Barbarians at the Gate: AI Meets Content Publishing

While the steady advance of applied artificial intelligence (AI) technology promises to dramatically increase the value of computers in daily life – for example, helping physicians to diagnose diseases in their patients by mining population health data – AI and machine learning in competitive intelligence and customer insights applications are on a collision course with copyright law and the content publishing industry.

Traditionally, a search engine delivers up a list of documents with a brief summary in response to a user query. For example, a user wants to know what the sales forecast is for personal cloud services in 2020. The search engine directs the user to one or more documents that contain the forecast; the user is enticed to read these documents by a snippet or summary within the search result. Then the user clicks through or downloads one or more of the documents to find answers to their questions. This click-through or download supports the publishers' business models in one way or another.

Enter AI and machine learning. What if the machine can "read" the documents, summarize the material, and extract answers to users' questions *without* requiring users to read the documents? The user types in a search query for a forecast of sales of personal cloud services in 2020 and gets back a response from the search engine "$4.9 billion dollars" – no document download required. The market research report or reports the data was taken from are not "consumed" by the user. Perhaps only one copy of the report containing the data was purchased and read by the machine, but every user in the organization can now get answers to questions from the material. Or perhaps multiple reports from different publishers contributed to the answer.

Providers of search technology are excited by this looming jump in the utility of search and are falling all over themselves to establish their machine learning solutions in the marketplace for application developers to use. For their part, content consumers, in their quest for automated insights, want their AI-enabled search engine to have unfettered access to the largest, most robust, broadest, and most authoritative content sets – but they don't care if their users have direct access to the underlying documents.

What does this do to a premium content provider's subscriber base, content licensing revenue, and advertising revenue stream? How does it affect a firm's intellectual property rights? What are the copyright implications of getting "answers" based on automated analyses of documents instead of providing links to documents? What is the meaning of a content license when the content is analyzed by a machine, the intelligence consolidated with informational bits from other publishers, and answers given directly without the need to ever download a document to a

Visit us at northernlight.com

user? What changes are required to licensing and subscription models of the future to deal with these issues?

In this brave new world of machine learning-based cognitive search, aggregators may become content publishers' best friends. When the cross-publisher AI application sits at the aggregator rather than the user organization, aggregators can report to publishers how frequently, by whom, and for what purpose their content is accessed. On the other hand, if a publisher sends full-text content directly to a user organization, the usage details are hidden behind the customer's firewall and content consumption is a complete mystery.

From a content pricing perspective, having machines as "users" fundamentally alters (or obviates) the traditional notion of a "seat." One possible solution to this business dilemma may be an "AI license" priced along the lines of an enterprise license. Such a model could help ensure fair value when the "user" is a machine that can digest and synthesize content to produce answers to specific questions on a scale far beyond what an individual human researcher could do. The assumption would be that when an AI machine learning application consumes a document, everyone in the organization consumes it, too.

We are reaching a point of discontinuity in terms of the perceived value of intelligent search summaries (created automatically by computers) versus the value of the rich source documents they draw upon. Indeed, many content consumers will ascribe greater value to the summary that offers readily digestible "answers" derived automatically from documents from many publishers than to the underlying source documents.

As a result, the AI industry is on a collision course with the content publishing industry. When the collision occurs, it will be a conflagration of major proportions. Like the Roman sentries that warned about "barbarians at the gate" as Attila the Hun approached, the publishing community would be wise to pay close attention to the role their content will play in AI / machine learning-based enterprise search applications going forward.

# # #

(8/17)

Visit us at northernlight.com