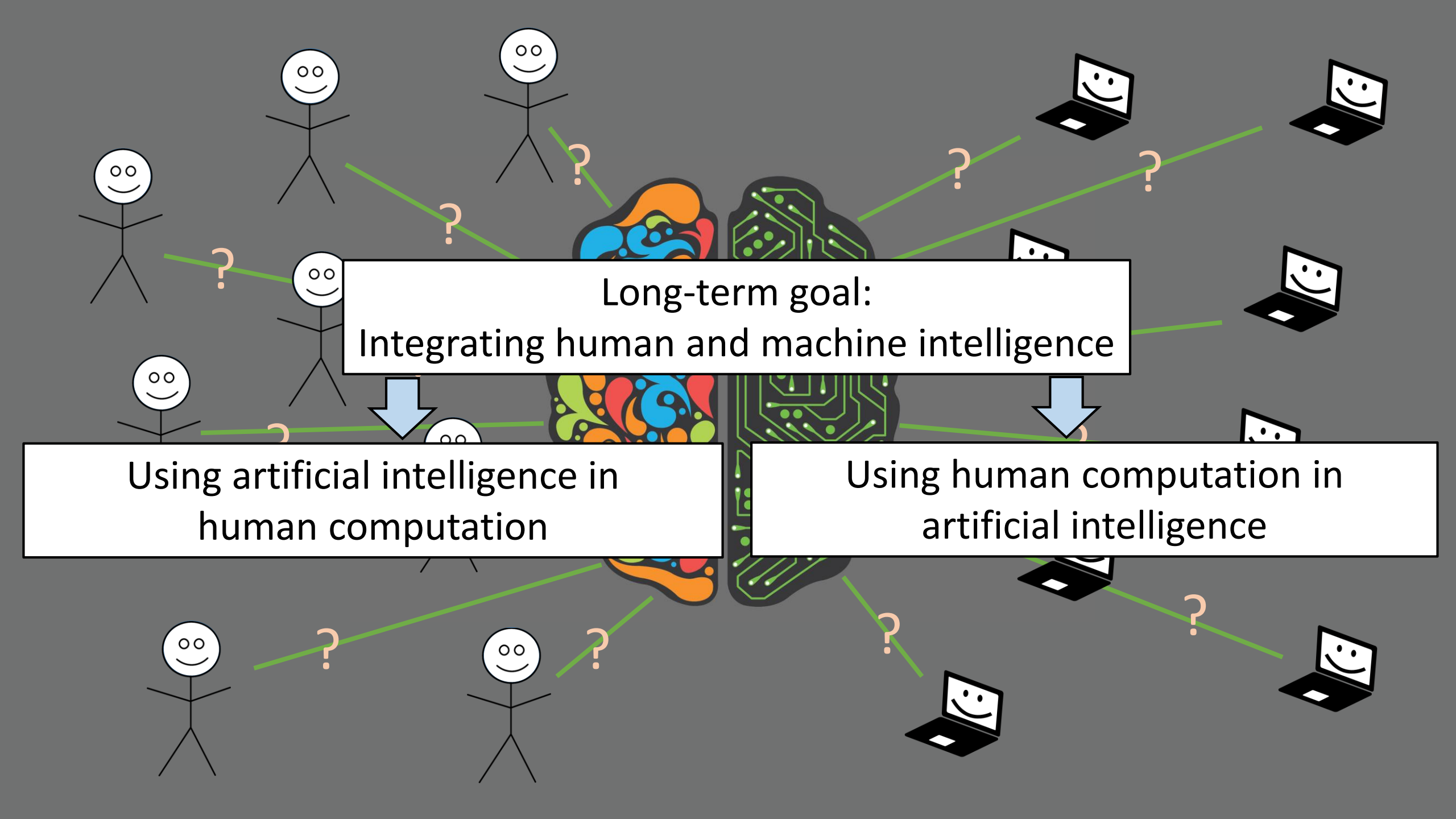
CS 5306
INFO 5306:
Crowdsourcing and Human Computation

Lecture 22

11/16/17

Haym Hirsh


Long-term goal:
Integrating human and machine intelligence
Using artificial intelligence in human computation
Using human computation in artificial intelligence

“artificial artificial intelligence”: Human Computation as “Fake” AI

New technology allows software to tap real human intelligence.

In the 50 years since John McCarthy coined the term *artificial intelligence*, much progress has been made toward identifying, understanding, and automating many classes of symbolic and computational problems that were once the exclusive domain of human intelligence. Much work remains in the field because humans still significantly outperform the most powerful computers at completing such simple tasks as identifying objects in photographs—something children can do even before they learn to speak.

with a fine-tooth comb and often scrapped because the cost of establishing and managing a network of skilled people to do the work outweighs the value of completing it. If software developers could programmatically access and incorporate human intelligence into their applications, a whole new class of innovative businesses and applications would be possible. This is the goal of Amazon Mechanical Turk:[1] to give software developers and businesses the power to use human intelligence as a core component of their applications and businesses. With Amazon Mechanical Turk, people are freer to innovate because they can now imbue software with real human intelligence.

In 1769, Wolfgang von Kempelen built an automaton that defeated many human opponents at chess. Known as "The Turk," the wooden mannequin toured the United States and Europe for many years, defeating such famous challengers as Benjamin Franklin, Napolean Bonaparte, and Edgar Allen Poe.[2] The secret to the automaton was, of course, a human chess master hidden inside. Like its namesake, Amazon's

tasks, and incorporate the answers into their software applications. To the application, the transaction looks very much like any remote procedure call: The application sends the request, and the service returns the results. In reality, a network of humans fuels this "artificial artificial intelligence" by coming to the Web site, searching for and completing tasks, and receiving payment for their work. This allows software developers to easily and economically build programs that tap into a worldwide, massively parallel, Internet-scale human workforce on an incremental, as-needed basis.

“AI gets a brain”
Barr, J. and Cabrera, L.F.
Queue, *4*(4), 2006

- “humans still significantly outperform the most powerful computers at completing such simple tasks as identifying objects in photographs—something children can do even before they learn to speak”
- “If software developers could programmatically access and incorporate human intelligence into their applications, a whole new class of innovative businesses and applications would be possible”
- “With Amazon Mechanical Turk, people are freer to innovate because they can now imbue software with real human intelligence.”

“AI gets a brain”
Barr, J. and Cabrera, L.F.
Queue, *4*(4), 2006

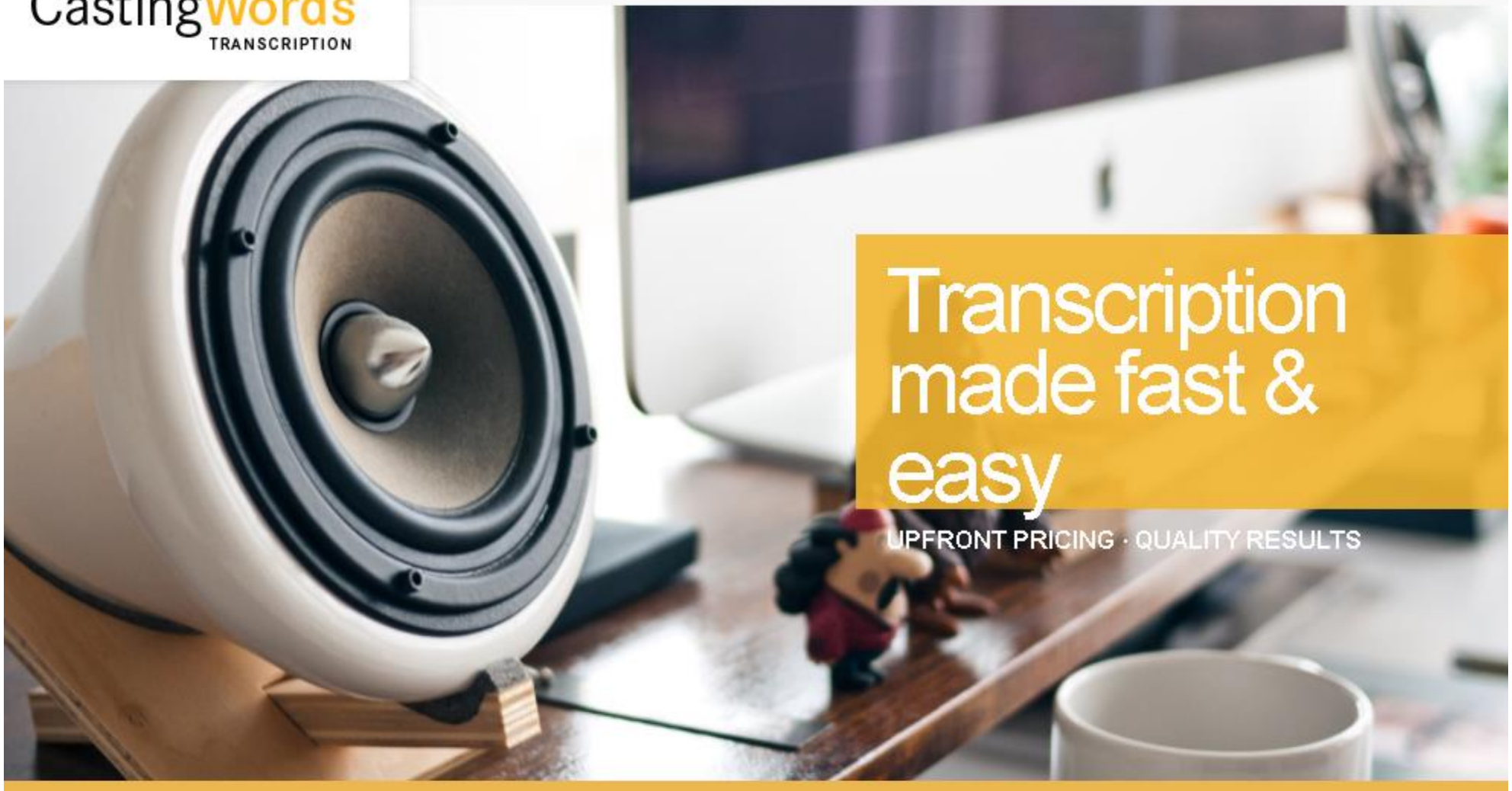
- “To the application, the transaction looks very much like any remote procedure call: The application sends the request, and the service returns the results.”
- “In reality, a network of humans fuels this “artificial artificial intelligence” by coming to the Web site, searching for and completing tasks, and receiving payment for their work.”
- “This allows software developers to easily and economically build programs that tap into a worldwide, massively parallel, Internet-scale human workforce on an incremental, as-needed basis.”

“AI gets a brain”
Barr, J. and Cabrera, L.F.
Queue, *4*(4), 2006

- Amazon internal uses:
  - Data quality: “data improvement and validation processes on its product catalog”
  - Japanese text orientation
  - Image selection: “After automatic processing has chosen several candidate images, human intelligence is used to choose the best possible image to represent each street address and business”
- Others
  - Podcast transcription: castingwords.com
  - Language translation: English-to-French and French-to-English translation
  - Data gathering: “lists of “Top 3” items (restaurants, theaters, and so forth) on a city-by-city basis”
  - Image tagging
  - Web site review: “answer a series of multiple choice questions”
  - Marketing survey
  - Sound verification
  - Facial image verification

CastingWords
TRANSCRIPTION

TRANSCRIPTION | SUPPORT | name@example.com password Log In | CART



Transcription made fast & easy

UPFRONT PRICING · QUALITY RESULTS

3 EASY STEPS

Choose a service to begin:

- **BUDGET** $1.00/minute
- **1 WEEK** $1.50/minute
- **1 DAY** $2.50/minute
- **INTERNATIONAL** $1.75/minute

Find HITs containing transcribe that pay at least $ 0.00 for which you are qualified require Master Qualification GO

Complete Profile Tasks to qualify for more HITs

Click here to add or update your profile information. By providing this information, you may qualify for HITs from Requesters looking for Workers like you.

HITs containing 'transcribe'

1-10 of 97 Results

Sort by: HITs Available (most first) GO! Show all details | Hide all details 1 2 3 4 5 › Next Items per Page: 10

Transcribe up to 35 Seconds of Media to Text - Earn up to $0.17 per HIT!! View a HIT in this group

Requester: Crowdsurf Support **HIT Expiration Date:** Nov 15, 2018 (51 weeks 6 days) **Reward:** $0.05
Time Allotted: 15 minutes

Transcribe US Elections Data from PDF View a HIT in this group

Requester: UofU RA **HIT Expiration Date:** Nov 29, 2017 (1 week 5 days) **Reward:** $0.05
Time Allotted: 60 minutes

Transcription of code-switched Hindi-English dialog speech data View a HIT in this group

Requester: ETS NLP Research **HIT Expiration Date:** Dec 1, 2017 (2 weeks 1 day) **Reward:** $0.35
Time Allotted: 2 hours

SmartScan - Merchant and Category View a HIT in this group

Requester: Expensify Infra **HIT Expiration Date:** Nov 17, 2017 (23 hours 57 minutes) **Reward:** $0.00
Time Allotted: 10 minutes

SmartScan - Amount and Currency View a HIT in this group

Requester: Expensify Infra **HIT Expiration Date:** Nov 17, 2017 (23 hours 57 minutes) **Reward:** $0.00
Time Allotted: 10 minutes

SmartScan - Date View a HIT in this group

Requester: Expensify Infra **HIT Expiration Date:** Nov 17, 2017 (23 hours 57 minutes) **Reward:** $0.00
Time Allotted: 10 minutes

Review, edit, and score the transcription of up to 35 seconds of media - Earn up to $0.14 per HIT! View a HIT in this group

Expensify

OUR PRODUCT

INTEGRATIONS

INDUSTRIES

PARTNERS

CUSTOMERS

SIGN IN

Expense reports that don't suck!

From receipt scanning to reimbursement, Expensify automates every step of the expense reporting process.

Email Address | SIGN UP FOR FREE

REQUEST A DEMO SEE OUR FEATURES

Preferred Partner Solution of CPA.COM an AICPA Company

.:: Merchant Name Guidelines ::.

Follow these instructions for entering the correct merchant on a receipt.

1. In most cases, enter the merchant name exactly as it is written on the receipt.
    - Any punctuation in the merchant name must be included exactly as it is on the receipt.
    - Some receipts may have a legal name, bank name, or atm name at the top of the receipt. The actual

1 / 6

Submit

Merchant

Category

** Most Commonly assigned categories for this merchant

Can't read receipt

Next

Expensify Infra [Receipt Categorizations]

Discussion in 'Requesters' started by starvingstatic, Sep 22, 2017.

Page 1 of 2 1 2 Next >

starvingstatic
Expensify Infra

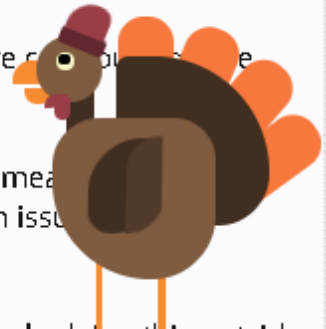
Hey all! It's been a while--last I chatted was over on mturkgrind.com. My more recent posts there were how I got an official job, and that job does things that might've transitioned well to MTurk (discussions starting here and in subsequent posts).

Well, great news--now that stuff *is* on Turk! If you need proof I work here, my name's Keagan and I'm on this page. Here's the HIT group we have up at the moment.

That said, I wanted to come here because I saw some pretty bad reviews on turkopticon and assume the difficulties we're having rolling this out are translating to those reviews, which then tend to propagate like wildfire. We're actually at the beginning of rolling out, and I really don't want this to be a permanent impression, so I'm here to serve as a bit of a liaison because I know how important it is to workers to have quality hits, and I know how important it is to Expensify to be a quality requester.

So--for those that have worked on these, how can we help? For those who haven't but are just looking now, how can we make these HITs better (e.g., instructions more clear, etc.)?

Essentially--how can I make everything as awesome as possible so this is a great HIT group that we c[illegible] e awesome pool of high-quality HIT workers to enjoy?



For now, one little disclaimer I want to pass along: we're still rolling this out and it's definitely not mea[illegible] mainstream at this time, so please give us a hot minute (a week or so I'd say?) to iron out the main issu[illegible] working on it, I promise!

Labeling Data

AI = Machine Learning

Machine Learning = Human-Labeled Data

Human-Labeled Data = Human Computation

Labeling Data

AI = Machine Learning

Machine Learning = Human-Labeled Data

Human-Labeled Data = Human Computation

AI = Human Computation

Labeling Data

AI = Machine Learning

Machine Learning = Human-Labeled Data

Human-Labeled Data = Human Computation

AI = Human Computation

(just get me more data)

Labeling Data

- Main application areas:
  - Computer vision
  - Natural language processing

Labeling Data: Images

- Kanefsky, B., Barlow, N.G. and Gulick, V.C., 2001. Can distributed volunteers accomplish massive data analysis tasks. *Lunar and Planetary Science*, *1*.
  - 800 people, 30,000 craters, 4 days
- 2004: ESP Game
  - 100,000 images with tags
- Westphal, A.J., Butterworth, A.L., Snead, C.J., Craig, N., Anderson, D., Jones, S.M., Brownlee, D.E., Farnsworth, R. and Zolensky, M.E., 2005. Stardust@ home: a massively distributed public search for interstellar dust in the stardust interstellar dust collector.
  - Videos, current top score 2369825 (max per item is 85)
- Raddick, J., Lintott, C.J., Schawinski, K., Thomas, D., Nichol, R.C., Andreescu, D., Bamford, S., Land, K.R., Murray, P., Slosar, A. and Szalay, A.S., 2007, December. Galaxy Zoo: an experiment in public science participation. In *Bulletin of the American Astronomical Society* (Vol. 39, p. 892).
  - >250,000 galaxies

Labeling Data: Images

- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 248-255). IEEE.
  - 2010: >14,000,000 images with Wordnet tags
- Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), pp.82-97.

Labeling Data: Natural Language

- 1989: Penn Treebank
- Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.Y., 2008, October. Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 254-263). Association for Computational Linguistics.
- Alonso, O., Rose, D.E. and Stewart, B., 2008, November. Crowdsourcing for relevance evaluation. In *ACM SigIR Forum*(Vol. 42, No. 2, pp. 9-15). ACM.

Labeling Data: Natural Language

- Callison-Burch, C., 2009, August. Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 286-295). Association for Computational Linguistics.
- Rada Mihalcea and Carlo Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the ACL- IJCNLP 2009
- Liu, F. and Liu, Y., 2009, August. From extractive to abstractive meeting summaries: Can it be done by sentence compression?. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 261-264). Association for Computational Linguistics.
- Alonso, O. and Mizzaro, S., 2009, July. Can we get rid of TREC assessors? Using Mechanical Turk for relevance assessment. In *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation* (Vol. 15, p. 16).
- Alonso, O. and Mizzaro, S., 2009, July. Relevance criteria for e-commerce: a crowdsourcing-based experimental analysis. In *Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval* (pp. 760-761). ACM.
- Alonso, O., 2009. Guidelines for designing crowdsourcing-based relevance experiments. In *Proceedings of the 32nd international acm sigir conference on research and development in information retrieval*.
- Kazai, G. and Milic-Frayling, N., 2009, July. On the evaluation of the quality of relevance assessments collected through crowdsourcing. In *SIGIR 2009 Workshop on the Future of IR Evaluation* (p. 21).
- Hsueh, P.Y., Melville, P. and Sindhwani, V., 2009, June. Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 workshop on active learning for natural language processing* (pp. 27-35). Association for Computational Linguistics.
- Ambati , V. and Carbonell, J. 2009. Proactive Learning for Building Machine Translation Systems for Minority Languages. HLT '09 Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing

Human Computation in Machine Learning

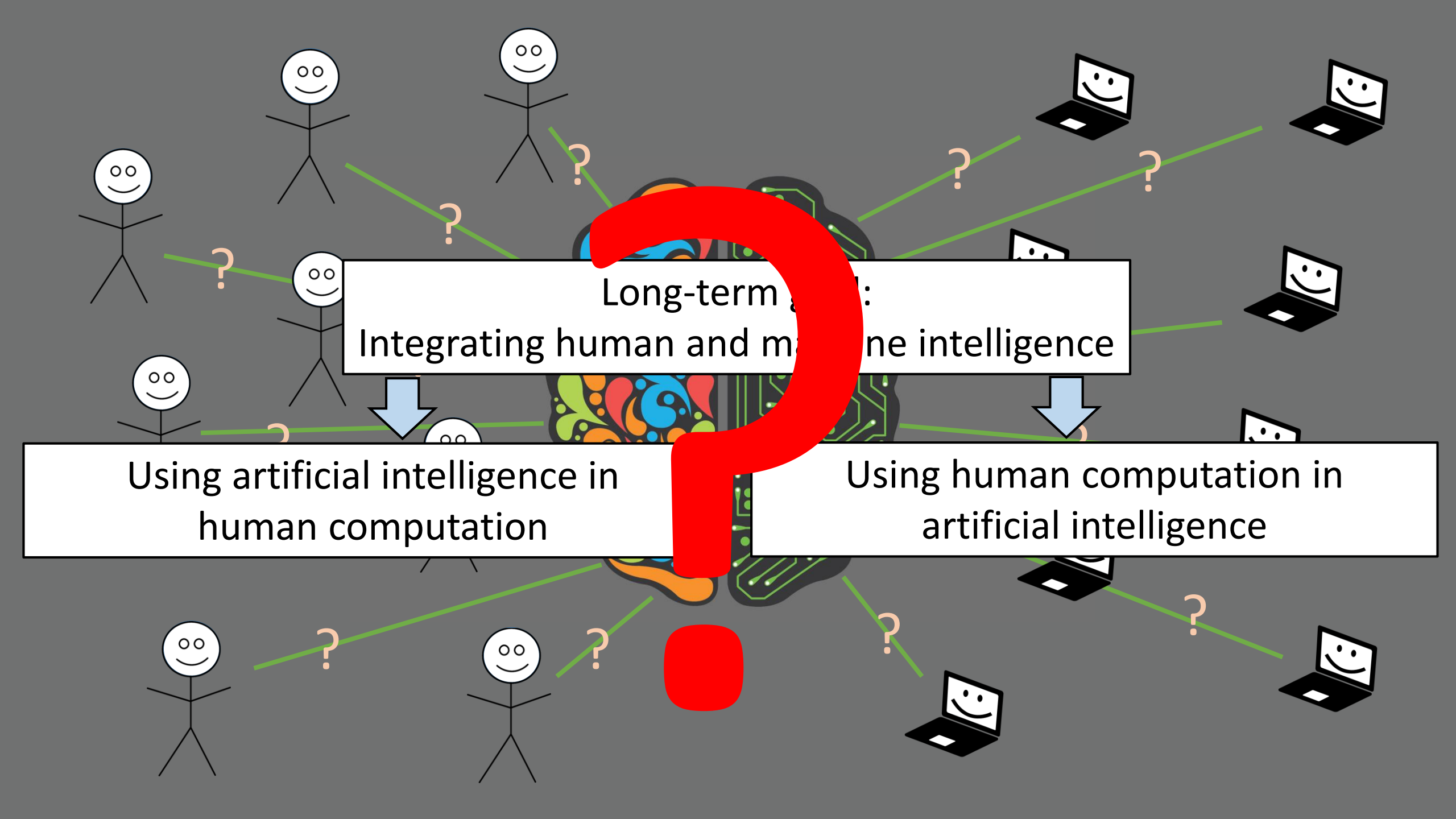
- Clustering:
  - Heikinheimo, H. and Ukkonen, A., 2013. The crowd-median algorithm. In *First AAAI Conference on Human Computation and Crowdsourcing*.
  - Tamuz, O., Liu, C., Belongie, S., Shamir, O. and Kalai, A.T., 2011. Adaptively learning the crowd kernel. *Proceedings of the 28th International Conference on Machine Learning*.

Machine Learning with Humans “In the Loop”

- Iterate between learning and human labeling
  - “Visual recognition with humans in the loop”, Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P. and Belongie, S., ECCV 2010.
- Game does machine learning behind the scene, guiding the game
  - “Game-powered machine learning”, Luke Barrington, Douglas Turnbull, and Gert Lanckriet, *Proceedings of the National Academy of Sciences*, *109*(17), pp.6411-6416, 2012.
- Iterate between learning and feature elicitation
  - “Crowdsourcing Feature Discovery via Adaptively Chosen Comparisons”, Chaudhuri, K. and Kalai, A., *HComp* 2015.

Computer Vision/Robotics with Human Computation

- Game with two players, one is human controller other “robot”, learn from how they interact
  - “Crowdsourcing human-robot interaction: New methods and system evaluation in a public environment”, Breazeal, C., DePalma, N., Orkin, J., Chernova, S. and Jung, M., *Journal of Human-Robot Interaction, 2013*
- Robot reinforcement learning with humans providing the feedback signal with clicks
  - “Robot reinforcement learning using crowdsourced rewards”, Penaloza, C.I., Chernova, S., Mae, Y. and Arai, T., *Proceedings of the IROS Workshop on Cloud Robotics*, 2013
- Use crowd for o dialog authoring, dialog editing, and nonverbal behavior authoring
  - “Semi-situated learning of verbal and nonverbal content for repeated human-robot interaction”, Leite, I., Pereira, A., Funkhouser, A., Li, B. and Lehman, J.F., *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016
- Get samples of natural language commands using AMT, get judgments of robot success using AMT
  - "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation“, Tellex, S. et al, AAAI 2011


Long-term goal:
Integrating human and machine intelligence
Using artificial intelligence in human computation
Using human computation in artificial intelligence

- “Generalized Task Markets for Human and Machine Computation”, Shahaf, D. and Horvitz, E., AAAI 2010.

From: AAAI-82 Proceedings. Copyright ©1982, AAAI (www.aaai.org). All rights reserved.

USER PARTICIPATION IN THE REASONING PROCESSES OF EXPERT SYSTEMS

Martha E. Pollack, Julia Hirschberg and Bonnie Webber

Department of Computer and Information Science
The Moore School
University of Pennsylvania, Philadelphia, Pa. 19104

We argue that expert systems, if they are to satisfy the legitimate needs of their users, must include dialogue capabilities as sophisticated as those proposed in current Natural Language research. In particular they must allow the user to direct the flow of the dialogue and must take into account the user's goals and expectations both in analyzing the user's statements and in providing appropriate responses. Our claims are corroborated by an analysis of transcripts of a "naturally occurring" expert system, a radio talk show in which callers ask an expert for financial advice. We present data demonstrating that user-expert dialogues are best viewed as a negotiation process, and we describe the exchanges that compose the dialogue in terms of the motivations, goals, strategies, and moves of the participants.*

To date, user participation in the reasoning processes of expert systems has been largely limited to probing expert reasoning or adding limited information. The user may only ask why the system requested more information and how it arrived at its advice. Research into extending the capabilities of expert systems [1,2,3] has so far failed to recognize the need to permit the full range of interactions possible when humans engage in the normal giving and getting of advice.

In order to delineate patterns of user-expert interaction in the reasoning process, we taped one week, February 1-5, 1982, of the radio talk show "Harry Gross: Speaking of Your Money" from station WCAU in Philadelphia. In this program listeners call in, introduce themselves and ask for financial advice, which the expert then attempts to provide. The ensuing discourse is a "naturally occurring" dialogue between a human user and a human expert. While the protocols we obtained should obviously not be seen as the ideal toward which all expert systems should aspire, despite the oral nature of this naturally occurring expert system, despite the entertainment element involved, and despite its domain specificity and the type of user toward which it is aimed, we believe that the major requirements of truly flexible and cooperative expert systems can be determined, in a principled way, from an explication of the relationship between human user input and human expert response.

From the radio show we collected twelve and one-half hours of user-expert interaction, involving 120 callers. An examination of these protocols reveals a regular pattern of interaction, which we describe as negotiation, the process whereby people arrive at a conclusion by means of a discussion. Rarely does a caller simply state a problem and passively listen to the expert's response. Rather the caller actively participates in the definition and resolution of the problem. Caller and expert must often negotiate to determine the statement of a problem the expert can solve and the statement of a solution the expert can support and the caller accept -- and, ideally, understand. They may also need to negotiate a common understanding of terminology, a common set of world or domain beliefs, or an acceptable justification for the solution.

Given this view of the interaction between user and expert as a negotiation process, one can characterize the particulars of each verbal exchange in terms of the participants' motivations, goals, strategies and moves . In order to generate a response appropriate to the user's participation, the expert must first recognize the user's motivation(s), goal(s) and strategy(s) through an examination of his/her move in the context of the discourse history and the user model, and then, based on that analysis and his own motivation(s), determine his own goal(s), strategy(s) and move.

When the user decides to participate s/he has some motive for doing so and some goal s/he hopes to achieve. Motivation answers the question "Why does the user decide to participate?", while goal answers "What does the user hope to achieve through participation?". The user then adopts some strategy by which s/he attempts to achieve the goal(s) s/he has set; so, strategy answers the question "How does the user attempt to achieve the goal?". This strategy is realized linguistically in an utterance, the user's move. For example, a user may be motivated by his/her surprise at an expert's advice: s/he may have thought of and rejected an answer because s/he believes it violates some specific constraint s/he wants met. His/her goal is then to gain assurance that the the expert's response meets that constraint. S/he might try to achieve this goal

*A longer version of this paper appears as University of Pennsylvania Technical Report #TR-CIS-82-9.

“Mixed-Initiative Interaction”

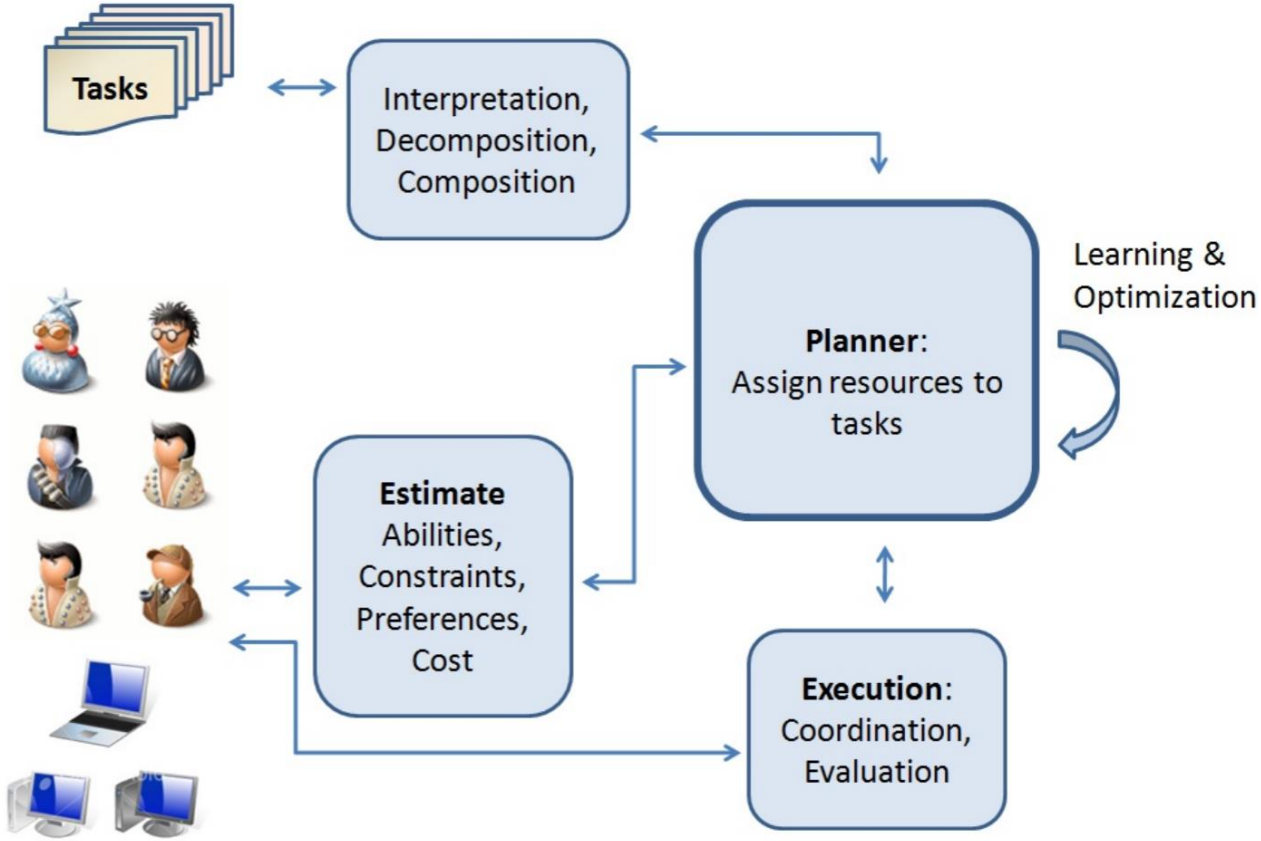
- Pollack, M.E., Hirschberg, J. and Webber, B., 1982. User participation in the reasoning processes of expert systems(pp. 356-361). *Proceedings AAAI*.
- Luff, P. and Frohlich, D.M., 1991. Mixed initiative interaction. *Knowledge based systems and legal applications*.
- …

“Human-Robot Interaction”

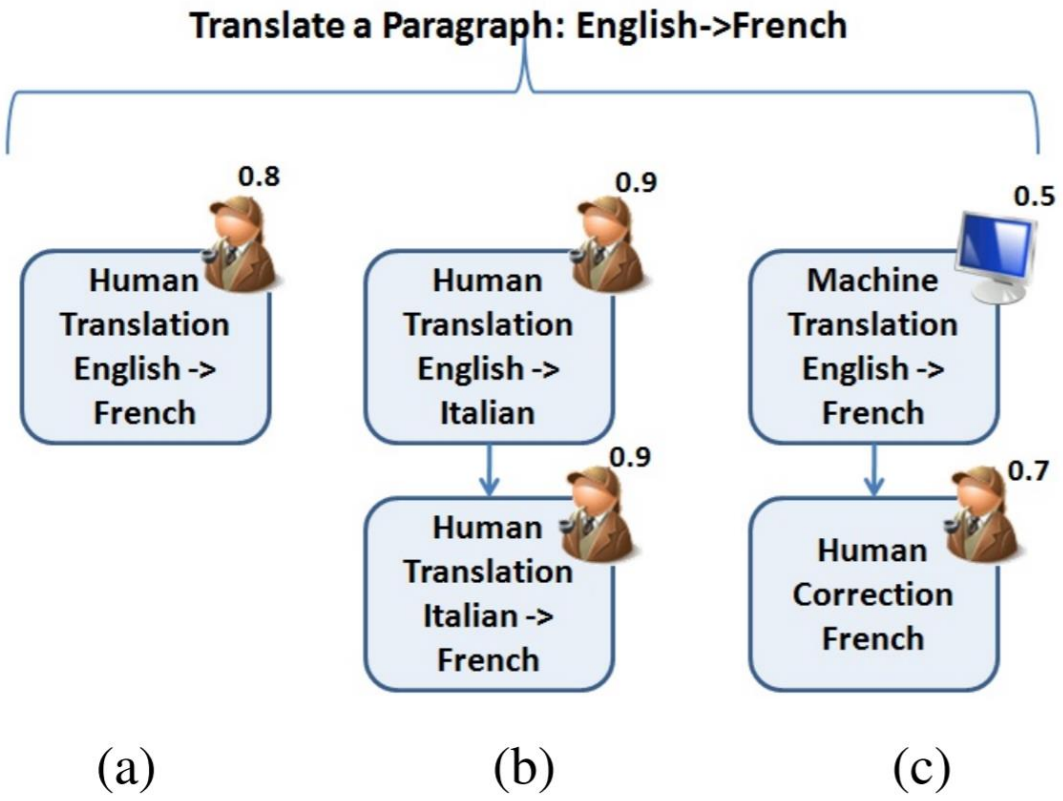
- Kazerooni, H., 1989, May. Human/robot interaction via the transfer of power and information signals. I. Dynamics and control analysis. In *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on* (pp. 1632-1640). IEEE.
- Kazerooni, H., 1990. Human-robot interaction via the transfer of power and information signals. *IEEE Transactions on systems, Man, and Cybernetics*, *20*(2), pp.450-463.
- …

- “Generalized Task Markets for Human and Machine Computation”, Shahaf, D. and Horvitz, E., AAAI 2010.

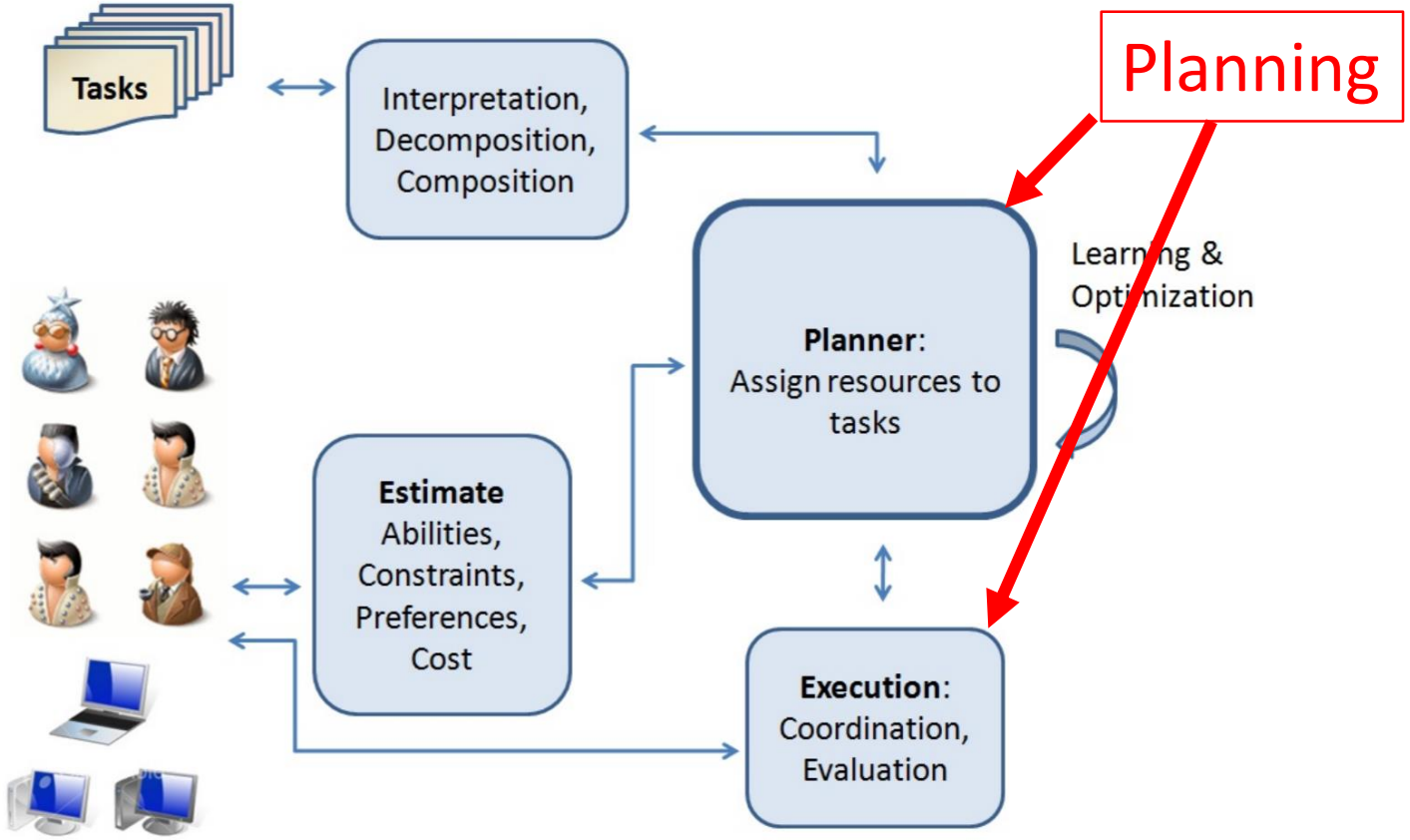
Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.



Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.



Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.



Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.

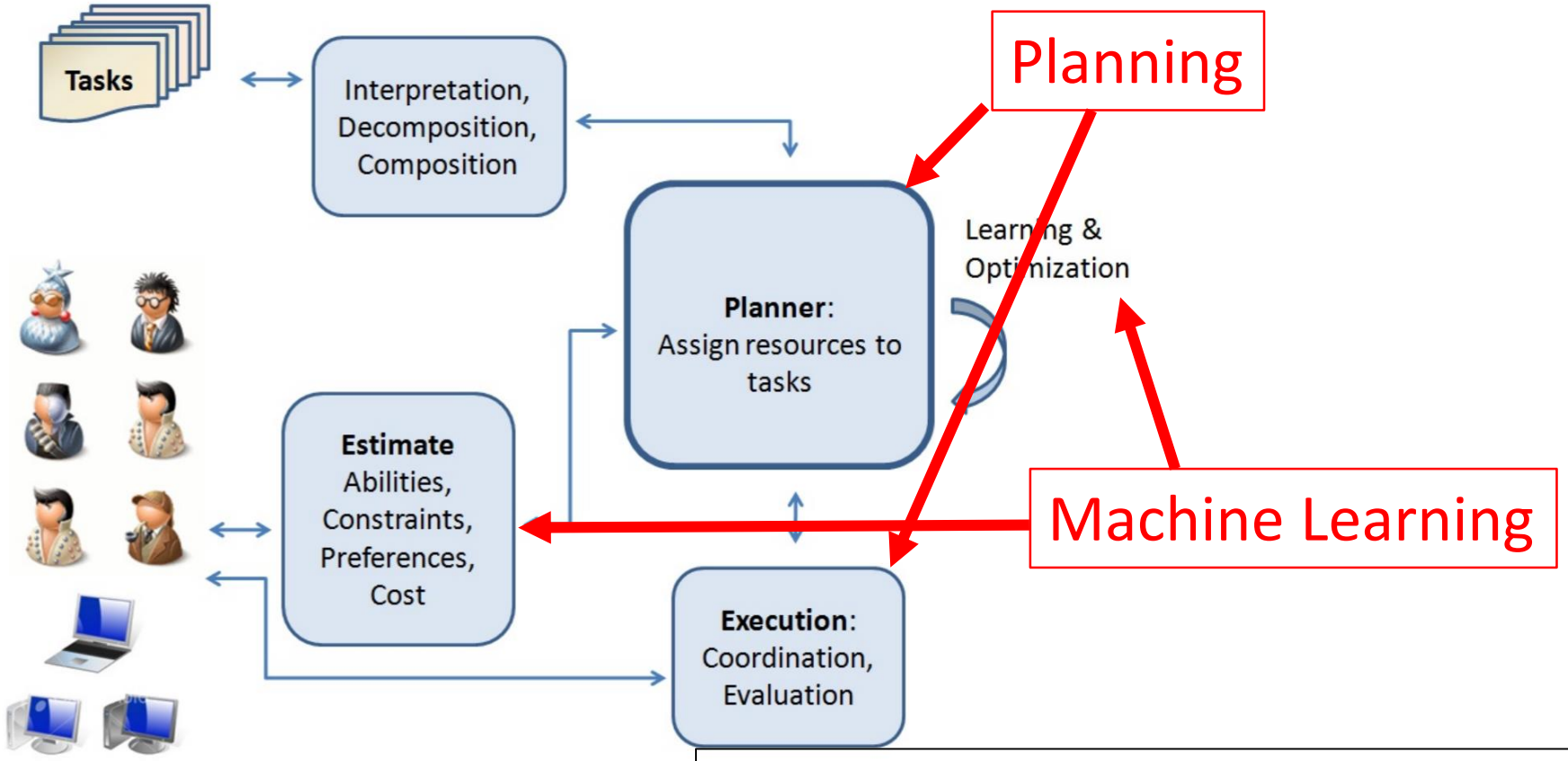
Claim 3.2. *A greedy strategy for selecting agents is a constant-factor approximation to the coalition problem (k is constant).*

Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.

Claim 3.2. *A greedy strategy for selecting agents is a constant-factor approximation to the coalition problem (k is constant).*

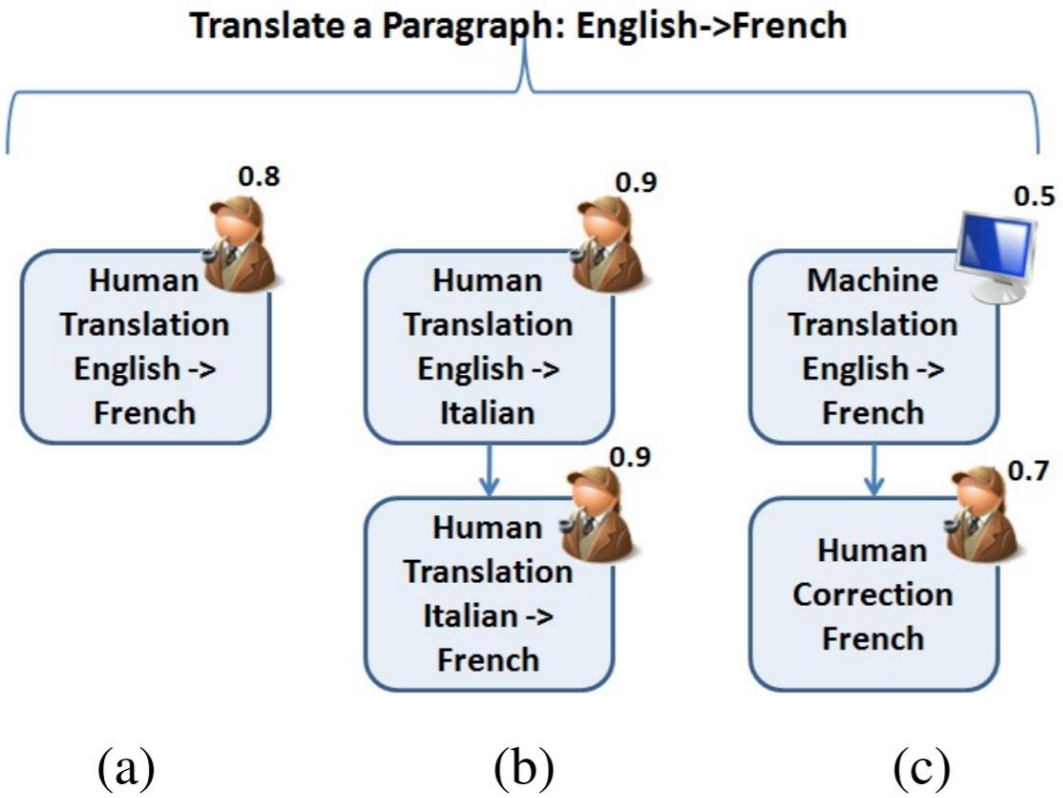
The number of possible coalitions considered in GTM plan generation is exponential in n. Thus, a natural way to reduce the search space is to restrict the maximal size of a coalition to k, thus reducing the number of coalitions to $O(n^k)$. Such a restriction is reasonable for language translation as most tasks do not require more than a few participants.

Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.



Using AI and Machine Learning for Human/Generalized Computation

Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.



Using Human Computation for AI

Shahaf, D. and Horvitz, E., 2010
Generalized Task Markets for Human and Machine Computation.
In *Proceedings National Conference on Artificial Intelligence*.

Question: Where are the markets?