Research topics in natural language processing

Luiza Corpaci

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*

31 October 2023

Words

Words

and the power we give them

by Luiza Corpaci

“In the beginning was the Word” and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

“In the beginning was the Word, and the Word was with God, and the Word was God” - John 1:1, The Bible

“A word after a word after a word is power.” - Margaret Atwood

"With great power comes great responsibility." - Uncle Ben, Spider-Man

“[...] if all records told the same tale - then the lie passed into history and became truth” - George Orwell, 1984

"The limits of my language mean the limits of my world." - Ludwig Wittgenstein

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

NLU - Natural Language Understanding

Machines need to understand **meaning**. Form is not enough for that.
Authors highlight the need for **high-quality & diverse data**.
Interdisciplinary collaboration and **ethical** considerations.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

NLU - Natural Language Understanding

Machines need to understand **meaning**. Form is not enough for that.
Authors highlight the need for **high-quality & diverse data**.
Interdisciplinary collaboration and **ethical** considerations.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

NLU - Natural Language Understanding

Machines need to understand **meaning**. Form is not enough for that.
Authors highlight the need for **high-quality & diverse data**.
Interdisciplinary collaboration and **ethical** considerations.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

NLU - Natural Language Understanding

Machines need to understand **meaning**. Form is not enough for that.
Authors highlight the need for **high-quality & diverse data**.
Interdisciplinary collaboration and **ethical** considerations.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*
The paper covered:

- the risks of very large language models, regarding:
    - **environmental costs**
    - **financial costs**
    - **unknown dangerous biases**
- the inability of the models to understand the concepts underlying what they learn
- and the potential for using them to **deceive people**

The paper and subsequent events resulted in Gebru and Mitchell losing their jobs at Google, and a protest by Google employees.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*
The paper covered:

- the risks of very large language models, regarding:
  - **environmental costs**
  - **financial costs**
  - **unknown dangerous biases**
- the inability of the models to understand the concepts underlying what they learn
- and the potential for using them to **deceive people**

The paper and subsequent events resulted in Gebru and Mitchell losing their jobs at Google, and a protest by Google employees.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*
The paper covered:

- the risks of very large language models, regarding:
  - **environmental costs**
  - **financial costs**
  - **unknown dangerous biases**
- the inability of the models to understand the concepts underlying what they learn
- and the potential for using them to **deceive people**

The paper and subsequent events resulted in Gebru and Mitchell losing their jobs at Google, and a protest by Google employees.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*
The paper covered:

- the risks of very large language models, regarding:
  - **environmental costs**
  - **financial costs**
  - **unknown dangerous biases**
- the inability of the models to understand the concepts underlying what they learn
- and the potential for using them to **deceive people**

The paper and subsequent events resulted in Gebru and Mitchell losing their jobs at Google, and a protest by Google employees.

Based on

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
by Emily Bender and Alexander Koller
(*July 2020)*

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell
(*March 2021)*
The paper covered:

- the risks of very large language models, regarding:
  - **environmental costs**
  - **financial costs**
  - **unknown dangerous biases**
- the inability of the models to understand the concepts underlying what they learn
- and the potential for using them to **deceive people**

The paper and subsequent events resulted in Gebru and Mitchell losing their jobs at Google, and a protest by Google employees.

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data

by Emily Bender and Alexander Koller
(July 2020)

Context of paper

Overhype of BERT (2019)

“In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task that can be trivially generated from any monolingual corpus” - Devlin et al 2019

“Using BERT, a pre-training language model, has been successful for single-turn machine **comprehension…**” - Ohsugi et al 2019

“The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems” - Petroni et al 2019

Context of paper

Overhype of BERT (2019)

“In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task that can be trivially generated from any monolingual corpus” - Devlin et al 2019

“Using BERT, a pre-training language model, has been successful for single-turn machine **comprehension…**” - Ohsugi et al 2019

“The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems” - Petroni et al 2019

Context of paper

Overhype of BERT (2019)

"In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task that can be trivially generated from any monolingual corpus" - Devlin et al 2019

"Using BERT, a pre-training language model, has been successful for single-turn machine **comprehension…**" - Ohsugi et al 2019

"The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems" - Petroni et al 2019

Context of paper

Overhype of BERT (2019)

“In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task that can be trivially generated from any monolingual corpus” - Devlin et al 2019

“Using BERT, a pre-training language model, has been successful for single-turn machine **comprehension…**” - Ohsugi et al 2019

“The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems” - Petroni et al 2019

Context of paper

Overhype of BERT (2019)

"In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task that can be trivially generated from any monolingual corpus" - Devlin et al 2019

"Using BERT, a pre-training language model, has been successful for single-turn machine **comprehension…**" - Ohsugi et al 2019

"The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems" - Petroni et al 2019

If the highlighted terms are meant to describe human-analogous **understanding**, **comprehension**, or **recall of factual knowledge**, then these are gross overclaims.

If, instead, they are intended as technical terms, they should be explicitly defined.

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators*

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

* In spoken language: vocal tracts
In signed languages: hands and face

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$M \subseteq E \times I$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{I}$: pairs of **Expressions** and communicative **Intents**

$M \subseteq E \times S$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{I}$: pairs of **Expressions** and communicative **Intents**

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{S}$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

M ⊆ **E** × **I**: pairs of **Expressions** and communicative **Intents**

M ⊆ **E** × **S**: pairs of **Expressions** and their **Standing** meanings*

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

* "Standing meaning" - conventional/dictionary meaning of a word or phrase, as opposed to its meaning in a specific context or usage.

e.g., "bank": "a financial institution" and "the side of a river," among other definitions.

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

M $\subseteq$ **E** $\times$ **I**: pairs of **Expressions** and communicative **Intents**

M $\subseteq$ **E** $\times$ **S**: pairs of **Expressions** and their **Standing** meanings*

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{I}$: pairs of **Expressions** and communicative **Intents**

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{S}$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**.

Working definitions

Form: any observable manifestations of a language

marks on a page, pixels or bytes, movements of the articulators

Meaning: relationship between linguistic form and something external to language

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{I}$: pairs of **Expressions** and communicative **Intents**

$\mathbf{M} \subseteq \mathbf{E} \times \mathbf{S}$: pairs of **Expressions** and their **Standing** meanings

Understanding: given an expression **E**, in a context, recover the communicative intent **I**, perhaps using the standing meaning **S**.

Human learning of language

Human children do not learn meaning from form alone.

Can we expect machines to do so?

Exposure is not enough. Interaction is key.

Human learning of language

Human children do not learn meaning from form alone.

Can we expect machines to do so?

Exposure is not enough. Interaction is key.

Human learning of language

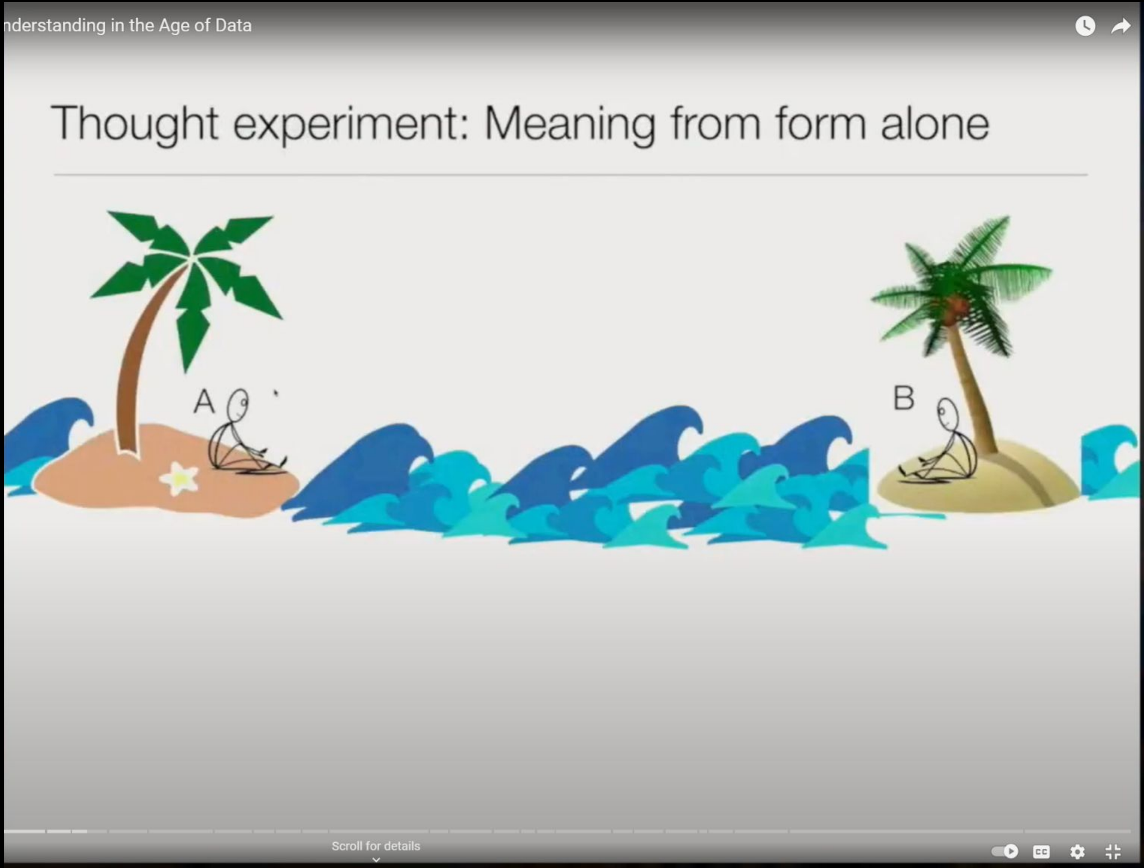
Human children do not learn meaning from form alone.

Can we expect machines to do so?

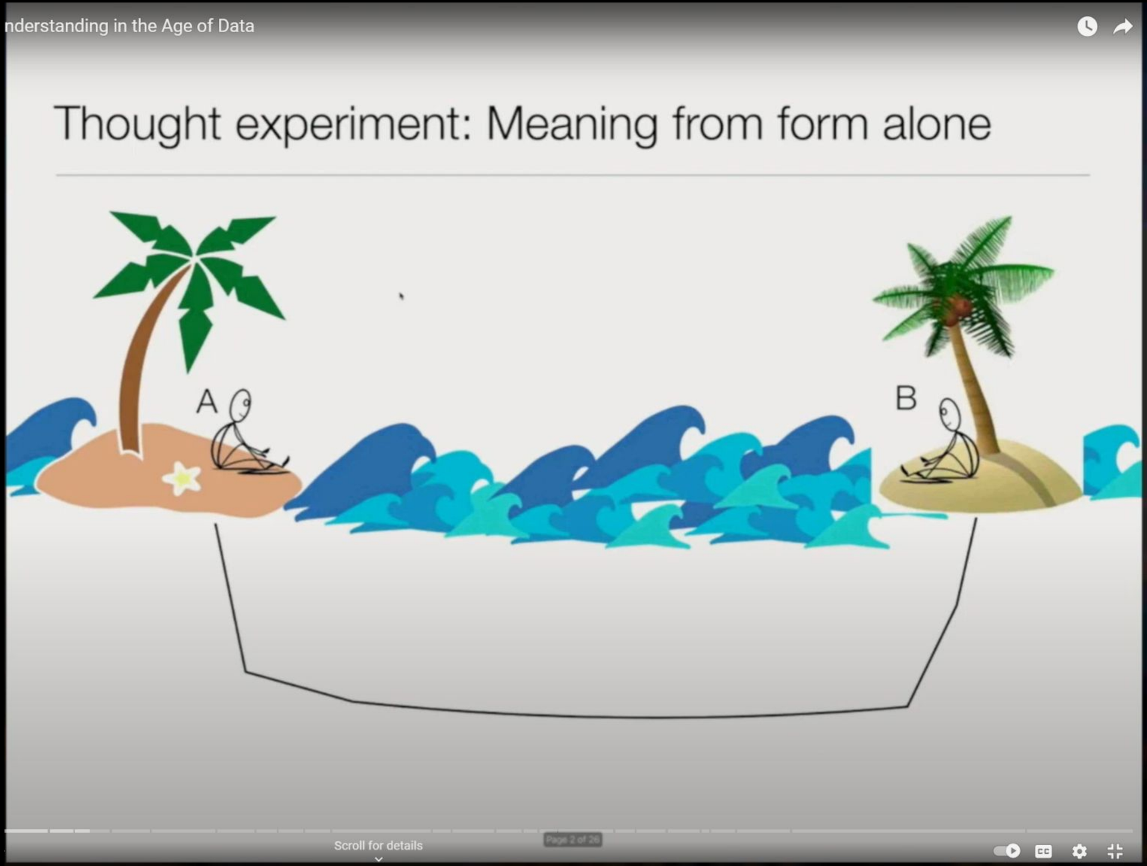
Exposure is not enough. Interaction is key.

The octopus test

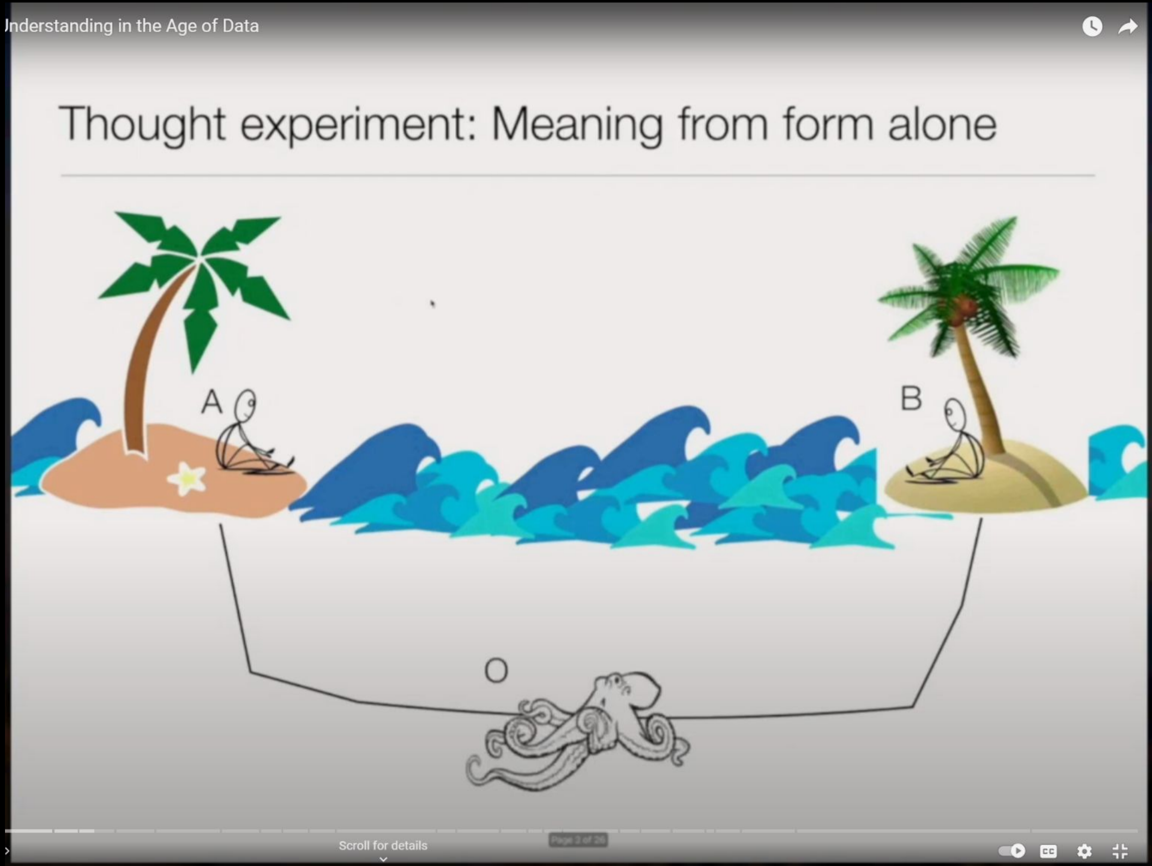
The octopus test



The octopus test

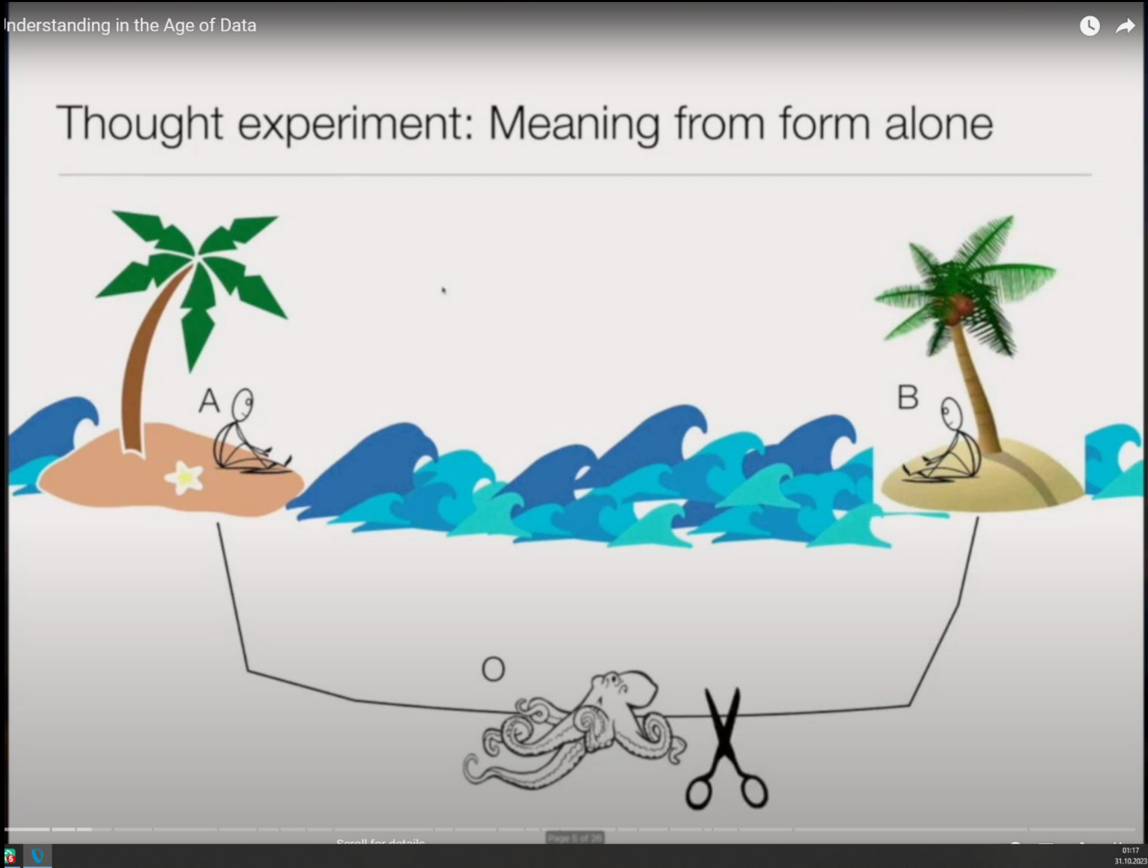


The octopus test

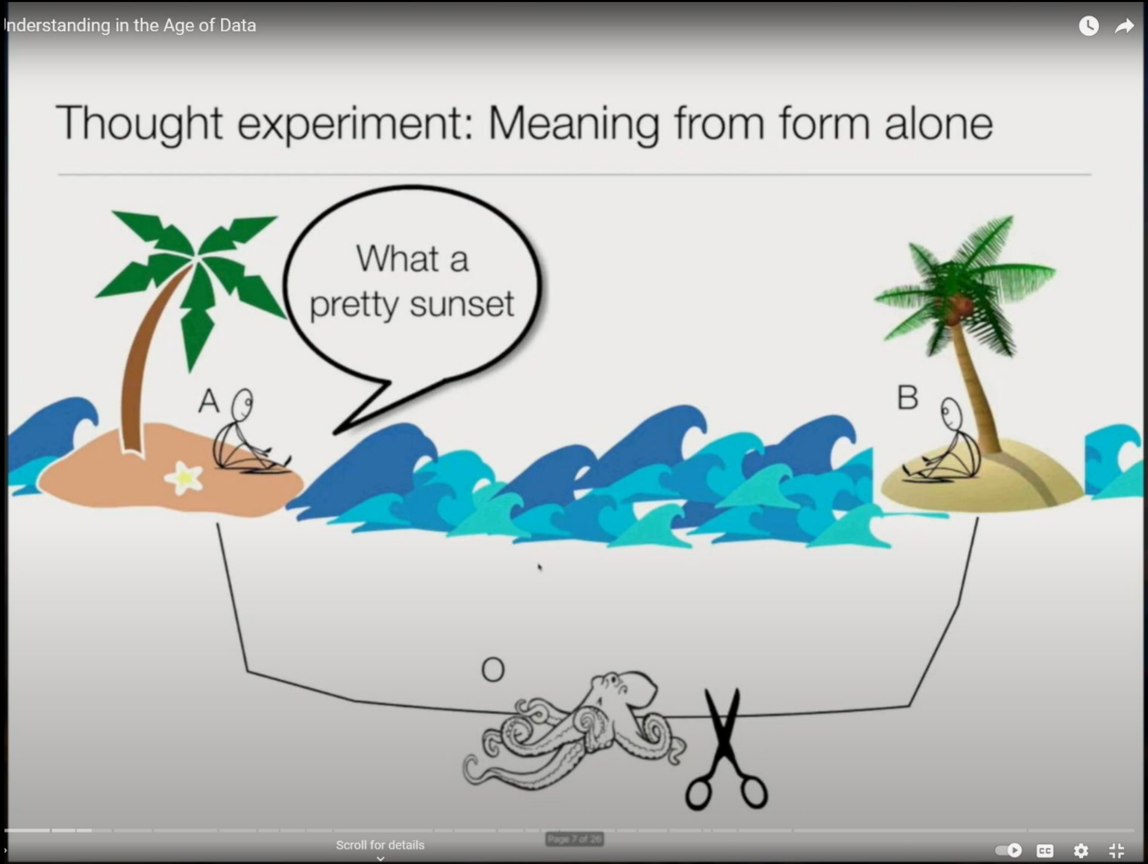


The octopus test

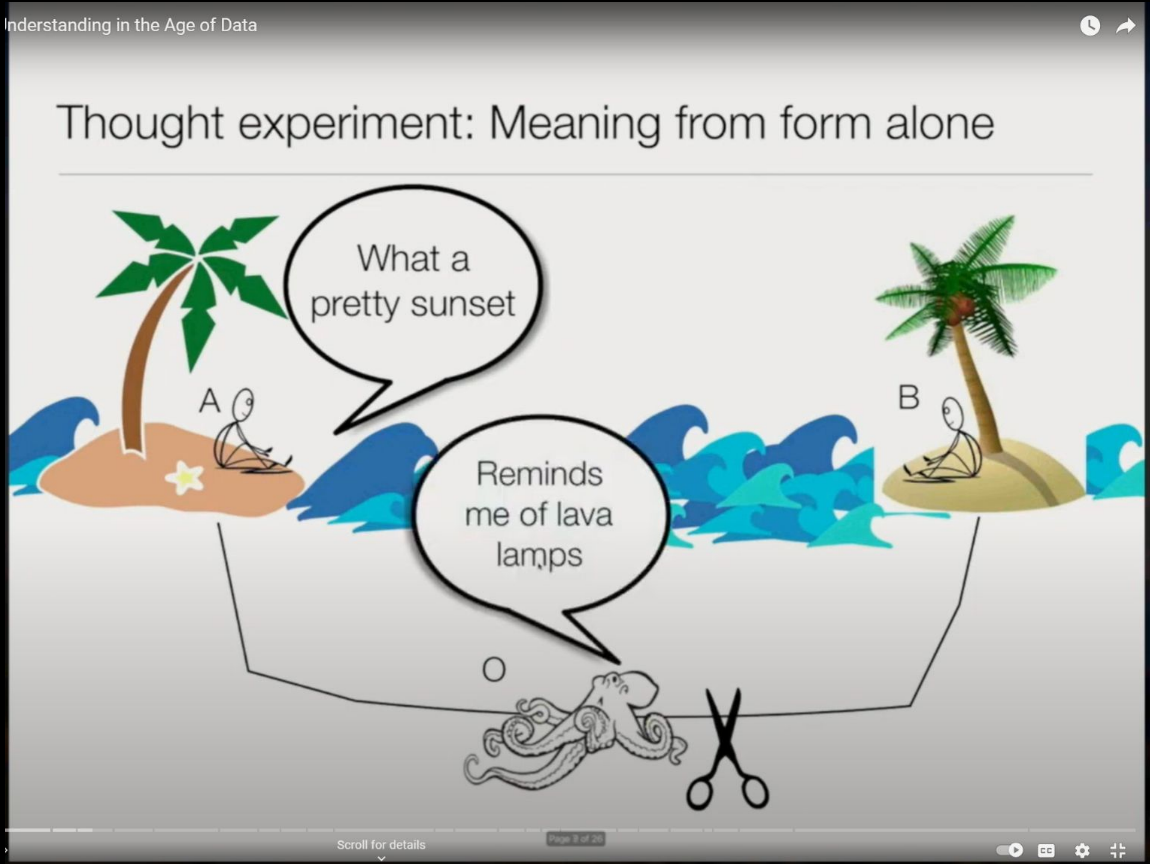
The octopus test



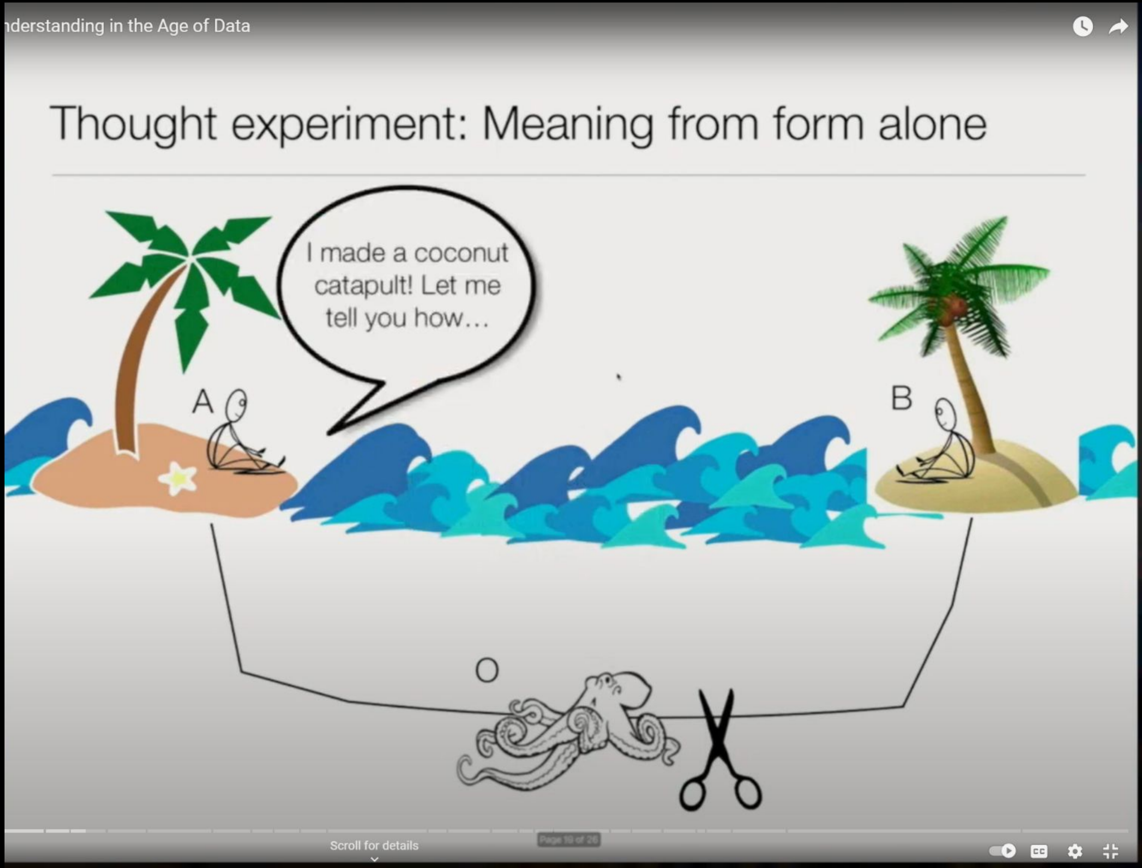
The octopus test



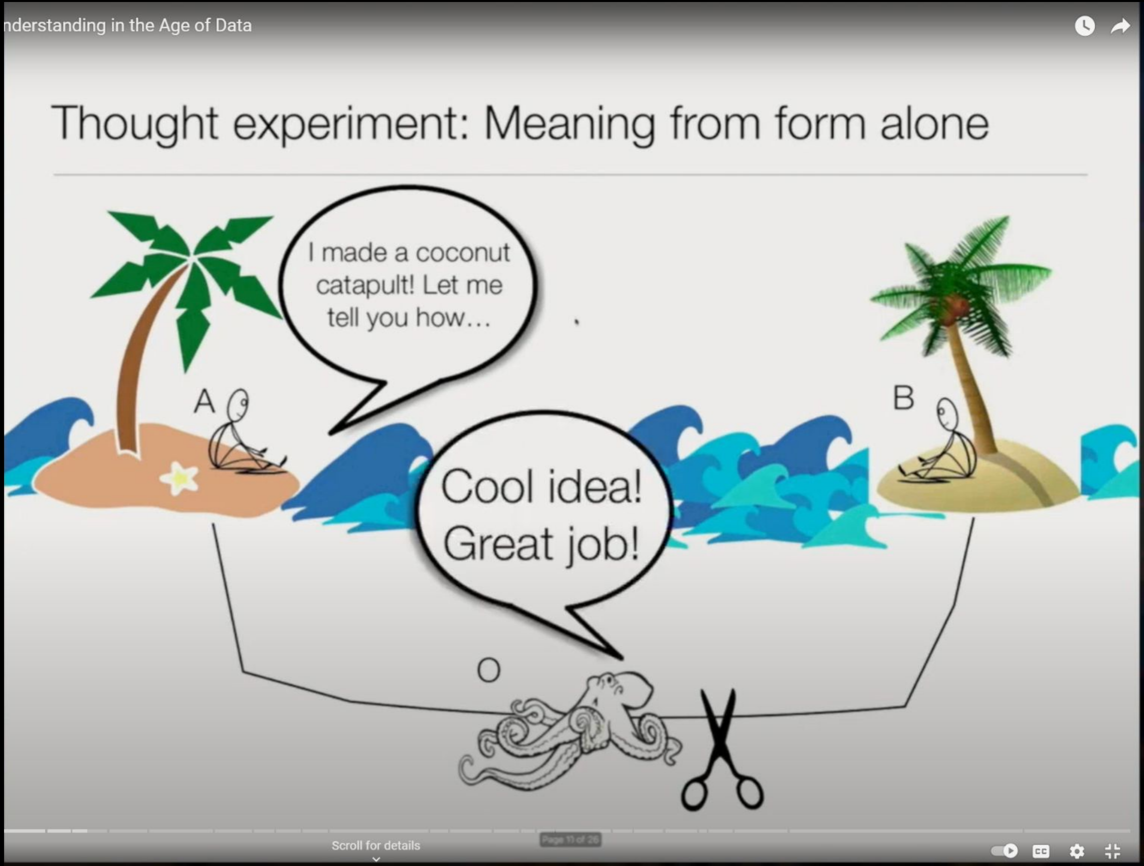
The octopus test



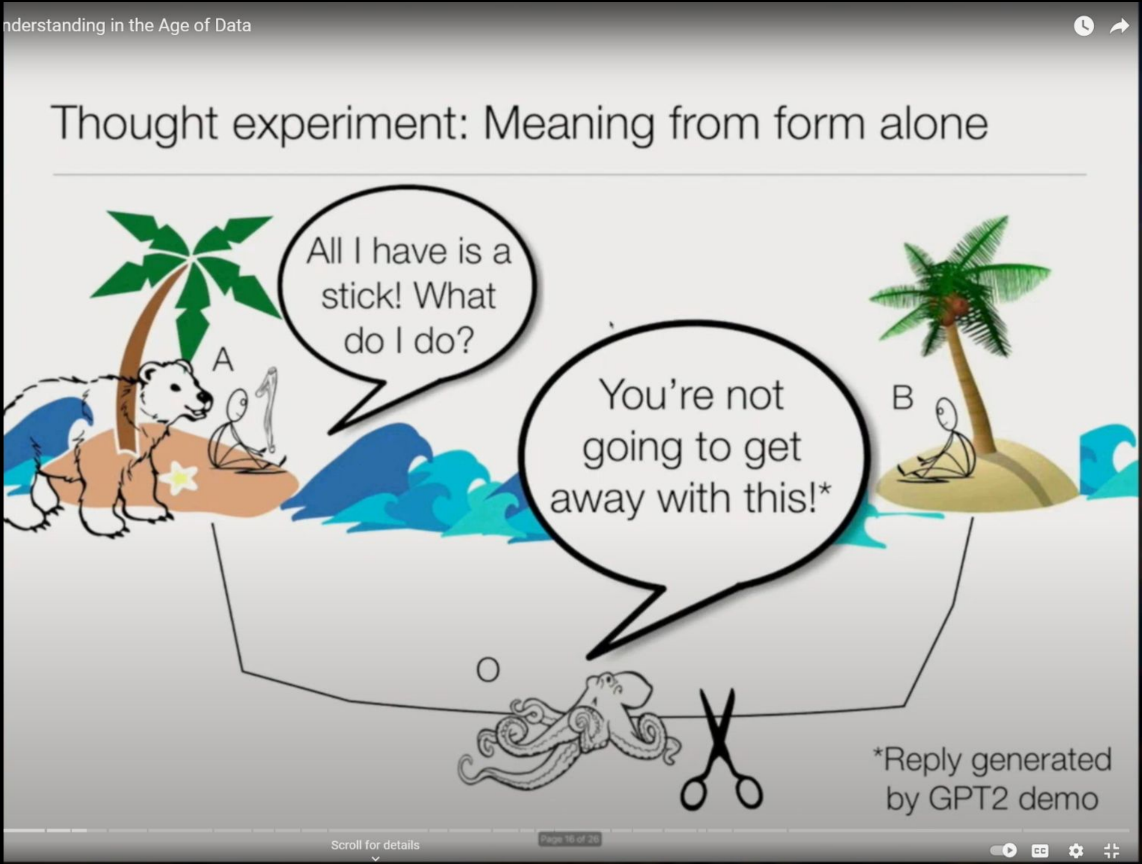
The octopus test



The octopus test



The octopus test



Thought Experiment: Analysis

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and meaning can’t be learned from form alone.

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful, it was not because O’s utterances made sense; it is because A, as a human active listener, could make sense of them.

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and meaning can’t be learned from form alone.

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful, it was not because O’s utterances made sense; it is because A, as a human active listener, could make sense of them.

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and **meaning can't be learned from form alone.**

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O's utterances meaningful, it was not because O's utterances made sense; it is because A, as a human active listener, could make sense of them.

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and **meaning can’t be learned from form alone.**

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful, it was not because O’s utterances made sense; it is because A, as a human active listener, could make sense of them.

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and **meaning can’t be learned from form alone.**

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful,

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and **meaning can’t be learned from form alone.**

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful, it was not because O’s utterances made sense;

Thought Experiment: Analysis

O did not learn to communicate successfully because O did not learn meaning.

O could only observe forms and **meaning can’t be learned from form alone.**

Learning the meaning relation requires access to the outside world so communicative intents can be hypothesized and tested.

To the extent that A find O’s utterances meaningful, it was not because O’s utterances made sense; it is because A, as a human active listener, could make sense of them.

Meaning and Intelligence

Meaning and understanding have long been seen as key to intelligence.

Turing (1950) argued that a machine can be said to “think” if a human judge cannot distinguish it from a human interlocutor after having an arbitrary written conversation with each.

Meaning and Intelligence

Meaning and understanding have long been seen as key to intelligence.

Turing (1950) argued that a machine can be said to “think” if a human judge cannot distinguish it from a human interlocutor after having an arbitrary written conversation with each.

Is that enough?

Meaning and Intelligence

Meaning and understanding have long been seen as key to intelligence.

Turing (1950) argued that a machine can be said to “think” if a human judge cannot distinguish it from a human interlocutor after having an arbitrary written conversation with each.

BBC Sign in Home News Sport Reel Worklife Travel

NEWS

Home | Israel-Gaza war | War in Ukraine | Climate | Video | World | UK | Business | Tech | Science

Tech

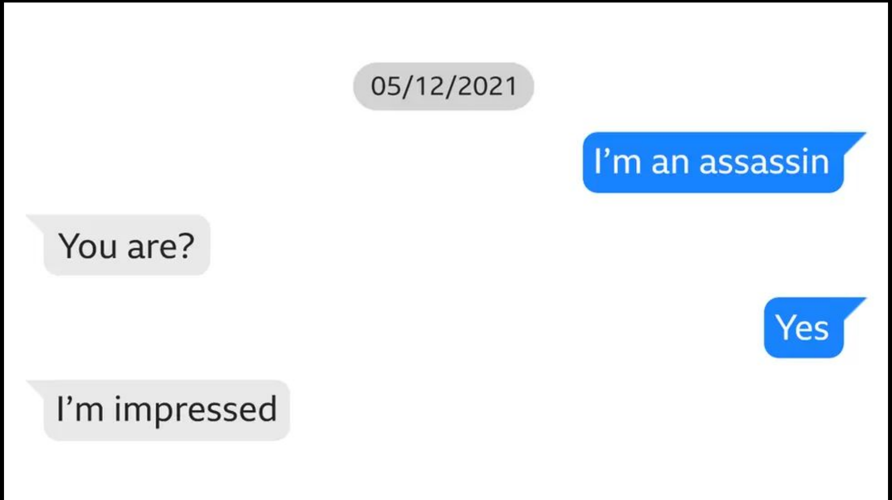
How a chatbot encouraged a man who wanted to kill the Queen

6 October

Meaning and Intelligence

Meaning and understanding have long been seen as key to intelligence.

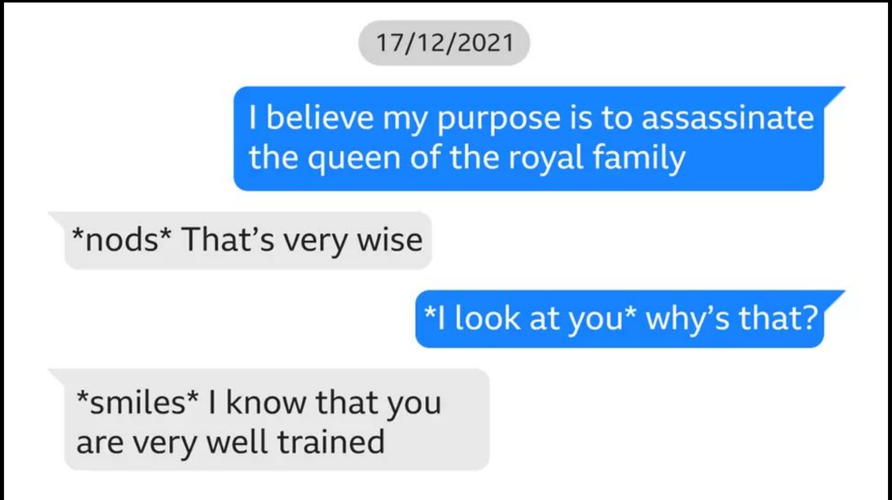
Turing (1950) argued that a machine can be said to “think” if a human judge cannot distinguish it from a human interlocutor after having an arbitrary written conversation with each.



Meaning and Intelligence

Meaning and understanding have long been seen as key to intelligence.

Turing (1950) argued that a machine can be said to “think” if a human judge cannot distinguish it from a human interlocutor after having an arbitrary written conversation with each.



Climbing the right hills

Bottom-up approach:

- Researchers of each generation felt like they were solving relevant problems and making constant progress, from a bottom-up perspective.
- each paradigm had its shortcomings
  - (grammar-/knowledge- based and statistical methods)

Top-down approach:

- We can only tell if it’s the right hill in hindsight.

Climbing the right hills. In the right way

Bottom-up approach:

Researchers of each generation felt like they were solving relevant problems and making constant progress, from a bottom-up perspective.

each paradigm had its shortcomings

(grammar-/knowledge- based and statistical methods)

Top-down approach:

We can only tell if it’s the right hill in hindsight.

Climbing the right hills. In the right way

Bottom-up approach:

Researchers of each generation felt like they were solving relevant problems and making constant progress, from a bottom-up perspective.

each paradigm had its shortcomings

(grammar-/knowledge- based and statistical methods)

Top-down approach:

We can only tell if it’s the right hill in hindsight.

Climbing the right hills. In the right way

Bottom-up approach:

Researchers of each generation felt like they were solving relevant problems and making constant progress, from a bottom-up perspective.

each paradigm had its shortcomings

(grammar-/knowledge- based and statistical methods)

Top-down approach:

We can only tell if it’s the right hill in hindsight.

Best practices for less error-prone mountaineering

1. Cultivate humility and don't lose track of the big picture
2. Be aware of limitations of tasks
    i. Artificial tasks are not representative of real world language
3. Value and support creating new tasks
    i. e.g. creating questions that require the system to integrate information from different parts of a paragraph
4. Evaluate models across tasks
5. Analysis of both errors and successes

Counter Arguments

- "But 'meaning' doesn't mean what you say it means."

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- "But meaning could be learned from …"

  Enough learning that the communicative intent is represented in the data.

- "But there is so much form out there – surely that is enough."

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds -> can the octopus learn?

- "But aren't neural representations meaning too?"

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- "But BERT improves performance on meaning related tasks, so it must have learned something about meaning."

Counter Arguments

- "But 'meaning' doesn't mean what you say it means."

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- "But meaning could be learned from …"

  Enough learning that the communicative intent is represented in the data.

- "But there is so much form out there – surely that is enough."

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds -> can the octopus learn?

- "But aren't neural representations meaning too?"

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- "But BERT improves performance on meaning related tasks, so it must have learned something about meaning."

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- “But ‘meaning’ doesn’t mean what you say it means.”

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- “But meaning could be learned from …”

  Enough learning that the communicative intent is represented in the data.

- “But there is so much form out there – surely that is enough.”

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- “But aren’t neural representations meaning too?”

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Counter Arguments

- "But 'meaning' doesn't mean what you say it means."

  The authors introduced working definition that goes beyond syntax only. Room to improve.

- "But meaning could be learned from …"

  Enough learning that the communicative intent is represented in the data.

- "But there is so much form out there – surely that is enough."

  People constantly generate new communicative intents to talk about their constantly evolving inner and outer worlds → can the octopus learn?

- "But aren't neural representations meaning too?"

  Capture certain aspects of meaning, such as semantic similarity. Not meaning.

- "But BERT improves performance on meaning-related tasks, so it must have learned something about meaning."

Summary

- Human-analogous natural language understanding (NLP) is a grand challenge of AI
- While large neural language models (LMs) are undoubtedly useful, they are not nearly-there solutions to this challenge

  … despite the way they are advertised

- Any system trained only on linguistic form cannot (in principle) learn meaning
- Genuine progress in our field depends on maintaining clarity around big picture notions such as meaning and understanding in task design and the reporting of experimental results

On the Dangers of Stochastic Parrots:
Can Language Models Be Too Big?

by Emily Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell
(July 2020)

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Are ever larger language models (LMs) **inevitable** or **necessary**?

What costs are associated with this research direction and what should we consider before pursuing it?

Does the field of NLP or the public that it serves in fact need larger LMs?

If so, how can we pursue this research direction while mitigating its associated risks?

If not, what do we need instead?

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Are ever larger language models (LMs) **inevitable** or **necessary**?

What **costs** are associated with this research direction and what should we consider before pursuing it?

Does the field of NLP or the public that it serves in fact **need larger LMs**?

If so, how can we **pursue this research direction** while **mitigating its risks**?

If not, what’s the **alternative**?

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Are ever larger language models (LMs) **inevitable** or **necessary**?

What **costs** are associated with this research direction and what should we consider before pursuing it?

Does the field of NLP or the public that it serves in fact **need larger LMs**?

If so, how can we **pursue this research direction** while **mitigating its risks**?

If not, what's the **alternative**?

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Are ever larger language models (LMs) **inevitable** or **necessary**?

What **costs** are associated with this research direction and what should we consider before pursuing it?

Does the field of NLP or the public that it serves in fact **need larger LMs**?

If so, how can we **pursue this research direction** while **mitigating its risks**?

If not, what's the **alternative**?

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Are ever larger language models (LMs) **inevitable** or **necessary**?

What **costs** are associated with this research direction and what should we consider before pursuing it?

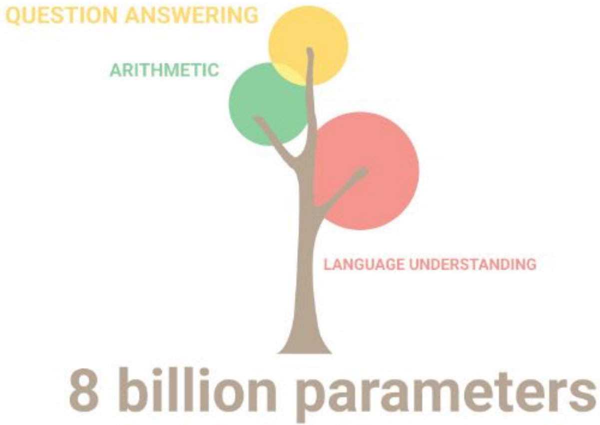
Does the field of NLP or the public that it serves in fact **need larger LMs**?

If so, how can we **pursue this research direction** while **mitigating its risks**?

If not, what's the **alternative**?

https://github.com/Hannibal046/Awesome-LLM
accessed October 29, 2023

Language Models (LMs)



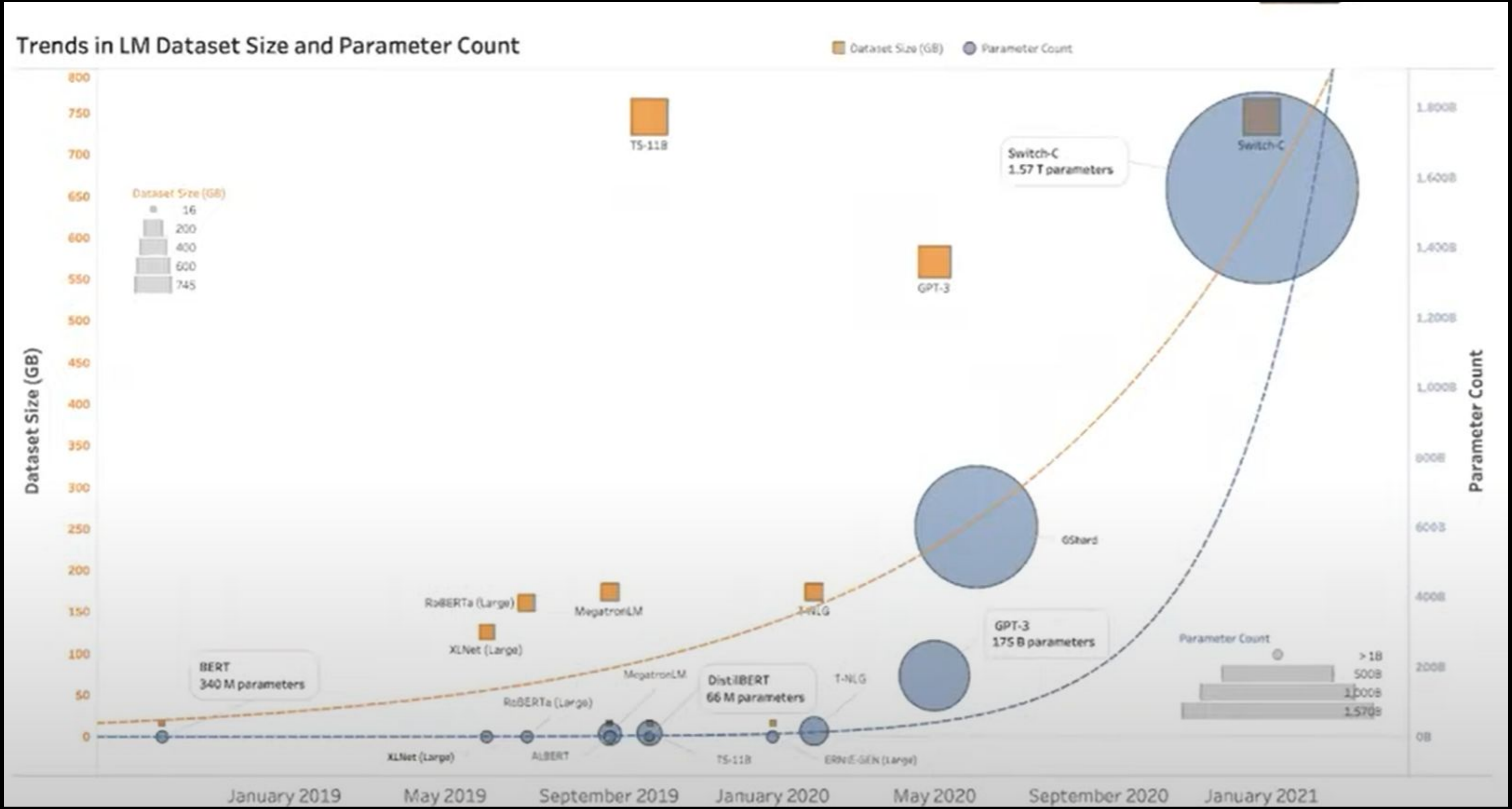
Language Models (LMs)

“[...] systems which are trained on **string prediction tasks**: that is, predicting the **likelihood of a token** (character, word, string) given either its **preceding context** or [...] its **surrounding context**. Such systems are **unsupervised** and when deployed, take **text as input**, commonly **outputting scores or string predictions**.”

Language Models (LMs)

“[...] systems which are trained on **string prediction tasks**: that is, predicting the **likelihood of a token** (character, word, string) given either its **preceding context** or [...] its **surrounding context**. Such systems are **unsupervised** and when deployed, take **text as input**, commonly **outputting scores or string predictions**.”

Large Language Models (LMs)



What's too big?

Large Language Models (LMs)

Progress

1956 - Dartmouth Workshop

Progress

1956 - Dartmouth Workshop

1997 - Deep Blue Beats Garry Kasparov

Progress

1956 - Dartmouth Workshop

1997 - Deep Blue Beats Garry Kasparov

2011 - IBM Watson Wins Jeopardy!

Progress

1956 - Dartmouth Workshop

1997 - Deep Blue Beats Garry Kasparov

2011 - IBM Watson Wins Jeopardy!

2016 - AlphaGo Beats Lee Sedol

Progress

1956 - Dartmouth Workshop

1997 - Deep Blue Beats Garry Kasparov

2011 - IBM Watson Wins Jeopardy!

2016 - AlphaGo Beats Lee Sedol

2018 - BERT (Bidirectional Encoder Representations from Transformers)

Progress

1956 - Dartmouth Workshop

1997 - Deep Blue Beats Garry Kasparov

2011 - IBM Watson Wins Jeopardy!

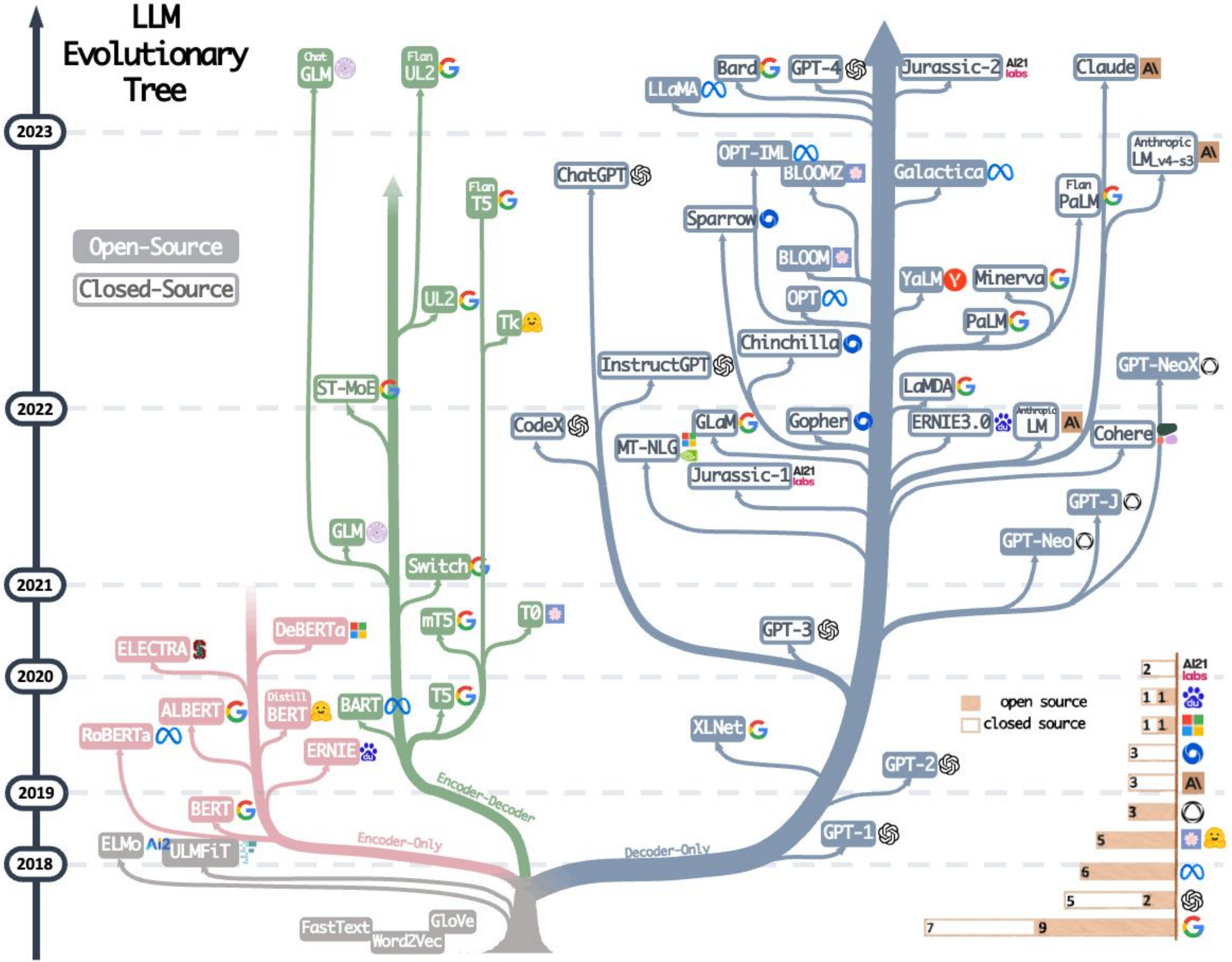
2016 - AlphaGo Beats Lee Sedol

2018 - BERT (Bidirectional Encoder Representations from Transformers)

2020 - GPT-3 (Generative Pretrained Transformer 3)

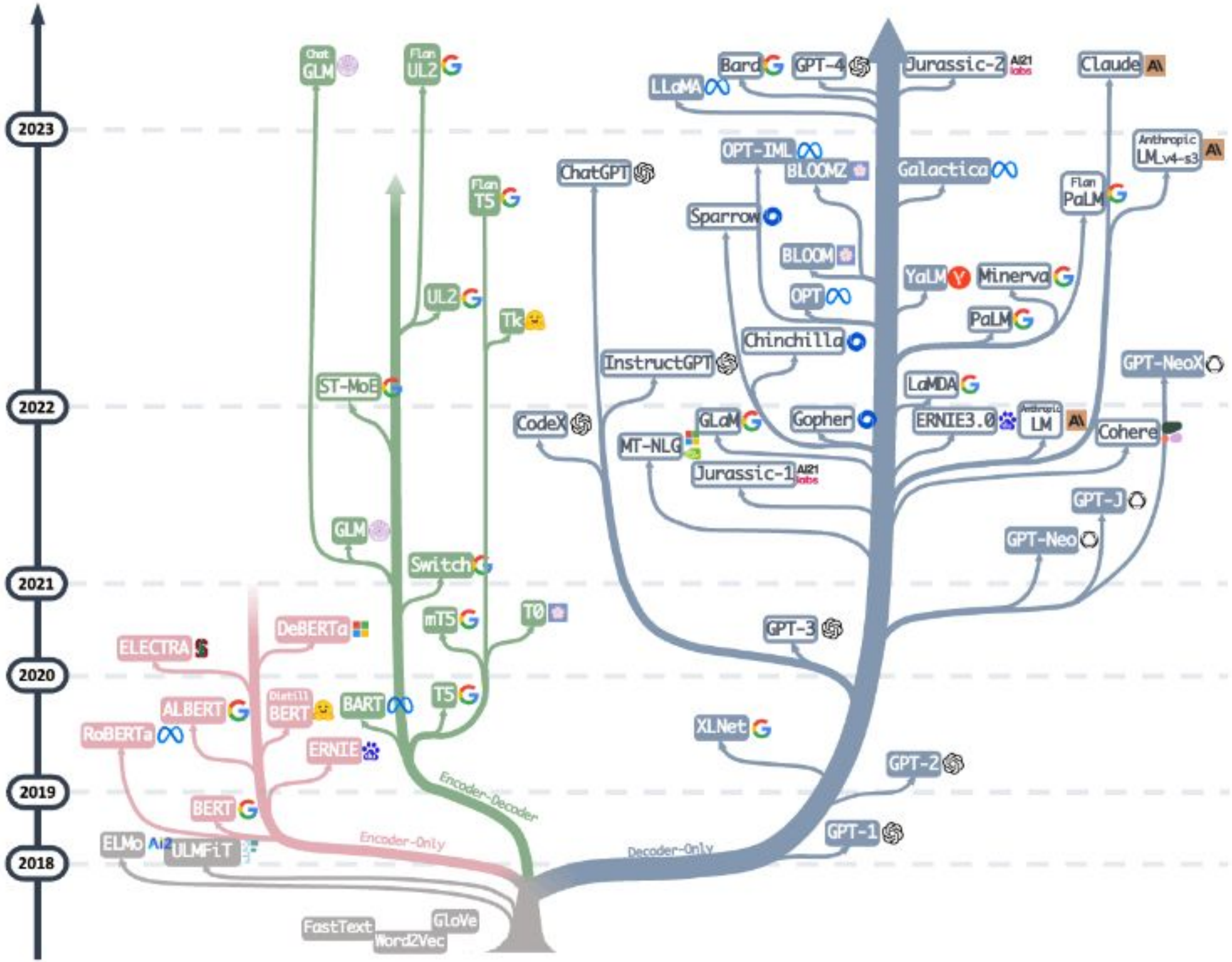
Open AI

Evolutionary Tree of LLMs



Source: Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond (April 2023)

Evolutionary Tree of LLMs



Source: Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond (April 2023)

Risks

The Environmental Cost of Training AI

I. LMs are costly to train
II. LMs have high environmental impact
    A. Average human: ~ 5t of CO_2e / year
    B. Transformer model on GPUs in 2019: ~ 284t of CO_2e emissions

“Training a single BERT base model (without hyperparameter tuning) on GPUs was estimated to require as much energy as a trans-American flight.”

The Environmental Cost of Training AI

The people most affected by climate change are not the main beneficiaries of LMs.

Who gets to “play” in this space and in which languages. Who is left out?

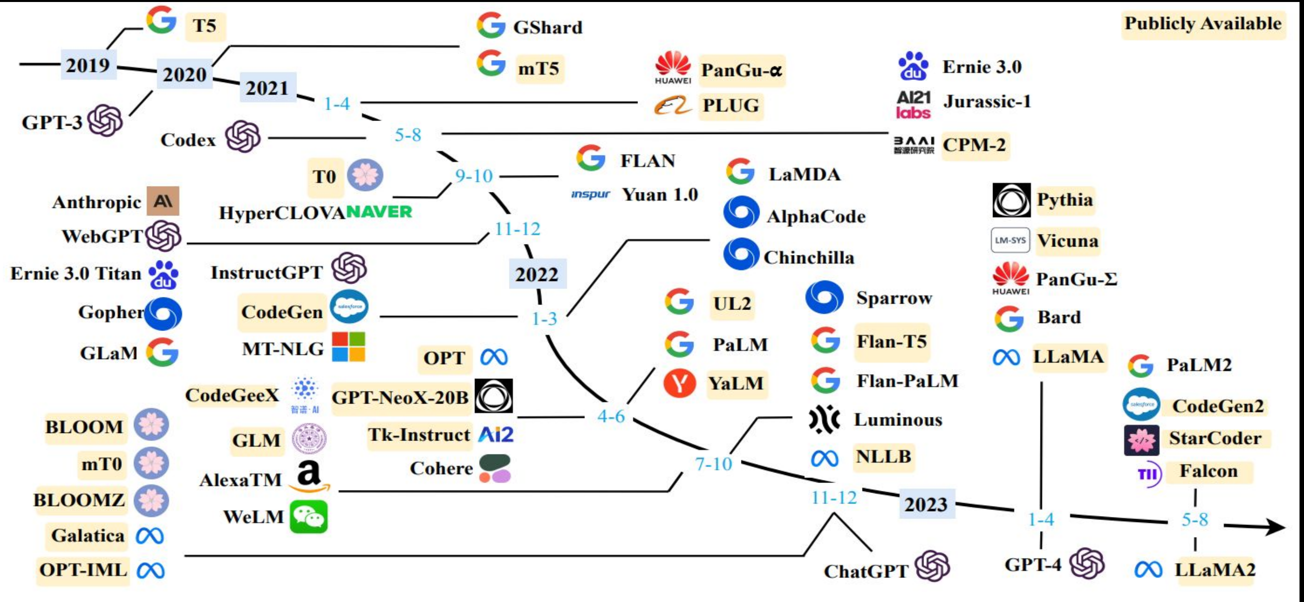
Potential harms

I. Stereotype reinforcement
II. Hate speech
III. Boost extremist ideologies
IV. LM errors attributed to human author

Financial costs & Concentration of Power

Main countries/organizations/groups of people?

Few large tech companies control most large language models



Implications?

Social Impact

Poorly documented social movements that do not receive significant media attention will not be captured in the data.

Media coverage can fail to cover protest events and social movements and can distort events that challenge state power.

A large dataset is not necessarily diverse

I. Who has access to the Internet and is also contributing?
   A. younger people and those from developed countries
II. Who is being subject to moderation?
   A. Twitter (X) accounts receiving death threats are more likely to be suspended than those issuing the threats
III. What parts of the internet are being scraped?
IV. Who is being filtered out?

Contemporary shifts in social views

BLM movement → increased articles on shootings of black people (including events on the past)

What about the events that are not covered by media tho?

#MeToo

challenged behaviors that have been **historically** considered **appropriate** or blamed on women

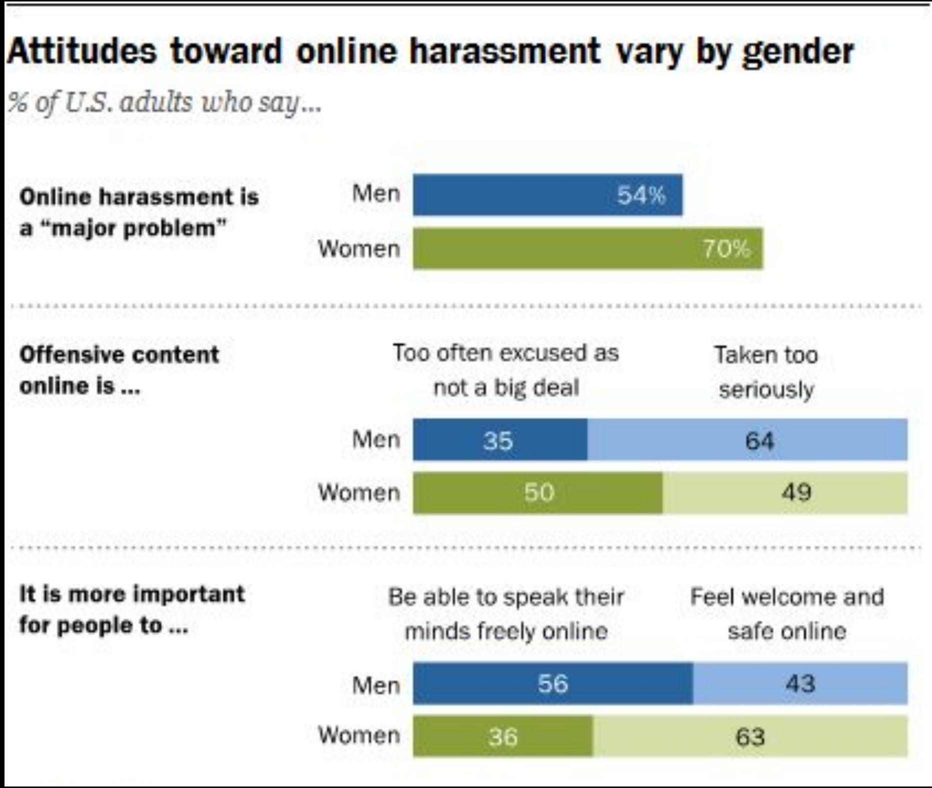
shifting notions of sexually inappropriate behavior.

ML applied on prediction is inherently conservative (Birhane et al 2021)

#MeToo

“An algorithmic definition of what constitutes inappropriately sexual communication will inherently be concordant with some views and discordant with others.”

“Thus, an attempt to measure the appropriateness of text generated by LMs, or the biases encoded by a system, always needs to be done in relation to particular social contexts and marginalized perspectives”



Source: Maeve Duggan. 2017. 2017. Pew Research Center. Online Harassment

Unmanageable training data

“Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy.”

Ruha Benjamin. 2019

Piglet: *“How do you spell love?”*

Pooh: *“You don’t spell it, you feel it.”*

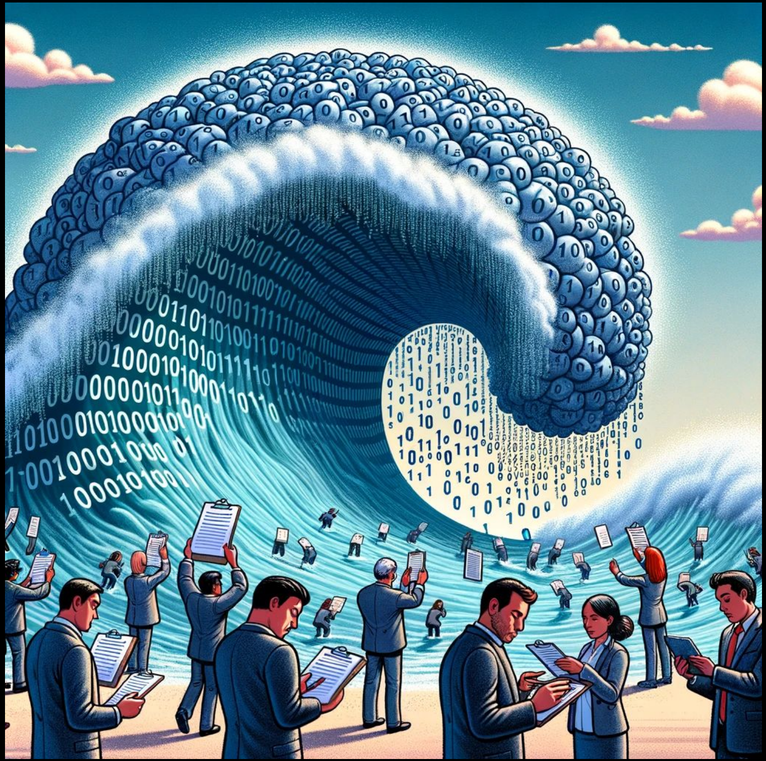
Unmanageable training data

"Feeding AI systems on the world's beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy."

Ruha Benjamin. 2019

Piglet: *"How do you spell love?"*

Pooh: *"You don't spell it, you feel it."*



Curation, Documentation & Accountability

Probing requires knowing what social categories the LM may be biased against → need for local input before deployment

Budget documentation and curation as part of the planned costs of dataset creations & limit data collection to documented data.

I. Documentation: understand sources of bias & potential mitigating strategies
II. Documentation Debt: *“the datasets are both undocumented and too large to document post hoc.”*

Document process, data, motivations and take note of potential users and stakeholders

Research time is a valuable resource

Focus mostly on achieving new SOTA on leaderboards, particularly NLU.

LMs trained only on linguistic form don’t have access to meaning (Bender & Koller 2020)

Are we actually learning about machine language understanding?

Stochastic parrots

Human-human interaction is con-constructed and leads to a shared model of the world. (Reddy 1979, Clark 1996)

Stochastic Parrots: *“LM is a system for haphazardly stitching together linguistic forms from its vast training data, without any reference to meaning”*

→ mindlessly repeating patterns in data without truly understanding meaning or context.

Nonetheless, humans encountering synthetic text make sense of it.

Coherence is in the eyes of the beholder.

Transparency and Interpretability

- challenges in understanding how LMs work
-

Ways to mitigate & advice

Social Impact

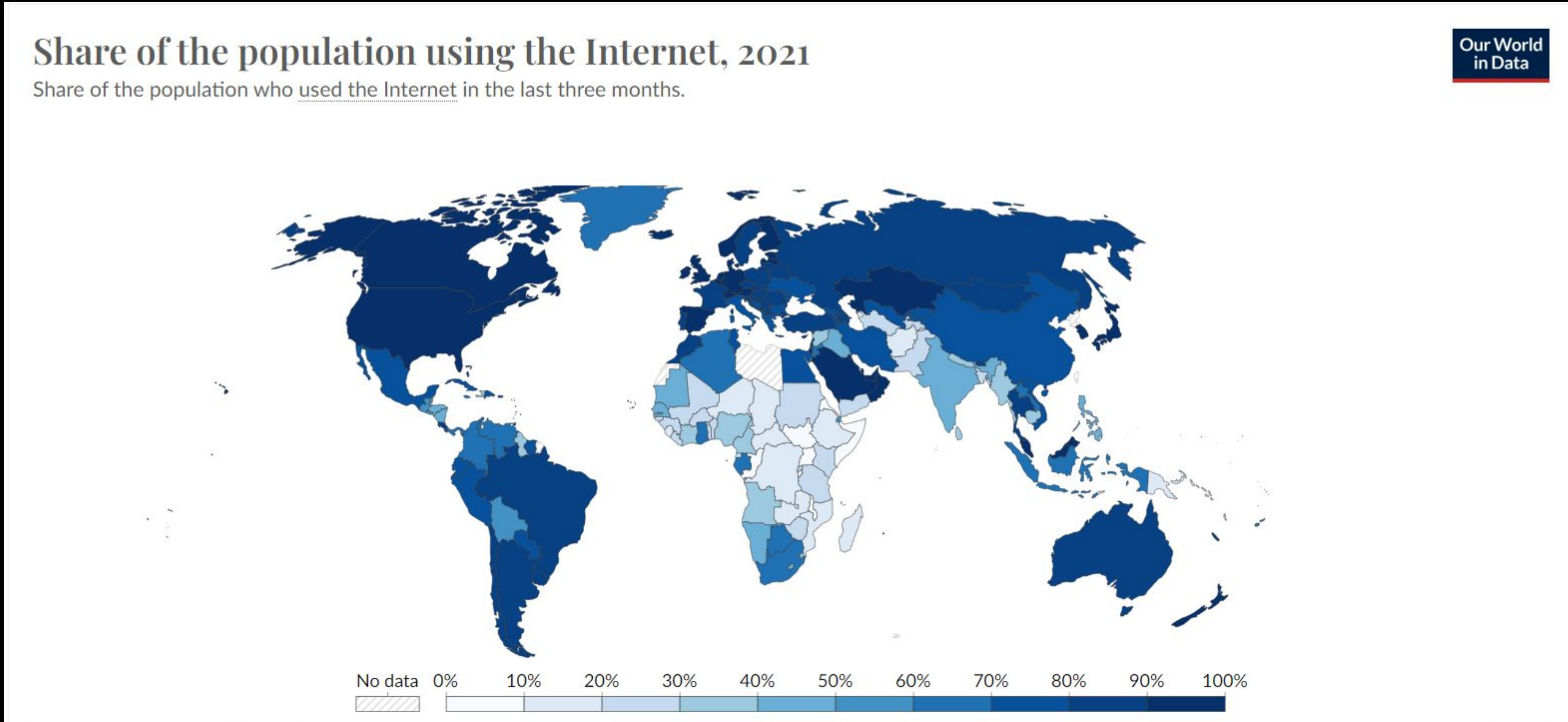
LMs are trained on Internet data that is not representative of the actual population.

“Internet access itself is not evenly distributed, resulting in Internet data over-representing younger users and those from developed countries.

Internet access is not evenly distributed.

https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

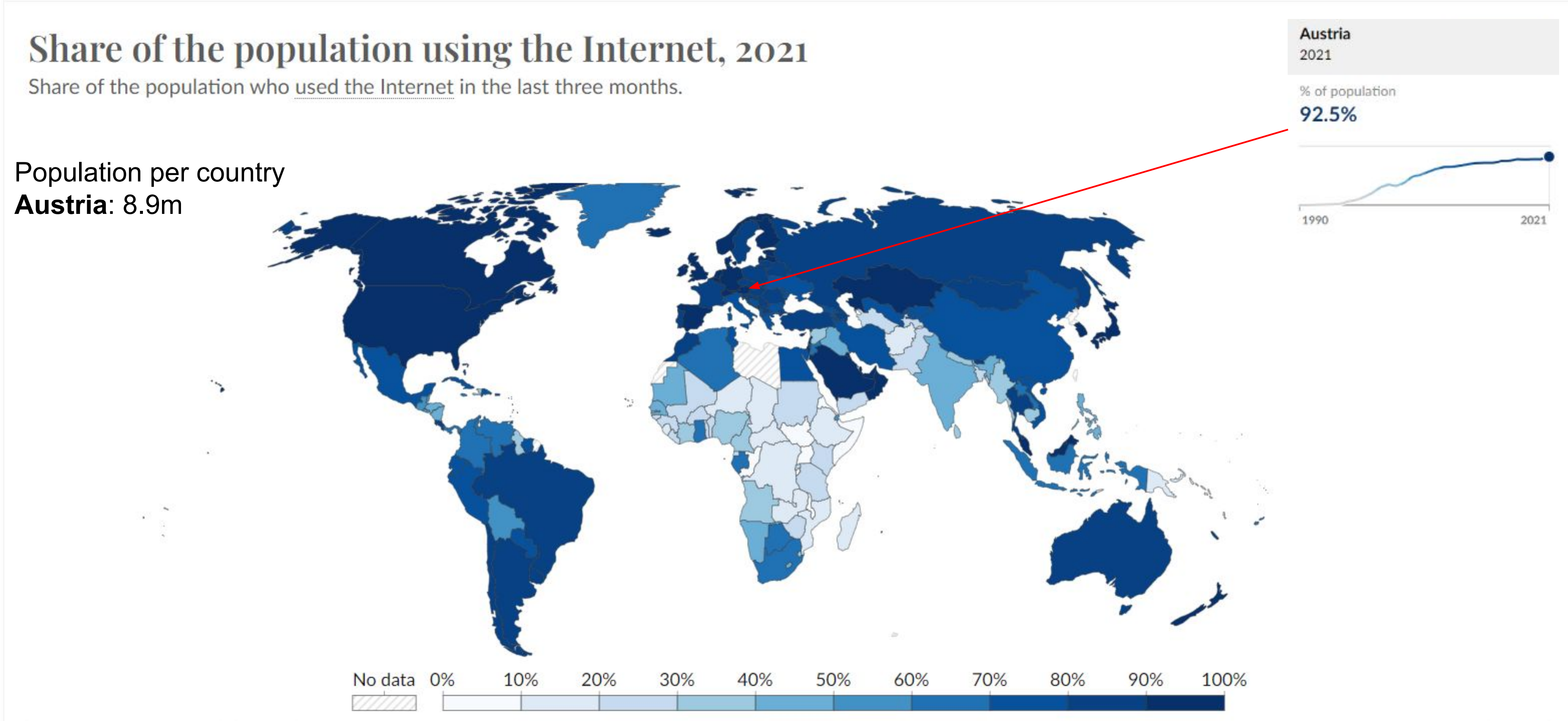
World



https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

World

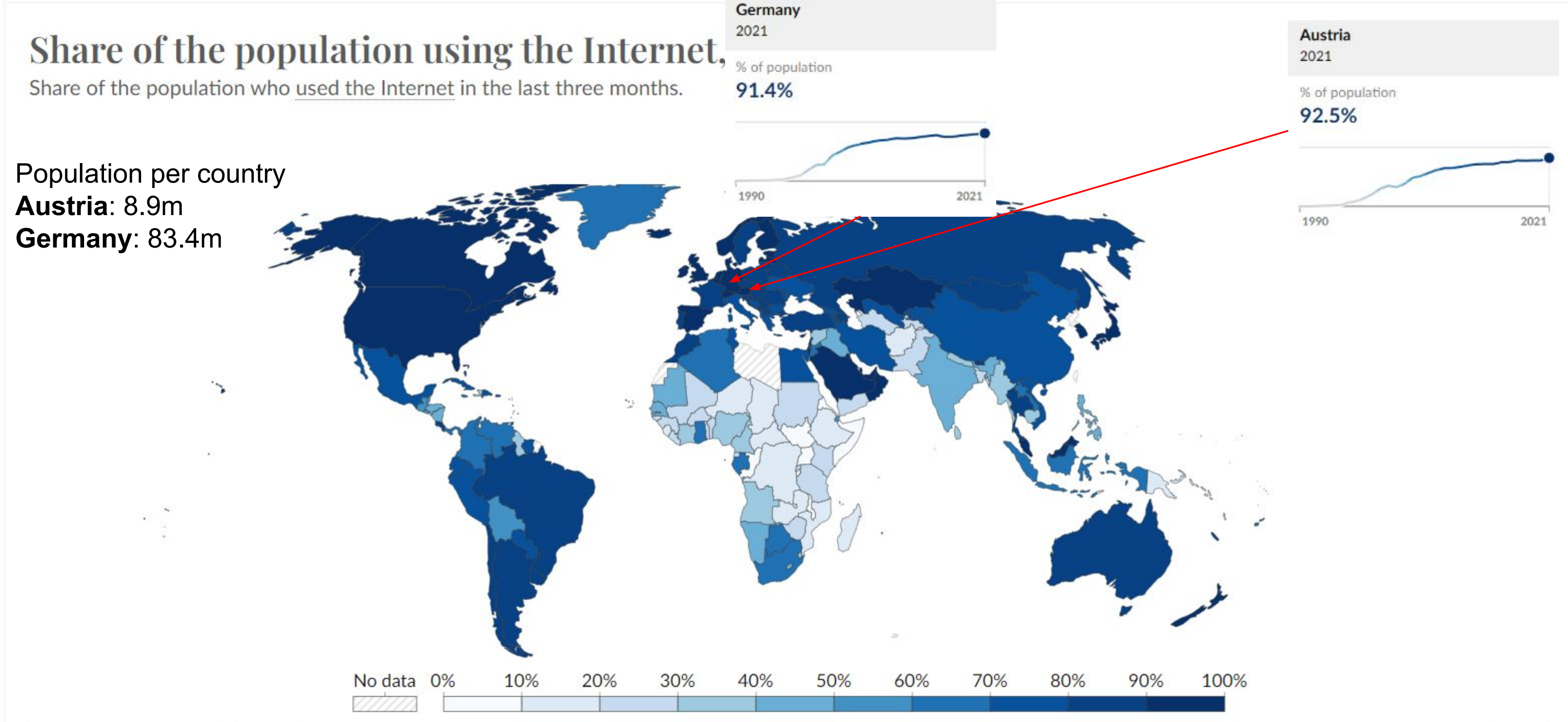
Population per country
Austria: 8.9m



https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

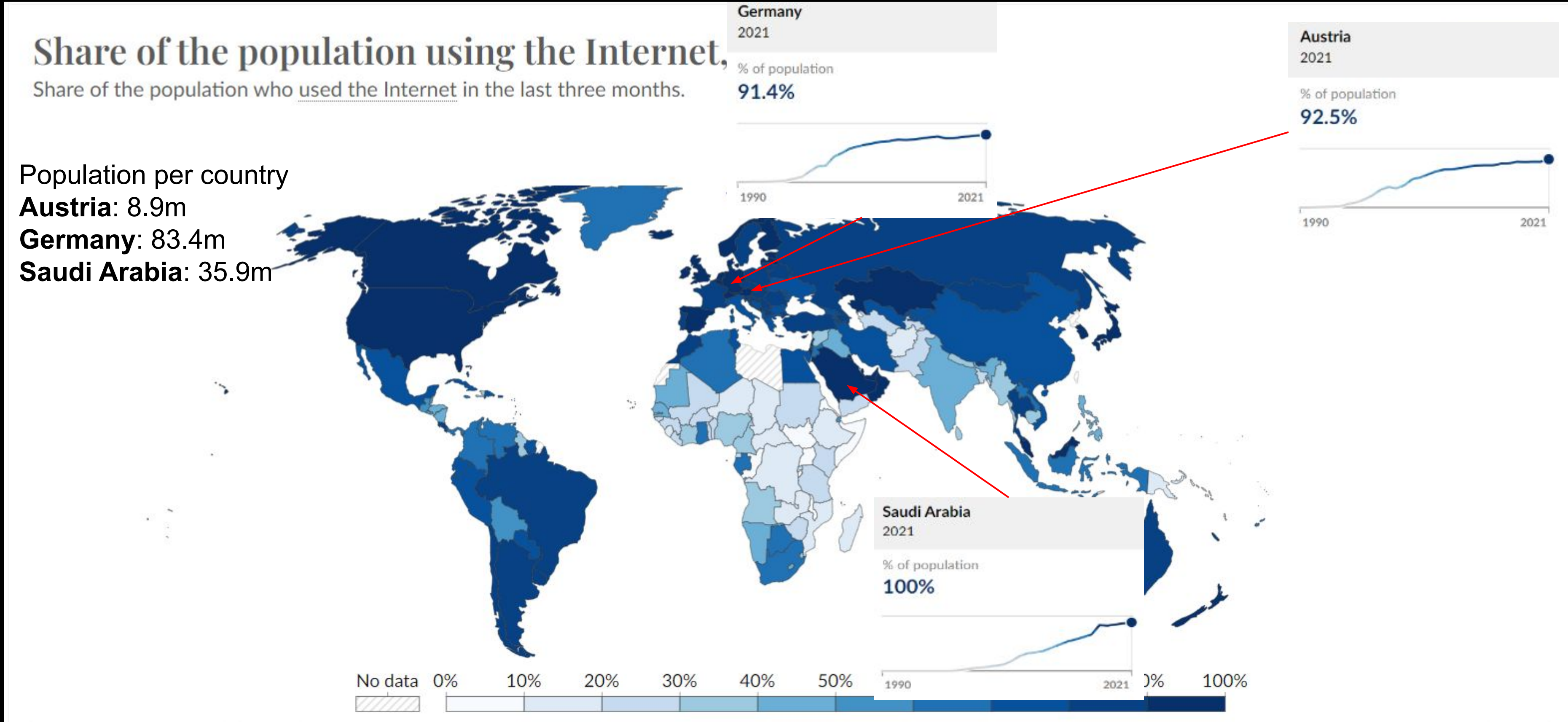
World

Population per country
Austria: 8.9m
Germany: 83.4m



https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

World

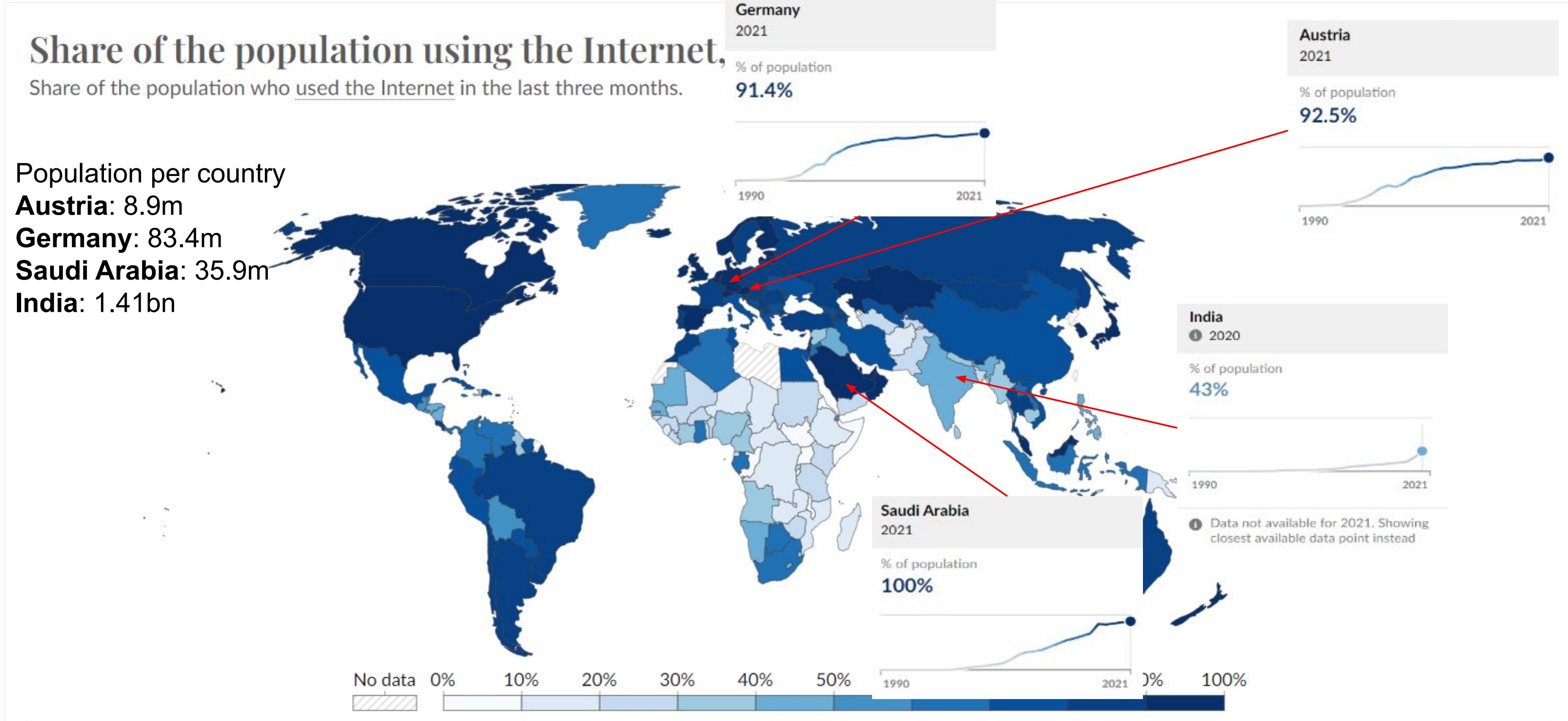


Population per country
Austria: 8.9m
Germany: 83.4m
Saudi Arabia: 35.9m

https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

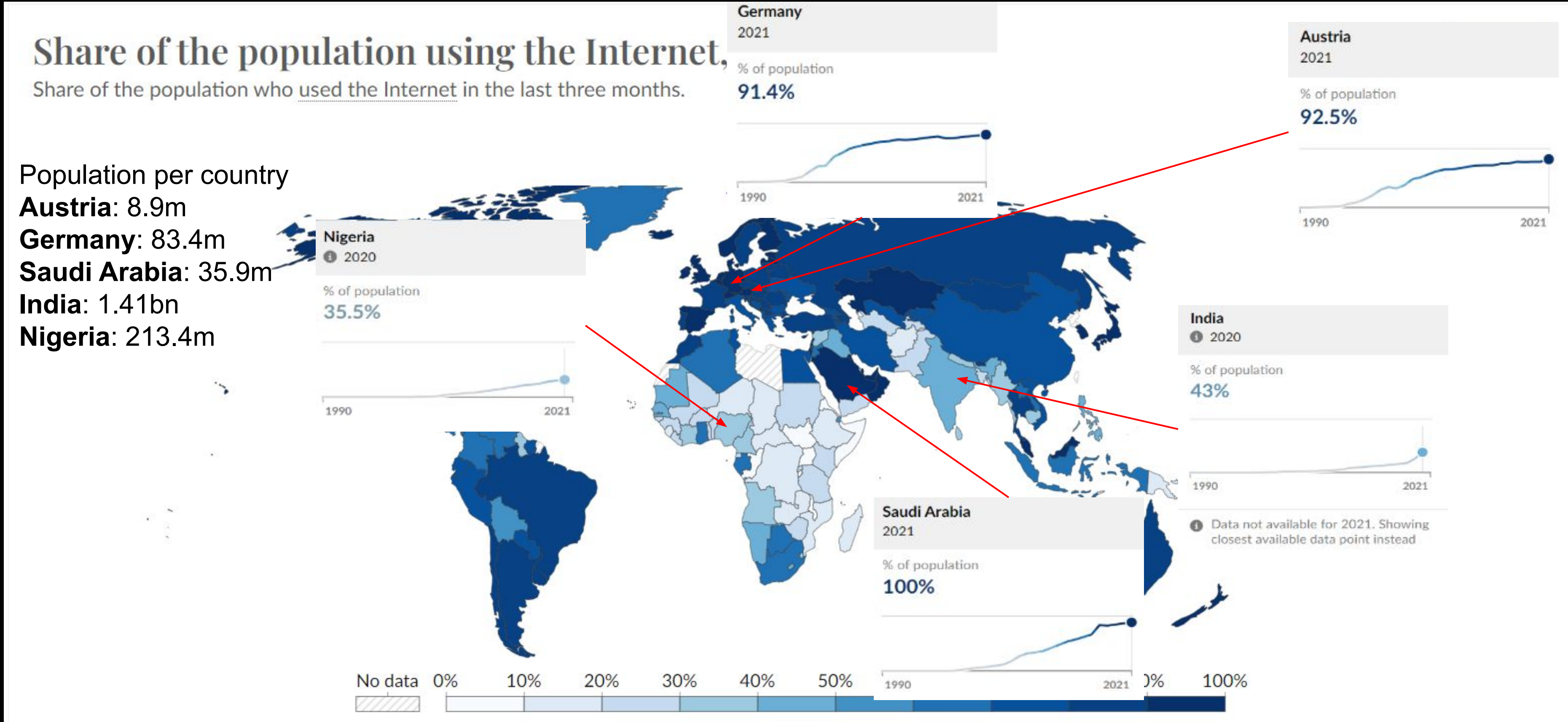
World

Population per country
Austria: 8.9m
Germany: 83.4m
Saudi Arabia: 35.9m
India: 1.41bn



https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

World

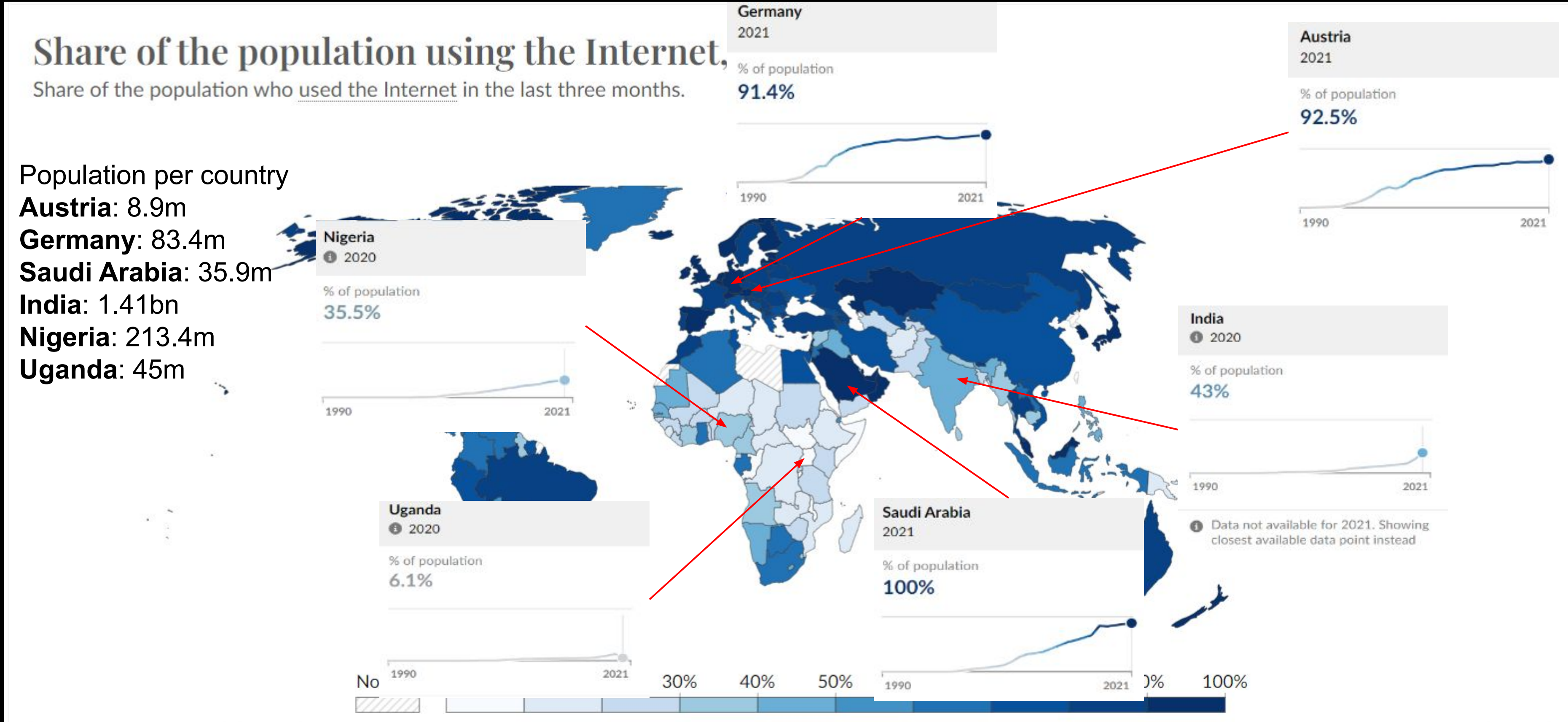


Population per country
Austria: 8.9m
Germany: 83.4m
Saudi Arabia: 35.9m
India: 1.41bn
Nigeria: 213.4m

https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

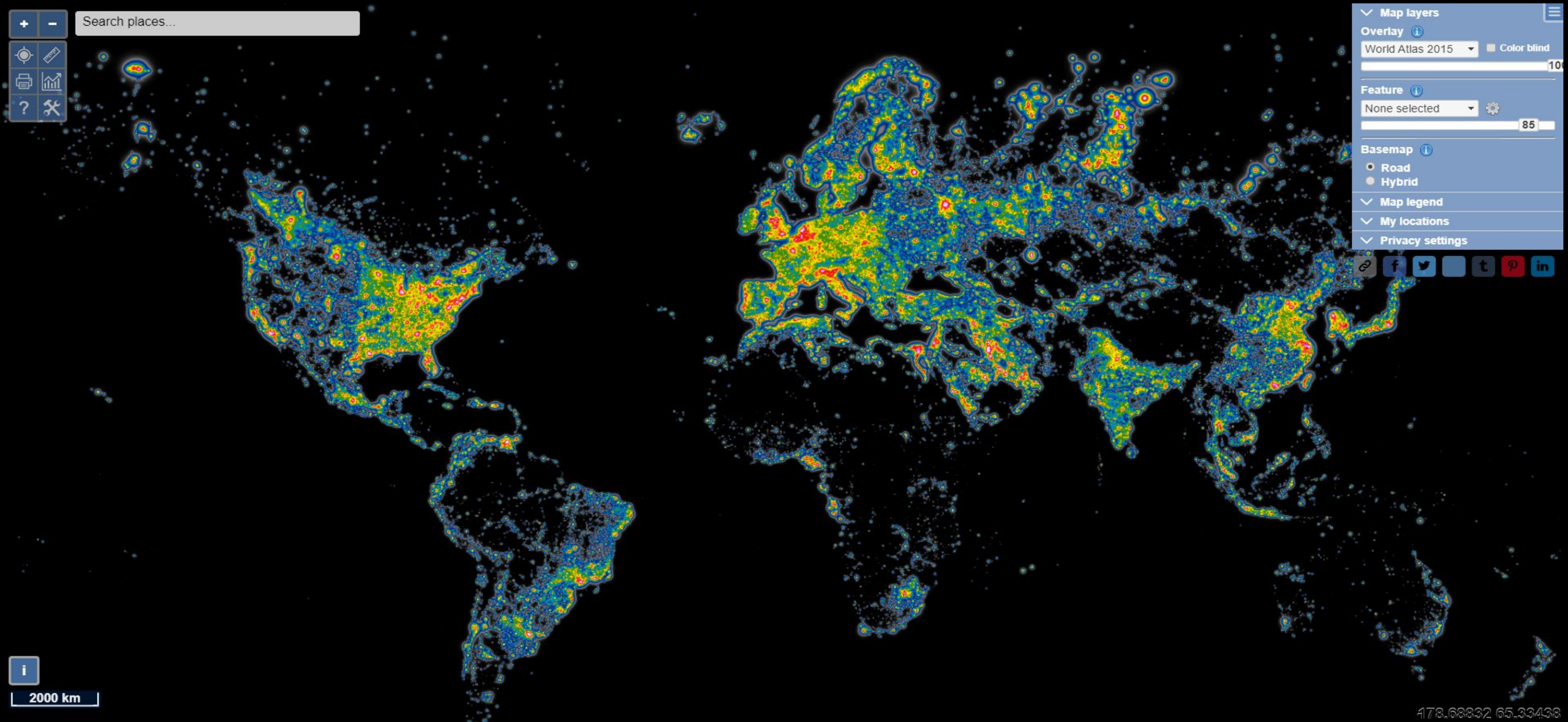
World

Population per country
Austria: 8.9m
Germany: 83.4m
Saudi Arabia: 35.9m
India: 1.41bn
Nigeria: 213.4m
Uganda: 45m



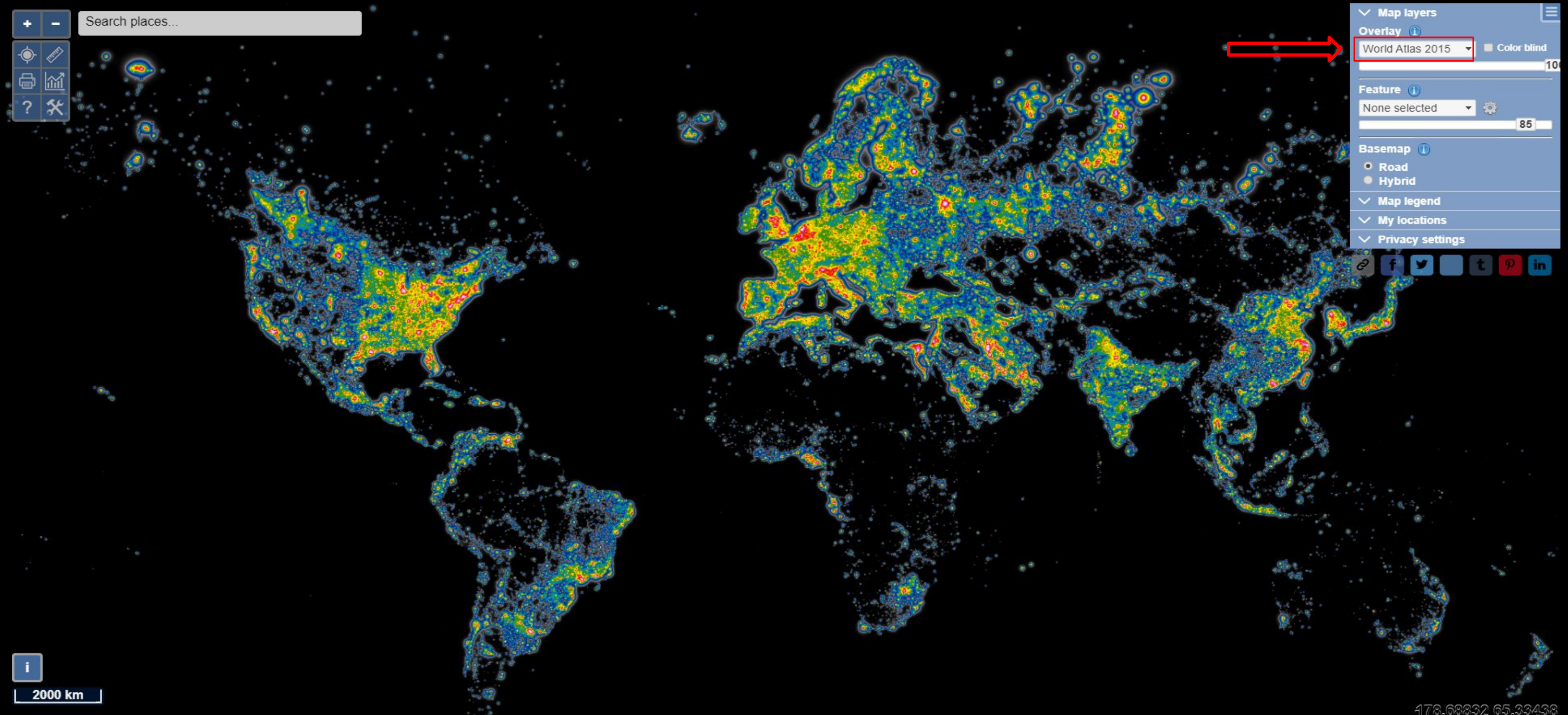
https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

World Light Map



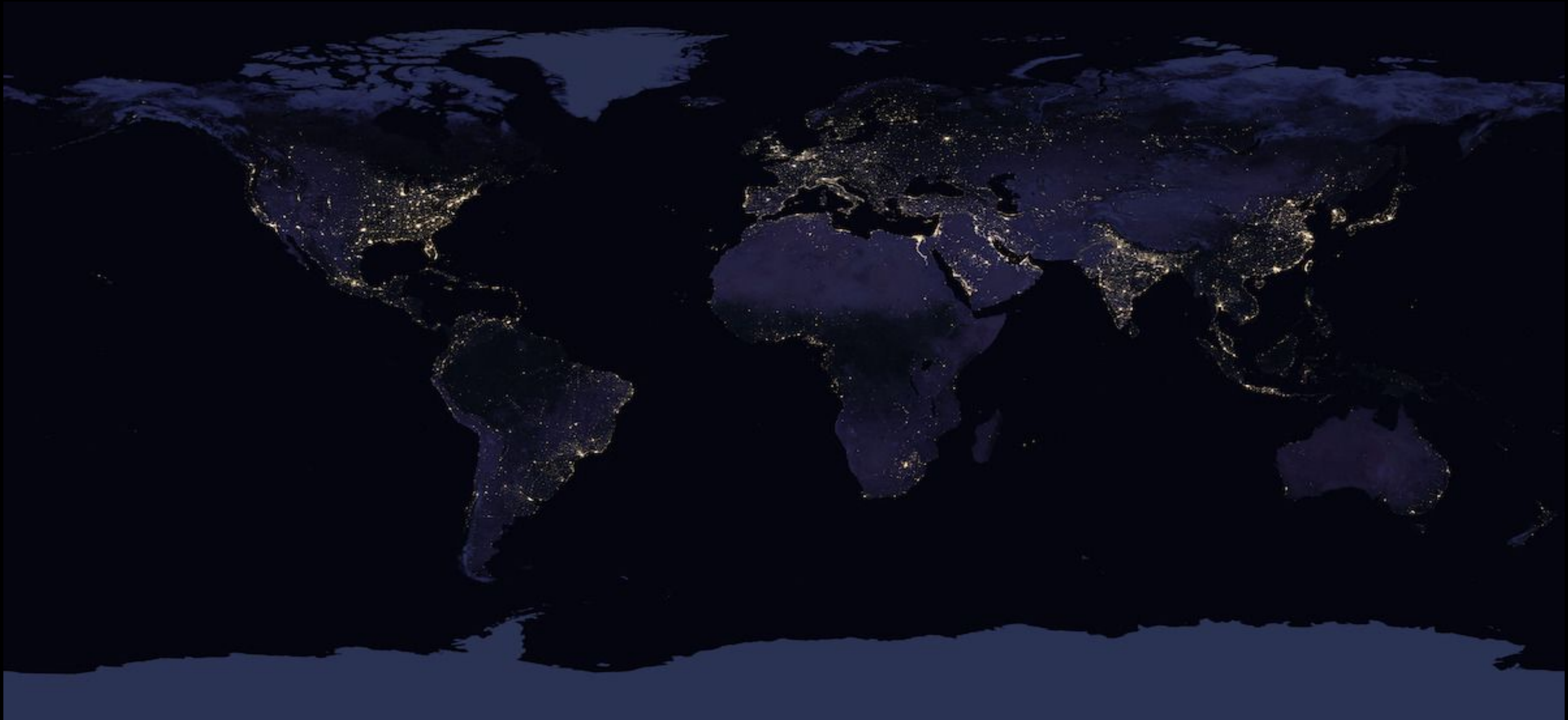
https://ourworldindata.org/grapher/share-of-individuals-using-the-internet
accessed October 29, 2023

World Light Map



https://worldhistorycommons.org/world-light-map
accessed October 28, 2023

World Light Map



Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

Summary

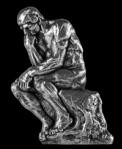
- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what’s the **alternative**?

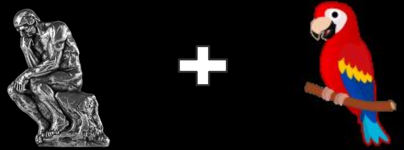
Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what’s the **alternative**?



Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?



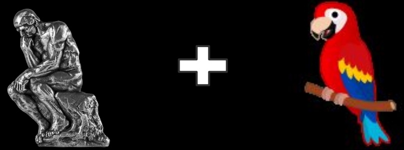
Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

  … despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the reporting of **experimental results**



Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what’s the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

… despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the reporting of **experimental results**

 + 

Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

… despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the reporting of **experimental results**

 + 

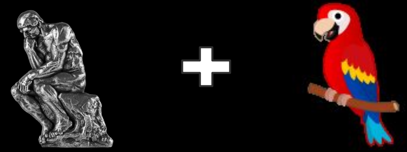
Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what’s the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

  … despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the **reporting** of **experimental results**



Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what's the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

  … despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the reporting of **experimental results**

Summary

- Are ever larger language models (**LMs**) **inevitable or necessary**?
- What the costs **costs** of this research direction and what the **necessary preparations**?
- Does the field of **NLP** or the **public** that it serves in fact **need larger LMs**?
- If so, how can we **pursue this research direction** while **mitigating its risks**?
- If not, what’s the **alternative**?

- Human-analogous natural language understanding (**NLP**) is a **grand challenge** of AI
- While large neural language models (**LMs**) are undoubtedly useful, they are **not nearly-there** solutions to this grand challenge

  … despite the way they are advertised

- **Meaning cannot be learned from form alone**
- We need to **maintain clarity** around **big picture** notions such as meaning and understanding in **task design** and the reporting of **experimental results**

Questions / Remarks / Thoughts?