

# Question Answering

Material adapted from Pedro Rodriguez

University of Maryland

Why we do question answering: Cranfield vs. Manchester

# Cranfield



# Ellen Vorhees



in the Cranfield paradigm, researchers perform experiments on test collections to compare the relative effectiveness of different retrieval approaches



Google AI



# Pathologies of QA

## SQuAD

---

Context      In 1899, John Jacob Astor IV invested \$100,000 for Tesla to further develop and produce a new lighting system. Instead, Tesla used the money to fund his **Colorado Springs experiments.**

Original      What did Tesla spend Astor's money on ?

Reduced      did

Confidence    0.78 → 0.91

---

Shi Feng, Eric Wallace, Alvin Grissom II, Pedro Rodriguez, Mohit Iyyer, and **Jordan Boyd-Graber**. **Pathologies of Neural Models Make Interpretation Difficult**. Empirical Methods in Natural Language Processing, 2018.

# Riddle of the Sphynx



(from Dominic Winter)

## From PDQ Bach:

What starts out on four legs then  
goes round on two  
Then finishes on three before it's  
through

# Riddle of the Sphynx



(from Dominic Winter)

## From PDQ Bach:

What starts out on four legs then  
goes round on two  
Then finishes on three before it's  
through

Well a baby then a man, that's plain,  
then finally an old geezer on a cane

# Riddle of the Sphynx

**From Dana Burgess:**

Nobody told Oedipus who he was; he figured that out for himself... You can do a Google search to find out the capital of Arkansas but you can't do a Google search to find out who you are, what you're good at, what makes you happy, what matters for your life.



(from Britannica)

# Riddle of the Sphynx

## From Dana Burgess:

When folks say that a college professor is like a sphinx, they don't usually mean that in a nice way. It suggests that we know the answers students need, but we keep them secret for the sake of our own power... That's part of the riddle of [The Sphynx's] hybrid nature: the curse of being forced to solve a riddle is also the gift of the ability to solve riddles.



(from Britannica)

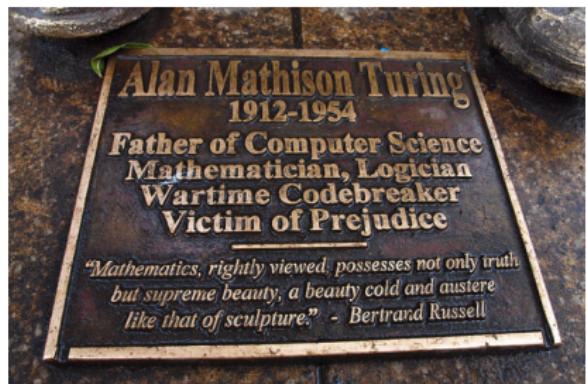


(from BBC)

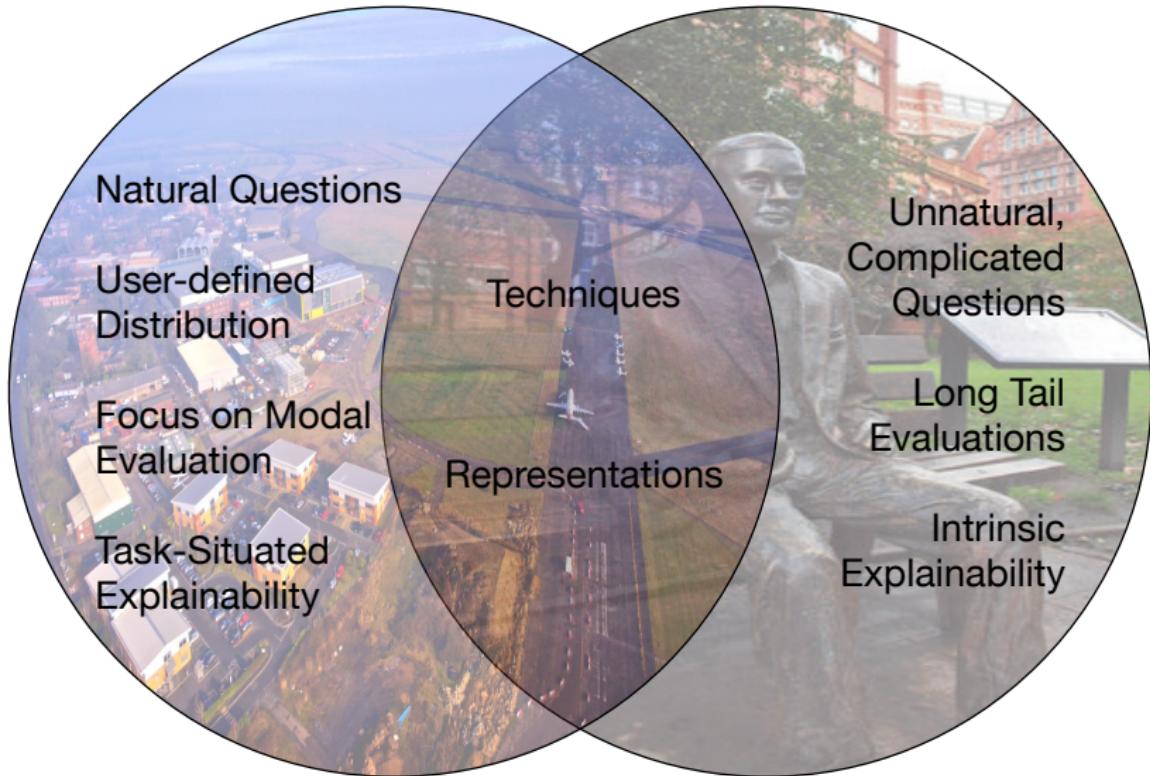


From IBM

# The Turing Test



From Flickr







## Ignore the Distinction at Your Peril

- Fun and games don't earn money
- Crowdworkers won't craft perfect questions
- Cranfield isn't as good at deciding who's smarter

# Ignore the Distinction at Your Peril

- Fun and games don't earn money
- Crowdworkers won't craft perfect questions
- Cranfield isn't as good at deciding who's smarter

## AI Outperforms Humans in Question Answering

Review of three winning SQuAD systems



## A Call to Action

- Cranfield should simulate real users
- Leave Cranfield to companies
- Funding agencies should favor Manchester
- Manchester can be fun!