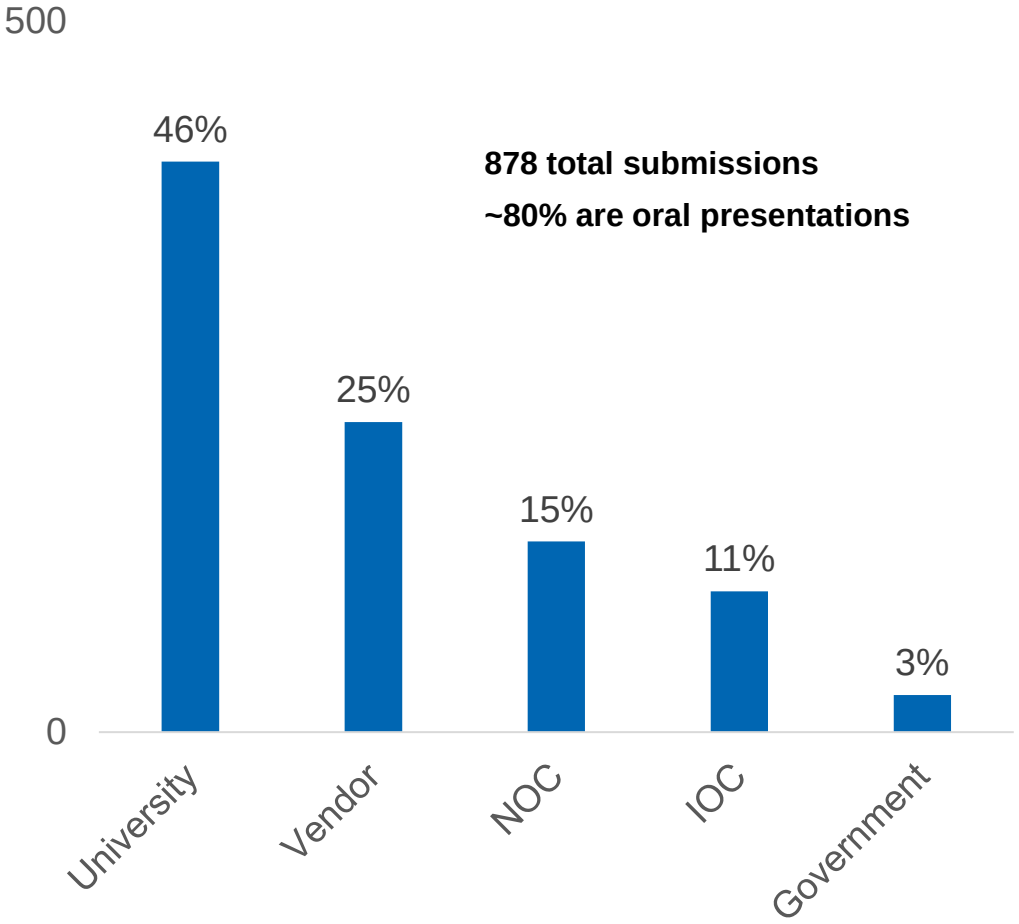
2023 Hackathon “Abstracters”

Steve Braun

Sergey Fomel

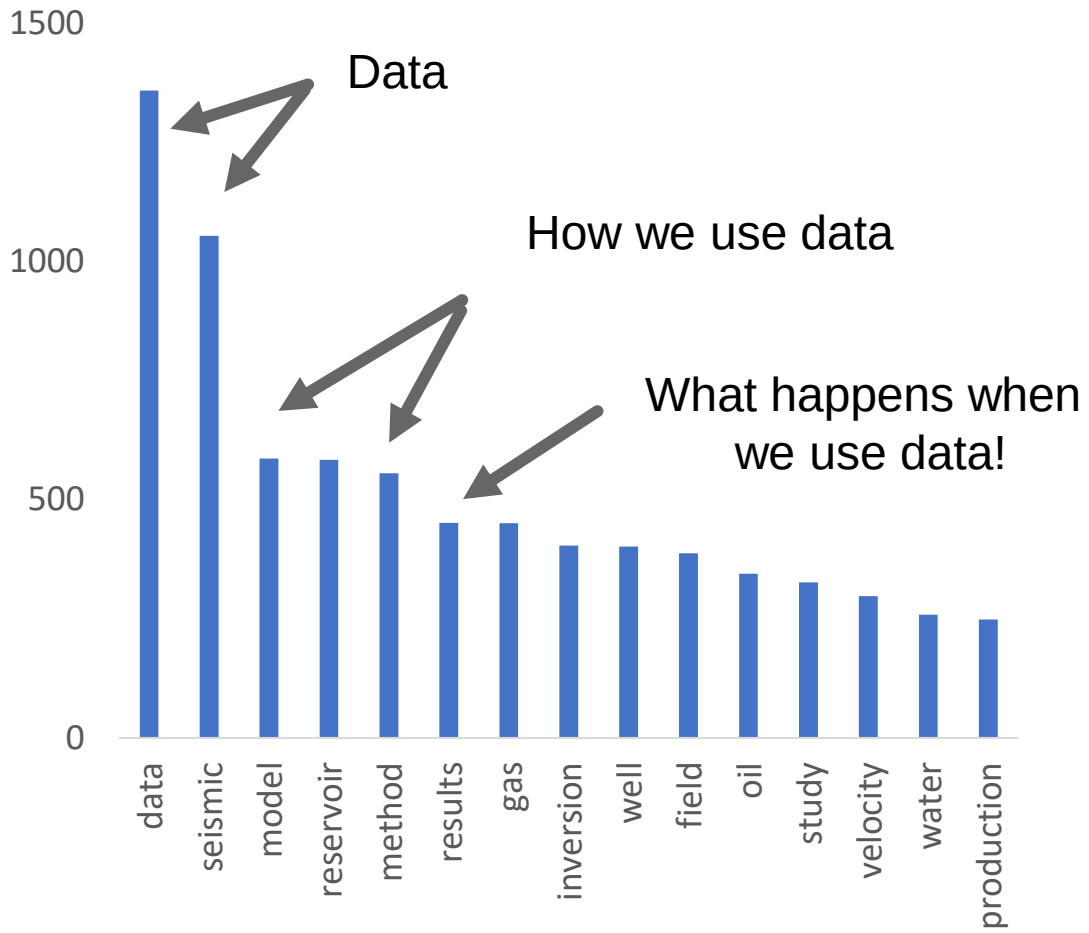
Preliminary data analysis: Who is presenting and about what?

Abstract count by organization type and presentation type*



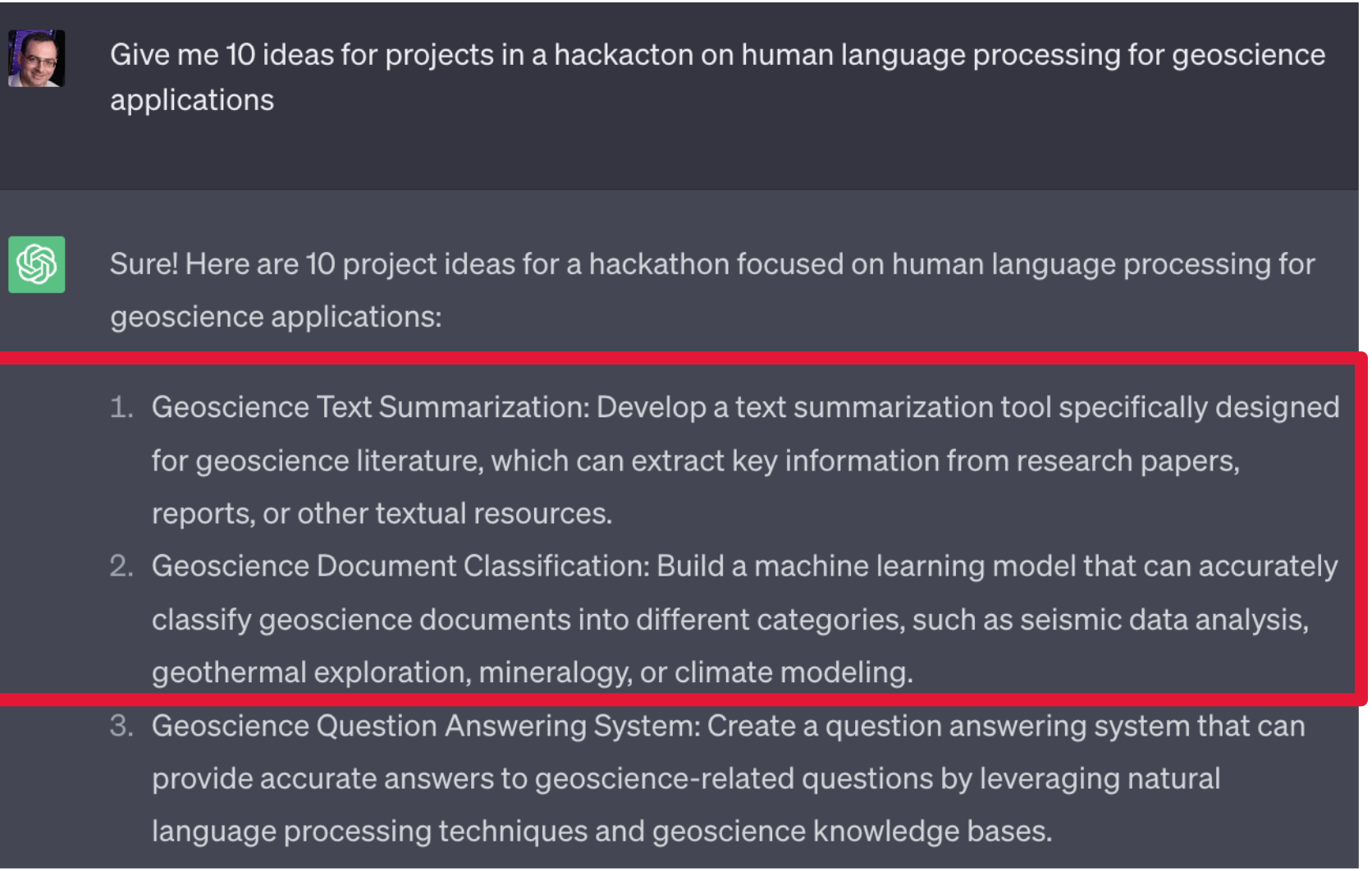
*Organization typed by first author on best effort basis

Top 15 words mentioned in abstract summaries



If ChatGPT is so smart maybe it should do the hackathon too?

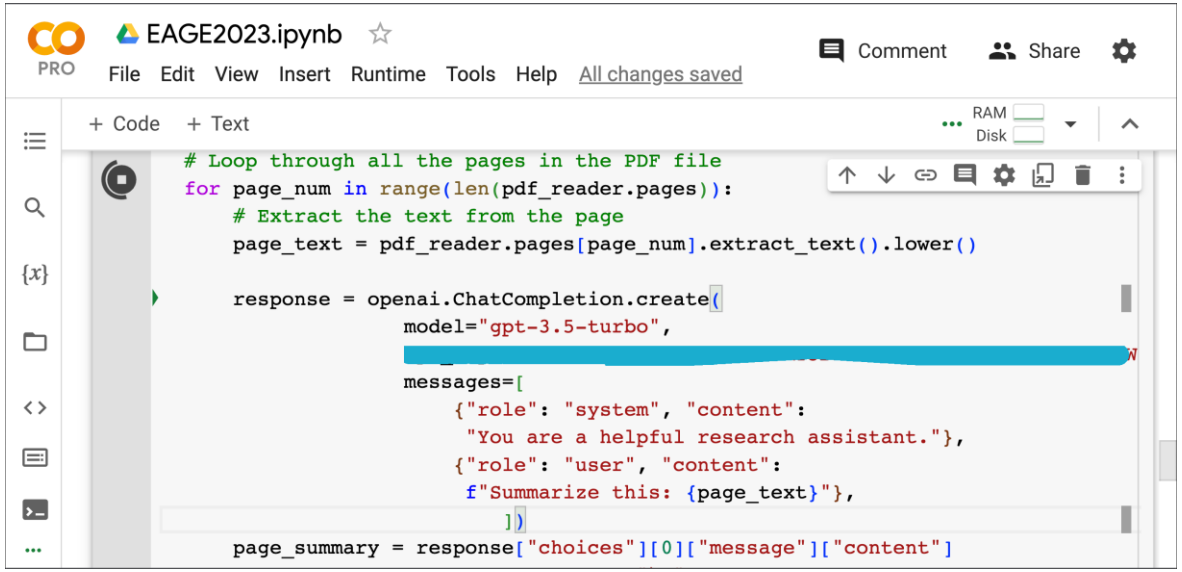
We decided to try these, with differing levels of success



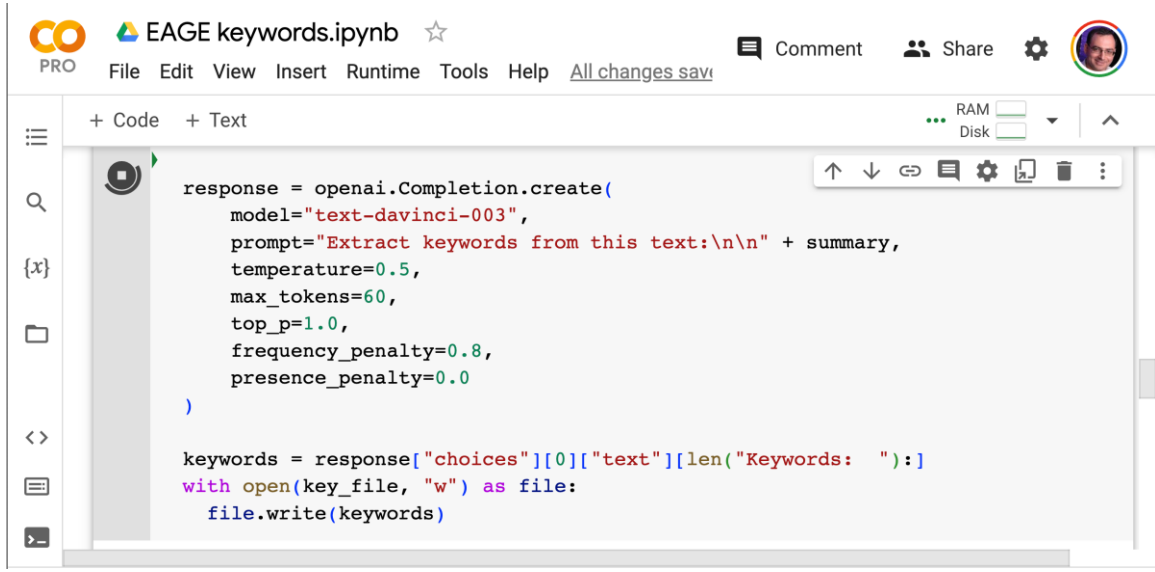
NRCAN, NPD have both released tools doing this in the last year

Followed templates from openAI for working with ChatGPT

Automating GPT3.5 summaries



Automating GPT3.5 key words



The code openAI gave us usually didn't work out of the box. However, light troubleshooting usually fixed the issues

Creating a text summary

What went into GPT3.5 summarization

L. Zhuo, W. Fei, L. Wei, J. Bergeron, G. Hennenfent

2023 EAGE ANNUAL 84TH CONFERENCE & EXHIBITION VIENNA | AUSTRIA

Evaluating 4D DAS VSP Acquisition at the Big Foot Field Gulf of Mexico

Introduction

The Big Foot field is located in the Walker Ridge area of the U.S. Gulf of Mexico, approximately 225 miles (360 km) south of New Orleans, La., in a water depth of approximately 5,200 feet (1,584 m) **(Figure 1a)**. The reservoir is a three-way closure sealed by the salt with the Miocene sand as the main reservoir **(Figure 1b)**. The Big Foot field production started in 2018. Field development plan includes a multi-well drilling program. 4D seismic imaging is considered for enhanced field development.

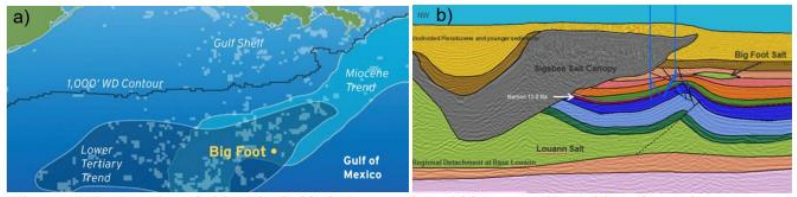


Figure 1 *The Big Foot field, U.S. Gulf of Mexico. a) Field location; b) Field geological section.*

Vertical Seismic Profiling (VSP) acquistion with Distributed Acoustic Sensing (DAS) technology is emerging as an attractive alternative for deep water fields in addressing challenges for 3D seismic imaging and 4D reservoir monitoring due to the lower cost, less interruption to production and the recent industry success (Zhan and Nahm, 2022, Kiyashchenko et al., 2019, Mateeva et al., 2017).

4D DAS VSP is considered as a potential solution for reservoir monitoring at the Big Foot field to further enhance oil recovery. There are critical questions to be answered for making the decision, including: 1) What is the chance of success for 4D DAS VSP at this field? 2) How frequently should the monitor survey be conducted? 3) What is the impact of the field noise to 4D signal? 4) Can we reduce the acquisition time and cost?

We performed a FD (Finite Difference) modeling study in 2022 to assess the feasibility of 4D DAS VSP acquisition at the Big Foot field to improve decisions for field development.

Method

FD modeling is a proven technology that has routinely demonstrated its value for Front-End-Loading (FEL) phase of evaluating seismic acquisition design to reduce cost and improve decision quality

What GPT3.5 came up with as a summary

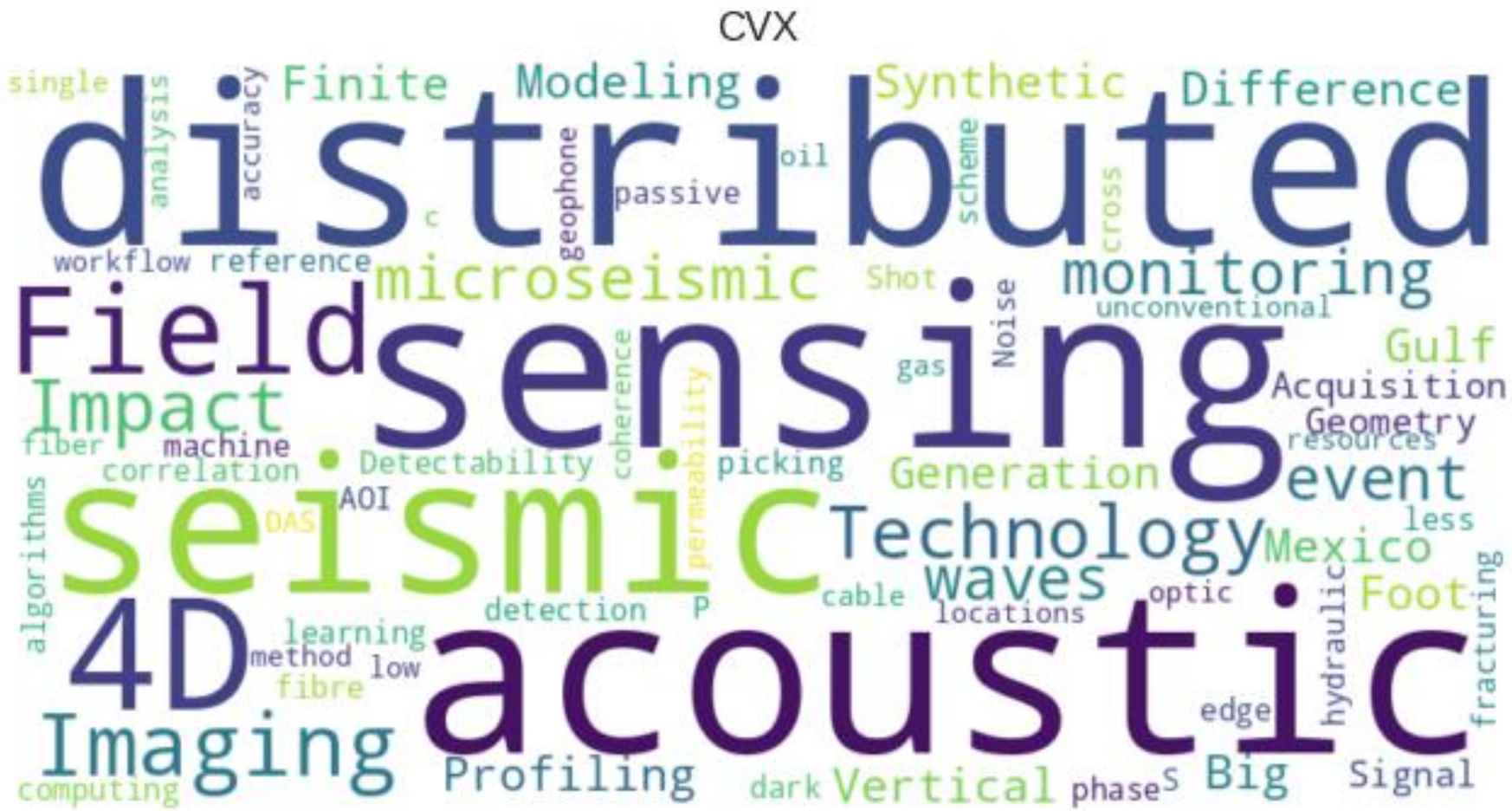
We performed a FD (Finite Difference) modeling study to evaluate the feasibility of DAS VSP acquisition for 4D reservoir monitoring at the Big Foot field, U.S. Gulf of Mexico.

The study demonstrates that: 1) DAS VSP acquisition has high chance of success for detecting 4D signal and effectively monitoring reservoir changes at this field. 2) Shot AOI at one well can be reduced without jeopardizing the 4D objective, which could save time and cost for field acquisition. However, the larger shot AOI is still needed for the other two wells. 3) Field shut-in noise and natural flow noise substantially reduced the S/N of the 3D gather. However, noise sensitivity test shows the shut-in noise is negligible, and natural flow only adds 1-2% of noise to 4D signal.

The learnings from this study significantly contributed to the decisions on planning DAS VSP acquisitions at the Big Foot field.

How did it do? Could be a good way to determine which presentations to go to

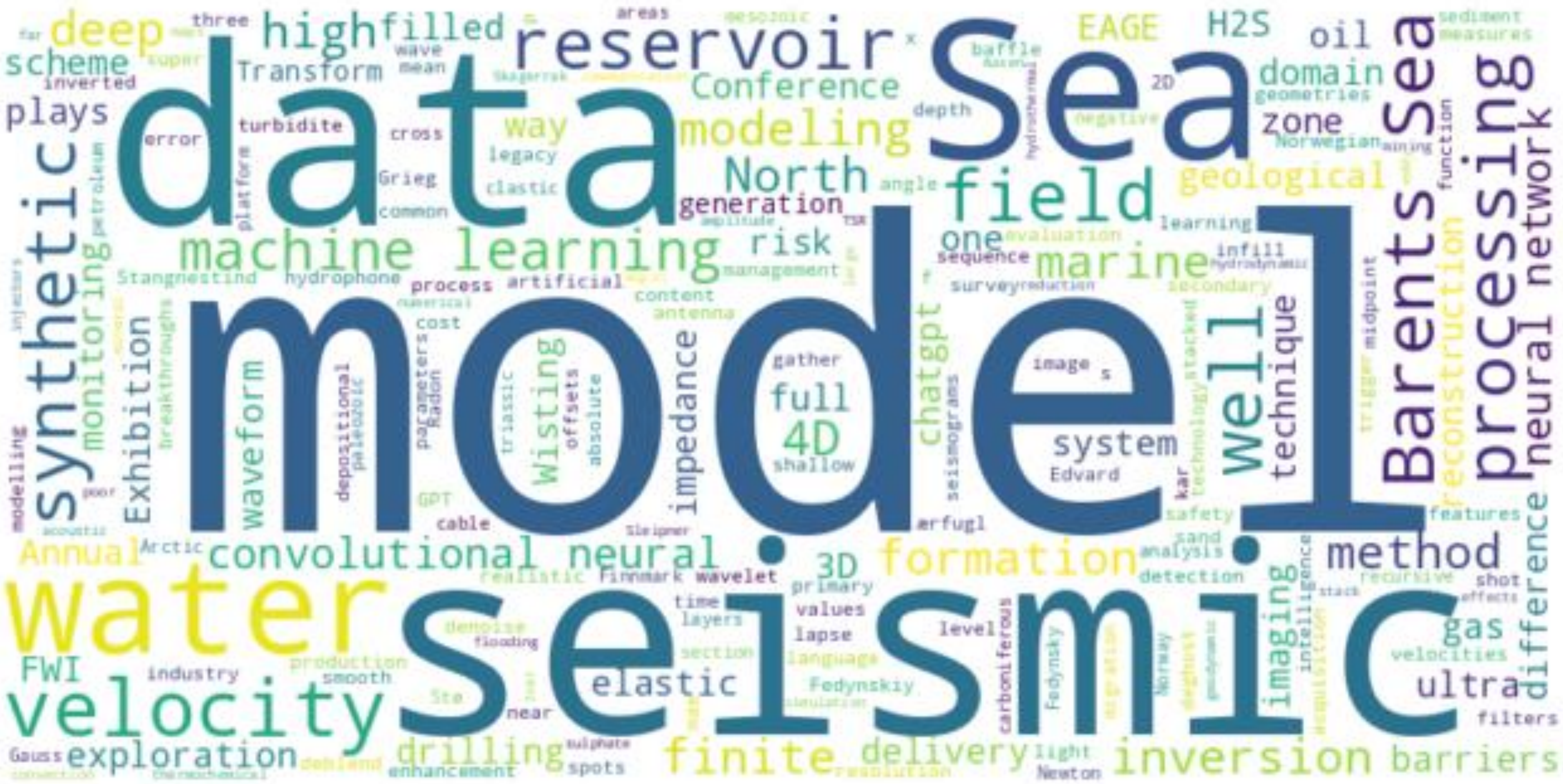
Using key words to understand what IOCs are sharing



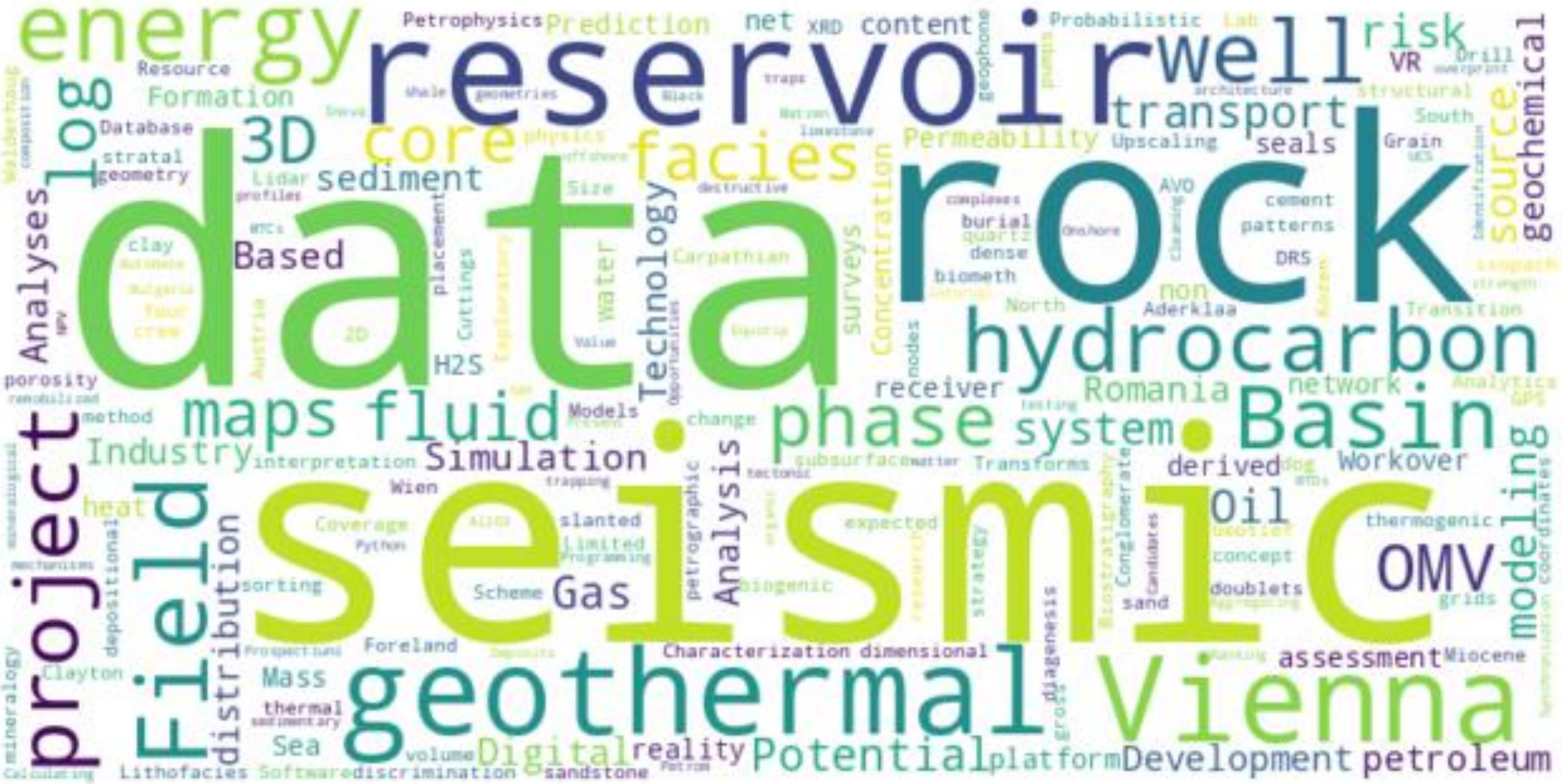
We also tried with document summaries, but non-industry terms made the clouds less useful
We created these word clouds for all companies

Is the next step to use these kinds of systems to *predict* what future research opportunities could be?

AkerBP

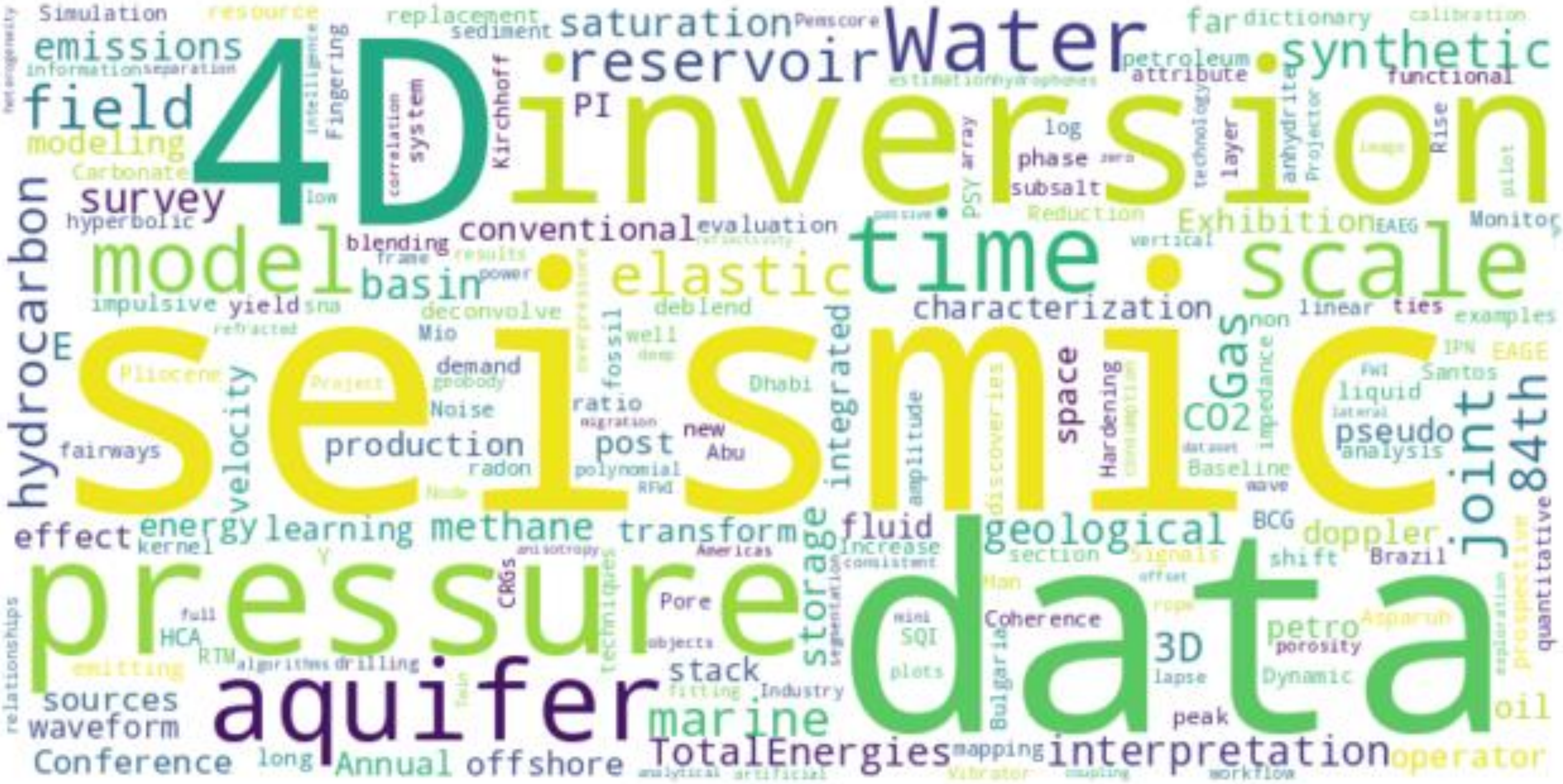

model
seismic
data
Sea
water
velocity
reservoir
Barents
processing
well
field
synthetic
machine learning
convolutional neural
modeling
North
marine
neural network
formation
finite
inversion
deep
high
filled
EAGE
H2S
oil
scheme
Transform
Conference
domain
plays
way
zone
geological
generation
risk
one
monitoring
Exhibition
waveform
Wisting
impedance
full
4D
chatgpt
system
technique
reconstruction
difference
method
3D
elastic
imaging
gas
ultra
FWI
Annual
exploration
drilling
delivery
barriers
turbidite
error
cross
legacy
Grieg
clastic
common
hydrophone
Stangnestind
process
artificial
content
antenna
cost
gather
image
shallow
Edvard
features
safety
sand
analysis
detection
primary
values
lapse
level
layers
time
wavelet
Finnmark
Arctic
industry
smooth
production
near
spots
enhancement
Gauss
filters
shot
velocities
Fedynskiy
section
language
carboniferous
Norway
Newton
light
inverted
three
areas
wave
mean
baffle
depth
angle
learning
infill
survey
sequence
evaluation
management
secondary
function
midpoint
Norwegian
geometries
negative
sediment
measures
2D
cable
GPT
modelling
depositional
parameters
offsets
absolute
triassic
seismograms
stacked
technology
intelligence
acquisition
recursive
Radon
Sleipner
erfugl
realistic
denoise
flooding
sulphate
resolution

OMV



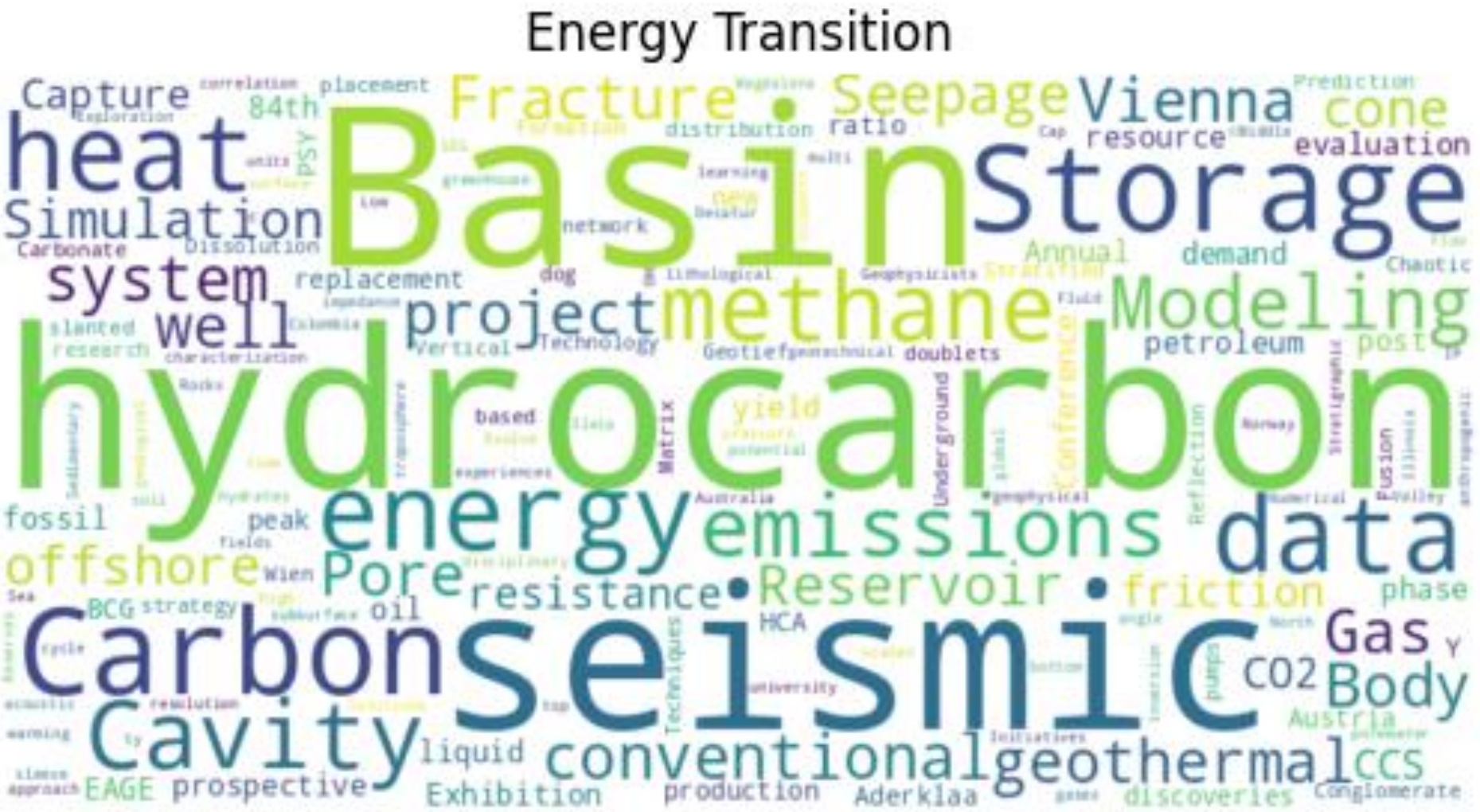
data
rock
seismic
reservoir
well
energy
geothermal
Vienna
Field
project
hydrocarbon
Basin
phase
fluid
maps
facies
core
3D
log
risk
transport
source
geochemical
modeling
OMV
Oil
Gas
system
Simulation
Industry
Analysis
Analyses
Technology
distribution
Potential
Development
petroleum
Digital
reality
platform
Romania
network
receiver
derived
Workover
assessment
sediment
Based
Formation
Prediction
content
net
Petrophysics
Probabilistic
Permeability
Upscaling
seals
Concentration
surveys
H2S
Models
method
porosity
interpretation
Characterization
dimensional
Foreland
Mass
thermal
Sea
Lithofacies
Software
discrimination
sandstone
volume
Scheme
slanted
limited
sorting
Clayton
mineralogy
depositional
heat
Coverage
Wien
change
subsurface
Transforms
expected
biogenic
petrographic
Conglomerate
Biostratigraphy
diagenesis
strategy
sand
concept
doublets
thermogenic
grids
Coordinates
Miocene
Resource
Database
stratal
geometry
clay
Lidar
Size
Carpathian
Cuttings
placement
water
non
North
Aderklaa
DRS
patterns
cement
burial
dense
biometh
AVO
Grain
South
structural
Drill
VR
XRD
geophone
Transition
Analytics
nodes
Austria
2D
Python
Exploratory

TOT



seismic
data
4D
inversion
pressure
Water
aquifer
time
scale
reservoir
model
elastic
synthetic
saturation
emissions
field
hydrocarbon
survey
modeling
Carbonate
Simulation
resource
replacement
sediment
information
separation
hyperbolic
conventional
evaluation
basin
blending
frame
results
power
impulsive
yield
deconvolve
deblend
characterization
Exhibition
Monitor
EAEG
vertical
linear
ties
examples
non
Gas
CO2
joint
84th
pseudo
analysis
liquid
Santos
EAGE
velocity
production
post
ratio
Noise
demand
Mio
well
fossil
Dhabi
Abu
new
integrated
amplitude
discoveries
space
Hardening
impedance
Baseline
wave
fairways
radon
polynomial
migration
Pliocene
Project
geobody
effect
energy
learning
methane
transform
storage
fluid
geological
BCG
doppler
kernel
anisotropy
Americas
Increase
consistent
section
Signals
shift
Brazil
quantitative
prospective
CRGs
techniques
Pore
Coherence
petro
porosity
3D
lapse
Dynamic
full
HCA
RTM
algorithms
drilling
emitting
objects
stack
mini
SQI
plots
Industry
fitting
Bulgaria
peak
oil
relationships
sources
waveform
marine
Conference
long
Annual
offshore
TotalEnergies
mapping
interpretation
operator
analytical
artificial
Vibrator
coupling
workflow
PI
Kirchhoff
system
correlation
Fingering
intelligence
low
log
phase
subsalt
Reduction
array
PSY
layer
anhydrite
Projector
Rise
pilot
technology
petroleum
attribute
functional
far
dictionary
calibration
Pemscore
heterogeneity

Securing a sustainable future together



We also did this for all other session themes: geology, geophysics, petroleum engineering, mining, etc

Which companies are presenting the most similar content?

Cluster analysis enables visual interrogation of company abstracts

EAGE Companies.ipynb

File Edit View Insert Runtime Tools Help

Comment

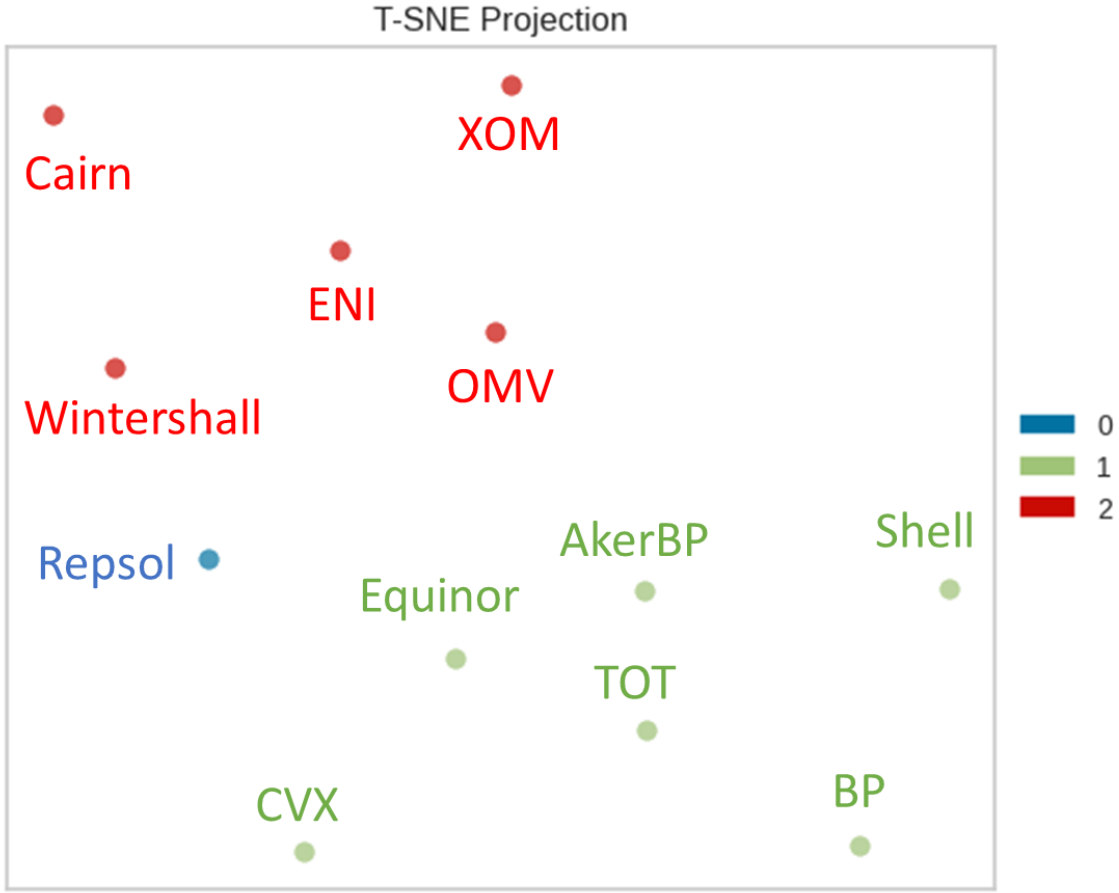
+ Code + Text

```
from sklearn.feature_extraction.text import TfidfVectorizer

objects = []
for company in companies:
  # extract keywords
  keywords = keytext[company].replace(',','')
  objects.append(keywords)

# vectorize using TF-IDF
vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform(objects)
```

✓ 0s completed at 3:20 AM

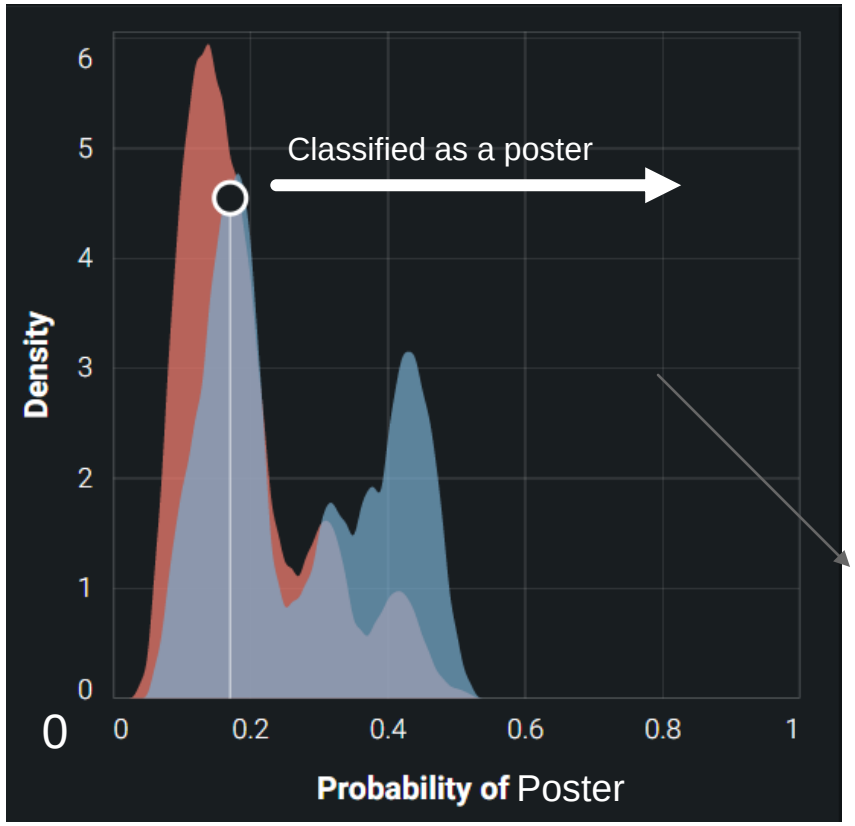


Could become the basis of a recommendation engine or to identify folks publishing anomalous work

Building a classifier to understand who gets a Poster Presentation

Results from holdout sample using a K-nearest neighbors classifier of document summaries

Look at the overlap between classes



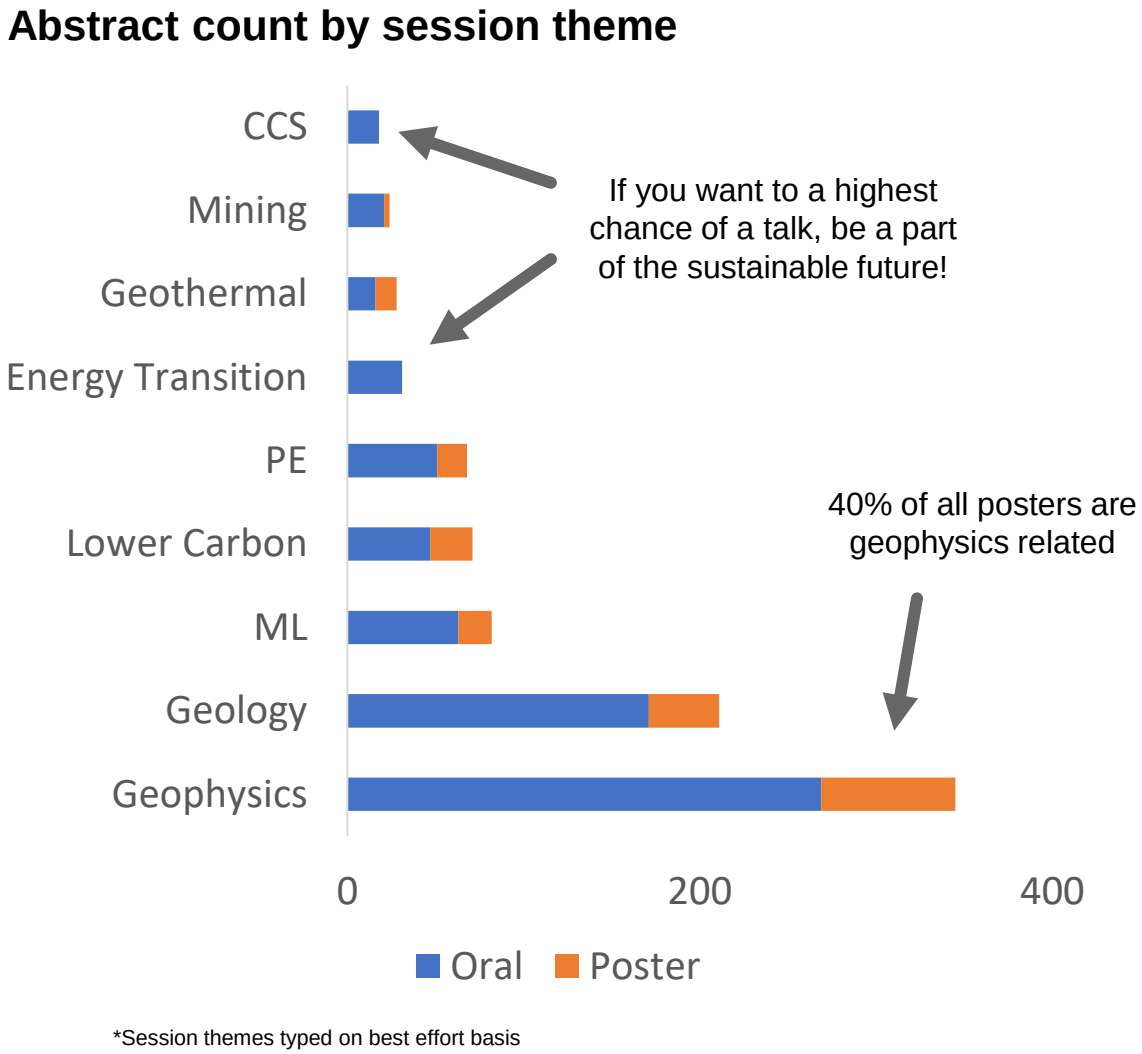
Confusion matrix shows low predictive power of model

Actual \ Predicted	Oral	Poster	
Oral	74 (TN)	64 (FP)	138
Poster	12 (FN)	26 (TP)	38
	86	90	176

While this is clearly not a production worthy model, there are still insights to be gained

Building a classifier to understand who gets a Poster Presentation

- Important to remember **predictions don't tell you reasons**
- For example, we examined the papers with the highest probability of being posters, many of them mentioned geophysics topics
- This could indicate, recalling back to our first data analysis that there were just more geophysics abstracts submitted than spaces for talks



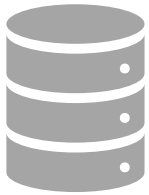
*Session themes typed on best effort basis

Conclusions

Large language models like ChatGPT are effective in working with human texts

Extracting summaries

Extracting keywords



Keywords enable data analysis on EAGE expanded abstracts

Word clouds

Clustering

Ideas for additional projects

Classification using supervised machine learning

- Predict oral vs poster
- Predict organization type