



14 Probability problems for acing Data Science interviews

Decrypting probability questions in data science interviews with style



Aakash Agrawal · Follow

Published in Towards Data Science · 7 min read · Jun 1, 2022



435



2





photo on unsplash.com by [@briansuman](#)

Questions on probability are very common in any Data Science interview. The questions might be challenging (and tricky) but are easily tractable if you have some practice and know basic formulas and concepts. In this blog, I share some practice questions (with solutions) on different concepts in probability.

Keywords: *Probability, Binomial distribution, Bayes theorem*

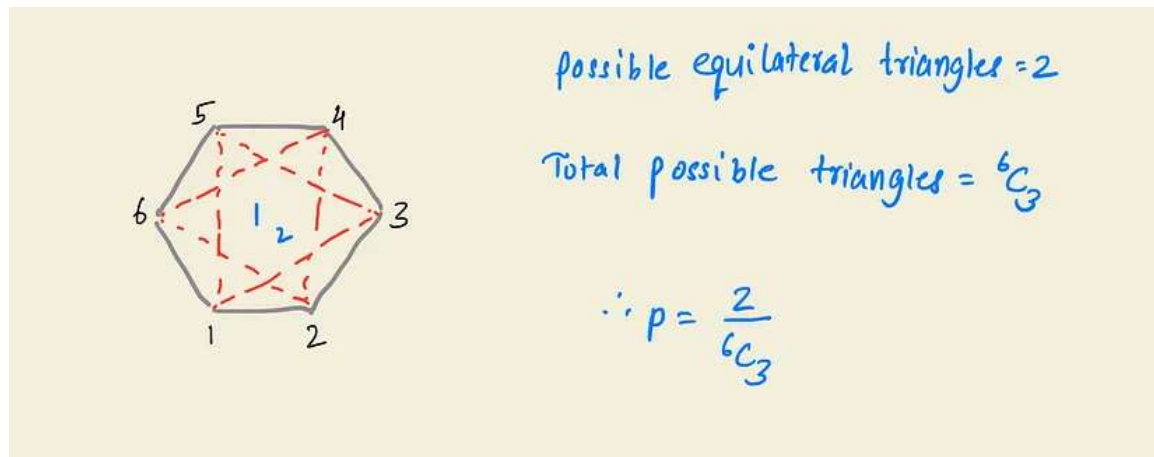
The blog assumes the reader knows basic probability formulas and concepts. Please refer to the reference section for some good reading materials on these concepts. I would recommend the readers to first try and solve the

questions themselves on a piece of paper and then head to the hints and solutions. Note: there might be many different ways of doing the same problem. The solution provided is just one of the ways (might be the only way).

Basic Probability

Q1. 3 vertices (corners) of a regular hexagon are randomly joined. What is the probability that an equilateral triangle is formed?

Note: In a regular hexagon all sides and angle measurements are equal. And an equilateral triangle has all three sides equal. *Answer = 0.1.*



Answer Q1. Image by the author.

Q2. 3 persons A, B, C independently fire at a target. What is the probability that (i) Exactly one of them hits the target, (ii) At least one of them hits the target?
Given: Probability of hitting the target. $P(A) = 1/6$, $P(B) = 1/4$, $P(C) = 1/3$.

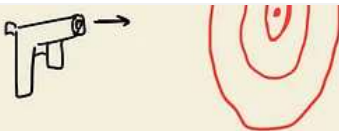
(i) Probability that exactly one of them hits the target requires the other two not to hit the target. We can readily see three cases for this event to happen. Finally, we calculate the probability by taking a union of those cases.

(ii) Probability that atleast one of them hits the target is solved by creating several cases and taking a union as we did in the earlier part. A much easier approach will be to calculate the negation of the same event and subtract it from 1. (Since the firings are independent $P(ABC)$ becomes $P(A)P(B)P(C)$).

Medium

Search

Write

A, B, C

(i) $P(\text{Exactly one of them hits target}) -$

CASES $\rightarrow A\bar{B}\bar{C}, \bar{A}\bar{B}C, \bar{A}B\bar{C}$

$\therefore P(A\bar{B}\bar{C} \cup \bar{A}\bar{B}C \cup \bar{A}B\bar{C})$

$= \frac{1}{6} \times \frac{3}{4} \times \frac{2}{3} + \frac{5}{6} \times \frac{1}{4} \times \frac{2}{3} + \frac{5}{6} \times \frac{3}{4} \times \frac{1}{3} \approx 0.43$

$\bar{B} \rightarrow P(B \text{ won't hit})$

$\cup \rightarrow \text{Union}$

Events are independent.

$\therefore P(ABC) = P(A)P(B)P(C)$

(ii) $P(\text{Atleast one of them hits target}) -$

(using 1-None approach)

$p = 1 - P(\bar{A}\bar{B}\bar{C}) = 1 - P(\bar{A})P(\bar{B})P(\bar{C})$

$= 1 - \frac{5}{6} \times \frac{3}{4} \times \frac{2}{3} \approx 0.58$

Q3. The probability that a teacher takes a surprise test is 0.55. If a student remains absent for two days. What is the probability that he misses exactly one test, and atmost one test?

(i) Similar to the previous question. (ii) Missing atmost 1 test means missing either 0 test or 1 test.

PS: this is similar to the Uber and Lyft problem.



Negation implies not taking the test.

$$P(T_1 \cap \bar{T}_2) = P(T_1) P(\bar{T}_2)$$

As the events are independent

$P(T_i)$ = Probability that teacher takes test on day i . = 0.55

(i) $P(\text{Missing Exactly 1 test}) \rightarrow$

$$P = P(T_1 \cap \bar{T}_2 \cup \bar{T}_1 \cap T_2)$$

$$= P(T_1 \cap \bar{T}_2) + P(\bar{T}_1 \cap T_2)$$

$$= 0.55 \times 0.45 + 0.45 \times 0.55$$

$$P = 0.495$$

(ii) $P(\text{Miss atmost 1 test}) =$

$$P(\text{Miss no test}) + P(\text{Miss 1 test})$$


$$P = 0.45 \times 0.45 + 0.495$$

$$P = 0.6975$$

Answer Q3. Image by the author.

Q4. A box contains 2 defective pens and 3 working pens. Pens are tested one by one until both defective ones are discovered. What is the probability that the testing procedure comes to an end at the end of (i) 2nd testing, (ii) 3rd testing?

For the test to come to an end at the end of two checks, the first two starting pens need to be defective. For the test to come to an end at the end of three checks, we can create cases and take the union.



$P(D)$ = probability that the pen is defective

(i) Testing procedure comes at the end of 2nd test
 $\Rightarrow P(\text{starting 2 pens defective})$
 2 Defective (D) $P(DD) = \frac{2}{5} \times \frac{1}{4}$
 3 Working (W)

(ii) Testing procedure comes at the end of 3rd test -
 CASES - WDD, DWD, WWW (remaining 2 defective).
 $= P(WDD) + P(DWD) + P(WWW)$ {Union}
 $= \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} + \frac{2}{5} \times \frac{3}{4} \times \frac{1}{3} + \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \approx 0.3$

Answer Q4. Image by the author.

Q5. If there are 30 people in a room, what is the probability that everyone has different birthdays? Assume 365 possible birthdays in a year.



Total possible ways 30 people can have birthdays = $(365)^{30}$

1st person: 365 ways to choose

2nd person: 364 ways to choose

⋮

30th person: 336 ways to choose -

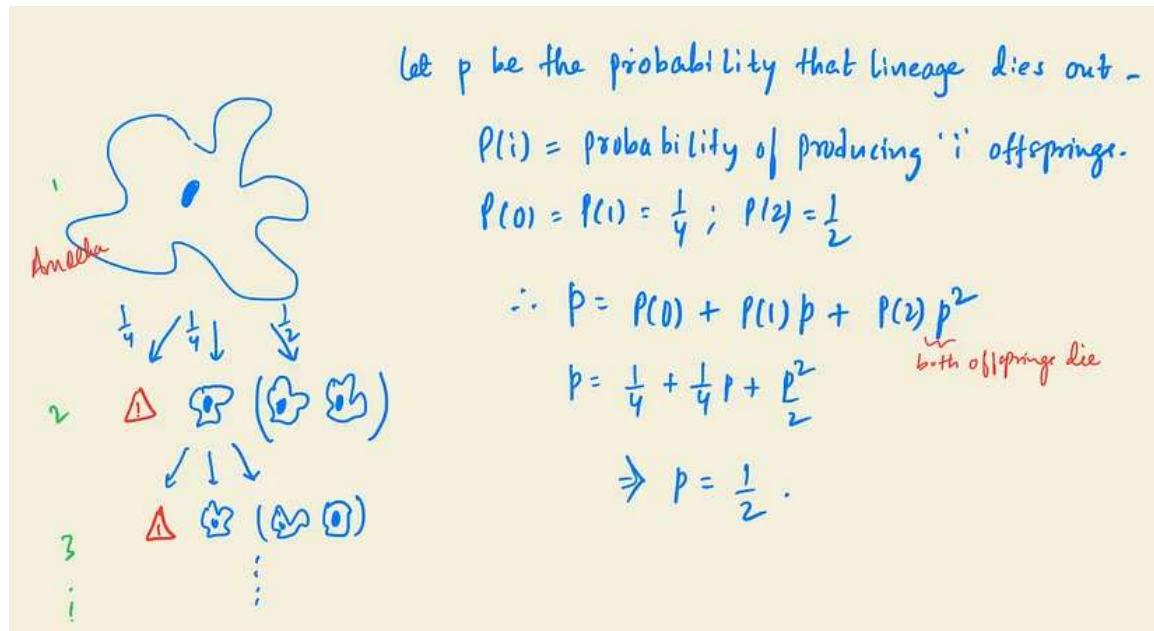
$$\therefore p = \frac{365 \cdot 364 \cdot \dots \cdot 336}{(365)^{30}}$$

Answer Q5. Image by the author.

Algebraic Problems

Q6. An amoeba has a 25%, 25%, and 50% chance of producing 0, 1, or 2 offspring, respectively. Each of the amoeba's descendants also has the same probabilities. What is the probability that the amoeba's lineage dies out?

For the amoeba lineage to die it needs to produce 0 offspring. If it produces 1 offspring then the child offspring must produce 0 offspring. Similarly for two children's offspring.



Answer Q6. Image by the author.

Q7. The entries in a 2×2 matrix are integers that are independently chosen for each entry. The probability that the entry is odd is p . If the probability that the value of the determinant is even is 0.5, find p .

The probability that the determinant is odd/even can be computed by making cases for odd/even and then taking a sum of the probabilities of those cases.

Consider $M = \begin{pmatrix} a & d \\ b & c \end{pmatrix}_{2 \times 2}$, the determinant is given by:
 $|M| = ad - bc$

given: $P(E) = 0.5 \Rightarrow P(O) = 0.5$ (probability that $|M|$ is odd).
[p is the probability that the entry is odd]
for $|M|$ to be odd we have 2 cases:
① $ad \rightarrow \text{odd}, bc \rightarrow \text{even} \Rightarrow |M| \rightarrow \text{odd}$
 $\therefore p_1 = p^2(1-p^2)$
② $ad \rightarrow \text{even}, bc \rightarrow \text{odd} \Rightarrow |M| \rightarrow \text{odd}$
 $\therefore p_2 = p^2(1-p^2)$
 $\therefore 2p^2(1-p^2) = P(O) = 0.5$
 $\Rightarrow p = \frac{1}{\sqrt{2}}$

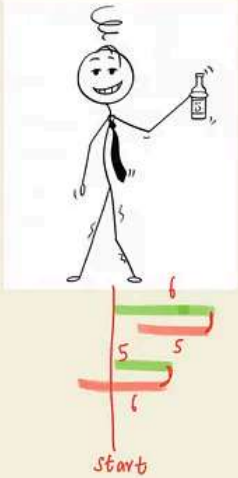
Answer Q7. Image by the author.

Binomial Distribution

Q8. A drunker takes either a step forward or backward. The probability that he takes a forward step is 0.4. Find the probability that at end of 11 steps he is 1 step away from the starting point?

Clearly, for the drunker to be 1 step away from start he can either take 5 steps forward (meaning 6 backward steps) hence ending 1 step behind the start OR he can take 6 forward (meaning 5 backward steps) hence ending 1

step front of the start. The final probability can be calculated by taking a union of the two events.



$P(F) = \text{probability of taking forward Step} = 0.4$
 $\therefore P(B) = 0.6$
 Backward step
 $\therefore P(1 \text{ step away}) = P(5 \text{ steps forward}) + P(6 \text{ steps forward})$
 $= p$
 $p = \binom{11}{5} 0.4^5 0.6^6 + \binom{11}{6} 0.4^6 0.6^5$
 $= 0.367$

Answer Q8. Image by the author.

Q9. A coin is twice as likely to land head as a tail in a series of independent tosses. Find the probability that 3rd head occurs on the 5th toss.

For the 3rd head to occur in the 5th toss, the earlier 2 heads can occur in any of the 4 tosses initially, which becomes a case of the binomial distribution.



$$P(H) = \frac{2}{3} \quad P(T) = \frac{1}{3}$$

The probability of 2 heads (2 successes) in 4 trials is given by:

$$\binom{4}{2} \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2 = p_1$$

Probability of a head in 5th trial = $\frac{2}{3} = p_2$

$$\begin{aligned} \therefore P(3^{\text{rd}} \text{ head in } 5^{\text{th}} \text{ trial}) &= p_1 \times p_2 \\ &= \binom{4}{2} \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2 \times \frac{2}{3} \\ &\approx 0.2 \end{aligned}$$

Answer Q9. Image by the author.

Law of Total Probability

Q10. A rich lady has 4 compartments in her purse. 1st compartment has 1 Rs (Rupees) and 2 Paise coins. 2nd has 2 Rs and 3 Paise coins. 3rd has 3 Rs and 4 Paise coins. 4th has 4 Rs and 6 Paise coins.

She selects a random compartment & draws a coin, what is the probability that the drawn coin is a rupee coin?



$P(C_i)$ = Probability that
the coin is drawn from
apartment i .

$P(R) = P(\text{Drawn coin is a Rupee coin})$ -

$$P(R) = P(C_1)P(R|C_1) + P(C_2)P(R|C_2) \\ + P(C_3)P(R|C_3) + P(C_4)P(R|C_4)$$

$$= \frac{1}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{2}{5} + \frac{1}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{4}{10} \\ \approx 0.47$$

Answer Q10. Image by the author.

Bayes Theorem

Q11. An HIV test is 99% accurate (both ways). Only 0.3% of the population is HIV+. What is the probability that a random person is HIV+ given that the person tests +?



$P(\text{person is HIV}+ \mid \text{person tests}+) = ?$
 applying Bayes Theorem

$$p = \frac{p(\text{tests}+ \mid \text{is HIV}+) p(\text{is HIV}+)}{p(\text{tests}+ \mid \text{is HIV}+) p(\text{is HIV}+) + p(\text{tests}+ \mid \text{is HIV}-) p(\text{is HIV}-)}$$

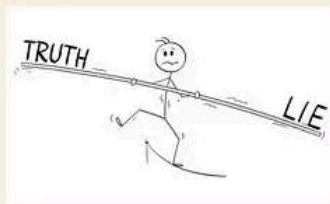
given: $p(\text{is HIV}+) = 0.3\%$

$\therefore p(\text{is HIV}-) = 99.7\%$

$$p = \frac{0.3\% \cdot 99\%}{0.3\% \cdot 99\% + 99.7\% \cdot 1\%} = 22.95\%$$

Answer Q11. Image by the author.

Q12. A speaks truth in 70% of cases, B in 50% of cases. Find the probability that they will speak the same thing while describing a certain event?



$p(E_1)$ is the probability that both A & B speak truth.

$p(A) = 0.7$ {probability that A speaks truth}

$p(B) = 0.5$

Define: E_1 : Both speak truth

E_2 : Both speak lie

E_3 : Both have same answer.

$\therefore p(E_3) = ??$

$$= p(E_1) p(E_3|E_1) + p(E_2) p(E_3|E_2)$$

$$= 0.7 \times 0.5 \times 1 + 0.3 \times 0.5 \times 1$$

$$\approx 0.5$$

Answer Q12. Image by the author.

Miscellaneous Card Problems

Q13. Cards are dealt one by one from a pack of 52 well-shuffled cards. What is the probability that exactly 'k' cards are dealt before the 1st ace appears?

We are indirectly looking for the probability that the 1st ace appears in the (k+1)th card. () is the standard combination's notation.



4 Aces

Total ways to choose 'k' cards from 52 cards -
 $= \binom{52}{k}$

Total ways to choose 'k' cards which are not Aces -
 $= \binom{48}{k}$

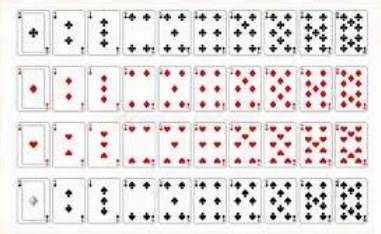
Ways to choose an Ace = $\binom{4}{1}$; This choice will be made after 'k' cards are dealt !!

$$\therefore p = \frac{\binom{48}{k} \times \binom{4}{1}}{\binom{52}{k} \binom{52-k}{1}}$$

Answer Q13. Image by the author.

Q14. All face cards are removed from a pack of 52 well-shuffled cards. From the remaining 40 cards, 4 cards are drawn randomly. What is the probability that 4 cards are from different suits and denominations?

Total suits = 4 (Spade, Hearts, Clubs, Diamonds); Total denominations = 13 (2, ..., 10, A, J, Q, K).



↑ Suits

→ denominations

Total ways to choose from 40 cards: $\binom{40}{4}$

Favourable Cases:

a) choosing 4 different denominations $\binom{10}{4}$

b) Ensuring that those 4 cards have different Suits - $\binom{4}{1}\binom{3}{1}\binom{2}{1}\binom{1}{1}$

Out of 4 cards; let's say 1st card will have the liberty to choose from any suit. $\binom{4}{1}$

\therefore Desired probability (p) = $\frac{\binom{10}{4}\binom{4}{1}\binom{3}{1}\binom{2}{1}\binom{1}{1}}{\binom{40}{4}}$

Answer Q14. Image by the author.

References

[1]. (Must Read) A summarised theory of probability concepts (includes laws and theorems): <https://www.cuemath.com/data/probability/>

[2]. Basic Probability concepts: <https://seeing-theory.brown.edu/basic-probability/index.html> (chance events, expectations, etc.)

[3]. Conditional Probability: <http://www.stat.yale.edu/Courses/1997-98/101/condprob.htm>

[4]. Law of Total probability: <https://youtu.be/7t9jyikrG7w>

[5]. Bayes theorem (3blue1brown): <https://youtu.be/HZGCoVF3YvM>

[6]. Binomial distribution (Khan Acad.): <https://youtu.be/WWv0RUxDfbs>

[7]. Permutations and Combinations: <https://youtu.be/XJnIdRXUi7A>

[8]. Some recommended YouTube channels on probability and statistics:
[@jbstatistics](#), [@TheOrganicChemistryTutor](#)

I hope it was some good practice on probability and that you enjoyed solving these questions. I would be glad to know any different approaches to solve the above problems. I would be happy to answer any doubts or discussions on any of the questions mentioned above. **Feedback** is highly appreciated. A clap 🖐️ is also a good feedback 😊. Good luck with your next data science interview. You can reach me via [Linkedin](#).

ThankYou!

Data Science

Probability

Interview

Statistics

Bayesian Statistics



Published in Towards Data Science

Following

772K Followers · Last published just now

Your home for data science and AI. The world's leading publication for data science, data analytics, data engineering, machine learning, and artificial intelligence professionals.



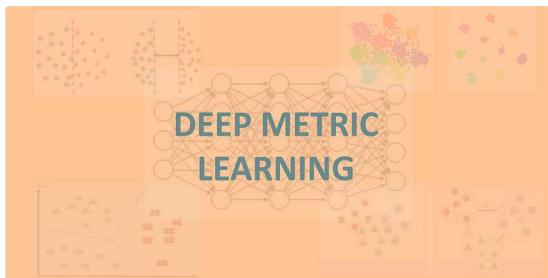
Written by Aakash Agrawal

Follow

195 Followers · 220 Following

Data Scientist @CRED | IIT Guwahati 2020

More from Aakash Agrawal and Towards Data Science



 In Towards Data Science by Aakash Agrawal

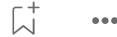


 In Towards Data Science by Cassie Kozyrkov

The Why and the How of Deep Metric Learning.

Diving deep into metric based deep learning.

Jan 11, 2021 🖱 548 💬 7



tds In Towards Data Science by Stephanie Kirmer

The Cultural Impact of AI Generated Content: Part 1

What happens when AI generated media becomes ubiquitous in our lives? How does...

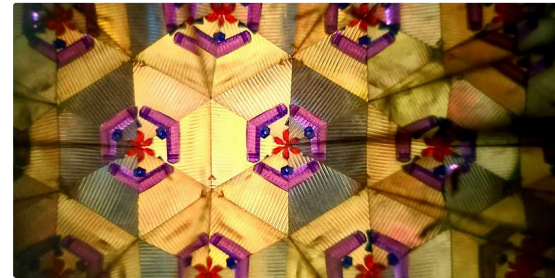
1d ago 🖱 233 💬 5



The Name That Broke ChatGPT: Who is David Mayer?

AI, privacy, human bias, prompting, the future of content, and how to hack a chatbot

★ 1d ago 🖱 452 💬 8



tds In Towards Data Science by Aakash Agrawal

Vision-Based Rep Counting in the Wild

A review of different approaches to vision-based rep counting

Feb 22, 2023 🖱 34



See all from Aakash Agrawal

See all from Towards Data Science

Recommended from Medium



FREQUENTLY ASKED
HYPOTHESIS TESTING
QUESTIONS FOR DATA
SCIENTIST INTERVIEWS



Vikash Singh

Frequently Asked Hypothesis Testing Questions for Data...

If you are preparing for a data science or statistical modelling role, brushing up on yo...



Sep 6



52



Saurav Naik

Likelihood Estimation

Here we will be understanding:

Jun 7



7



Lists



Predictive Modeling w/ Python

20 stories · 1700 saves



Coding & Development

11 stories · 927 saves



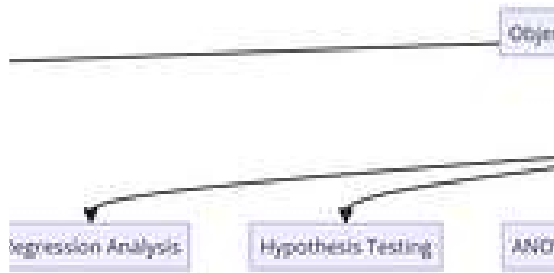
Practical Guides to Machine Learning


10 stories · 2068 saves



ChatGPT prompts

50 stories · 2317 saves



 Prem Vishnoi(cloudvala)

Probability and Statistics in Machine Learning

Introduction:

★ Aug 19 🖱 6

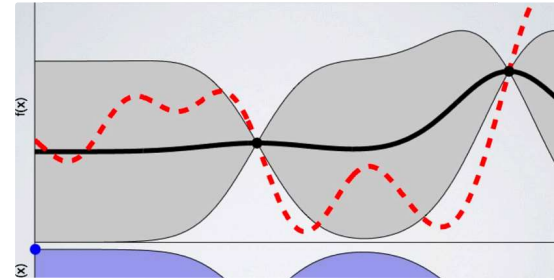


 Crystal X

Correlation analysis as a crucial part of statistical hypothesis...

For the past week or so I have been researching hypothesis testing. During this...

★ Nov 26 🖱 8

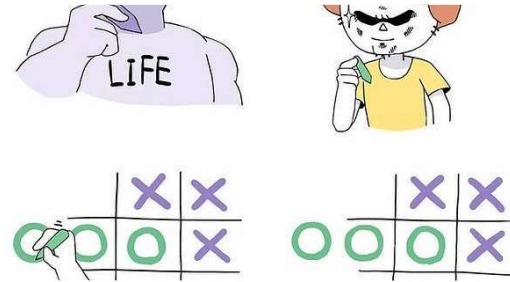


 M CHO

Bayesian Optimization

Bayesian Optimization concerns the problem of maximizing expensive black-box functions.

★ Jul 21 🖱 63



 Amit Yadav

Top 11 Linear Regression Interview Questions (With Answers)

If you think you need to spend \$2,000 on a 120-day program to become a data scientist...

★ Jul 21 🖱 5



[See more recommendations](#)

[Help](#) [Status](#) [About](#) [Careers](#) [Press](#) [Blog](#) [Privacy](#) [Terms](#) [Text to speech](#) [Teams](#)