencies, and documents created by agencies could be reworked to enable AI processing. AI could then perform tasks that currently require tedious or error prone human labor. In summary, such standards could bridge the gap between AI capabilities and user's needs.	

July 1, 2019

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)