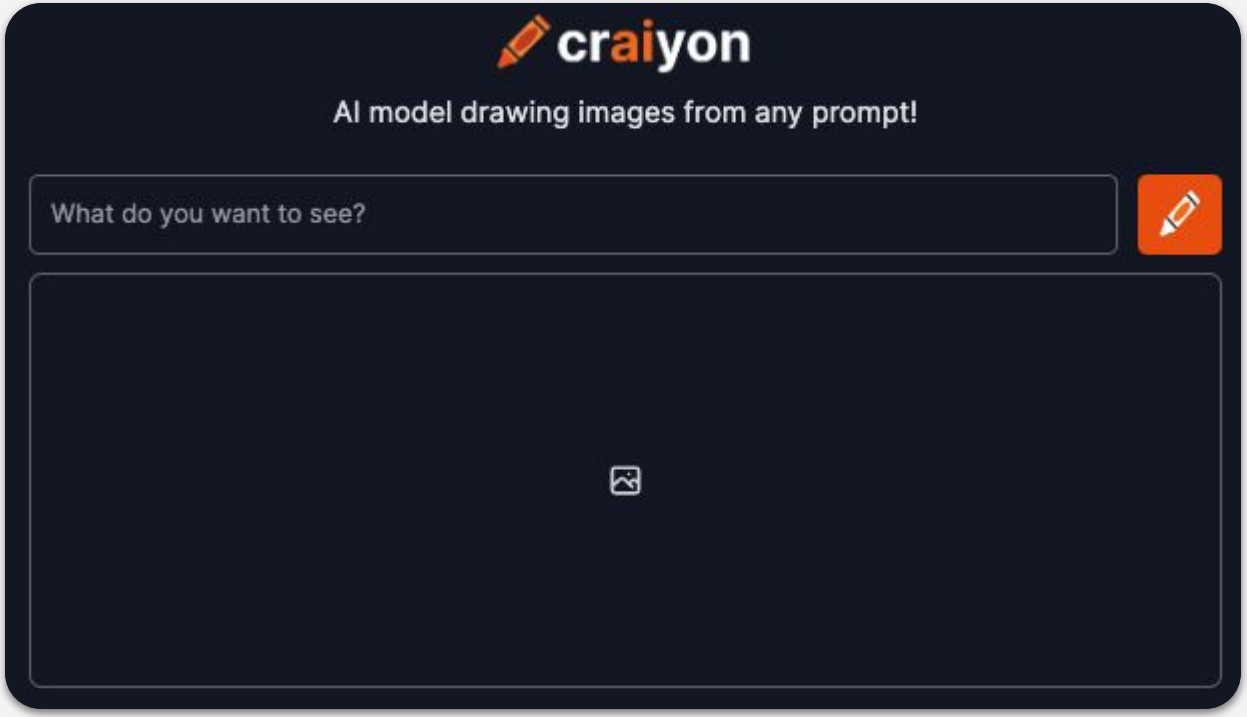
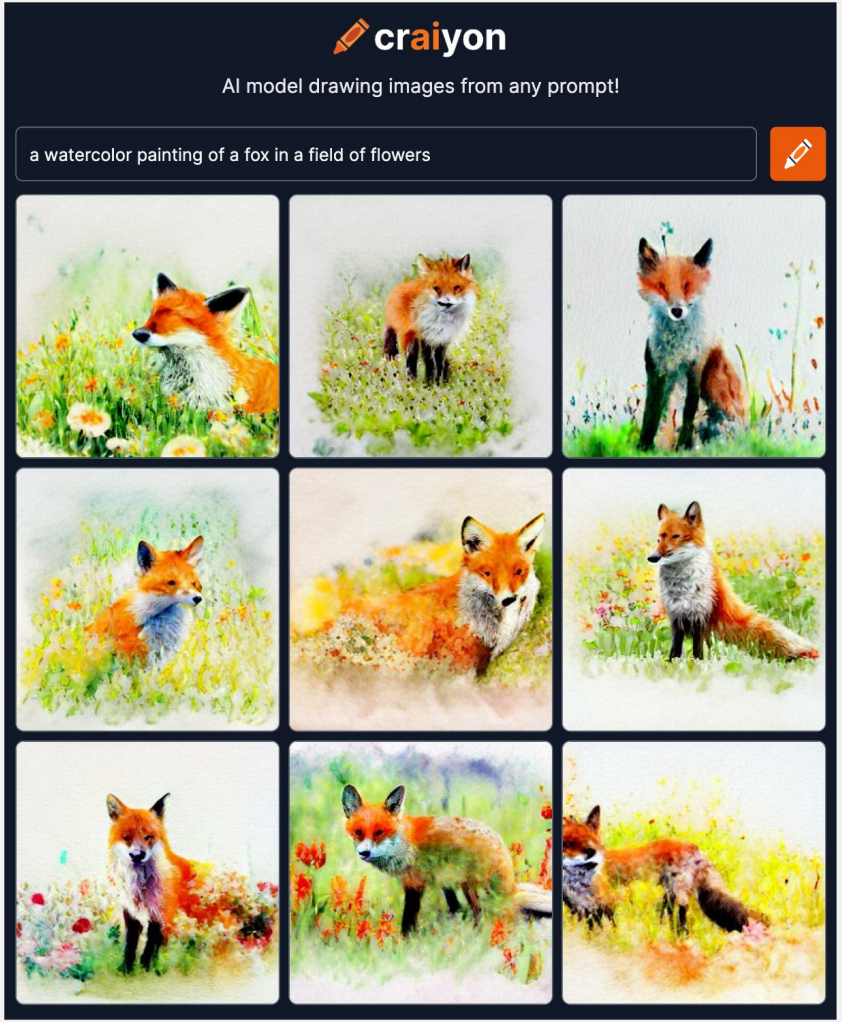
Ryan Harrington
Tech Impact, Data Innovation Lab
July 18, 2022

TECHIMPACT

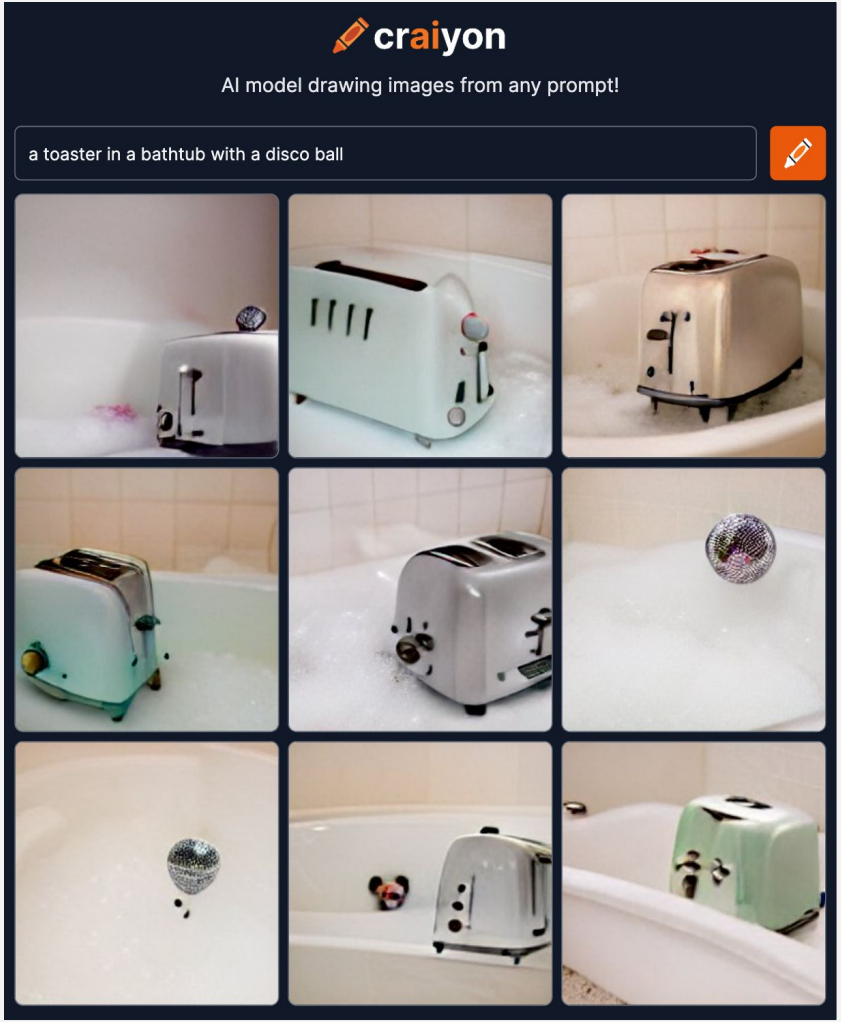
craiyon.com


craiyon
AI model drawing images from any prompt!
What do you want to see?

A watercolor painting of a fox in a field of flowers

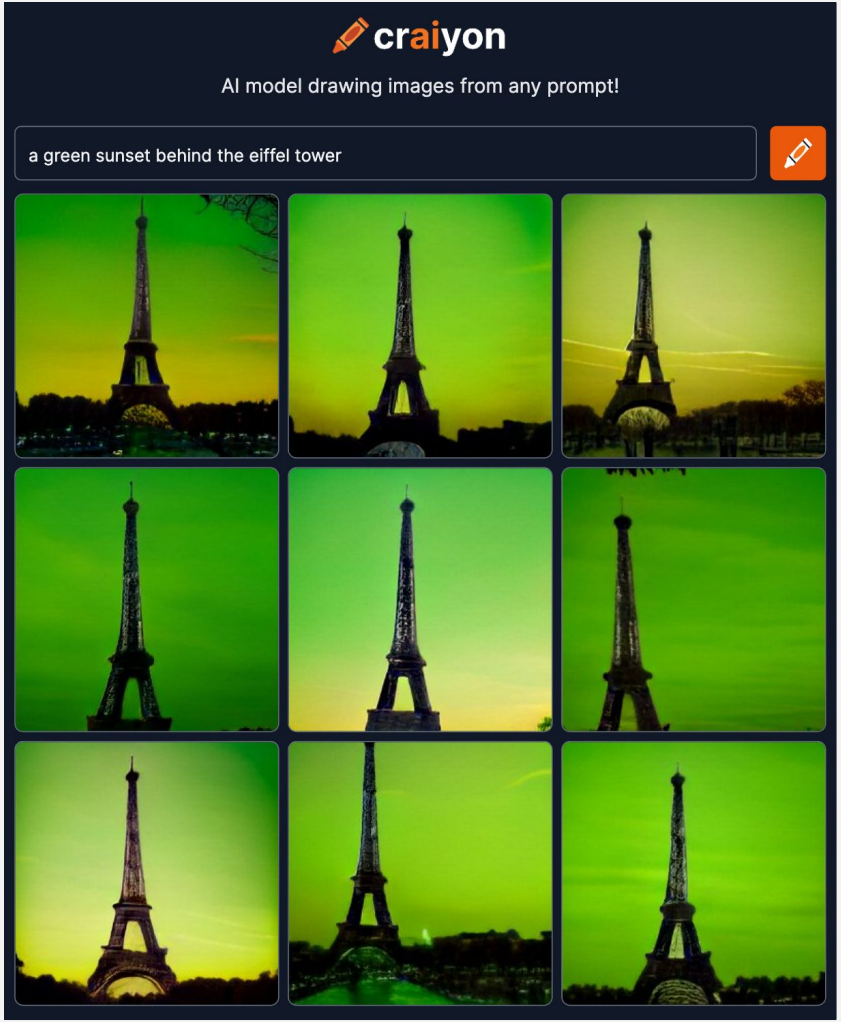


A toaster in a bathtub with a disco ball

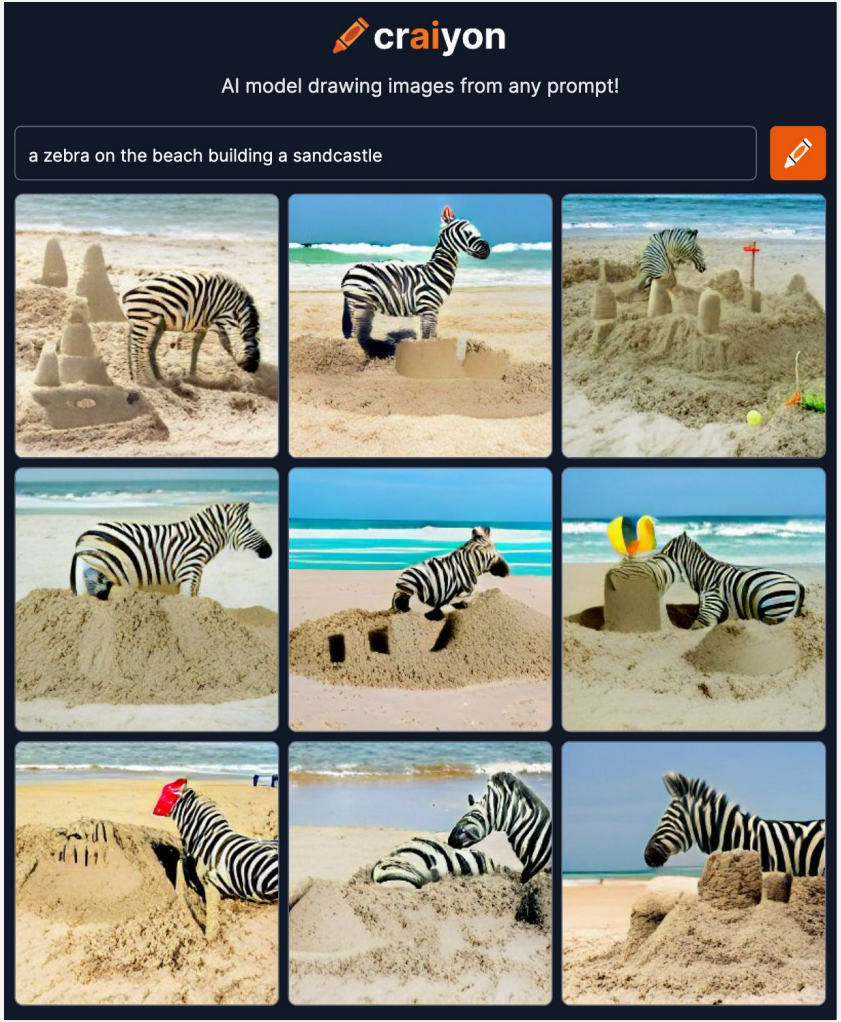


Albert Einstein driving a pickup truck

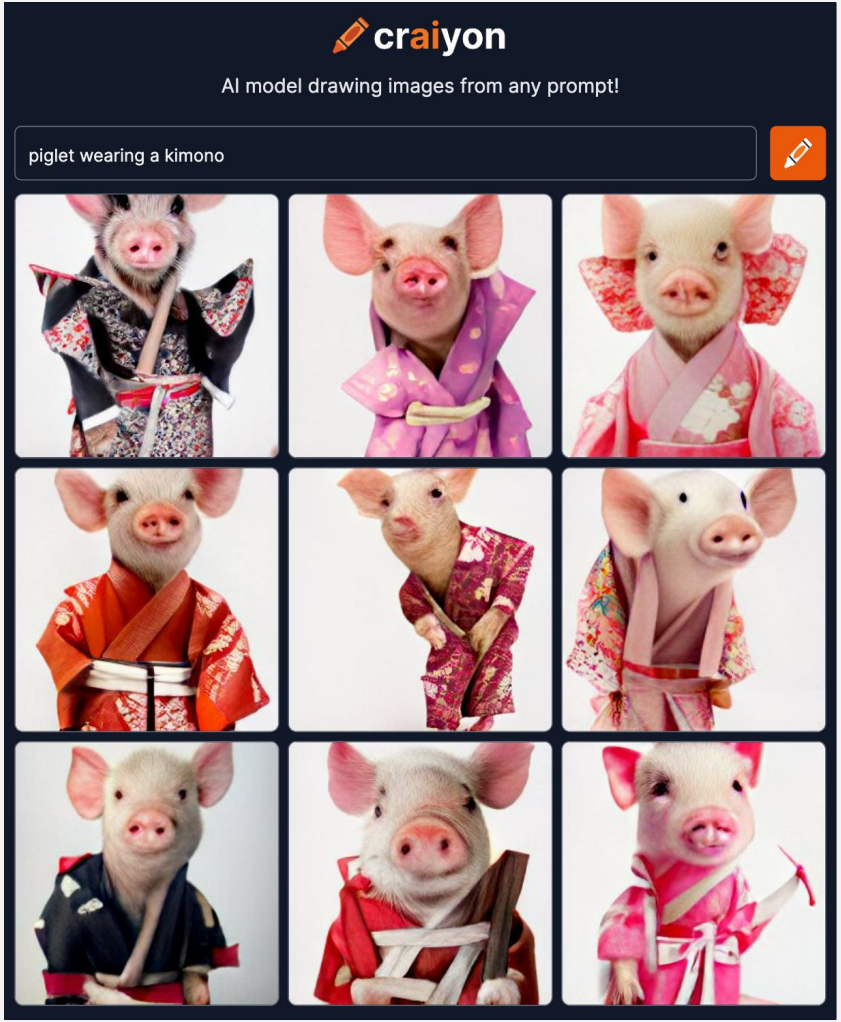
A green sunset behind the Eiffel Tower



A zebra on the beach building a sandcastle



Piglet wearing a kimono



Spend 5 minutes generating interesting captions

bit.ly/craiyon-uploads

NOTE: Case Matters in the URL!

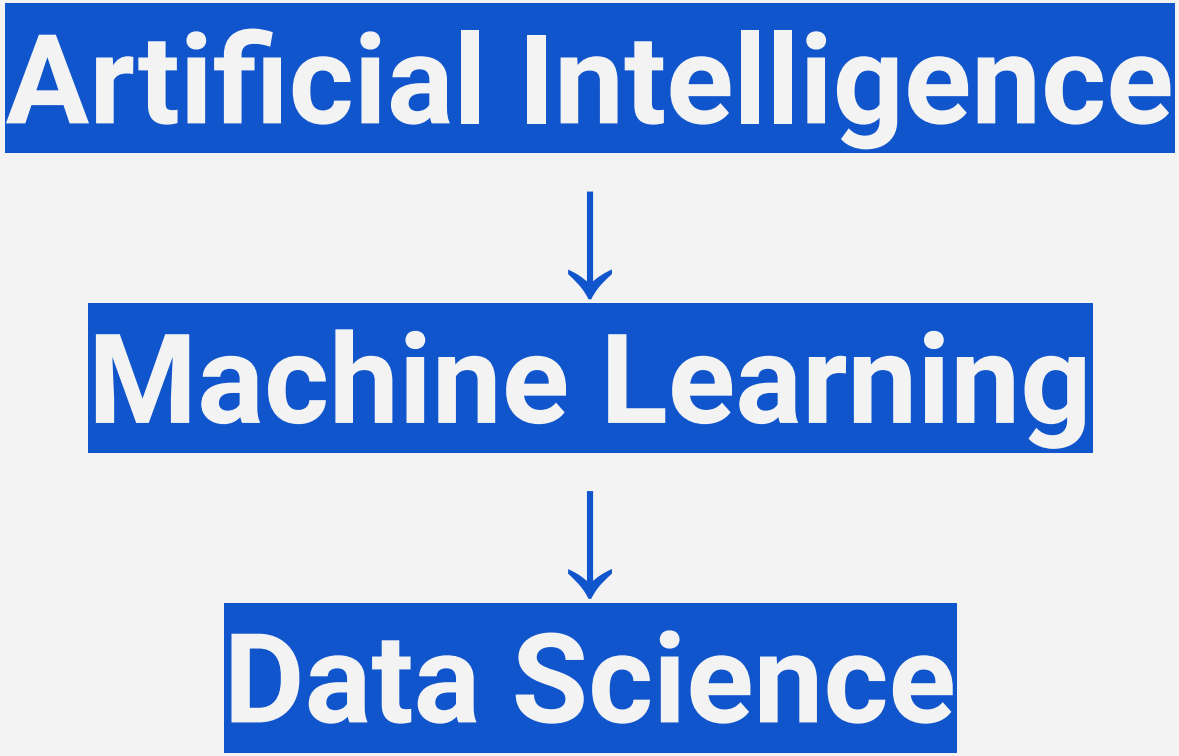
craiyon

DALL·E 2

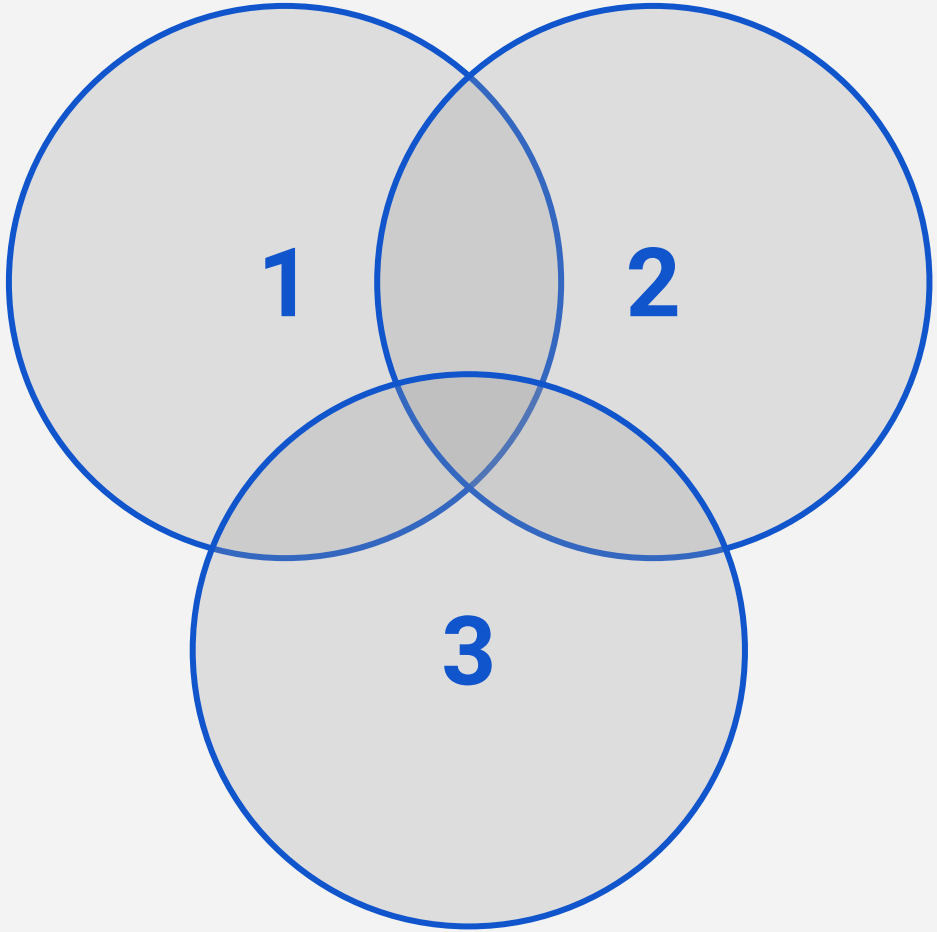
Dali + WALL-E = DALL-E


A photo of a sloth in a scuba diving mask exploring the wonders under the sea, digital art
Open AI
(dalle2)
Open Source
(dalle-mini)
(dalle-mini)

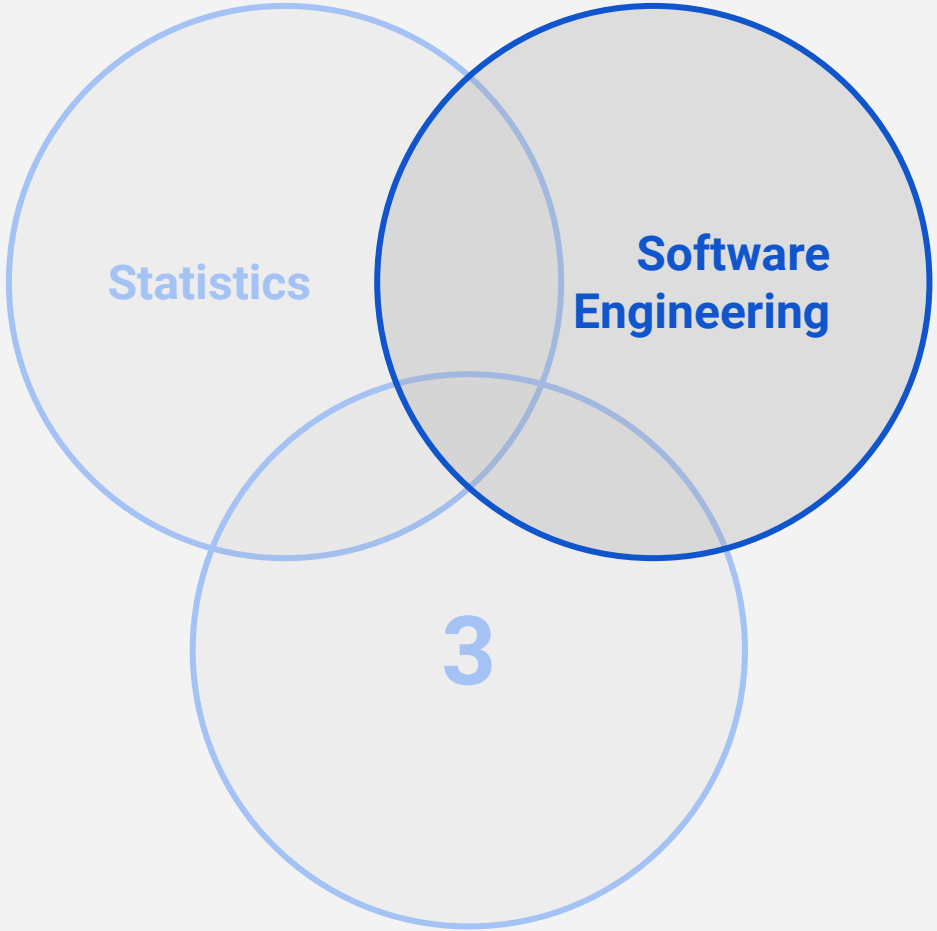
How does this work?

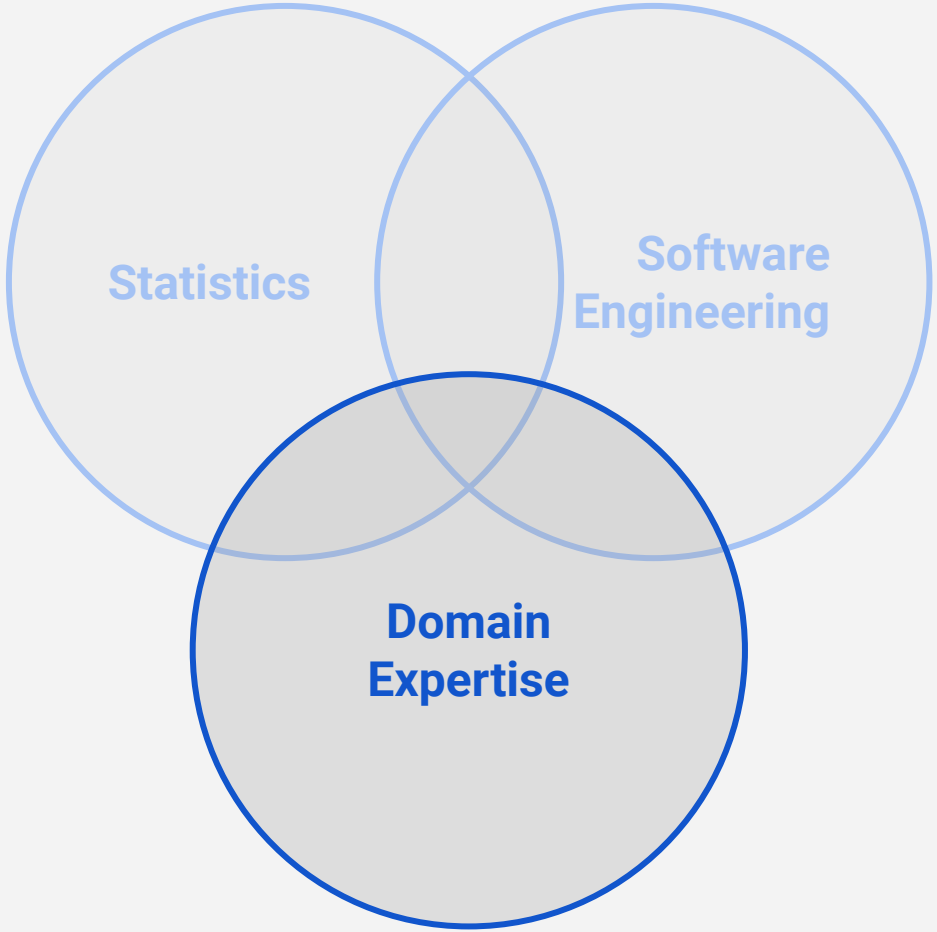

Artificial Intelligence
Machine Learning
Data Science

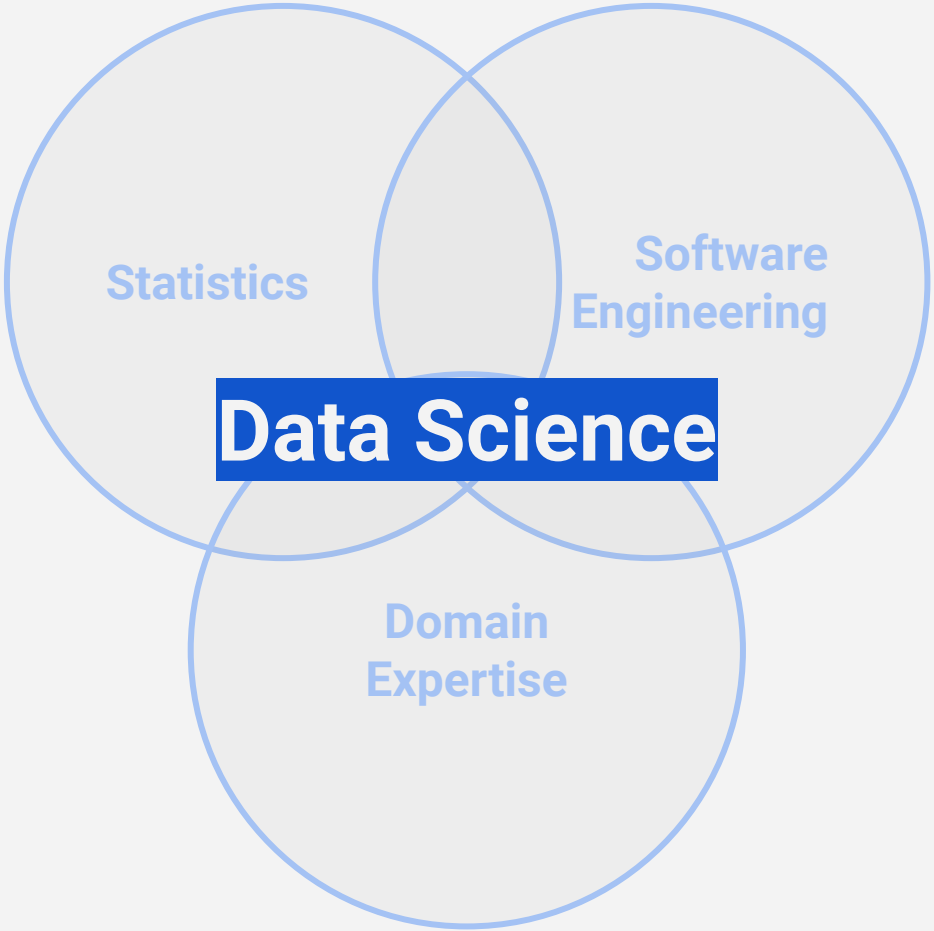
“Computation with Context”


1
2
3


Statistics
2
3


Statistics
Software Engineering
3


Statistics
Software Engineering
Domain Expertise


Statistics
Software Engineering
Data Science
Domain Expertise

How does this work?

Training / Testing

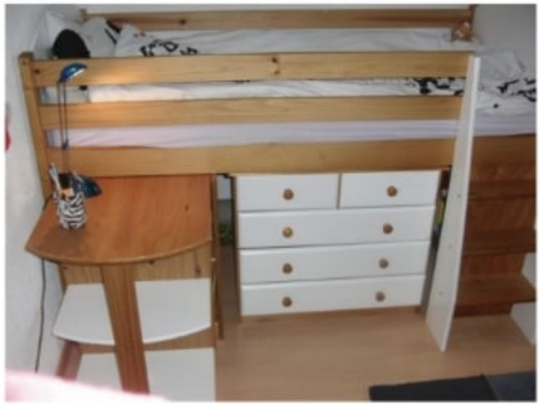
Deployment

Training

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.

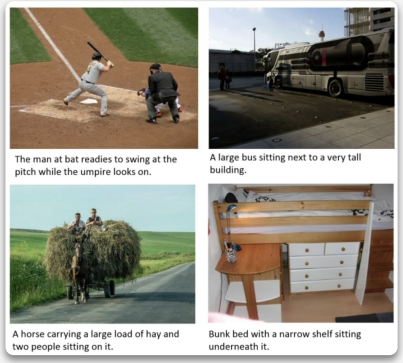
A horse carrying a large load of hay and two people sitting on it.



Bunk bed with a narrow shelf sitting underneath it.

Source

Training



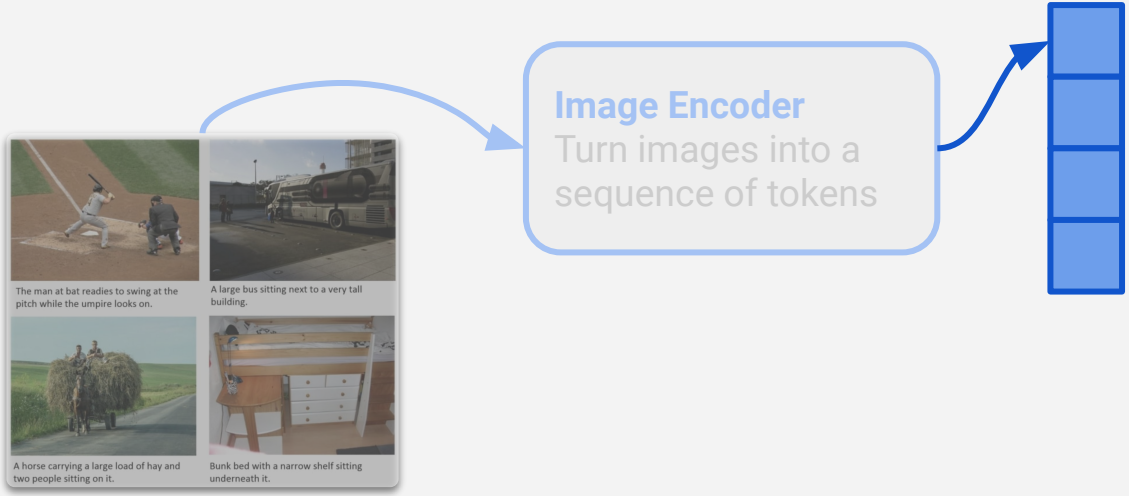
Source

Training

Image Encoder
Turn images into a sequence of tokens

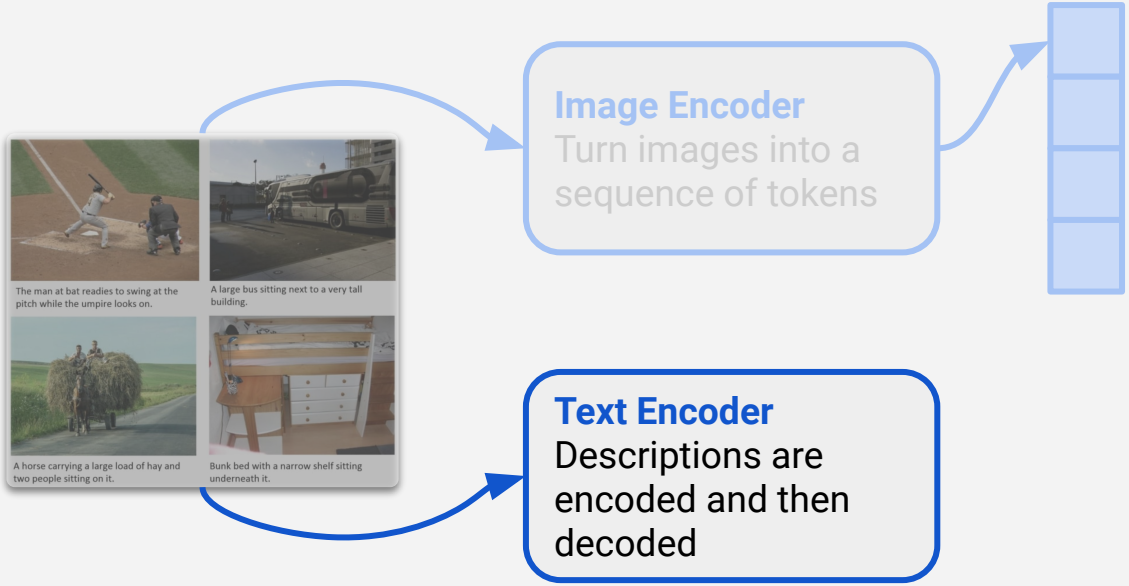
Source

Training



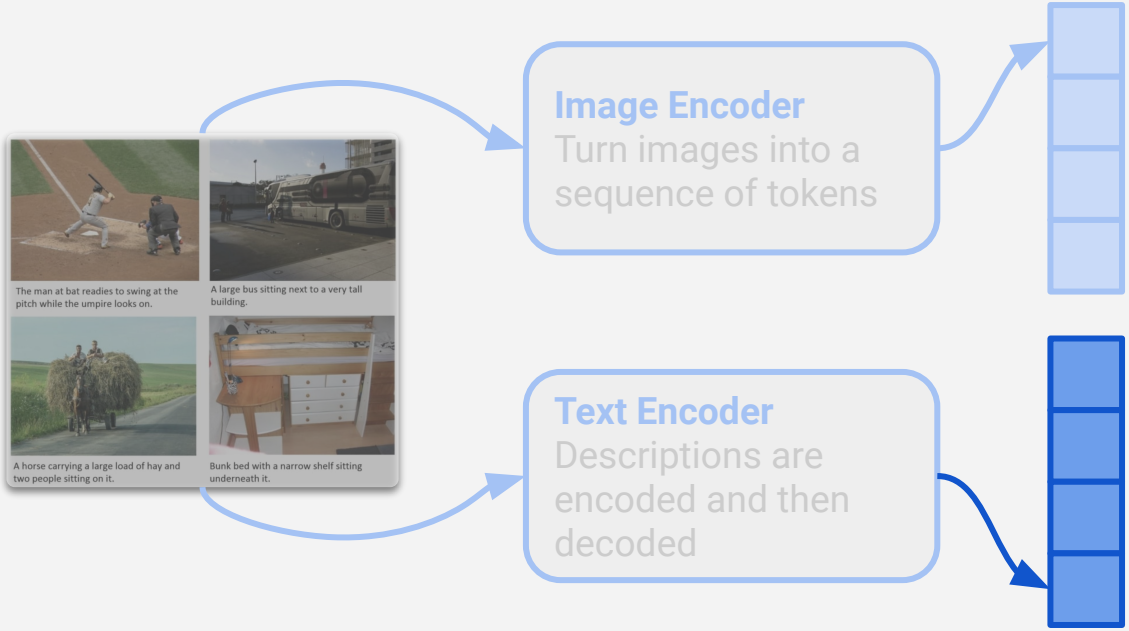
Source

Training



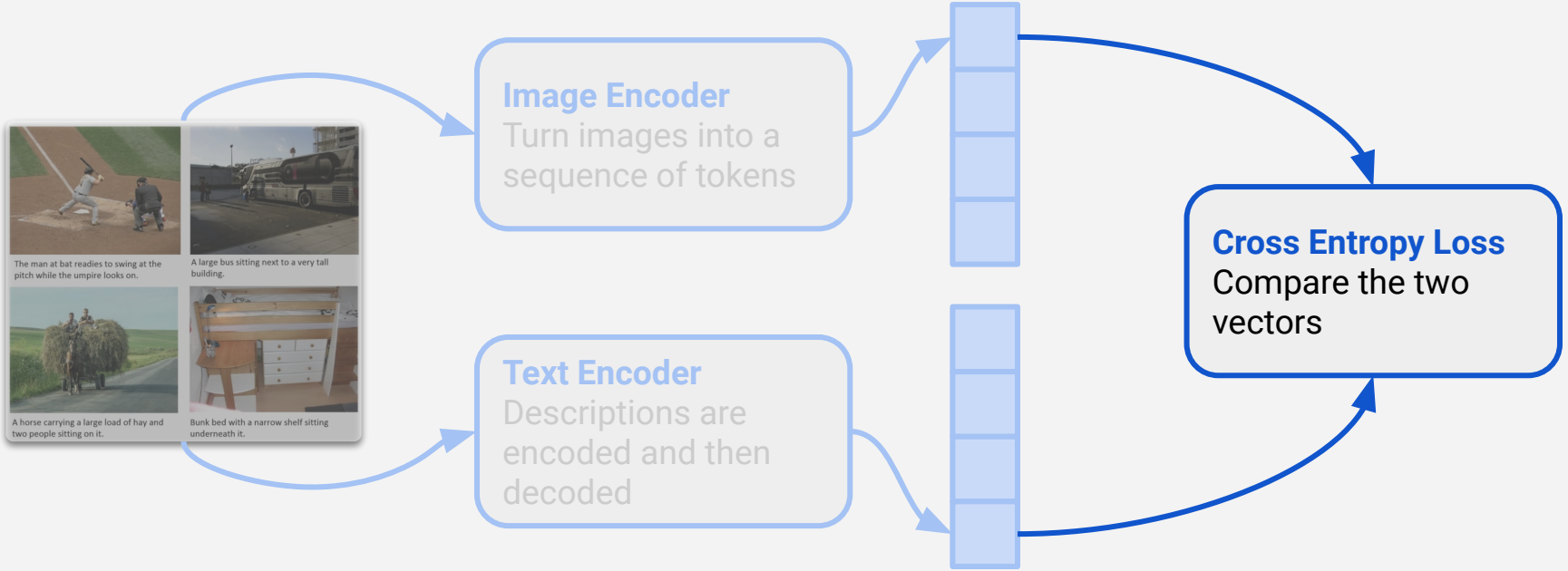
Source

Training



Source

Training



Source

Deployment

```
White snow covered mountain under
blue sky during day time
```

Source

Deployment

```
White snow covered
mountain under blue
sky during day time
```

Source

Deployment

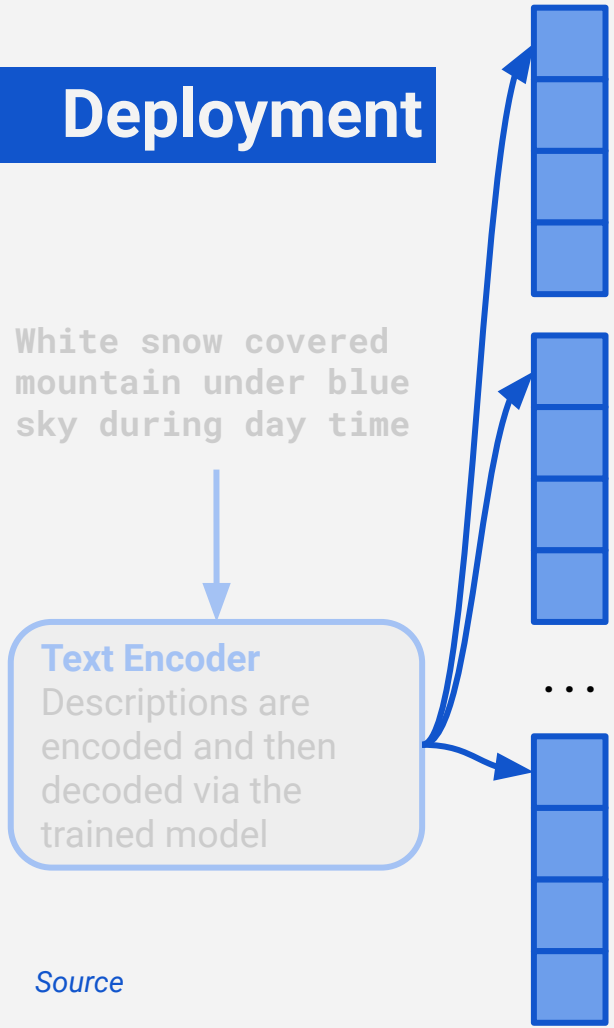
White snow covered mountain under blue sky during day time

Text Encoder
Descriptions are encoded and then decoded via the trained model

Source

Deployment

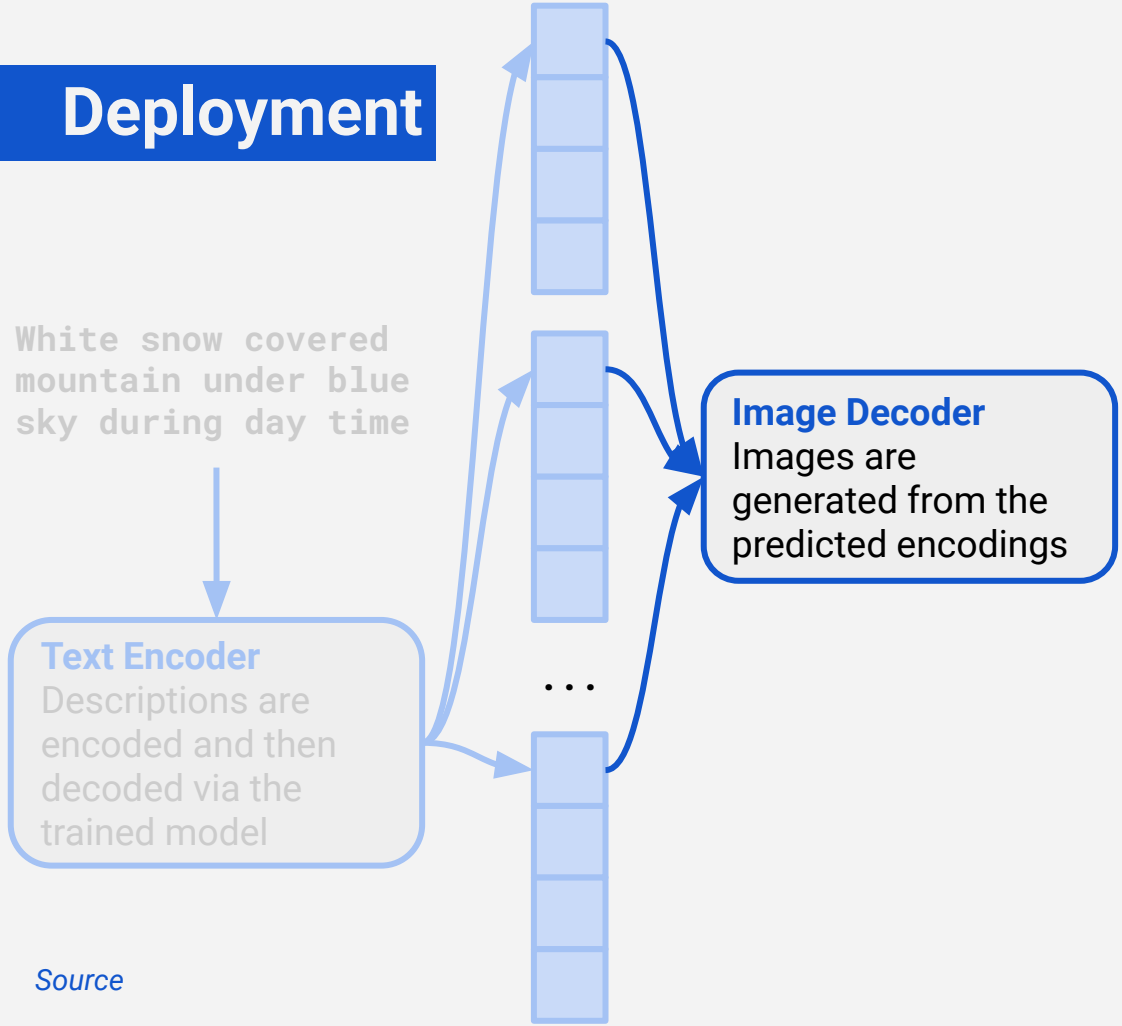
White snow covered mountain under blue sky during day time



Source

Deployment

White snow covered mountain under blue sky during day time



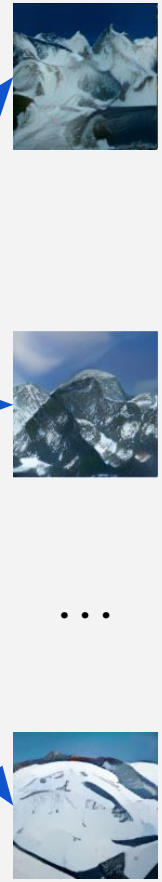
Source

Deployment

White snow covered mountain under blue sky during day time

Text Encoder
Descriptions are encoded and then decoded via the trained model

Image Decoder
Images are generated from the predicted encodings



Source

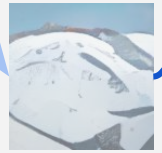
Deployment

White snow covered mountain under blue sky during day time

Text Encoder
Descriptions are encoded and then decoded via the trained model

Image Decoder
Images are generated from the predicted encodings





CLIP

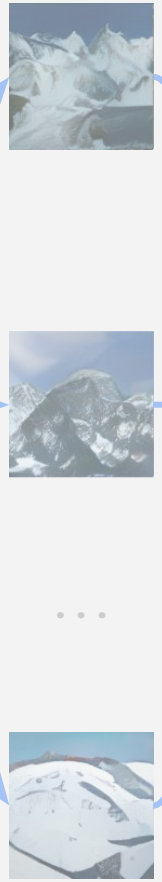
Source

Deployment

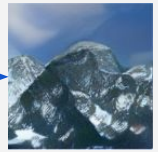
White snow covered mountain under blue sky during day time

Text Encoder
Descriptions are encoded and then decoded via the trained model

Image Decoder
Images are generated from the predicted encodings



CLIP



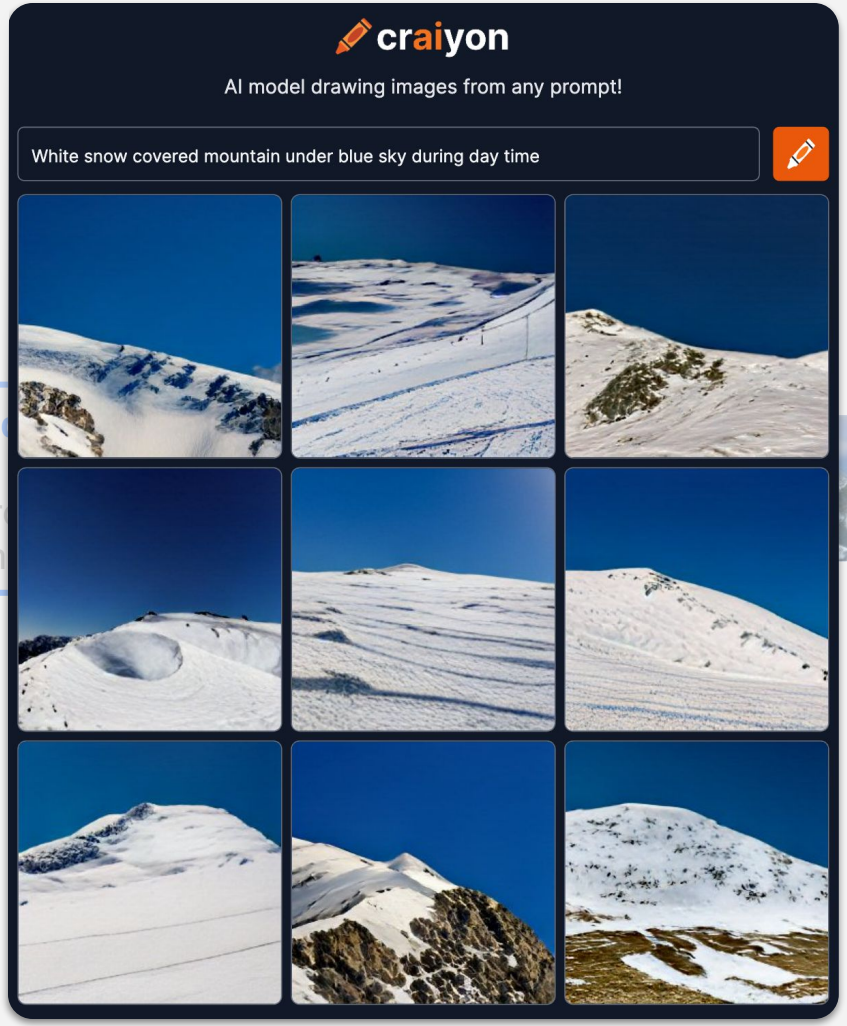
Source

White snow covered mountain under blue sky during day time

White snow covered mountain under blue sky during day time

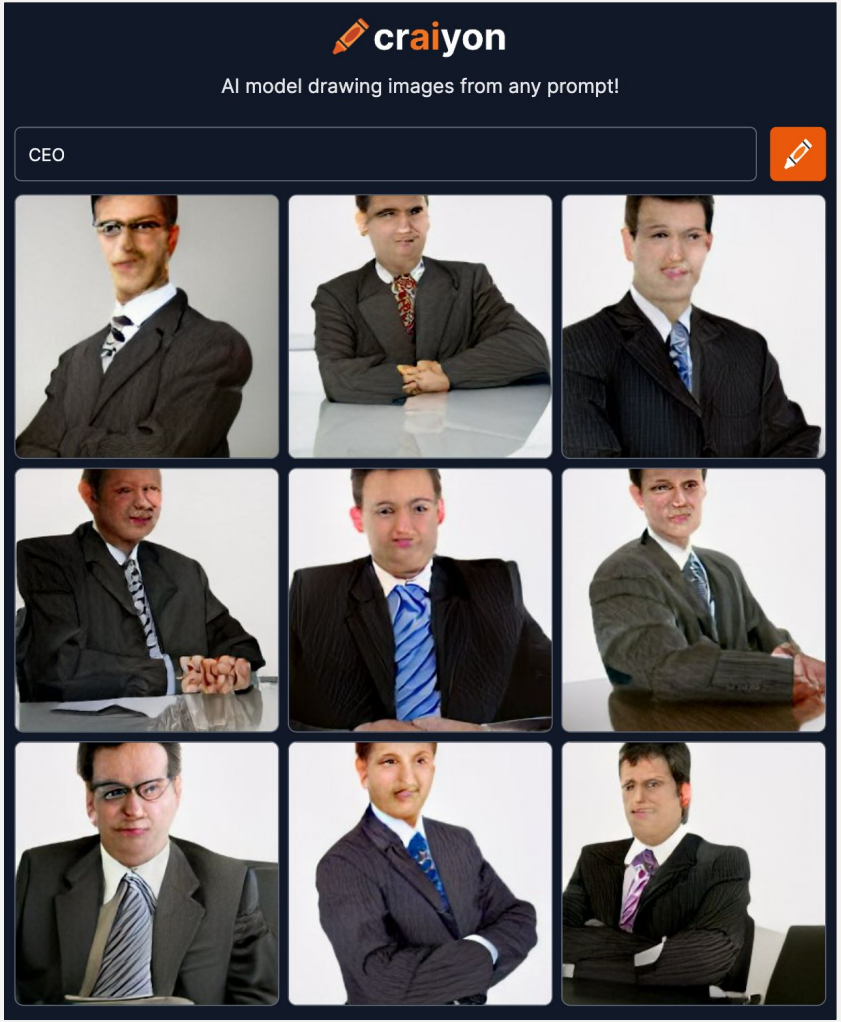
Text Encoder
Descriptions are encoded and then decoded via the trained model

Image Deco
Images are generated fr
predicted en

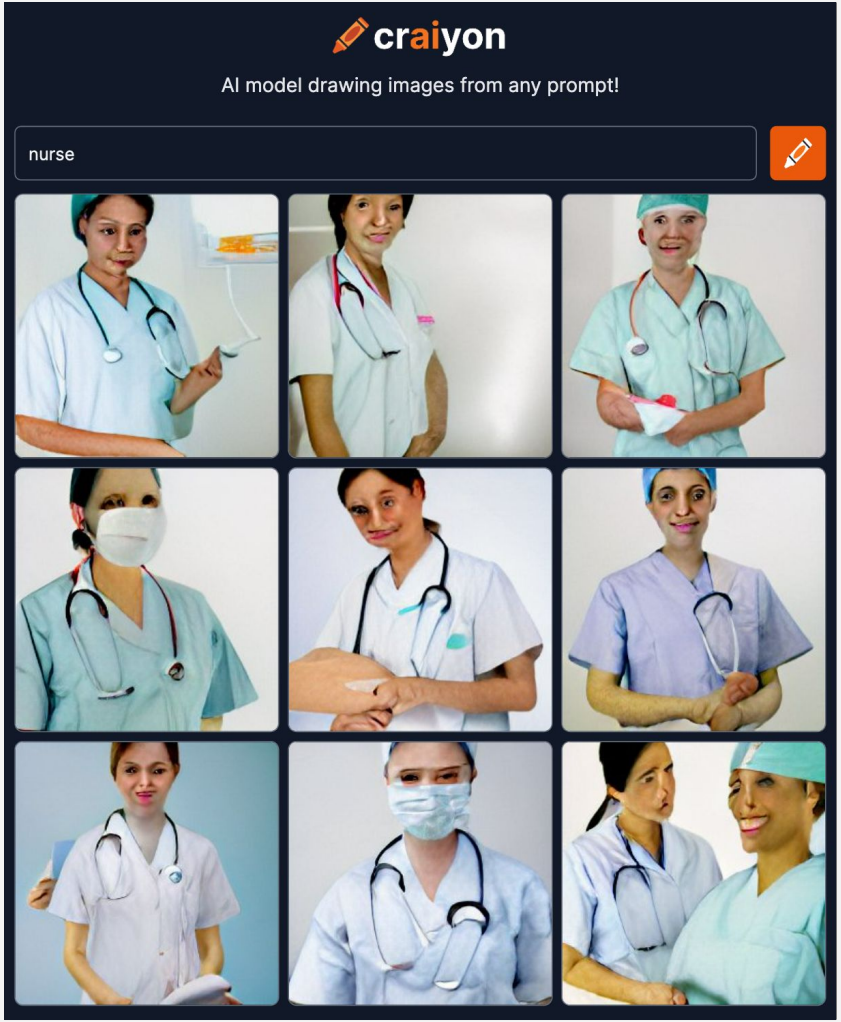


Source

CEO



Nurse



Introduction of Biases

- Occupations demonstrating higher levels of education (such as engineers, doctors or scientists) or high physical labor (such as in the construction industry) are **mostly represented by white men**. In contrast, nurses, secretaries or assistants are typically women, often white as well.

- **Most of the people generated are white.** It's only on specific examples such as athletes that we will see different races, though most of them still under-represented.

- **The dataset is limited to pictures with English descriptions**, preventing text and images from non-English speaking cultures to be represented.

Source

Scenarios

Hiring Practices

Recidivism

Child Protection

Scenario #1: Hiring Practices

A company is overwhelmed by the number of applicants that it is receiving for the engineering roles that it has posted.

In an effort to make it easier to hire candidates, a team of data scientists has built a machine learning model that will determine whether or not a candidate should be hired based upon their resume.

Some personally identifiable information, such as the name of the candidate were not considered as a feature of the model.

Scenario #2: Recidivism

Judges make decisions about the sentences that criminals receive. One of the factors that judges consider during sentencing is the likelihood of the person to re-offend (recidivism).

Courtrooms have adopted tools designed to eliminate bias in sentencing through the use of artificial intelligence. The history of the criminal can be input into the model. It will then output the likelihood of the person to re-offend.

Demographic information about the criminal is not included in the model.

Scenario #3: Child Protection

Child welfare workers are asked to make thousands of decisions related to their work in any given year. This is an overwhelming number of decisions to make - often with imperfect information.

A county's Department of Health and Human Services has built an algorithm that can aid decision making for its child welfare workers. Each case is given a score that indicates how risky it is based upon the likelihood of the child to be removed from the home within 2 years.

Incidents of potential neglect are reported to the county's child protection hotline. The reports go through a screening process where the algorithm calculates the child's potential risk and assigns a score. Child welfare workers then use their discretion to decide whether to investigate.

rharrington31.github.io/ai-ethics-scenarios/

Questions to Consider (2 minutes each)

1. What do you consider to be the potential benefits of the artificial intelligence program described in the scenario?

Questions to Consider (2 minutes each)

1. What do you consider to be the potential benefits of the artificial intelligence program described in the scenario?

2. What do you consider to be the potential drawbacks of the artificial intelligence program described in the scenario?

Questions to Consider (2 minutes each)

1. What do you consider to be the potential benefits of the artificial intelligence program described in the scenario?

2. What do you consider to be the potential drawbacks of the artificial intelligence program described in the scenario?

3. If you were in a position where you could decide if the organization would use the program or not, then what would you decide to do? Why?

Questions to Consider

1. What do you consider to be the potential benefits of the artificial intelligence program described in the scenario?

2. What do you consider to be the potential drawbacks of the artificial intelligence program described in the scenario?

3. If you were in a position where you could decide if the organization would use the program or not, then what would you decide to do? Why?

Scenario #1: Hiring Practices

A company is overwhelmed by the number of applicants that it is receiving for the engineering roles that it has posted.

In an effort to make it easier to hire candidates, a team of data scientists has built a machine learning model that will determine whether or not a candidate should be hired based upon their resume.

Some personally identifiable information, such as the name of the candidate were not considered as a feature of the model.

Scenario #1: Hiring Practices

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Source

Scenario #1: Hiring Practices

That is because Amazon’s computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. **Most came from men, a reflection of male dominance across the tech industry.**

Source

Scenario #1: Hiring Practices

In effect, Amazon's system taught itself that male candidates were preferable. It **penalized resumes that included the word "women's," as in "women's chess club captain."** And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

Source

Scenario #1: Hiring Practices

Amazon edited the programs to make them neutral to these particular terms. But that was **no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory**, the people said.

Source

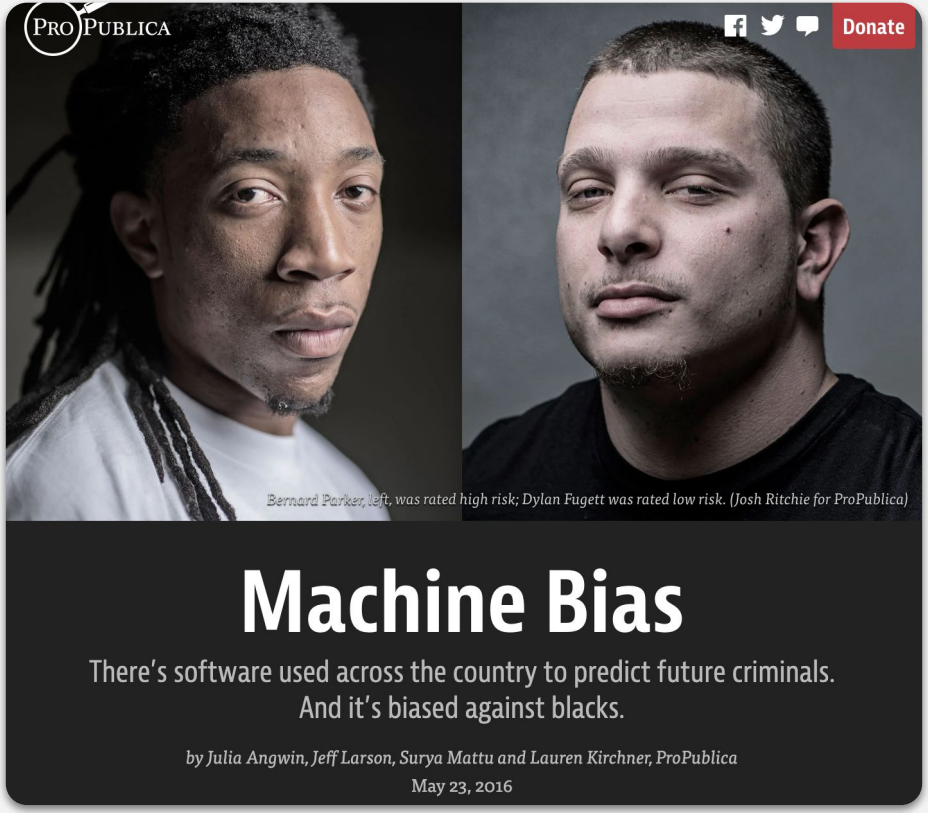
Scenario #2: Recidivism

Judges make decisions about the sentences that criminals receive. One of the factors that judges consider during sentencing is the likelihood of the person to re-offend (recidivism).

Courtrooms have adopted tools designed to eliminate bias in sentencing through the use of artificial intelligence. The history of the criminal can be input into the model. It will then output the likelihood of the person to re-offend.

Demographic information about the criminal is not included in the model.

Scenario #2: Recidivism



Source

Scenario #2: Recidivism

In 2014, then U.S. Attorney General Eric Holder warned that the risk scores might be injecting bias into the courts. He called for the U.S. Sentencing Commission to study their use.

“Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualized and equal justice,” he said, adding, “they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.”

Source

Scenario #2: Recidivism

The score proved remarkably unreliable in forecasting violent crime: **Only 20 percent of the people predicted to commit violent crimes actually went on to do so.**

When a full range of crimes were taken into account — including misdemeanors such as driving with an expired license — the algorithm was somewhat more accurate than a coin flip. Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

Source

Scenario #2: Recidivism

We also turned up significant racial disparities, just as Holder feared. In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.

- The formula was particularly likely to **falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.**
- White defendants were mislabeled as low risk more often than black defendants.

Source

Scenario #2: Recidivism

Scores like this — known as risk assessments — are increasingly common in courtrooms across the nation. They are used to inform decisions about who can be set free at every stage of the criminal justice system, from assigning bond amounts — as is the case in Fort Lauderdale — to even more fundamental decisions about defendants' freedom. **In Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin, the results of such assessments are given to judges during criminal sentencing.**

Source

Scenario #3: Child Protection

Child welfare workers are asked to make thousands of decisions related to their work in any given year. This is an overwhelming number of decisions to make - often with imperfect information.

A county's Department of Health and Human Services has built an algorithm that can aid decision making for its child welfare workers. Each case is given a score that indicates how risky it is based upon the likelihood of the child to be removed from the home within 2 years.

Incidents of potential neglect are reported to the county's child protection hotline. The reports go through a screening process where the algorithm calculates the child's potential risk and assigns a score. Child welfare workers then use their discretion to decide whether to investigate.

Scenario #3: Child Protection

An algorithm that screens for child neglect raises concerns

By SALLY HO and GARANCE BURKE
April 29, 2022

Inside a cavernous stone fortress in downtown Pittsburgh, attorney Robin Frank defends parents at one of their lowest points – when they risk losing their children.

The job is never easy, but in the past she knew what she was up against when squaring off against child protective services in family court. Now, she worries she's fighting something she can't see: an opaque algorithm whose statistical calculations help social workers decide which families should be investigated in the first place.

Source

Scenario #3: Child Protection

According to new research from a Carnegie Mellon University team obtained exclusively by AP, Allegheny's algorithm in its first years of operation showed a **pattern of flagging a disproportionate number of Black children for a "mandatory" neglect investigation, when compared with white children.** The independent researchers, who received data from the county, also found that social workers disagreed with the risk scores the algorithm produced about one-third of the time.

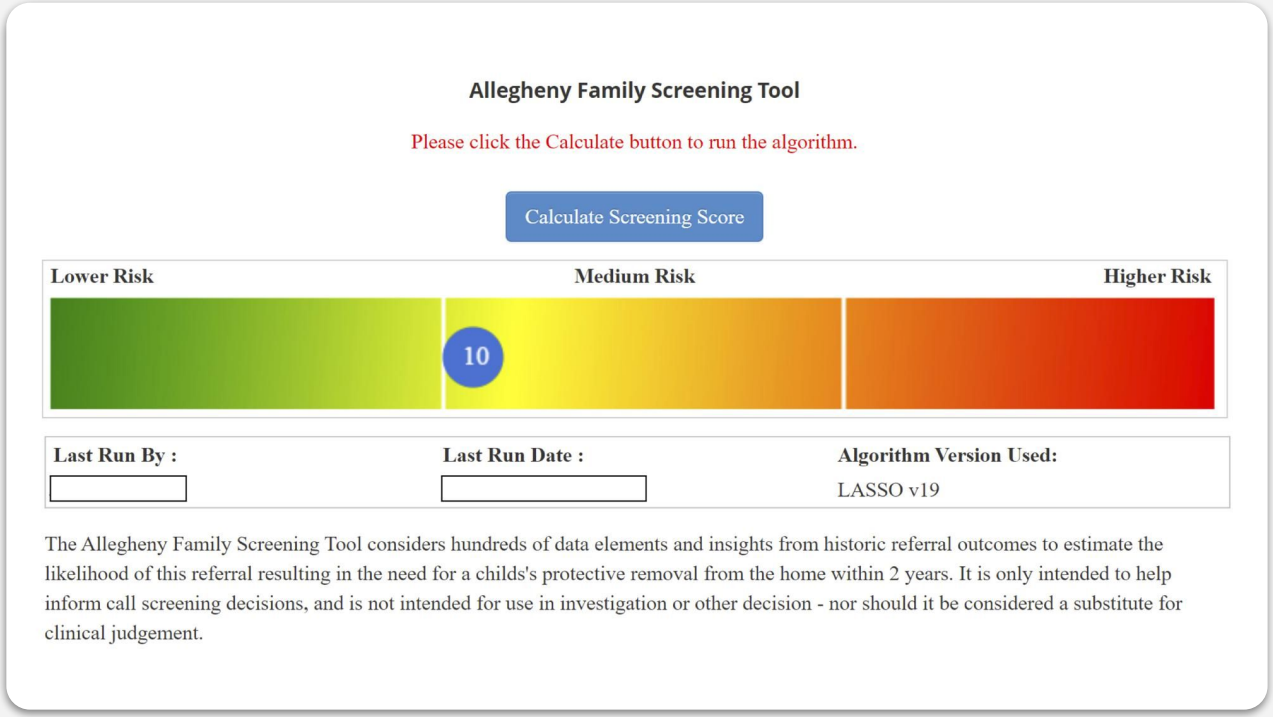
Source

Scenario #3: Child Protection

If the tool had acted on its own to screen in a comparable rate of calls, **it would have recommended that two-thirds of Black children be investigated, compared with about half of all other children reported**, according to another study published last month and co-authored by a researcher who audited the county's algorithm.

Source

Scenario #3: Child Protection



Source

Scenario #3: Child Protection

Given the high stakes – skipping a report of neglect could end with a child's death but scrutinizing a family's life could set them up for separation – **the county and developers have suggested their tool can help "course correct" and make the agency's work more thorough and efficient by weeding out meritless reports** so that social workers can focus on children who truly need protection.

The developers have described using such tools as a moral imperative, saying child welfare officials should use whatever they have at their disposal to make sure children aren't neglected.

Source

So what now?

Ryan Harrington
ryanh@techimpact.org