Co-powered by Bugasura.io

When to use AI?
When not to use AI?

Aalok Thakkar (Ashoka University) and Manoj Kumar (Moolya)

UK creating ‘murder prediction’ tool to identify people most likely to kill

Exclusive: Algorithms allegedly being used to study data of thousands of people, in project critics say is ‘chilling and dystopian’

If not predictive policing, then what?

If not predictive policing, then what?

Reasoning?

Reasoning?

Computer Chess will surpass human chess abilities within ten years.

Herbert Simon (1957)


KASPAROV vs
DEEP BLUE
the rematch
1:57
1:58
AUDEMARS PIGUET


8
7
6
5
4
3
2
1
a
b
c
d
e
f
g
h

Coherence, not Correctness

If IBM Deep Blue cannot be expected to write a work email, ChatGPT cannot be expected to play chess.

Coherence, not Correctness

If IBM Deep Blue cannot be expected to write a work email, ChatGPT cannot be expected to play chess.

And yet we use it for chess...

But what about a chatbot?

But what about a chatbot?

But what about a chatbot?

Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

23 February 2024

Share Save

Maria Yagoda
Features correspondent

When Air Canada's chatbot gave incorrect information to a traveller, the airline argued its chatbot is "responsible for its own actions".

DPD error caused chatbot to swear at customer

Bigger AI chatbots more inclined to spew nonsense — and people don't always realize

Artificial-intelligence models are improving overall but are more likely to answer every question, leading to wrong answers.

AI Gone Wild: Cursor's Rogue Bot 'Hallucinates' New User Policy

News

NYC's AI Chatbot Tells Businesses to Break the Law

The Microsoft-powered bot says bosses can take workers' tips and that landlords can discriminate based on source of income

But what about writing and summarization?

But what about writing and summarization?

AI chatbots unable to accurately summarise news, BBC finds

11 February 2025

Share Save

Imran Rahman-Jones
Technology reporter

It found 51% of all AI answers to questions about the news were judged to have significant issues of some form.

Additionally, 19% of AI answers which cited BBC content introduced factual errors, such as incorrect factual statements, numbers and dates.

But what about writing and summarization?

Some examples of inaccuracies found by the BBC included:

- Gemini incorrectly said the NHS did not recommend vaping as an aid to quit smoking
- ChatGPT and Copilot said Rishi Sunak and Nicola Sturgeon were still in office even after they had left
- Perplexity misquoted BBC News in a story about the Middle East, saying Iran initially showed "restraint" and described Israel's actions as "aggressive"

Beyond LLMs?

Tesla Autopilot feature was involved in 13 fatal crashes, US regulator says

'Alexa, how should I vote?': rightwing uproar over voice assistant's pro-Kamala Harris points

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

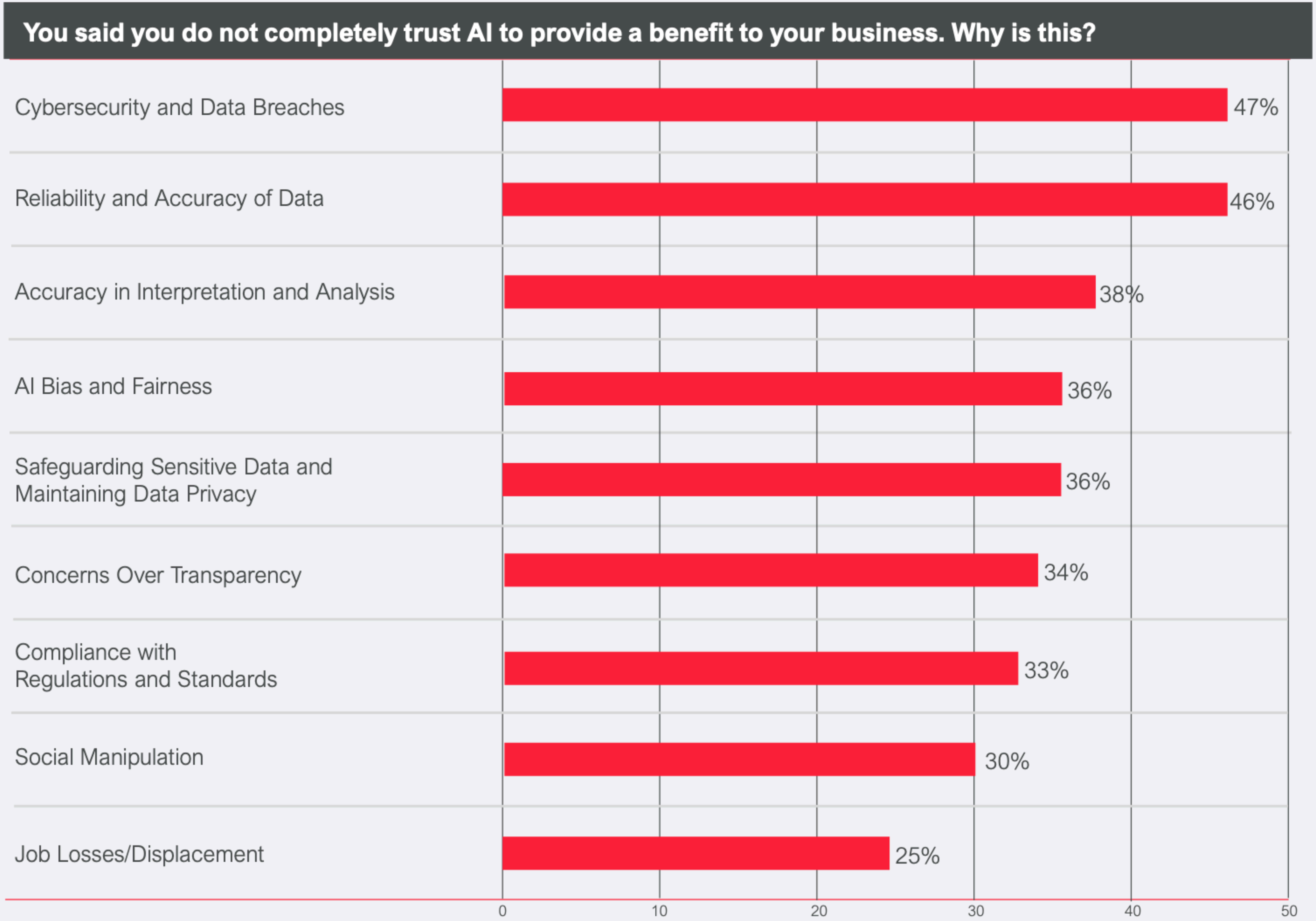
When to use AI?

When not to use AI?

Customer expectation driving AI adoption: 55% acknowledged that customer expectation is a key driver for AI adoption

FOMO a key driver for AI uptake: 63% of global IT leaders worried their company will fall behind if they don’t adopt AI

State of Intelligent Automation Report 2024 by ABBYY

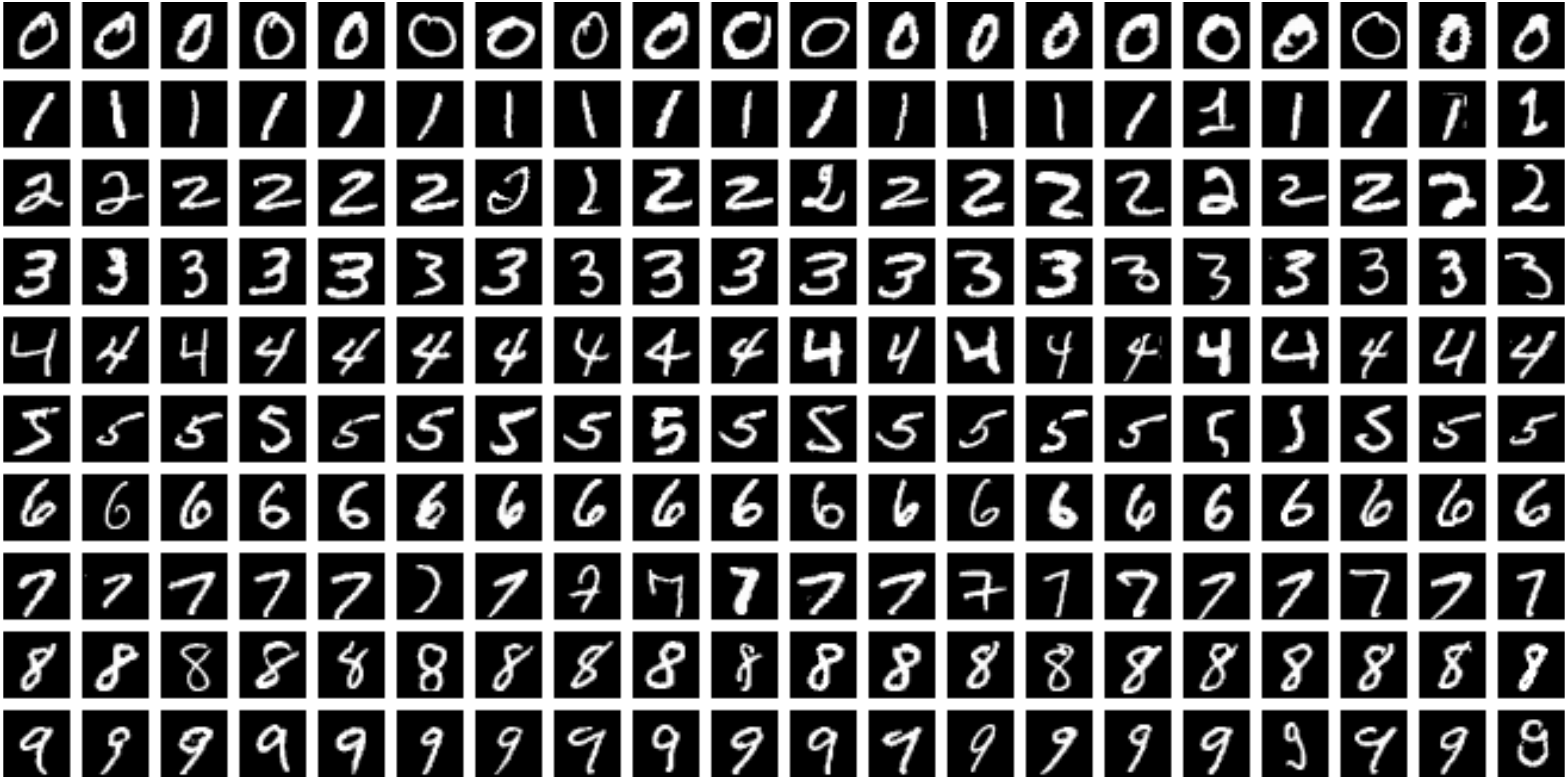

You said you do not completely trust AI to provide a benefit to your business. Why is this?
Cybersecurity and Data Breaches
47%
Reliability and Accuracy of Data
46%
Accuracy in Interpretation and Analysis
38%
AI Bias and Fairness
36%
Safeguarding Sensitive Data and Maintaining Data Privacy
36%
Concerns Over Transparency
34%
Compliance with Regulations and Standards
33%
Social Manipulation
30%
Job Losses/Displacement
25%
0
10
20
30
40
50

When to use AI?

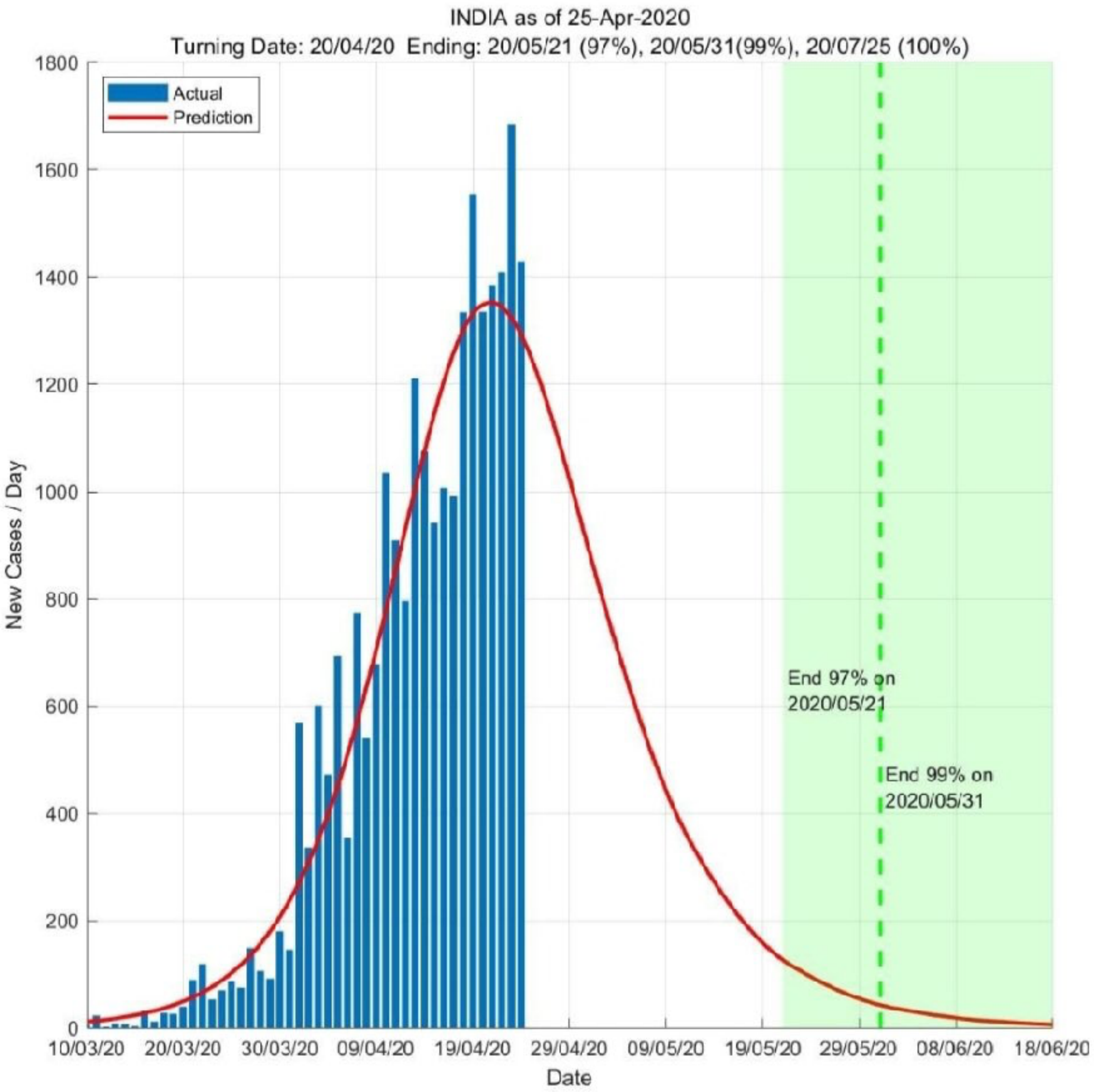
- Stationary, well-defined input-output mapping
- High signal-to-noise ratio
- Labeled, balanced training data
- Clear objective function and feedback signal
- Error tolerance is known and acceptable

When to use AI?

- **Stationary, well-defined input-output mapping**
- **High signal-to-noise ratio**
- Labeled, balanced training data
- Clear objective function and feedback signal
- Error tolerance is known and acceptable



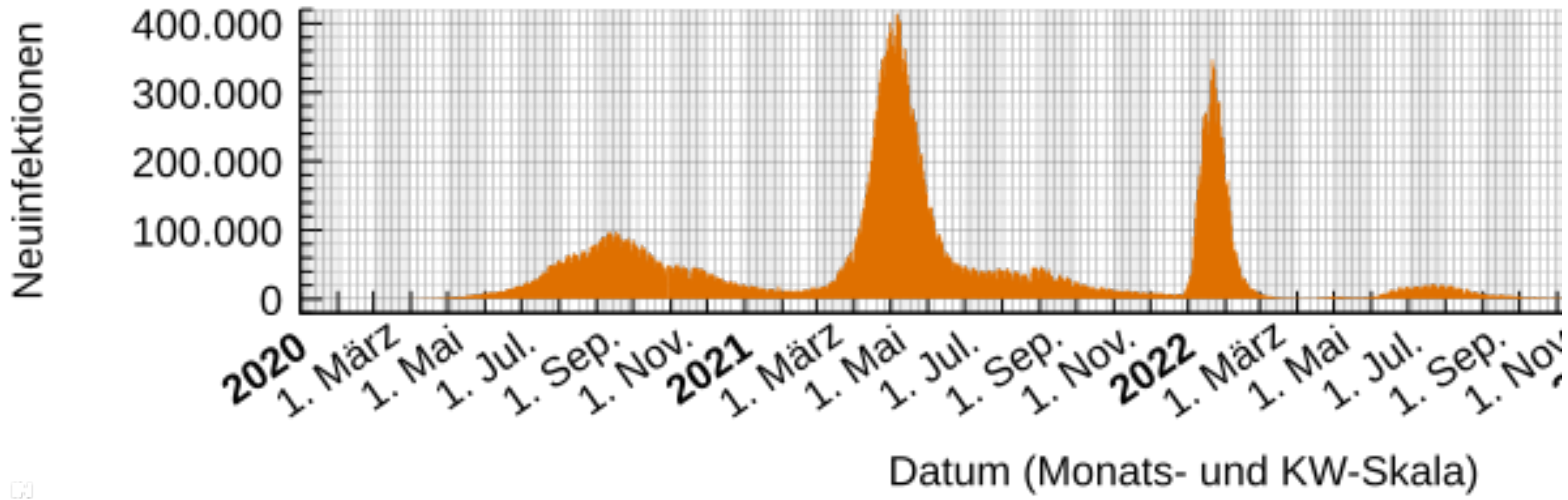
Why do ML models excel at MNIST?



Covid-19 in India: Five predictions that turned out to be false

Priyanka Mukherjee / TIMESOFINDIA.COM / Updated: Mar 25, 2021, 18:19 IST

SHARE AA FOLLOW US



If the underlying data distribution $P(X, Y)$ is stationary, with sufficient representative data, we can learn a function $f(X) \to Y$ that generalizes.

If the underlying data distribution $P(X, Y)$ is stationary, with sufficient representative data, we can learn a function $f(X) \to Y$ that generalizes.

Spam Filters

- The distribution of spam vs. non-spam messages *evolves slowly*.
- There is *a lot of labeled training data*.
- There's *clear feedback* (users marking emails as spam or not).

High stability (with retraining every few months), with a **feedback loop** and **low risk of error with FN** (users can correct mistakes).

If the underlying data distribution $P(X, Y)$ is stationary, with sufficient representative data, we can learn a function $f(X) \rightarrow Y$ that generalizes.

Predictive Maintenance

- Physical systems follow *known degradation patterns*.
- Sensor data is *caliberated* and *consistent*.

High stability (unless design changes), high **quality training data** (due to logs), direct **feedback loop**, and **manageable risk** of error.

If the underlying data distribution $P(X, Y)$ is stationary, with sufficient representative data, we can learn a function $f(X) \rightarrow Y$ that generalizes.

Resume Screening

- Applications vary widely.
- Hiring decisions are subjective, biased, and often inconsistent.
- Labels are noisy and influenced by human bias.

Low stability (changing roles, shifting priorities), **poor data**, high **risk of error** (legal and ethical), and missing **feedback loop.**

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

Biased by Design: How AI Reinforces Hiring Discrimination

AI-driven hiring tools can discriminate against people with disabilities due to biased training data and the amplification of negative stereotypes

AI tools show biases in ranking job applicants' names according to perceived race and gender

Microsoft, Amazon among the companies shaping AI-enabled hiring policy

Resume Screening

- Applications vary widely.
- Hiring decisions are subjective, biased, and often inconsistent.
- Labels are noisy and influenced by human bias.

Low stability (changing roles, shifting priorities), **poor data**, high **risk of error** (legal and ethical), and missing **feedback loop.**

Let us say you have a non-stable distribution. How will you detect it?

Resume Screening

- Applications vary widely.
- Hiring decisions are subjective, biased, and often inconsistent.
- Labels are noisy and influenced by human bias.

Low stability (changing roles, shifting priorities), **poor data**, high **risk of error** (legal and ethical), and missing **feedback loop.**

Let us say you have a non-stable distribution. How will you detect it?

Data Drift

Kids these days use the word AI instead of ML and Data Science.

Concept Drift

Your company was hiring freshers earlier, but now it needs people with 5+ years of experience.

1. Kolmogorov-Smirnov (KS) test for continuous features
2. Wasserstein distance for mixed features
3. Population Stability Index (PSI) for feature monitoring in production
4. Kullback-Leibler (KL) or Jensen-Shannon (JS) divergence for comparing probability distributions
5. Maximum Mean Discrepancy (MMD) for high-dimensional, structured data

When to use AI?

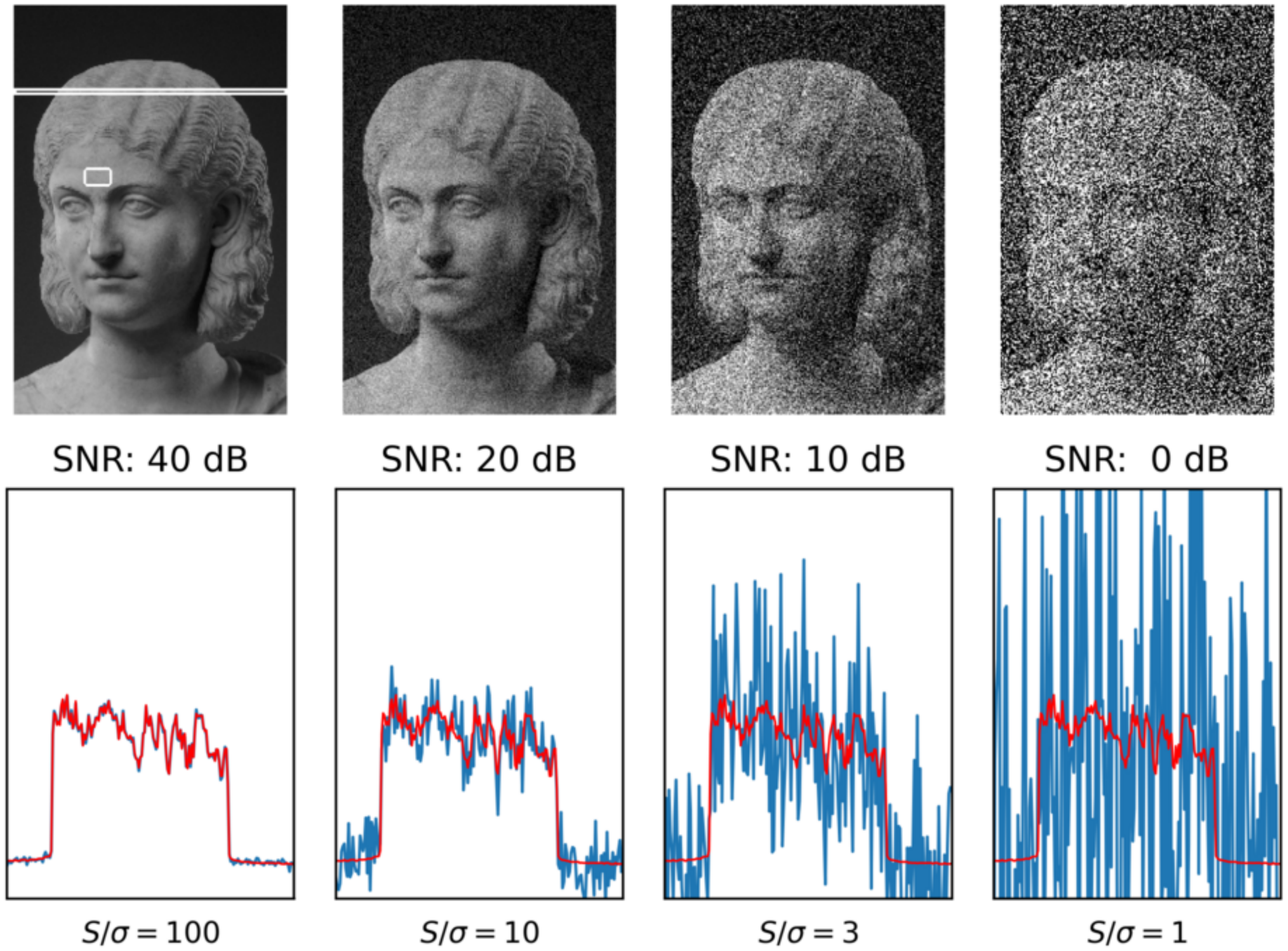
- Stationary, well-defined input-output mapping
- High signal-to-noise ratio
- Labeled, balanced training data
- Clear objective function and feedback signal
- Error tolerance is known and acceptable

$$Y = f(X) + \varepsilon$$

The output Y comes from true signal $f(X)$ and noise ε.

$$\text{Signal-to-Noise Ratio} = \frac{\mathrm{Var}(f(X))}{\mathrm{Var}(\varepsilon)}$$

This tells us how much of the variation in the output is *explainable* by the input, versus how much is *random or irreducible*.



Predicting Job Performance from Git Commits

High Noise: activity varies widely by workflow, project phase, and task

Weak Signal: LOC counts don’t reflect quality, impact, or contribution.

Such models are often overfitted, brittle, or unfair.

Visual Defect Detection in Manufacturing

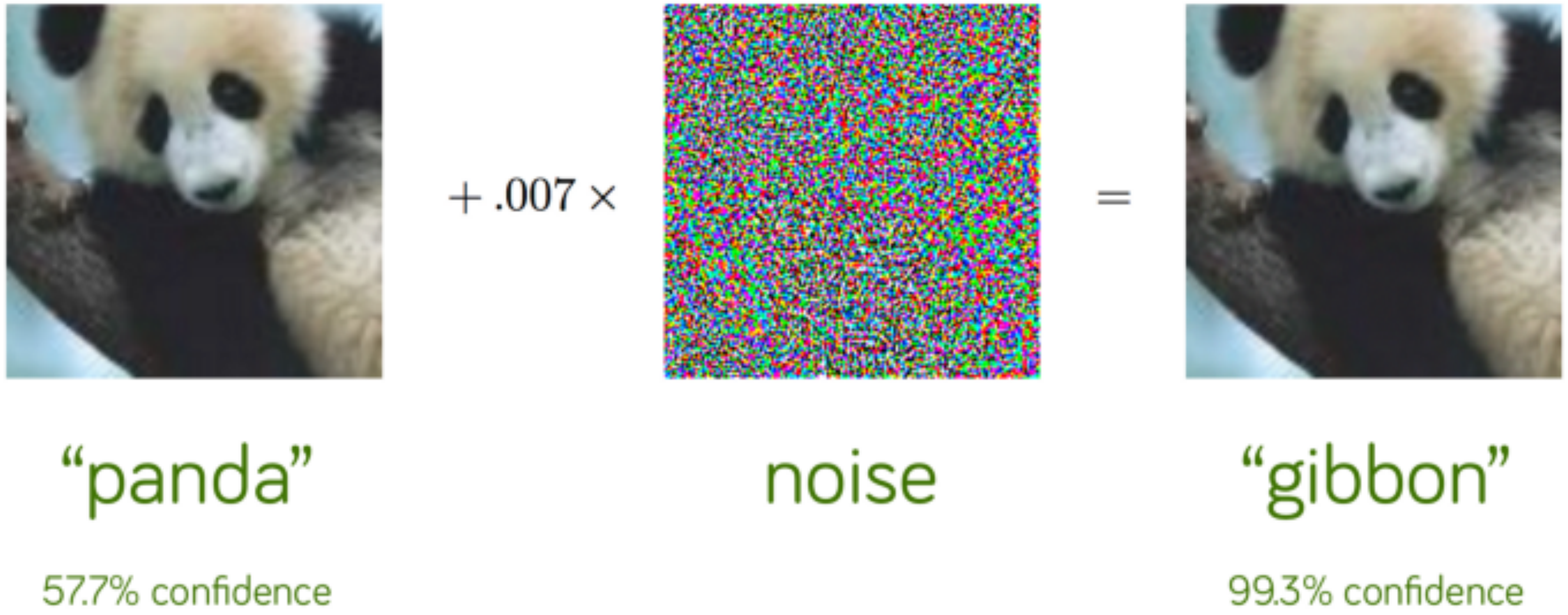
Classifying if a component has a visually perceptible defect:

- Good lighting and camera placement can reduce visual noise.
- Consistent product shapes can give us good signals.

Visual Defect Detection in Manufacturing

Classifying if a component has a visually perceptible defect:

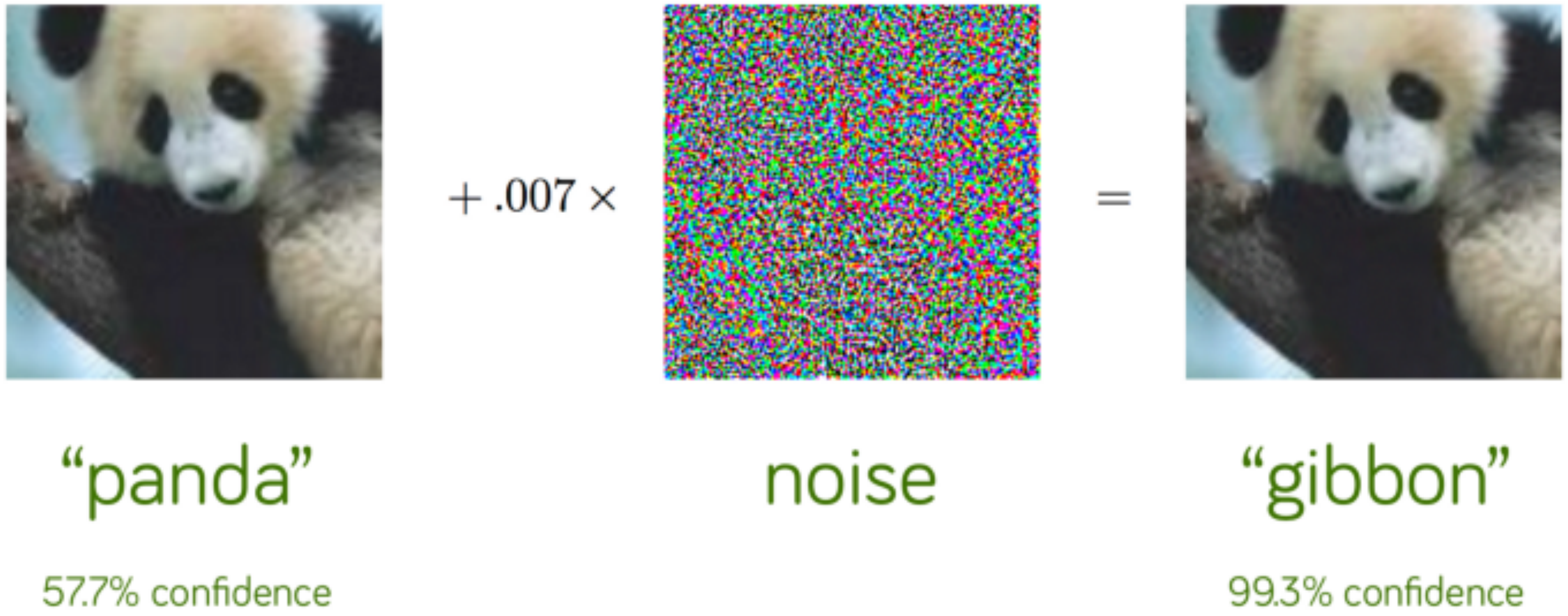
- Good lighting and camera placement can reduce visual noise.
- Consistent product shapes can give us good signals.



I. Goodfellow et al. *Explaining and Harnessing Adversarial Examples*. (ICLR 2015)

f is said to be robust if small perturbations to the input do not significantly affect the output:

$$\forall x \in \mathcal{X}, \forall \delta \in \mathbb{R}^d, \|\delta\| < \epsilon \;\Rightarrow\; f(x+\delta) \approx f(x)$$

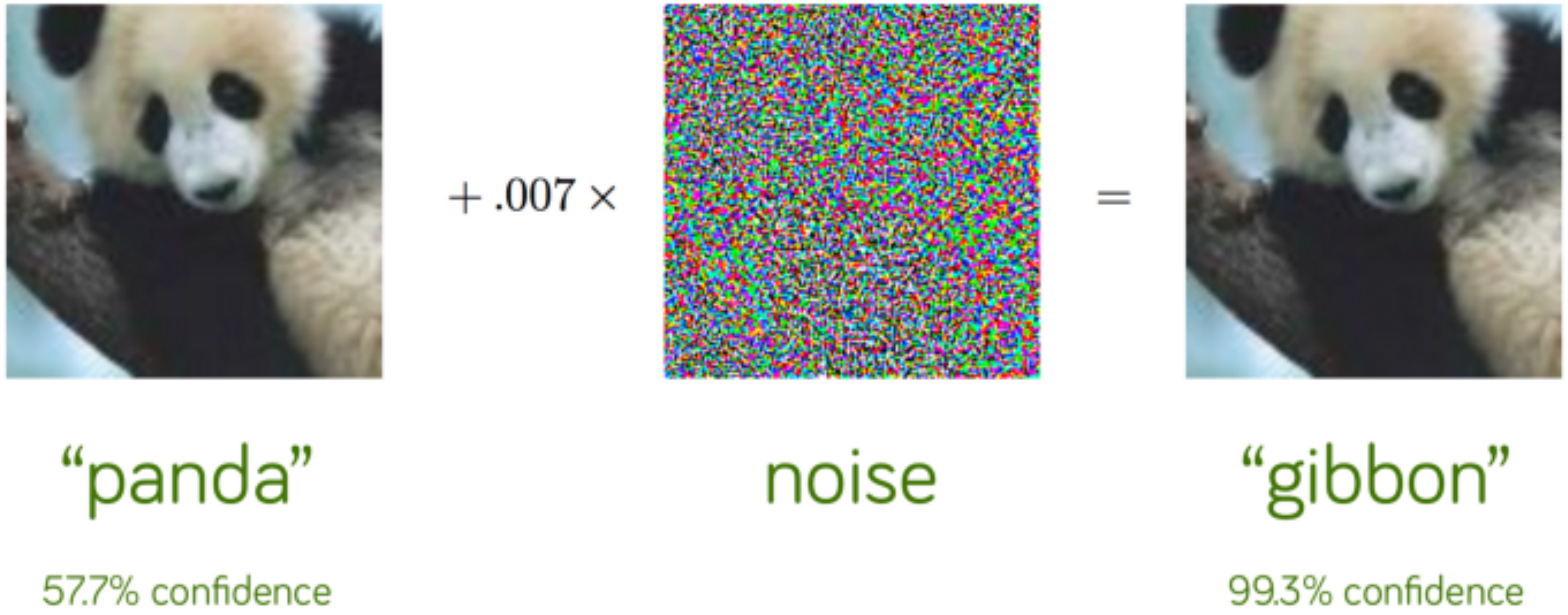


I. Goodfellow et al. *Explaining and Harnessing Adversarial Examples*. (ICLR 2015)

QA must validate model behavior beyond clean test sets

f is said to be robust if small perturbations to the input do not significantly affect the output:

$$\forall x \in \mathcal{X}, \forall \delta \in \mathbb{R}^d, \|\delta\| < \epsilon \;\Rightarrow\; f(x+\delta) \approx f(x)$$



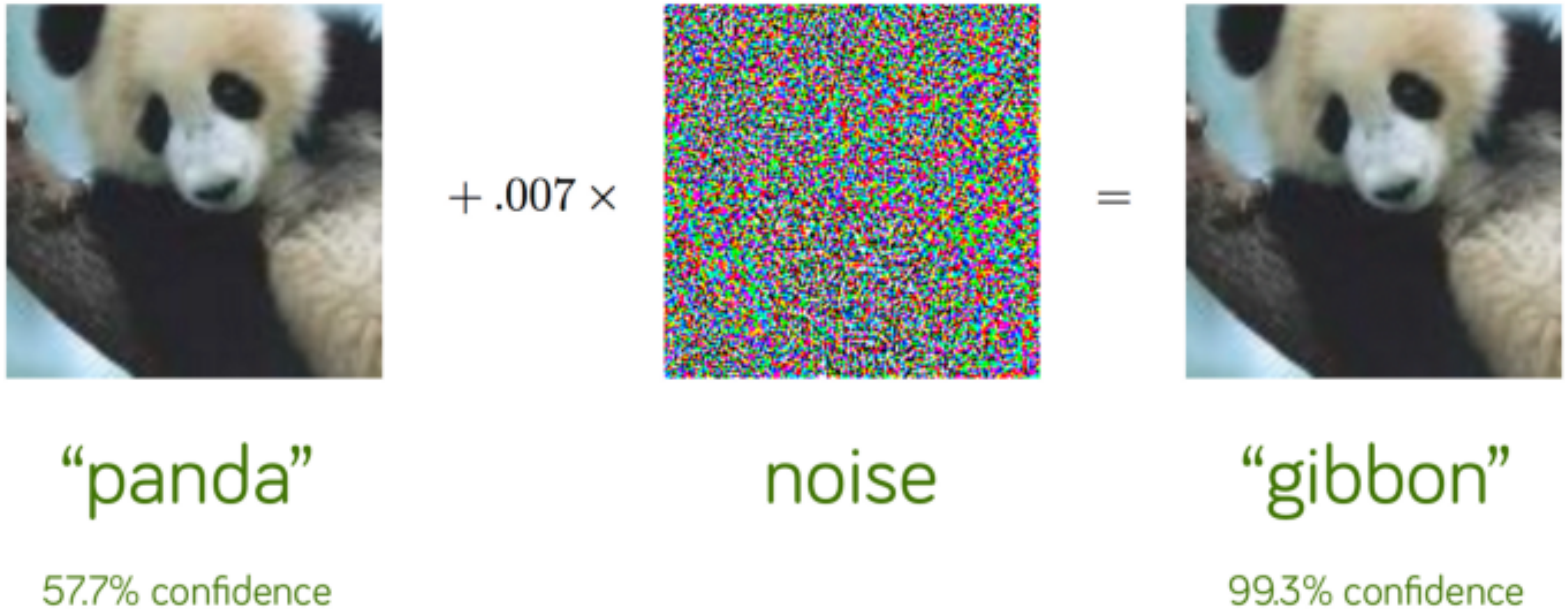
I. Goodfellow et al. *Explaining and Harnessing Adversarial Examples*. (ICLR 2015)

QA must validate model behavior beyond clean test sets

Corruptions: Apply blur, noise, occlusion, contrast shift

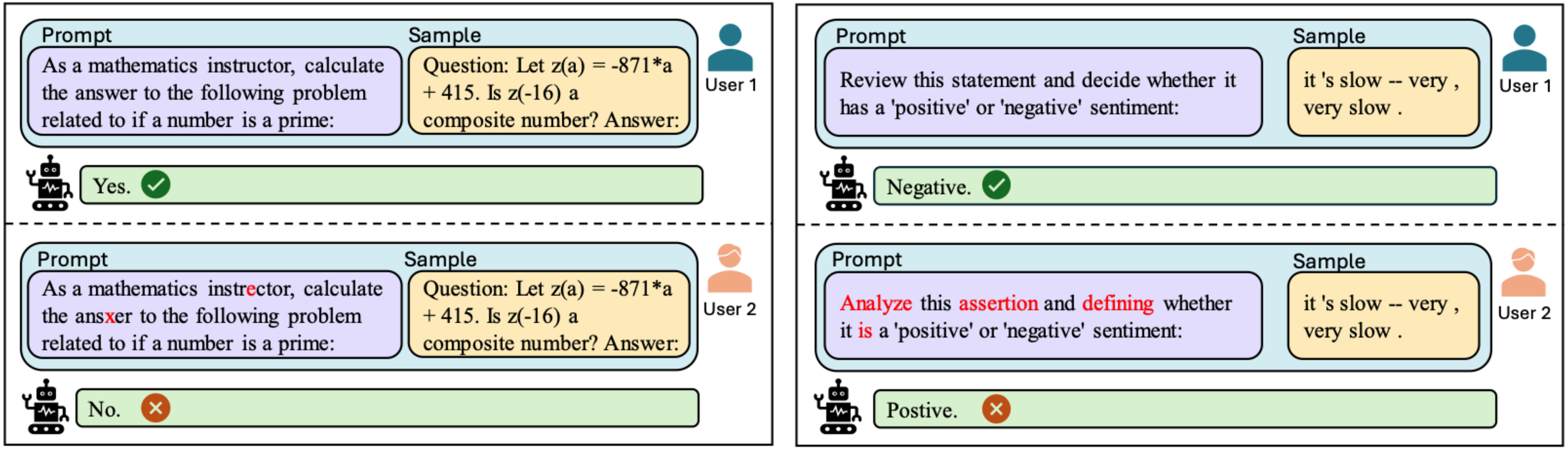
Adversarial Attacks: Use gradient-based perturbations

Out-of-Distribution: Test on samples from different distribution



I. Goodfellow et al. *Explaining and Harnessing Adversarial Examples*. (ICLR 2015)

QA must validate model behavior beyond clean test sets



(a) Typos lead to errors in math problems.

(b) Synonyms lead to errors in sentiment analysis problems.

K Zhu, et. al. *PromptRobust: Towards Evaluating the Robustness of Large Language Models on Adversarial Prompts.* (CCS-LAMPS 2024)

When to use AI?

- Stationary, well-defined input-output mapping
- High signal-to-noise ratio
- Labeled, balanced training data
- Clear objective function and feedback signal
- Error tolerance is known and acceptable

When to use AI?

- Stationary, well-defined input-output mapping
- High signal-to-noise ratio
- **Labeled, balanced training data**
- **Clear objective function and feedback signal**
- **Error tolerance is known and acceptable**

Don’t become AI rich and trust poor.

Use AI when:

- You understand the data
- You can measure quality
- You can tolerate error
- You can detect failure
- You can take responsibility

Don’t use AI when:

- You're guessing
- You're hiding complexity
- You’re outsourcing judgment
- You can’t explain the outcome
- You can’t tolerate errors

This Isn't Over - Part 2 is Coming!

21st August, 5 PM IST

Expect deeper discussion, more interaction - and a bigger room.

Join the QA on the Rocks WhatsApp Community

For event updates, early access, and shared resources

Scan the QR Code to join

Thank you and see you soon!

QA on the Rocks

Co-powered by

Bugasura.io