



AI Profiles: An Interview with Thomas Dietterich

Marion Neumann (Washington University in St. Louis; m.neumann@wustl.edu)

DOI: [10.1145/3320254.3320258](https://doi.org/10.1145/3320254.3320258)

Introduction

Welcome to the eighth interview in our series profiling senior AI researchers. This month we are especially happy to interview our SIGAI advisory board member, Thomas Dietterich, Director of Intelligent Systems at the Institute for Collaborative Robotics and Intelligence Systems (CoRIS) at Oregon State University.



Figure 1: Tom Dietterich

Biography

Dr. Dietterich (AB Oberlin College 1977; MS University of Illinois 1979; PhD Stanford University 1984) is Professor Emeritus in the School of Electrical Engineering and Computer Science at Oregon State University, where he joined the faculty in 1985. Dietterich is one of the pioneers of the field of Machine Learning and has authored more than 200 refereed publications and two books. His research is motivated by challenging real world problems with a special focus on ecological science, ecosystem management, and sustainable development. He is best known for his work on ensemble methods in machine learning including the development of error-correcting output coding. Dietterich has also invented important reinforcement learning algorithms including the MAXQ method for hierarchical reinforcement learning. Dietterich has

Copyright © 2019 by the author(s).

devoted many years of service to the research community. He served as President of the Association for the Advancement of Artificial Intelligence (2014-2016) and as the founding president of the International Machine Learning Society (2001-2008). Other major roles include Executive Editor of the journal *Machine Learning*, co-founder of the *Journal for Machine Learning Research*, and Program Chair of AAAI 1990 and NIPS 2000. Dietterich is a Fellow of the ACM, AAAI, and AAAS.

Getting to Know Tom Dietterich

When and how did you become interested in CS and AI?

I learned to program in Basic in my early teens; I had an uncle who worked for GE on their time-sharing system. I learned Fortran in high school. I tried to build my own adding machine out of TTL chips around that time too. However, despite this interest, I didn't really know what CS was until I reached graduate school at the University of Illinois. I first engaged with AI when I took a graduate assistant position with Ryszard Michalski on what became machine learning, and I took an AI class from Dave Waltz. I had also studied philosophy of science in college, so I had already thought a bit about how we acquire knowledge from data and experiment.

What would you have chosen as your career if you hadn't gone into CS?

I had considered going into foreign service, and I have always been interested in policy issues. I might also have gone into technical management. Both of my brothers have been successful technical managers.

What do you wish you had known as a Ph.D. student or early researcher?

I wish I had understood the importance of strong math skills for CS research. I was a software engineer before I was a computer

science researcher, and it took me a while to understand the difference. I still struggle with the difference between making an incremental advance within an existing paradigm versus asking fundamental questions that lead to new research paradigms.

What professional achievement are you most proud of?

Developing the MAXQ formalism for hierarchical reinforcement learning.

What is the most interesting project you are currently involved with?

I'm fascinated by the question of how machine learning predictors can have models of their own competence. This is important for making safe and robust AI systems. Today, we have ML methods that give accurate predictions in aggregate, but we struggle to provide point-wise quantification of uncertainty. Related to these questions are algorithms for anomaly detection and open category detection. In general, we need AI systems that can work well even in the presence of "unknown unknowns".

Recent advances in AI led to many success stories of AI technology undertaking real-world problems. What are the challenges of deploying AI systems?

AI systems are software systems, so the main challenges are the same as with any software system. First, are we building the right system? Do we correctly understand the users' needs? Have we correctly expressed user preferences in our reward functions, constraints, and loss functions? Have we done so in a way that respects ethical standards? Second, have we built the system we intended to build? How can we test software components created using machine learning? If the system is adapting online, how can we achieve continuous testing and quality assurance? Third, when ML is employed, the resulting software components (classifiers and similar predictive models) will fail if the input data distribution changes. So we must monitor the data distribution and model the process by which the data are being generated.

This is sometimes known as the problem of "model management". Fourth, how is the deployed system affecting the surrounding social and technical system? Are there unintended side-effects? Is user or institutional behavior changing as a result of the deployment?

One promising approach is combining humans and AI into a collaborative team. How can we design such a system to successfully tackle challenging high-risk applications? Who should be in charge, the human or the AI?

I have addressed this in a recent short paper (Robust Artificial Intelligence and Robust Human Organizations. *Frontiers of Computer Science*, 13(1): 1-3). To work well in high-risk applications, human teams must function as so-called "High reliability organizations" or HROs. When we add AI technology to such teams, we must ensure that it contributes to their high reliability rather than disrupting and degrading it. According to organizational researchers, HROs share five main practices: (a) continuous attention to anomalous and near-miss events, (b) seeking diverse explanations for such events, (c) maintaining continuous situational awareness, (d) practicing improvisational problem solving, and (e) delegating decision making authority to the team member who has the most expertise about the specific decision regardless of rank. AI systems in HROs must implement these five practices as well. They must be constantly watching for anomalies and near misses. They must seek multiple explanations for such events (e.g., via ensemble methods). They must maintain situational awareness. They must support joint human-machine improvisational problem solving, such as mixed-initiative planning. And they must build models of the expertise of each team member (including themselves) to know which team member should make the final decision in any situation.

You ask "Who is in charge?" I'm not sure that is the right question. Our goal is to create human-machine teams that are highly reliable as a team. In an important sense, this means every member of the team has responsibility for robust team performance. However, from an ethical standpoint, I think the human team leader should have ultimate responsibility. That task of taking action in a specific situ-

ation could be delegated to the AI system, but the team leader has the moral responsibility for that action.

Moving towards transforming AI systems into high-reliable organizations, how can diversity help to achieve this goal?

Diversity is important for generating multiple hypotheses to explain anomalies and near misses. Experience in hospital operating rooms is that often it is the nurses who first detect a problem or have the right solution. The same has been noted in nuclear power plant operations. Conversely, teams often fail when they engage in “group think” and fixate on an incorrect explanation for a problem.

How do you balance being involved in so many different aspects of the AI community?

I try to stay very organized and manage my time carefully. I use a machine learning system called TAPE (Tagging Assistant for Productive Email) developed by my collaborator and student Michael Slater to automatically tag and organize my email. I also take copious notes in OneNote. Oh, and I work long hours...

What was your most difficult professional decision and why?

The most difficult decision is to tell a PhD student that they are not going to succeed in completing their degree. All teachers and mentors are optimistic people. When we meet a new student, we hope they will be very successful. But when it is clear that a student isn't going to succeed, that is a deep disappointment for the student (of course) but also for the professor.

What is your favorite AI-related movie or book and why?

I really don't know much of the science fiction literature (in books or films). My favorite is 2001: A Space Odyssey because I think it depicts most accurately how AI could lead to bad outcomes. Unlike in many other stories, HAL doesn't “go rogue”. Instead, HAL creatively achieves the objective programmed by

its creators. Unfortunately, as a side effect, it kills the crew.



Help us determine who should be in the AI Matters spotlight!

If you have suggestions for who we should profile next, please feel free to contact us via email at aimatters@sigai.acm.org.