|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| hello  helllo  helo  hello  hello  helo  hello  hello  hello  ello  elo  hello  hello  helo  heloo  hello  hello  ello  heloo  heloo  hello  ello  hello  hello  ello  heloo  hello  helloo  hello  helllo  hello  hello  ello  heloo  hello  ello  heloo  helloo  hello  helllo  hello  hello  Hàng 43: rỗng. end of string: eos | |  | | --- | | 98 | | 39 | | 4 | | 1. 93 – e | | 22 – h | | 12 – h | | 22- h | | 6 – h | | 80 – h | | 68 – h | | 1. 52 - h | | 73 – h | | 66 – h | | 15 - h | | 35 – h | | 43 – h | | 97 – e | | 70 – h | | 43 - h | | 1. 57 - h | | 69 – h | | 85 – e | | 15 – h | | 97 – e | | 57 – h | | 1. 63 – h | | 10 – h | | 24 – h | | 32 – h | | 100 – eos | | 47 – h | | 9 – h | | 66 – h | | 99 - e | |

h: 35/43 = 81.4% (đếm tay, có tổng cộng 35 mẫu chữ h đầu tiên/tổng 43 mẫu)

e: 7/43 = 16.2%

l: 0

o: 0

eos: 1/43 = 2.32%

---

random từ 1-100, nếu rời vào 0-81 là h, 82 -> 98 là e, 99-100 là eos

→

E

p(?|e) = Nếu biết ký tự đầu tiên là chữ e, cho biết xác suất

char[0] = e

char[1]

h: 0

e: 0

l: 7/7

o: 0

eos: 0

---

el

p(?|”el”)

h: 0

e: 0

l: 6/7

o: 1/7

eos: 0

---

ell

p(?|”ell”)

h: 0

e: 0

l: 0

o: 6/6

eos: 0

===

Lần 1: 93 22 12 22 6

e

el

ell p( l | el) = 85.71%

ello p( o | ell) = 100%

ello<eos> p(<eos> | ello) = 100%

**Output: ello**

Lần 2: 52 73 66 15 35 43

h p( h | null) = 83.33%

e p( e | h) = 100%

l p( l | he) = 100%

l p( l | hel) = 61.9% (26/42)

o p(o | hell) = 88.46% (23/26)

<eos> p(<eos> | hello) = 91.3% (21/23)

**Output: hello**

Lần 3: 57 69 85 15 97 57

h p( h | null) = 83.33%

e

l

l p( l | hel) = 61.9%

l p( l | hell) = 11.54% (3/26) do 97 > 88.46

o p( o | helll) = 100% (3/3)

<eos> p(<eos> | helllo) = 100%

**Output: helllo**

-> p(n? | n-1 - gram). Cho chuỗi n-1 gram trước đó. XS của n là chữ gì?

**47:55** Đây là kết quả của GPT-2, GPT-3. Nhập đoạn văn đầu tiên, sau đó máy sẽ tự xuất ra. Cho input là đoạn văn Harry Potter, máy tự xuất ra các đoạn văn tương tự cách viết của truyện HP.

**Chạy từ lần 4: 63 10 24 32 100 47 9 66 99**

1) Nhập vào chữ “he”. XS ra chữ gì? he**L**

2) Nhập vào chữ “o”. Xuất ra gì?

+ “o”. Bắt đầu bằng chữ o, XS các chữ kia là bao nhiêu?

h: 0%

e: 0%

l: 0%

o: 0%

p(o|eos) = 0% -> Báo “Bạn đã nhập sai”

+ “ho” -> Bạn đã nhập sai

Đầu tiên xét n-1 gram, check từng step xem chữ có hợp lý chưa. Check h -> check ho,

P(?|h)

h: 0%

e: 100%

l: 0%

o: 0%

eos: 0%

Quy trình check:

P(h|eos) = 81%

P(o|h) = 0% -> quá thấp nên nhập “ho” không có nghĩa. Báo nhập sai, Warning

Tiếp theo, Correction

**Dựa vào đây:**

**h: 35/43 = 81.4%**

**e: 7/43 = 16.2%**

**l: 0**

**o: 0**

**eos: 1/43 = 2.32%**

P(o|eos) = 0% -> Tùy vào ngưỡng % chính xác mà sẽ đưa ra gợi ý phù hợp. VD: Trên 50% -> gợi ý nhập h.

Thấy P(e|h) = 100% -> Gợi ý he thay vì ho