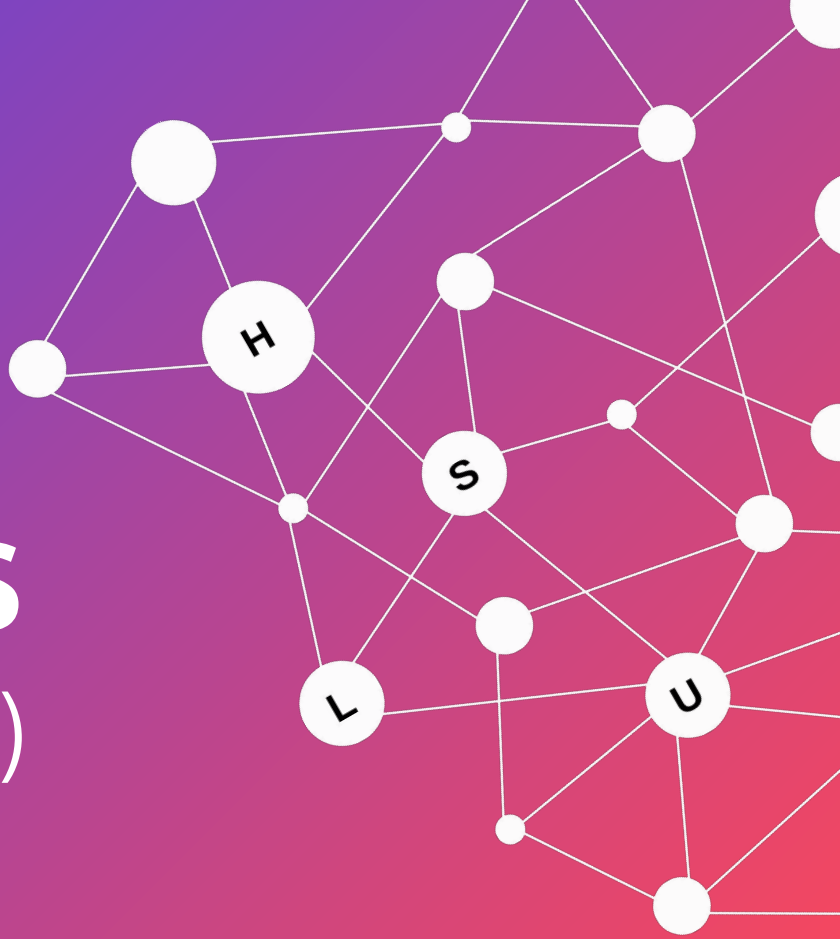
Why AI Loves Vectors

(And You Should Too!)



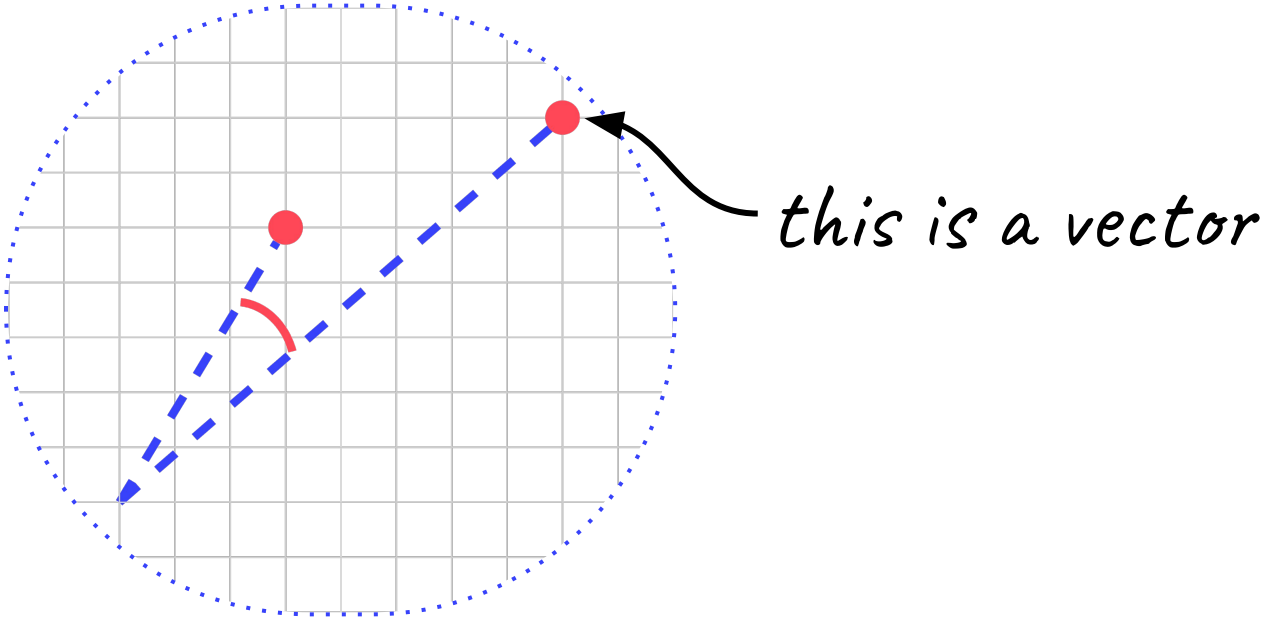
Andrei Cătălin Coman
andrei.coman@idiap.ch

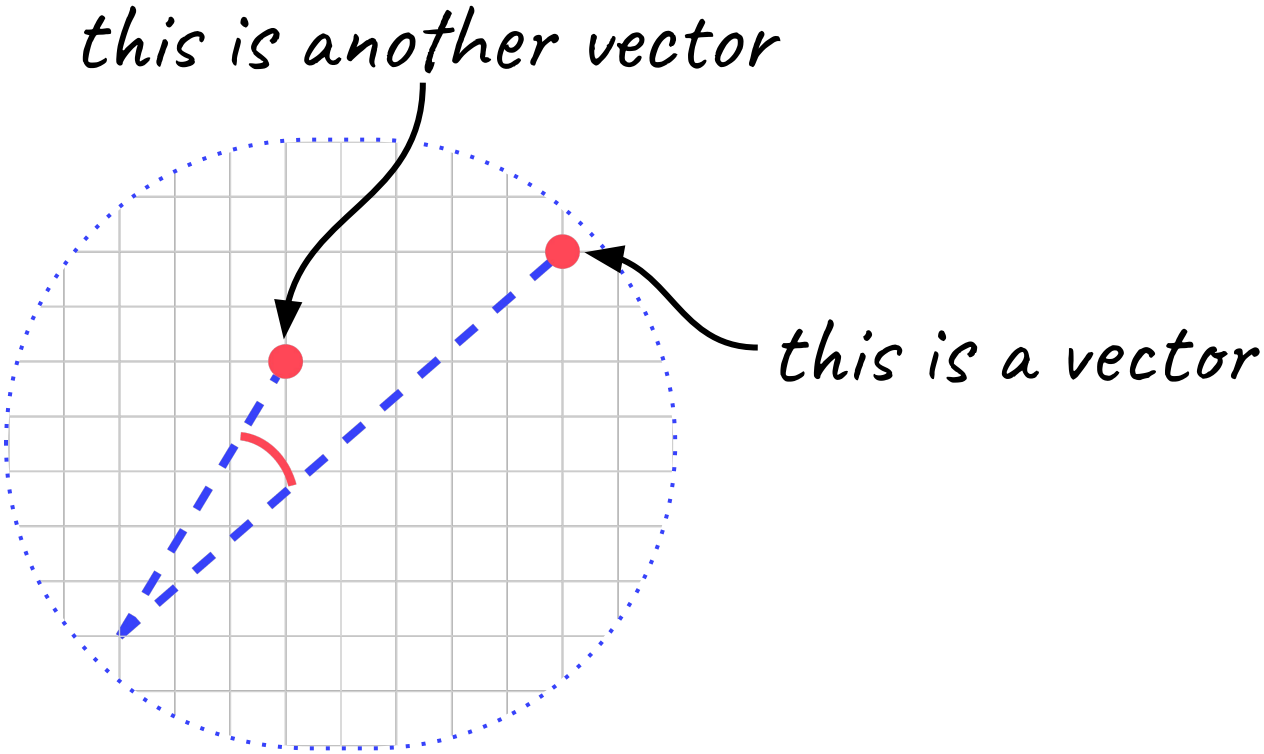
HSLU Hochschule Luzern idiap RESEARCH INSTITUTE EPFL

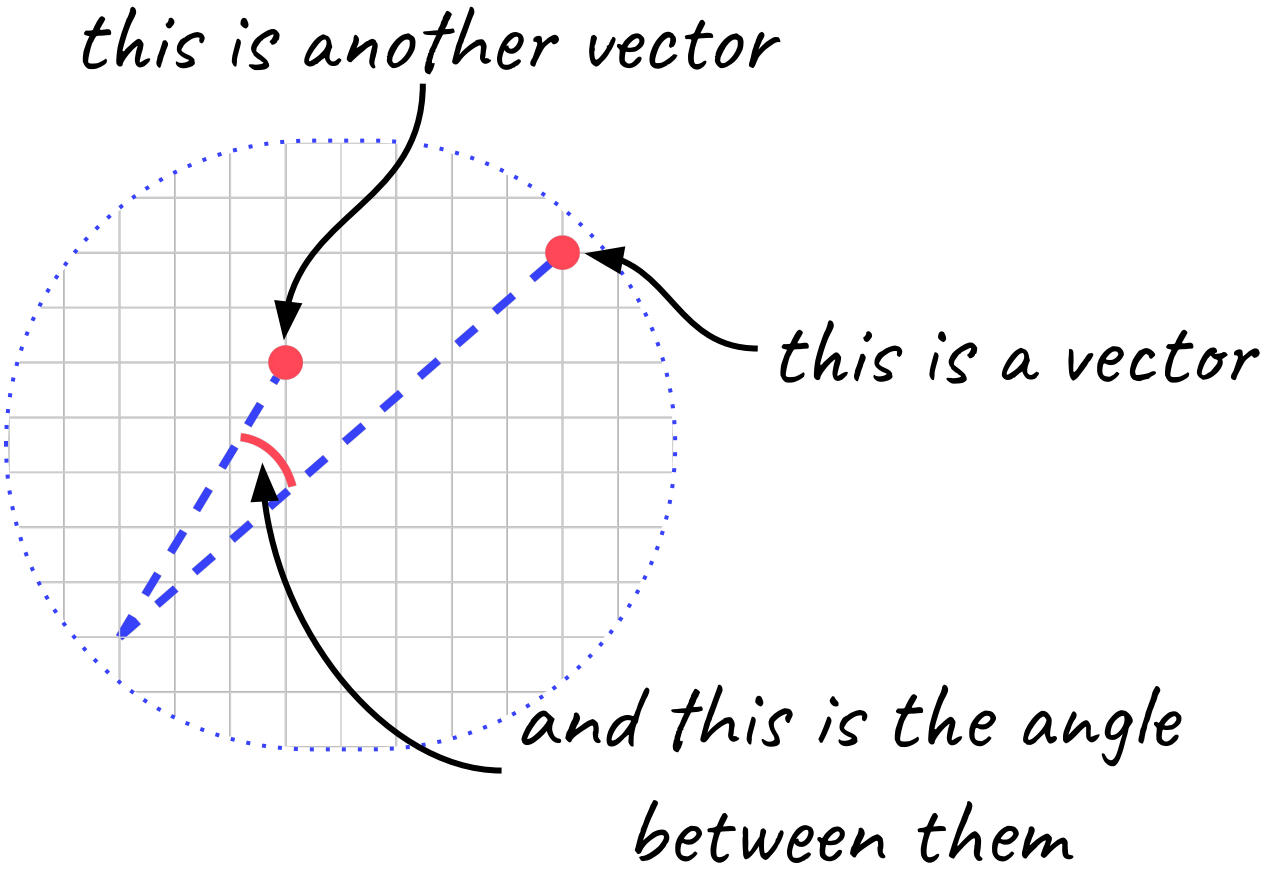
We are witnessing an AI revolution.

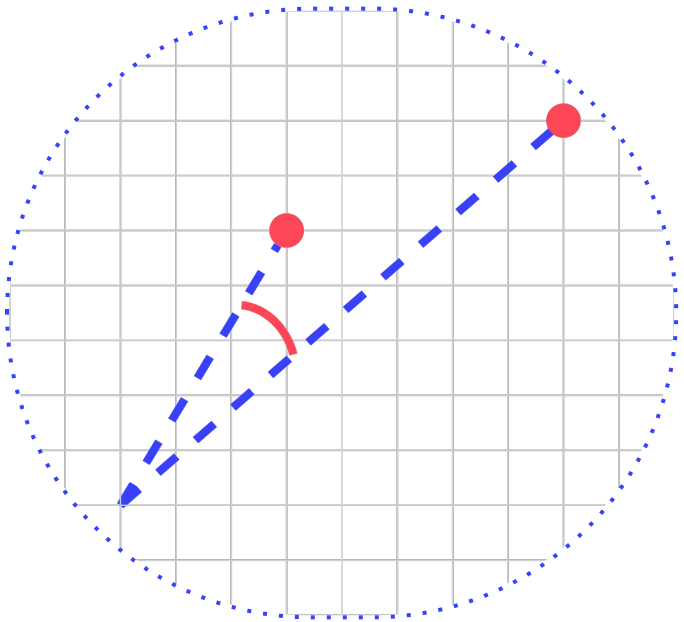
You can understand it, build it, and shape it.

You have access to tools, knowledge, and opportunities.


this is a vector

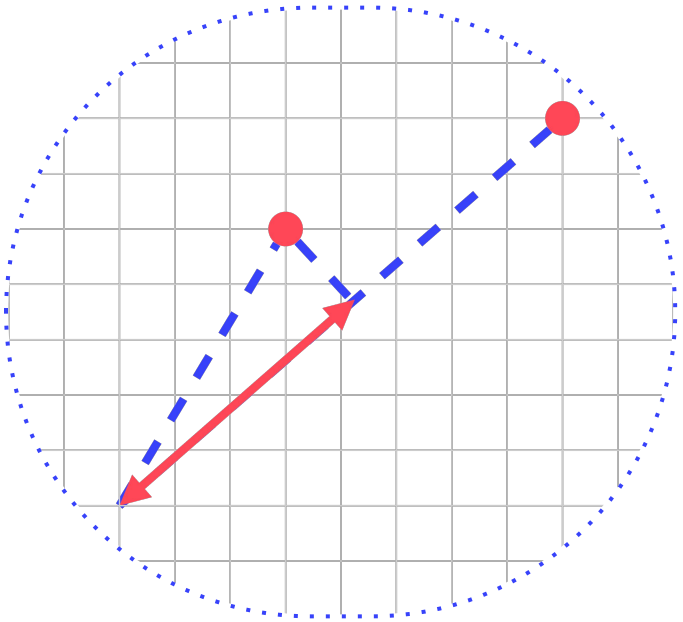

this is another vector
this is a vector


this is another vector
this is a vector
and this is the angle
between them



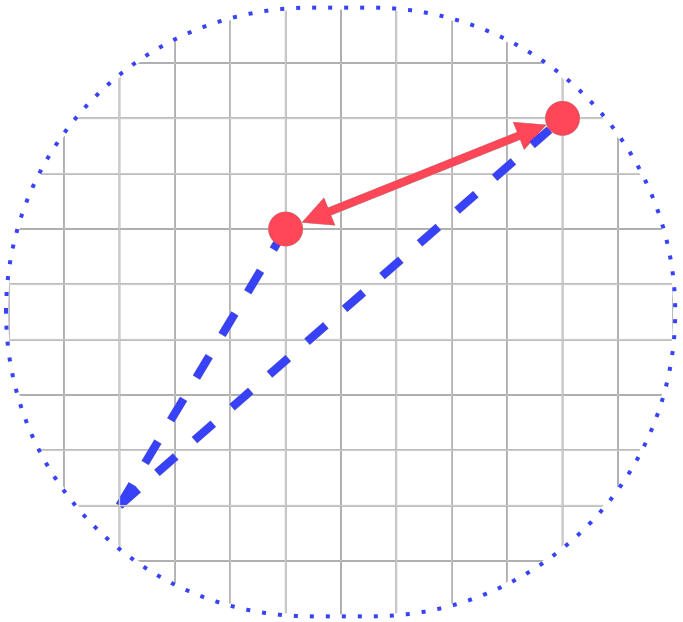
Cosine Distance

$$1 - \frac{A \cdot B}{||A|| \cdot ||B||}$$



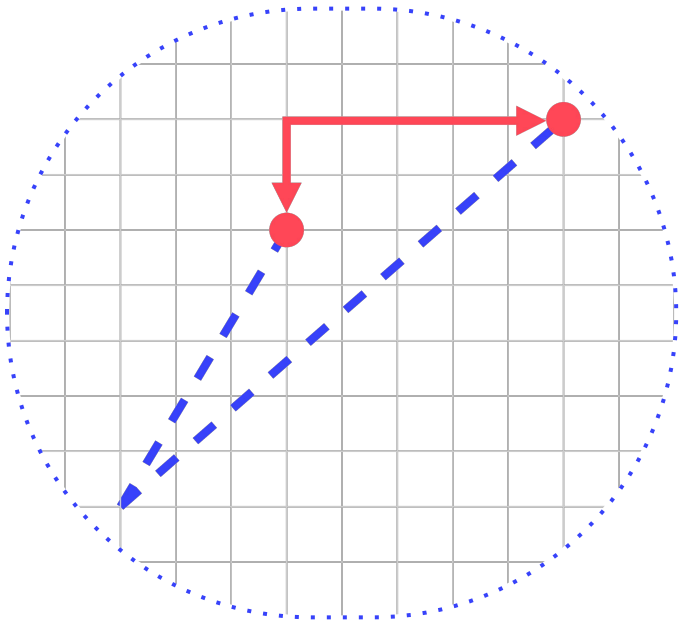
Dot Product

$$A \cdot B = \sum_{i=1}^{n} A_i B_i$$



Euclidean (L2)

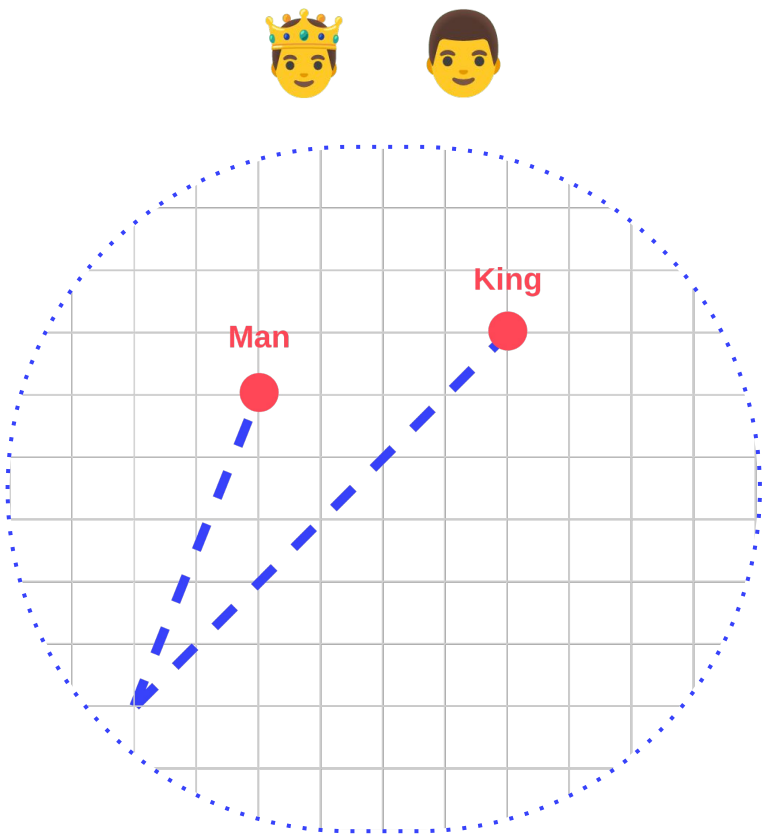
$$\sqrt{\sum_{i=1}^{n}(A_i - B_i)^2}$$

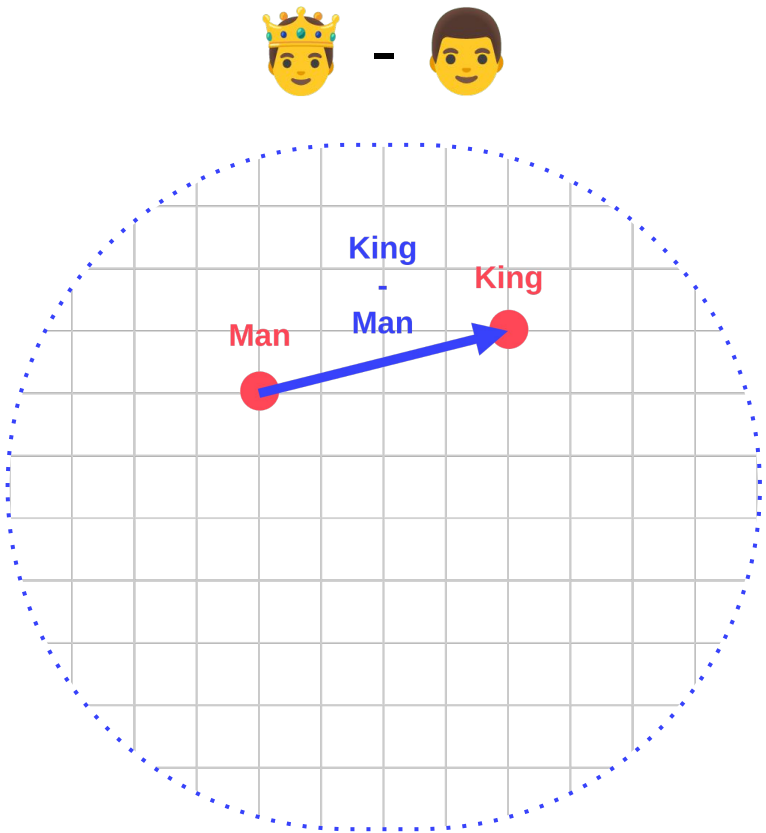


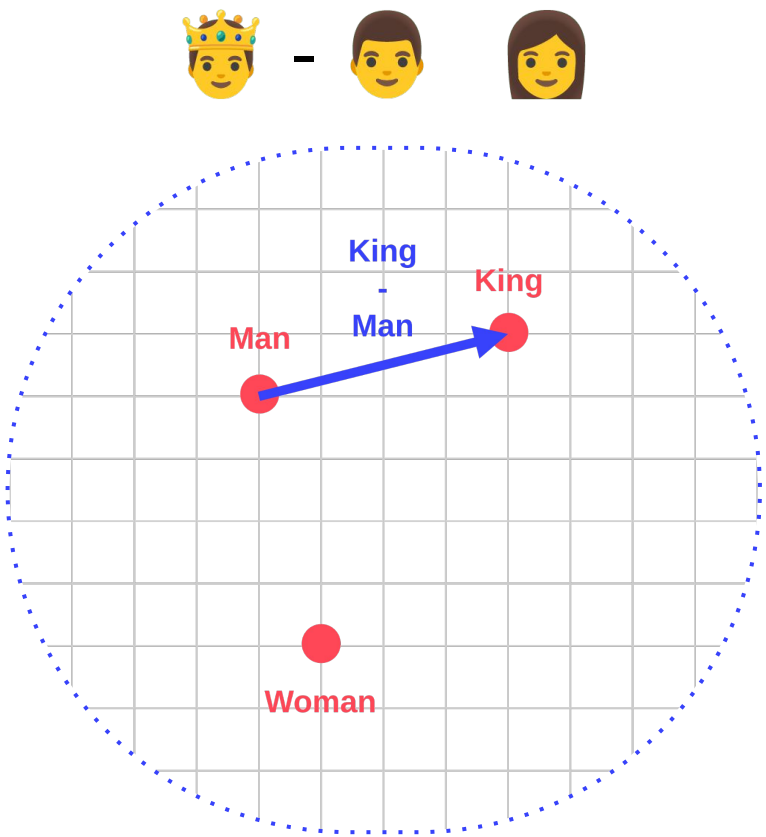
Manhattan (L1)

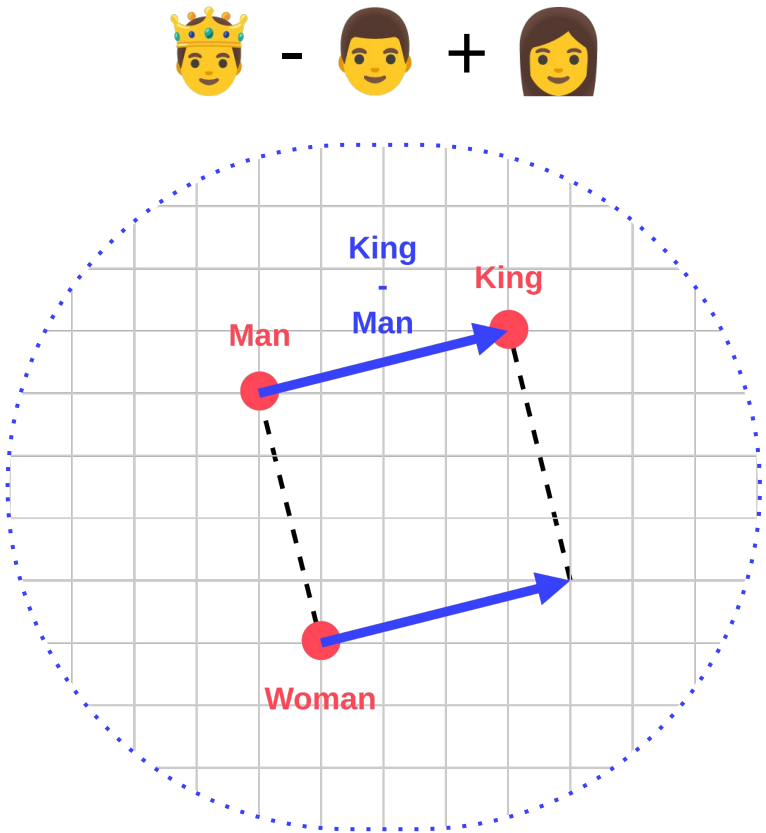
$$\sum_{i=1}^{n} |A_i - B_i|$$

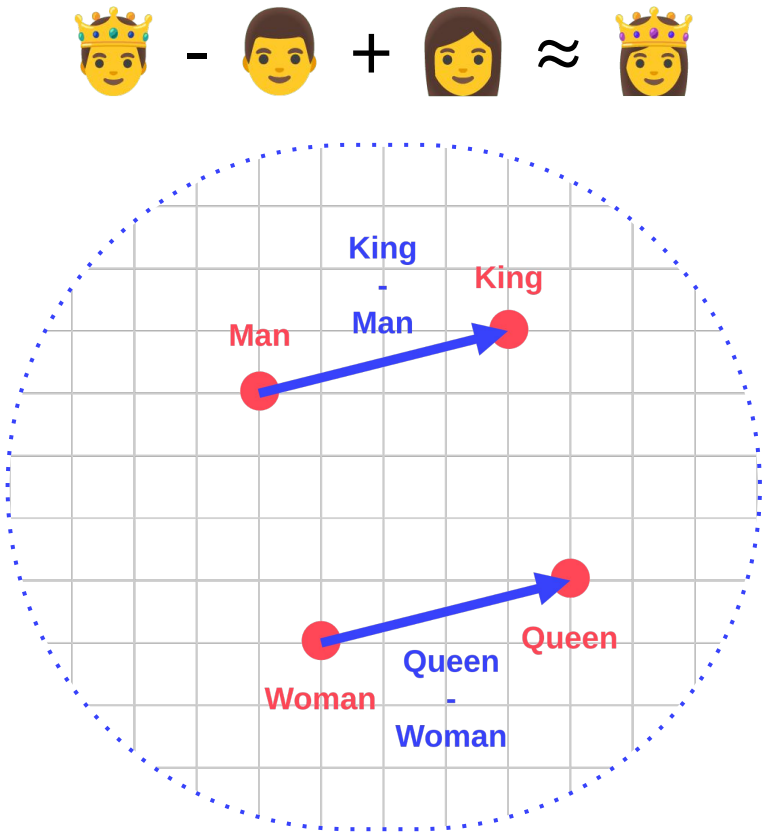
Quantify the relation between
vector A and vector B

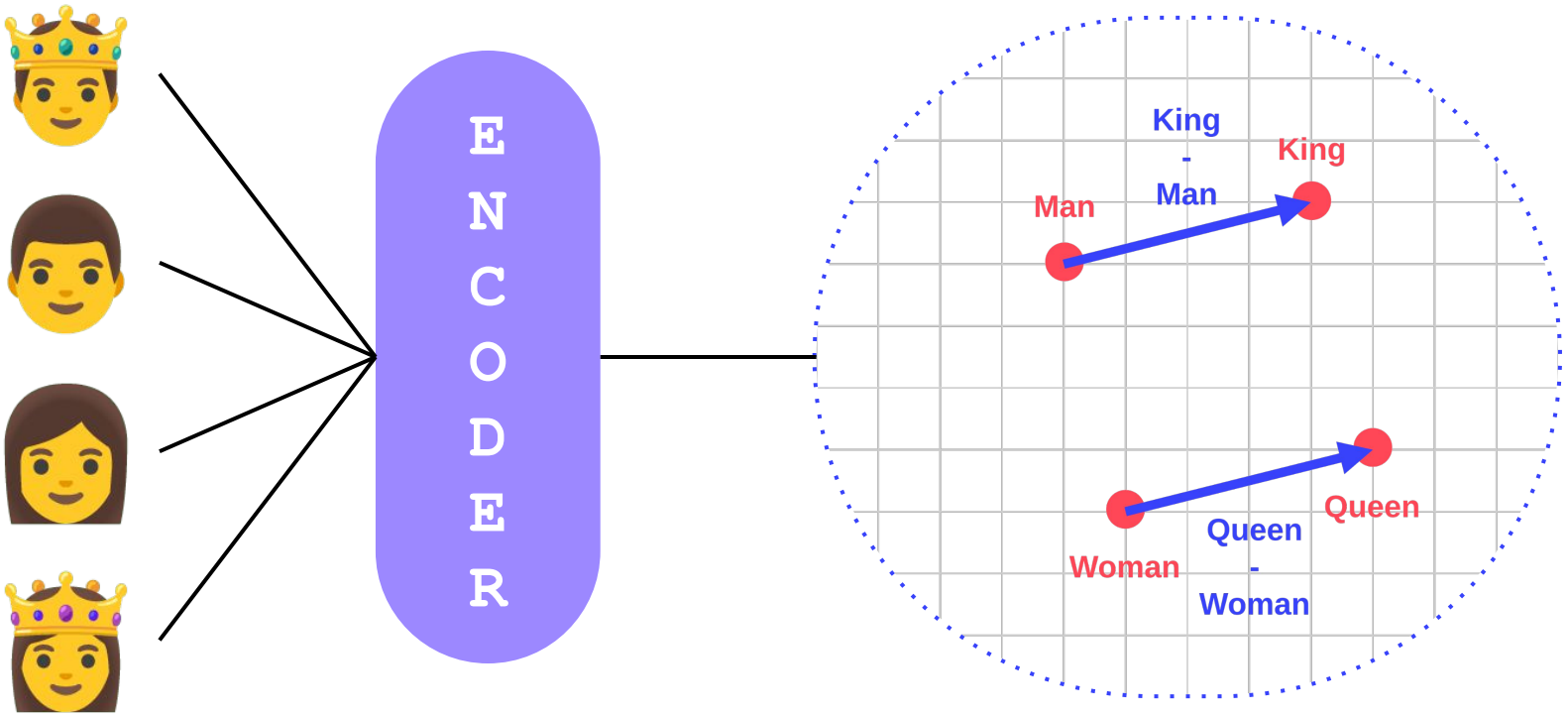

👑
👨
King
Man


King
-
Man
King
Man


King
-
Man
King
Man
Woman


King
-
Man
King
Man
Woman


King
-
Man
King
Man
Queen
Queen
-
Woman
Woman

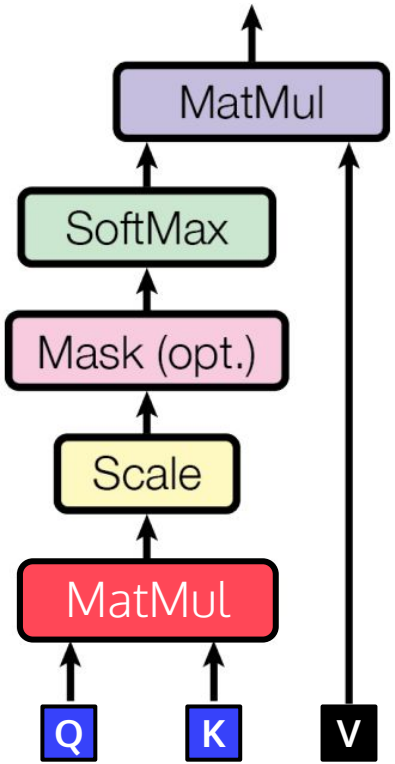

ENCODER
King
-
Man
King
Man
Queen
-
Woman
Queen
Woman

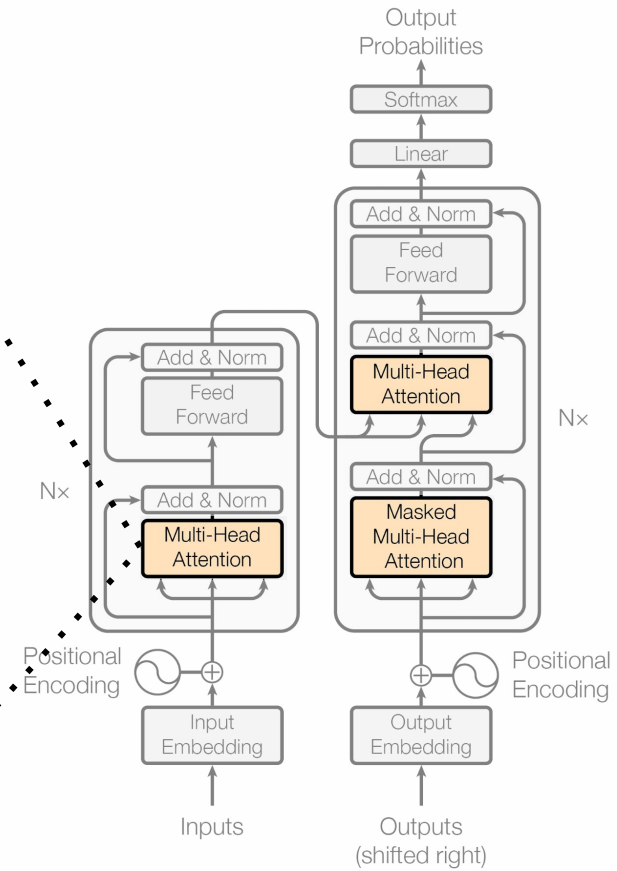
Quantify the relation between head and tail

does magic for

head relation tail

Quantify the relation between query and key


MatMul
SoftMax
Mask (opt.)
Scale
MatMul
Q
K
V


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

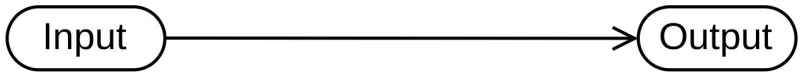
Vectors

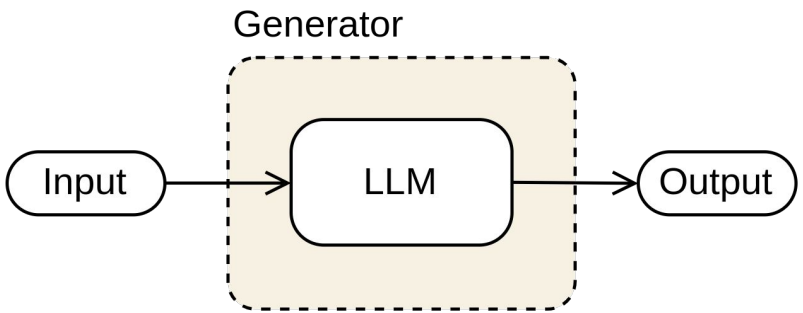
encode meaning.

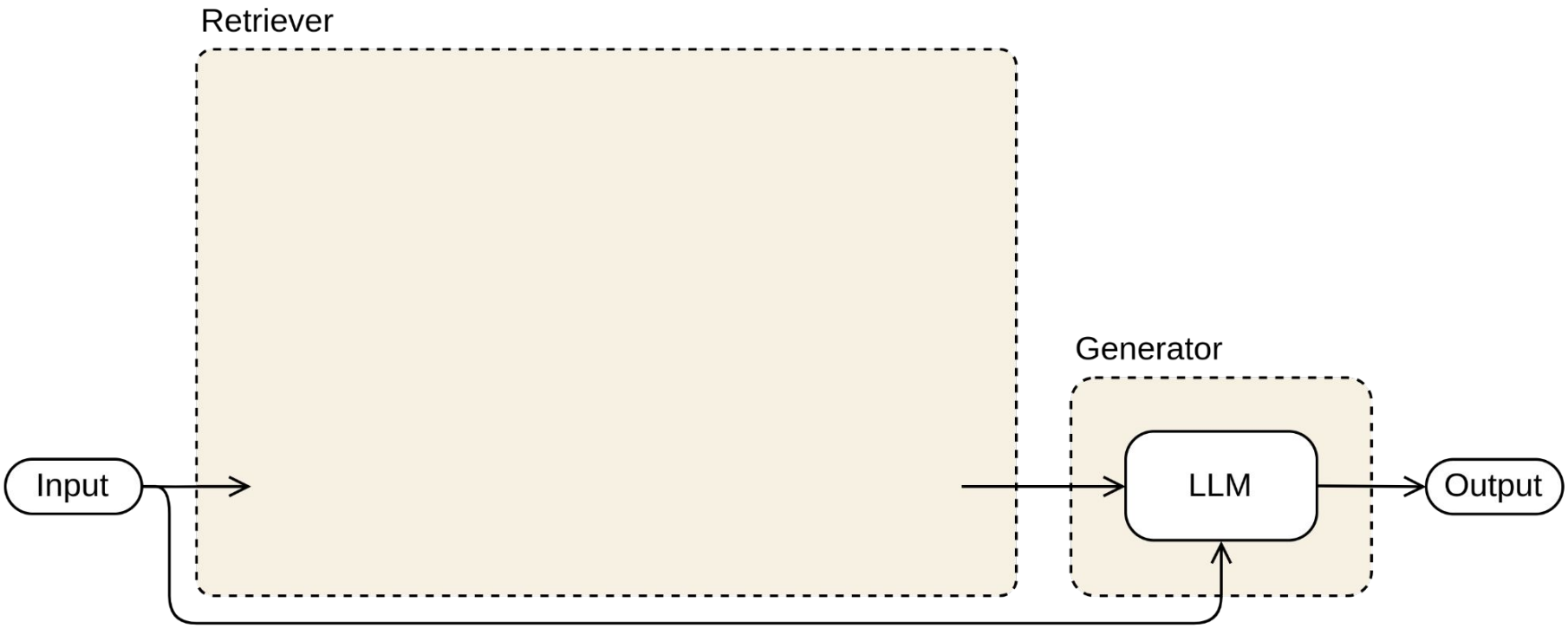
are the language AI “thinks” in.

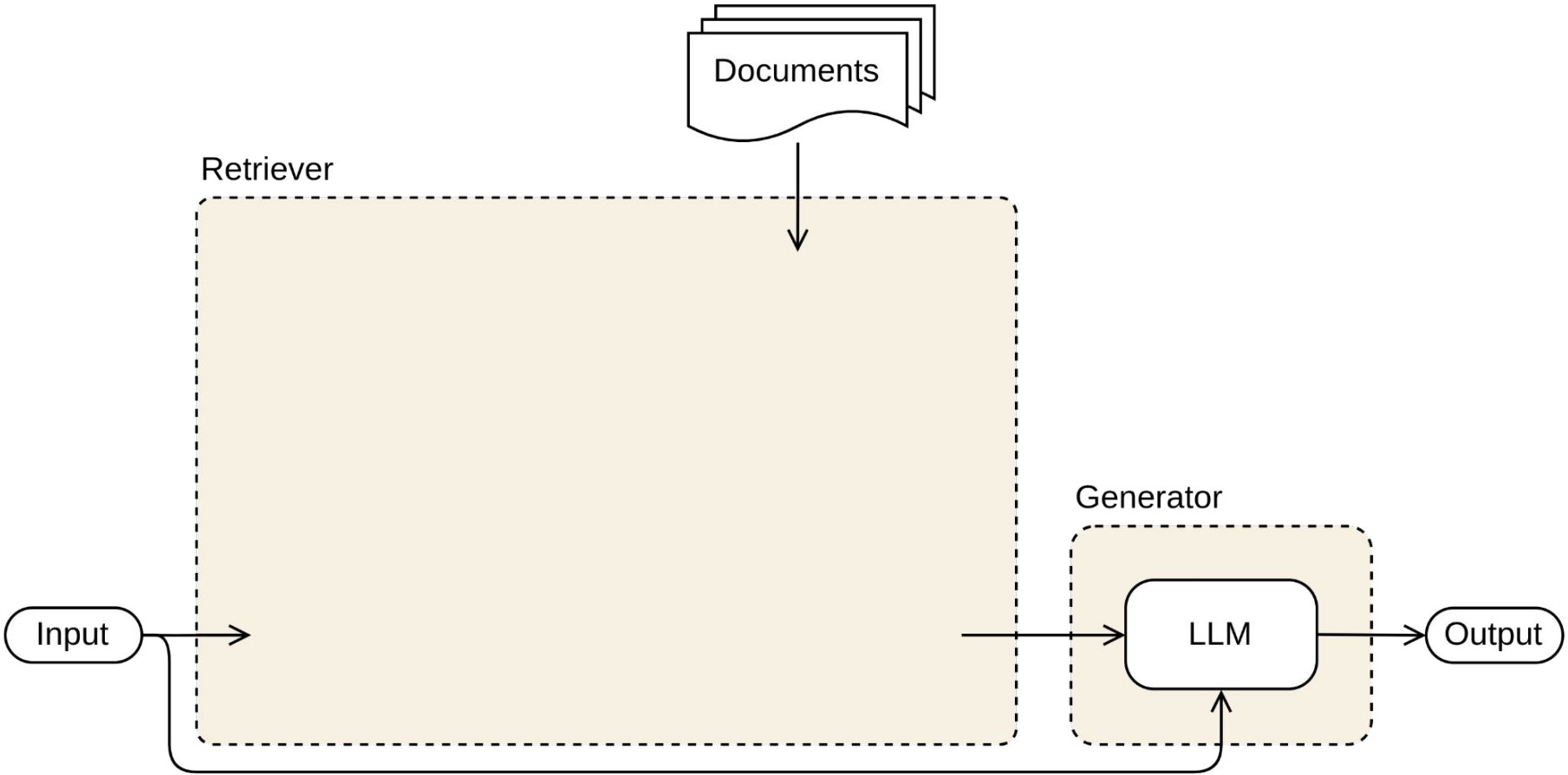
are how language becomes computable.

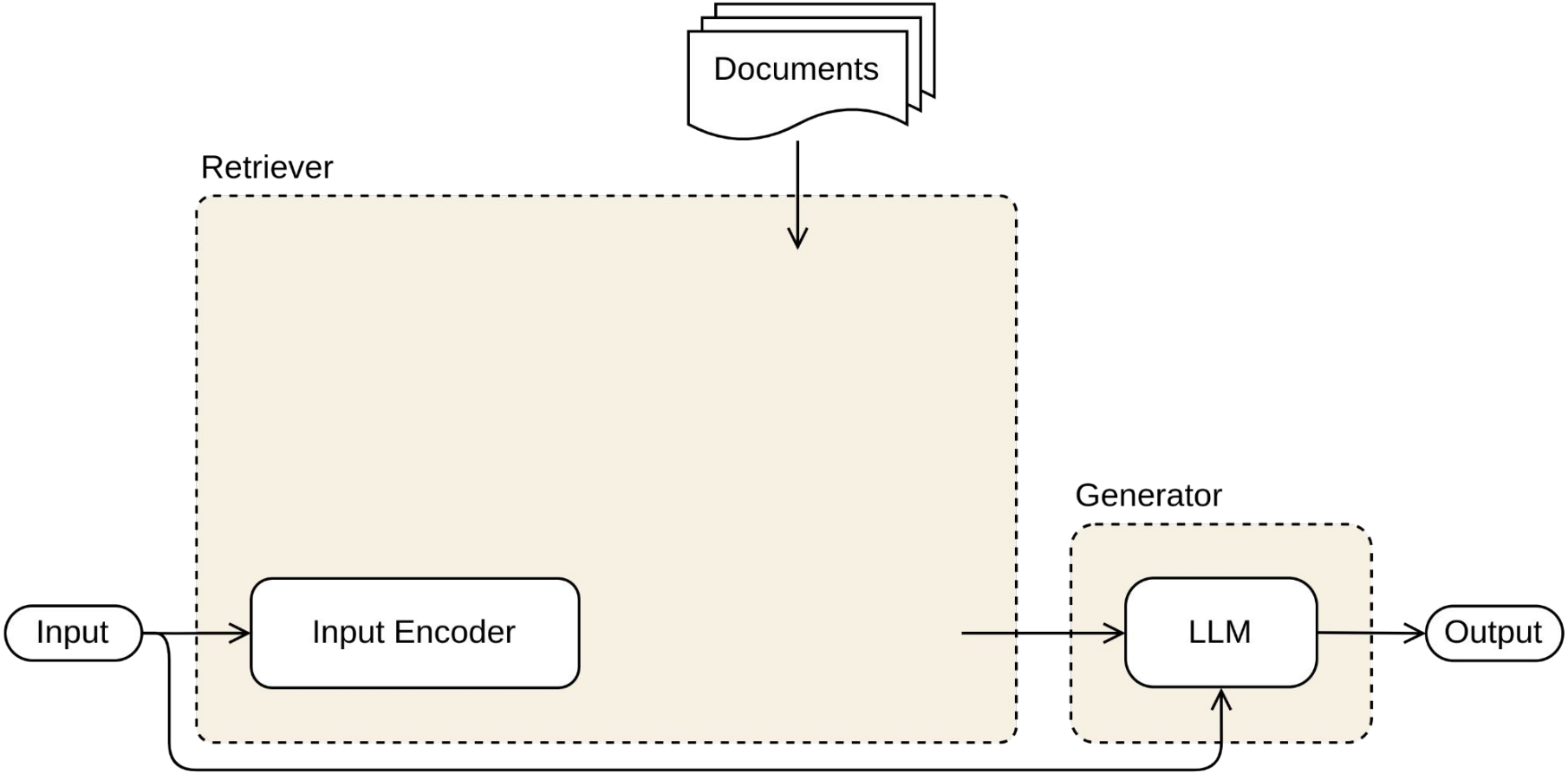
are the quiet engine behind your favourite apps.

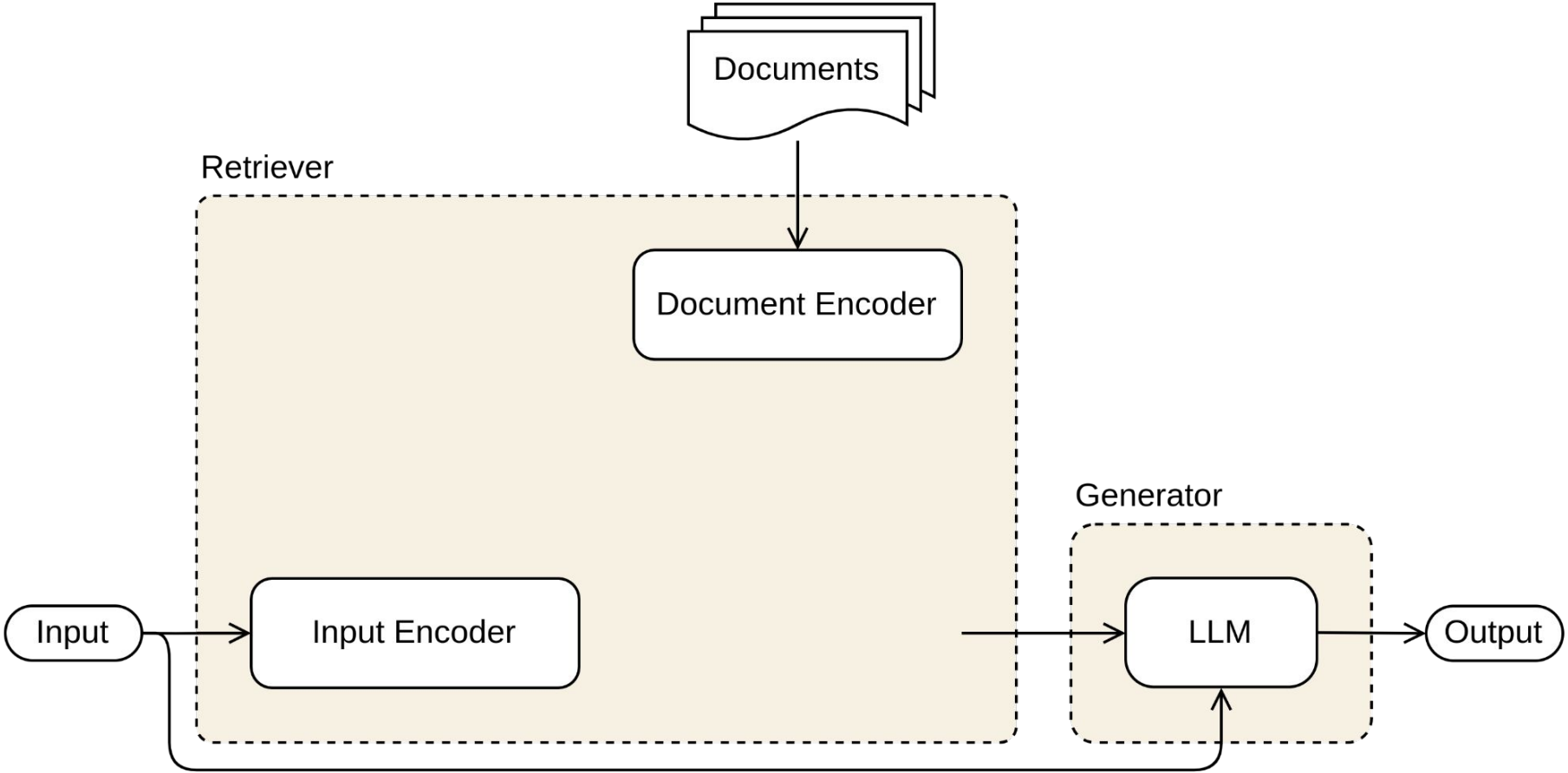

Input
Output

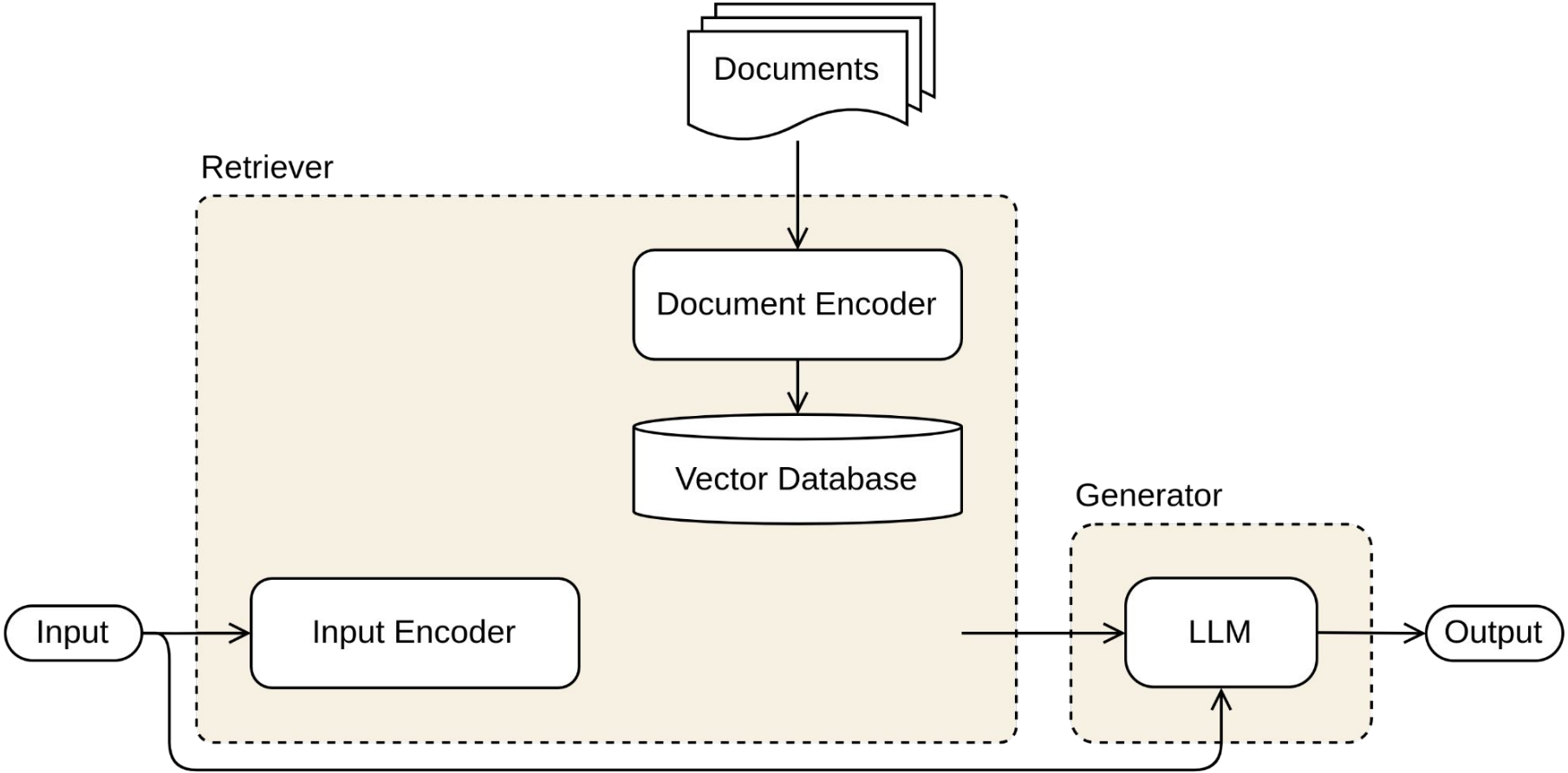

Generator
Input
LLM
Output

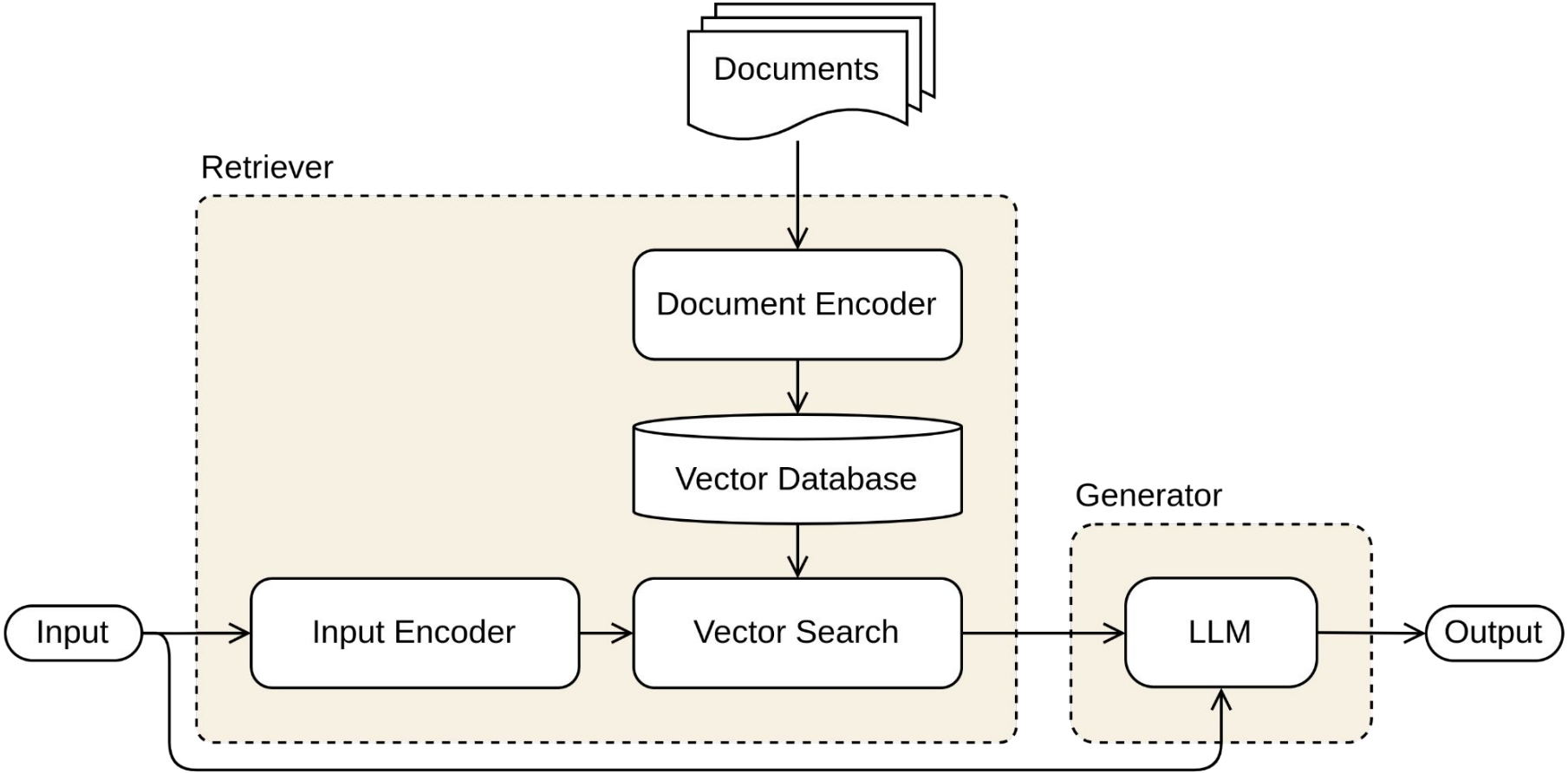

Retriever
Generator
Input
LLM
Output

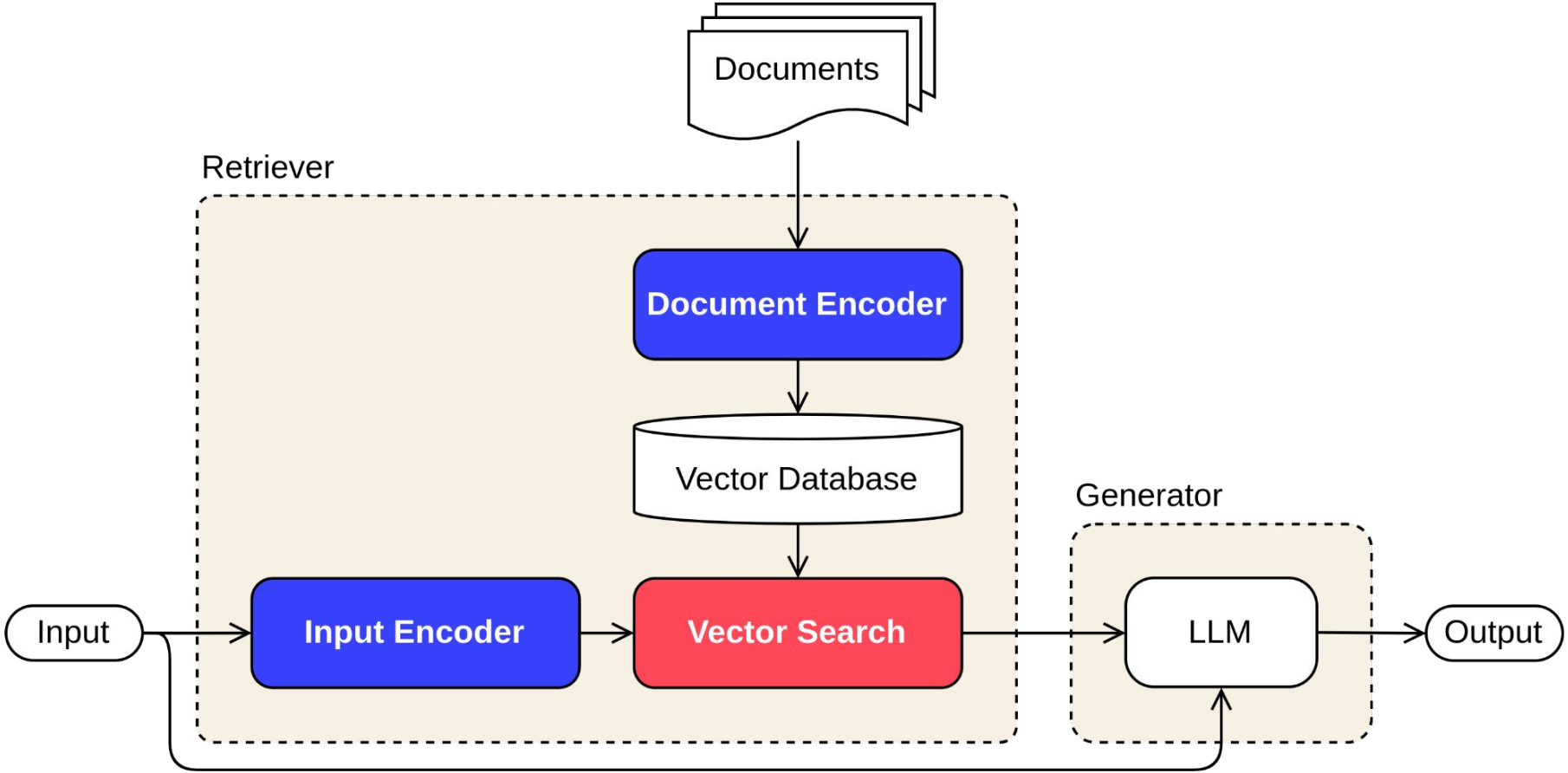

Documents
Retriever
Generator
Input
LLM
Output

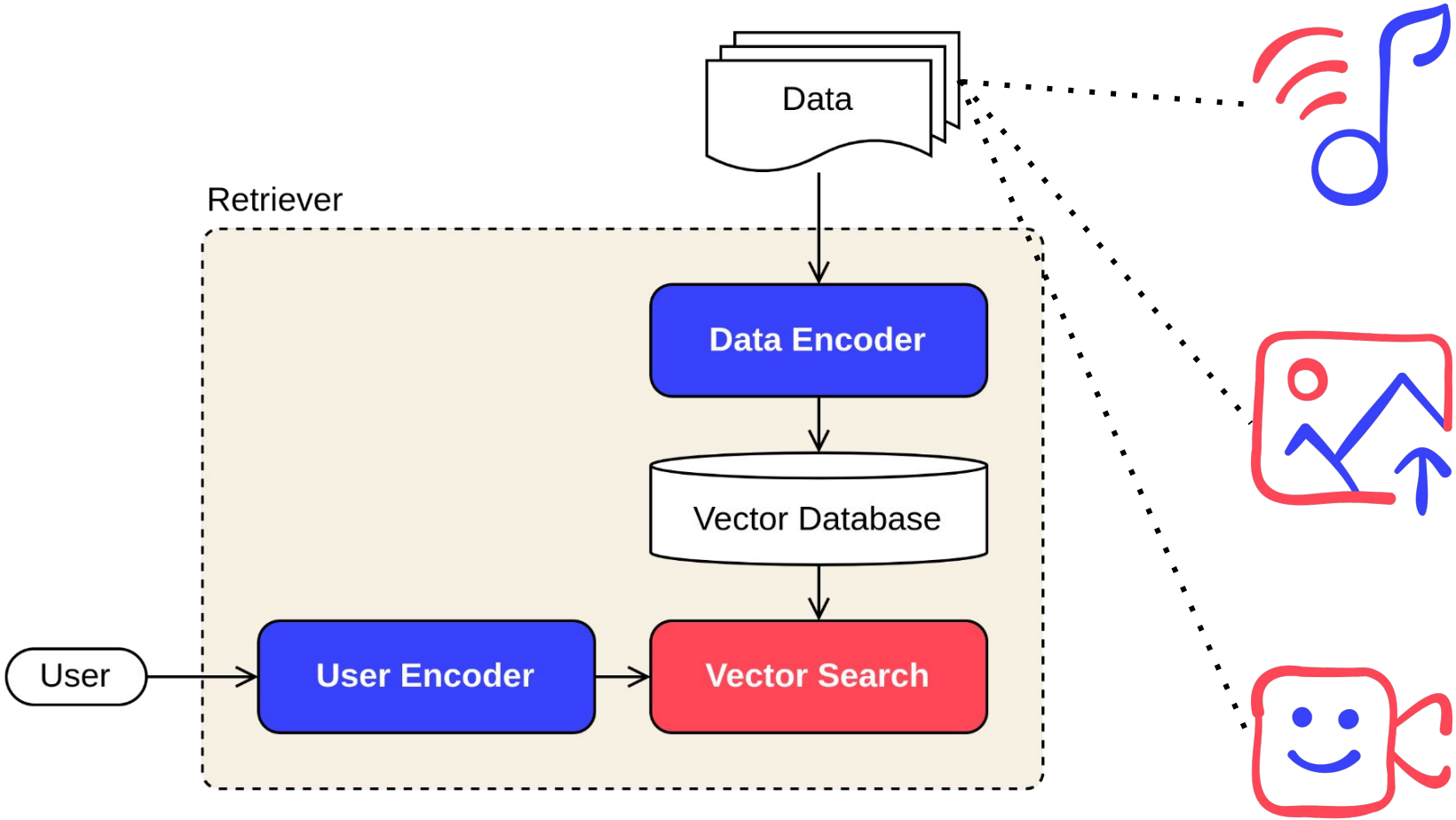

Documents
Retriever
Generator
Input
Input Encoder
LLM
Output

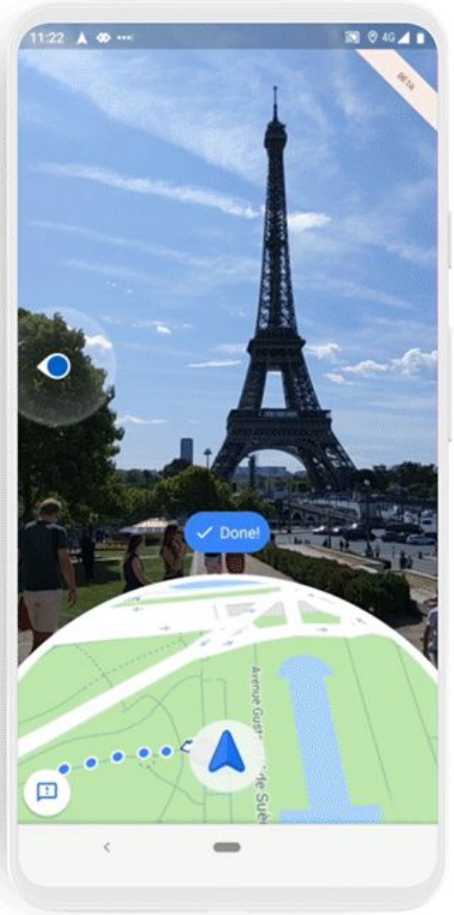

Documents
Retriever
Document Encoder
Generator
Input
Input Encoder
LLM
Output


Documents
Retriever
Document Encoder
Vector Database
Generator
Input
Input Encoder
LLM
Output


Documents
Retriever
Document Encoder
Vector Database
Generator
Input
Input Encoder
Vector Search
LLM
Output


Documents
Retriever
Document Encoder
Vector Database
Generator
Input
Input Encoder
Vector Search
LLM
Output


Data
Retriever
Data Encoder
Vector Database
User
User Encoder
Vector Search


11:22
Done!

Verified Artist

Linkin Park

49,668,327 monthly listeners

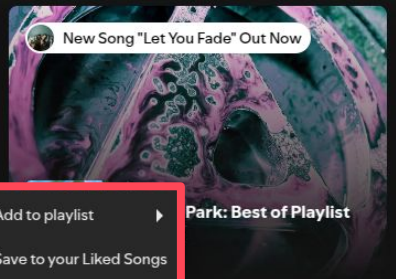
Following

Popular

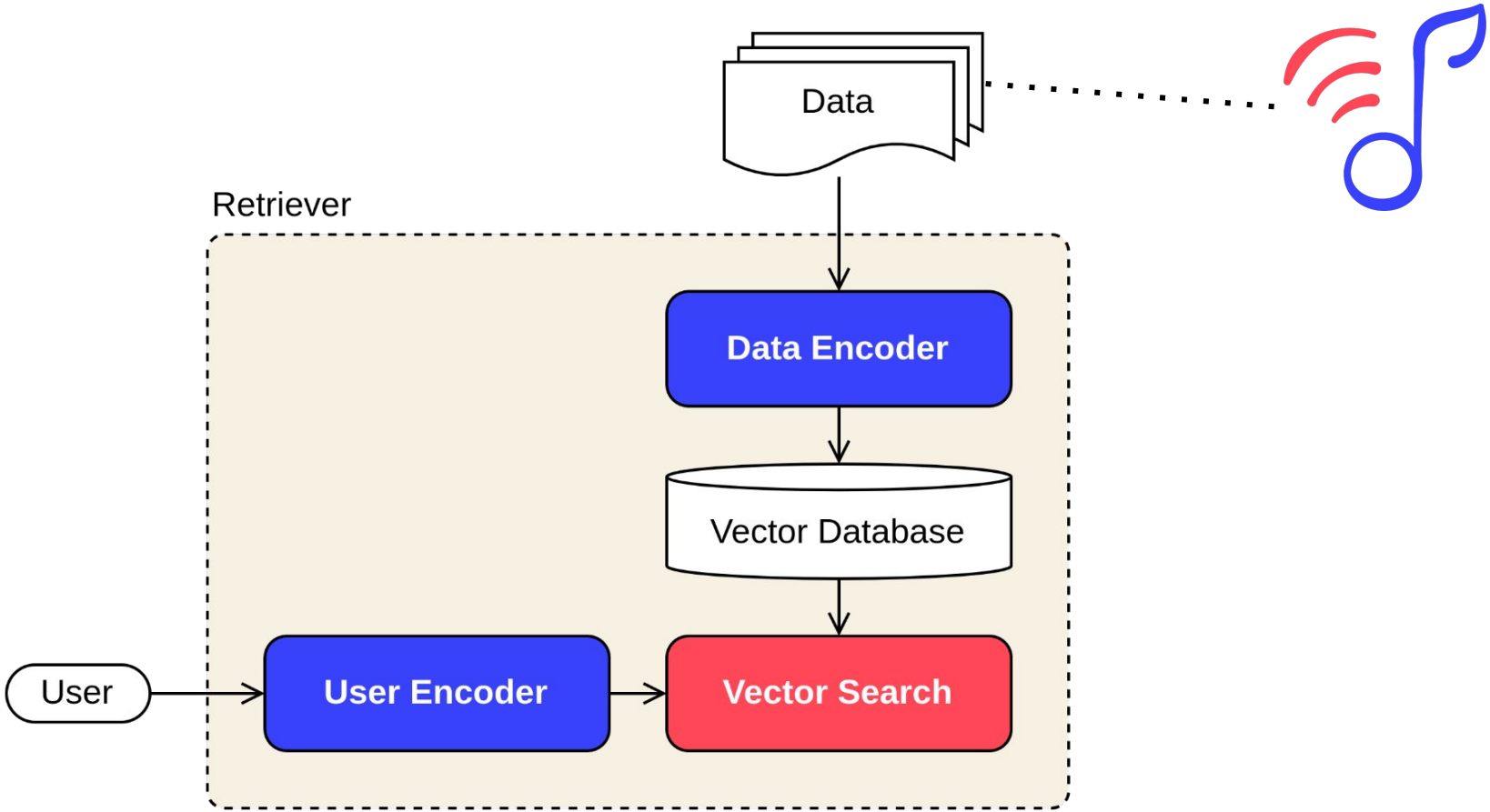
1	In the End Music video	2,568,068,328	3:36	
2	Numb Music video	2,245,600,977	3:07	
3	The Emptiness Machine E Music video	549,895,238	3:10	•••
4	Faint Music video	1,007,938,211	2:42	
5	One Step Closer Music video	959,730,959	2:37	

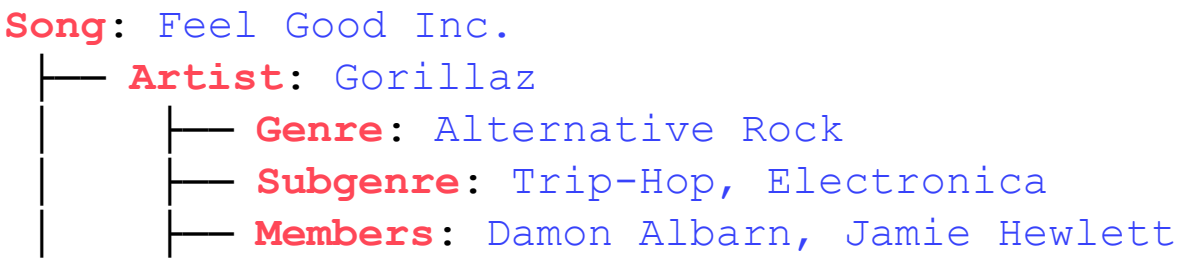
See more

Artist pick



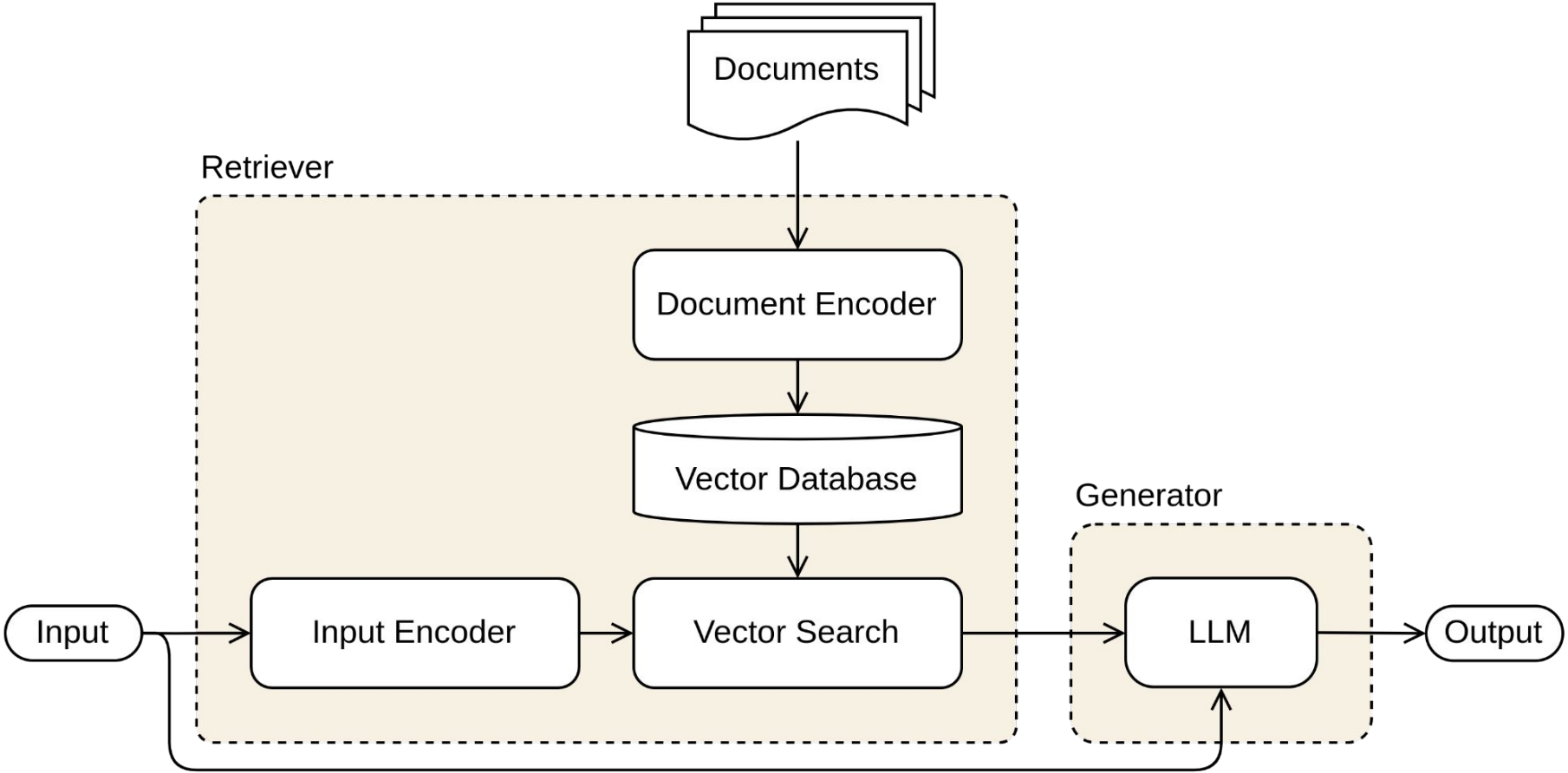
- Add to playlist
- Save to your Liked Songs
- Add to queue
- Go to song radio
- Go to album

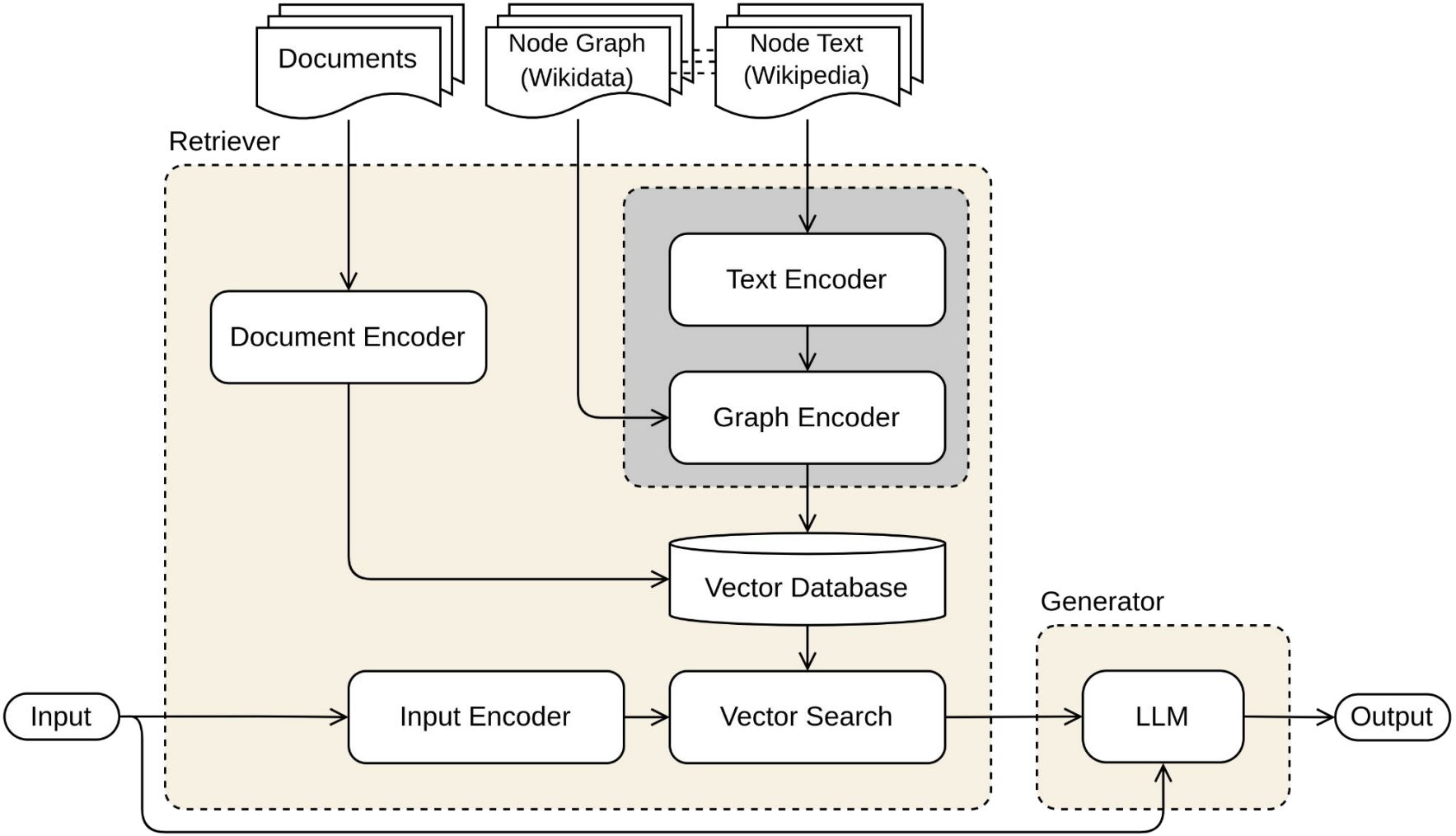

Data
Retriever
Data Encoder
Vector Database
User
User Encoder
Vector Search

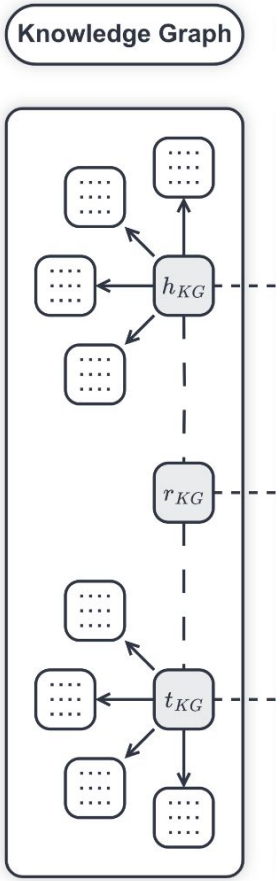

Song: Feel Good Inc.
Artist: Gorillaz
Genre: Alternative Rock
Subgenre: Trip-Hop, Electronica
Members: Damon Albarn, Jamie Hewlett

Song: Feel Good Inc.
- **Artist**: Gorillaz
  - **Genre**: Alternative Rock
  - **Subgenre**: Trip-Hop, Electronica
  - **Members**: Damon Albarn, Jamie Hewlett
- **Featuring**: De La Soul
- **Album**: Demon Days
- **Release Year**: 2005
- **Labels**: Parlophone, Virgin
- **Themes**: Anti-consumerism, Isolation, Media Culture
- **Mood**: Energetic, Dark, Playful
- **Lyrics**: **[embedded text]**


Song: Feel Good Inc.
Artist: Gorillaz
Genre: Alternative Rock
Subgenre: Trip-Hop, Electronica
Members: Damon Albarn, Jamie Hewlett
Featuring: De La Soul
Album: Demon Days
Release Year: 2005
Labels: Parlophone, Virgin
Themes: Anti-consumerism, Isolation, Media Culture
Mood: Energetic, Dark, Playful
Lyrics: [embedded text]
Audio Features: [embedded audio]
BPM: 139
Key: D Minor
Energy: High
Danceability: Medium-High


Documents
Retriever
Document Encoder
Vector Database
Generator
Input
Input Encoder
Vector Search
LLM
Output

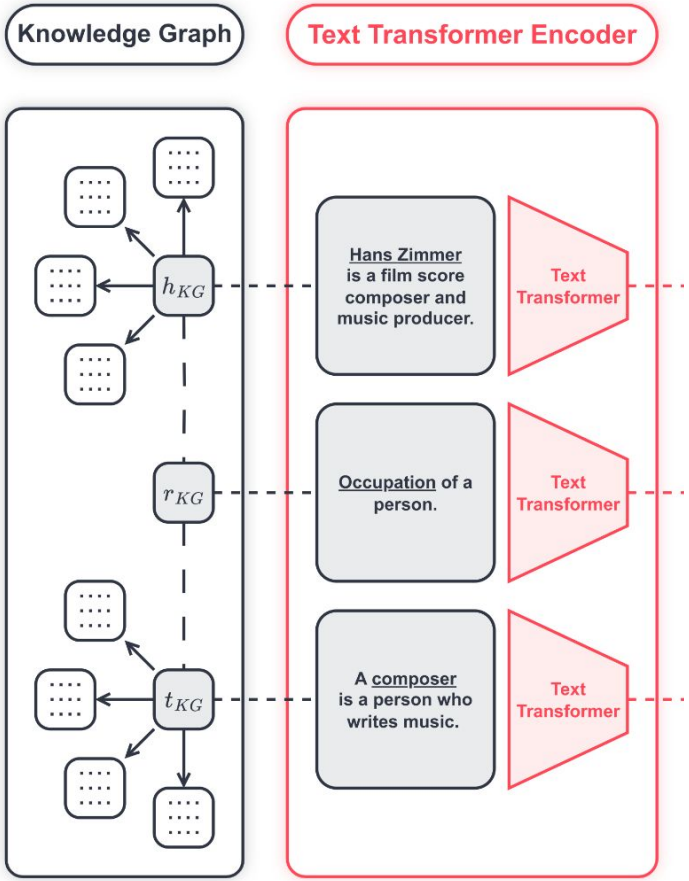

Documents
Node Graph (Wikidata)
Node Text (Wikipedia)
Retriever
Text Encoder
Document Encoder
Graph Encoder
Vector Database
Generator
Input
Input Encoder
Vector Search
LLM
Output



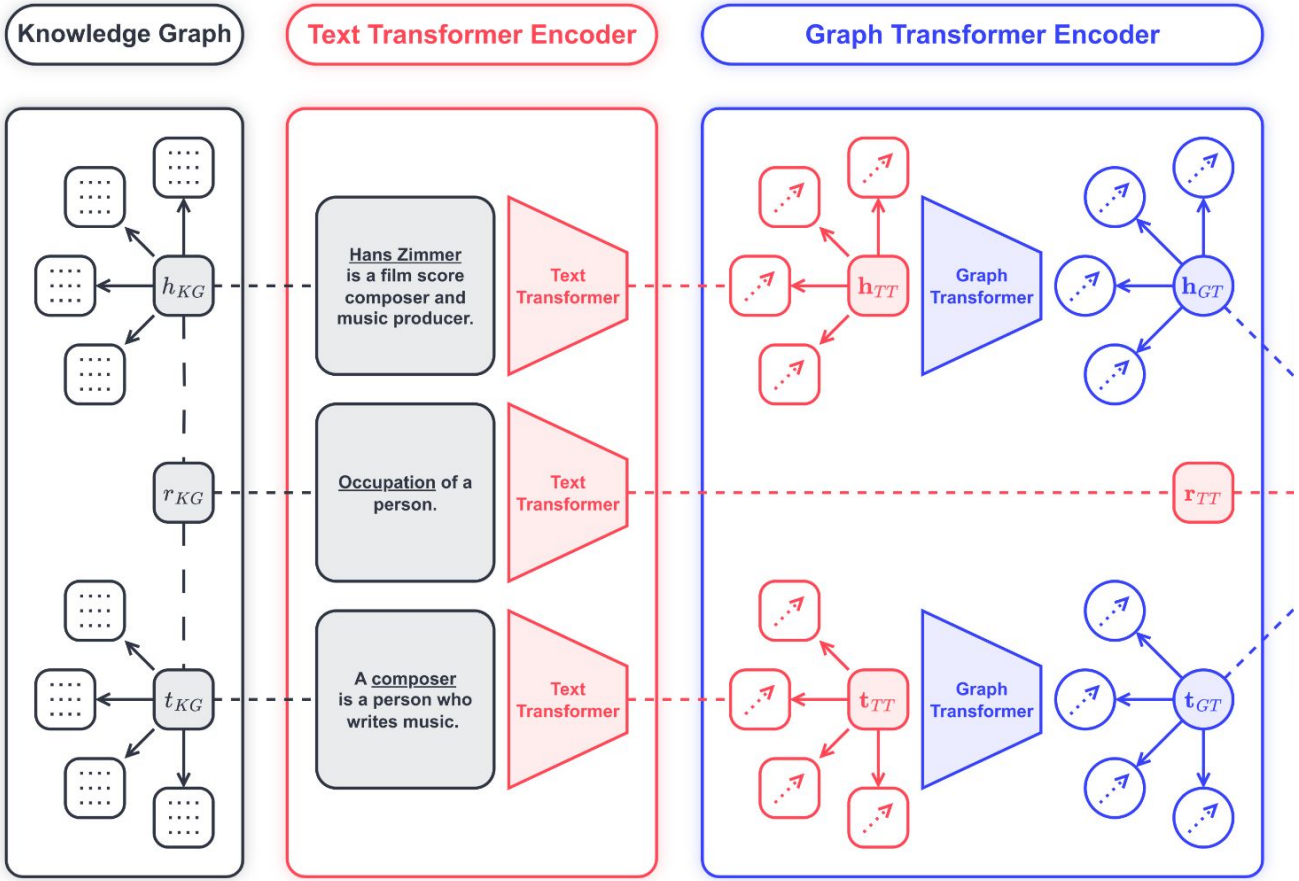
Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.

does magic for

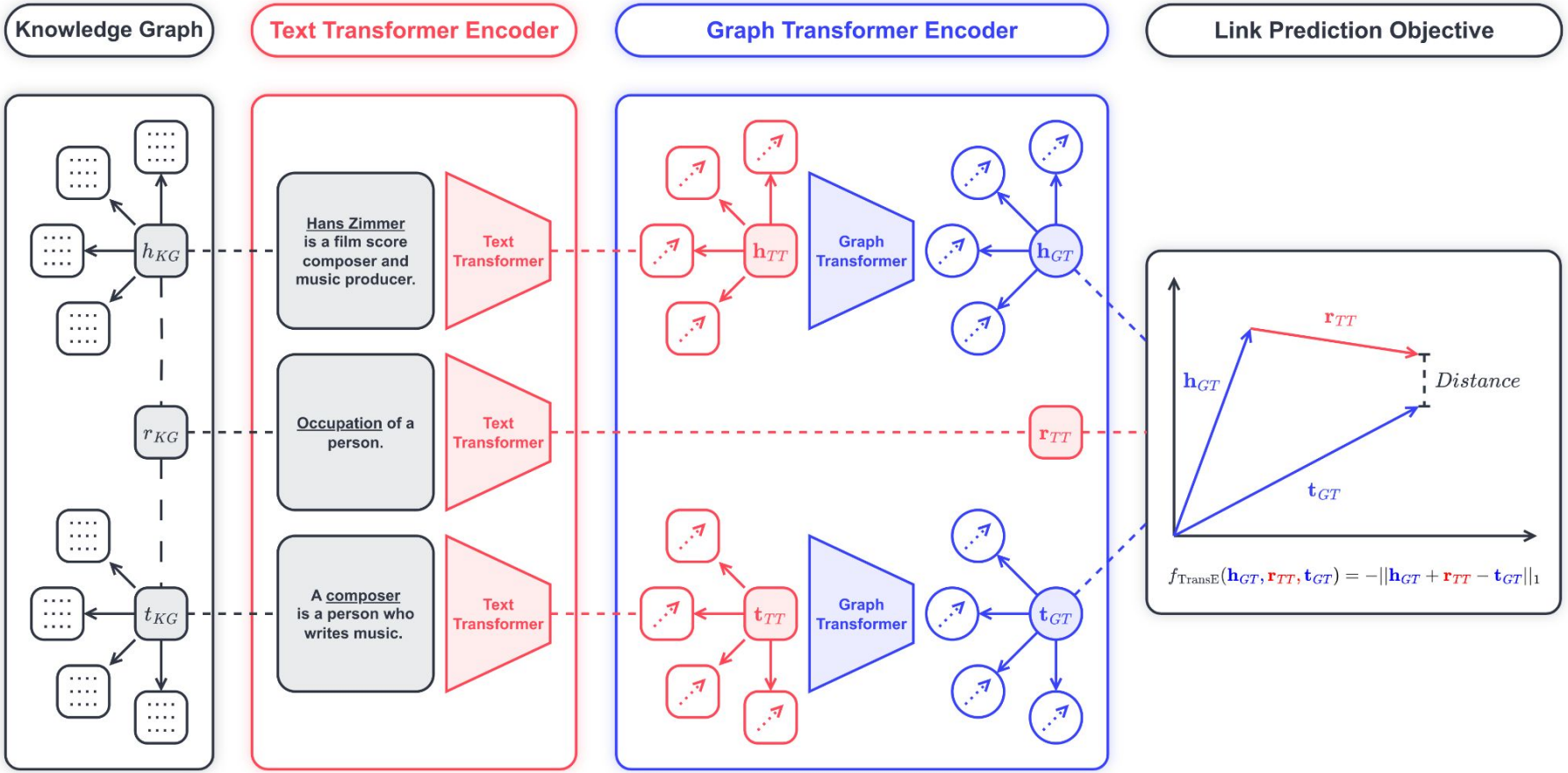
head relation tail



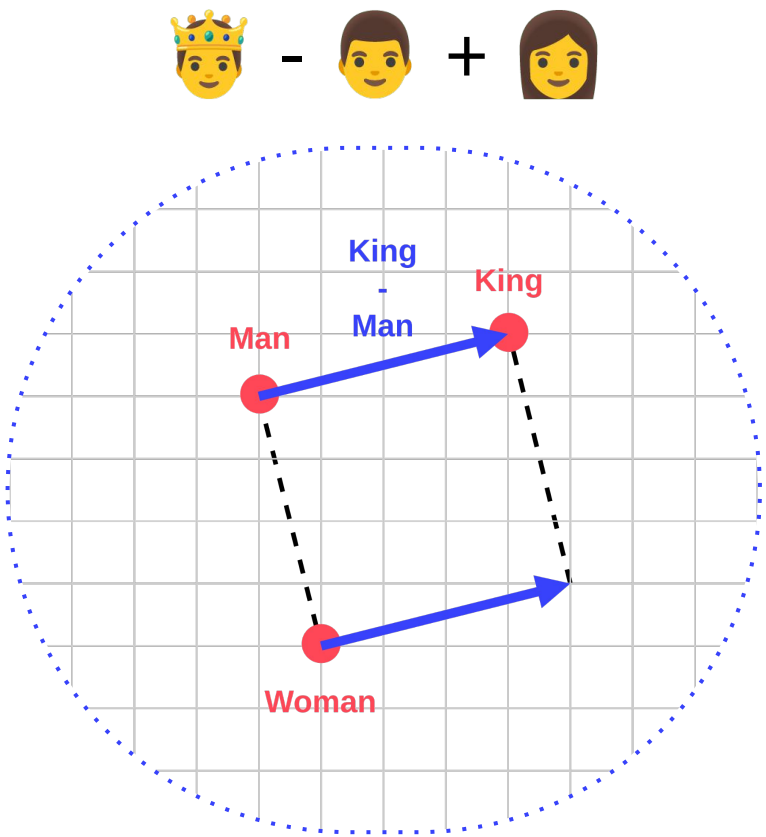
Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.

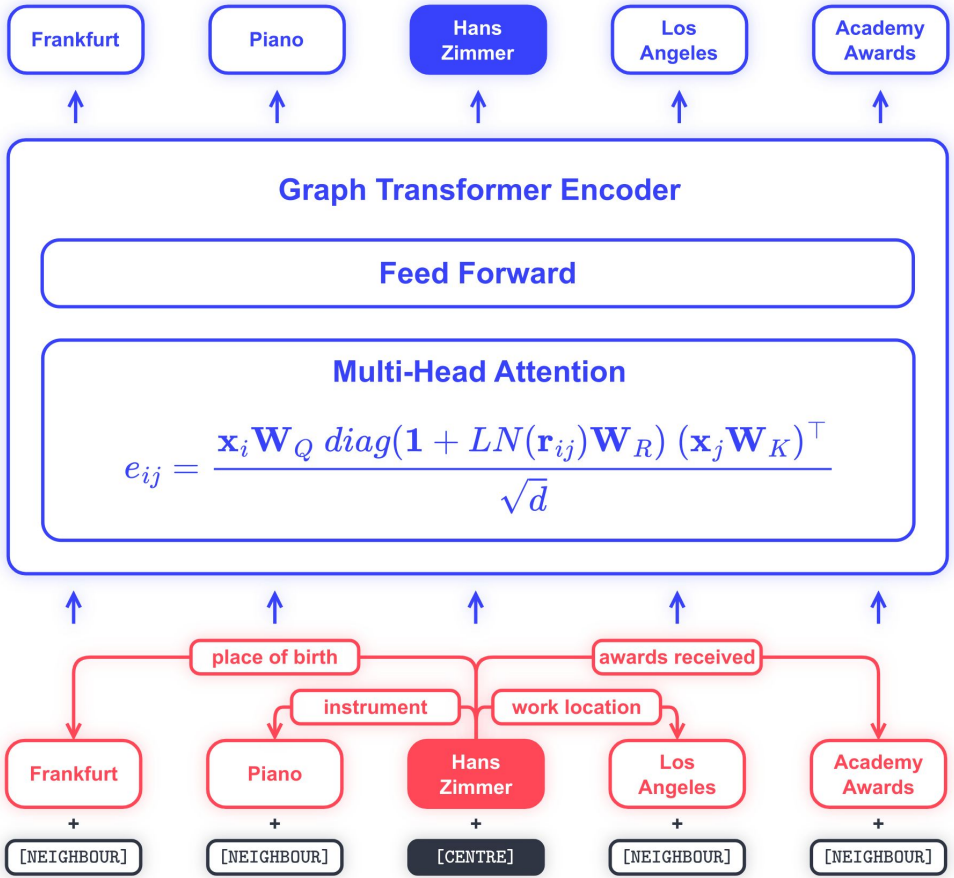


Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.



Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.


King
-
Man
King
Man
Woman



Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.

Model	WN18RR$_{IND}$ MRR	H@1	H@3	H@10	FB15k-237$_{IND}$ MRR	H@1	H@3	H@10
Text-Only Models								
DKRL$^{*}_{BERT_{BASE}}$	0.139	0.048	0.169	0.320	0.144	0.084	0.151	0.263
BOW$^{*}_{BERT_{BASE}}$	0.180	0.045	0.244	0.450	0.173	0.103	0.184	0.316
BLP$^{*}_{BERT_{BASE}}$	0.285	0.135	0.361	0.580	0.195	0.113	0.213	0.363
FnF-T$_{BERT_{BASE}}$	**0.373**	**0.238**	**0.442**	**0.647**	**0.266**	**0.174**	**0.297**	**0.453**

Table 1: WN18RR$_{IND}$ and FB15k-237$_{IND}$ test set results. *Daza et al. (2021); *Markowitz et al. (2022).

Model	WN18RR$_{IND}$ MRR	H@1	H@3	H@10	FB15k-237$_{IND}$ MRR	H@1	H@3	H@10
Text-Only Models								
DKRL$^{*}_{BERT_{BASE}}$	0.139	0.048	0.169	0.320	0.144	0.084	0.151	0.263
BOW$^{*}_{BERT_{BASE}}$	0.180	0.045	0.244	0.450	0.173	0.103	0.184	0.316
BLP$^{*}_{BERT_{BASE}}$	0.285	0.135	0.361	0.580	0.195	0.113	0.213	0.363
FnF-T$_{BERT_{BASE}}$	**0.373**	**0.238**	**0.442**	**0.647**	**0.266**	**0.174**	**0.297**	**0.453**
FnF-T$_{BERT_{MEDIUM}}$	0.342	0.213	0.405	0.603	0.253	0.164	0.281	0.431
FnF-T$_{BERT_{SMALL}}$	0.320	0.197	0.379	0.572	0.239	0.152	0.265	0.411
FnF-T$_{BERT_{MINI}}$	0.268	0.156	0.318	0.498	0.204	0.128	0.223	0.354
FnF-T$_{BERT_{TINY}}$	0.193	0.098	0.230	0.385	0.164	0.100	0.176	0.289

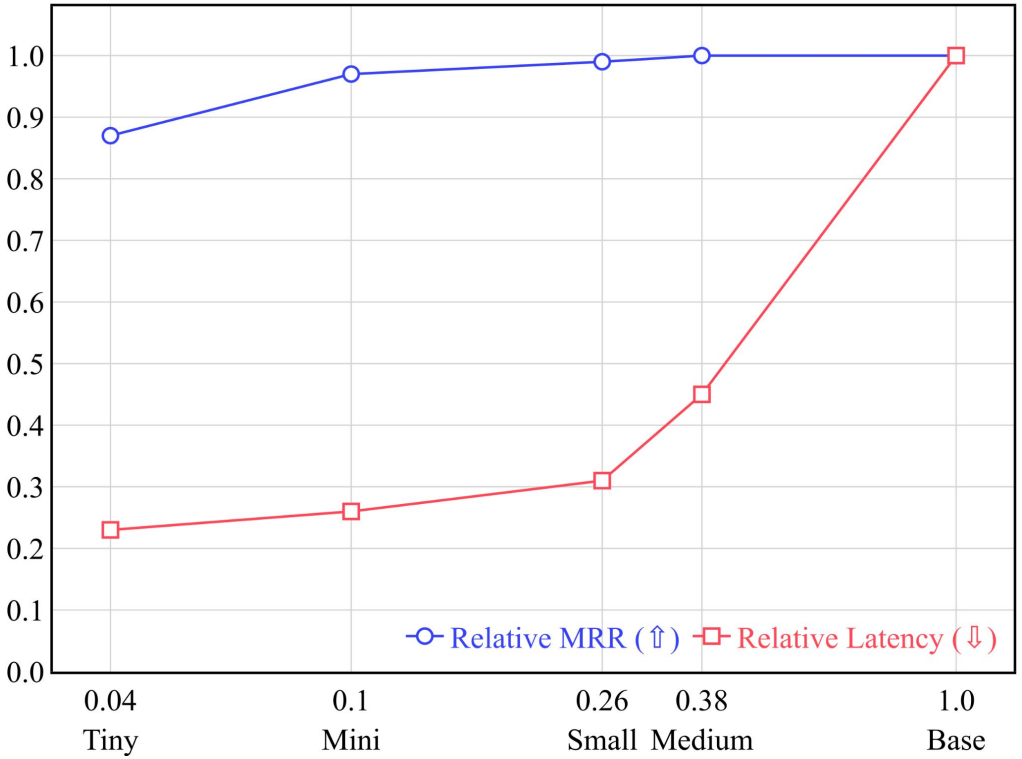
Table 1: WN18RR$_{IND}$ and FB15k-237$_{IND}$ test set results. *Daza et al. (2021); *Markowitz et al. (2022).

Model	WN18RR$_{IND}$ MRR	H@1	H@3	H@10	FB15k-237$_{IND}$ MRR	H@1	H@3	H@10
Text-Only Models								
DKRL$^{*}_{BERT_{BASE}}$	0.139	0.048	0.169	0.320	0.144	0.084	0.151	0.263
BOW$^{*}_{BERT_{BASE}}$	0.180	0.045	0.244	0.450	0.173	0.103	0.184	0.316
BLP$^{*}_{BERT_{BASE}}$	0.285	0.135	0.361	0.580	0.195	0.113	0.213	0.363
FnF-T$_{BERT_{BASE}}$	**0.373**	**0.238**	**0.442**	**0.647**	**0.266**	**0.174**	**0.297**	**0.453**
FnF-T$_{BERT_{MEDIUM}}$	0.342	0.213	0.405	0.603	0.253	0.164	0.281	0.431
FnF-T$_{BERT_{SMALL}}$	0.320	0.197	0.379	0.572	0.239	0.152	0.265	0.411
FnF-T$_{BERT_{MINI}}$	0.268	0.156	0.318	0.498	0.204	0.128	0.223	0.354
FnF-T$_{BERT_{TINY}}$	0.193	0.098	0.230	0.385	0.164	0.100	0.176	0.289
Structure-Informed Models								
StAR$^{\star}_{BERT_{BASE}}$	0.321	0.192	0.381	0.576	0.163	0.092	0.176	0.309
StATIK$^{\star}_{BERT_{BASE}}$	0.516	0.425	0.558	0.690	0.224	0.143	0.248	0.381
FnF-TG$_{BERT_{BASE}}$	0.732	0.652	0.785	**0.875**	0.316	0.214	0.350	**0.524**

Table 1: WN18RR$_{IND}$ and FB15k-237$_{IND}$ test set results. *Daza et al. (2021); ⋆Markowitz et al. (2022).

Model	WN18RR$_{IND}$ MRR	H@1	H@3	H@10	FB15k-237$_{IND}$ MRR	H@1	H@3	H@10
Text-Only Models								
DKRL$^{*}_{BERT_{BASE}}$	0.139	0.048	0.169	0.320	0.144	0.084	0.151	0.263
BOW$^{*}_{BERT_{BASE}}$	0.180	0.045	0.244	0.450	0.173	0.103	0.184	0.316
BLP$^{*}_{BERT_{BASE}}$	0.285	0.135	0.361	0.580	0.195	0.113	0.213	0.363
FnF-T$_{BERT_{BASE}}$	**0.373**	**0.238**	**0.442**	**0.647**	**0.266**	**0.174**	**0.297**	**0.453**
FnF-T$_{BERT_{MEDIUM}}$	0.342	0.213	0.405	0.603	0.253	0.164	0.281	0.431
FnF-T$_{BERT_{SMALL}}$	0.320	0.197	0.379	0.572	0.239	0.152	0.265	0.411
FnF-T$_{BERT_{MINI}}$	0.268	0.156	0.318	0.498	0.204	0.128	0.223	0.354
FnF-T$_{BERT_{TINY}}$	0.193	0.098	0.230	0.385	0.164	0.100	0.176	0.289
Structure-Informed Models								
StAR$^{\star}_{BERT_{BASE}}$	0.321	0.192	0.381	0.576	0.163	0.092	0.176	0.309
StATIK$^{\star}_{BERT_{BASE}}$	0.516	0.425	0.558	0.690	0.224	0.143	0.248	0.381
FnF-TG$_{BERT_{BASE}}$	0.732	0.652	0.785	**0.875**	0.316	0.214	0.350	**0.524**
FnF-TG$_{BERT_{MEDIUM}}$	**0.737**	**0.661**	**0.789**	0.873	0.314	0.214	0.353	0.515
FnF-TG$_{BERT_{SMALL}}$	0.727	0.648	0.781	0.867	**0.316**	**0.216**	**0.354**	0.518
FnF-TG$_{BERT_{MINI}}$	0.713	0.632	0.768	0.857	0.302	0.204	0.337	0.502
FnF-TG$_{BERT_{TINY}}$	0.638	0.543	0.700	0.808	0.288	0.195	0.318	0.475

Table 1: WN18RR$_{IND}$ and FB15k-237$_{IND}$ test set results. *Daza et al. (2021); ⋆Markowitz et al. (2022).



Coman, Andrei C., et al. "Fast-and-Frugal Text-Graph Transformers are Effective Link Predictors." To be presented at ACL 2025.

Input

Breakout is an arcade game developed and published by Atari , Inc. , released on May 13 , 1976 . It was conceptualized by Nolan Bushnell and Steve Bristow , influenced by the 1972 Atari arcade game Pong , and built by Steve Wozniak aided by Steve Jobs . Breakout was the basis and inspiration for certain aspects of the Apple II personal computer . [...]

Input

Breakout is an arcade game developed and published by **Atari , Inc.** , released on **May 13 , 1976** . It was conceptualized by **Nolan Bushnell** and **Steve Bristow** , influenced by the 1972 **Atari** arcade game **Pong** , and built by **Steve Wozniak** aided by **Steve Jobs** . **Breakout** was the basis and inspiration for certain aspects of the **Apple II** personal computer . [...]

Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024

Input

Breakout is an arcade game developed and published by Atari , Inc. , released on May 13 , 1976 . It was conceptualized by Nolan Bushnell and Steve Bristow , influenced by the 1972 Atari arcade game Pong , and built by Steve Wozniak aided by Steve Jobs . Breakout was the basis and inspiration for certain aspects of the Apple II personal computer . [...]

Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024

Output

Subject: Breakout **Object**: Atari

Relation: developer

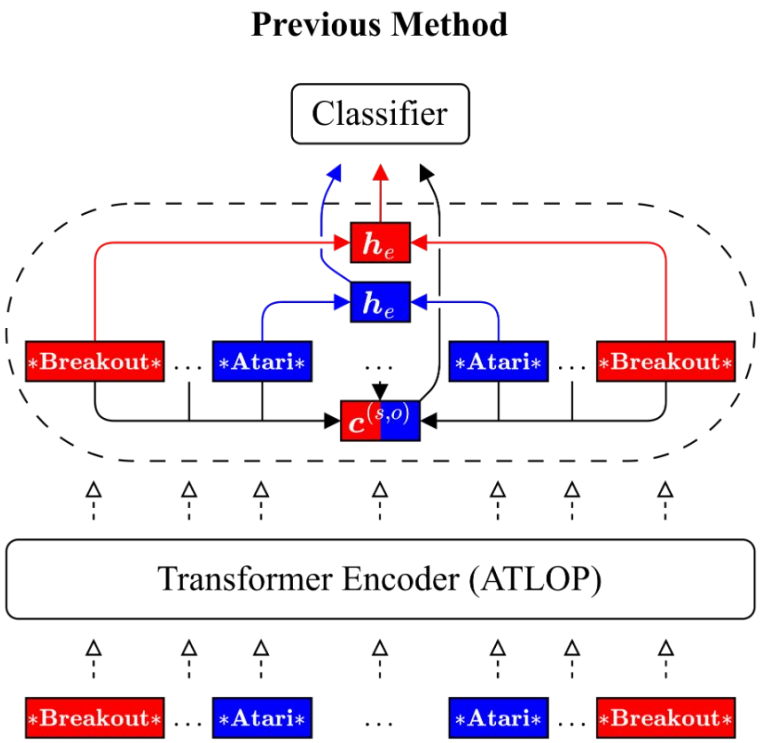
Subject: Breakout **Object**: Atari

Relation: publisher

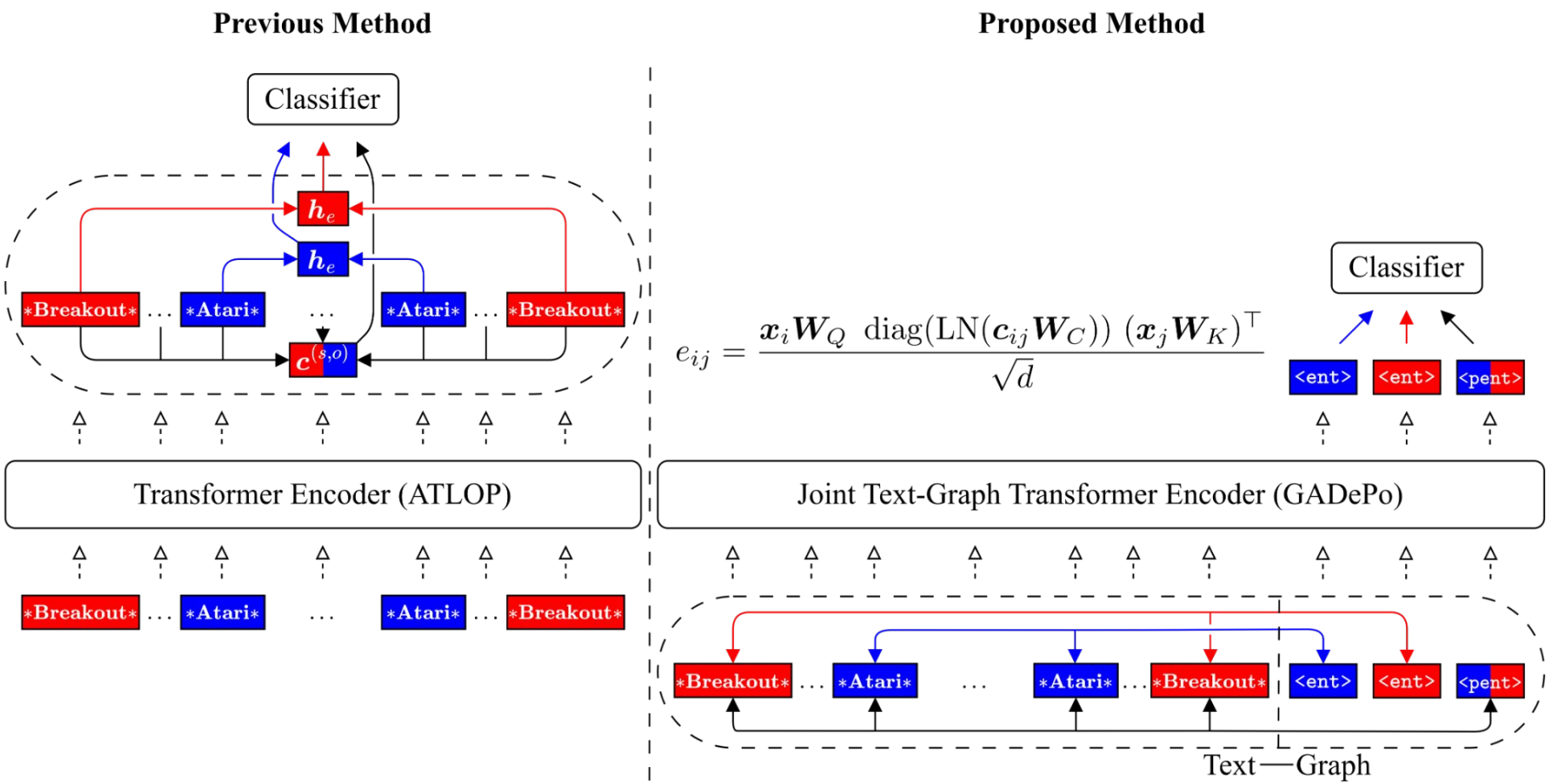
Input

Breakout is an arcade game developed and published by Atari , Inc. , released on May 13 , 1976 . It was conceptualized by Nolan Bushnell and Steve Bristow , influenced by the 1972 Atari arcade game Pong , and built by Steve Wozniak aided by Steve Jobs . Breakout was the basis and inspiration for certain aspects of the Apple II personal computer . [...]

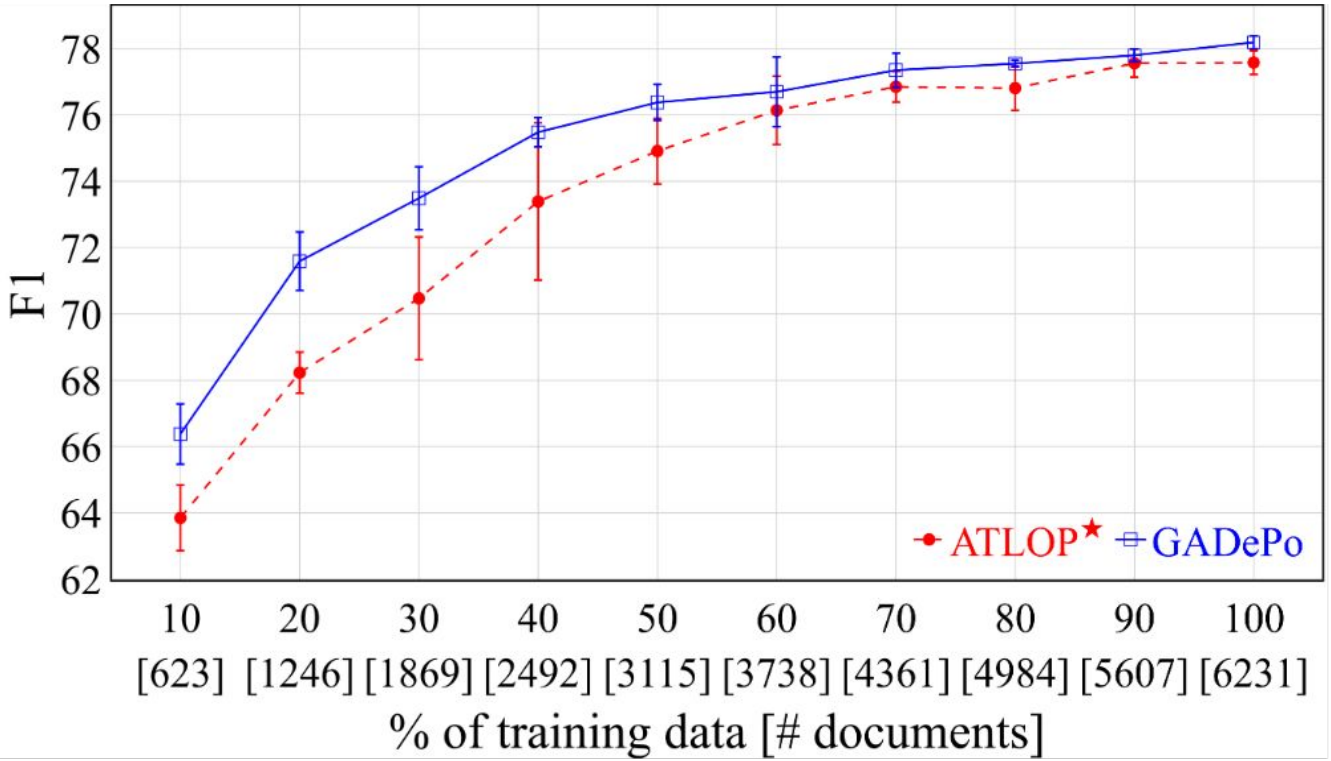
Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024



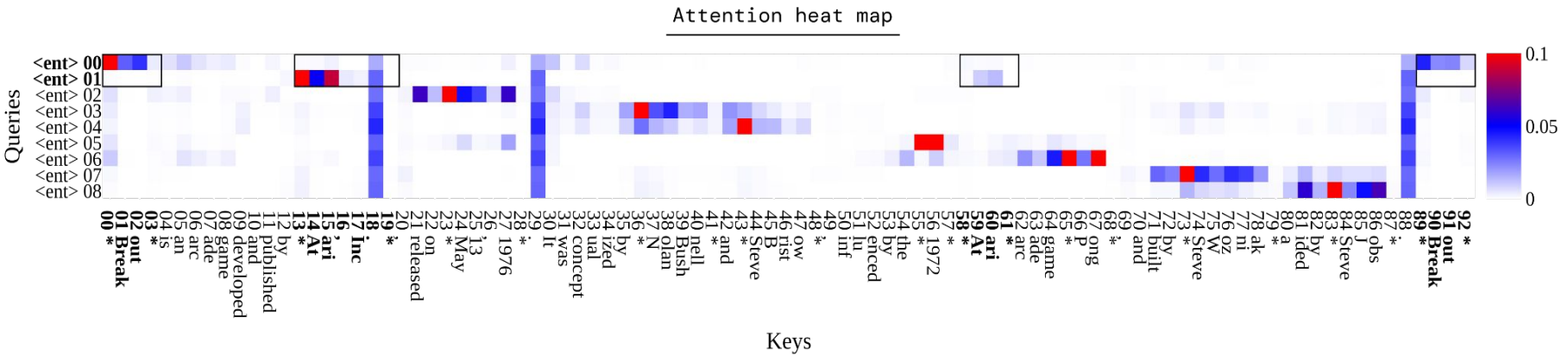
Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024



Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024



Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024



Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024

Output

Subject: Breakout **Object**: Atari

Relation: developer

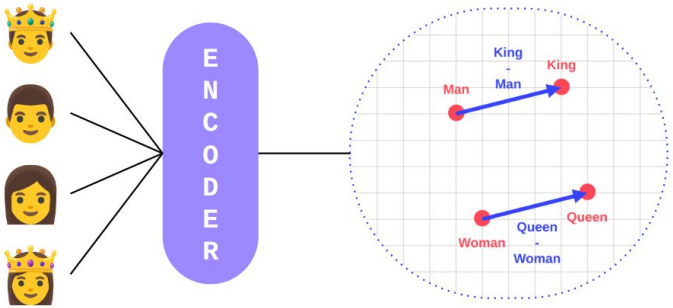
Subject: Breakout **Object**: Atari

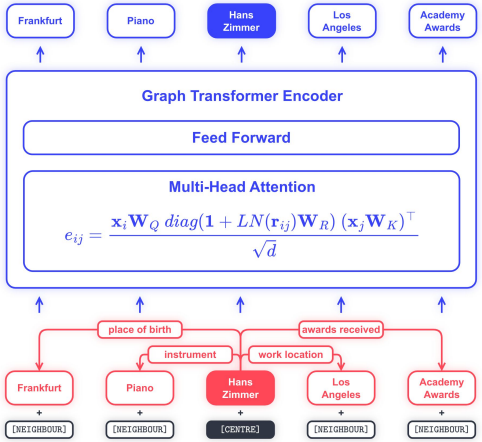
Relation: publisher

Input

Breakout is an arcade game developed and published by Atari , Inc. , released on May 13 , 1976 . It was conceptualized by Nolan Bushnell and Steve Bristow , influenced by the 1972 Atari arcade game Pong , and built by Steve Wozniak aided by Steve Jobs . Breakout was the basis and inspiration for certain aspects of the Apple II personal computer . [...]

Coman, Andrei C., et al. "GADePo: Graph-assisted declarative pooling transformers for document-level relation extraction." KnowledgeNLP 2024





Proposed Method