

Adimen Artifizial Sortzailea

Historiatik...

...HISTERIARA

Mikel Artetxe

Reka AI
HiTZ – IXA (EHU)

Hizkuntza ereduak

Historiatik...

...HISTERIARA

Mikel Artetxe

Reka AI
HiTZ – IXA (EHU)

Hizkuntza ereduak

Hizkuntza ereduak

- Izan bedi x testu bat:

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

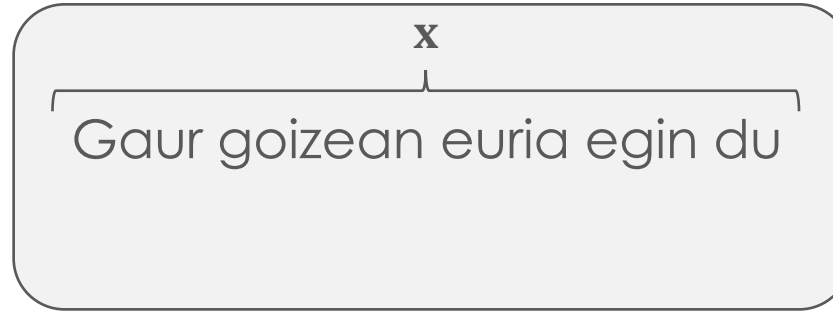
$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$

Gaur goizean euria egin du

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

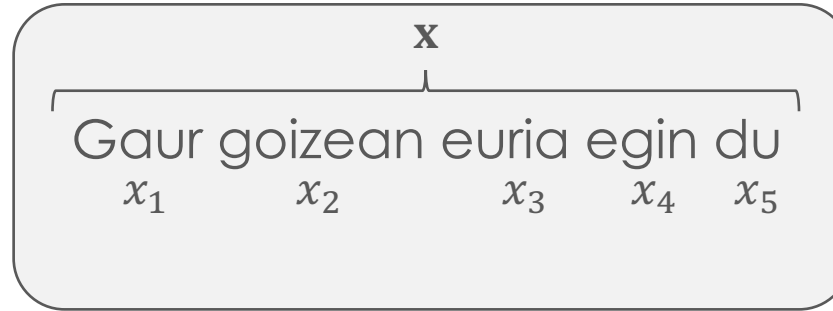
$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

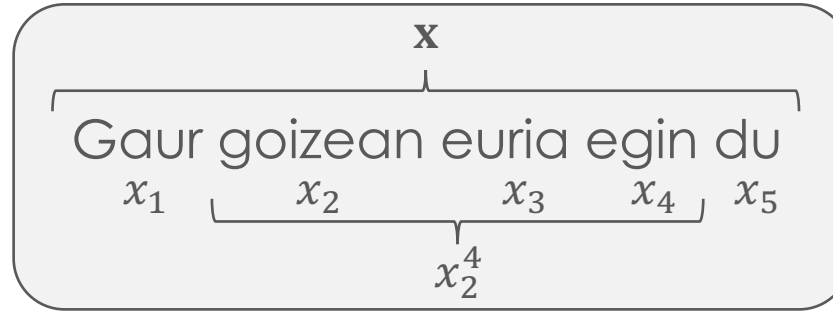
$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

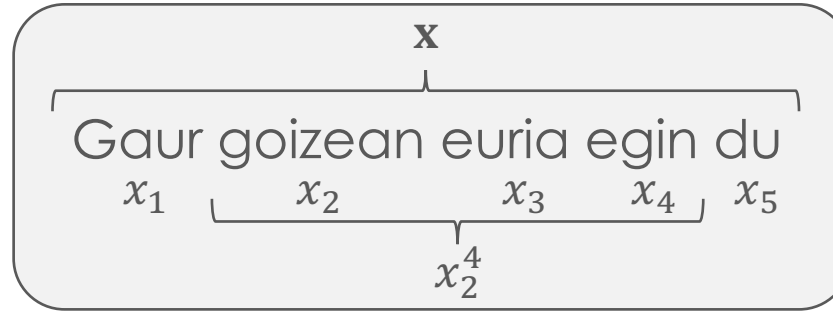
$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$

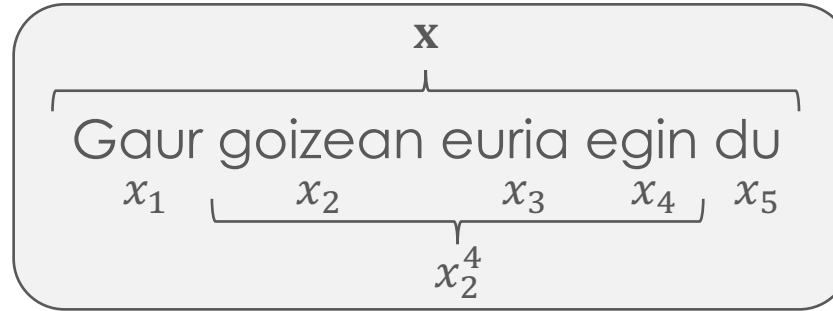


- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$

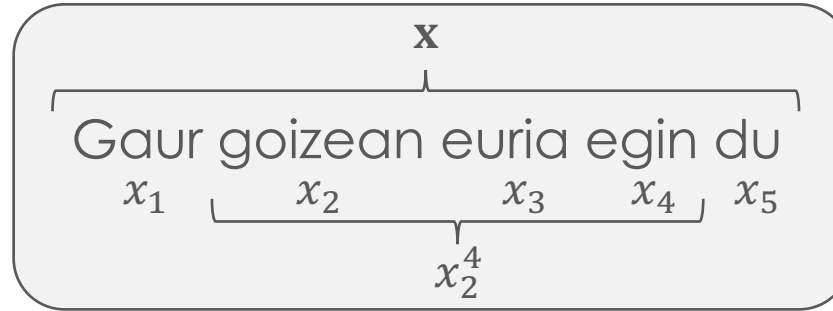


- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



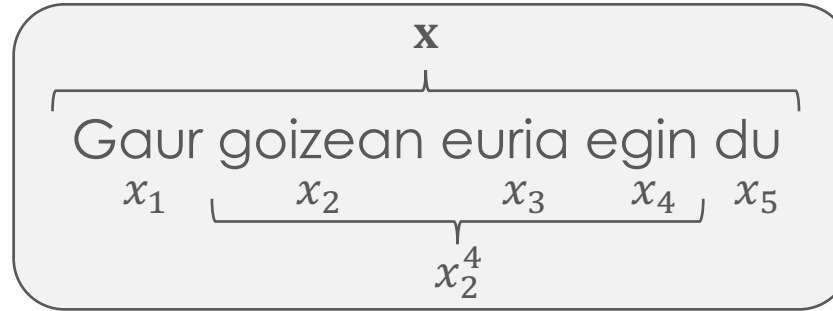
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1})$$

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



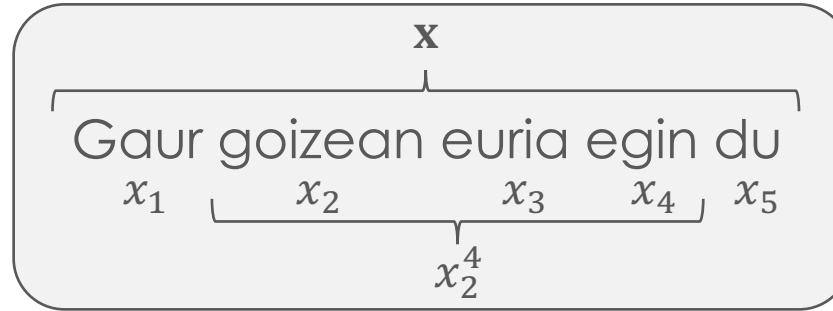
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



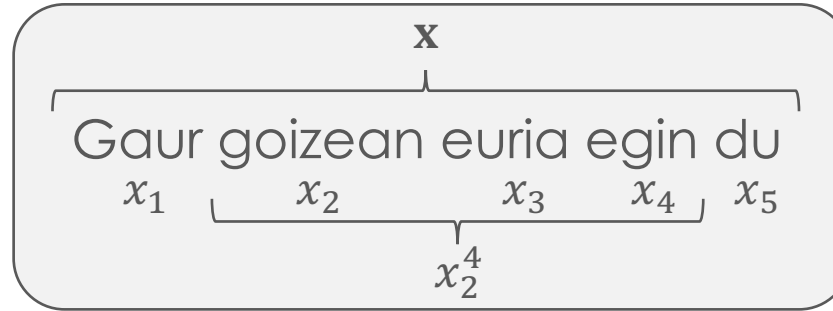
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

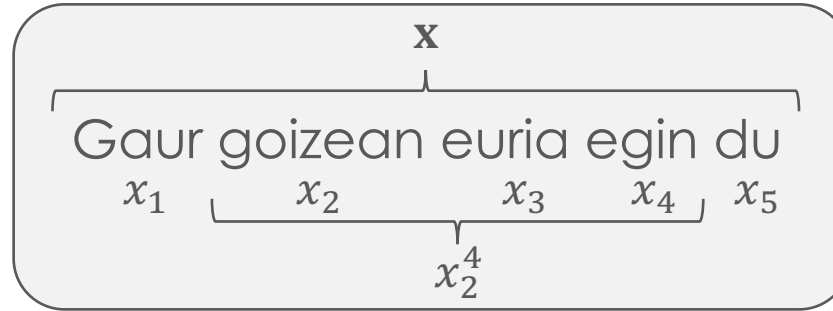
$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

↓

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

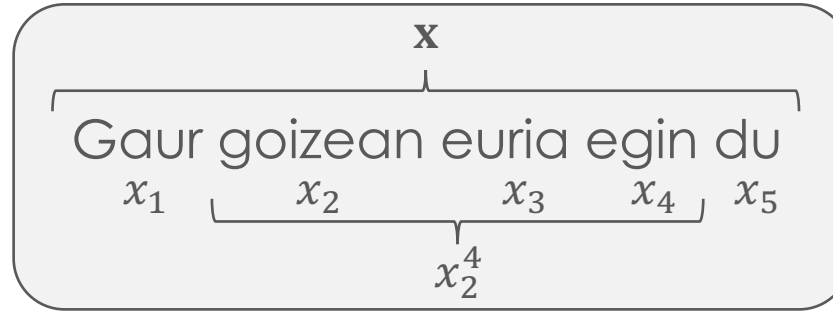
$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi

Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$

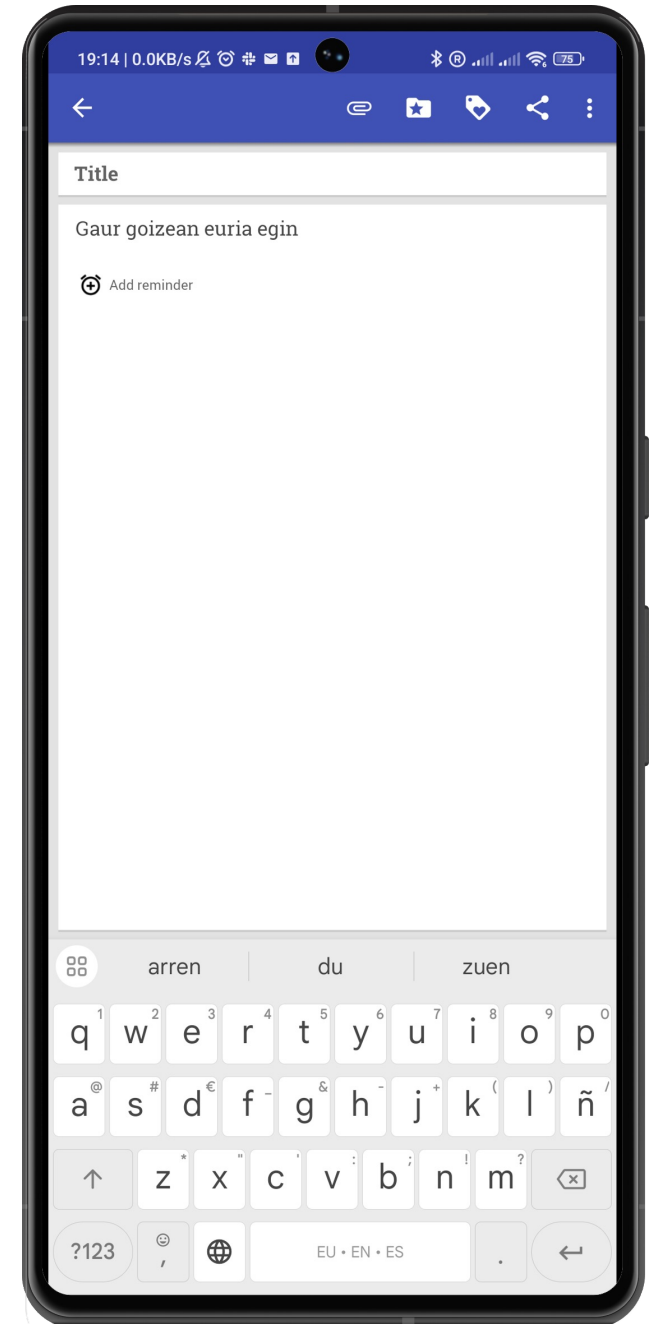


- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$



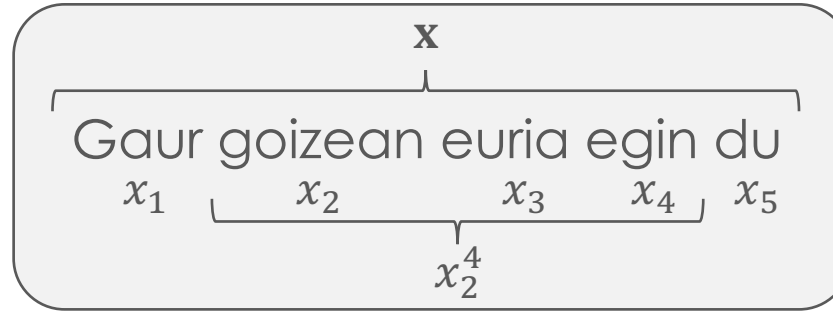
testu partzial bat
emanda hurrengo
hitza aurreikusi



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

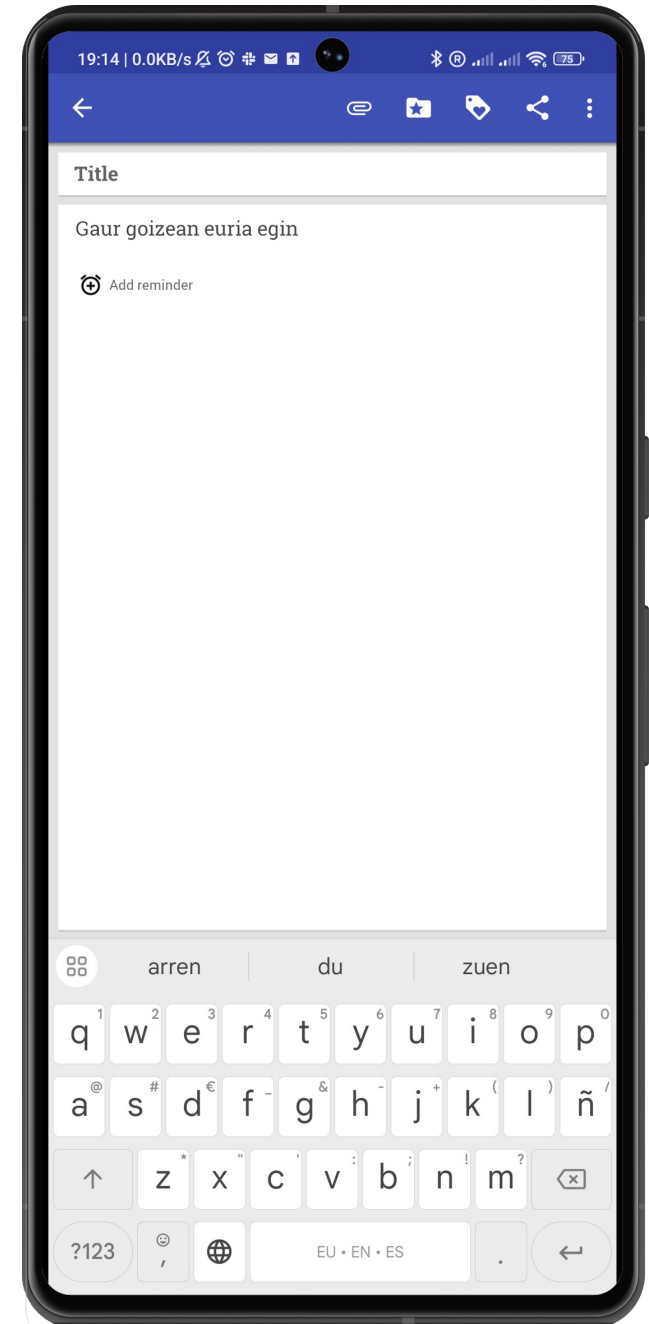
$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

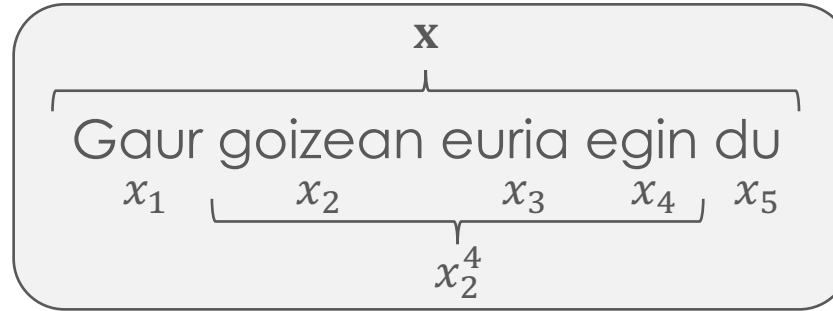
testu partzial bat
emanda hurrengo
hitza aurreikusi \Rightarrow



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



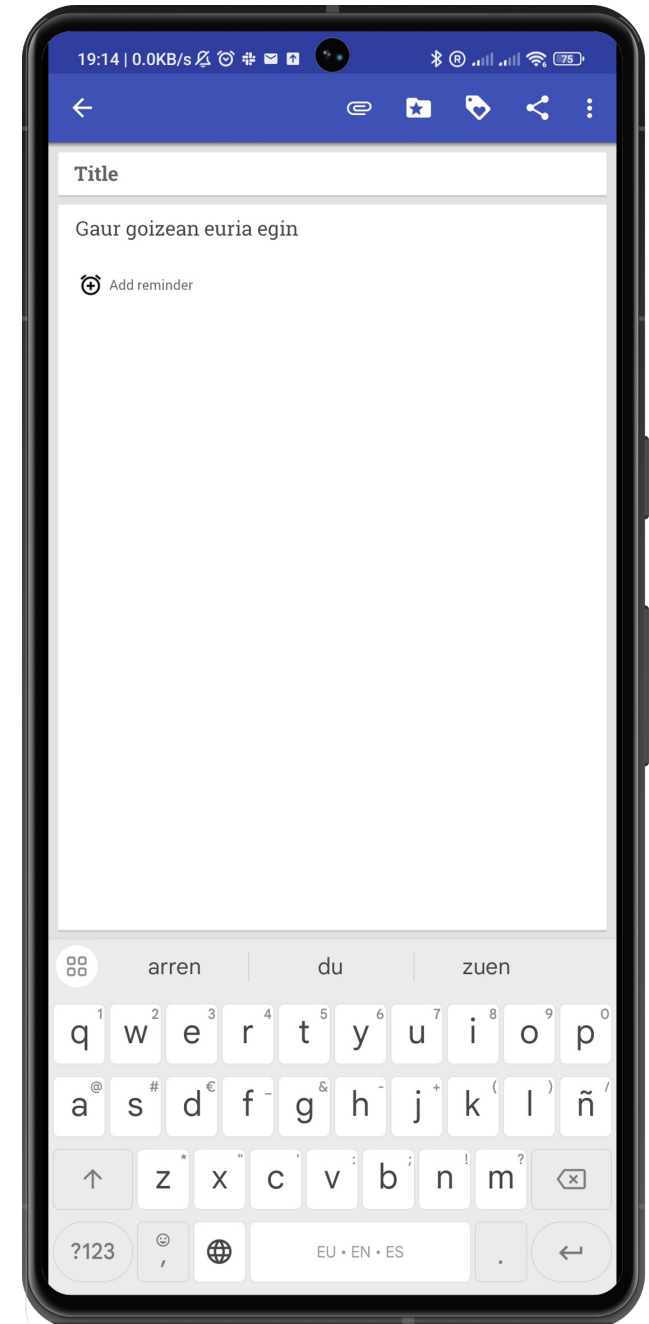
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



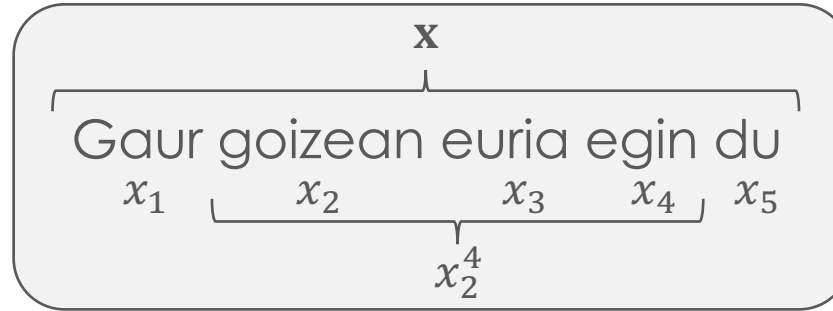
banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



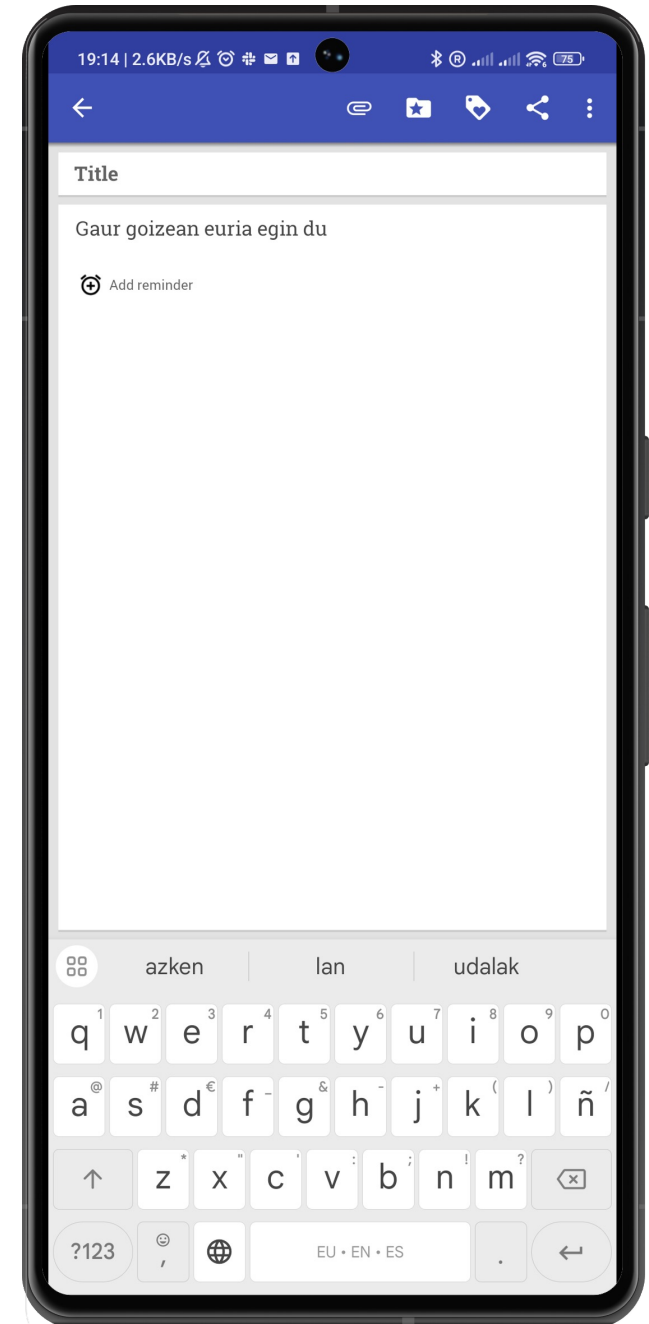
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



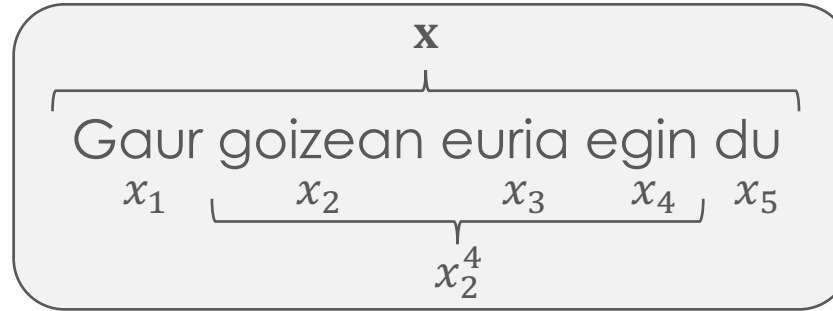
banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



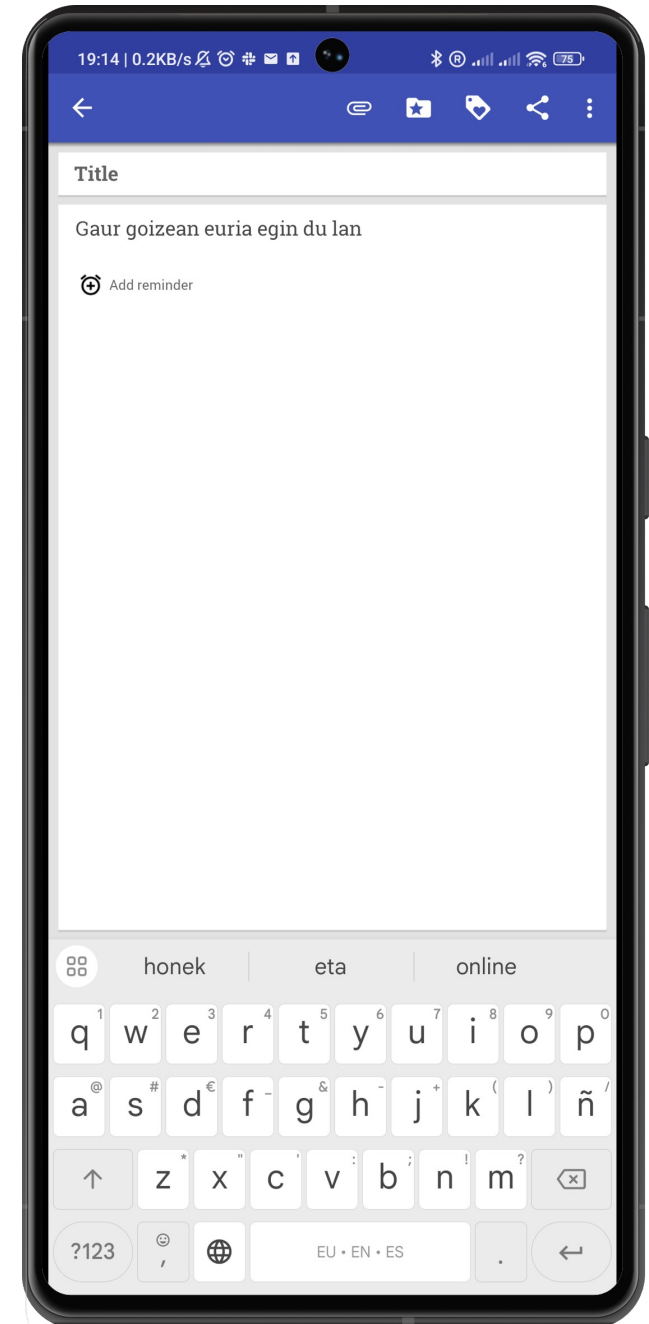
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



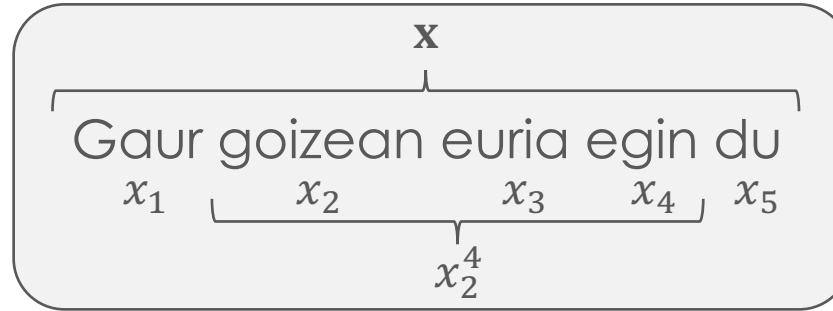
banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



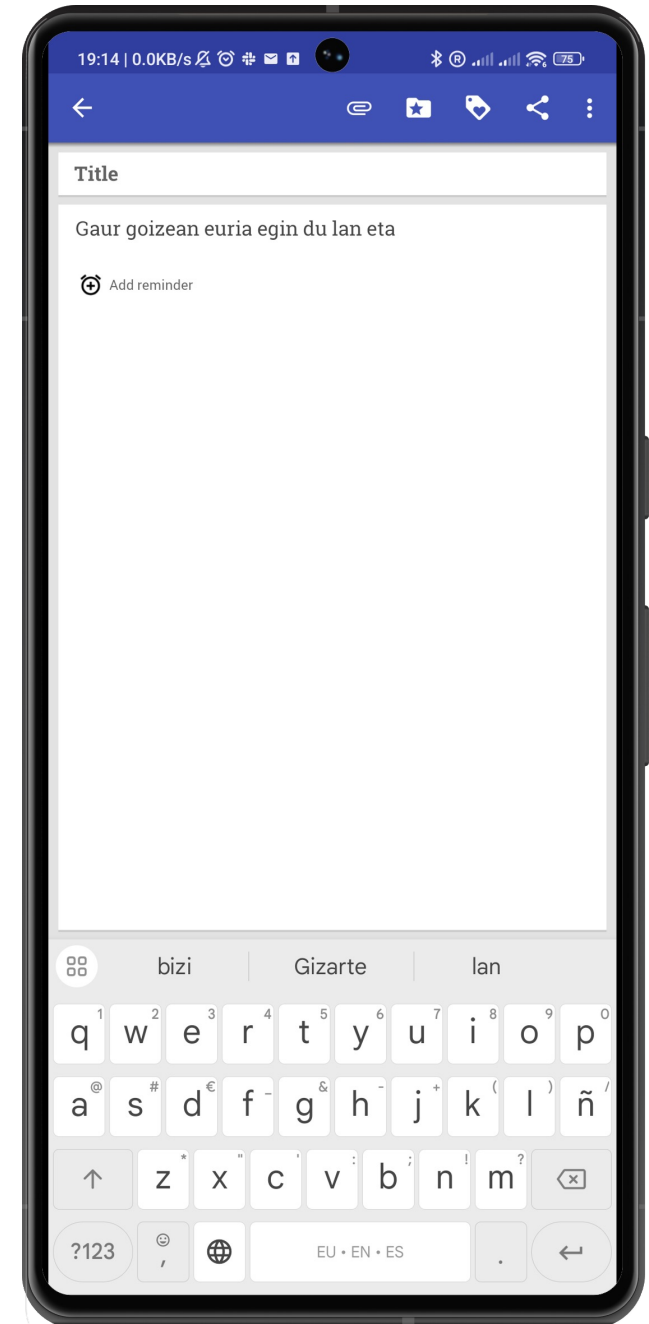
- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



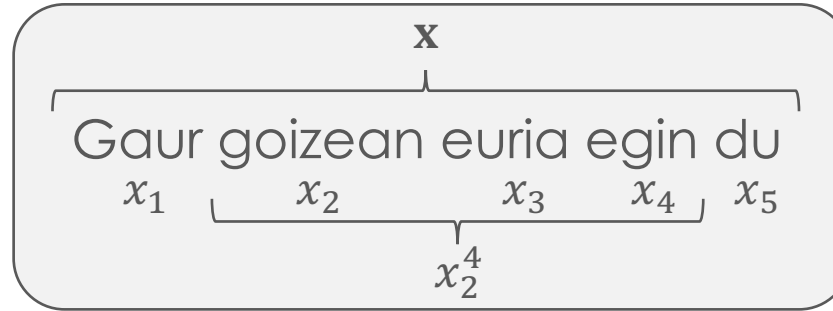
banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



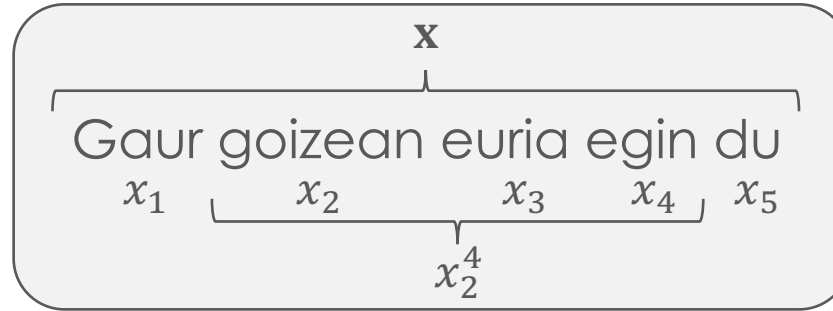
banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

- Izan bedi \mathbf{x} testu bat:

$$\mathbf{x} = x_1^n = (x_1, x_2, \dots, x_n)$$



- Hizkuntza eredu batek $p(\mathbf{x})$ modelatzen du
- Formulazio autoerregresiboa:

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | x_1^{i-1}) = p(x_1) \times p(x_2 | x_1) \times \dots \times p(x_n | x_1, x_2, \dots, x_{n-1})$$

testu partzial bat
emanda hurrengo
hitza aurreikusi



banaketa hau
laginduz testua sor
dezakegu!



Hizkuntza ereduak

Historiatik...

...HISTERIARA

Mikel Artetxe

Reka AI
HiTZ – IXA (EHU)

Historiatik...

LEHEN HIZKUNTZA
EREDUA

1948

Historiatik...

LEHEN HIZKUNTZA
EREDUA



1948



HISTORIAUREA

1948

LEHEN HIZKUNTZA
EREDUA

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

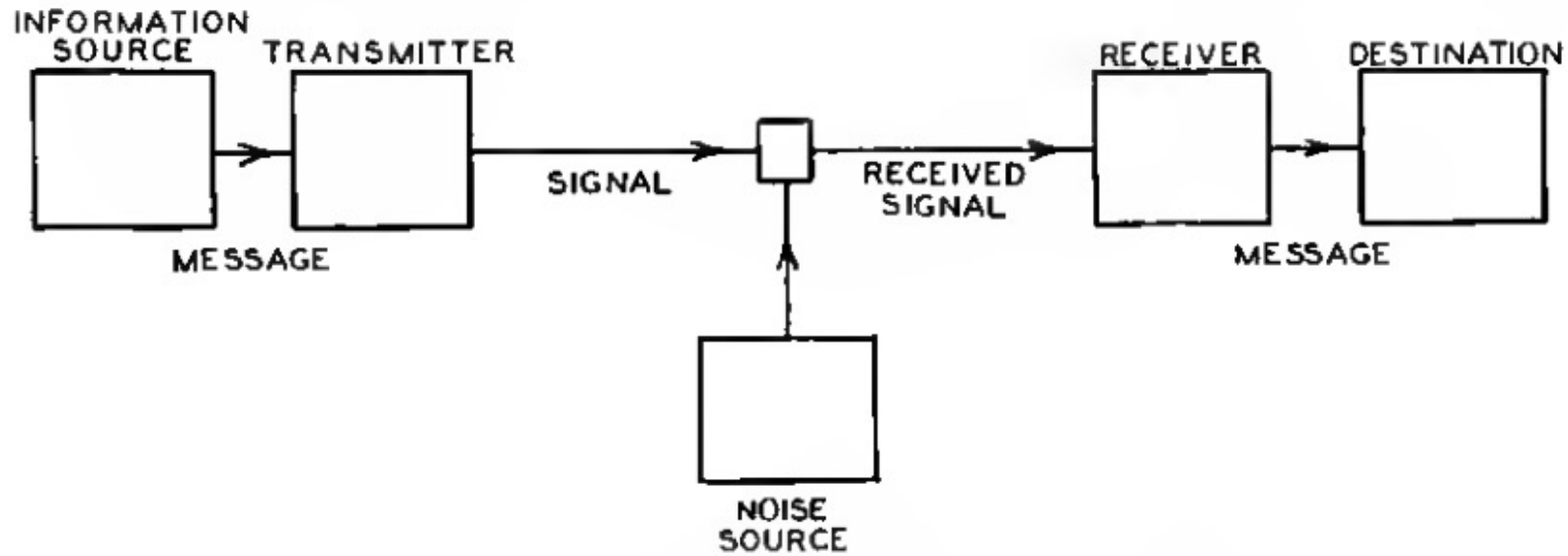
Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

A Mathematical Theory of Communication

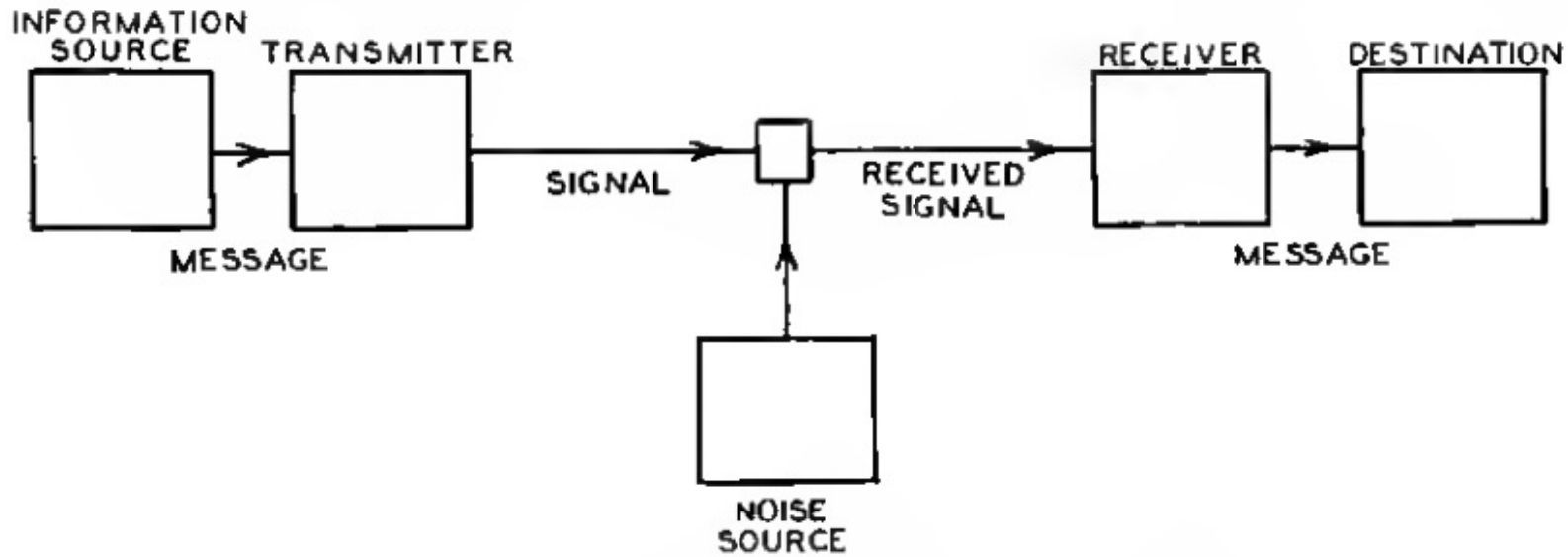
By C. E. SHANNON



A Mathematical Theory of Communication

By C. E. SHANNON

K A I X O



1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

Prediction and Entropy of Printed English

By C. E. SHANNON

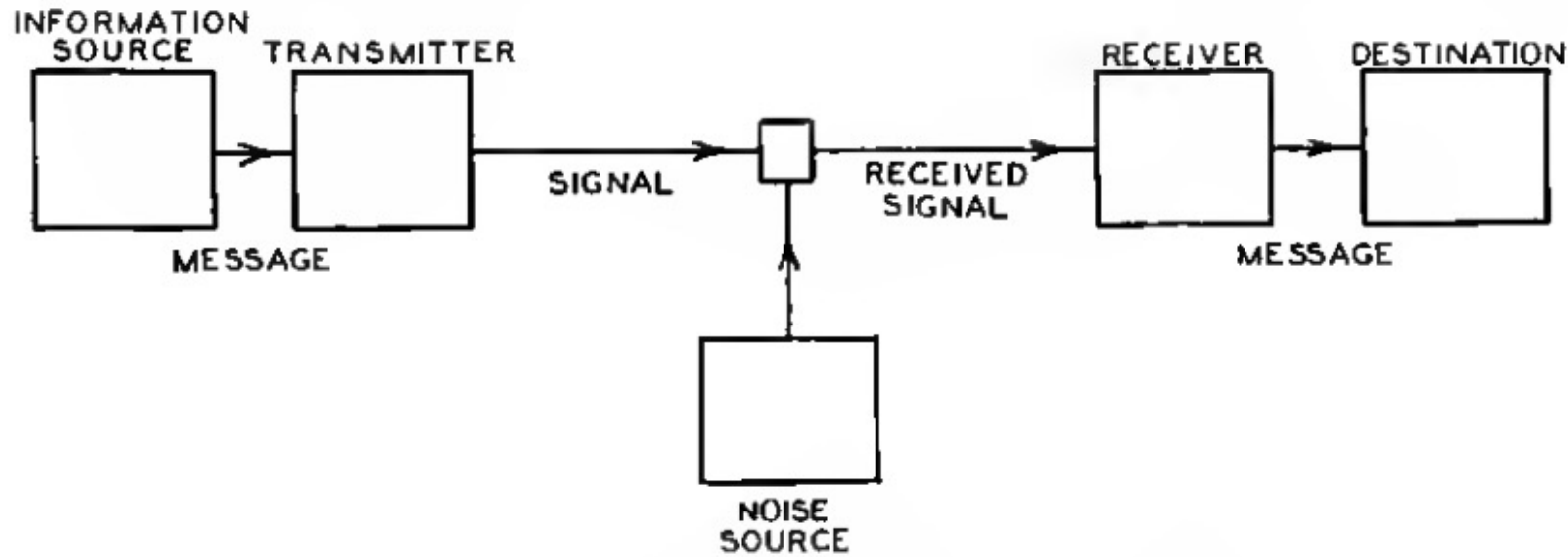
(Manuscript Received Sept. 15, 1950)

A Mathematical Theory of Communication

By C. E. SHANNON

K A I X O

5 byte?



1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

Prediction and Entropy of Printed English

By C. E. SHANNON

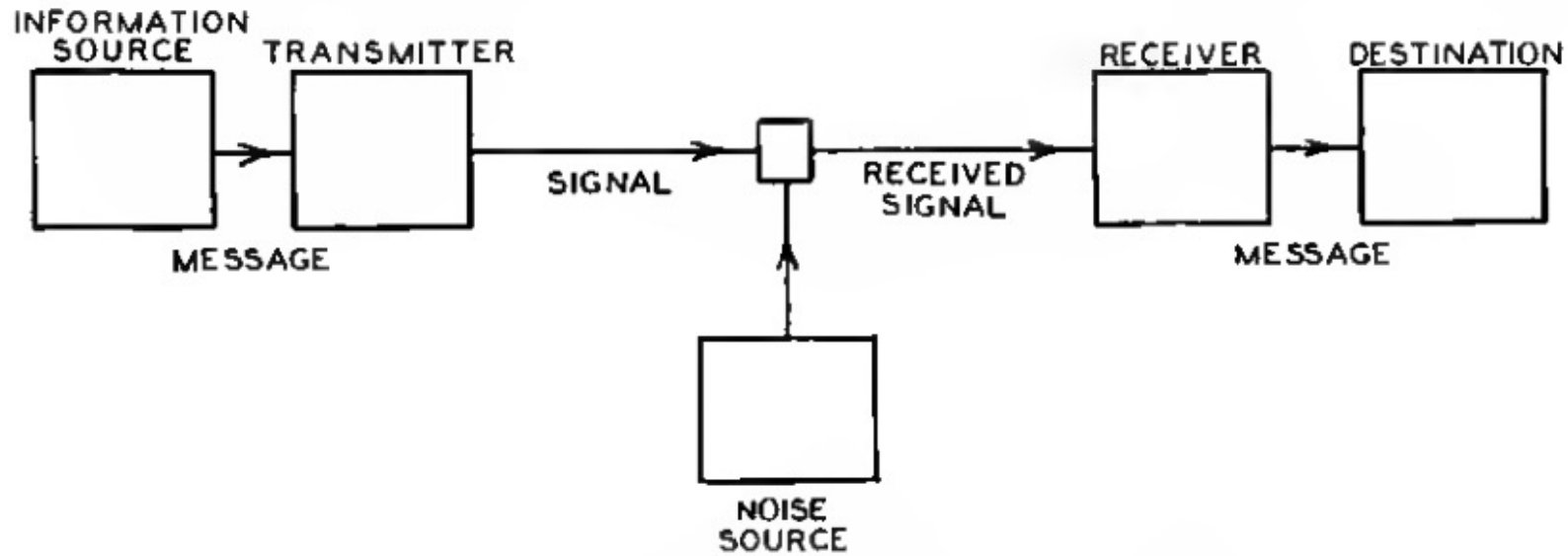
(Manuscript Received Sept. 15, 1950)

A Mathematical Theory of Communication

By C. E. SHANNON

K A I X _

5 byte?



1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

Prediction and Entropy of Printed English

By C. E. SHANNON

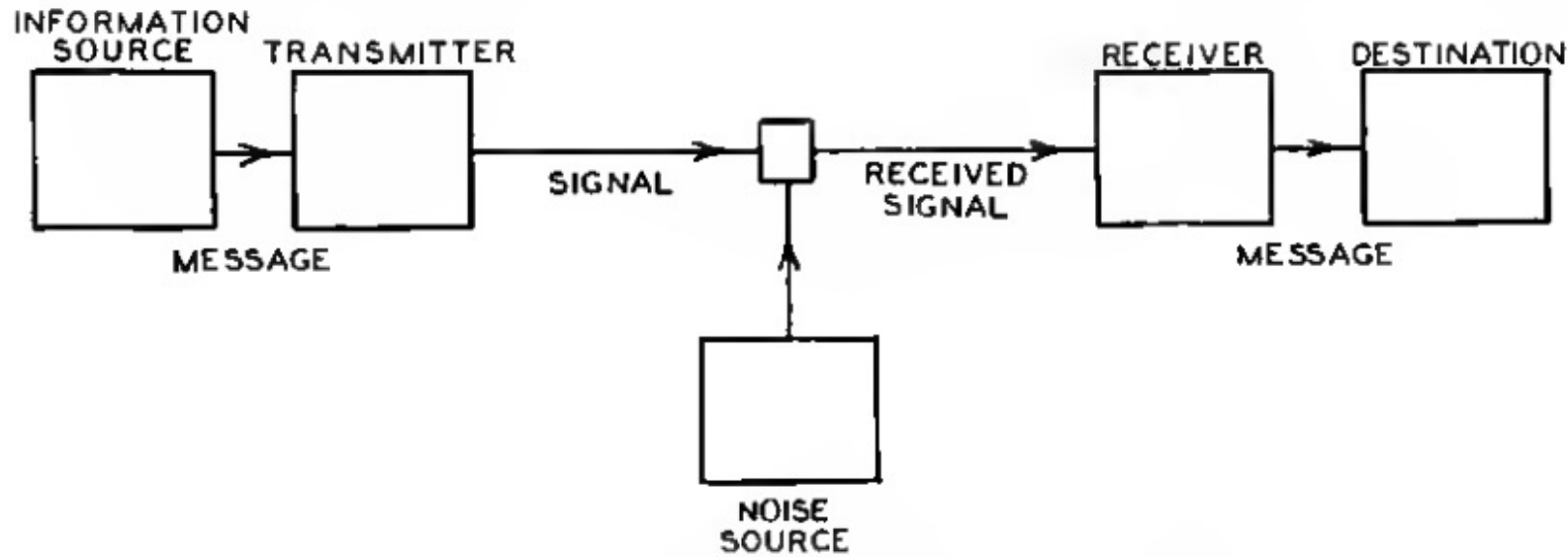
(Manuscript Received Sept. 15, 1950)

A Mathematical Theory of Communication

By C. E. SHANNON

K A I X _ { O
W

5 byte?



1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

K A I X _ { O
W

5 byte?

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

K A I X - { O
W

5 byte?

$$p(O|K A I X) \gg p(W|K A I X)$$

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

A Mathematical Theory of Communication

By C. E. SHANNON

K A I X - $\left\{ \begin{array}{l} O \\ W \end{array} \right.$

5 byte?

$$p(O|K A I X) \gg p(W|K A I X)$$

Kodeketa eraginkorrago bat diseina dezakegu jarraipen probableenentzat bit gutxiago erabiliz!

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu

1948

LEHEN HIZKUNTZA
EREDUA

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})}$$

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})} \quad p(\text{zuen} | \text{egin})$$

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})} p(\text{zuen} | \text{egin})$$

Zeremonia bat antolatu zuten etxekoek Margaritari paparrekoa entregatzeko. Egongeletako batean egin zuten, ez handiengan, baina erabat serio. Familiako asko bildu ziren egongelara: aita eta ama zeuden, hiru zerbitzari, sukaldaria, alabarik txikiak, umeak oraindik, amaren lehengusinetako bat... Kalerako jantzita egin zuten, eta Margarita bera goitik behera lotsatuta zegoen, berrogeita hamalau urteko.

Etxeko andreak jarri zion paparreko orratza jakan, berokian. Senarrak egin behar zuten baina beldur zen, ez ote zion behar ez den lekuren bat ukituko. Izerditan egon zen Margarita denbora osoan, behar baino bero handiagoa zen beti etxe horietan, leihoak ere itxita zeuden; zenbat eta mamu gehiago etxe batean, orduan eta leiho gutxiago... Kristalak guztiz gardenak ere ez.

Hirutan ahalegindu zen andrea orratza Margaritaren paparrean sartzen, baina traba egiten zion zerbaitek, orratzak ez zuten aurrera egiten. Izerditan zituen eskuak, dardarka ere bai, dardara txikia. Hirugarrenean irten zen orratza oihaletik. Andreak irribarre egin zuen, Margaritak ere bai, malko bat, gizonak kea bota zuen ahotik... Inoizko goizen heldu zen Margarita Maguregi etxera egun hartan.

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})} p(\text{zuen} | \text{egin})$$

Zeremonia bat antolatu zuten etxekoek Margaritari paparrekoa entregatzeko. Egongeletako batean egin zuten, ez handiengan, baina erabat serio. Familiako asko bildu ziren egongelara: aita eta ama zeuden, hiru zerbitzari, sukaldaria, alabarik txikiak, umeak oraindik, amaren lehengusinetako bat... Kalerako jantzita egin zuten, eta Margarita bera goitik behera lotsatuta zegoen, berrogeita hamalau urteko.

Etxeko andreak jarri zion paparreko orratza jakan, berokian. Senarrak egin behar zuen baina beldur zen, ez ote zion behar ez den lekuren bat ukituko. Izerditan egon zen Margarita denbora osoan, behar baino bero handiagoa zen beti etxe horietan, leihoak ere itxita zeuden; zenbat eta mamu gehiago etxe batean, orduan eta leiho gutxiago... Kristalak guztiz gardenak ere ez.

Hirutan ahalegindu zen andrea orratza Margaritaren paparrean sartzen, baina traba egiten zion zerbaitek, orratzak ez zuen aurrera egiten. Izerditan zituen eskuak, dardarka ere bai, dardara txikia. Hirugarrenean irten zen orratza oihaletik. Andreak irribarre egin zuten, Margaritak ere bai, malko bat, gizonak kea bota zuten ahotik... Inoizko goizen heldu zen Margarita Maguregi etxera egun hartan.

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})} p(\text{zuen} | \text{egin})$$

Zeremonia bat antolatu zuten etxekoek Margaritari paparrekoa entregatzeko. Egongeletako batean egin zuten, ez handiengan, baina erabat serio. Familiako asko bildu ziren egongelara: aita eta ama zeuden, hiru zerbitzari, sukaldaria, alabarik txikiak, umeak oraindik, amaren lehengusinetako bat... Kalerako jantzita egin zuten, eta Margarita bera goitik behera lotsatuta zegoen, berrogeita hamalau urteko.

Etxeko andreak jarri zion paparreko orratza jakan, berokian. Senarrak egin behar zuten baina beldur zen, ez ote zion behar ez den lekuren bat ukituko. Izerditan egon zen Margarita denbora osoan, behar baino bero handiagoa zen beti etxe horietan, leihoak ere itxita zeuden; zenbat eta mamu gehiago etxe batean, orduan eta leihogutxiago... Kristalak guztiz gardenak ere ez.

Hirutan ahalegindu zen andrea orratza Margaritaren paparrean sartzen, baina traba egiten zion zerbaitek, orratzak ez zuten aurrera egiten. Izerditan zituen eskuak, dardarka ere bai, dardara txikia. Hirugarrenean irten zen orratza oihaletik. Andreak irribarre egin zuen, Margaritak ere bai, malko bat, gizonak kea bota zuen ahotik... Inoizko goizen heldu zen Margarita Maguregi etxera egun hartan.

- Kodeketa optimoak $\sum_{i=1}^n \log p(x_i | x_1^{i-1})$ bit behar ditu
 - Geroz eta hizkuntza eredu hobea, orduan eta gutxiago!
- k . ordenako Markoven katea: $p(x_i | x_1^{i-1}) \approx p(x_i | x_{i-(k-1)}^{i-1})$
 - 3. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{euria egin})$
 - 2. ordena: $p(\text{zuen} | \text{Atzo goizean euria egin}) \approx p(\text{zuen} | \text{egin})$
- n -grama ereduak maiztasun kontaktak erabiliz:

$$p(x_i | x_{i-(k-1)}^{i-1}) = \frac{\text{count}(x_{i-(k-1)}^i)}{\text{count}(x_{i-(k-1)}^{i-1})}$$

$$p(\text{zuen} | \text{egin}) = 1/4 = 0.25$$

Zeremonia bat antolatu zuten etxekoek Margaritari paparrekoa entregatzeko. Egongeletako batean **egin zuten**, ez handiengan, baina erabat serio. Familiako asko bildu ziren egongelara: aita eta ama zeuden, hiru zerbitzari, sukaldaria, alabarik txikiak, umeak oraindik, amaren lehengusinetako bat... Kalerako jantzita **egin zuten**, eta Margarita bera goitik behera lotsatuta zegoen, berrogeita hamalau urteko.

Etxeko andreak jarri zion paparreko orratza jakan, berokian. Senarrak **egin behar** zuen baina beldur zen, ez ote zion behar ez den lekuren bat ukituko. Izerditan egon zen Margarita denbora osoan, behar baino bero handiagoa zen beti etxe horietan, leihoak ere itxita zeuden; zenbat eta mamu gehiago etxe batean, orduan eta leiho gutxiago... Kristalak guztiz gardenak ere ez.

Hirutan ahalegindu zen andrea orratza Margaritaren paparrean sartzen, baina traba egiten zion zerbaitek, orratzak ez zuen aurrera egiten. Izerditan zituen eskuak, dardarka ere bai, dardara txikia. Hirugarrenean irten zen orratza oihaletik. Andreak irribarre **egin zuen**, Margaritak ere bai, malko bat, gizonak kea bota zuen ahotik... Inoizko goizen heldu zen Margarita Maguregi etxera egun hartan.

Historiatik...

LEHEN HIZKUNTZA
EREDUA



1948



HISTORIAUREA

Historiatik...

LEHEN HIZKUNTZA
EREDUA



1948



HISTORIAUREA



N-GRAMEN AROA

1948

N-GRAMEN AROA

1948+

N-GRAMEN AROA

1948+

N-GRAMEN AROA

- Aurrerapen tekniko gutxi

1948+

N-GRAMEN AROA

- Aurrerapen tekniko gutxi
- Aplikazioen garapena noisy channel formulazioaren eskutik

1948+

N-GRAMEN AROA

- Aurrerapen tekniko gutxi
- Aplikazioen garapena noisy channel formulazioaren eskutik

$$\operatorname{argmax}_x p(x|y) = \operatorname{argmax}_x p(y|x) \times p(x)$$

1948

N-GRAMEN AROA

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

1948

2000



HISTORIAUREA



N-GRAMEN AROA

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

1948

2000



HISTORIAUREA



N-GRAMEN AROA

2000

LEHEN EREDU
NEURONALA

2000

LEHEN EREDU
NEURONALA

A Neural Probabilistic Language Model

Yoshua Bengio*, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

2000

LEHEN EREDU
NEURONALA

A Neural Probabilistic Language Model

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez

2000

LEHEN EREDU NEURONALA

A Neural Probabilistic Language Model

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengio,y.ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du
 - Hitz-bektoreak eta funtzioaren parametroak elkarrekin ikasten dira entrenamendu-testuaren probabilitatea maximizatzen

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du
 - Hitz-bektoreak eta funtzioaren parametroak elkarrekin ikasten dira entrenamendu-testuaren probabilitatea maximizatzeko
- n-gramen hainbat muga gainditu

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du
 - Hitz-bektoreak eta funtzioaren parametroak elkarrekin ikasten dira entrenamendu-testuaren probabilitatea maximizatzeko
- n-gramen hainbat muga gainditu
 - Hitzen arteko antzekotasunak baliatu

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du
 - Hitz-bektoreak eta funtzioaren parametroak elkarrekin ikasten dira entrenamendu-testuaren probabilitatea maximizatzeko
- n-gramen hainbat muga gainditu
 - Hitzen arteko antzekotasunak baliatu
 - *Sparsity* edo sakabanaketa arazoa saihestu

Yoshua Bengio, Réjean Ducharme and Pascal Vincent
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
{bengioy,ducharme,vincentp}@iro.umontreal.ca

- $p(x_i | x_1^{i-1})$ zuzenean modelatu sare neuronal baten bidez
 - Hitz bakoitza bektore baten bidez errepresentatu
 - Funtzio parametrizatu batek testuinguruko hitz-bektoreak hartu eta hurrengo hitzaren probabilitate banaketa ematen du
 - Hitz-bektoreak eta funtzioaren parametroak elkarrekin ikasten dira entrenamendu-testuaren probabilitatea maximizatzeko
- n-gramen hainbat muga gainditu
 - Hitzen arteko antzekotasunak baliatu
 - *Sparsity* edo sakabanaketa arazoa saihestu
 - Testuinguru “mugagabea”

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

1948

2000



HISTORIAUREA



N-GRAMEN AROA

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012



HISTORIAUREA



N-GRAMEN AROA

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012



HISTORIAUREA



N-GRAMEN AROA

2012

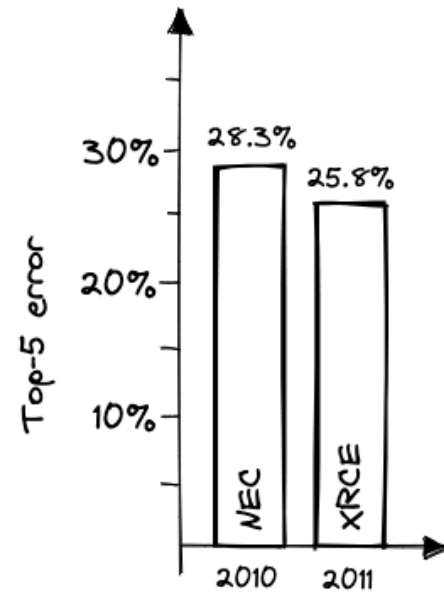
IKASKETA
SAKONA

2012

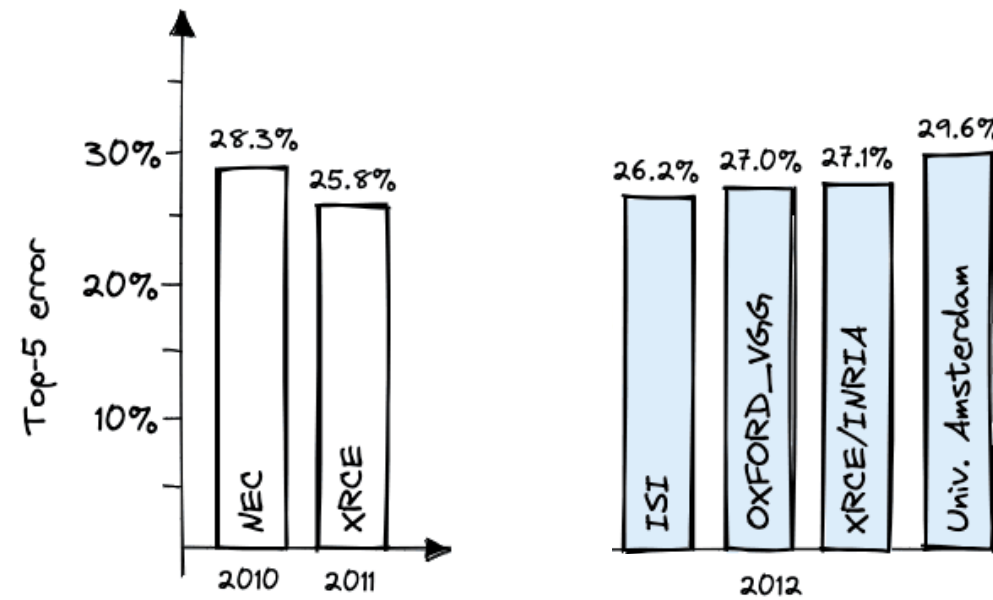
IKASKETA
SAKONA

ImageNet ataza partekatua

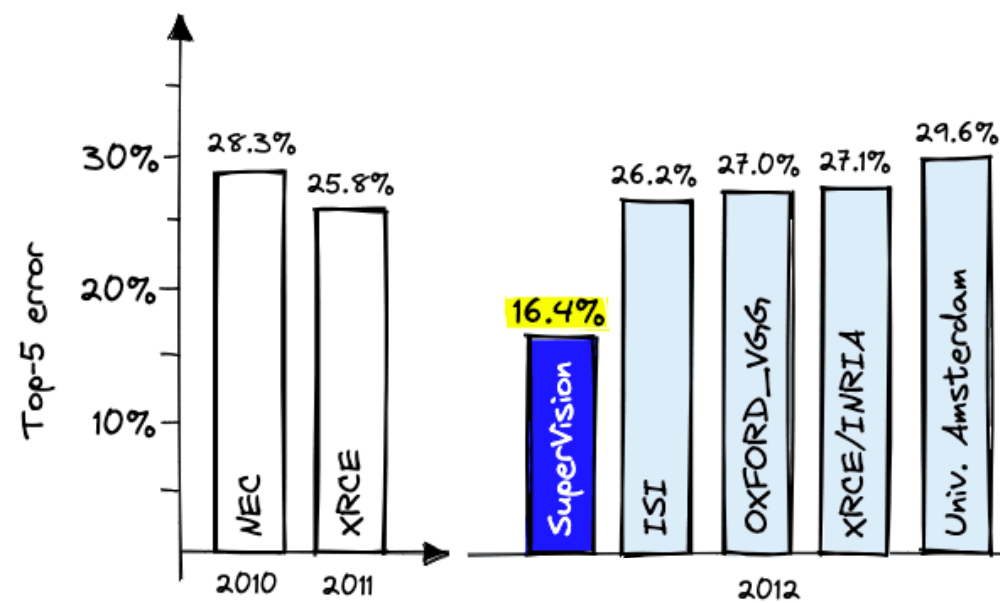
ImageNet ataza partekatua



ImageNet ataza partekatua



ImageNet ataza partekatua



Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012



HISTORIAUREA



N-GRAMEN AROA

Historiatik...

LEHEN EREDU
NEURONALA

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA

2012

MODELATZE
NEURONALAREN AROA

2012+

MODELATZE
NEURONALAREN AROA

2012+

MODELATZE
NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♠] John Hale^{♠♥}
Dani Yogatama[♠] Stephen Clark[♠] Phil Blunsom^{♣♣}**

[♠]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro,cdyer,jthale,dyogatama,clarkstephen,pblunsom}@google.com

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♣] John Hale^{♣♥}
Dani Yogatama[♣] Stephen Clark[♣] Phil Blunsom^{♣♣}**

[♣]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro, cdyer, jthale, dyogatama, clarkstephen, pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang*, Zihang Dai*, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University

{zhiliny, dzihang, rsalakhu, wcohen}@cs.cmu.edu

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♣] John Hale^{♣♥}
Dani Yogatama[♣] Stephen Clark[♣] Phil Blunsom^{♣♣}**

[♣]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro,cdyer,jthale,dyogatama,clarkstephen,pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang*, Zihang Dai*, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University

{zhiliny,dzihang,rsalakhu,wcohen}@cs.cmu.edu

- Diseinu experimental tipikoa:

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♣] John Hale^{♣♥}
Dani Yogatama[♣] Stephen Clark[♣] Phil Blunsom^{♣♣}**

[♣]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro, cdyer, jthale, dyogatama, clarkstephen, pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang^{*}, Zihang Dai^{*}, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University
{zhiliny, dzihang, rsalakhu, wcohen}@cs.cmu.edu

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♠] John Hale^{♠♥}
Dani Yogatama[♠] Stephen Clark[♠] Phil Blunsom^{♣♣}**

[♠]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro, cdyer, jthale, dyogatama, clarkstephen, pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang^{*}, Zihang Dai^{*}, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University
{zhiliny, dzihang, rsalakhu, wcohen}@cs.cmu.edu

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea

2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♠] John Hale^{♠♥}
Dani Yogatama[♠] Stephen Clark[♠] Phil Blunsom^{♣♣}**

[♠]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

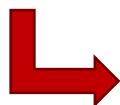
[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro,cdyer,jthale,dyogatama,clarkstephen,pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang*, Zihang Dai*, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University
{zhiliny,dzihang,rsalakhu,wcohen}@cs.cmu.edu

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea



2012+

MODELATZE NEURONALAREN AROA

- Zein da arkitektura edota ikasketa teknika egokiena hizkuntza modelatzeko?

LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better

**Adhiguna Kuncoro^{♣♣} Chris Dyer[♠] John Hale^{♠♥}
Dani Yogatama[♠] Stephen Clark[♠] Phil Blunsom^{♣♣}**

[♠]DeepMind, London, UK

[♣]Department of Computer Science, University of Oxford, UK

[♥]Department of Linguistics, Cornell University, NY, USA

{akuncoro,cdyer,jthale,dyogatama,clarkstephen,pblunsom}@google.com

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang*, Zihang Dai*, Ruslan Salakhutdinov, William W. Cohen
School of Computer Science
Carnegie Mellon University
{zhiliny,dzihang,rsalakhu,wcohen}@cs.cmu.edu


- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea



**Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**

2012+

MODELATZE NEURONALAREN AROA

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea
-  **Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**

2012+

MODELATZE NEURONALAREN AROA

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea



**Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**



2012+

MODELATZE NEURONALAREN AROA

BADAKIT DATU GEHIAGO EDO
EREDU HANDIAGOAK
ERABILITA EMAITZA HOBEAK
LOR DAITEZKEELA

- Diseinu esperimental tipikoa:
 - *Baseline* edo oinarriko sistema
 - Proposatutako teknika darabilen sistema baliokidea



**Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**



2012+

MODELATZE NEURONALAREN AROA

BAINA HORI
INGENIARITZA DA,
EZ ZIENTZIA!!!

BADAKIT DATU GEHIAGO EDO
EREDU HANDIAGOAK
ERABILITA EMAITZA HOBEAK
LOR DAITEZKEELA

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea



**Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**



2012+

MODELATZE NEURONALAREN AROA

BAINA HORI
INGENIARITZA DA,
EZ ZIENTZIA!!!

BADAKIT DATU GEHIAGO EDO
EREDU HANDIAGOAK
ERABILITA EMAITZA HOBEAK
LOR DAITEZKEELA

ETA SATURATZEN
BUKATZEN DU!!!

- Diseinu experimental tipikoa:
 - *Baseline* edo oinarrizko sistema
 - Proposatutako teknika darabilen sistema baliokidea



**Ez da munduko eredurik onena izan behar,
baselinea gainditzea nahikoa hipotesia frogatzeko!**



2012

MODELATZE
NEURONALAREN AROA

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012

2019



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012

2019



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA

2019

GPT-2
GPT-3

2019

GPT-2
GPT-3

Scaling Laws for Neural Language Models

Jared Kaplan *

Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*

OpenAI
sam@openai.com

Tom Henighan

OpenAI
henighan@openai.com

Tom B. Brown

OpenAI
tom@openai.com

Benjamin Chess

OpenAI
bchess@openai.com

Rewon Child

OpenAI
rewon@openai.com

Scott Gray

OpenAI
scott@openai.com

Alec Radford

OpenAI
alec@openai.com

Jeffrey Wu

OpenAI
jeffwu@openai.com

Dario Amodei

OpenAI
damodei@openai.com

2019

GPT-2
GPT-3

- Eskalatzearen hobekuntzak aurreangarriak dira

Scaling Laws for Neural Language Models

Jared Kaplan *

Johns Hopkins University, OpenAI

jaredk@jhu.edu

Sam McCandlish*

OpenAI

sam@openai.com

Tom Henighan

OpenAI

henighan@openai.com

Tom B. Brown

OpenAI

tom@openai.com

Benjamin Chess

OpenAI

bchess@openai.com

Rewon Child

OpenAI

rewon@openai.com

Scott Gray

OpenAI

scott@openai.com

Alec Radford

OpenAI

alec@openai.com

Jeffrey Wu

OpenAI

jeffwu@openai.com

Dario Amodei

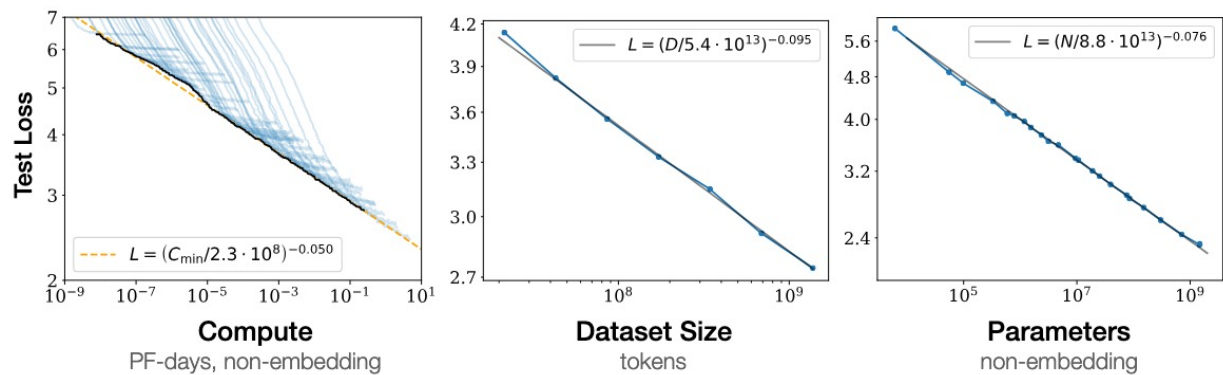
OpenAI

damodei@openai.com

2019

GPT-2
GPT-3

- Eskalatzearen hobekuntzak aurreangarriak dira



Scaling Laws for Neural Language Models

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan
OpenAI
henighan@openai.com

Tom B. Brown
OpenAI
tom@openai.com

Benjamin Chess
OpenAI
bchess@openai.com

Rewon Child
OpenAI
rewon@openai.com

Scott Gray
OpenAI
scott@openai.com

Alec Radford
OpenAI
alec@openai.com

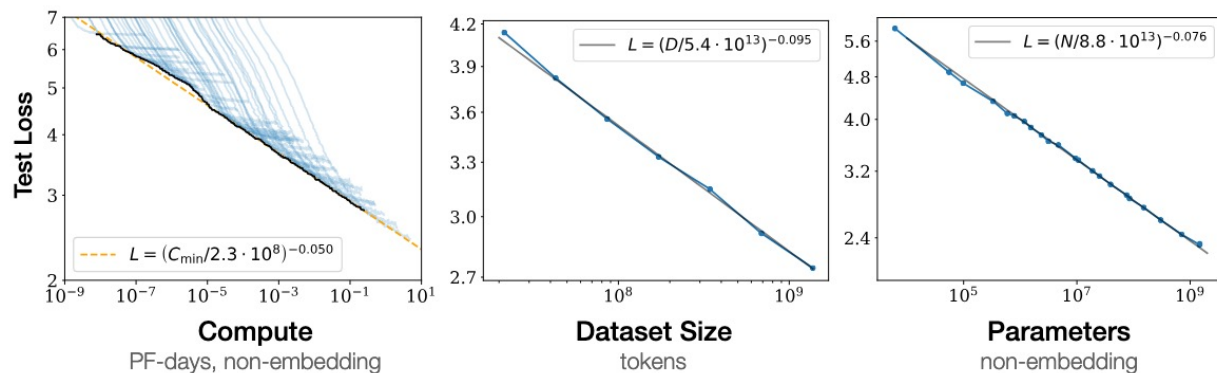
Jeffrey Wu
OpenAI
jeffwu@openai.com

Dario Amodei
OpenAI
damodei@openai.com

2019

GPT-2
GPT-3

- Eskalatzearen hobekuntzak aurreangarriak dira



- Eskalatzea zientzia (ere) bada!

Scaling Laws for Neural Language Models

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan
OpenAI
henighan@openai.com

Tom B. Brown
OpenAI
tom@openai.com

Benjamin Chess
OpenAI
bchess@openai.com

Rewon Child
OpenAI
rewon@openai.com

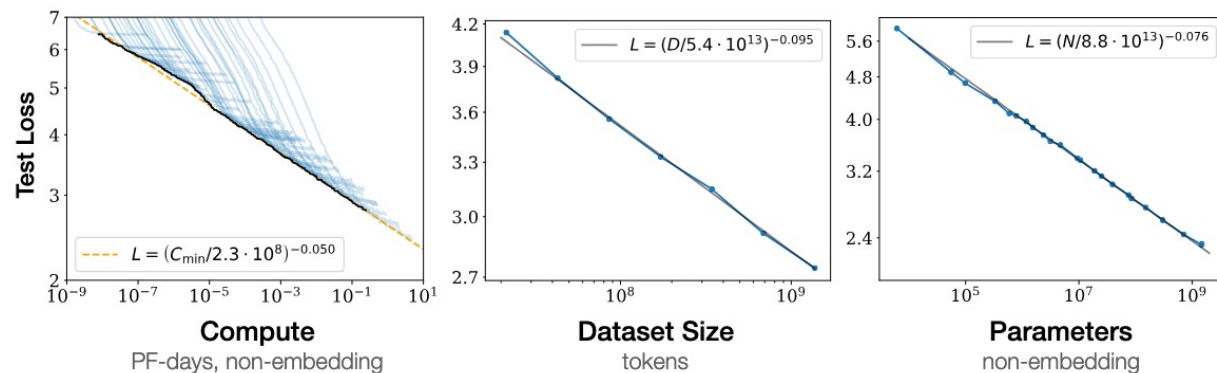
Scott Gray
OpenAI
scott@openai.com

Alec Radford
OpenAI
alec@openai.com

Jeffrey Wu
OpenAI
jeffwu@openai.com

Dario Amodei
OpenAI
damodei@openai.com

- Eskalatzearen hobekuntzak aurreangarriak dira



- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobeak eraikitzeko estrategia segurua

Scaling Laws for Neural Language Models

Jared Kaplan *

Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*

OpenAI
sam@openai.com

Tom Henighan

OpenAI

henighan@openai.com

Tom B. Brown

OpenAI

tom@openai.com

Benjamin Chess

OpenAI

bchess@openai.com

Rewon Child

OpenAI

rewon@openai.com

Scott Gray

OpenAI

scott@openai.com

Alec Radford

OpenAI

alec@openai.com

Jeffrey Wu

OpenAI

jeffwu@openai.com

Dario Amodei

OpenAI

damodei@openai.com

2019

GPT-2
GPT-3

Scaling Laws for Neural Language Models

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan
OpenAI
henighan@openai.com

Tom B. Brown
OpenAI
tom@openai.com

Benjamin Chess
OpenAI
bchess@openai.com

Rewon Child
OpenAI
rewon@openai.com

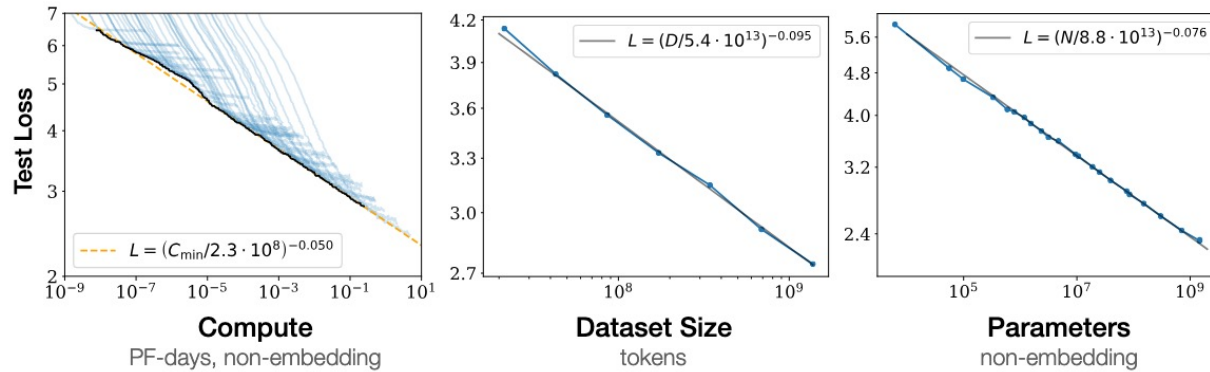
Scott Gray
OpenAI
scott@openai.com

Alec Radford
OpenAI
alec@openai.com

Jeffrey Wu
OpenAI
jeffwu@openai.com

Dario Amodei
OpenAI
damodei@openai.com

- Eskalatzearen hobekuntzak aurreangarriak dira

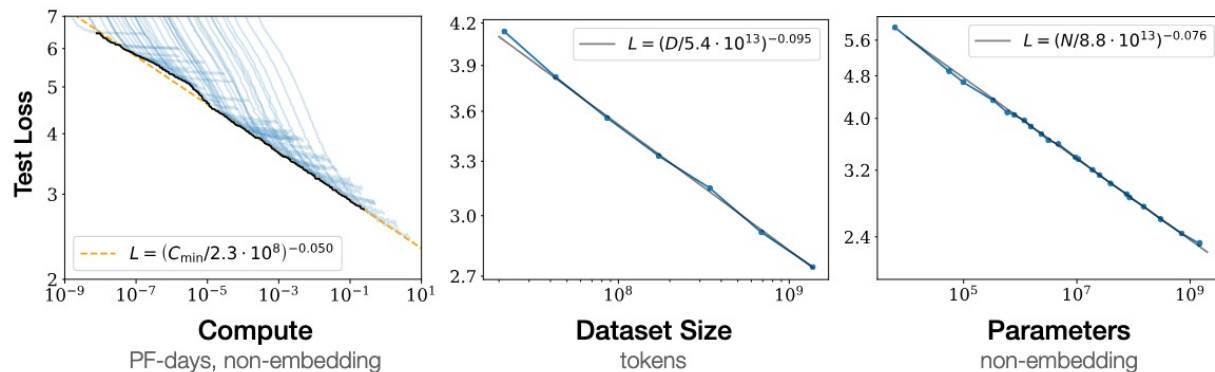


- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobekiak eraikitzekeo estrategia segurua ... **baina zertarako?**

2019

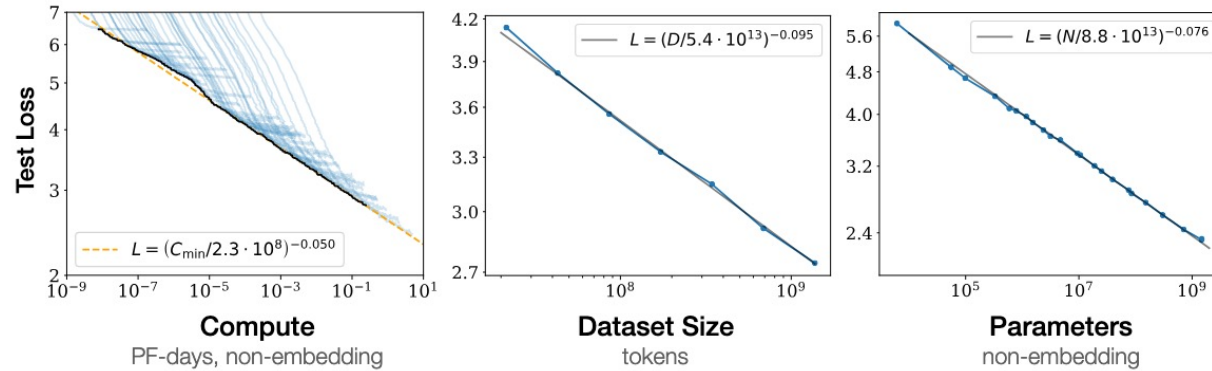
GPT-2
GPT-3

- Eskalatzearen hobekuntzak aurreangarriak dira



- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobeak eraikitzeko estrategia segurua ... **baina zertarako?**

- Eskalatzearen hobekuntzak auresangarriak dira



- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobeak eraikitzeko estrategia segurua ... **baina zertarako?**

2019

GPT-2
GPT-3

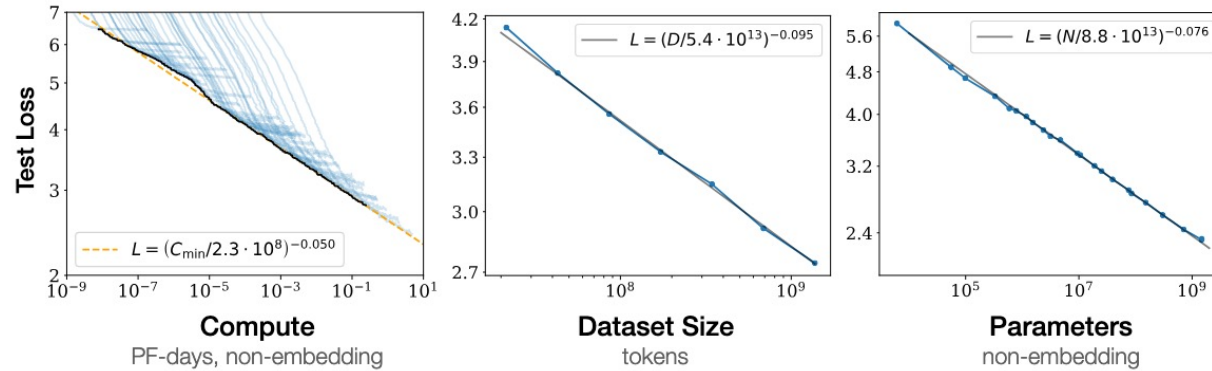
Language Models are Unsupervised Multitask Learners

Alec Radford *1 Jeffrey Wu *1 Rewon Child1 David Luan1 Dario Amodei **1 Ilya Sutskever **1

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan1	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Matusz Litwin Scott Gray
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

- Eskalatzearen hobekuntzak auresangarriak dira



- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobekiak eraikitzeko estrategia segurua ... **baina zertarako?**

2019

GPT-2
GPT-3

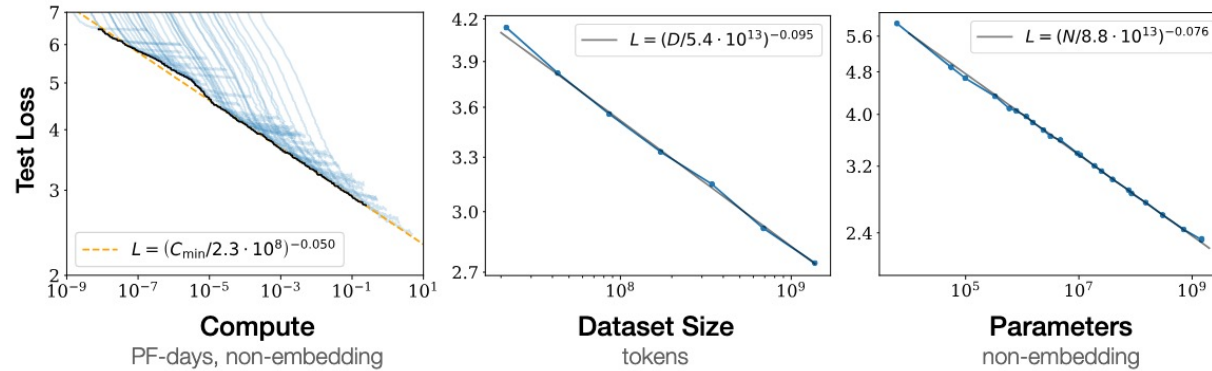
Language Models are Unsupervised Multitask Learners

Alec Radford *1 Jeffrey Wu *1 Rewon Child1 David Luan1 Dario Amodei **1 Ilya Sutskever **1

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan1	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Matusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

- Eskalatzearen hobekuntzak aurreangarriak dira



- Eskalatzea zientzia (ere) bada!
- Hizkuntza eredu hobekuntza erakitzekeo estrategia segurua ... **baina zertarako?**

Hobetu ahala hizkuntza ereduak gaitasun berriak ikasten dituzte!

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford ^{*1} Jeffrey Wu ^{*1} Rewon Child ¹ David Luan ¹ Dario Amodei ^{**1} Ilya Sutskever ^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford ^{*1} Jeffrey Wu ^{*1} Rewon Child ¹ David Luan ¹ Dario Amodei ^{**1} Ilya Sutskever ^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford ^{*1} Jeffrey Wu ^{*1} Rewon Child ¹ David Luan ¹ Dario Amodei ^{**1} Ilya Sutskever ^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford ^{*1} Jeffrey Wu ^{*1} Rewon Child ¹ David Luan ¹ Dario Amodei ^{**1} Ilya Sutskever ^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____



2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

} **patata**

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{ patata
dut

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{
patata
dut
du

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{
patata
dut
du

Ikuspegi modernoa (eskala handia)

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{
patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

AEBko hiriburua _____

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ {
 patata
 dut
 du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduak askotariko ezagutza biltzen dute

AEBko hiriburua _____ {

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ {
patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

AEBko hiriburua _____ {
patata

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ {
patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

AEBko hiriburua _____ {
patata
New York da

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
 Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
 Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
 Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
 Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
 Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
 Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
 Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ {
 patata
 dut
 du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

AEBko hiriburua _____ {
 patata
 New York da
 Washington da

2019

GPT-2
GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown^{*} Benjamin Mann^{*} Nick Ryder^{*} Melanie Subbiah^{*}
Jared Kaplan¹ Prafulla Dhariwal¹ Arvind Neelakantan¹ Pranav Shyam¹
Girish Sastry¹ Amanda Askell¹ Sandhini Agarwal¹ Ariel Herbert-Voss¹
Gretchen Krueger¹ Tom Henighan¹ Rewon Child¹ Aditya Ramesh¹
Daniel M. Ziegler¹ Jeffrey Wu¹ Clemens Winter¹
Christopher Hesse¹ Mark Chen¹ Eric Sigler¹ Mateusz Litwin¹ Scott Gray¹
Benjamin Chess¹ Jack Clark¹ Christopher Berner¹
Sam McCandlish¹ Alec Radford¹ Ilya Sutskever¹ Dario Amodei¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{
patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown [*]	Benjamin Mann [*]	Nick Ryder [*]	Melanie Subbiah [*]
Jared Kaplan ¹	Prafulla Dhariwal ¹	Arvind Neelakantan ¹	Pranav Shyam ¹
Girish Sastry ¹	Amanda Askell ¹	Sandhini Agarwal ¹	Ariel Herbert-Voss ¹
Gretchen Krueger ¹	Tom Henighan ¹	Rewon Child ¹	Aditya Ramesh ¹
Daniel M. Ziegler ¹	Jeffrey Wu ¹	Clemens Winter ¹	
Christopher Hesse ¹	Mark Chen ¹	Eric Sigler ¹	Mateusz Litwin ¹ Scott Gray ¹
Benjamin Chess ¹	Jack Clark ¹	Christopher Berner ¹	
Sam McCandlish ¹	Alec Radford ¹	Ilya Sutskever ¹	Dario Amodei ¹

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ {
 patata
 dut
 du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____ {

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{

patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{

patata

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{

patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{

patata
726

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{

patata
dut
du

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{

patata
726
868

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown [*]	Benjamin Mann [*]	Nick Ryder [*]	Melanie Subbiah [*]
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{	patata
	dut
	du

- Hizkuntza ereduak helburu zehatzeko aplikazioen osagaitzat

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{	patata
	726
	868

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown [*]	Benjamin Mann [*]	Nick Ryder [*]	Melanie Subbiah [*]
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Matusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{	patata
	dut
	du

- Hizkuntza ereduak helburu zehatzeko aplikazioen osagaitzat

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{	patata
	726
	868

- Hizkuntza ereduek interes propioa helburu orokorreko aplikaziotzat

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan ¹	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin
Benjamin Chess	Jack Clark	Christopher Berner	
Sam McCandlish	Alec Radford	Ilya Sutskever	Dario Amodei

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduek ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____

{

patata
dut
du

- Hizkuntza ereduak helburu zehatzeko aplikazioen osagaitzat

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduek askotariko ezagutza biltzen dute

546+322= _____

{

patata
726
868

- Hizkuntza ereduek interes propioa helburu orokorreko aplikaziotzat
 - Edozein ataza hizkuntzaren modelatze modura formula daiteke!

Ikuspegi klasikoa (eskala txikia)

- Hizkuntza ereduak ezagutza linguistikoa biltzen dute (nagusiki)

Gaur goizean euria egin _____ { **patata**
dut
du

- Hizkuntza ereduak helburu zehatzeko aplikazioen osagaitzat

Ikuspegi modernoa (eskala handia)

- Hizkuntza ereduak askotariko ezagutza biltzen dute

546+322= _____ { **patata**
726
868

- Hizkuntza ereduak interes propioa helburu orokorreko aplikaziotzat
 - Edozein ataza hizkuntzaren modelatze modura formula daiteke!

Gaztelania: Esta mañana ha llovido.
Euskara: _____

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012

2019



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

1948

2000

2012

2019



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA



LERROKATZE ETA
ESKALAREN AROA

2019

LERROKATZE ETA
ESKALAREN AROA

2019+

LERROKATZE ETA
ESKALAREN AROA

2019+

LERROKATZE ETA
ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

2019+

LERROKATZE ETA ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

Zein da AEBko hiriburua? _____

2019+

LERROKATZE ETA
ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

Zein da AEBko hiriburua? _____



2019+

LERROKATZE ETA
ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

Zein da AEBko hiriburua? _____

patata

2019+

LERROKATZE ETA
ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

Zein da AEBko hiriburua? _____

{
patata
New York

2019+

LERROKATZE ETA
ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute

Zein da AEBko hiriburua? _____

patata

New York

Washington

2019+

LERROKATZE ETA ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute
- Baina ezagutza hori azalera zaila izan daiteke!

Zein da AEBko hiriburua? _____

patata

New York

Washington

2019+

LERROKATZE ETA ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute
- Baina ezagutza hori azalera zaila izan daiteke!

Zein da AEBko hiriburua? _____

patata

New York

Washington

Eta Kanadako?

2019+

LERROKATZE ETA ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute
- Baina ezagutza hori azalera zaila izan daiteke!

Zein da AEBko hiriburua? _____

patata

New York

Washington

Eta Kanadako?

Auskalo!

2019+

LERROKATZE ETA ESKALAREN AROA

- Eskalatu ahala, hizkuntza ereduak ezagutza handia biltzen dute
- Baina ezagutza hori azalera zaila izan daiteke!

Zein da AEBko hiriburua? _____

patata

New York

Washington

Eta Kanadako?

Auskalo!

- Nola “domestikatu” hizkuntza ereduak?

2019+

LERROKATZE ETA
ESKALAREN AROA

2019+ | LERROKATZE ETA
ESKALAREN AROA

1. FASEA: PROMPT INGENIARITZA

2019+

LERROKATZE ETA
ESKALAREN AROA

1. FASEA: PROMPT INGENIARITZA

- Ataza ezberdinak hizkuntza eredu bati formulatzeko modu egokia bilatu

2019+

LERROKATZE ETA ESKALAREN AROA

1. FASEA: PROMPT INGENIARITZA

- Ataza ezberdinak hizkuntza eredu bati formulatzeko modu egokia bilatu

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou

Google Research, Brain Team
{jasonwei,dennyzhou}@google.com

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

1. FASEA: PROMPT INGENIARITZA

- Ataza ezberdinak hizkuntza eredu bati formulatzeko modu egokia bilatu

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma

Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou

Google Research, Brain Team
{jasonwei,dennyzhou}@google.com

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Large Language Models are Zero-Shot Reasoners

Takeshi Kojima

The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu

Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ❌

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✅

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is

(Output) 8 ❌

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: Let's think step by step.

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✅

2019+

LERROKATZE ETA
ESKALAREN AROA

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)

[USER]: Kaixo!

[AI]: Kaixo! Zertan lagun zaitzaket?

[USER]: Zein da AEBko hiriburua?

[AI]: AEBko hiriburua Washington da.

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)

[USER]: Zein da Kanadako hiriburua?

[AI]: _____

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)

2019+

LERROKATZE ETA
ESKALAREN AROA

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)
- Baina...

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)
- Baina...
 - Kalitatezko galdera-erantzunak biltzea ez da erraza!

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)
- Baina...
 - Kalitatezko galdera-erantzunak biltzea ez da erraza!
 - Erantzun zuzenen espazioa zabala da eta guztiak ikastea ez da erraza (eta ezta beharrezkoa ere)!

2. FASEA: IKASKETA GAINBEGIRATUA

- Galdera-erantzunen multzo bat bildu (txtat formatuan)
- Hizkuntza ereduaren ikasketa jarraitu multzo horren gainean
- Ezagutza lehenagotik dauka, ikasten dena interfazea (nagusiki)
- Baina...
 - Kalitatezko galdera-erantzunak biltzea ez da erraza!
 - Erantzun zuzenen espazioa zabala da eta guztiak ikastea ez da erraza (eta ezta beharrezkoa ere)!

[USER]: Kontatu txiste bat

[AI]: _____

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz

[USER]: Zein da AEBko hiriburua?

[AI]: _____

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz

[USER]: Zein da AEBko hiriburua?

[AI]: _____

A aukera

Ez daukat

arrastorik ere!

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz

[USER]: Zein da AEBko hiriburua?

[AI]: _____

A aukera

Ez daukat
arrastorik ere!

B aukera

AEBko hiriburua
New York da.

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz

[USER]: Zein da AEBko hiriburua?

[AI]: _____

A aukera

Ez daukat
arrastorik ere!

B aukera

AEBko hiriburua
New York da.

C aukera

AEBko hiriburua
Washington da.

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz
- Anotatzaile batek erantzun onena aukeratzen du

[USER]: Zein da AEBko hiriburua?

[AI]: _____

A aukera

Ez daukat
arrastorik ere!

B aukera

AEBko hiriburua
New York da.

C aukera

AEBko hiriburua
Washington da.

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz
- Anotatzaile batek erantzun onena aukeratzen du

[USER]: Zein da AEBko hiriburua?

[AI]: _____

A aukera

Ez daukat
arrastorik ere!

B aukera

AEBko hiriburua
New York da.

C aukera

AEBko hiriburua
Washington da.

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz
- Anotatzaile batek erantzun onena aukeratzen du

2019+

LERROKATZE ETA
ESKALAREN AROA

3. FASEA: PREFERENTZIEN BIDEZKO IKASKETA

- Galdera bakoitzeko hainbat erantzun posible sortu hizkuntza eredu bat erabiliz
- Anotatzaile batek erantzun onena aukeratzen du
- Hizkuntza ereduari anotatzaileak aukeratzea probableagoa den erantzunak sortzen irakatsi (PPO, DPO...)

2019

LERROKATZE ETA
ESKALAREN AROA

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

CHATGPT

1948

2000

2012

2019

2022



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA



LERROKATZE ETA
ESKALAREN AROA

Historiatik...

LEHEN EREDU
NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA
EREDUA

IKASKETA
SAKONA

CHATGPT

1948

2000

2012

2019

2022



HISTORIAUREA



N-GRAMEN AROA



MODELATZE
NEURONALAREN AROA



LERROKATZE ETA
ESKALAREN AROA



HISTERIAREN
AROA

...HISTERIARA

LEHEN EREDU NEURONALA

GPT-2
GPT-3

LEHEN HIZKUNTZA EREDUA

IKASKETA SAKONA

CHATGPT

1948

2000

2012

2019

2022



HISTORIAUREA

N-GRAMEN AROA

MODELATZE NEURONALAREN AROA

LERROKATZE ETA ESKALAREN AROA

HISTERIAREN AROA

...HISTERIARA

MODELATZE
NEURONALAREN AROA

HISTERIAREN
AROA

...HISTERIARA

MODELATZE
NEURONALAREN AROA

BREAKING THE SOFTMAX BOTTLENECK:
A HIGH-RANK RNN LANGUAGE MODEL

Zhilin Yang*, **Zihang Dai***, **Ruslan Salakhutdinov**, **William W. Cohen**
School of Computer Science
Carnegie Mellon University
{zhiliny, dzihang, rsalakhu, wcohen}@cs.cmu.edu

HISTERIAREN
AROA

...**HISTERIARA**

**MODELATZE
NEURONALAREN AROA**

**BREAKING THE SOFTMAX BOTTLENECK:
A HIGH-RANK RNN LANGUAGE MODEL**

Zhilin Yang*, **Zihang Dai***, **Ruslan Salakhutdinov**, **William W. Cohen**
School of Computer Science
Carnegie Mellon University
{zhiliny, dzihang, rsalakhu, wcohen}@cs.cmu.edu

**HISTERIAREN
AROA**

GPT-4 Technical Report

OpenAI*

...HISTERIARA

MODELATZE
NEURONALAREN AROA

HISTERIAREN
AROA

...HISTERIARA

MODELATZE NEURONALAREN AROA

- Ideiak garrantzitsuak

HISTERIAREN AROA

...HISTERIARA

MODELATZE NEURONALAREN AROA

- Ideiak garrantzitsuak

HISTERIAREN AROA

- Exekuzioa garrantzitsua

...HISTERIARA

MODELATZE NEURONALAREN AROA

- Ideiak garrantzitsuak
- Proiektu txikiak

HISTERIAREN AROA

- Exekuzioa garrantzitsua

...HISTERIARA

MODELATZE NEURONALAREN AROA

- Ideiak garrantzitsuak
- Proiektu txikiak

HISTERIAREN AROA

- Exekuzioa garrantzitsua
- Proiektu erraldoiak

...HISTERIARA

...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...

...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!

...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



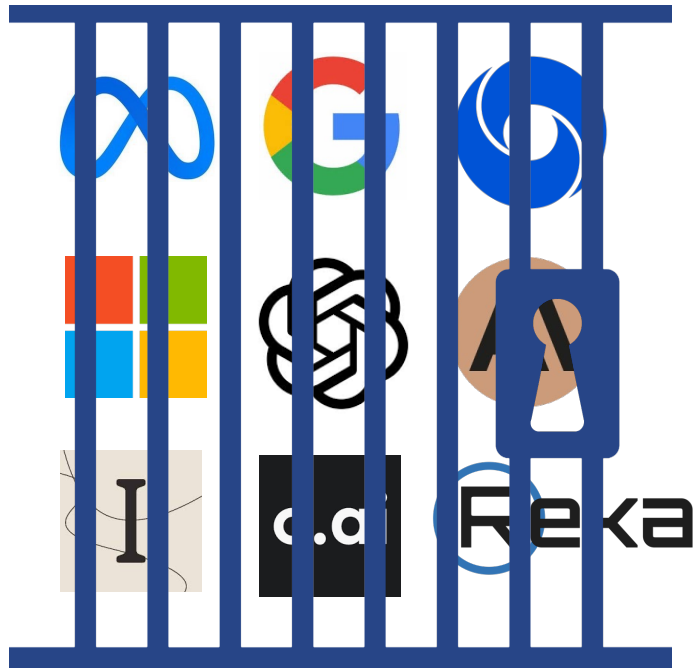
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



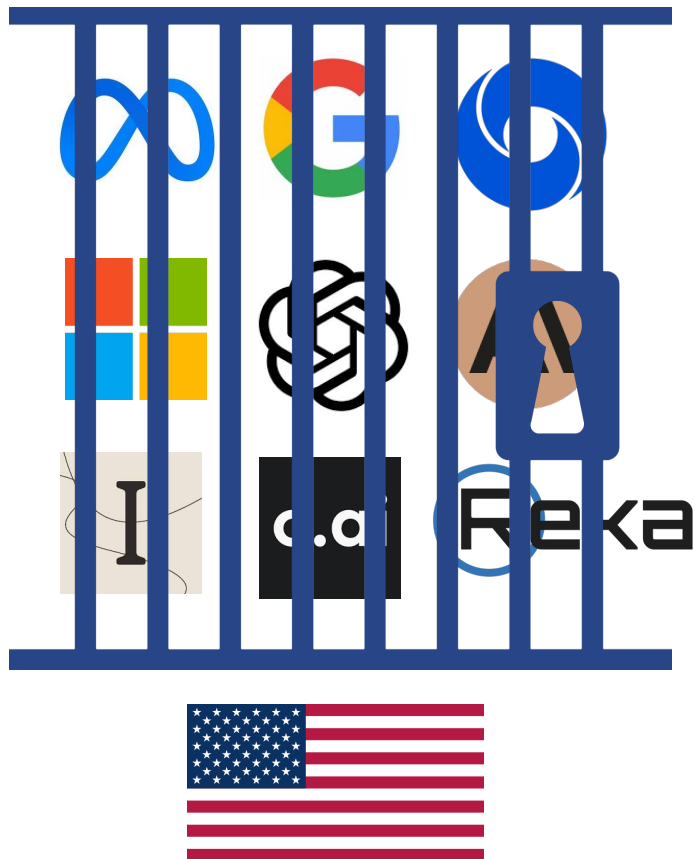
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



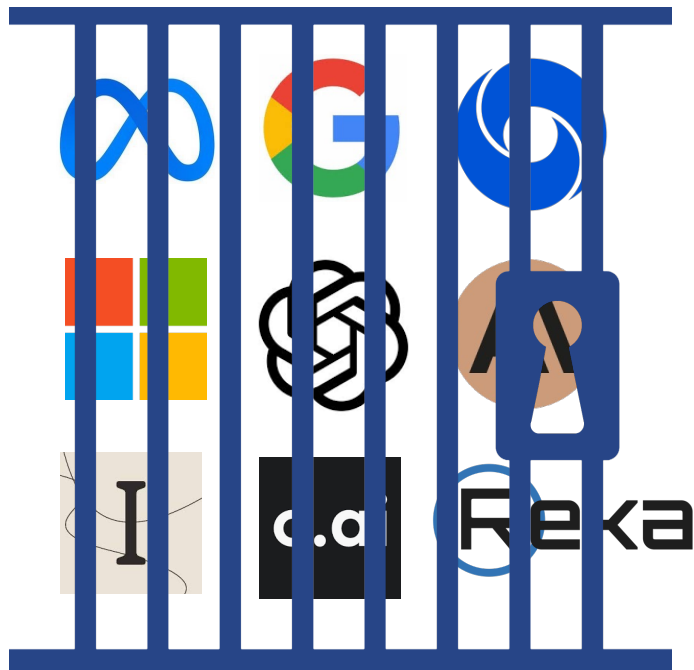
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



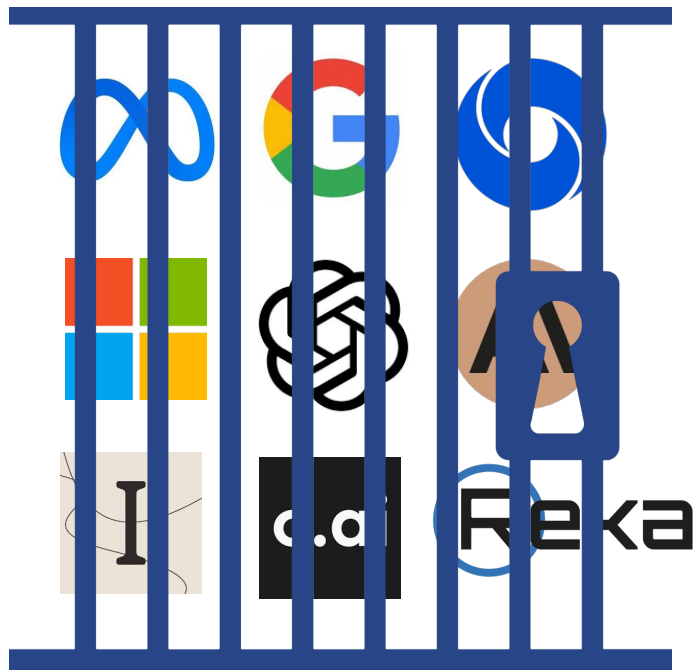
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



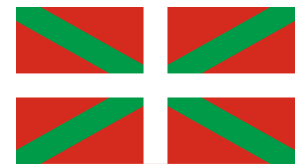
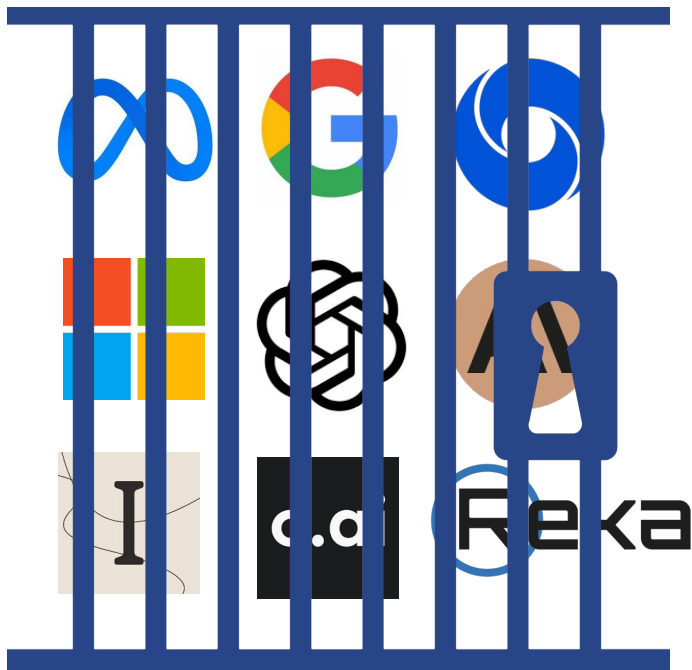
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



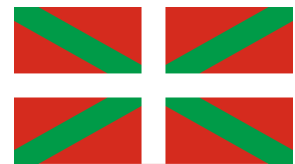
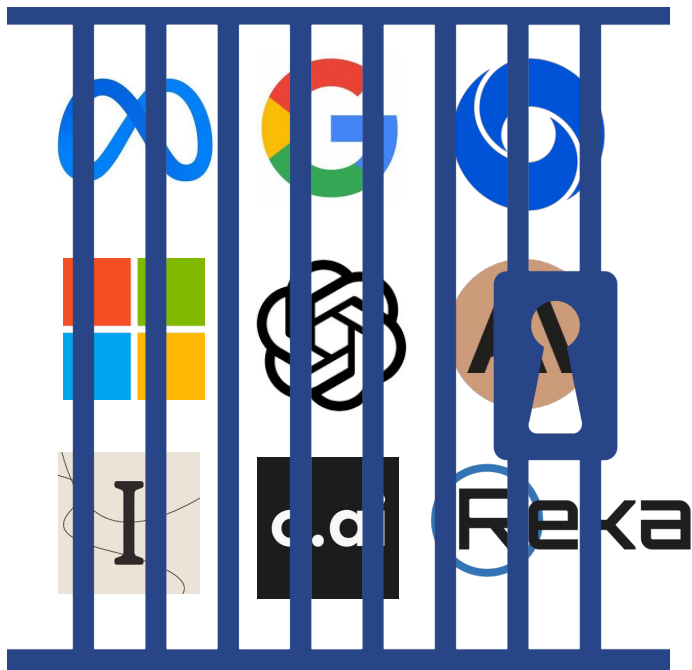
...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



...HISTERIARA

- Garai berri hau hobeto egokitzen da erraldoi teknologikoetara...
- ...baina erraldoi teknologikoak ere sakonki eraldatu dira!



ESKERRIK ASKO!

 mikel@reka.ai

 [@artetxem](https://twitter.com/artetxem)