Ways of knowing for AI: (chat)bots as interfaces for machine learning

Q. Vera Liao
Michael Muller

IBM Research
- drive down desired information paths
- dynamic and adaptive interface
- personified/anthropomorphic designs
- relatively low cost for language understanding and knowledge base
Knowledge elicitation for machine learning

Towards more **efficient** learning algorithms that **interact** with people

- **Interactive**
  - e.g. *interactive ML* allows domain experts to examine the model output and **incrementally update** the knowledge input

- **Proactive**
  - e.g. *active learning* relies on the learning algorithm to **request knowledge for selected items** (e.g. most uncertain ones)

- **Rich forms of knowledge**
  - e.g. *weak supervision* allows higher-level knowledge input in rules and heuristics

---

interactive

proactive

rich forms of knowledge

I have learned to classify messages with 85% accuracy. Try testing me a billing related question?

The following message is asking about billing, right?

What are some common phrases people use when talking about billing?

Why is this message about billing? Can you explain the pattern to me?
Design issues (an ongoing list)

Design goals: **articulation**, **efficiency**, **robustness**, **engagement**

- **Knowledge elicitation techniques**: “prompts”, dialogue structure, probes, etc.

- **Bridging natural conversational behaviors and formal input for learning algorithms** (socially guided machine learning Thomaz & Breazel 2008)

- **Bot persona**: mental model, mental model, mental model, mental model
Teacheable agent: a naive student model?

- A natural fit for “learner”
- Leveraging teaching interactions and teaching/learning theories
- But, leading to behavioral biases
- May not be a common interface for diverse models and human knowledge sources

(Cakmak & Thomaz, 2012) (Bradesko et al., 2010)
Examples of knowledge needed for ML models

- **Type of knowledge**
  - Tacit
  - Explicit
  
  **Values**
  - Knowledge structure, reasoning process
  - Domain concepts
  - Instance labels

Examples of HCI methods

- **Type of design knowledge**
  - Open-ended
  - Close-ended
  - Speculative
  - Inferential
  - Descriptive
  - Evaluative

- **Design fictions**
  - Interview, verbal protocol, task analysis
  - Card sorting, structured interview, survey

- **Questionnaire, AB testing**
A common knowledge input interface for ML:
A toolbox of knowledge elicitation methods with the goal of learning the task domain, understanding diverse stakeholders, and ultimately, optimization centered around user needs and societal benefits.

- Best practices in elicitation techniques and procedures
- Novel “probes”
- Overcoming biases
- Ethical considerations in interacting with human subjects
HCI researcher bot?