# Must Recommender Systems be Simplistic to be Trustable?

**Markus Stolze**
IBM Research,
Watson Research Center
1101 Kitchawan Rd, Route 134,
Yorktown Heights, NY 10598
mgstolze@us.ibm.com

Position statement for the workshop on
***Usable Artificial Intelligence at CHI 2008***


CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 978-1-60558-012-8/08/04.

## Abstract
Research shows that recommender system need to provide understandable explanations to be trustable. We propose a method to achieve such explanations without retreating to simplistic reasoning models.

## Keywords
Recommender Systems, Trust, Machine Learning, Explanation

## ACM Classification Keywords
H.3.3 [Information Search and Retrieval]: Information Filtering; H5.2 [Information interfaces and presentation]: User Interfaces; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces.

## Position Statement
Designing usable interfaces for an AI algorithm is challenging. The user interface and algorithm have to be designed in such a way that users know what input they are expected to provide, how and when that input should (or can) be provided, how the system output should be interpreted, and when the output can be trusted. Well documented challenges include the difficulty of building systems that clearly define their boundary of competency and provide gracefully degrading support under boundary conditions. Maglio et al. (2003) describe an instructive counter example in

which an airline accident was caused by the fact that a control system of an MD-83 airplane compensated for the slow deterioration of a plane component without notifying the pilots. The plane seemed to be in perfectly normal condition until the automatic system could no longer compensate for the abnormal condition. Under these conditions also the pilot could not keep the plane from spiraling out of control. A system that would have degraded more gracefully and informed pilots of the deteriorating condition had provided pilots with more time to plan for an emergency landing.

A related challenge is to make AI systems trustable for users. In our research on eCommerce Recommender systems we found that a key requirement for users to trust recommendations is the ability of the system to provide useful explanations how users' stated needs are fulfilled by the suggested products (Stolze & Nart, 2004, Felix et al. 2001). We found that users had problems trusting the Recommender system if it just provided a ranked list of products that did not indicate why exactly a given product was better than another in fulfilling the user's stated needs. In particular, we identified the desire of users to learn enough about a domain to convince themselves which product best matches their needs (Stolze & Ströbel, 2004). The importance of explanations is also confirmed by studies of other types of Recommender systems (Herlocker et al., 2000; Spiekermann & Parachiv, 2002; Sinha & Swearingen, 2002; O'Donovan & Smyth, 2005; Pu & Chen, 2006). Similar findings are reported in related domains such as mixed initiative planning and cognitive assistance (Cortellessa & Cesta, 2006; Glass et al., 2007). Explanations are also seen as a useful tool to support students and users to learning about a new domain (Patokorpi, 2007).

The requirement for making recommendations explainable and traceable can create an interesting dilemma. Results of algorithms will be easier to explain if they use a less sophisticated reasoning mechanism and rely on explicitly stated rules (or cases) instead of statistical analysis of dynamically collected information. This implies that there is the danger that these algorithms are basing their recommendation on rules that are very general and possibly outdated. In a pointed way one could therefore claim that system designers have to choose between two unsatisfying choices. The first option is to select an algorithm that is "smart" and low maintenance as it derives recommendations from statistical analysis of collected data. But this makes the algorithm difficult to explain and therefore difficult to be trusted by users. The second option is to select an algorithm that is less "smart", uses rules or cases that require manual maintenance and might outdate more quickly. But the way recommendations have been derived can be explained to the user, which makes it easier to be trusted.

In some domains this dilemma can be avoided. Stumpf et al. (2007) report on a study in the domain of e-mail classification. The best performing machine learning model was a decision tree. This model also was well suited to generate the rule-oriented explanations that where preferred by users over other forms of explanations. For explaining task execution of cognitive assistance agents McGuinnes et al. (2007) have developed an architecture that can provides different types of explanations depending on specific user needs. Their solution provides a uniform way of encoding task execution as well as deductive reasoning and learning components as task processing justifications. Thus, explanations are directly generated by introspection of the models and algorithms used for generating recommendations.

We expect that for some recommender systems independent models for learning and explanation will be most appropriate. For example, a recommender system could use a Bayesian model for learning and generating recommendations, but it could use a rule base in combination with abductive reasoning (Peng & Reggia) to generate plausible explanations that are easy to understand by users. This way recommendations can be generated from up-to-date statistical information while explanations can be generated based on a manually maintained knowledge base that links recommendations to established and well understood principles. Cases that cannot be properly explained with the current explanation model point to the need for updating the explanation model. We expect that the resulting system will be able to provide trustable recommendations without having to rely on over-simplified reasoning models. Empirical studies in real-world application domains will be needed to determine the usefulness and practicability of such a hybrid approach.

## References


[1] Cortellessa, G. and Cesta, A. (2006) Evaluating Mixed-Initiative Systems: An Experimental Approach. ICAPS-06

[2] O'Donovan, D, Smyth, B. (2005) Trust in recommender systems, Proceedings of the 10th international conference on Intelligent user interfaces, January 10-13, 2005, San Diego, California, USA

[3] Felix, D. , Niederberger, C. , Steiger, P. , Stolze M., (2001) Feature-Oriented vs. Needs-Oriented Product Access for Non-Expert-Online Shoppers, Proceedings of the IFIP Conference on Towards The E-Society: E-Commerce, E-Business, E-Government, p.399-406, October 03-05, 2001

[4] Glass, A., McGuinness, D. L., Wolverton, M. (2007) Toward Establishing Trust in Adaptive Agents, Proceedings of the International Conference on Intelligent User Interfaces (IUI-2007), ACM Press, 2007.

[5] Herlocker, J., Konstan, J., and Riedel, J. 2000. Explaining collaborative filtering recommendations. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW’00)

[6] McGuinness, D.L., Glass, A., Wolverton, M., and Pinheiro da Silva, P. Explaining Task Processing in Cognitive Assistants that Learn. Proceedings of the 20th International FLAIRS Conference (FLAIRS-20), 2007.

[7] Patokorpi, E. (2007). Logic of Sherlock Holmes in Technology Enhanced Learning. Educational Technology & Society, 10 (1), 171-185.

[8] Peng, Y., Reggia, J. A., (1990) Abductive Inference Models for Diagnostic Problem-Solving. Springer-Verlag, 1990.

[9] Pu, P., Chen, L. (2006) Trust building with explanation interfaces, Proceedings of the 11th international conference on Intelligent user interfaces, January 29-February 01, 2006, Sydney, Australia

[10] Sinha, R. , Swearingen, K. (2002) The role of transparency in recommender systems, CHI '02 extended abstracts on Human factors in computing systems, April 20-25, 2002, Minneapolis, Minnesota, USA

[11] Spiekermann, S., Parachiv, C. (2002) Motivating Human-Agent Interaction: Transferring Insights from Behavioral Marketing to Interface Design. Electronic Commerce Research, Vol. 2, No. 3, pp. 255-285.

[12] Stolze, M. Nart, F. (2004) Well-integrated needs-oriented recommender components regarded as helpful. Conference on Human Factors in Computing Systems extended abstracts (CHI ‘04)

[13] Stolze, M., Ströbel, M. (2004) Recommending as personalized teaching: towards credible needs-based ecommerce recommender systems, Designing personalized user experiences in eCommerce, Kluwer Academic Publishers, Norwell, MA, 2004.

[14] Stumpf, S., Rajaram, V., Li, L., Burnett, M., Dietterich, T., Sullivan, E., Drummond, R., and Herlocker, J. (2007) Toward Harnessing User Feedback for Machine Learning. Proceedings of the International Conference on Intelligent User Interfaces (IUI-2007), ACM Press, 2007.