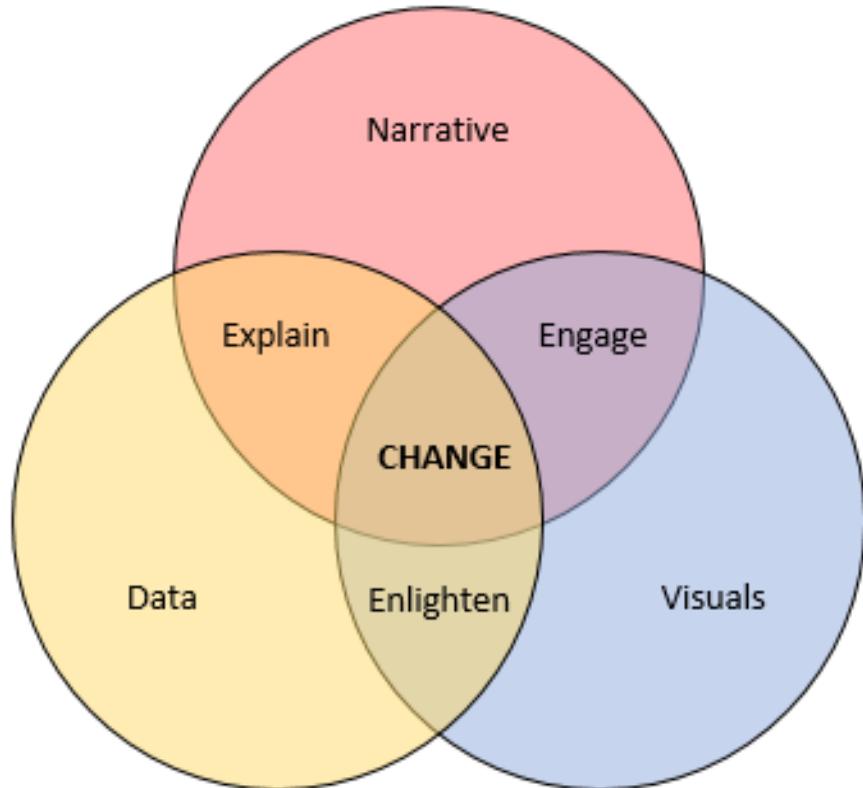




Using Natural Language Processing to  
improve our claims handling process.

March 28, 2019

# Data Science at The General® – Who We Are: The Intersection of Change



## Driving change through:

### People

- Data Science Manager
- Data Scientists (2)
- Machine Learning Engineers (3)

### Technology

- Cloud computing (AWS, Snowflake, Spark)
- Software (DataRobot, Alteryx)
- Open source software (R, Python)
- Bitbucket, Jira

### Process

- Code documentation and versioning
- Project questionnaires and regular check-ins
- Results reporting
- Project prioritization



# Crash course in types of auto insurance coverage

- Liability/Property Damage (LI/PD)
- Personal Injury Protection/Medical Payments (PIP/MP)
- Uninsured Motorist/Underinsured Motorist (UM/UIM)
- Collision (COL)
- Comprehensive (COMP)
- Bodily Injury (BI)



# “Claims” vs “Exposures”

- Claims are the “parents”, Exposures are the “children”
- Separate exposure for each coverage type and each incident
- TG customer files claim after accident with two people injured
  - 1 claim with multiple exposures
  - Presumably exposures opened for collision, possibly property damage
  - Exposure opened for each injured party (2)
    - Reserving is set separately for each exposure



# “Reserves” vs “Paid Losses”

- Reserve is the initial amount we think we will pay on the claim
- “Loss reserving” is handled by our actuarial team, and reserves could take months if not years to resolve
- Once a payment is made on the loss, that amount moves from reserves to paid losses
- When reserves equal zero, a claim is considered closed (but could be open again in the future)
- Once a claim is closed, we consider the amount paid “ultimate paid loss”



# Claim groups at The General®

- Fast Track/Total Loss
- Field Ops
- Casualty
- Large Loss Unit
- PIP/Med Pay
- NARBI



# The NARBI team

- Non-Attorney Represented Bodily Injury
- Specialized team that works to get in front of Attorney Rep'ed claims
- Why is this so important?

	Reserved amt.	No. days to settle	Overhead (\$100/day)	Attorney fees	Settlement amt.	Cost to TG	Payment to insured
With representation	\$ 3,000	30	\$ 3,000	\$ 1,000.00	\$ 3,000	\$ 6,000	\$ 2,000.00
Without representation	\$ 3,000	3	\$ 300	\$ -	\$ 2,500	\$ 2,800	\$ 2,500.00



# Predicting Attorney Involvement

- Predicting when a claim will involve an attorney is relatively easy:
  - Multiple vehicles involved
  - Severe injuries (anything involving a fracture and worse)
- Predicting when a claim will involve an attorney AND we can prevent it is hard:
  - Location Location Location
  - NARBI team won't take on claim if:
    - Coverage issue (driver not on policy, vehicle not on policy)
    - No injury exposure (not worth the team's time)
    - Claimant already repped at time of referral
    - Special Investigations Unit (fraud team) involved



# Progress so far

- Currently have a model in flight that refers claims we think will become Attorney Rep'ed AND we can make an impact on directly to the NARBI team
- Predictive model includes claim attributes and NLP, plus filters to remove claims we know for sure the team will not accept
- NARBI supervisor reviews claims we identify and accepts or rejects them for NARBI team
  - Supervisor uses domain knowledge to make final call
  - Claim Acceptability Rate (CAR)
  - Currently 60-65%; we'd like to see this closer to 75%
  - Low hanging fruit has already been found!
- Current estimated savings of \$750k-\$1.4M annually



# How can you help us?

- Primary question of interest: Who was at fault?
- Secondary questions of interest:
  - What Claim Group should the claim be routed to?
  - Can we predict severity type (minor, moderate, major, life threatening, death)
  - What was the cause of the loss?



# The data!

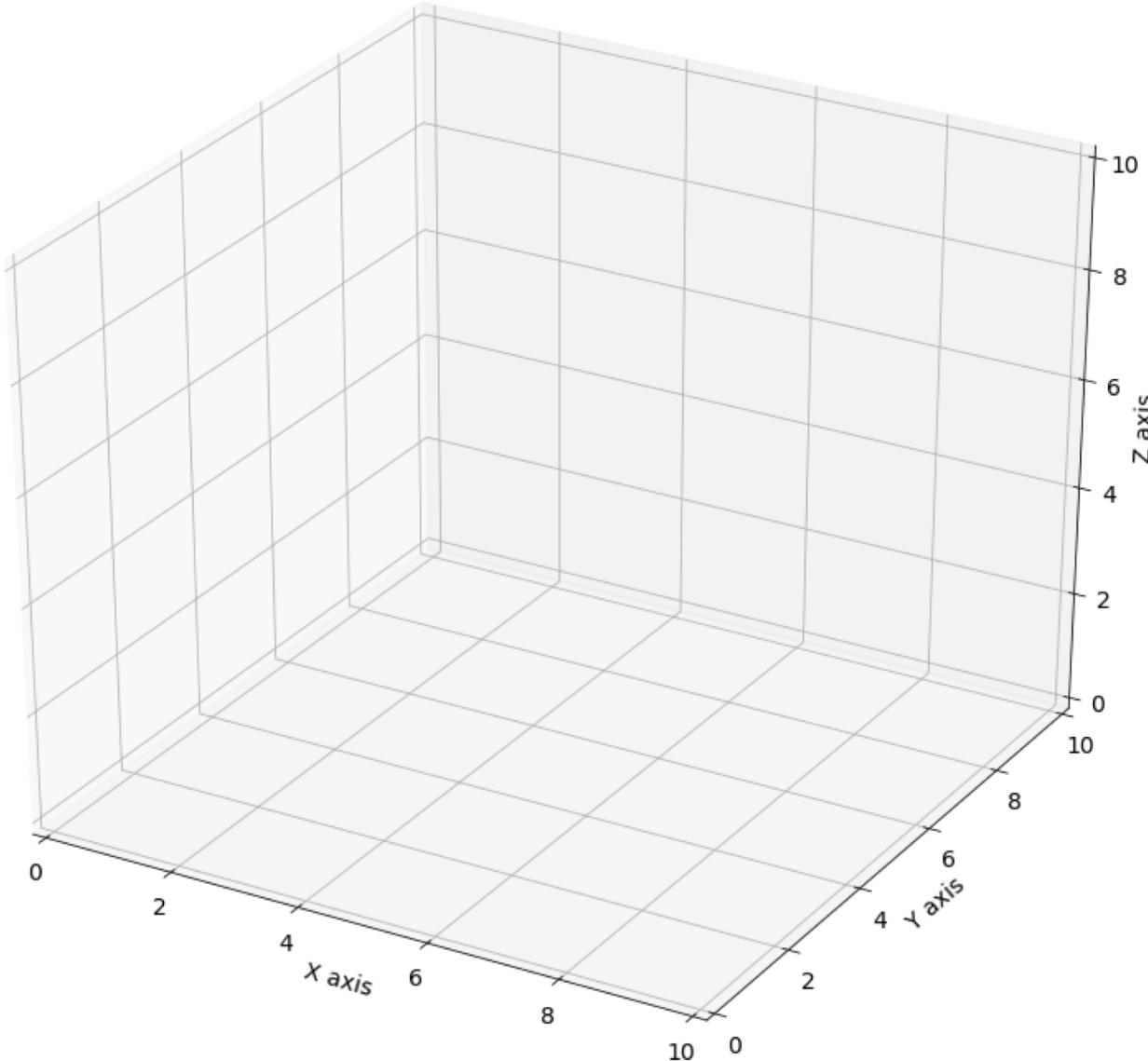
- ~100k BI exposures that have been closed between 1/1/17 and 12/31/18
- Includes an identifier for both Claim and Exposures
- Who was at fault
- Severity Type Name (none, minor, moderate, major, life threatening, death)
- Claim Group (internal claim team)
- Loss Cause Name
- Five text fields
  - Accident description
  - Injury description
  - Damage description
  - Claim Level body
  - Exposure Level body



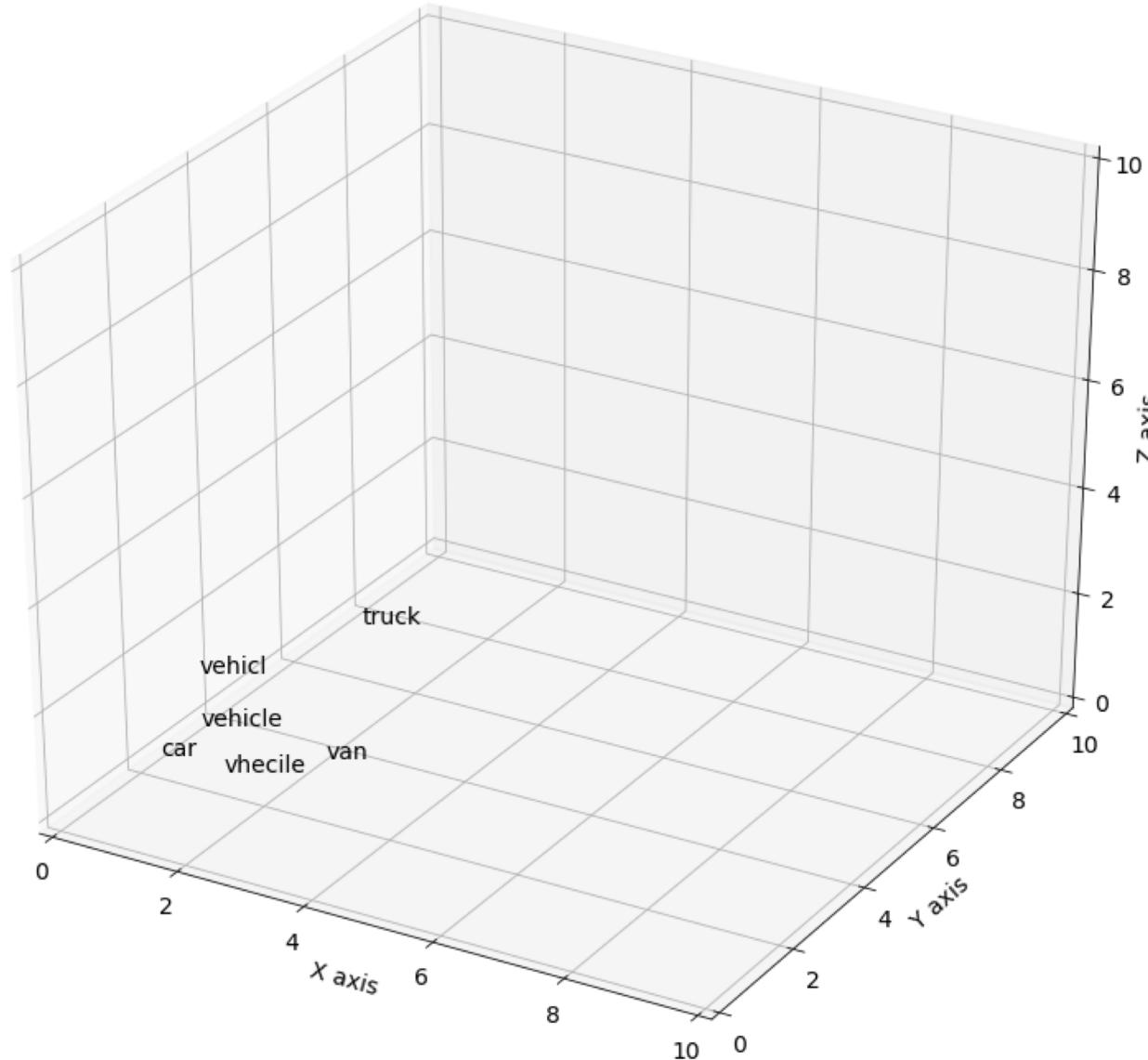
# Natural Language Processing



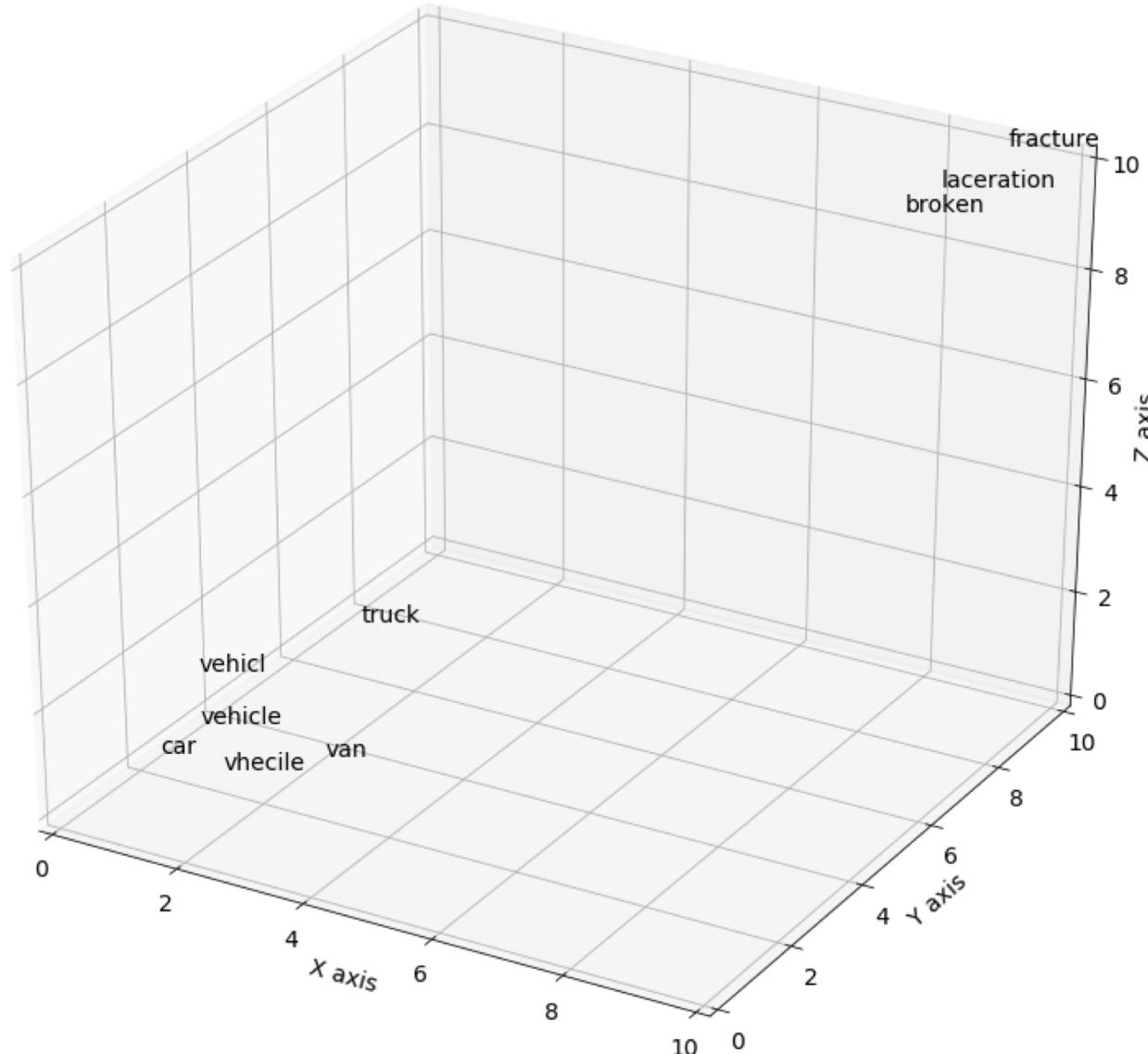
# Word embeddings



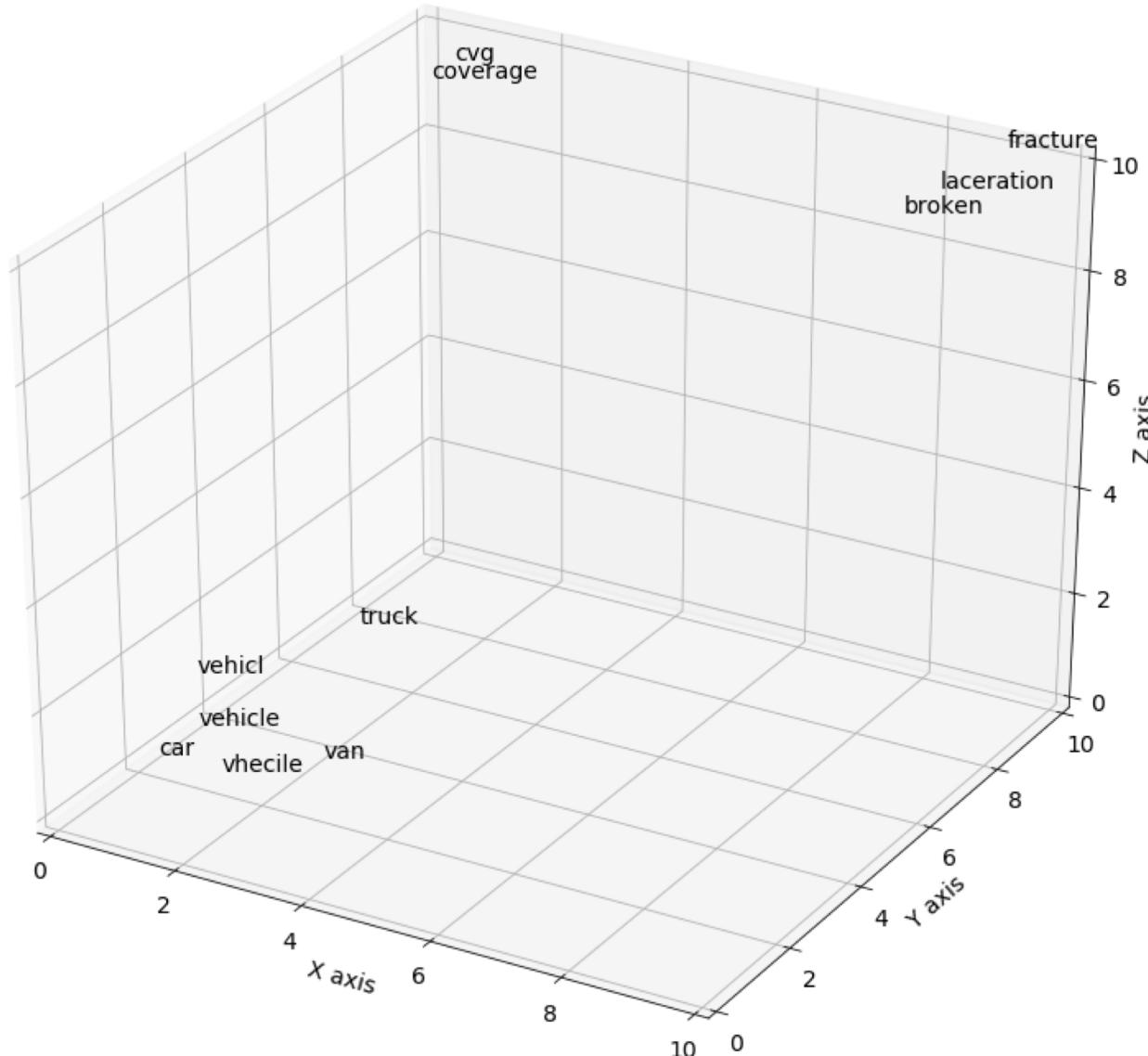
# Word embeddings



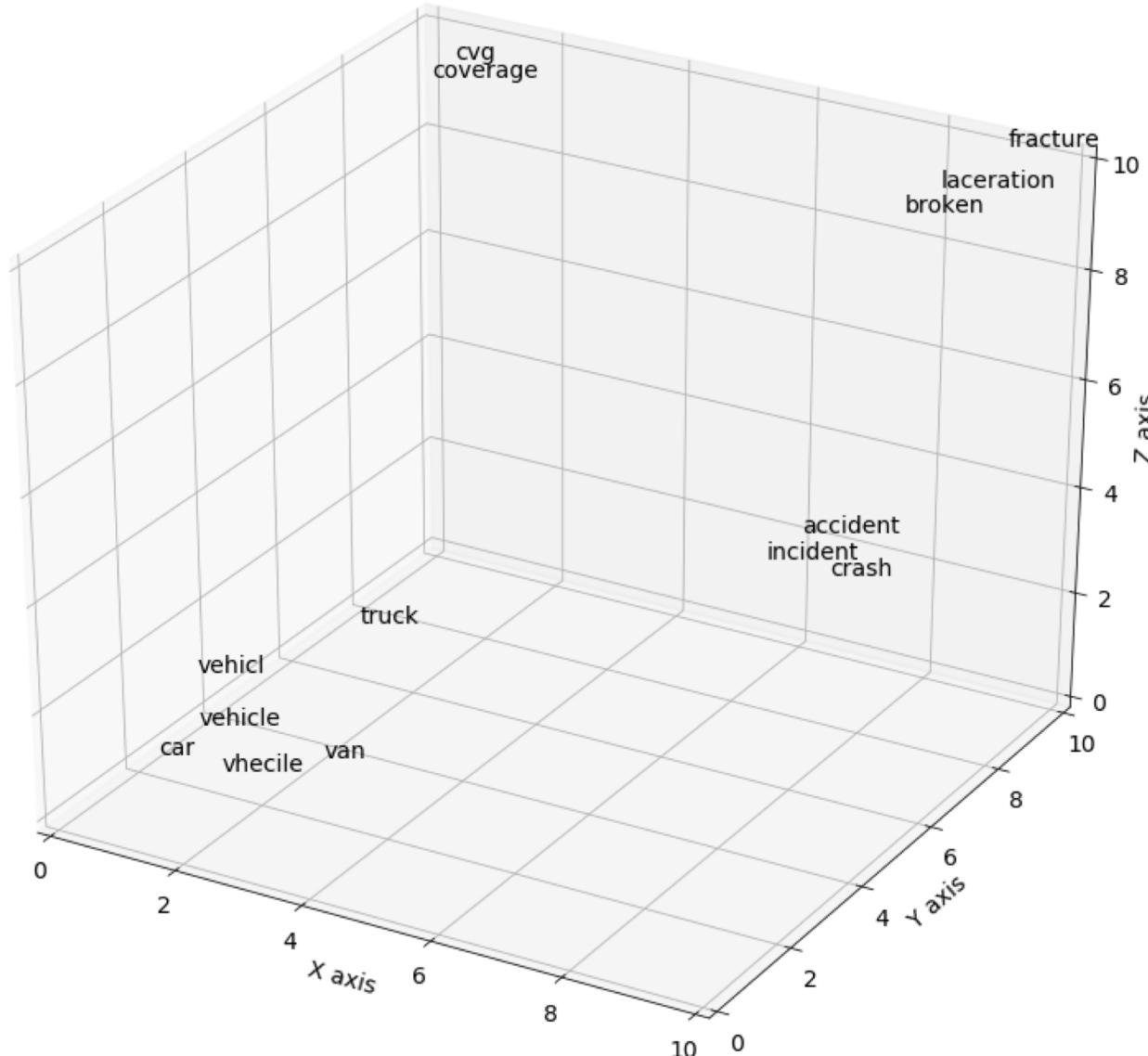
# Word embeddings



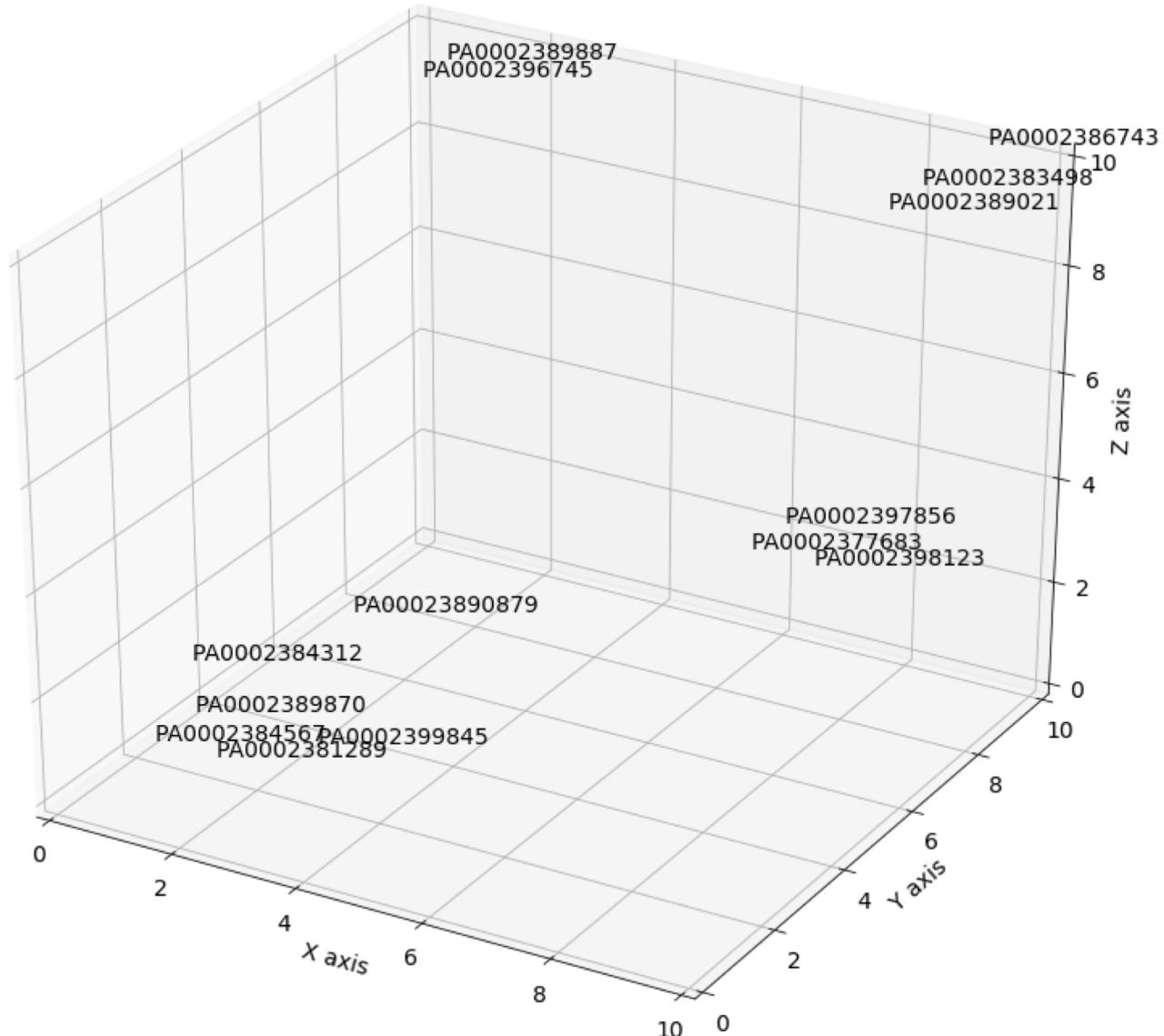
# Word embeddings



# Word embeddings



# Document Embeddings



# Learned Features with Concept Detection

- Attorney Representation
- Multiple vehicles involved
- Multiple people involved
- Fracture and other injuries
- Severity of injuries



# Learned Features with Semantic Role Labeling (RnD)



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?

CV            r/e            IV    in    parking lot



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?

CV

r/e

IV in parking lot



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?

Claimant rear-ended insured in parking lot



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?

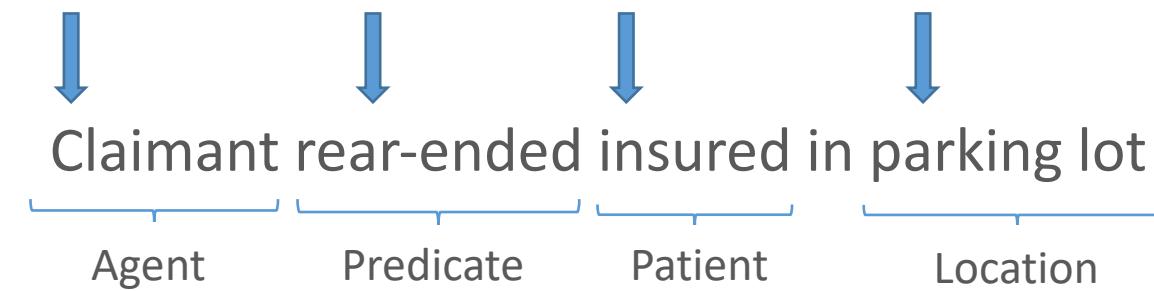


Claimant rear-ended insured in parking lot



# Learned Features with Semantic Role Labeling (RnD)

Who did what to whom at where?



# Connect With Us



[www.autoregressed.com](http://www.autoregressed.com)

 @AutoRegressed



-  @TmthyDobbins
-  tmthyjames.github.io
-  [linkedin.com/in/tim-dobbins](https://linkedin.com/in/tim-dobbins)
-  [tdobbins@thegeneral.com](mailto:tdobbins@thegeneral.com)



-  @thestatsnerd
-  [linkedin.com/in/morganchristopher](https://linkedin.com/in/morganchristopher)
-  [cmorgan@thegeneral.com](mailto:cmorgan@thegeneral.com)

