



Chair, Won Kim Phd
won.kim@cyberdb.com

“Data Mining” Is NOT Against Civil Liberties

*Executive Committee on
ACM Special Interest Group on Knowledge Discovery and Data Mining*

June 30, 2003 (revised July 28, 2003)

We the undersigned are directors on the Executive Committee of ACM's SIGKDD (www.acm.org/sigkdd). ACM's SIGKDD (Special Interest Group on Knowledge Discovery and Data Mining) is the worldwide professional society created in 1998 to foster advances in the science and technology of automatic discovery of knowledge from stored databases. Leaders and many members of SIGKDD are experts on data mining, and SIGKDD is the best source of information and opinions about data mining that are grounded on technical expertise.

We write this letter because of the recent public portrayals of several US Government projects, including DARPA's Terrorism Information Awareness (TIA, formerly known as Total Information Awareness) project, the Transportation Security Administration's Computer Assisted Passenger Prescreening System (CAPPS II), and the FBI's Trilogy project that appear to create the misguided impression that these projects are synonymous with data mining technology, and that data mining technology is potentially a major threat to civil liberties and privacy. We are concerned that recent newsmedia reports and a recent press release from the US ACM Public Policy Committee (see http://www.acm.org/usacm/Letters/tia_final.html) may have contributed to this misimpression. We are concerned that the proposed S. 188 Data Mining Moratorium Act of 2003 (see <http://feingold.senate.gov/~feingold/releases/03/01/2003116745.html>) does not reflect a sound understanding of data mining science, technology or applications. Finally, we are concerned that the public debate has not distinguished between the research and development of data mining technology and the application and use of these technologies by specific agencies on specific data for specific purposes.

We believe that there are many important and legitimate policy issues that arise in the context of these projects and proposals, and that these issues involve a multitude of technologies and uses, far broader than those involving data mining. These issues are so important to society that it is essential that policy decisions be well-informed and carefully considered. We are offering to contribute our scientific expertise to the public debate so that any policy decisions will be well-informed by accurate information about data mining science, technology and applications. We are not, however, advocating any particular policy positions.

We believe that the current debate that portrays these Government projects as “developing massive data mining systems” is misleading and injurious to the large scientific community working on the research and development of data mining technology. Data mining is but one of many technologies that may be used in these projects. Other technologies include database management, online analytical processing, speech recognition, image (face, iris, fingerprint, etc.) recognition, natural language understanding and translation, data warehousing, data integration, information retrieval, etc. Does it make sense to attempt to outlaw any or all of these? Clearly the intent should be to regulate use and define policy, not restrict or hinder development of a technology or a field of science. Data mining technology is

**ASSOCIATION FOR COMPUTING MACHINERY
SPECIAL INTEREST GROUP ON KNOWLEDGE, DISCOVERY & DATA MINING**

www.acm.org/sigkdd/



Chair, Won Kim Phd
won.kim@cyberdb.com

an analysis technique, and is separate from issues of data collection and data aggregation. It can find patterns in medical drug design experiments, in biological micro-arrays, and many more life-saving applications, as well as countless business and scientific applications.

As any technology, data mining technology is not perfect, despite its long heritage, rooted in statistical analysis and artificial intelligence. The results of data mining indeed often include false positives and false negatives. However, the best (and perhaps only) way to overcome the "limitations" of data mining techniques is to do more research in data mining, including areas like data security and privacy-preserving data mining, which are actually active and growing research areas.

The main goal of this letter is to help differentiate the data mining technology from data collection and specific applications in specific domains. We believe that the most significant sources of danger to civil liberties are the unnecessary and unauthorized collection of data, and misuse of collected data, including the use of wrong data, the use of data in unauthorized ways, the wrong and unauthorized dissemination of data, and reaching wrong conclusions from data.

Since 1994, ACM SIGKDD has held an annual international research conference, KDD (Knowledge Discovery and Data Mining). The KDD-2003 conference will be held in Washington, D.C. on August 24-27. We believe it is an excellent opportunity for government representatives and policy makers to learn first hand what data mining really is, and equally importantly, what it is not. For the conference, we have assigned two Government Relations Co-Chairs as contacts to help facilitate these interactions. Details of KDD-2003 are available at <http://www.acm.org/sigkdd/kdd2003/>

Each of the directors on the Executive Committee of ACM SIGKDD is willing to be available as the scientific experts to support the news media and the policy community with accurate technical information about data mining.

Won Kim (won.kim@cyberdb.com), Chair, SIGKDD (CEO, Cyber Database Solutions)
Rakesh Agrawal (ragrawal@almaden.ibm.com), Director,
(IBM Almaden Research Center)
Christos Faloutsos (christos@cs.cmu.edu) Director, (Prof., CMU)
Usama Fayyad (fayyad@acm.org), Director (CEO, DigiMine)
Jiawei Han (hanj@cs.uiuc.edu), Director (Prof., University of Illinois)
Gregory Piatetsky-Shapiro (gps@kdnuggets.com), Director, (CEO, KDnuggets)
Daryl Pregibon (daryl@research.att.com), Director, (AT&T Labs-Research)
Ramasamy Uthurusamy (samy@gm.com), Director, (General Motors)

Note: This letter should not be interpreted as representing the views or policies of the organizations with which the Directors are affiliated.

**ASSOCIATION FOR COMPUTING MACHINERY
SPECIAL INTEREST GROUP ON KNOWLEDGE, DISCOVERY & DATA MINING**
www.acm.org/sigkdd/

Cyber Databases Solutions, 3445 Executive Center Drive, Suite 256, Austin, TX 78731 USA
Tel: +1-512-349-9757 * Fax: +1-512-349-9763