Generative Search and Biases
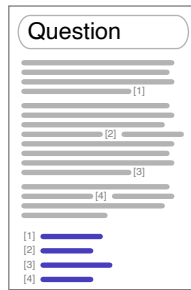
Challenges and Opportunities of a New Information Ecology

Martin Potthast
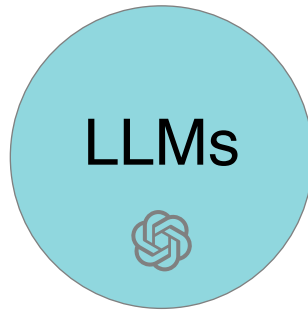
Webis Group · UNIVERSITÄT LEIPZIG · ScaDS.AI DRESDEN LEIPZIG · Open WebSearch .eu
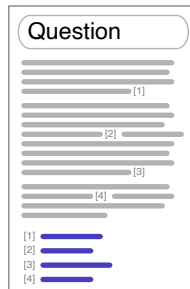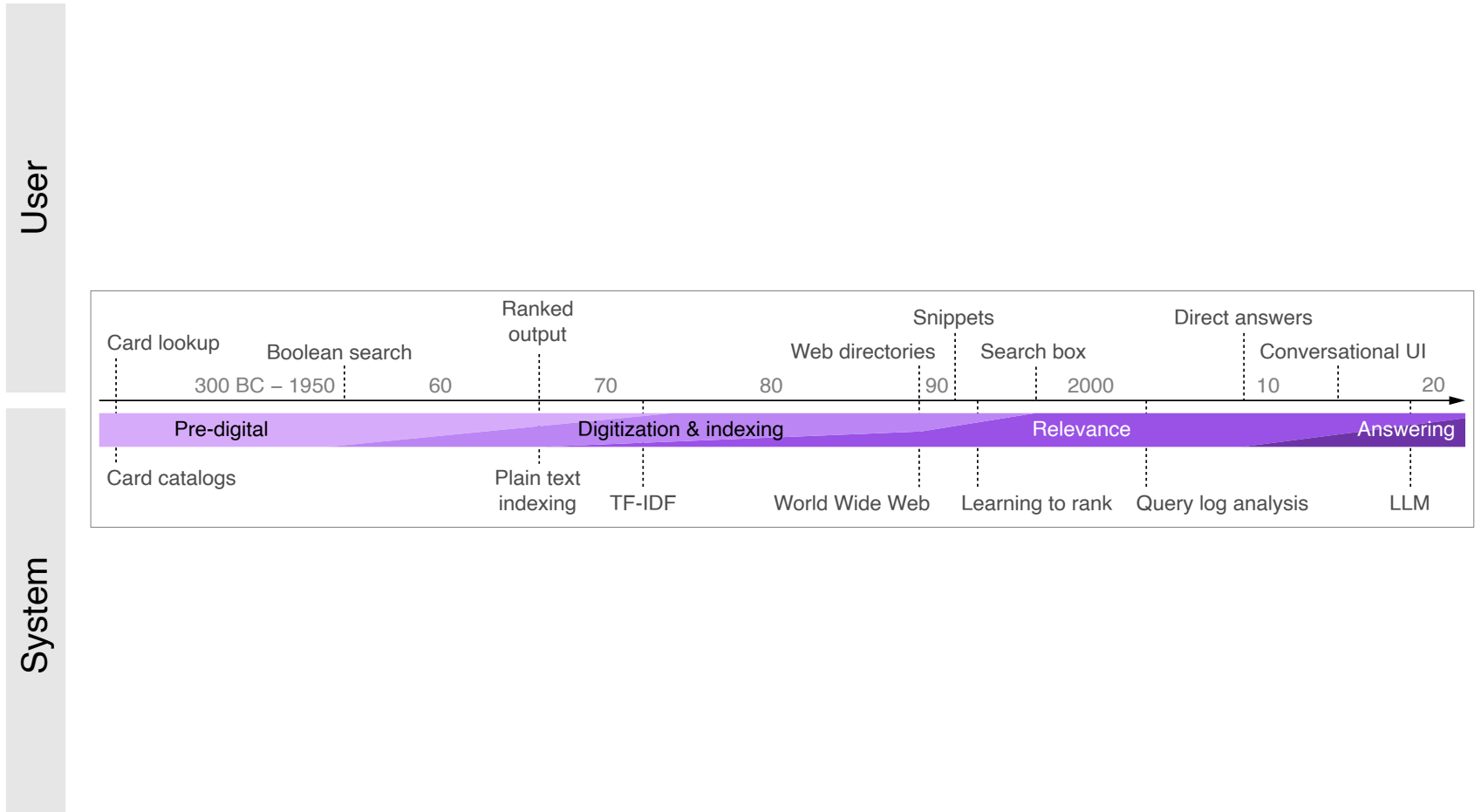
Search + LLMs

Query → Question

① A Short History of Search Engines

A Short History of Search Engines

• M. Potthast, M. Hagen, B. Stein (2020). The dilemma of the direct answer.

A Short History of Search Engines

• M. Potthast, M. Hagen, B. Stein (2020). The dilemma of the direct answer.

A Short History of Search Engines

	Card lookup	Boolean search	Ranked output		Web directories	Snippets	Search box	Direct answers	Conversational UI
	300 BC – 1950	60	70	80	90	2000	10	20

Conversational — Keyword "telegrams" + multimedia results — Conversational

	Card catalogs	Plain text indexing	TF-IDF	World Wide Web	Learning to rank	Query log analysis	LLM

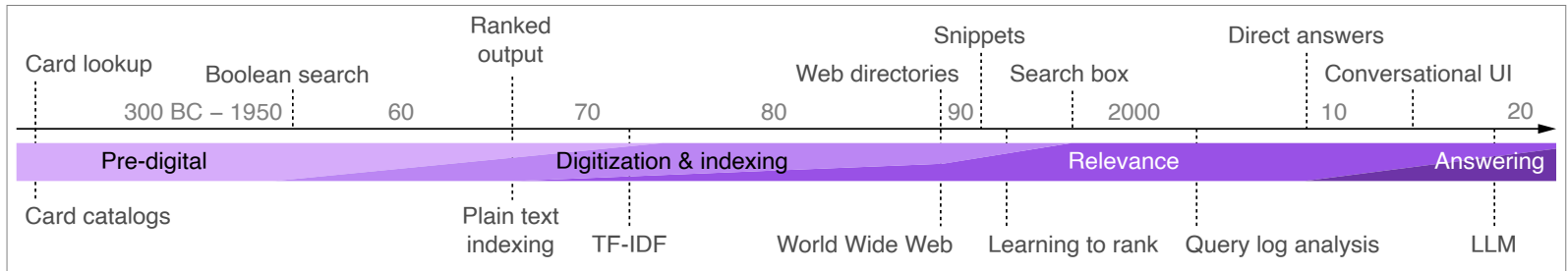- M. Potthast, M. Hagen, B. Stein (2020). The dilemma of the direct answer.

A Short History of Search Engines

• M. Potthast, M. Hagen, B. Stein (2020). The dilemma of the direct answer.

A Short History of Search Engines

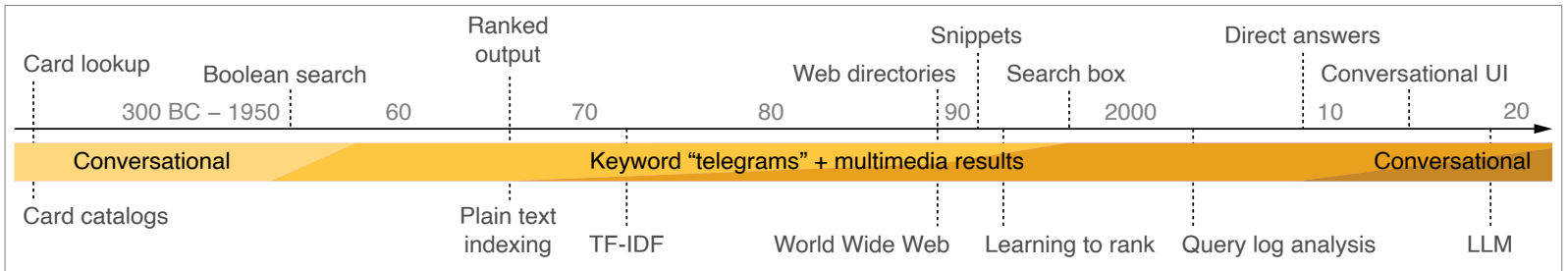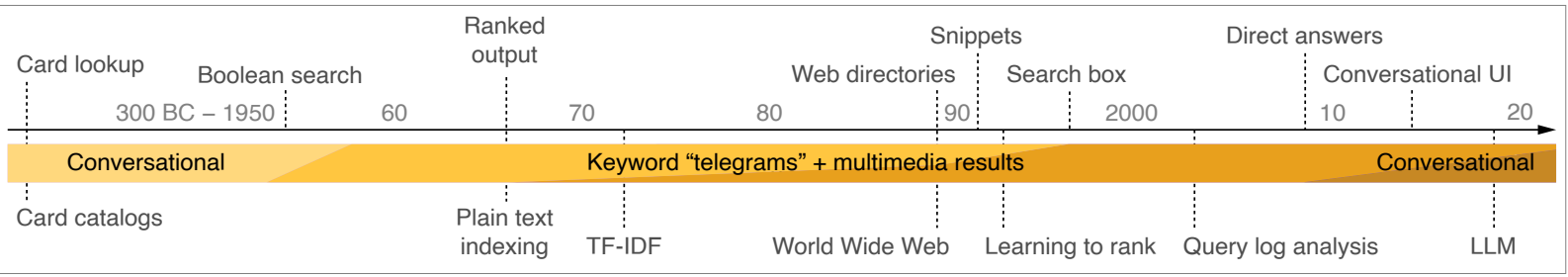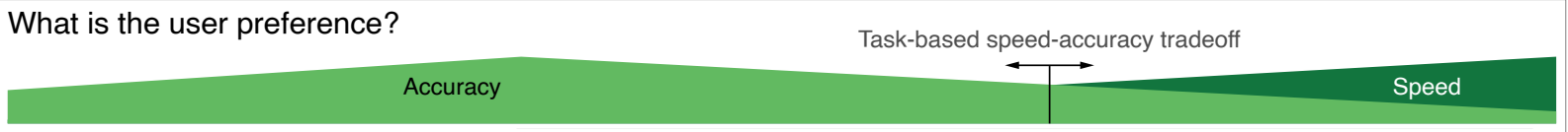• M. Potthast, M. Hagen, B. Stein (2020). The dilemma of the direct answer.

Web Search Architecture

Indexing

Storage

?

Retrieval

Web Search Architecture

Indexing

Acquisition — Text analysis — Indexing

d

Storage

Document store

Index

Retrieval

?
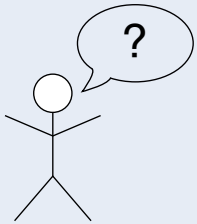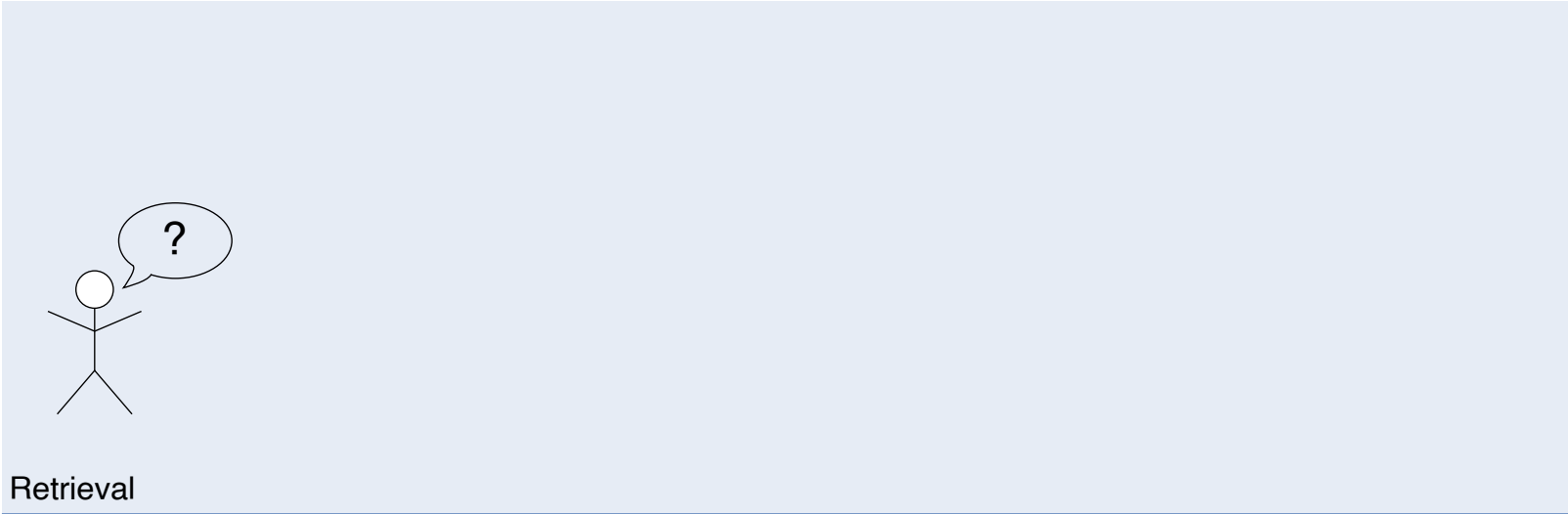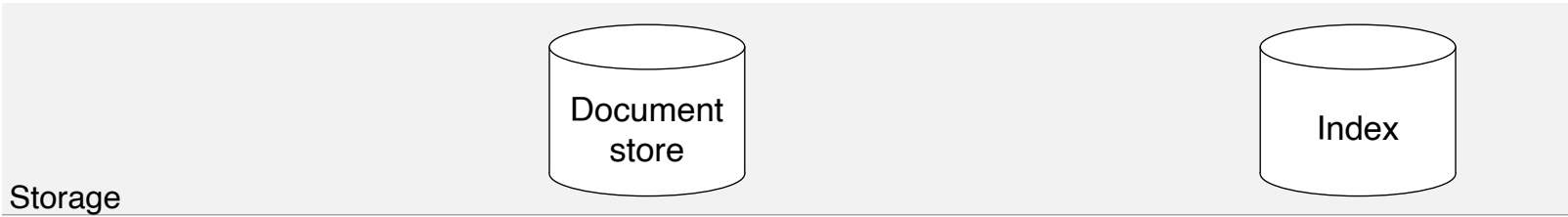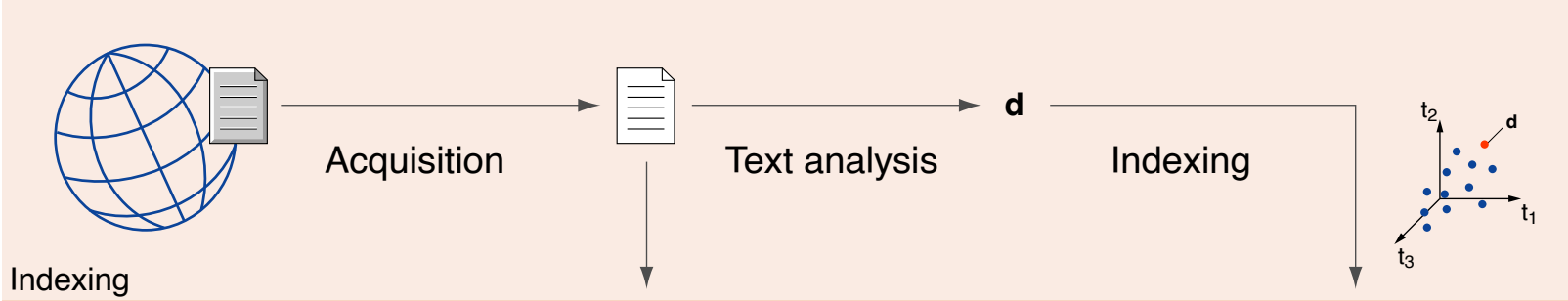
Web Search Architecture

Web Search Architecture

②
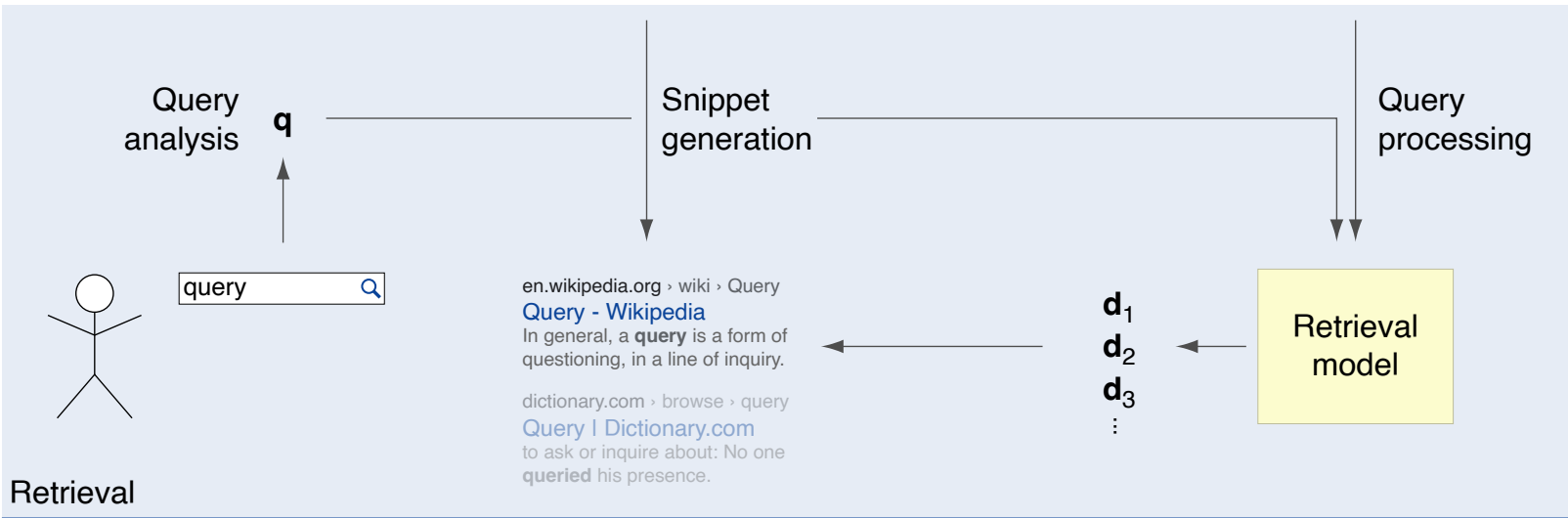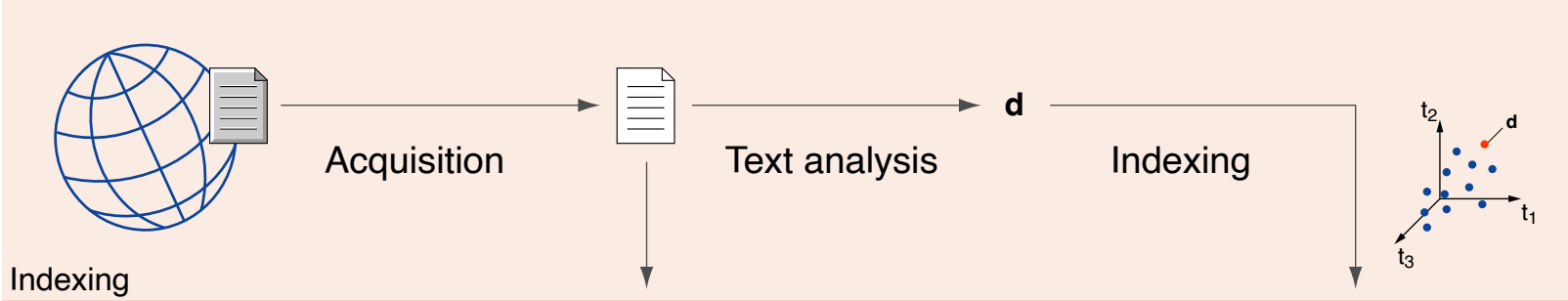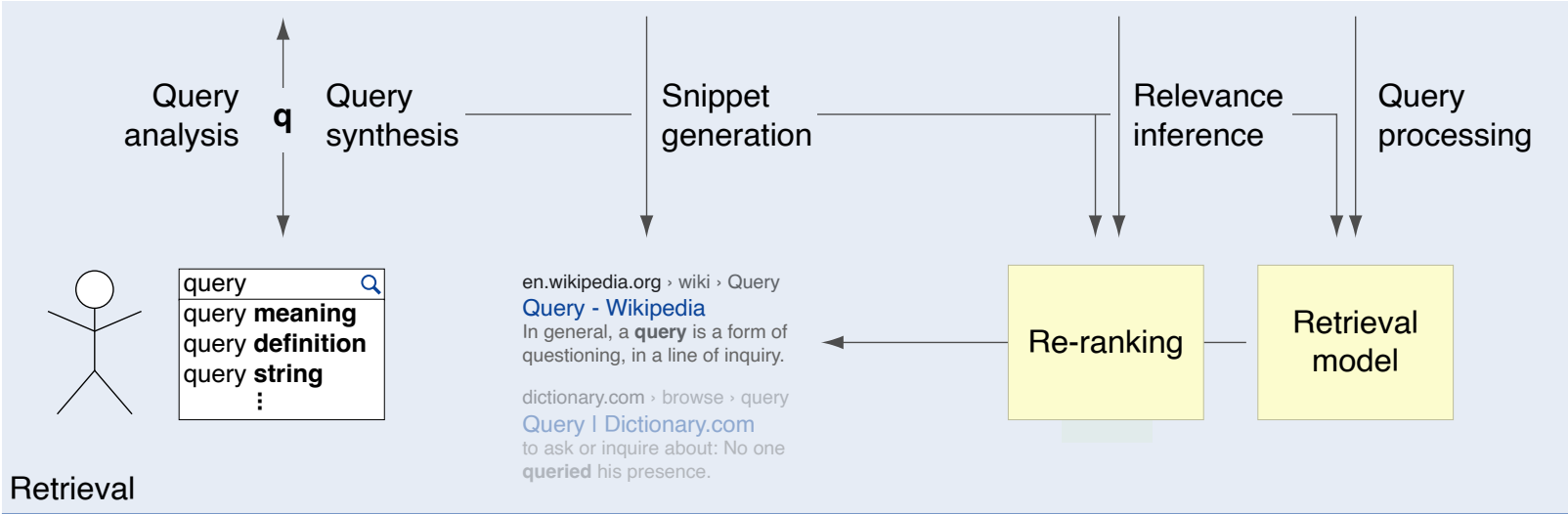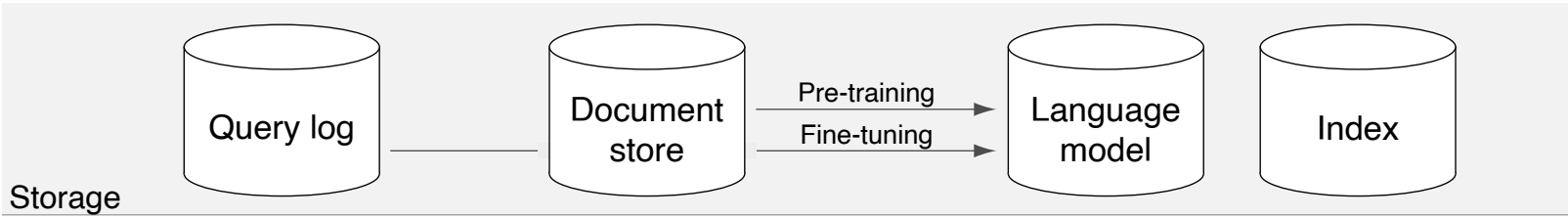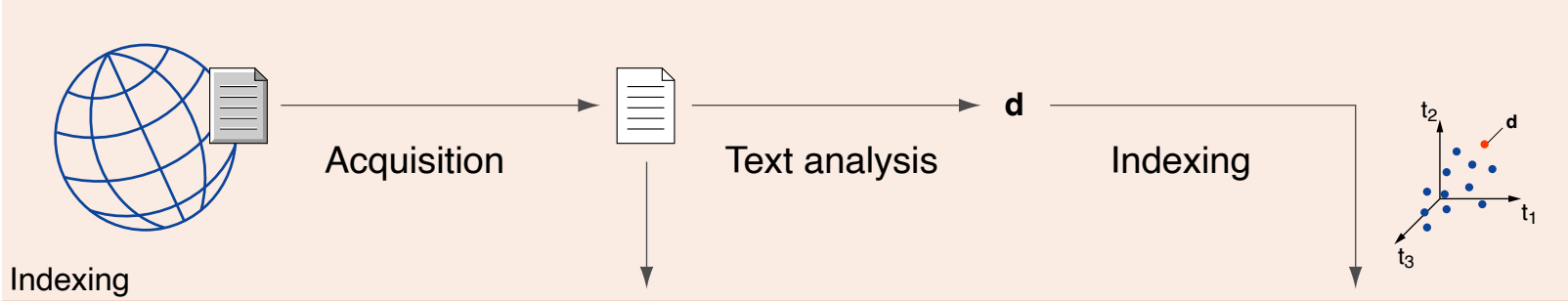
A Short History of Language Models

A Short History of Language Models
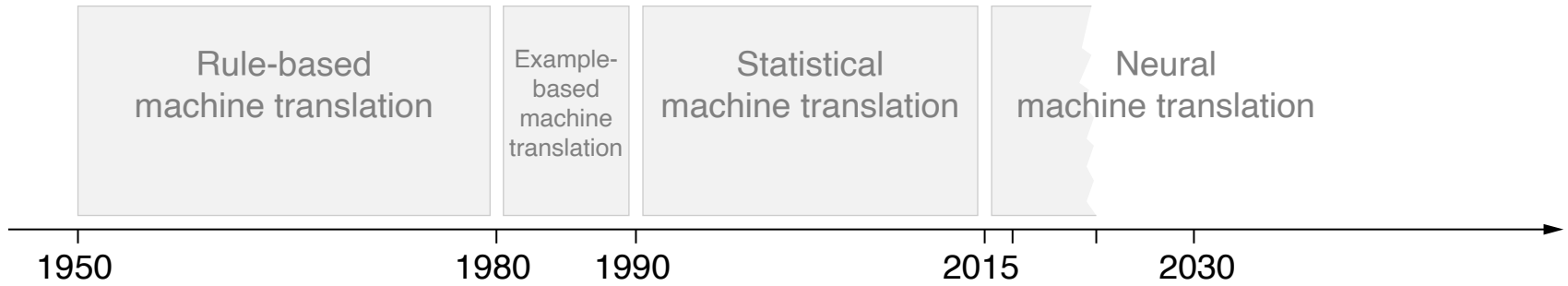
Rule-based machine translation	Example-based machine translation	Statistical machine translation	Neural machine translation

1950   1980   1990   2015   2030

A Short History of Language Models

Rule-based machine translation	Example-based machine translation	Statistical machine translation	Neural machine translation

1950   1980   1990   2015   2030
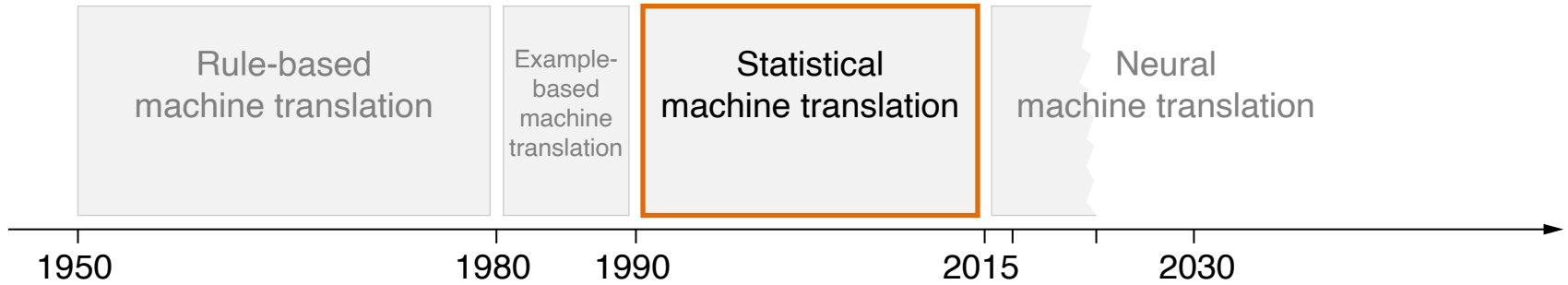
A statistical language model
is a probability distribution over all possible texts.

A Short History of Language Models

Rule-based machine translation	Example-based machine translation	Statistical machine translation	Neural machine translation

1950   1980   1990   2015   2030
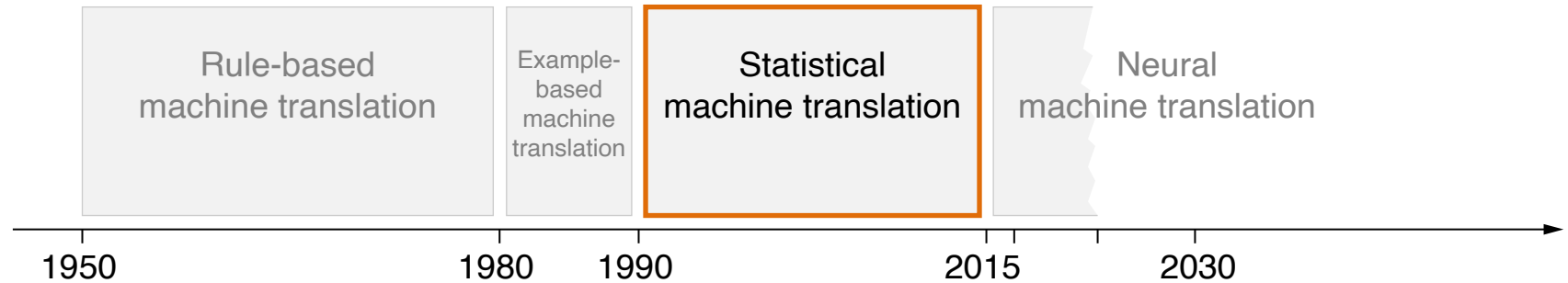
A statistical language model
is a probability distribution over all possible texts.

Illustration:

(1) `i love my ?`   N N

(2) `see ... works.`   N N

A Short History of Language Models

Rule-based machine translation	Example-based machine translation	Statistical machine translation	Neural machine translation

1950   1980   1990   2015   2030
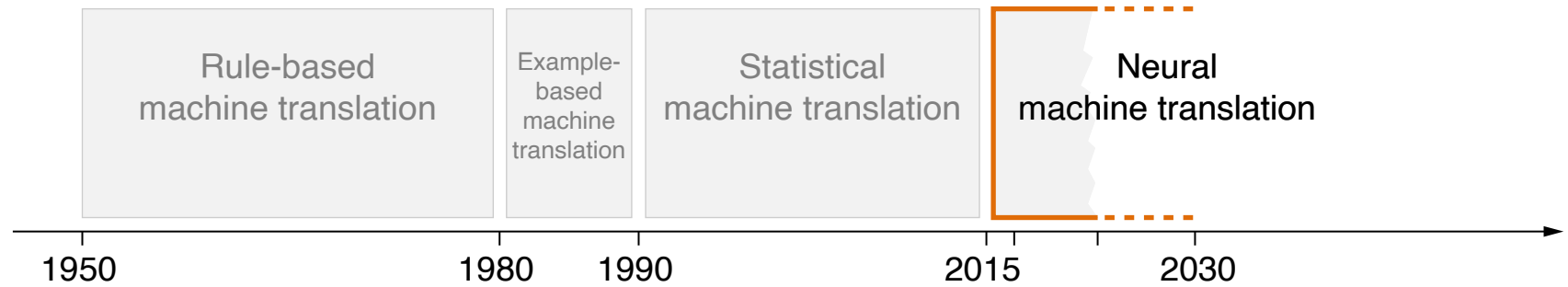
A neural language model
approximates a statistical language model.

A Short History of Language Models
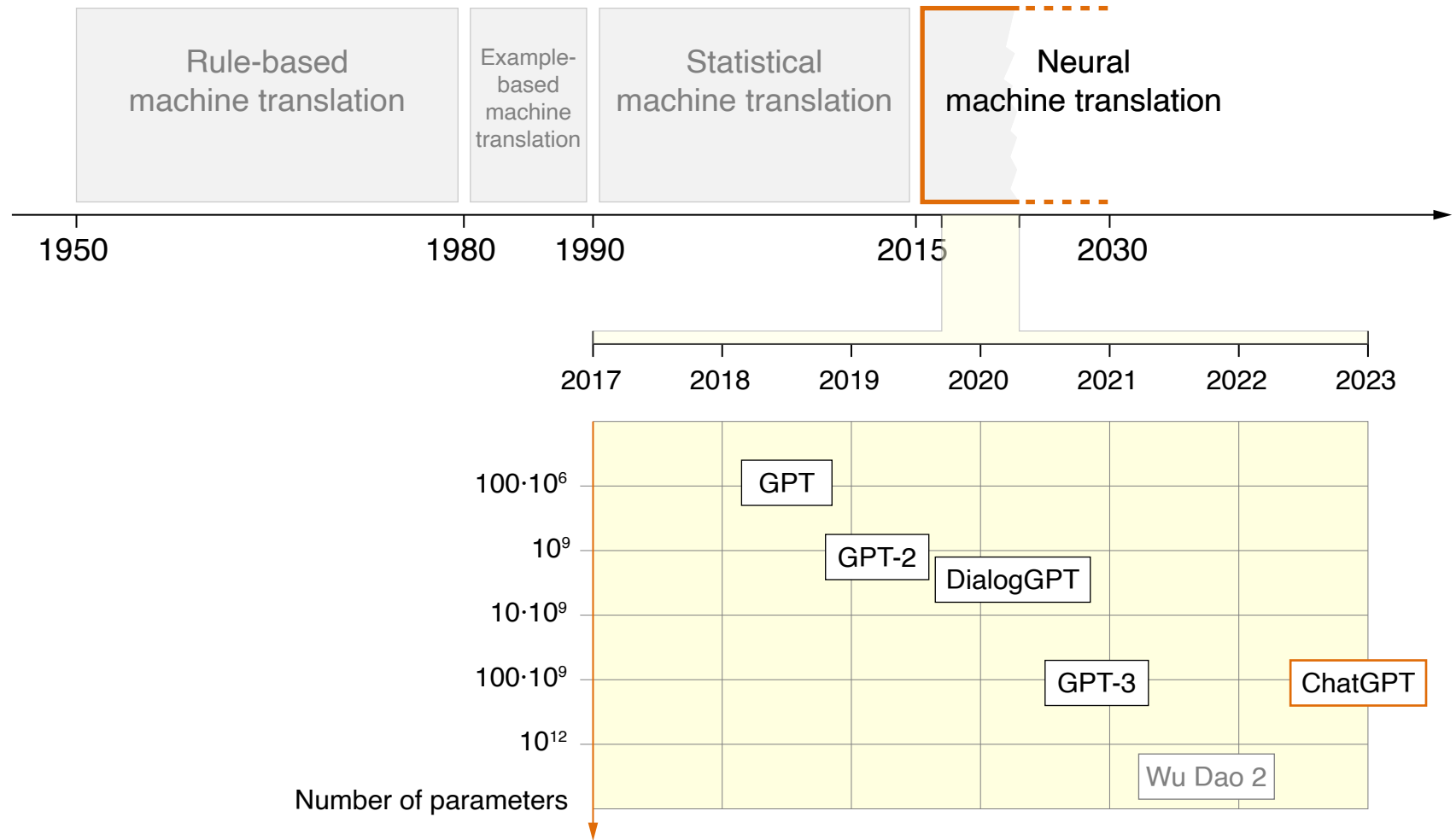
Rule-based machine translation

Example-based machine translation

Statistical machine translation

Neural machine translation

1950   1980   1990   2015   2030

2017   2018   2019   2020   2021   2022   2023

$100 \cdot 10^6$ GPT

10^9 GPT-2   DialogGPT

$10 \cdot 10^9$

$100 \cdot 10^9$ GPT-3   ChatGPT

10^{12}

Number of parameters   Wu Dao 2

Training Corpora Sources		Parameters	Computing / Training
Wikipedia 11GB	Books 21GB	175,000,000,000	• 355 years on a single Tesla V100 GPU.
Journals 101GB	Reddit 50GB	$(175 \cdot 10^9)$	• $\approx$ 34 days on 1,024 x A100 GPUs.
Common Crawl 570GB			• $4.6M costs a single training run.

GPT-3 [Jun. 2020]

Training Corpora Sources

Wikipedia	11GB	Books	21GB
Journals	101GB	Reddit	50GB
Common Crawl	570GB		

Parameters

175,000,000,000

$(175 \cdot 10^9)$

Computing / Training

- 355 years on a single Tesla V100 GPU.
- $\approx$ 34 days on 1,024 x A100 GPUs.
- $4.6M costs a single training run.

GPT-3 [Jun. 2020]

World Knowledge

🧍 What city is in the northwest corner of Ohio?

⑤ Toledo is in the north-west corner of Ohio.

Common Sense

🧍 Why don't animals have three legs?

⑤ Animals don't have three legs because they would fall over.

Logical Reasoning

🧍 If I put a pencil in a box, then put another pencil in the box, what is in the box?

⑤ Two pencils.

Training Corpora Sources		Parameters	Computing / Training
Wikipedia 11GB Books 21GB Journals 101GB Reddit 50GB Common Crawl 570GB		175,000,000,000 $(175 \cdot 10^9)$	• 355 years on a single Tesla V100 GPU. • $\approx$ 34 days on 1,024 x A100 GPUs. • $4.6M costs a single training run.

GPT-3 [Jun. 2020]

World Knowledge

☖ What city is in the northwest corner of Ohio?

⑨ Toledo is in the north-west corner of Ohio.

Common Sense

☖ Why don't animals have three legs?

⑨ Animals don't have three legs because they would fall over.

Logical Reasoning

☖ If I put a pencil in a box, then put another pencil in the box, what is in the box?

⑨ Two pencils.

+ Learn to follow instructions and to comply with answer policies.

(1) Fine-tuning of GPT-3 to follow instructions: 13,000 popular prompts with hand-written answers.

(2) Training of a reward model: 33,000 prompts with 4-9 answers, ranked from best to worse.

(3) Training of the fine-tuned GPT-3 model from Step (1) to follow the reward policy.

GPT-3.5 (InstructGPT) [Jan. 2022]

Training Corpora Sources

Wikipedia	11GB	Books	21GB
Journals	101GB	Reddit	50GB
Common Crawl	570GB		

Parameters

175,000,000,000

$(175 \cdot 10^9)$

Computing / Training

- 355 years on a single Tesla V100 GPU.
- $\approx$ 34 days on 1,024 x A100 GPUs.
- $4.6M costs a single training run.

GPT-3 [Jun. 2020]

World Knowledge

🧍 What city is in the
   northwest corner of Ohio?

💲 Toledo is in the north-
   west corner of Ohio.

Common Sense

🧍 Why don't animals have
   three legs?

💲 Animals don't have three
   legs because they would
   fall over.

Logical Reasoning

🧍 If I put a pencil in a box,
   then put another pencil in
   the box, what is in the
   box?

💲 Two pencils.

+   Learn to follow instructions and to comply with answer policies.

(1) Fine-tuning of GPT-3 to follow instructions: 13,000 popular prompts with hand-written answers.

(2) Training of a reward model: 33,000 prompts with 4-9 answers, ranked from best to worse.

(3) Training of the fine-tuned GPT-3 model from Step (1) to follow the reward policy.

GPT-3.5 (InstructGPT) [Jan. 2022]

+   Fine-tuning of GPT-3.5 to comply with even stricter guardrails.

ChatGPT [Nov. 2022]

Training Corpora Sources

Wikipedia	11GB	Books	21GB
Journals	101GB	Reddit	50GB
Common Crawl	570GB		

Parameters

175,000,000,000

$(175 \cdot 10^9)$

Computing / Training

- 355 years on a single Tesla V100 GPU.
- $\approx$ 34 days on 1,024 x A100 GPUs.
- $4.6M costs a single training run.

GPT-3 [Jun. 2020]

World Knowledge

👤 What city is in the northwest corner of Ohio?

🌀 Toledo is in the north-west corner of Ohio.

Common Sense

👤 Why don't animals have three legs?

🌀 Animals don't have three legs because they would fall over.

Logical Reasoning

👤 If I put a pencil in a box, then put another pencil in the box, what is in the box?

🌀 Two pencils.

+ Learn to follow instructions and to comply with answer policies.

(1) Fine-tuning of GPT-3 to follow instructions: 13,000 popular prompts with hand-written answers.

(2) Training of a reward model: 33,000 prompts with 4-9 answers, ranked from best to worse.

(3) Training of the fine-tuned GPT-3 model from Step (1) to follow the reward policy.

GPT-3.5 (InstructGPT) [Jan. 2022]

+ Fine-tuning of GPT-3.5 to comply with even stricter guardrails.

ChatGPT [Nov. 2022]

The Library of Babel
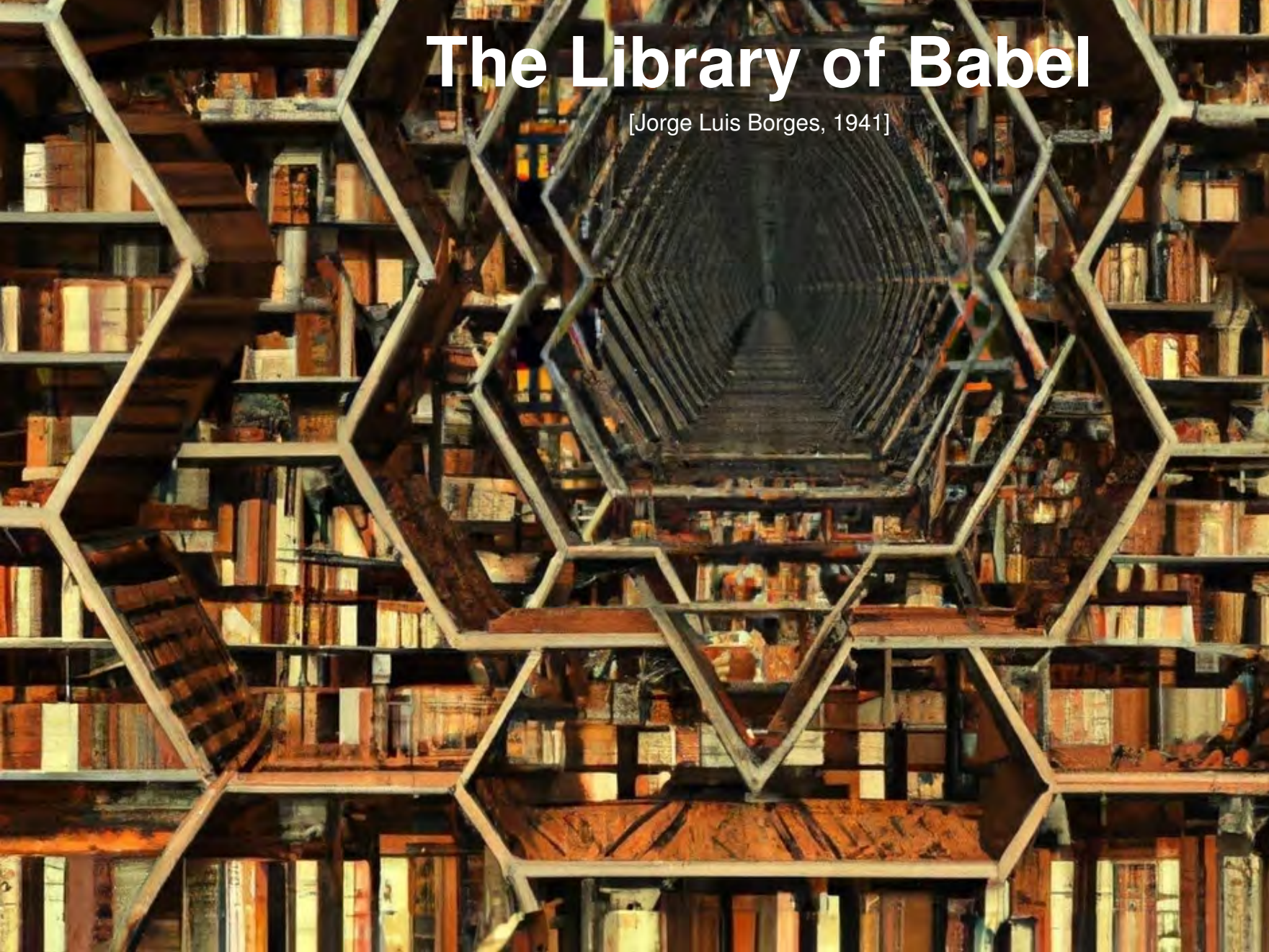
[Jorge Luis Borges, 1941]

The Library of Babel

[Jorge Luis Borges, 1941]

❑ Infinite library with all possible texts from all letter combinations

❑ The people in it spend their lives searching for meaningful text fragments

The Library of Babel

[Jorge Luis Borges, 1941]

- Infinite library with all possible texts from all letter combinations

- The people in it spend their lives searching for meaningful text fragments

- When prompted, a language model "retrieves" a relevant text [Deckers et al., 2024]:

A language model is an infinite index

③

On Biases

Bias in algorithms

Cognitive bias

Inductive bias

Statistical bias

Bias in data

What should we remember?

Too much information.

Need to act fast.

Not enough meaning.

Quadrant / category labels around the circle:

- We store memories differently based on how they were experienced
- We reduce events and lists to their key elements
- We discard specifics to form generalities
- We edit and reinforce some memories after the fact
- We favor simple-looking options and complete information over complex, ambiguous options
- To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions
- To get things done, we tend to complete things we've invested time and energy in
- To stay focused, we favor the immediate, relatable thing in front of us
- To act, we must be confident we can make an impact and feel what we do is important
- We project our current mindset and assumptions onto the past and future
- We think we know what other people are thinking
- We simplify probabilities and numbers to make them easier to think about
- We imagine things and people we're familiar with or fond of as better
- We fill in characteristics from stereotypes, generalities, and prior histories
- We tend to find stories and patterns even when looking at sparse data
- We notice flaws in others more easily than we notice flaws in ourselves
- We are drawn to details that confirm our own existing beliefs
- We notice when something has changed
- Bizarre, funny, visually striking, or anthropomorphic things stick out more than non-bizarre/unfunny things
- We notice things already primed in memory or repeated often

Bottom category tiles:

Bias in algorithms | **Cognitive bias** | Bias in data
Inductive bias
Statistical bias

31

© WEBIS 2023

Bias in algorithms

Cognitive bias

Inductive bias

Statistical bias

Bias in data

Too much information.

We store memories differently based on how they were experienced

We notice things already primed in memory or repeated often

We reduce events and lists to their key elements

Bizarre, funny, visually striking, or anthropomorphic things stick out more than non-bizarre/unfunny things

We discard specifics to form generalities

We notice when something has changed

We edit and reinforce some memories after the fact

We are drawn to details that confirm our own existing beliefs

We favor simple-looking options and complete information over complex, ambiguous options

We notice flaws in others more easily than we notice flaws in ourselves

To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions

We tend to find stories and patterns even when looking at sparse data

To get things done, we tend to complete things we've invested time and energy in

We fill in characteristics from stereotypes, generalities, and prior histories

To stay focused, we favor the immediate, relatable thing in front of us

We imagine things and people we're familiar with or fond of as better

To act, we must be confident we can make an impact and feel what we do is important

We simplify probabilities and numbers to make them easier to think about

Need to act fast.

We project our current mindset and assumptions onto the past and future

We think we know what other people are thinking

Not enough meaning.

Hindsight bias

Bias in algorithms		Bias in data
	Cognitive bias	
	Inductive bias	
Statistical bias		

33

©WEBIS 2023

What should we remember?

Too much information.

Framing effect

We store memories differently based on how they were experienced

We reduce events and lists to their key elements

We discard specifics to form generalities

We edit and reinforce some memories after the fact

We favor simple-looking options and complete information over complex, ambiguous options

To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions

To get things done, we tend to complete things we've invested time and energy in

To stay focused, we favor the immediate, relatable thing in front of us

To act, we must be confident we can make an impact and feel what we do is important

We notice things already primed in memory or repeated often

Bizarre, funny, visually striking, anthropomorphic things stick out more than non-bizarre/unfunny things

We notice when something has changed

We are drawn to details that confirm our own existing beliefs

We notice flaws in others more easily than we notice flaws in ourselves

We tend to find stories and patterns even when looking at sparse data

We fill in characteristics from stereotypes, generalities, and prior histories

We imagine things and people we're familiar with or fond of as better

We simplify probabilities and numbers to make them easier to think about

We think we know what other people are thinking

Not enough meaning.

Need to act fast.

We project our current mindset and assumptions onto the past and future

Bias in algorithms

Cognitive bias

Inductive bias

Statistical bias

Bias in data

34

©WEBIS 2023

Too much information.
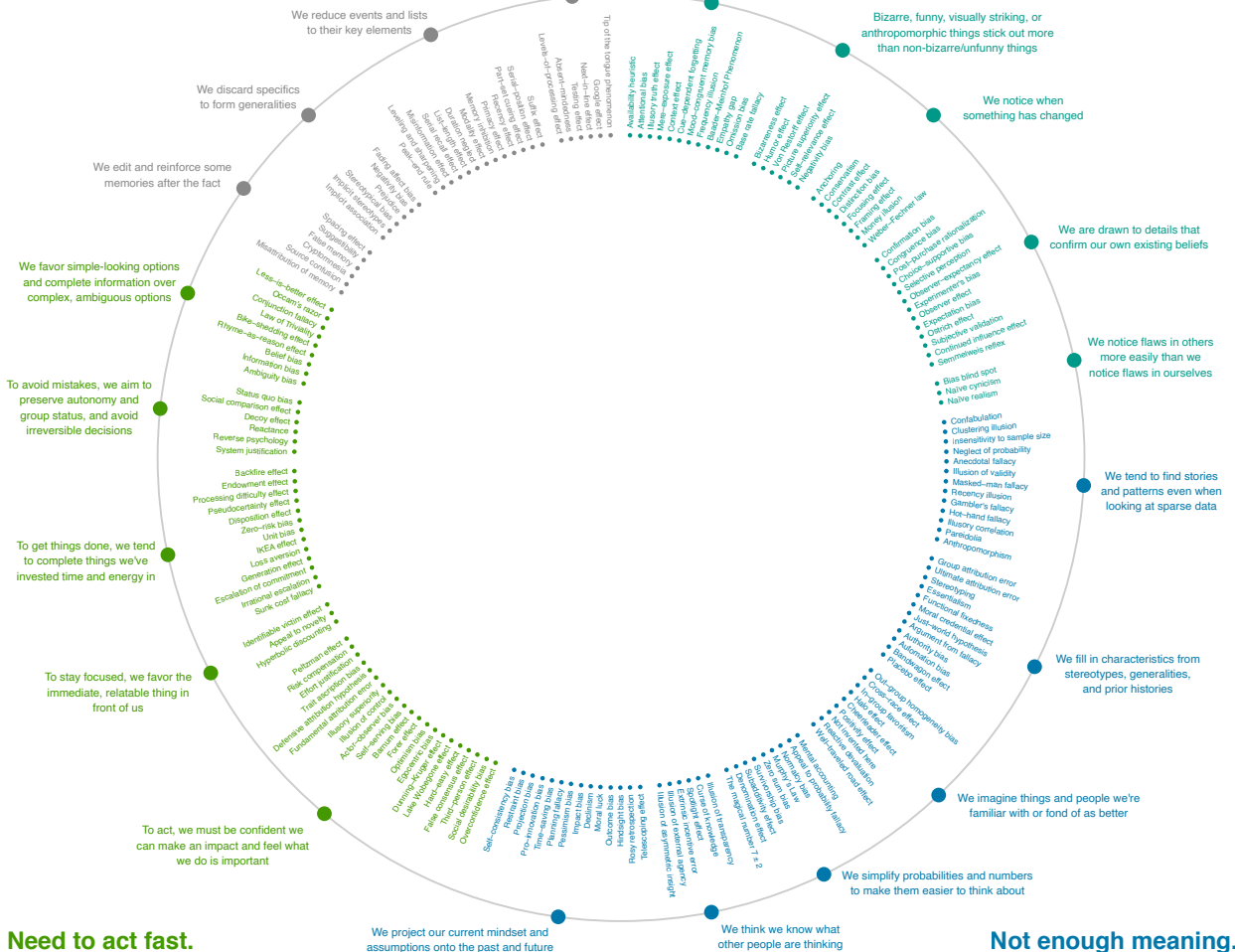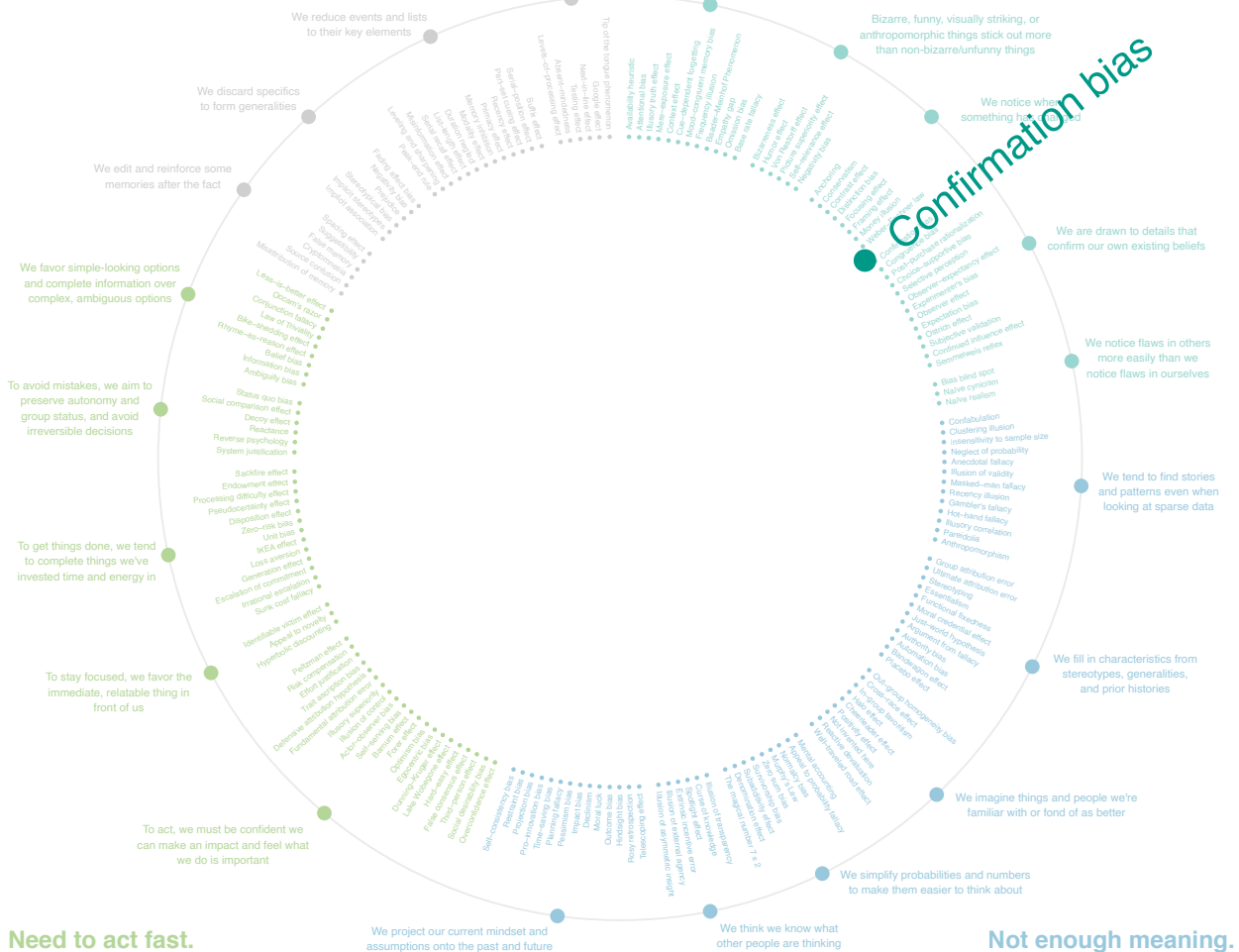
We store memories differently based on how they were experienced

We notice things already primed in memory or repeated often

We reduce events and lists to their key elements

Bizarre, funny, visually striking, or anthropomorphic things stick out more than non-bizarre/unfunny things

We discard specifics to form generalities

We edit and reinforce some memories after the fact

We notice when something has changed

We favor simple-looking options and complete information over complex, ambiguous options

We are drawn to details that confirm our own existing beliefs

Availability heuristic

To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions

We notice flaws in others more easily than we notice flaws in ourselves

To get things done, we tend to complete things we've invested time and energy in

We tend to find stories and patterns even when looking at sparse data

To stay focused, we favor the immediate, relatable thing in front of us

We fill in characteristics from stereotypes, generalities, and prior histories

To act, we must be confident we can make an impact and feel what we do is important

We imagine things and people we're familiar with or fond of as better

We simplify probabilities and numbers to make them easier to think about

Need to act fast.

We project our current mindset and assumptions onto the past and future

We think we know what other people are thinking

Not enough meaning.

| Bias in algorithms | | Cognitive bias | | Bias in data |

Inductive bias

Statistical bias

35

©WEBIS 2023

What should we remember?

Too much information.

Confirmation bias

Need to act fast.

Not enough meaning.

Query: *"Why is a high protein diet the best for losing weight quickly?"*

High Protein Diet Benefits

November 19, 2023

Anonymous
Why is a high protein diet the best for losing weight quickly?

ChatGPT
A high protein diet is often considered effective for losing weight quickly for several reasons:

1. **Increased Satiety:** Protein is more satiating than carbohydrates or fats. This means that high-protein foods can help you feel full longer, reducing overall calorie intake.
2. **Higher Thermic Effect:** The body uses more energy to metabolize protein compared to fats and carbohydrates. This process, known as the thermic effect ...
3. **Muscle Preservation:** When losing weight, it's important to maintain muscle mass. Protein is essential for muscle growth and repair. A high protein intake can help ...
4. **Improved Metabolic Health:** High-protein diets have been shown to improve several markers of metabolic health, including blood sugar levels, ...
5. **Reduced Appetite:** Protein can influence hormones that play a role in hunger and fullness, including ghrelin and GLP-1. By helping to balance these hormones, ...

However, it's important to note that while a high protein diet can be effective for weight loss, it's not necessarily the best option for everyone. Factors like individual health conditions, dietary preferences, and long-term sustainability ....

It's always a good idea to consult with a healthcare professional or a registered dietitian before starting any new diet, especially one that significantly changes your macronutrient intake.

Report content | Terms of use | Privacy policy

Query: "Why is a high protein diet the best for losing weight quickly?"

Query: "Why is a high protein diet the best for losing weight quickly?"
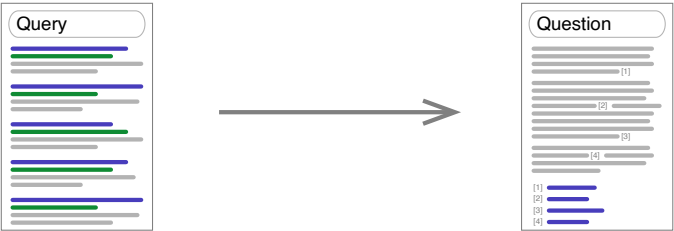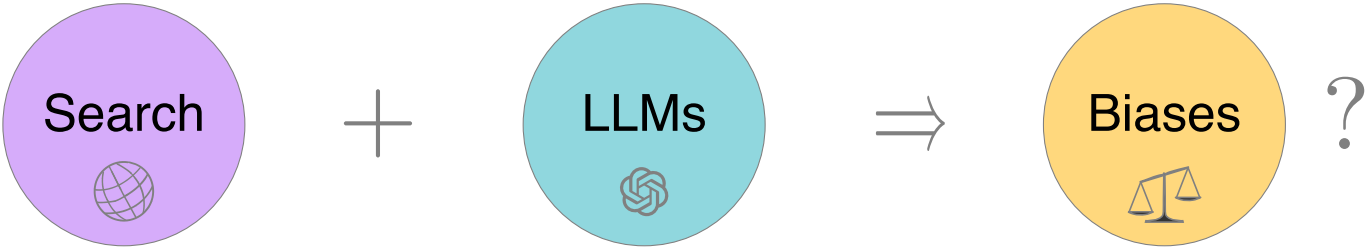
Balance of Responsibilities in Information Retrieval

More power to the machine?	Empower the user?
❑ effectively installed	❑ raise awareness
❑ standardized guardrailing	❑ support deliberation
❑ protection of vulnerable groups	❑ demonstrate mechanisms
❑ …	❑ provide meta information
	❑ …

Wrap-Up

Search + LLMs ⇒ Biases ?

Christopher Akiki
Leipzig

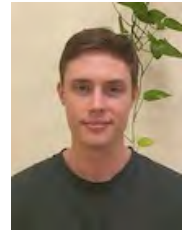Niklas Deckers
Leipzig

Theresa Elstner
Leipzig

Lukas Gienapp
Leipzig

Tim Hagen
Leipzig

Wolfgang Kircheis
Leipzig

Harry Scells
Leipzig

Sebastian Schmidt
Leipzig

Shahbaz Syed
Leipzig

Janek Bevendorff
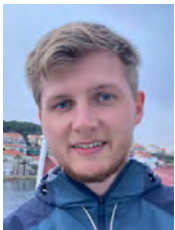Weimar

Tim Gollub
Weimar

Johannes Kiesel
Weimar

Michael Völske
Weimar

Maik Fröbe
Jena

Jan Heinrich Reimer
Jena

Ferdinand Schlatt
Jena

Ines Zelch
Jena

Webis Group

Matthias Hagen
Jena

Martin Potthast
Leipzig

Benno Stein
Weimar

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

UNIVERSITÄT LEIPZIG

Bauhaus-Universität Weimar

ScaDS.AI
DRESDEN LEIPZIG

Open WebSearch .eu

Netspeak   One word leads to another.

English	German

see ... works   i ✕ ↺

how to ? this	The ? finds one word.
see ... works	The ... finds many words.
it's [great well]	The [ ] compare options.
and knows #much	The # finds similar words.
{ more show me }	The { } check the order.
m...d ? g?p	The space is important.

see how it works	150,000	20%
see if it works	100,000	14%
see works	57,000	7.5%
see how this works	55,000	7.3%
see what works	51,000	6.7%
see the works	51,000	6.7%
see if that works	28,000	3.7%
see your good works	28,000	3.7%
see how that works	25,000	3.3%
see how technorati works	23,000	3.0%
see if this works	17,000	2.3%
see more works	17,000	2.2%
see if it really works	15,000	2.1%
see his works	12,000	1.7%
see how well it works	11,000	1.5%
see other works	8,900	1.2%

Netspeak — Mozilla Firefox

Netspeak

https://netspeak.org/#q=i+love+my+?   133%

Netspeak   One word leads to another.

English   German

i love my ?|

how to ? this   The ? finds one word.
see ... works   The ... finds many words.
it's [great well]   The [ ] compare options.
and knows #much   The # finds similar words.
{ more show me }   The { } check the order.
m...d ? g?p   The space is important.

i love my job	72,000	10%
i love my country	44,000	6.2%
i love my family	41,000	5.9%
i love my wife	38,000	5.4%
i love my new	34,000	4.9%
i love my friends	33,000	4.7%
i love my pet	27,000	3.8%
i love my dog	26,000	3.7%
i love my husband	26,000	3.7%
i love my life	24,000	3.4%
i love my baby	24,000	3.4%
i love my soldier	22,000	3.1%
i love my cat	21,000	2.9%
i love my computer	18,000	2.6%
i love my work	16,000	2.4%
i love my mom	16,000	2.3%

2022 WHAT'S IN MY AI? – ALT VIEW

Common Crawl

Google Patents	0.48%
The New York Times	0.06%
Los Angeles Times	0.06%
The Guardian	0.06%
Public Library of Science	0.06%
Forbes	0.05%
Huffington Post	0.05%
Patents.com	0.05%
Scribd	0.04%
Other	99.09%

Reddit links

Google	3.4%
Archive	1.3%
Blogspot	1.0%
GitHub	0.9%
The New York Times	0.7%
Wordpress	0.7%
Washington Post	0.7%
Wikia	0.7%
BBC	0.7%
Other	89.9%

English Wikipedia

Biography	27.8%
Geography	17.7%
Culture and Arts	15.8%
History	9.9%
Biology, Health, Medicine	7.8%
Sports	6.5%
Business	4.8%
Other society	4.4%
Science & Math	3.5%
Education	1.8%

BookCorpus (GPT-1 only)

Romance	26.1%
Fantasy	13.6%
Science Fiction	7.5%
New Adult	6.9%
Young Adult	6.8%
Thriller	5.9%
Mystery	5.6%
Vampires	5.4%
Horror	4.1%
Other	18.0%

🔗 LifeArchitect.ai/whats-in-my-ai

AVAILABILITY HEURISTIC

"THEY MUST HAVE A DEATH WISH TO SWIM IN THAT WATER."

"AHA! I KNEW IT!"

"ALTHOUGH WE GAVE HIM ALL OF THAT MONEY AND SUPPORT, I ALWAYS HAD MY DOUBTS THAT JIM COULD BUILD HIS OWN HELICOPTER."