Introduction to Artificial Intelligence with Python

Language

Natural Language Processing

Natural Language Processing

- automatic summarization
- information extraction
- language identification
- machine translation
- named entity recognition
- speech recognition
- text classification
- word sense disambiguation
- ...

Syntax

"Just before nine o'clock Sherlock Holmes stepped briskly into the room."

"Just before Sherlock Holmes nine o'clock stepped briskly the room."

"I saw the man on the mountain
with a telescope."

Semantics

"Just before nine o'clock Sherlock Holmes stepped briskly into the room."

"Sherlock Holmes stepped briskly into the room just before nine o'clock."

"A few minutes before nine, Sherlock Holmes walked quickly into the room."

"Colorless green ideas sleep furiously."

CALIFORNIA

Big rig carrying fruit crashes on 210 Freeway, creates jam

By JOSEPH SERNA MAY 20, 2013 | 6:35 AM

Monday's morning commute started off horribly for drivers in the San Gabriel Valley when a big rig carrying fruit overturned on the 210, blocking lanes in both directions in Monrovia for most of the morning.

The big rig crashed through the center divider just before 5 a.m. near Myrtle Avenue. Three westbound lanes and two eastbound lanes will be blocked until about 9:15 a.m., according to the California Highway Patrol.

Westbound traffic appeared to be backed up to the 605 freeway.

The trailer is estimated to weigh about 35,000 pounds, according to the CHP.

Natural Language Processing

Syntax

formal grammar

a system of rules for generating sentences in a language

Context-Free Grammar

she saw the city

N	V	D	N
she	saw	the	city

N → she | city | car | Harry | ...

D → the | a | an | ...

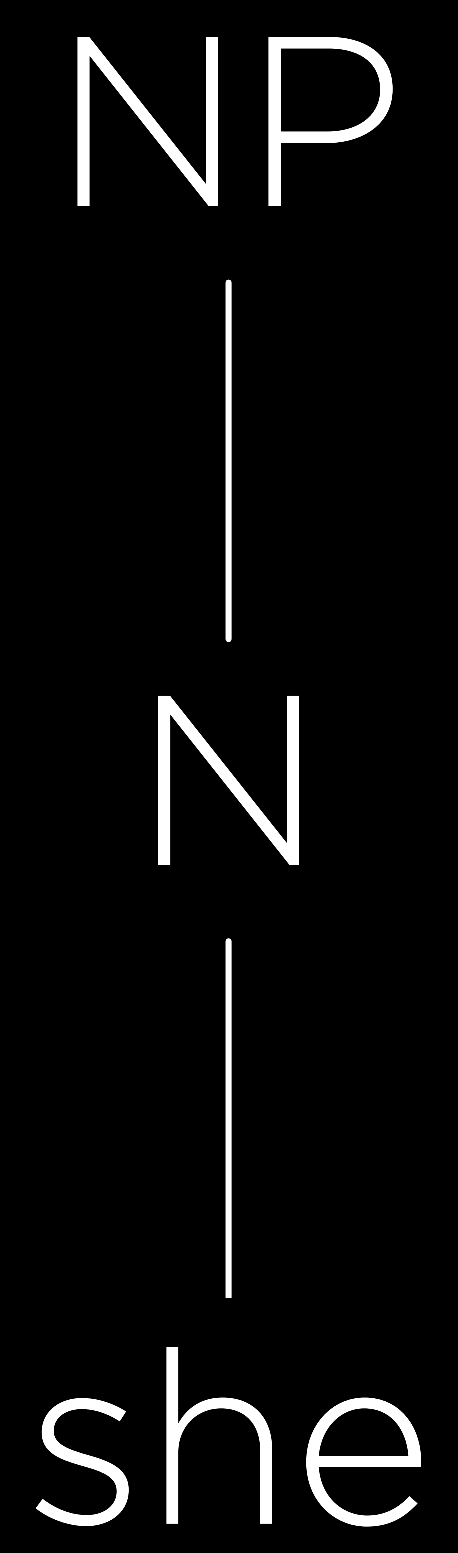
V → saw | ate | walked | ...

P → to | on | over | ...

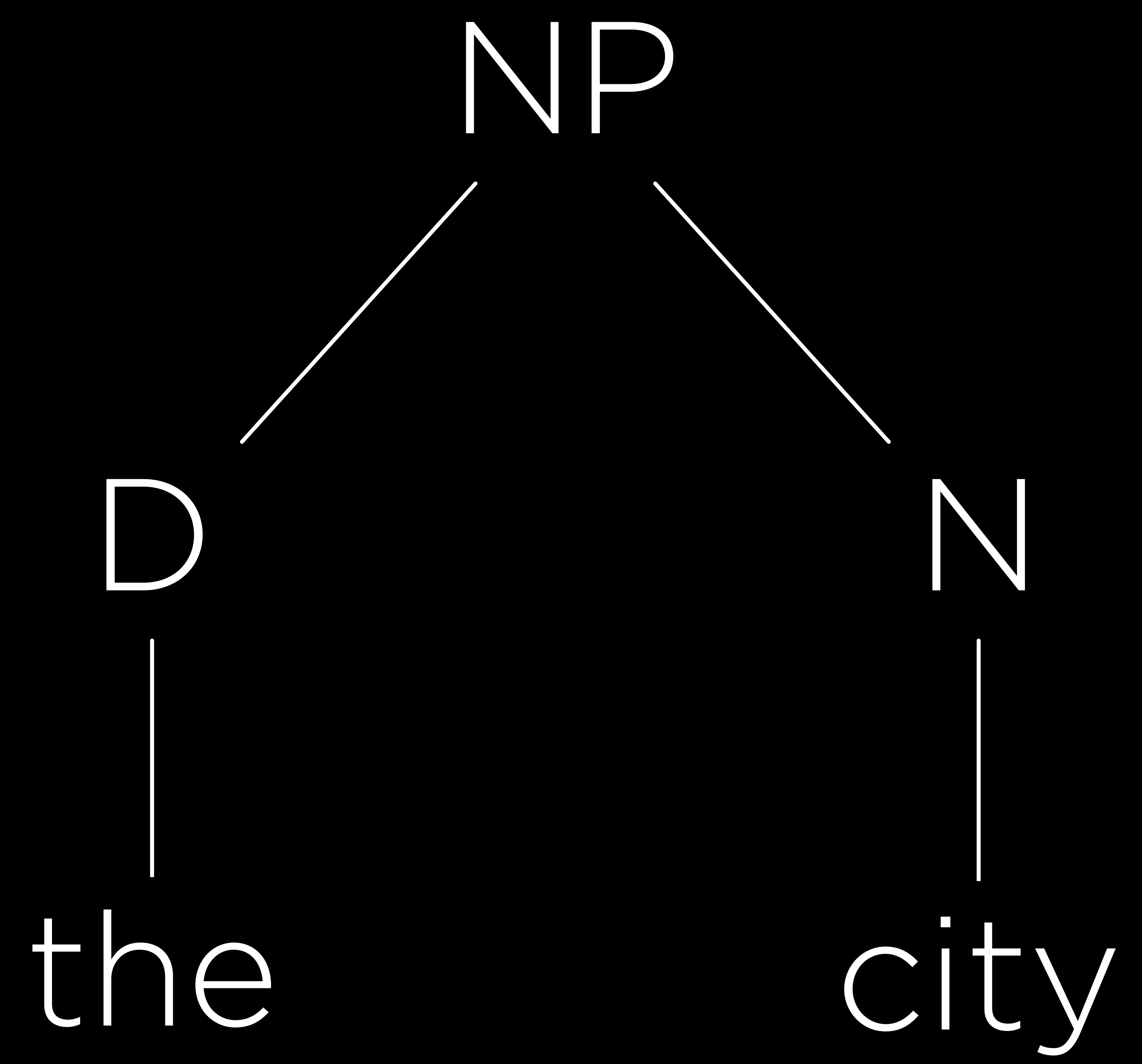
ADJ → blue | busy | old | ...

$$NP \rightarrow N \mid D\ N$$

NP → N | D N

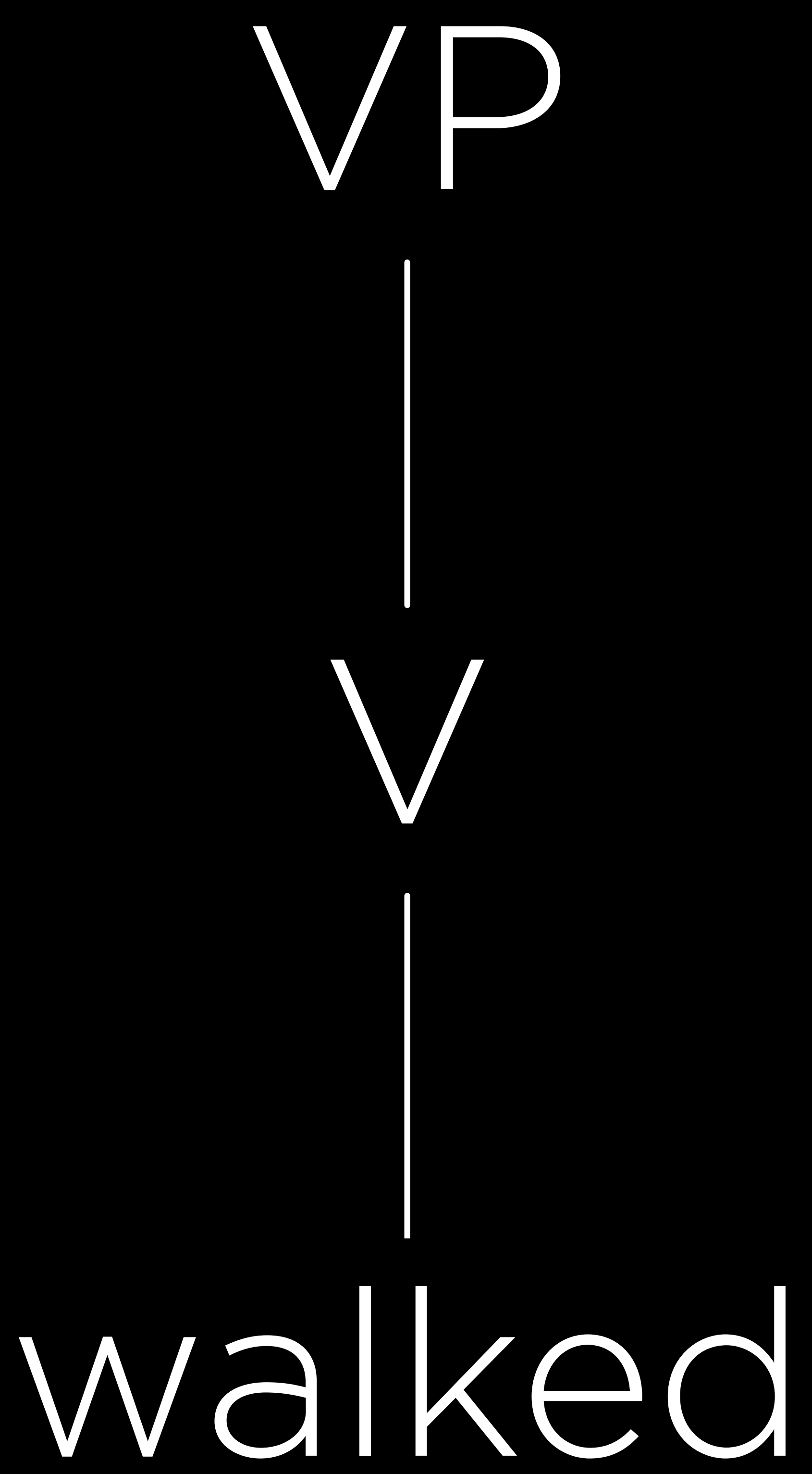


NP → N | D N

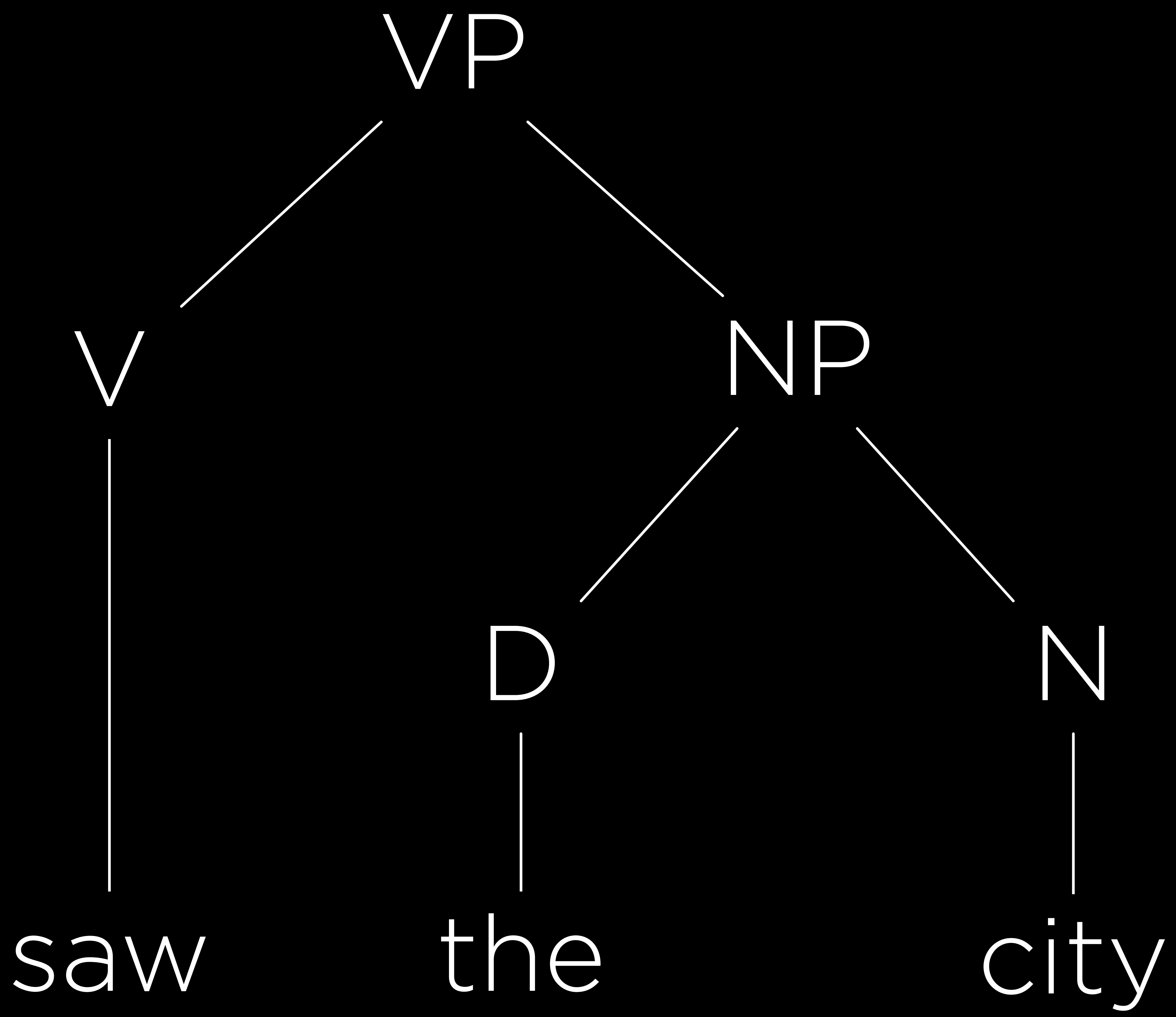


$$VP \rightarrow V \mid V \; NP$$

$$VP \rightarrow V \mid V\ NP$$

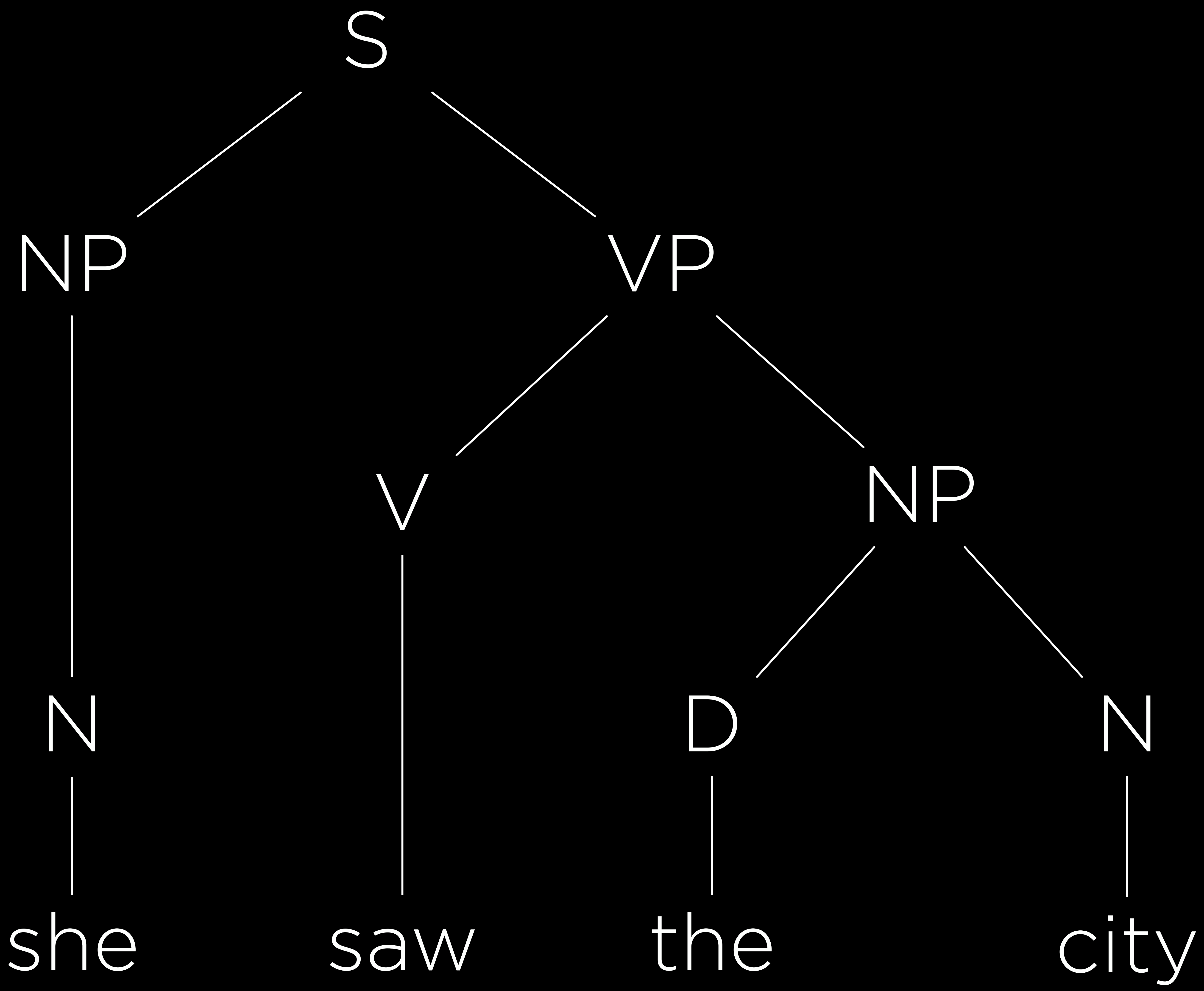


$$VP \rightarrow V \mid V \ NP$$



$$S \rightarrow NP \; VP$$

S → NP VP



nltk

n-gram

a contiguous sequence of n items
from a sample of text

character *n*-gram

a contiguous sequence of *n* characters
from a sample of text

word n-gram

a contiguous sequence of n words
from a sample of text

unigram

a contiguous sequence of 1 item
from a sample of text

bigram

a contiguous sequence of 2 items
from a sample of text

trigrams

a contiguous sequence of 3 items
from a sample of text

"How often have I said to you that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"**How often have** I said to you that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"How **often have I** said to you that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"How often **have I said** to you that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"How often have **I said to** you that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"How often have I **said to you** that when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

"How often have I said **to you that** when you have eliminated the impossible whatever remains, however improbable, must be the truth?"

tokenization

the task of splitting a sequence of
characters into pieces (tokens)

word tokenization

the task of splitting a sequence of characters into words

sentence tokenization

the task of splitting a sequence of characters into sentences

"Whatever remains, however improbable, must be the truth."

"Whatever remains, however improbable, must be the truth."

```
["Whatever", "remains,", "however",
 "improbable,", "must", "be", "the",
 "truth."]
```

"Whatever remains, however improbable, must be the truth."

```
["Whatever", "remains,", "however",
 "improbable,", "must", "be", "the",
 "truth."]
```

"Whatever remains, however improbable, must be the truth."

```
["Whatever", "remains", "however",
 "improbable", "must", "be", "the",
 "truth"]
```

"Just before nine o'clock Sherlock Holmes stepped briskly into the room."

"Just before nine **o'clock** Sherlock Holmes stepped briskly into the room."

"He was dressed in a sombre yet rich style, in black frock-coat, shining hat, neat brown gaiters, and well-cut pearl-grey trousers."

"He was dressed in a sombre yet rich style, in black **frock-coat**, shining hat, neat brown gaiters, and **well-cut pearl-grey** trousers."

"I cannot waste time over this sort of fantastic talk, Sherlock. If you can catch the man, catch him, and let me know when you have done it."

"**I cannot waste time over this sort of fantastic talk, Sherlock.** If you can catch the man, catch him, and let me know when you have done it."

"I cannot waste time over this sort of fantastic talk, Sherlock. **If you can catch the man, catch him, and let me know when you have done it.**"

"I cannot waste time over this sort of fantastic talk, Sherlock. If you can catch the man, catch him, and let me know when you have done it."

"I cannot waste time over this sort of fantastic talk, Mr. Holmes. If you can catch the man, catch him, and let me know when you have done it."

"**I cannot waste time over this sort of fantastic talk, Mr.** Holmes. If you can catch the man, catch him, and let me know when you have done it."

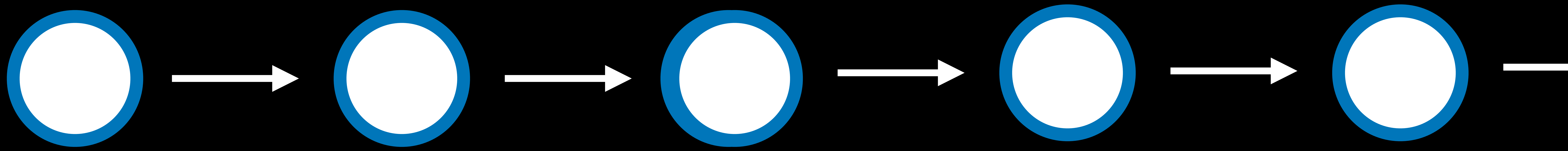
"I cannot waste time over this sort of fantastic talk, Mr. **Holmes.** If you can catch the man, catch him, and let me know when you have done it."

"I cannot waste time over this sort of fantastic talk, Mr. Holmes. **If you can catch the man, catch him, and let me know when you have done it.**"

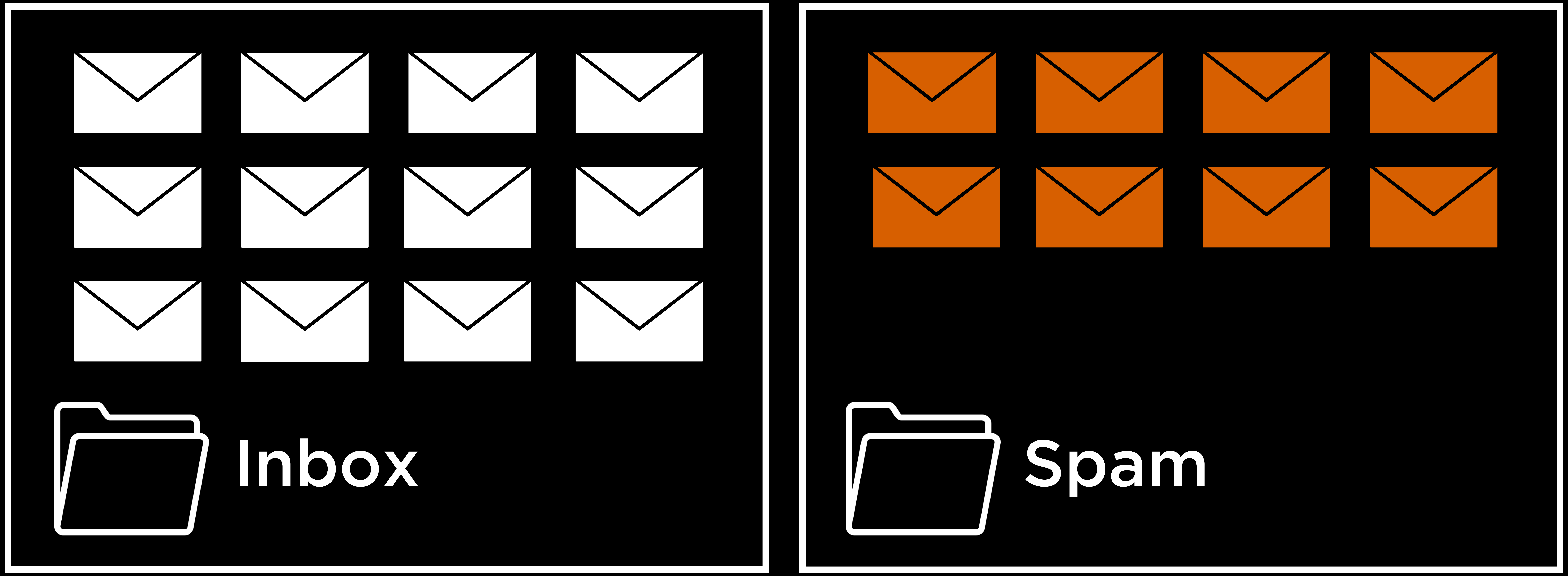
"I cannot waste time over this sort of fantastic talk, Mr. Holmes. If you can catch the man, catch him, and let me know when you have done it."

"I cannot waste time over this sort of fantastic talk, Mr. Holmes," he said. "If you can catch the man, catch him, and let me know when you have done it."

Markov Models



Text Categorization


Inbox
Spam


"My grandson loved it! So much fun!"

"Product broke after a few days."

"One of the best games I've played in a long time."

"Kind of cheap and flimsy, not worth it."

😀 "My grandson loved it! So much fun!"

🙁 "Product broke after a few days."

😀 "One of the best games I've played in a long time."

🙁 "Kind of cheap and flimsy, not worth it."

😀 "My grandson **loved** it! So much **fun**!"

🙁 "Product **broke** after a few days."

😀 "One of the **best** games I've played in a long time."

🙁 "Kind of **cheap** and **flimsy**, not worth it."

bag-of-words model

model that represents text as an unordered collection of words

Naive Bayes

Bayes' Rule

$$P(b \mid a) = \frac{P(a \mid b)\, P(b)}{P(a)}$$

$$P(\text{Positive})$$

$$P(\text{Negative})$$

P(😀)

P(🙁)

"My grandson loved it!"

$P(\text{😀})$

$$P(\text{😀} \mid \text{"my grandson loved it"})$$

$$P(😀 \mid \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

$$P(\text{😀} \mid \text{"my", "grandson", "loved", "it"})$$

$$P(\text{😀} \mid \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

equal to

$$\frac{P(\text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"} \mid \text{😀})\, P(\text{😀})}{P(\text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})}$$

$$P(😀 \mid \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

proportional to

$$P(\text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"} \mid 😀)\, P(😀)$$

$$P(\text{😀} \mid \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

proportional to

$$P(\text{😀}, \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

$$P(😀 \mid \text{"my"}, \text{"grandson"}, \text{"loved"}, \text{"it"})$$

naively proportional to

$$P(😀)P(\text{"my"} \mid 😀)P(\text{"grandson"} \mid 😀)$$
$$P(\text{"loved"} \mid 😀)\, P(\text{"it"} \mid 😀)$$

$$P(\text{😀}) = \frac{\text{number of positive samples}}{\text{number of total samples}}$$

$$P(\text{"loved"} \mid 😀) = \frac{\text{number of positive samples with "loved"}}{\text{number of positive samples}}$$

$$P(😀)P(\text{"my"} \mid 😀)P(\text{"grandson"} \mid 😀)$$
$$P(\text{"loved"} \mid 😀)\ P(\text{"it"} \mid 😀)$$

😀	🙁
0.49	0.51

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

$P(\text{😀})P(\text{"my"} \mid \text{😀})P(\text{"grandson"} \mid \text{😀})$
$P(\text{"loved"} \mid \text{😀})\ P(\text{"it"} \mid \text{😀})$

😀	🙁
0.49	0.51

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

P(😀)P("my" | 😀)P("grandson" | 😀)
P("loved" | 😀) P("it" | 😀)

😀	🙁
0.49	0.51

😀 0.00014112

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

$P(😀)P(\text{"my"} \mid 😀)P(\text{"grandson"} \mid 😀)$
$P(\text{"loved"} \mid 😀)\ P(\text{"it"} \mid 😀)$

😀	🙁
0.49	0.51

😀 0.00014112

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

P(🙁)*P*("my" | 🙁)*P*("grandson" | 🙁)
P("loved" | 🙁) *P*("it" | 🙁)

😀	🙁
0.49	0.51

😀 0.00014112

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

P(🙁)P("my" | 🙁)P("grandson" | 🙁)
P("loved" | 🙁) P("it" | 🙁)

😀	🙁
0.49	0.51

😀 0.00014112

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

$$P(🙁)P(\text{"my"} \mid 🙁)P(\text{"grandson"} \mid 🙁)$$
$$P(\text{"loved"} \mid 🙁)\ P(\text{"it"} \mid 🙁)$$

😀	🙁
0.49	0.51

😀 0.00014112

🙁 0.00006528

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

$$P(\text{🙁})P(\text{"my"} \mid \text{🙁})P(\text{"grandson"} \mid \text{🙁})$$
$$P(\text{"loved"} \mid \text{🙁})\ P(\text{"it"} \mid \text{🙁})$$

😀	🙁
0.49	0.51

😀 0.00014112

🙁 0.00006528

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

P(🙁)*P*("my" | 🙁)*P*("grandson" | 🙁)
P("loved" | 🙁) *P*("it" | 🙁)

😀	🙁
0.49	0.51

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

😀 0.6837

🙁 0.3163

$P($🙁$)P(\text{"my"} \mid $🙁$)P(\text{"grandson"} \mid $🙁$)$
$P(\text{"loved"} \mid $🙁$)\ P(\text{"it"} \mid $🙁$)$

😀	🙁
0.49	0.51

	😀	🙁
my	0.30	0.20
grandson	0.01	0.02
loved	0.32	0.08
it	0.30	0.40

P(🙁)P("my" | 🙁)P("grandson" | 🙁)

P("loved" | 🙁) P("it" | 🙁)

😀	🙁
0.49	0.51

	😀	🙁
my	0.30	0.20
grandson	0.00	0.02
loved	0.32	0.08
it	0.30	0.40

$P(🙁)P(\text{"my"} \mid 🙁)P(\text{"grandson"} \mid 🙁)$
$P(\text{"loved"} \mid 🙁)\ P(\text{"it"} \mid 🙁)$

😀	🙁
0.49	0.51

😀 0.00000000

🙁 0.00006528

	😀	🙁
my	0.30	0.20
grandson	0.00	0.02
loved	0.32	0.08
it	0.30	0.40

additive smoothing

adding a value α to each value in our distribution to smooth the data

Laplace smoothing

adding 1 to each value in our distribution: pretending we've seen each value one more time than we actually have

information retrieval

the task of finding relevant documents in response to a user query

topic modeling

models for discovering the topics for a set of documents

term frequency

number of times a term appears in a document

function words

words that have little meaning on their own, but are used to grammatically connect other words

function words

am, by, do, is, which, with, yet, …

content words

words that carry meaning independently

content words

algorithm, category, computer, ...

inverse document frequency

measure of how common or rare a word is across documents

inverse document frequency

$$\log \frac{TotalDocuments}{NumDocumentsContaining(word)}$$

tf-idf

ranking of what words are important in a document by multiplying term frequency (TF) by inverse document frequency (IDF)

Semantics

information extraction

the task of extracting knowledge from documents

"When Facebook was founded in 2004, it began with a seemingly innocuous mission: to connect friends. Some seven years and 800 million users later, the social network has taken over most aspects of our personal and professional lives, and is fast becoming the dominant communication platform of the future."

Harvard Business Review, 2011

"Remember, back when Amazon was founded in 1994, most people thought his idea to sell books over this thing called the internet was crazy. A lot of people had never even hard of the internet."

Business Insider, 2018

"When **Facebook** was founded in **2004**, it began with a seemingly innocuous mission: to connect friends. Some seven years and 800 million users later, the social network has taken over most aspects of our personal and professional lives, and is fast becoming the dominant communication platform of the future."

Harvard Business Review, 2011

"Remember, back when **Amazon** was founded in **1994**, most people thought his idea to sell books over this thing called the internet was crazy. A lot of people had never even hard of the internet."

Business Insider, 2018

"When **Facebook** was founded in **2004**, it began with a seemingly innocuous mission: to connect friends. Some seven years and 800 million users later, the social network has taken over most aspects of our personal and professional lives, and is fast becoming the dominant communication platform of the future."

Harvard Business Review, 2011

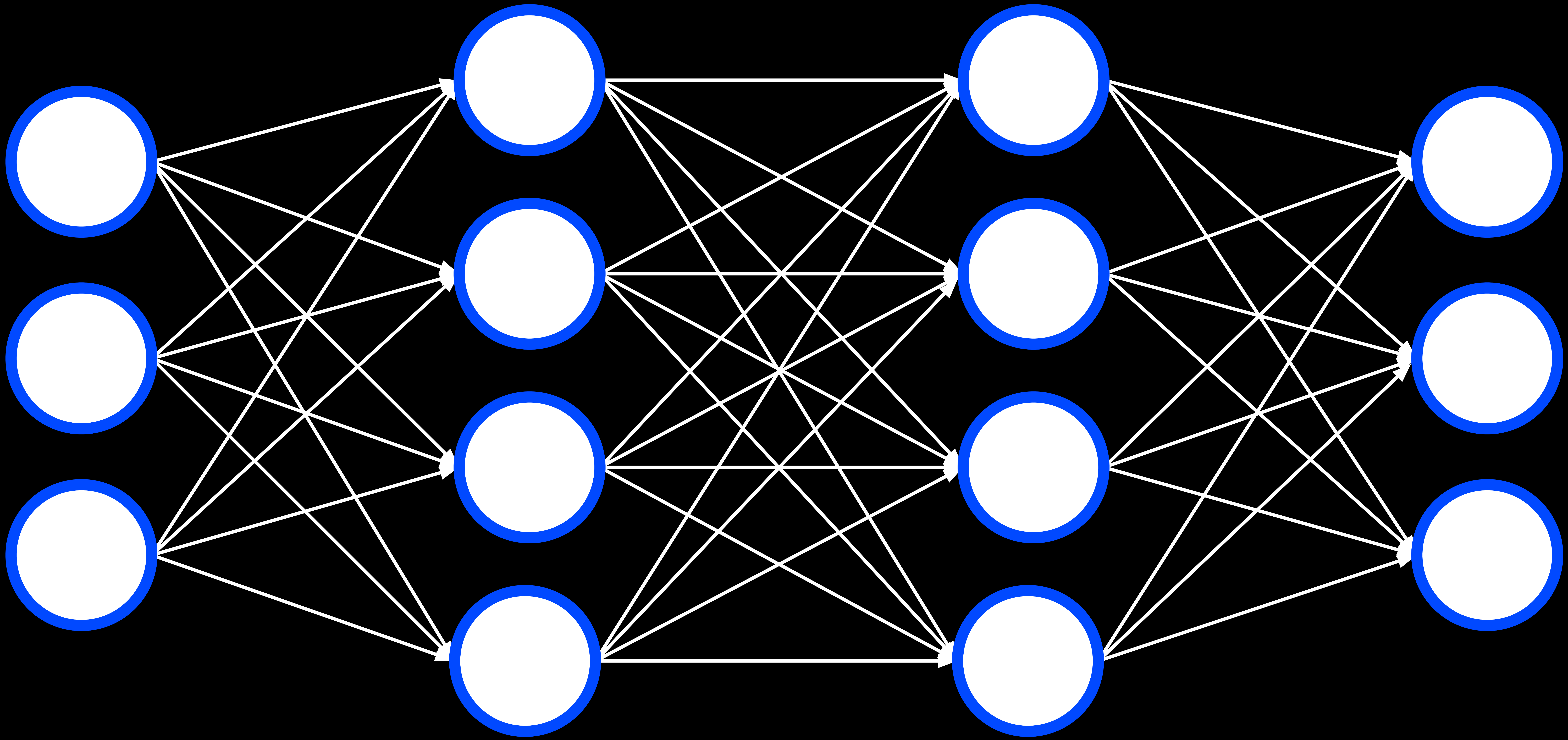
"Remember, back when **Amazon** was founded in **1994**, most people thought his idea to sell books over this thing called the internet was crazy. A lot of people had never even hard of the internet."

Business Insider, 2018

When {company} was founded in {year},

WordNet

Word Representation



"He wrote a book."

```
he [1, 0, 0, 0]
wrote [0, 1, 0, 0]
a [0, 0, 1, 0]
book [0, 0, 0, 1]
```

one-hot representation

representation of meaning as a vector with a single 1, and with other values as 0

"He wrote a book."

```
he [1, 0, 0, 0]
wrote [0, 1, 0, 0]
a [0, 0, 1, 0]
book [0, 0, 0, 1]
```

"He wrote a book."

```
he [1, 0, 0, 0, 0, 0, 0, 0, ...]
wrote [0, 1, 0, 0, 0, 0, 0, ...]
a [0, 0, 1, 0, 0, 0, 0, 0, ...]
book [0, 0, 0, 1, 0, 0, 0, ...]
```

"He wrote a book."

"He authored a novel."

```
wrote [0, 1, 0, 0, 0, 0, 0, 0, 0]
authored [0, 0, 0, 0, 1, 0, 0, 0, 0]

book [0, 0, 0, 0, 0, 0, 1, 0, 0]
novel [0, 0, 0, 0, 0, 0, 0, 0, 1]
```

distribution representation

representation of meaning distributed across multiple values

"He wrote a book."

```
he [-0.34, -0.08, 0.02, -0.18, 0.22, ...]
wrote [-0.27, 0.40, 0.00, -0.65, -0.15, ...]
a [-0.12, -0.25, 0.29, -0.09, 0.40, ...]
book [-0.23, -0.16, -0.05, -0.57, ...]
```

"You shall know a word
by the company it keeps."

J. R. Firth, 1957

for		he	ate

for	**breakfast**	he	ate

for	**lunch**	he	ate

for	**dinner**	he	ate

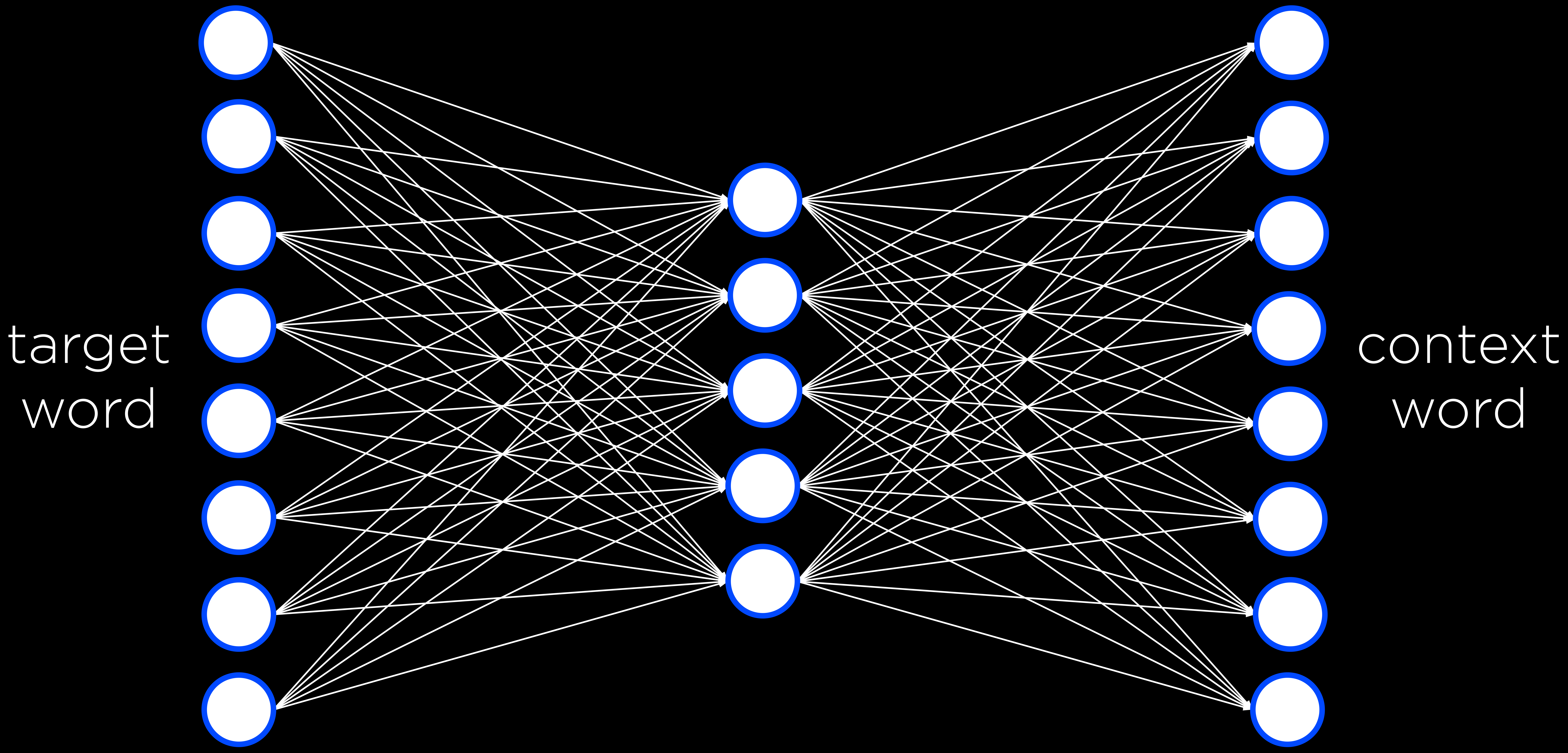
for		he	ate

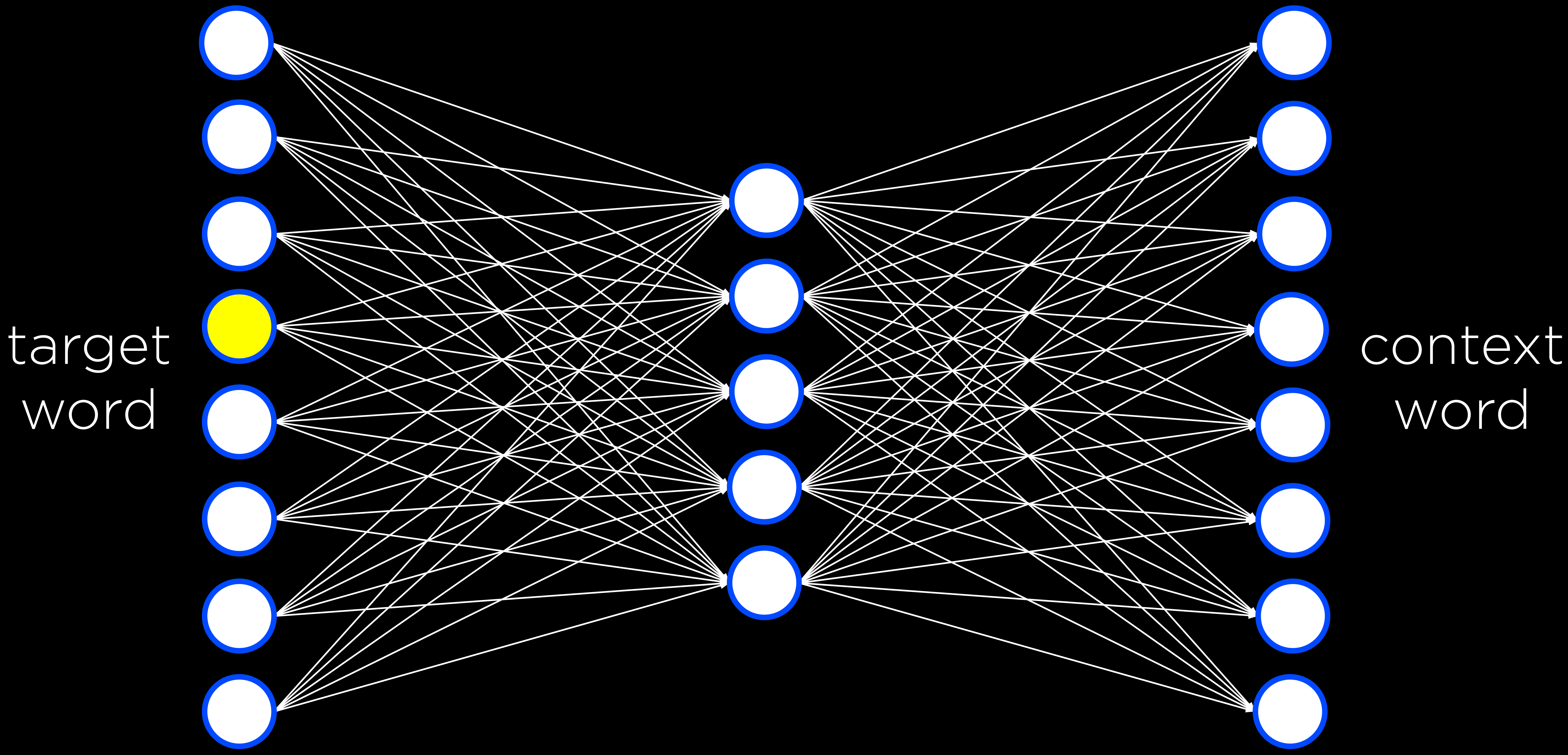
word2vec

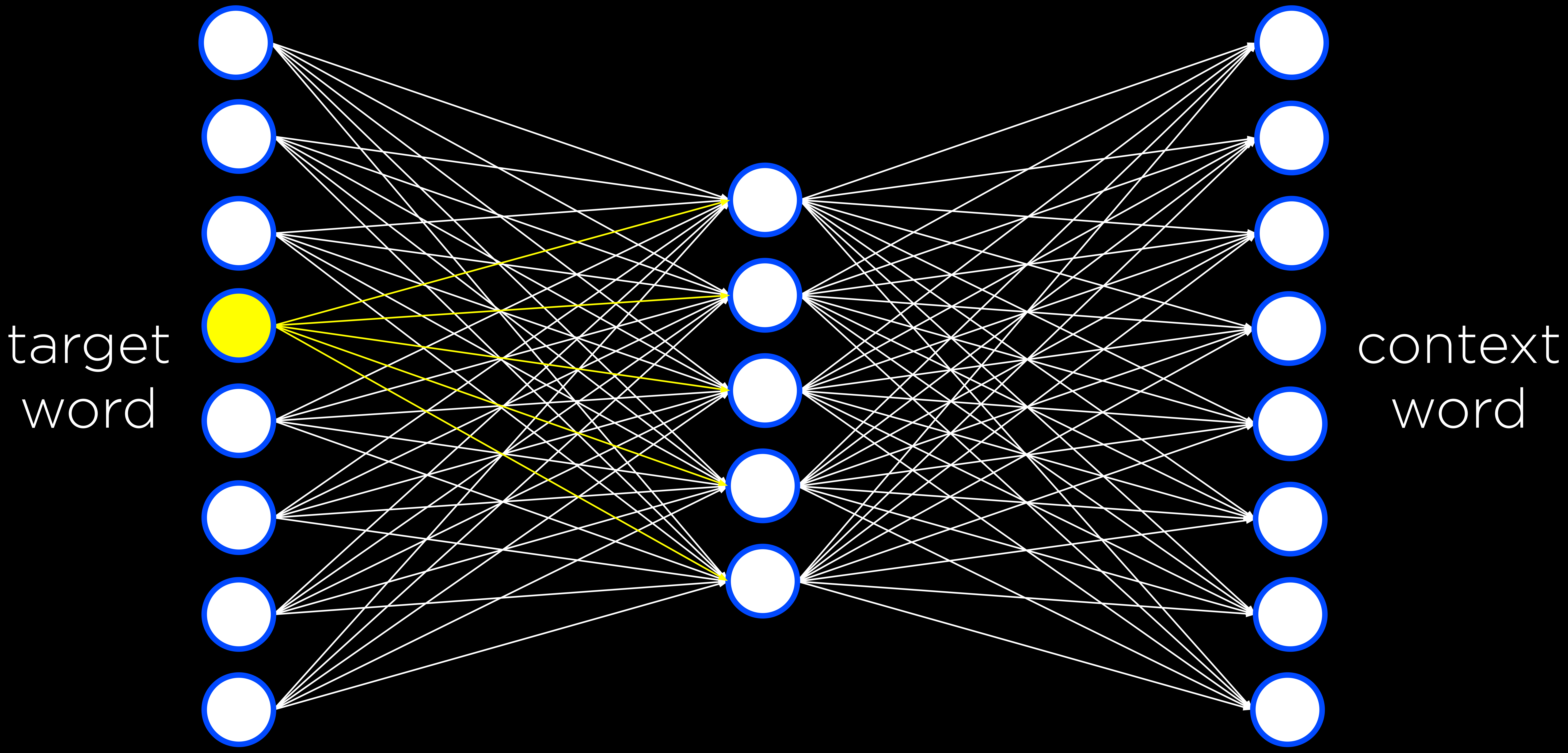
model for generating word vectors

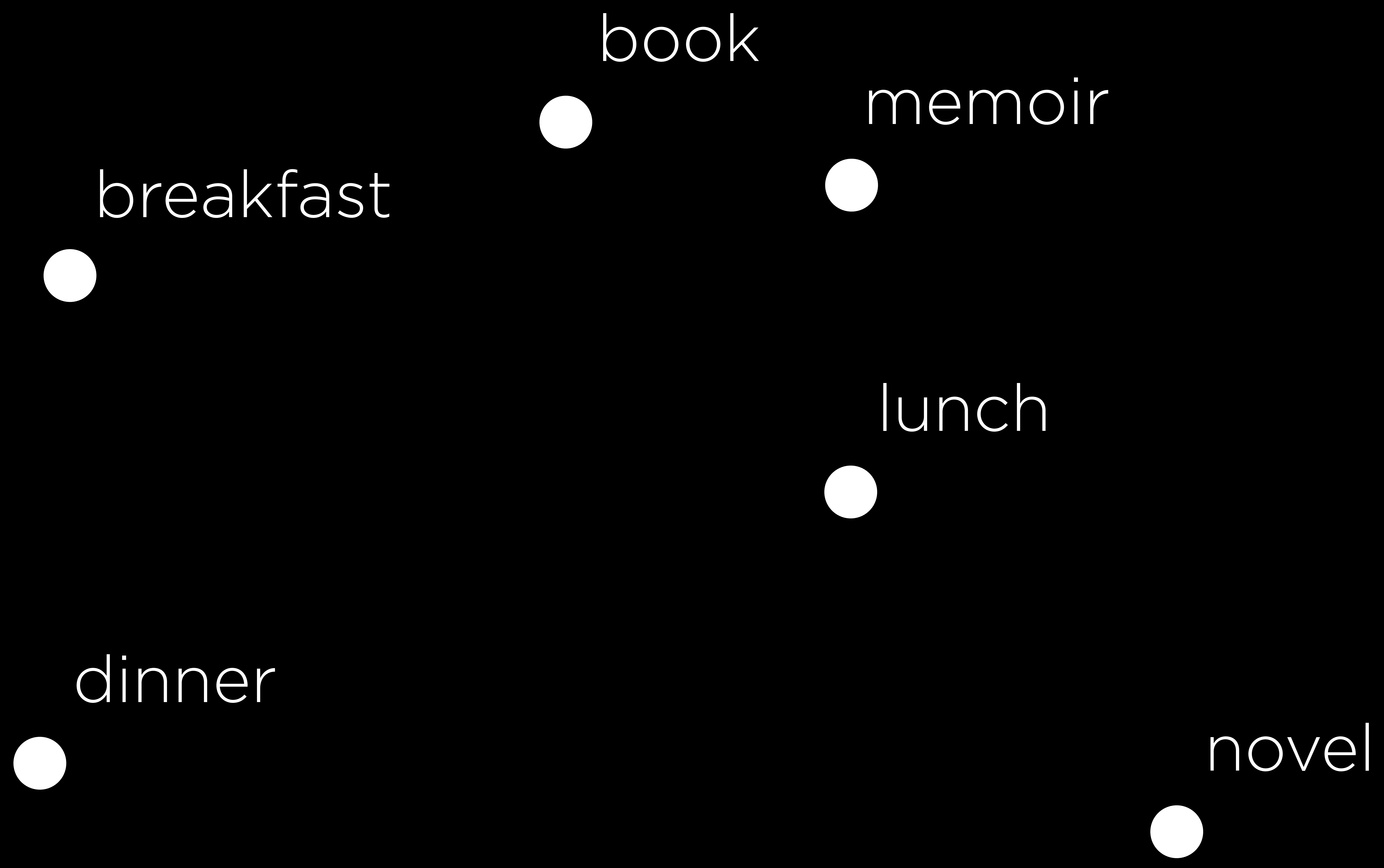
skip-gram architecture

neural network architecture for predicting context words given a target word

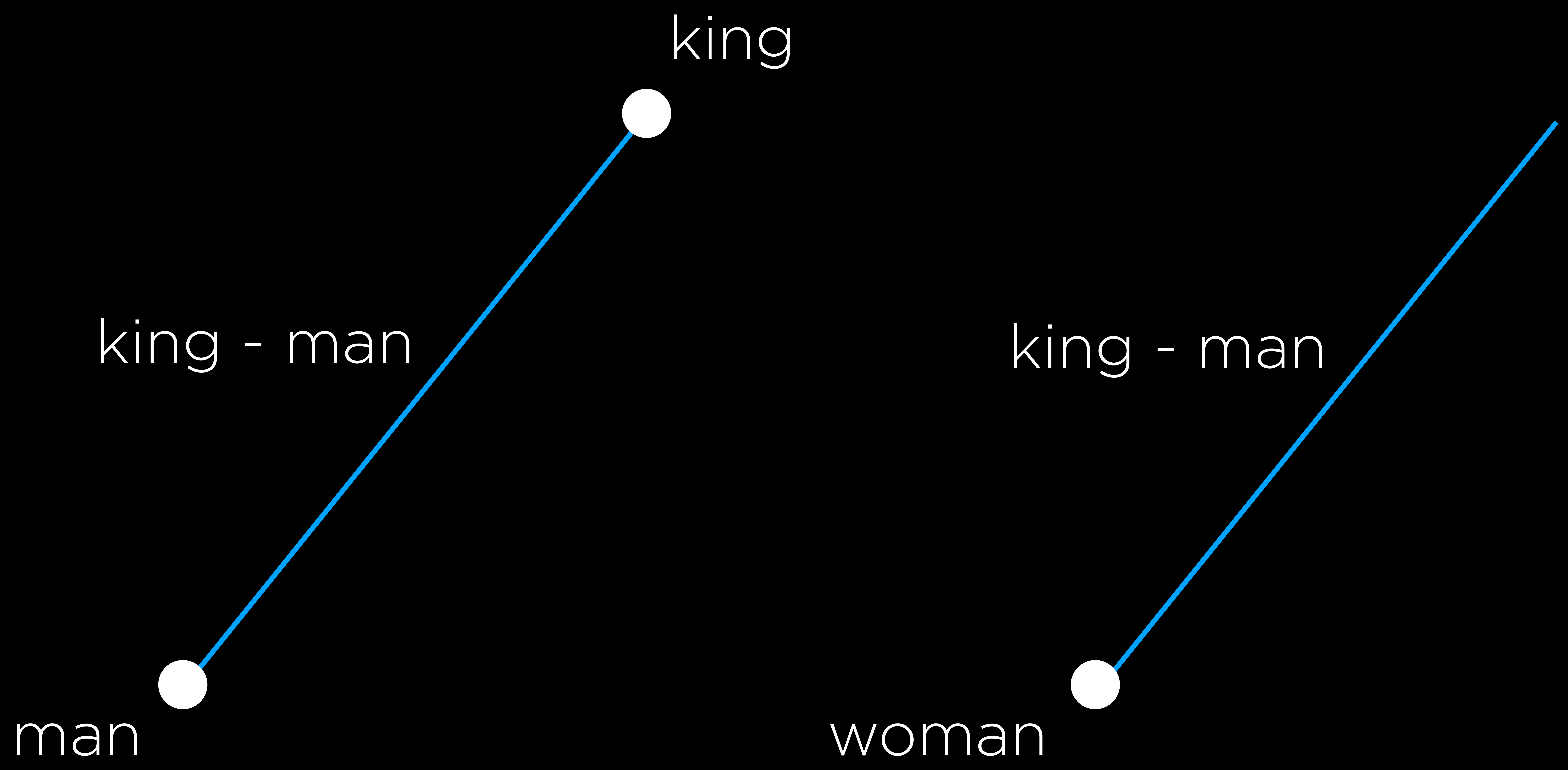

target
word
context
word

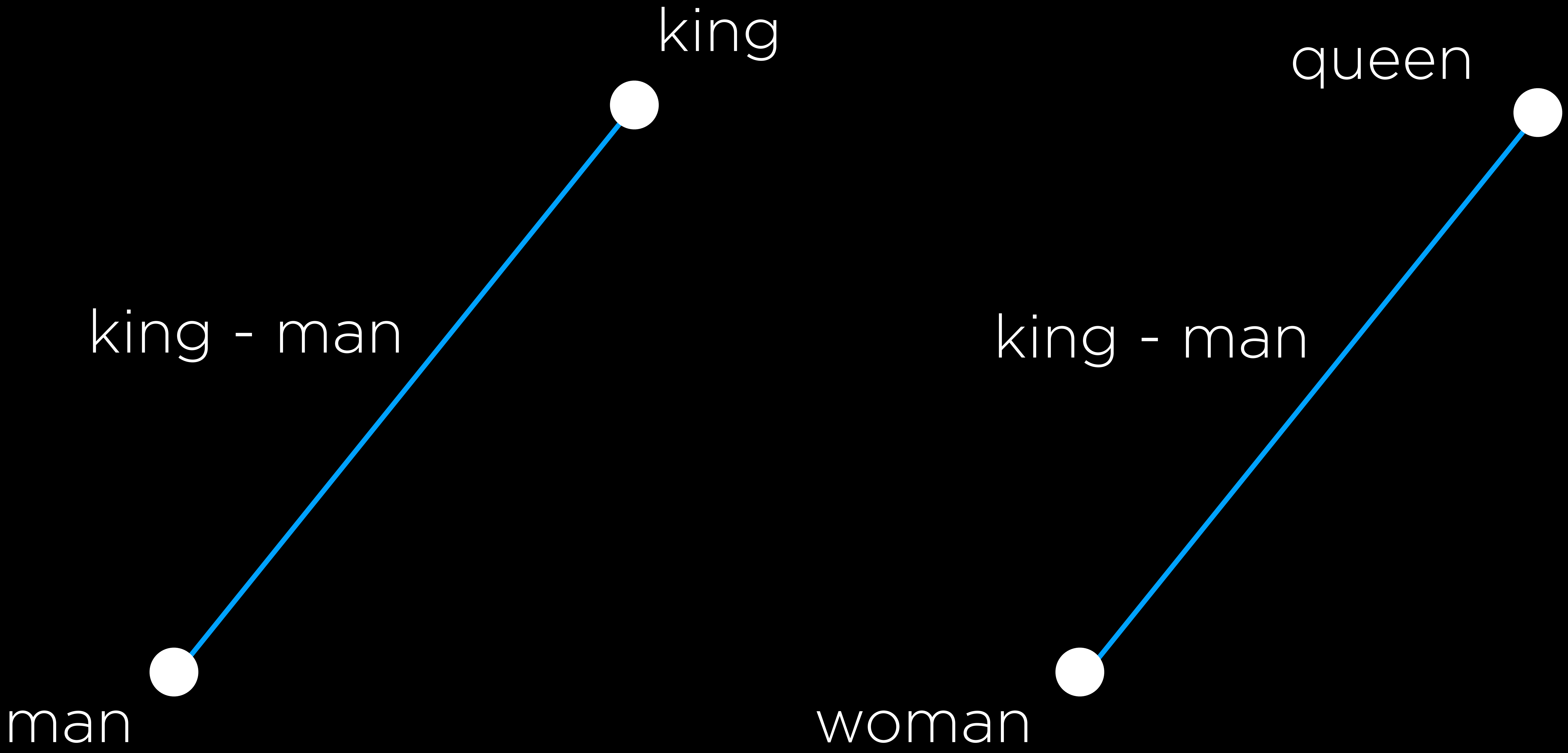

target
word
context
word


target
word
context
word


book
memoir
breakfast
lunch
dinner
novel


memoir
book
breakfast
novel
lunch
dinner

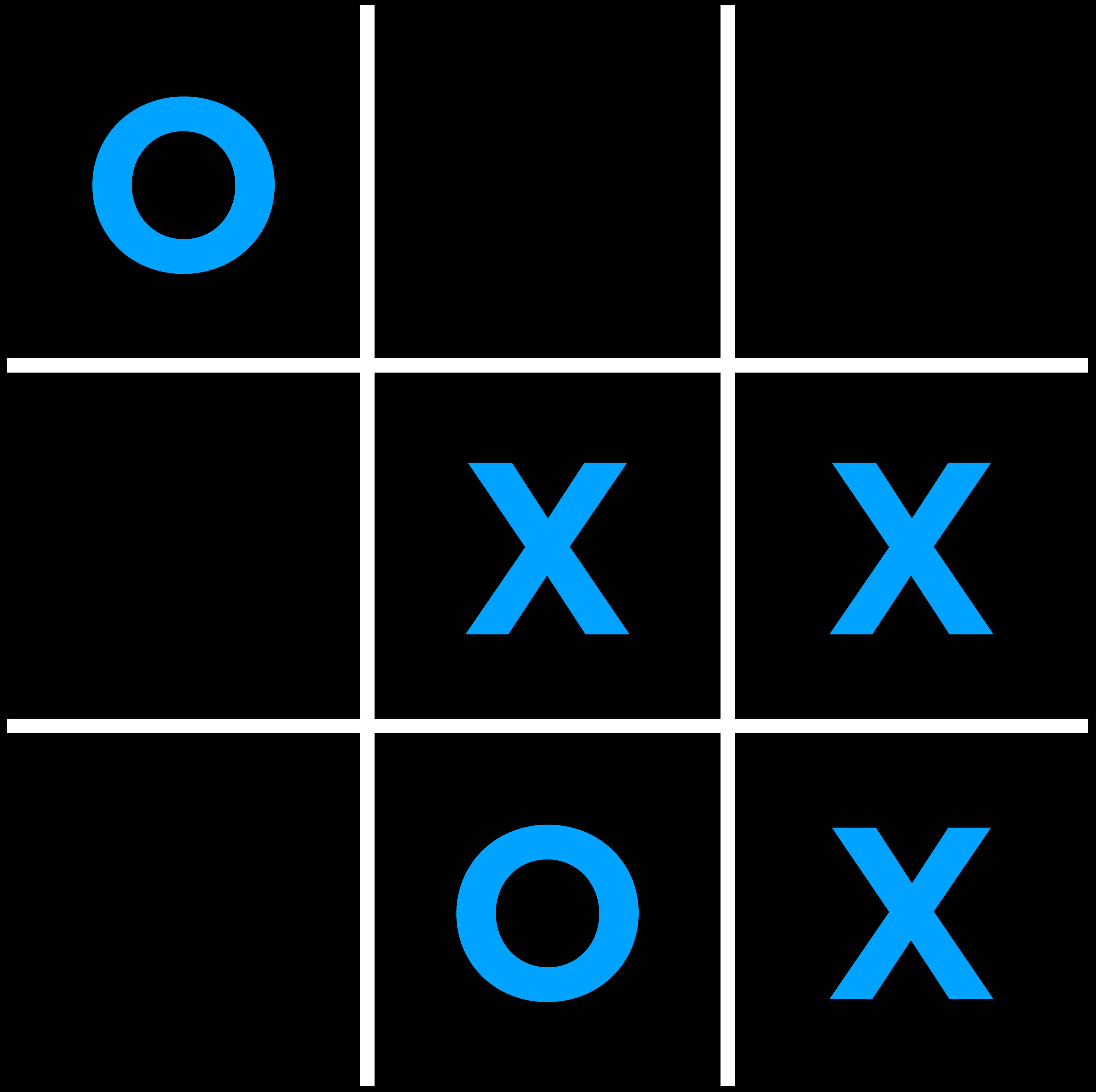

king
king - man
man
king - man
woman


king
queen
king - man
king - man
man
woman

Language

Artificial Intelligence

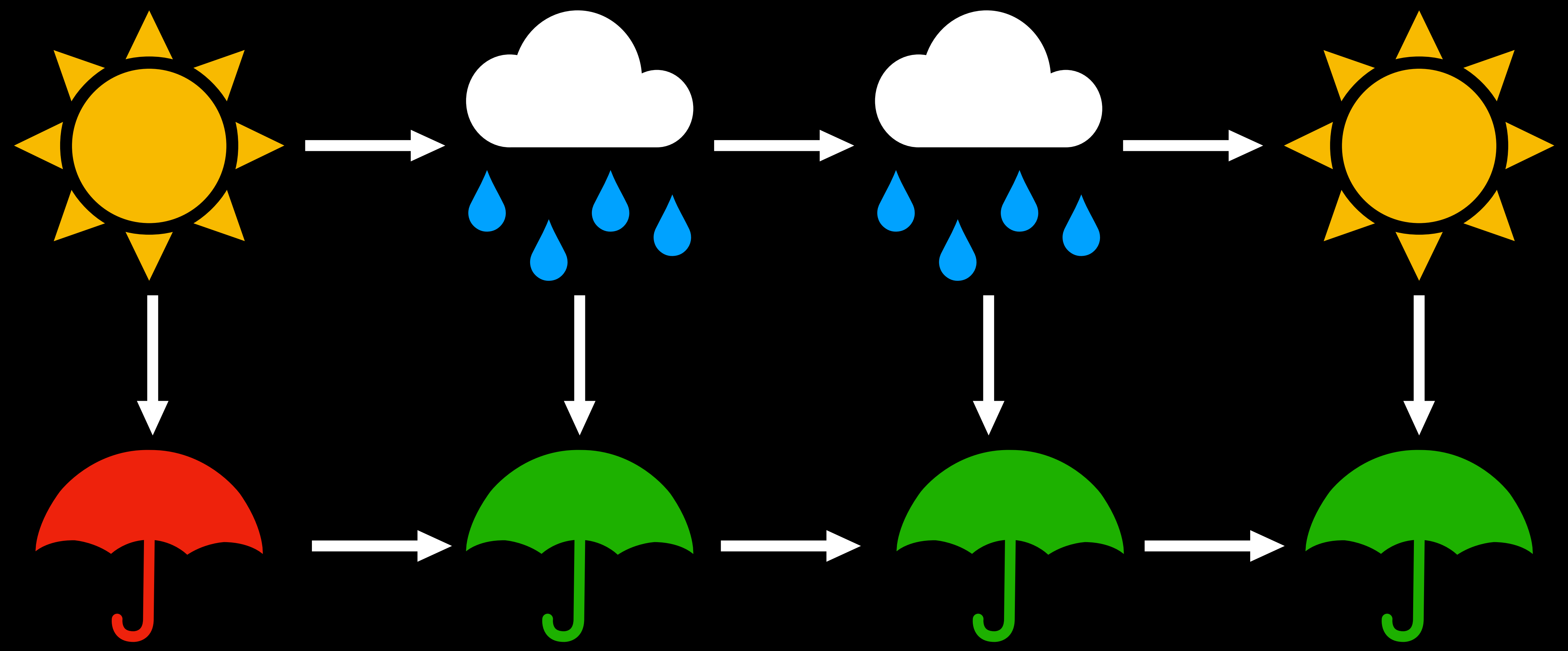
Search



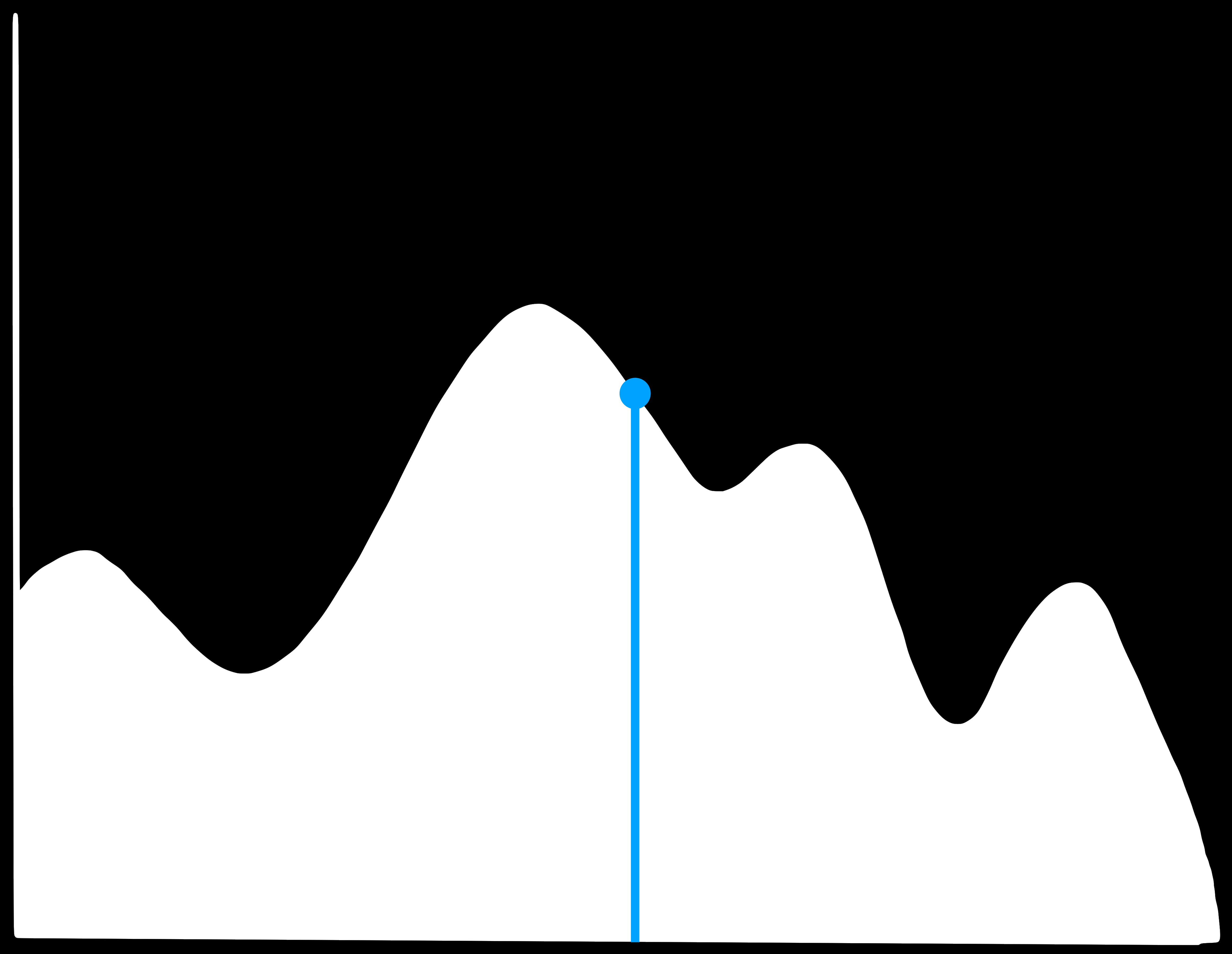
Knowledge

$$\frac{\begin{array}{c} P \rightarrow Q \\ P \end{array}}{Q}$$

Uncertainty

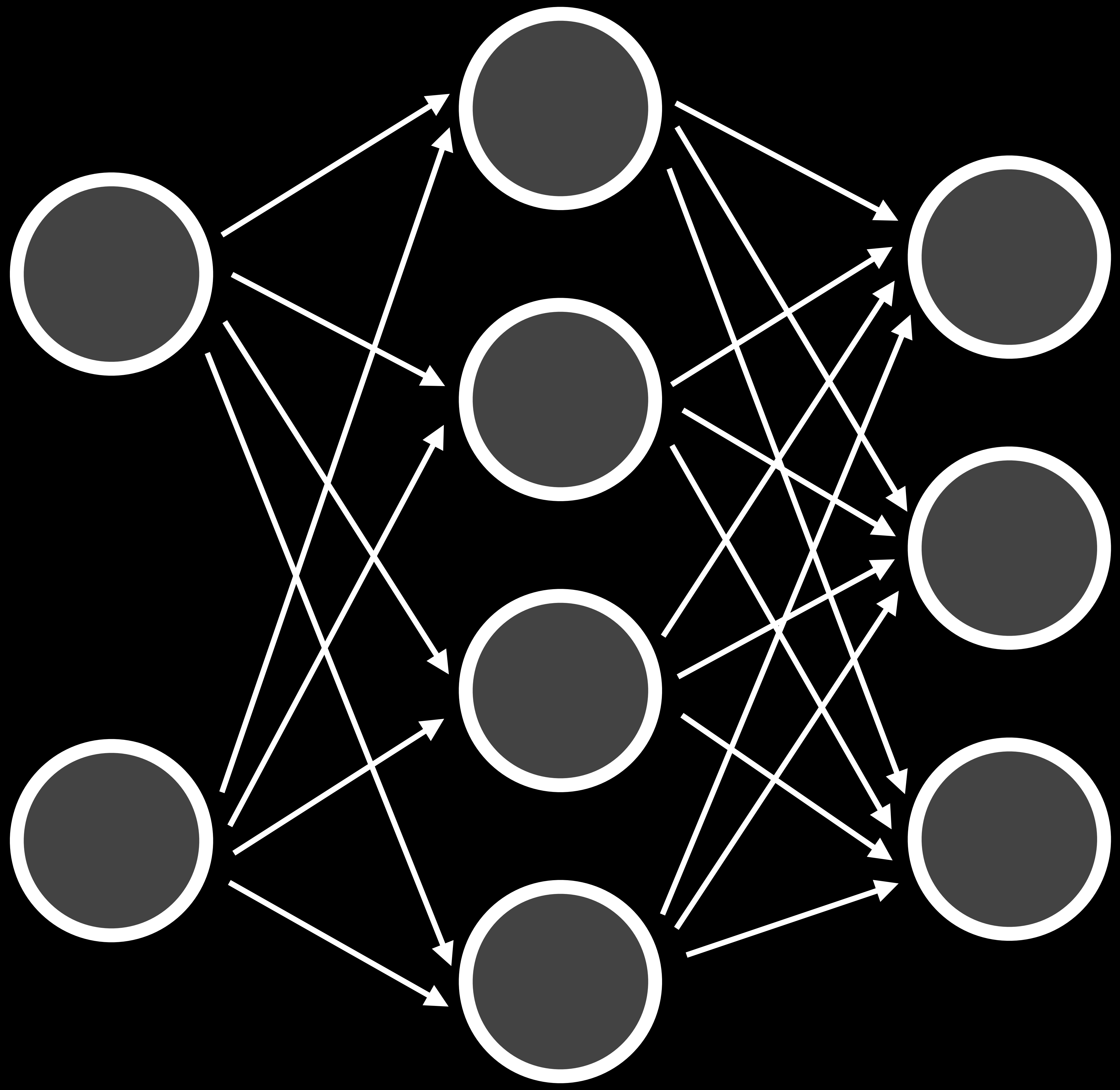


Optimization

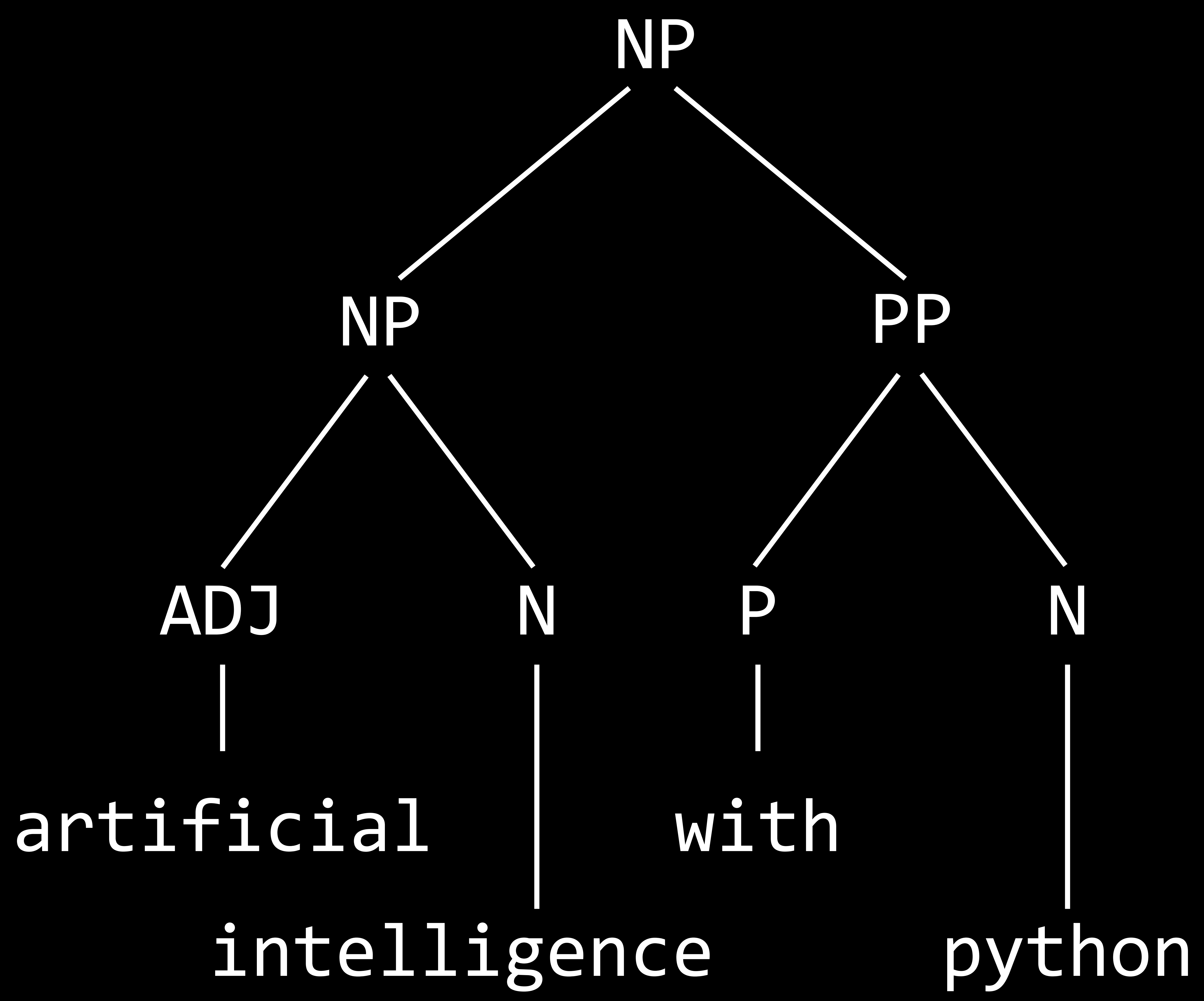


Learning

Neural Networks



Language



Introduction to

Artificial Intelligence

with Python