Scenario:

You want release statistical data for analysis or machine learning without compromising the privacy of the individuals in the dataset.

(e.g. medical, financial, retail data)

(Central) Differential Privacy

Differential privacy is a mathematical framework designed to **control** the amount of **privacy loss** suffered in a system. Its fundamental promise is that users are not affected by conclusions drawn from a study they contributed their data to any more adversely than they would be if they did not participate at all. Data is obfuscated by **'random statistical noise'**, controlled by the privacy loss parameter **ε** (epsilon), that gets added to the data. The noise **maintains underlying patterns** in the data but **hides** any **identifying information**, so an individual's privacy is not compromised.

Use cases

- Machine learning
- Analysing data while preserving privacy

Known Applications

- United States 2020 Census
- Google's COVID-19 Search Trends Dataset

Advantages

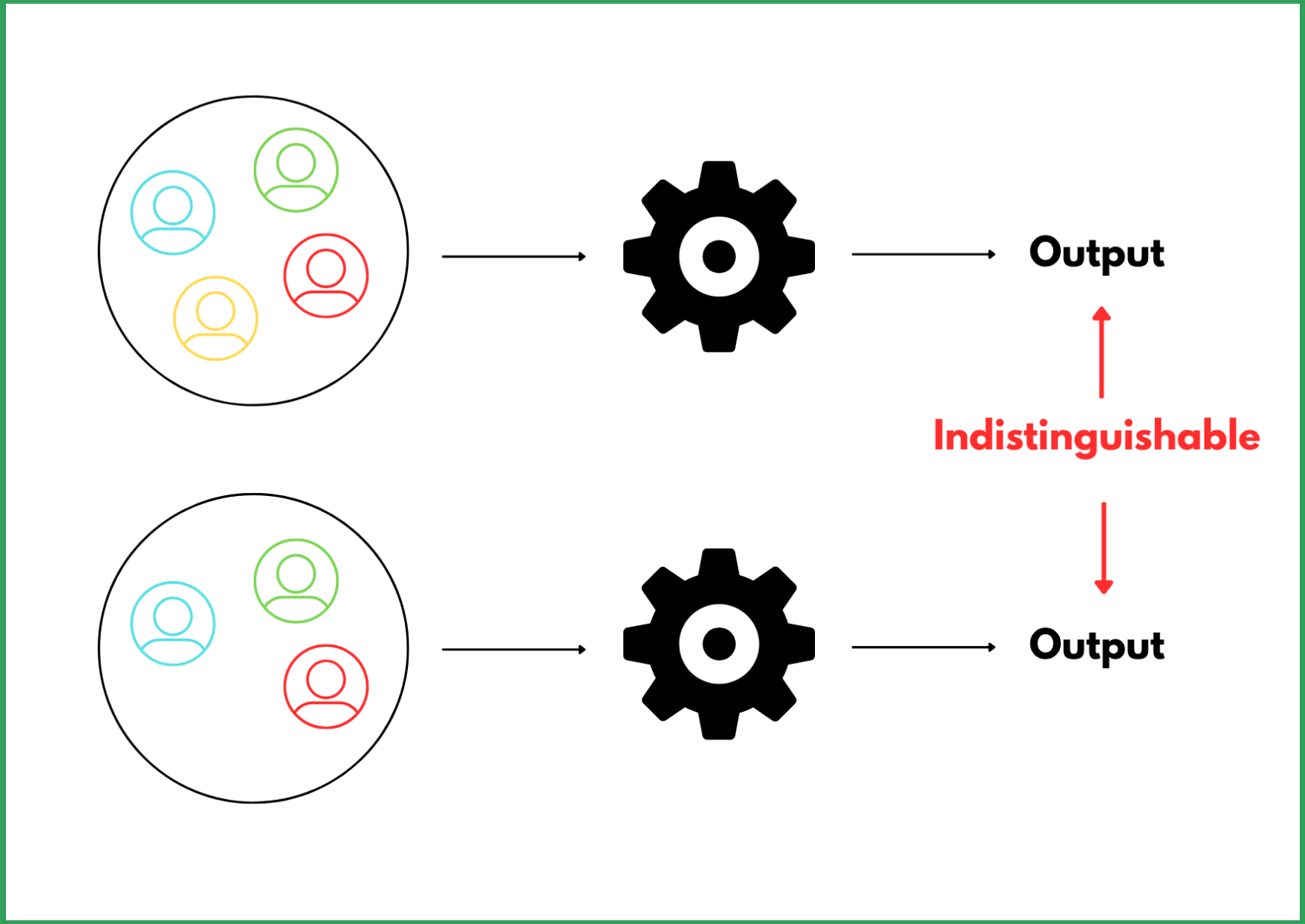
- Privacy guarantee holds regardless of background information available
- The result produced by combining the output of two differentially private algorithms is still differentially private
- Public access to data without impacting privacy of the individuals it is based on
- Provides greater accuracy compared to 'local differential privacy'

Limitations

- Privacy guarantee may be too strong for some data releases
- Does not prevent statistical inference drawn from results
- Does not perform as well on small datasets

Scan for further reading recommendations

(Central) Differential Privacy



Scenario:

You want to use data collected from an end-user's device to learn about a group of users without actually collecting their personal data.

Local Differential Privacy

In Local Differential Privacy (DP), individuals in a dataset **ensure their own privacy** by applying **noise** to their data **before** they send it to the curator. The noise **maintains underlying patterns** in the data but hides any identifying information, so **an individual's privacy is not compromised**. By adding noise themselves, individuals do not have to trust a third-party with their unprotected data. However, this comes at the cost of a large amount of total noise attributed to the aggregated dataset. This can be mitigated by using **higher values of ε.** (Refer to the 'Differential Privacy' card for a more in-depth discussion.)

Use cases

- User data stored on their device
- Protection more important than accuracy

Known Applications

- Apple's Emoji Suggestion Algorithm
- Google's Spanish Word Prediction Algorithm

Advantages

- Privacy guarantee holds regardless of background information available
- The result produced by combining the output of two differentially private algorithms is still differentially private
- Users do not need to trust a third-party to keep their data secure
- Provides a greater privacy guarantee compared to 'central differential privacy'

Limitations

- Not useful when trying to find interactions between multiple variables
- Poorer accuracy when data aggregated compared to central differential privacy's guarantees

Scan for further reading recommendations

Scenario:

Differential privacy is not a suitable framework to apply to your dataset

(e.g. your dataset is too small, too much utility is lost etc.)

K-anonymity

K-anonymity is a **de-identification** model designed to **protect the privacy of individuals in a data release**. To satisfy k-anonymity, information about a specific individual in a data release must not be distinguishable from **multiple other individuals** who also appear in the release. This is controlled by the value **'k'**. For example, if k is set to 3, there are at least 3 individuals who **have the exact same record**. To achieve anonymity, the data custodian needs to protect a set of attributes they define as **quasi-identifiers** (QI), which are **any attributes that an adversary might have access to that can be used in combination to identify an individual**.

Use cases

- No background information is available
- Differential privacy is impractical

Advantages

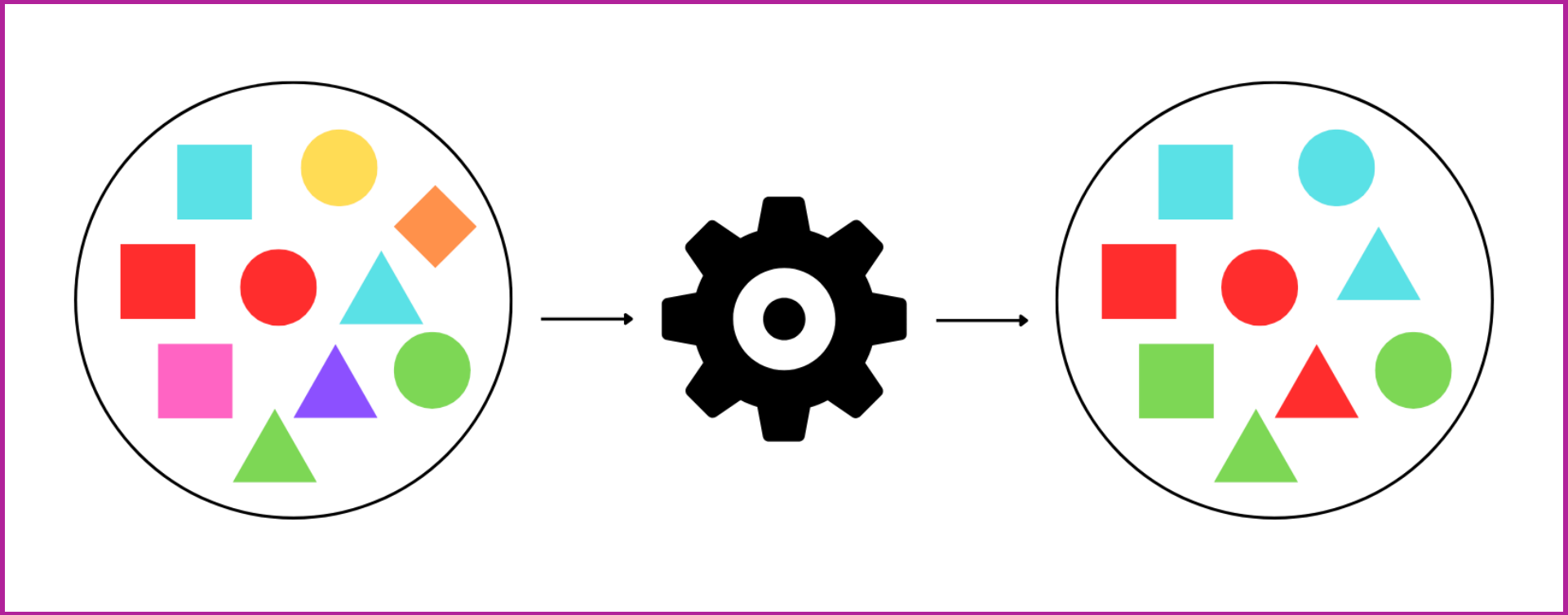
- Provides some privacy protection - using k-anonymity is better than using nothing to protect your data
- Data utility is maintained if anonymization is done correctly
- Easier to implement than differential privacy, and has been used in practice for longer
- Works well enough on small, tabular datasets

Limitations

- Does not perform well on high-dimensional datasets
- Does not provide enough protection against all possible background information available
- Its privacy guarantees are not preserved if two releases based on the same underlying data are combined
- **Should not be first choice for a privacy model due to its weaknesses**

Scan for further reading recommendations

K-anonymity



Scenario:

You want to combine the data you have collected with data from a different organisation without either of you having to reveal your data to each other.

(e.g. to use in research or when data compliance issues are a concern)

Multi-Party Computation (MPC)

Multi-Party Computation (MPC) is a **cryptographic protocol** that allows participants to compute an output based on their combined inputs **without revealing** what their inputs are to each other. While other cryptographic protocols are designed to hide data from an external adversary, MPC also **hides an individual's data from anyone that is not the individual themself**. As a result, MPC aims to guarantee that it is **impossible** to gain information about an individual's input **beyond what can be learned from the output**. MPC can be used in combination with 'differential privacy' to protect the interests of both data owners and subjects.

Use cases

- Communicating highly sensitive data
- Data mining

Known Applications

- Businesses collaborating to gain industry insights

Advantages

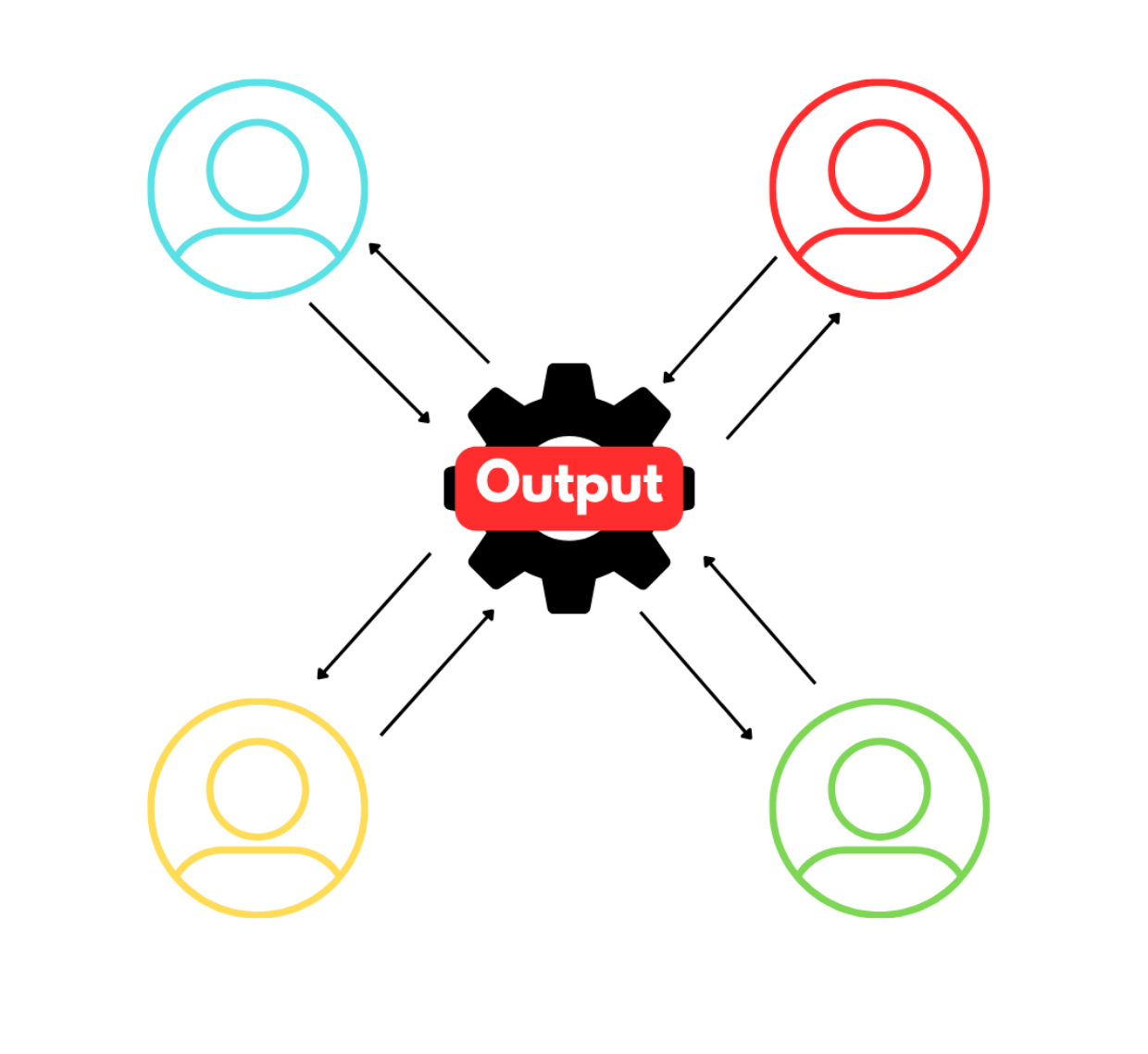
- Reduces trade-off between data privacy and utility
- Reveals only the final result instead of the entire database
- Data analysts can run computations on encrypted data and still draw meaningful conclusions
- Secure data sharing in healthcare and business industries
- Results are accurate

Limitations

- Can be compromised if some participating parties decide to collude with each other to expose other participants' secret values
- Significant communication overhead, making it slow to run, especially on small datasets

Scan for further reading recommendations

Multi-Party Computation (MPC)



Scenario:

You have vulnerable groups in your data that you want to make sure are properly protected by your privacy framework OR you want a clearer way to work out which privacy-preserving values will protect your data.

Demographic Coherence Enforcement

Inspired by criticisms about how abstract differential privacy is, Demographic Coherence Enforcement (DCE) functions as a **'harms-aware'** framework for **reasoning about the privacy impact** of a data release. It fundamentally aims to stop predictors from making **'demographically incoherent'** predictions on data that could be used to **guess something harmful** about an individual based on their presence in a dataset, even if it is not clear if the prediction is inaccurate. DCE can be used with differential privacy to **suggest a meaningful ε** (epsilon) **value** that takes into account specific constraints a data curator wants to put in place.

Use cases

- If you want a result that is easier to interpret and assess because it maps to your use cases

Known Applications

- N/A - was only developed in 2025

Advantages

- Based around a predictor's capacity to harm, which is more consistent and easier to reason about than differential privacy
- Can be used with differential privacy to improve guarantees
- Theoretically easier to implement than differential privacy
- Makes it so you can argue you are protecting your sensitive data in the right way

Limitations

- Still in early days of development
- Privacy guarantees are not as strong as those provided by differential privacy
- Unclear if the result of combining two outputs from a DCE algorithm is still private

Scan for further reading recommendations