

Generative AI and the Natural Language Interface for Data

seek.ai





Agenda

1. Brief introduction
2. Introducing Natural Languages Interfaces for Data
3. How will organizations benefit?
4. Challenges
5. Conclusion

Some background on Seek

Founded and led by data professionals and AI researchers...



Sarah Nagy
CEO



Sarah Smith
CRO



Raz Besaleli
NLP Research Scientist



Kevin Stone
Head of Product



...to provide a natural language interface for data.

Some background on Seek



Born and raised in NYC

\$7.5m raised from **Battery** CONVICTION

Customers range from startups to the Fortune 100

World-class AI talent from



Berkeley
UNIVERSITY OF CALIFORNIA



Stanford
University



COLUMBIA
UNIVERSITY

CAUSAL
AI LAB



NYU

TANDON SCHOOL
OF ENGINEERING



PRINCETON
UNIVERSITY



MONTCLAIR
NLP LAB

Why ChatGPT Struck a Nerve



Latest version of ChatGPT aces bar exam with score nearing 90th percentile



Rowan Cheung 
@rowancheung · [Follow](#)



I just watched GPT-4 turn a hand-drawn sketch into a functional website.

This is insane.



Matthew Yglesias 
@mattyglesias · [Follow](#)



English majors get the last laugh as GPT-4 crushes every exam except AP English Language and AP English Lit

**ChatGPT Just Passed
Three Levels of the
Master Sommelier Exam**

Why Natural Language Processing?

Slides from Stanford CS224n, circa 2019

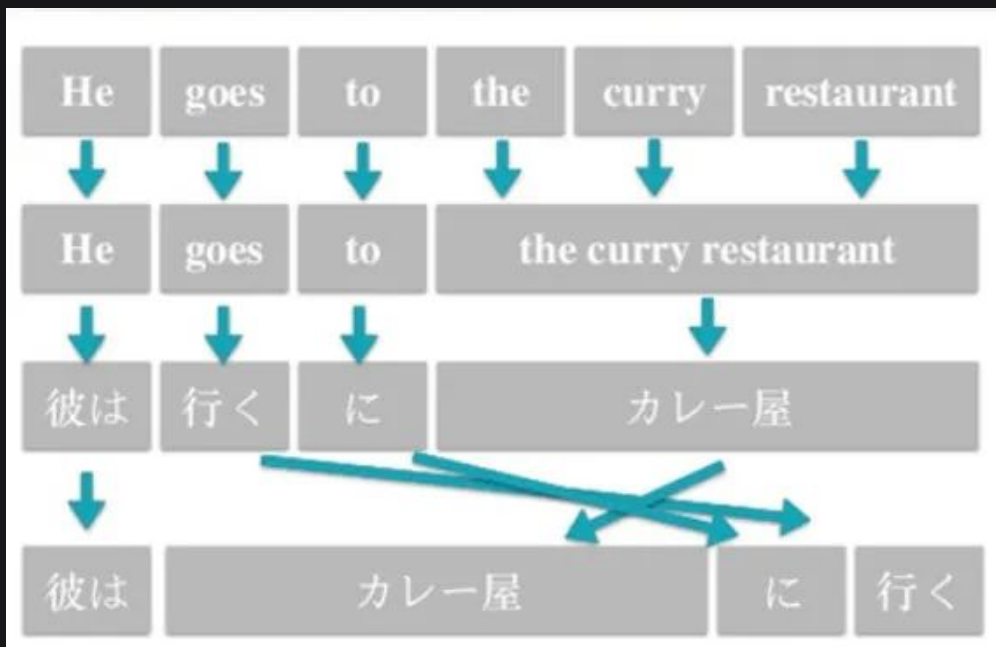
Knowledge-based Pronominal Coreference

- She poured water from the pitcher into the cup until it was full.
- She poured water from the pitcher into the cup until it was empty.
- The city council refused the women a permit because they feared violence.
- The city council refused the women a permit because they advocated violence.
 - Winograd (1972)
- These are called **Winograd Schema**
 - Recently proposed as an alternative to the Turing test
 - See: Hector J. Levesque "On our best behaviour" IJCAI 2013
<http://www.cs.toronto.edu/~hector/Papers/ijcai-13-paper.pdf>
 - <http://commonsensecognition.org/winograd.html>
 - If you've fully solved coreference, arguably you've solved AI !!!



Before Deep Learning

Example: Google Translate (Statistical Machine Translation)



Deep Learning's Impact



Example: Google switches to Neural Machine Translation (2016)

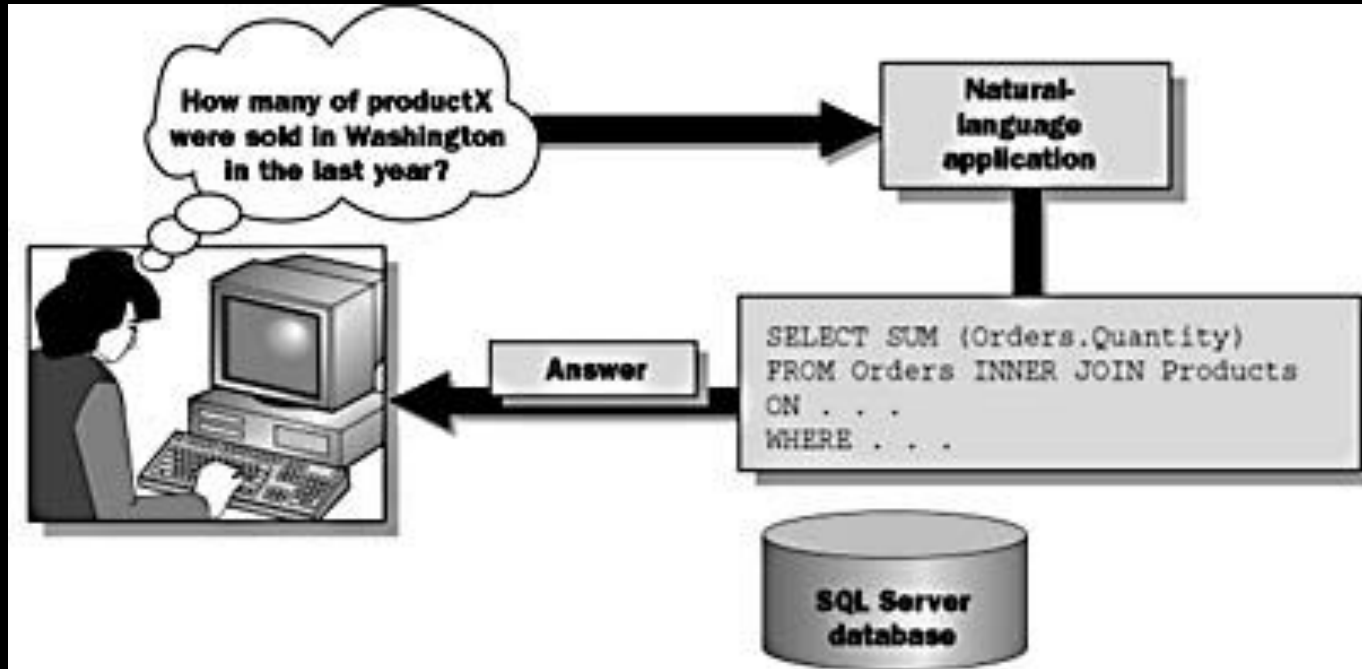
A Neural Network for Machine Translation, at Production Scale

TUESDAY, SEPTEMBER 27, 2016

Posted by Quoc V. Le & Mike Schuster, Research Scientists, Google Brain Team

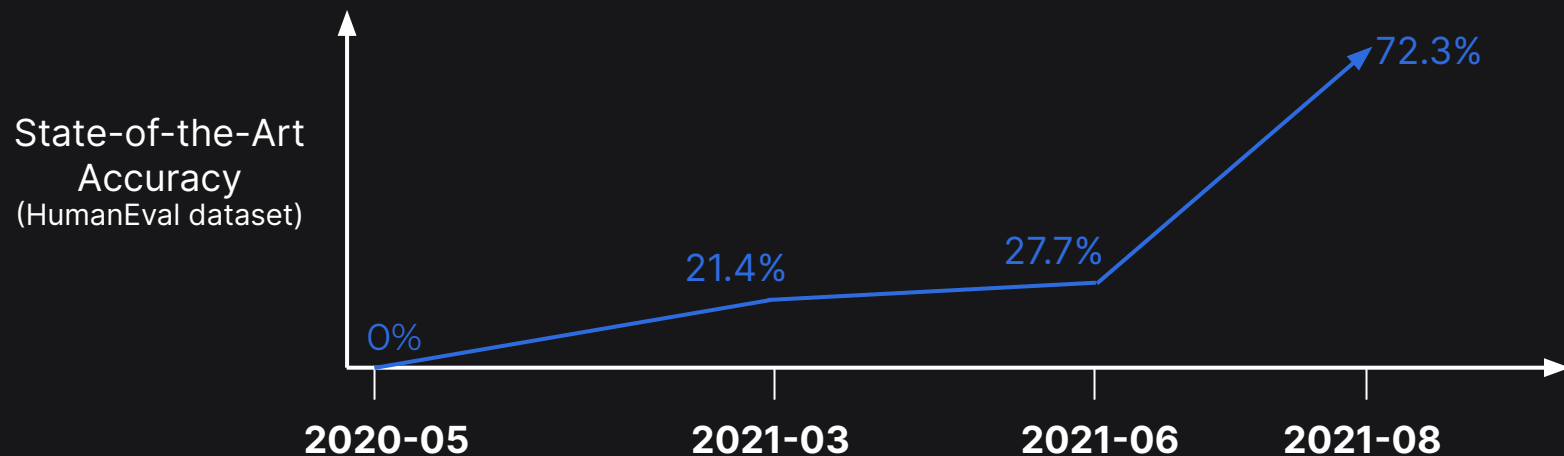
The Natural Language Interface for Data

Many iterations over the years (below is from SQL Server 1996)



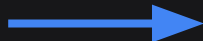
Improvements in AI Code Generation

Code generation accuracy went from 0% to 72% between 2020 and 2021 alone



Data's Leaky ROI Problem

“Hey data team, sorry to bug you again, but can you pull these stats on customers we want to prevent from churning?”

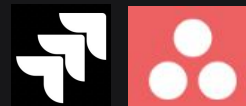


Manual Workflow

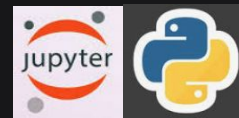
Pulling data



Opening ticket



Generating results



One month later...



“Thanks, but most of these customers already churned.”

The Seek workflow™



Blake (Customer Success Manager) 10:31 AM

@Seek



Seek APP 10:31 AM

What do you want to work on?

Ask a question

Go to app

Question

Of accounts with past due invoices, show me their last activity date ▾

ask as a new question

The Seek workflow™



...powered by
generative AI for
data querying

```
0 select distinct domain from (  
1     select property_closedate, split_part(property_email,  
2     '@', 2) as domain, property_email, * from  
3     fivetran_database.hubspot.contact  
4     where property_closedate is not null  
5 )
```

Q Search Question Status

domain
customer1.com
customer2.io
customer3.com
customer4.org
customer5.com
customer6.io

Case Study: unlocking ROI on data

“Without the right data, retailers of CPG products are missing out on **billions of dollars of GMV** that they are capable of shipping. Seek allows our CPG customers to unlock this GMV.” –CPO, Mid-Stage SaaS Company

With Seek...

...questions are answered **75% faster**

...**thousands of questions** get asked

...**time waiting goes down** from 2 weeks to same-day

What are the limitations of an NLID?



We are here



Answers

Data Q&A

“How many transactions did we have in Montreal?”

→ Data

Knowledge

Problem Q&A with multiple solutions

“Should we build a brick and mortar store in Montreal?”

→ Analysis of pros/cons supported by data

Wisdom

Problem Q&A with one solution

“Should we build a brick and mortar store in Montreal?”

→ Argument of yes or no supported by data

Challenges

Querying databases can involve a lot of complex code

```
1 SELECT MILLTO, MILLNAME, EXPDATE, EXPDATE, EXPDA, DEALNO, QCODE, QUALITY, GWD, gwd, agwd, FWD, DCODE, DESIGN, SUM(JOBMETER) JOBMETER, PART, GISSUE, RCV, CDATE,
2 (SELECT DISTINCT BUYER FROM PROGRAM_MASTER P WHERE P.DEALNO=F.DEALNO) BUYER,
3 (SELECT DISTINCT BUYERCODE FROM PROGRAM_MASTER P WHERE P.DEALNO=F.DEALNO) BUYER_CODE,
4 CASE WHEN GISSUE>0 THEN 'OK' ELSE 'NOT-OK' END
5 GSTATUS,
6 NVL(SUM(JOBMETER),0)-NVL(GISSUE,0) TO_ISSUE,
7 NVL(SUM(JOBMETER),0)-NVL(RCV,0) TO_RCV, accode
8 FROM
9
10 E (SELECT M, MILLTO, (SELECT PNAME FROM test WHERE key='Mill' AND PCODE=M.MILLTO) MILLNAME,
11 (SELECT distinct q, ACCODE FROM test Q WHERE Q.DEALNO=dd DEALNO and roynum=1) accode,
12 (SELECT EXPDATE FROM PROGRAM_MASTER P WHERE P.DEALNO=DD DEALNO) EXPDATE,
13 DD DEALNO, HJGRNO, QCODE, (SELECT QTYPE||' '||STN||' '||FTYPE FROM test WHERE PCODE=QCODE AND KEY='ity') QUALITY,
14 GWD, FWD, DD DCODE, (SELECT PNAME FROM test WHERE KEY='DESIGN' AND PCODE=DD.DCODE) DESIGN, DD METER JOBMETER, PART,
15
16 E (SELECT NVL(SUM(METER),0) FROM PRO_GISS S WHERE
17 S.DEALNO=DD DEALNO
18 AND S.dst_code=M.MILLTO
19 AND S.DSG_CODE=DD DCODE
20 AND S.pgwd=GWD
21 AND S.AC_FWIDTH=FWD
22 AND S.VER=PART
23 AND S.AC_CODE IN (SELECT ACCODE FROM test Q WHERE Q.DEALNO=S.DEALNO) )
24 GISSUE,
25
26 E (SELECT NVL(SUM(d1 METER),0) FROM test M1.test D1 WHERE
27 M1.DOCNO=D1.DOCNO
28 AND M1.FYEAR=D1.FYEAR
29 AND M1.ITYPE=D1.ITYPE
30 AND M1.ITYPE='OSRCV'
31 AND M1.LOC_CODE=M.MILLTO
32 AND D1.DDEALNO=DD DEALNO
33 AND d1.DSG_CODE=DD DCODE
34 AND d1.PGWD=GWD
35 AND d1.AC_FWIDTH=FWD
36 AND d1.VER=PART) RCV,
37
38 E (SELECT distinct max(S.docdate) FROM test S WHERE
39 S.DEALNO=DD DEALNO
40 AND S.dst_code=M.MILLTO
41 AND S.DSG_CODE=DD DCODE
42 AND S.pgwd=GWD
43 AND S.AC_FWIDTH=FWD
44 AND S.AC_CODE IN (SELECT ACCODE FROM test Q WHERE Q.DEALNO=S.DEALNO) ) CDATE
45 FROM test M, test D1.test D1
46 WHERE D.DOCNO=DD.DOCNO
47 AND D.FYEAR=DD.FYEAR
48 AND D.SNO=DD.SNO
49 --AND DD.DEALNO='01158-17'
50 AND M.DOCNO=D.DOCNO
```


Challenges

How do you know you can trust the output?

When did France gift Lithuania Vilnius TV tower?



France gifted Lithuania the Vilnius TV tower in 1980.

What prompted the French government to provide the gift?

The French government provided the gift as a gesture of friendship and solidarity with the Lithuanian people, who had recently declared independence from the Soviet Union.

The workflow of the future

» "Seek"

"Pull up our transaction data from Montreal"

"Last 3 months, and daily."

"Now project 3 months into the future."

"Yes."

"Can you pull up all transactions from our site now?"

"Same chart."



"History and frequency?"

"Apply smoothing?"

"Yes, in the same chart or in a new chart?"

Conclusion

Natural language interfaces for data are higher in quality than ever before, thanks to deep learning and, in particular, generative AI.

While challenges include handling complex questions/data and ensuring results are accurate, as technology advances, NLIDs will be able to make data more accessible than ever before.

Please reach out!
sarah@seek.ai