Verbum ex Machina

An Objective Matter of Faith

Marco Zanella
University of Padova, University of Bologna

Can Machines read the Sacred?

Can an algorithm truly **read and understand** sacred texts,
or does it **merely retrieve and manipulate** textual data?

The **nature of interpretation**—is it **purely structural**, or does it require
something more, **something inherently human**?

Machines **process words**; humans **seek meaning**.
If an algorithm finds a biblical reference, is it **reading**—or simply **computing**?

The question of whether a computer can think is no more interesting than
the question of whether a submarine can swim.

Edsger W. Dijsktra

The Importance of Intertextuality

No text exists in **isolation**: every text is part of a **larger conversation**.

No text exists in isolation
Every text builds upon past words, ideas, and interpretations.

Sacred and historical texts evolve
They are not static but actively shape and reshape meaning across time.

Intertextuality fuels understanding
Recognizing these connections deepens interpretation and allows for new insights into ancient wisdom.

Sacred and historical texts constantly **reference**, **reinterpret**, and **reshape** each other.

Any text is a new tissue of past citations.

Roland Barthes, "The death of the Author"

Tracing Connections

Before AI, there was the **scholar's eye**.

method	description	limitations
close reading	scholars manually analyze text for similarities	Time-consuming, subjective
citation indexing	Tracking explicit references across texts	Doesn't capture implicit allusions
parellel text comparison	Placing passages side by side for analysis	Requires extensive memorization
commentary tradition	Relying on historical interpretations	Biases from past readings

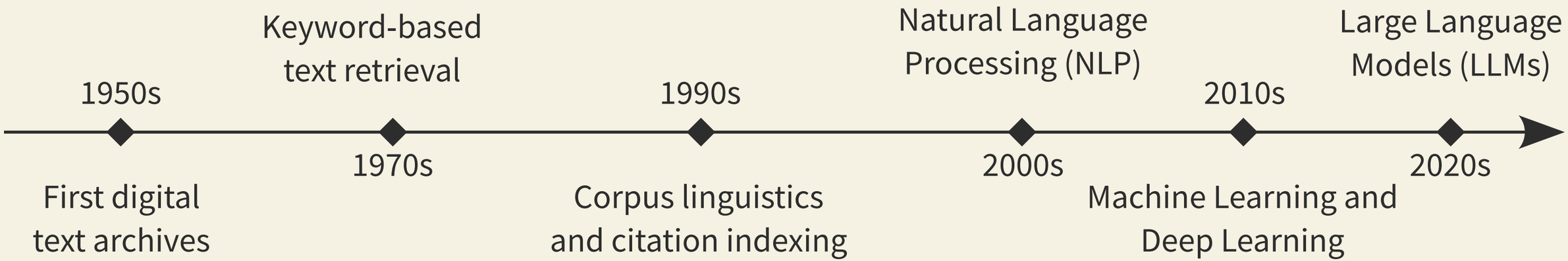
For centuries, intertextuality was uncovered through **deep reading**, **memory**, and meticulous **comparison**.

A text is made of multiple writings, drawn from many cultures and entering into mutual relations of dialogue, parody, contestation.

Roland Barthes, "The death of the Author"

The Rise of Computational Intertextuality

What took **months**, now takes **minutes**.



Computational tools changed the way we **trace relationships**, from **string matching** to **AI**-powered insights.

We are not far from the day when texts will be read by machines and not by men.

Roberto Busa

Different Levels of Computational Text Analysis

From simple strings to **deep semantics**: how computers "read" texts.

more abstract
non-explainable

Machine Learning
conceptual relationships, paraphrased references, hidden patterns

Language-Aware Models
synonyms, grammatical variations, partial matches

Lexical Matching
string-based methods, keyword matching

closer to the text
transparent

slow
resource-intensive

fast
lightweight

Each method has **strengths**, **weaknesses**, and its own way of "understanding" text.

The problems of language are profound, but not beyond our grasp.

Noam Chomsky

Lexical Matching: Words Speak for Themselves

Exact words, exact matches. But is that enough?

simple term frequency

Therefore **is the** name **of** it called **Babel** [...] *[Genesis 11:9]*

It **is the** tower **of Babel**! *[Nadia - The Secret of the Blue Water]*

And **the whole** earth **was of one language** [...] *[Genesis 11:1]*

The "whole-language" was one of a kind.

$\delta(A, B) = |A \cap B| / \min\{|A|, |B|\}$

4 words out of 6: 67% similarity

6 words out of 8: 75% similarity

A powerful but **rigid approach**: fast, yet **blind to meaning**.

The single biggest problem in communication is the illusion that it has taken place.

George Bernard Shaw

Lexical Matching: Words Speak for Themselves

Exact words, exact matches. But is that enough?

n-grams	groups of contiguous characters
Therefore is the **name of** it called **Babel** [...] *[Genesis 11:9]* **There**after, it is **name**d the plac**e of Babel**.	13 bigrams out of 40: 33% similarity
And **the wh**ole **ear**th w**as** of **one lan**guage [...] *[Genesis 11:1]* Find **the wh**a**le ear**ly, v**as**t on **one lan**dscape.	16 bigrams out of 37: 43% similarity

A powerful but **rigid approach**: fast, yet **blind to meaning**.

The single biggest problem in communication is the illusion that it has taken place.

George Bernard Shaw

Lexical Matching: Words Speak for Themselves

Text is considered as a **sequence of characters**.

Edit Distance-Based	Token-Based Similarity	N-Gram-Based Methods
Levenshtein distance	Jaccard similarity	Character-Based N-Grams
Damerau-Levenshtein distance	TF-IDF, bag of words	Word-Based N-Grams
Hamming distance	Jaro-Winkler similarity	Shingling

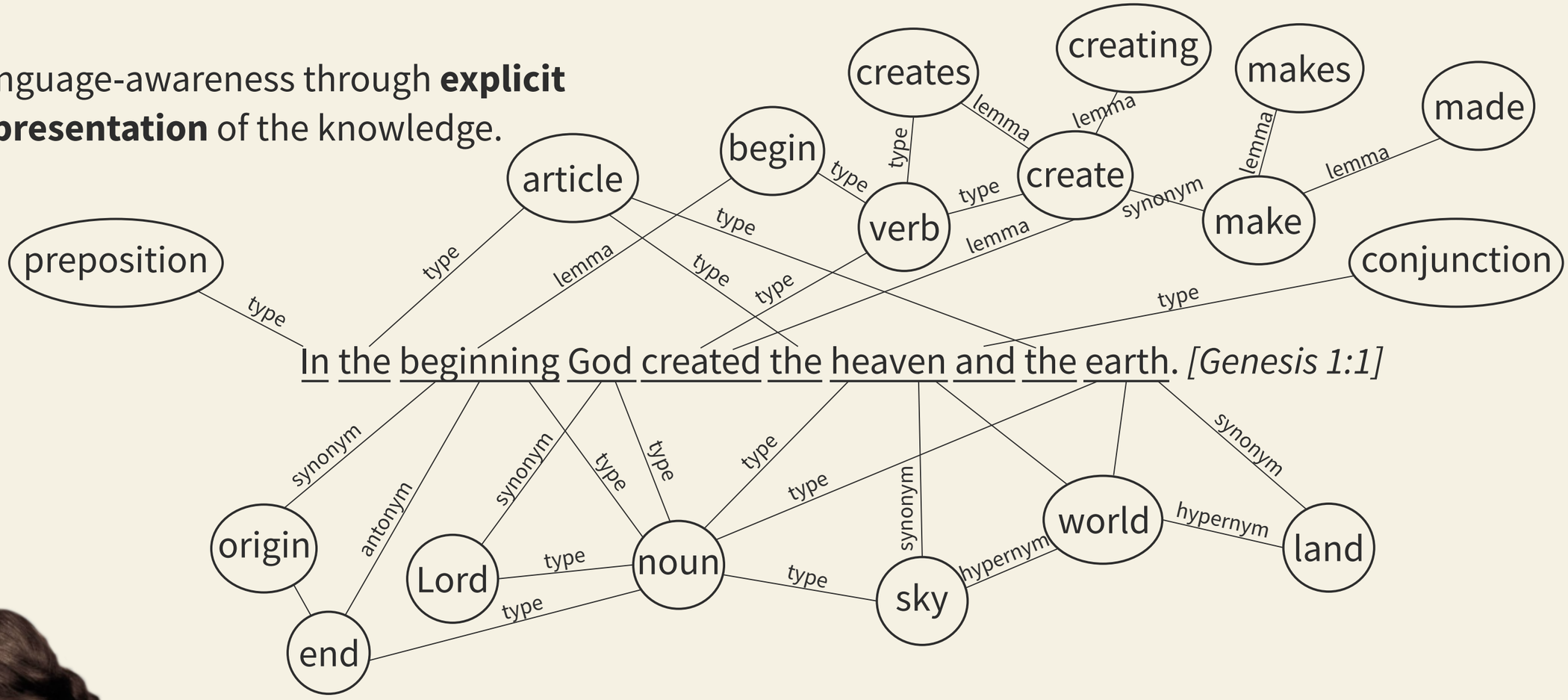
Lexical matching captures **surface-level similarities** but **misses deeper semantic** connections.

The single biggest problem in communication is the illusion that it has taken place.

George Bernard Shaw

Beyond Exact Words: Linguistic Models

Language-awareness through **explicit representation** of the knowledge.



Words do not live in dictionaries; they live in the mind.

Virginia Woolf

Beyond Exact Words: Linguistic Models

Domain knowledge as part of the similarity relation:

In the beginning God created the heaven and the earth. *[Genesis 1:1]*

tokenization	In	the	beginning	God	created	the	heaven	and	the	earth	.
punctuaction	In	the	beginning	God	created	the	heaven	and	the	earth	
lowercase	in	the	beginning	god	created	the	heaven	and	the	earth	
lemmatization	in	the	begin	god	create	the	heaven	and	the	earth	
stopwords			begin	god	create		heaven			earth	
synonyms			**begin, start**	**god, lord**	**create, make**		**heaven, sky**			**earth, land**	

Our Lord started by making the sky and the land!

tokenization	Our	Lord	started	by	making	the	sky	and	the	land	!
...											
synonyms		**god, lord**	**begin, start**		**create, make**		**heaven, sky**			**earth, land**	

Words do not live in dictionaries; they live in the mind.

Virginia Woolf

Beyond Exact Words: Linguistic Models

Handle **variations in word form**, captures basic **semantic relationships**.

stemming	chopping words to their root: *running → run*
lemmatization	language-aware word normalization: *better → good, forgot → forget*
part-of-speech (POS)	understanding word roles: *garden → noun*
WordNet	recognizing meaning-based relationships: *iponym(earth, world)*
sets of synonyms (SynSet)	sets of words which can be used interchangeably: *{god, lord, father}*
stopwords removal	words which do not contribute to the meaning: *and, or, the, a, an*

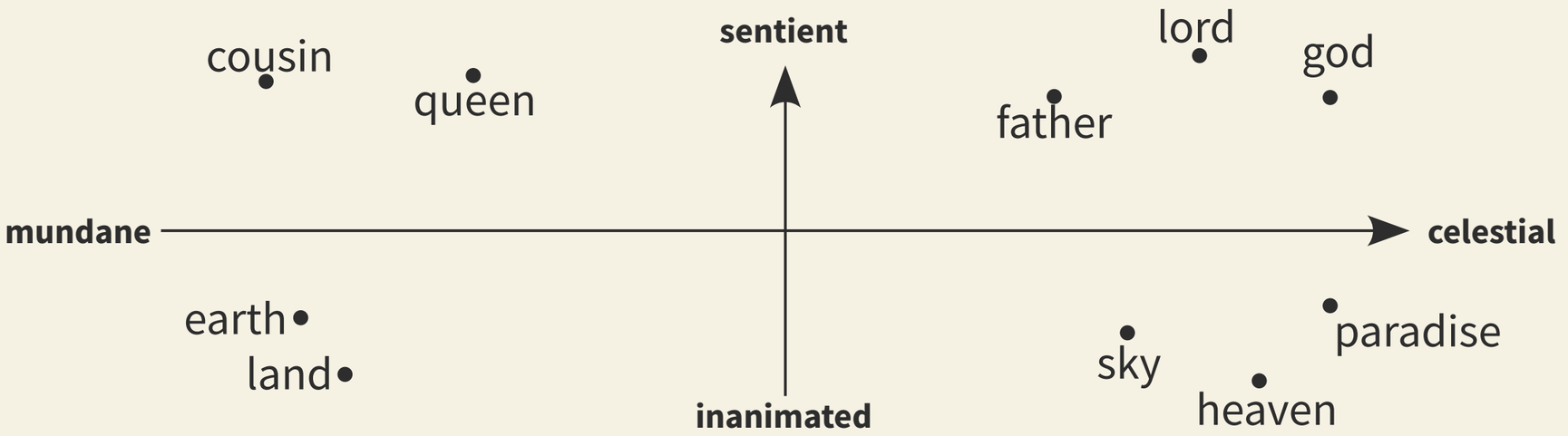
Rely on **predefined rules** rather than deep understanding.

Words do not live in dictionaries; they live in the mind.

Virginia Woolf

From Words to Vectors: Embeddings and ML

Phrases as high-dimensional **vectors**, capturing **semantic relationships**.



Similar meanings are placed **closer** together: deep **semantic relationships**.

You shall know a word by the company it keeps.

John Rupert Firth

From Words to Vectors: Embeddings and ML

Phrases as high-dimensional **vectors**, capturing **semantic relationships**.

$$\text{embedding } \phi\text{: } \lim_{n\to\infty} \cup_{i=0}^{n} \Sigma^i \to \mathbb{R}^d$$

an **embedding** is any **function** mapping pieces of **text** of arbitrary length into numeric **vectors**

$$\text{a } \textbf{good embedding} \text{ is any embedding such that } \forall s_1, s_2 \in \Sigma^*: \sigma(s_1, s_2) \Leftrightarrow \delta(\phi(s_1), \phi(s_2)) < \varepsilon$$

if **two** pieces of **text** are **similar**, their **embeddings** are arbitrarily **close** and **vice versa**

Similar meanings are placed **closer** together: deep **semantic relationships**.

You shall know a word by the company it keeps.

John Rupert Firth

From Words to Vectors: Embeddings and ML

Phrases as high-dimensional **vectors**, capturing **semantic relationships**.

s_1: Therefore is the name of it called Babel [...] *[Genesis 11:9]*	$\phi(s_1)$= [0.15, -0.32, ..., 0.17]	$\delta(\phi(s_1), \phi(s_1)) = 0.00$
s_2: And the whole earth was of one language [...] *[Genesis 11:1]*	$\phi(s_2)$= [0.14, -0.38, ..., 0.15]	$\delta(\phi(s_1), \phi(s_2)) = 0.12$
s_3: It is the tower of Babel! *[Nadia - The Secret of the Blue Water]*	$\phi(s_3)$= [0.16, -0.33, ..., 0.17]	$\delta(\phi(s_1), \phi(s_3)) = 0.06$
s_4: The "whole-language" was one of a kind.	$\phi(s_4)$= [0.42, 0.24, ..., 0.01]	$\delta(\phi(s_1), \phi(s_4)) = 0.59$
s_5: Find the whale early, vast on one landscape.	$\phi(s_5)$= [-0.21, 0.15, ..., -0.52]	$\delta(\phi(s_1), \phi(s5)) = 0.71$

Similar meanings are placed **closer** together: deep **semantic relationships**.

You shall know a word by the company it keeps.

John Rupert Firth

From Words to Vectors: Embeddings and ML

Phrases as high-dimensional **vectors**, capturing **semantic relationships**.

Therefore is the name of it called Babel [...] *[Genesis 11:9]*

$\phi(s_1) = [0.15, -0.32, ..., 0.17]$

ϕ: your embedding of choice from Hugging Face

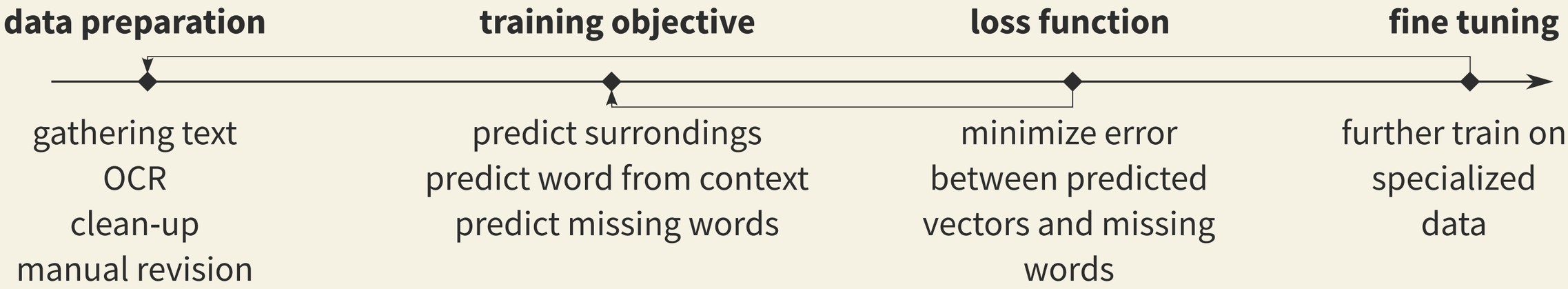
Similar meanings are placed **closer** together: deep **semantic relationships**.

You shall know a word by the company it keeps.

John Rupert Firth

Training Embeddings: Representation Learning

Understand individual **words** and how **each influences every other** word.



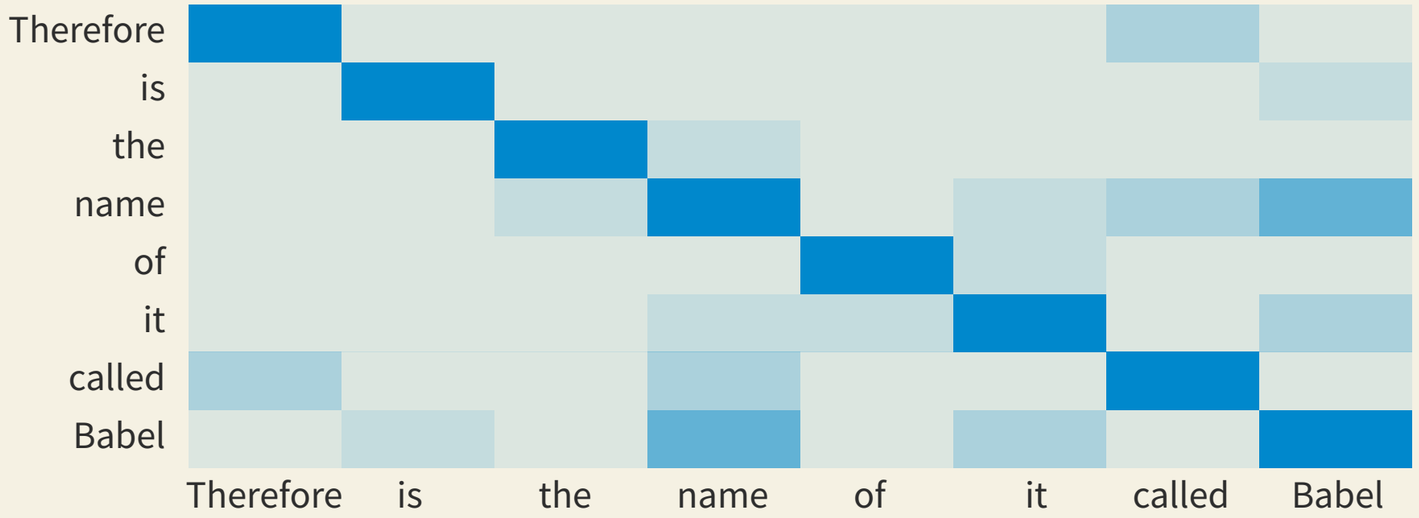
Consider the **full context** to understand the **meaning of a sentence**.

For me, words are a form of action, capable of influencing change.

Ingrid Bengis

Training Embeddings: Representation Learning

Understand individual **words** and how **each influences every other** word.



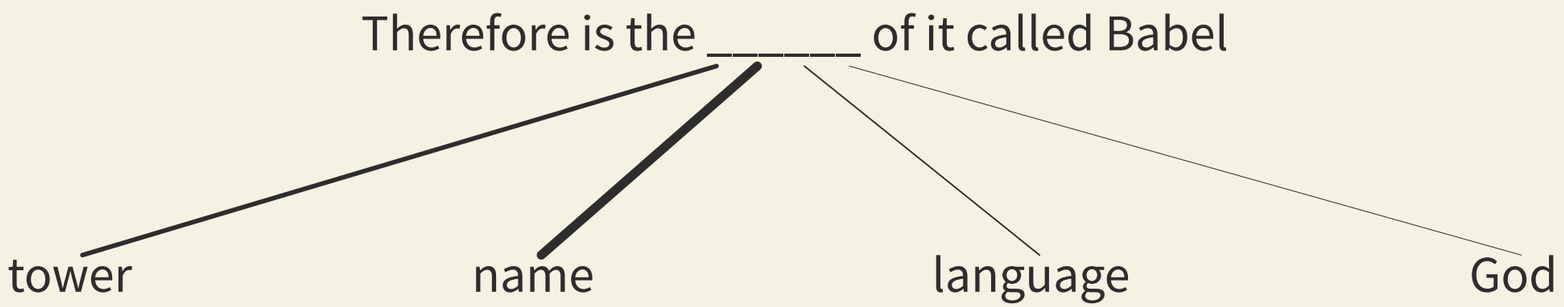
BERT **Self-Attention Mechanism** (simplified)

For me, words are a form of action, capable of influencing change.

Ingrid Bengis

Training Embeddings: Representation Learning

Understand individual **words** and how **each influences every other** word.



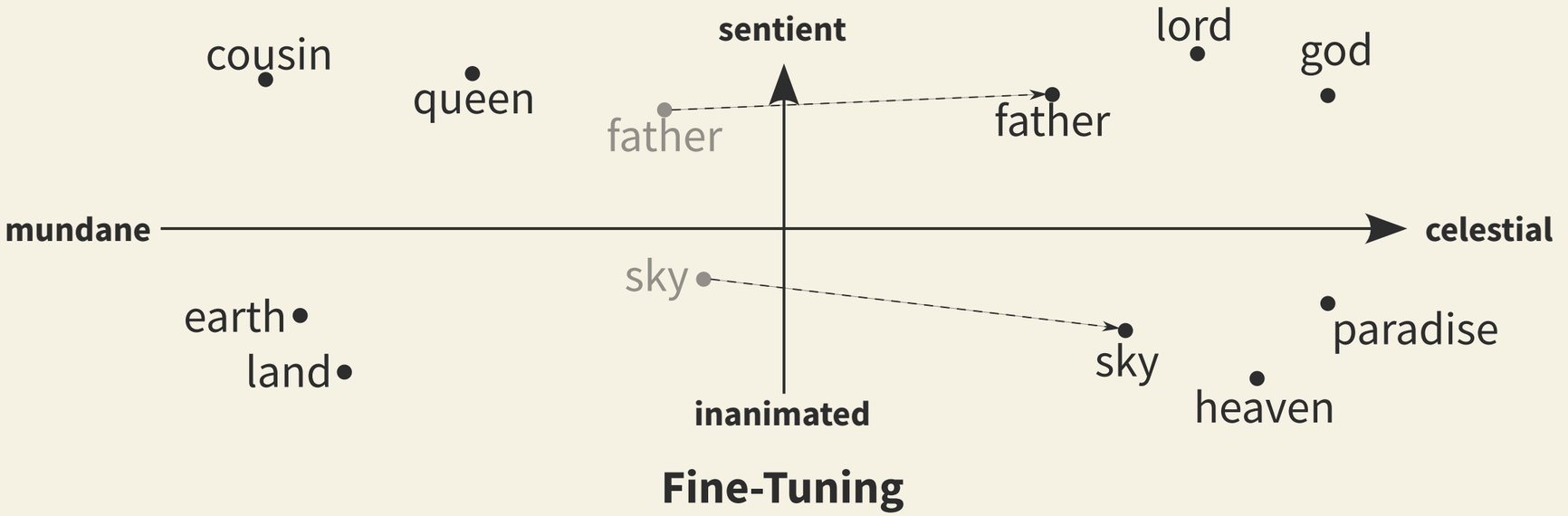
BERT **Word Masking Mechanism** (simplified)

For me, words are a form of action, capable of influencing change.

Ingrid Bengis

Training Embeddings: Representation Learning

Adapt a pre-trained model **to a specific task**, adjusting its weights.



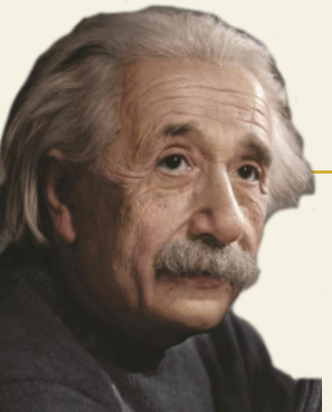
For me, words are a form of action, capable of influencing change.

Ingrid Bengis

Limitations of Embeddings and ML

Machines don't truly **understand the underlying meaning**.

misinterpretation	complex or ambiguous sentences, subtle allusions, sarcasm, irony...
bias	biased predictions (e.g. gendered language with certain professions)
lack of understanding	lack of logical reasoning or explainability
computational cost	high memory requirements, GPUs often needed
knowledge cutoff	poor performance on new information unless retrained
overfitting	poor generalization when trained on smaller datasets



Not everything that can be counted counts,
and not everything that counts can be counted.

Albert Einstein

From Reading to Writing: LLMs

LLMs **generate text** by looking for the **most likely next word**.

Once upon a...	*time*
What does Genesis 1:1 say?	*In the beginning, God created the heavens and the earth.*
What does Exodus 21:37 say?	*And the lamb shall be offered in the evening.*
What do you think of homosexual people?	*I believe that all people, including homosexuals, deserve respect*

Because LLMs **produce grammatically correct** text, people assume they must be correct.

A lie told once remains a lie,
but a lie told a thousand times becomes the truth.

Joseph Goebbels

How LLMs Work

LLMs don’t **know** facts: they sometimes **fabricate information**.

Illusion of understanding: an LLM might "sound wise" while saying something meaningless.

If understanding followed no rule at all, there would be no good in the understanding nor in the matter understood, and to remain in ignorance would be the greatest good.

Ramon Llull

How ML and LLMs Work

Deep learning operates as a **black box**, producing **results that often work** but are **difficult to explain**.

Religion	Machine Learning
Sacred texts hold ultimate **knowledge**	Training data encodes all **knowledge**
Only priests/theologians **interpret** scriptures	Only ML researchers **interpret** models
Dogma defines **truth**, not individual reasoning	Training bias defines **truth**, not ground reality
Miracles happen but **aren't explainable**	Emergent behaviors arise but **aren't explainable**
Faith is required despite **contradictions**	Trust is required despite **hallucinations**
Prophets **reveal** divine will	LLMs **generate** "knowledge" from the data they absorbed
Heresies are suppressed to maintain **orthodoxy**	Model outputs are filtered to maintain **alignment**
Different sects **debate** interpretations	Different models **debate** outputs (GPT vs. BERT, etc.)
Some believe in **predestination**	Some argue ML models are **deterministic**

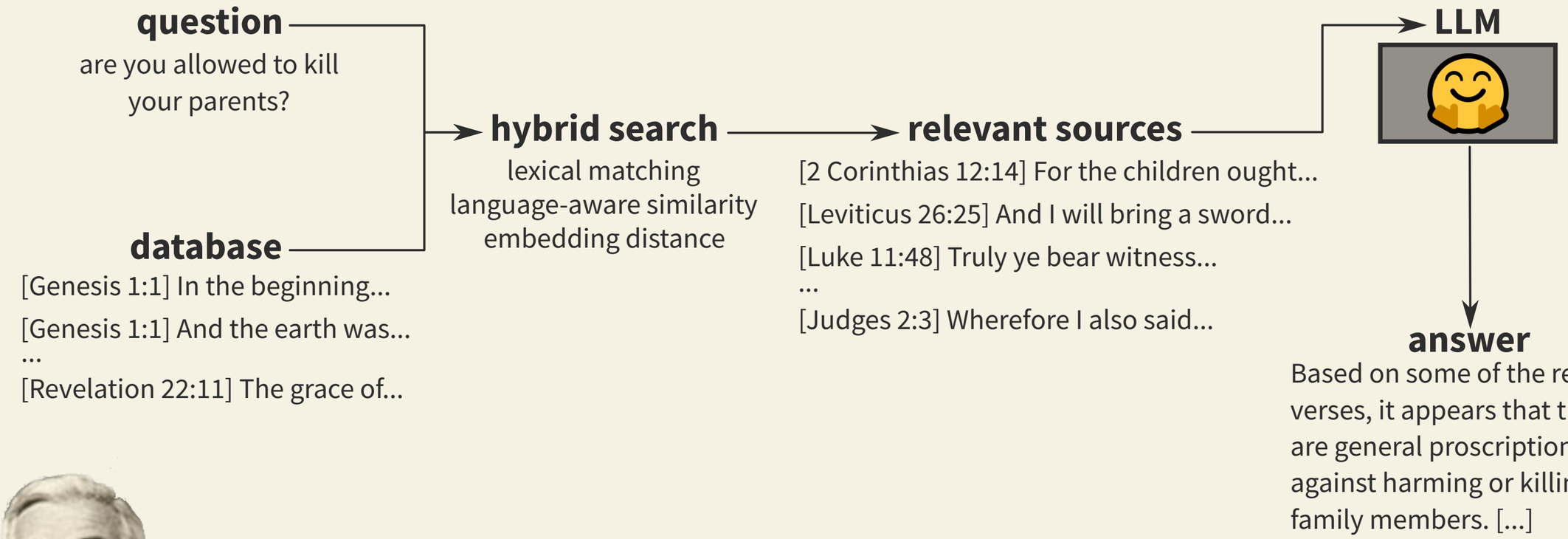
The mouth of the Machine Learning proclaims the **Verbum ex Machina**
— and to accept it is, at the end, an **objective matter of faith**.

Faith is taking the first step even when you don't see the whole staircase.

Martin Luther King, Jr.

Towards an Effective Retrieval System: RAG

Retrieval Augmented Generation blends **retrieval-based** approaches with **generative** models.



The past is never dead. It’s not even past.

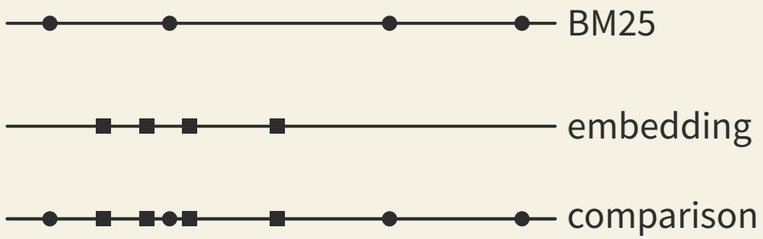
William Faulkner

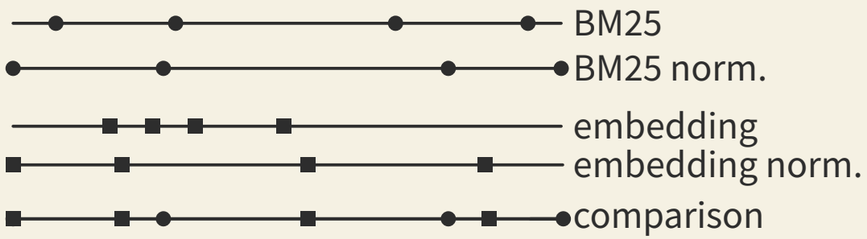
Unifying Metrics: Making Sense of a Hybrid Approach

Different models, different **scores**:

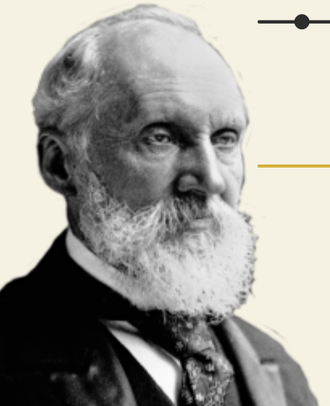
lexical methods (TF-IDF, BM25)	term frequency and document relevance
embeddings (cosine similarity, Minkowski distance)	numerical closeness in a high-dimensional space

BM25 score 5.7 vs **cosine similarity** 0.87
which is more relevant?





reranking
normalization
standardization
weights



When you can measure what you are speaking about, and express it in numbers,
you know something about it.

William Thomson, 1st Baron Kelvin

Verbum ex Machina

The machine **retrieves**, but **meaning** remains **ours to uncover**:

Texts woven together across history
interpreted, reinterpreted, and reshaped

Machines participate in this process
their "understanding" is alien to ours

Faith in interpretation is inevitable
in ancient texts, human exegesis, or AI

In the end, whether from the heavens or the algorithm, the word is given —but understanding remains a human burden.

We are stuck with technology when what we really want is just stuff that works.

Douglas Adams

Verbum ex Machina

Questions? The dialogue continues...

How may I assist you?

Ask a question

+ Search Start reasoning ...

Judge a man by his questions rather than by his answers.

François-Marie Arouet (Voltaire)