

Artificial Intelligence Evidence



Heather L. King
heather@koonsfuller.com

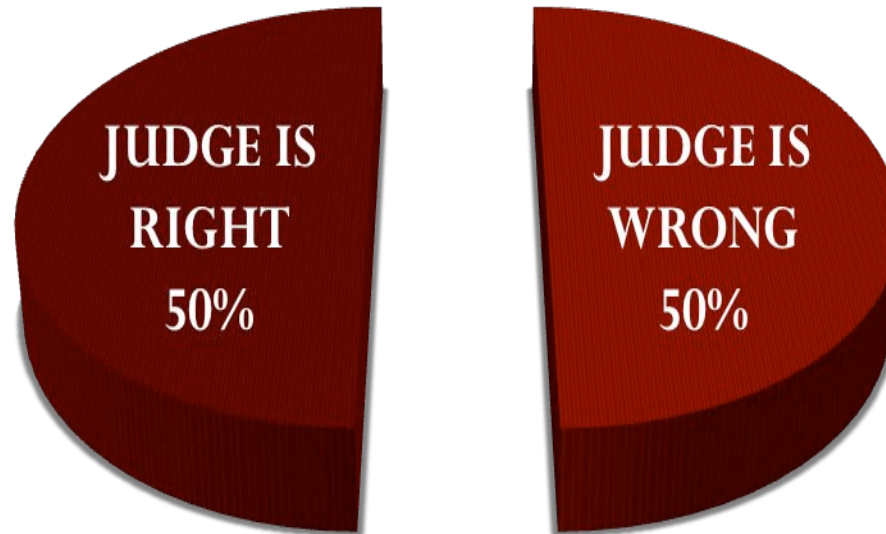


heather@koonsofuller.com

PDF or PPT



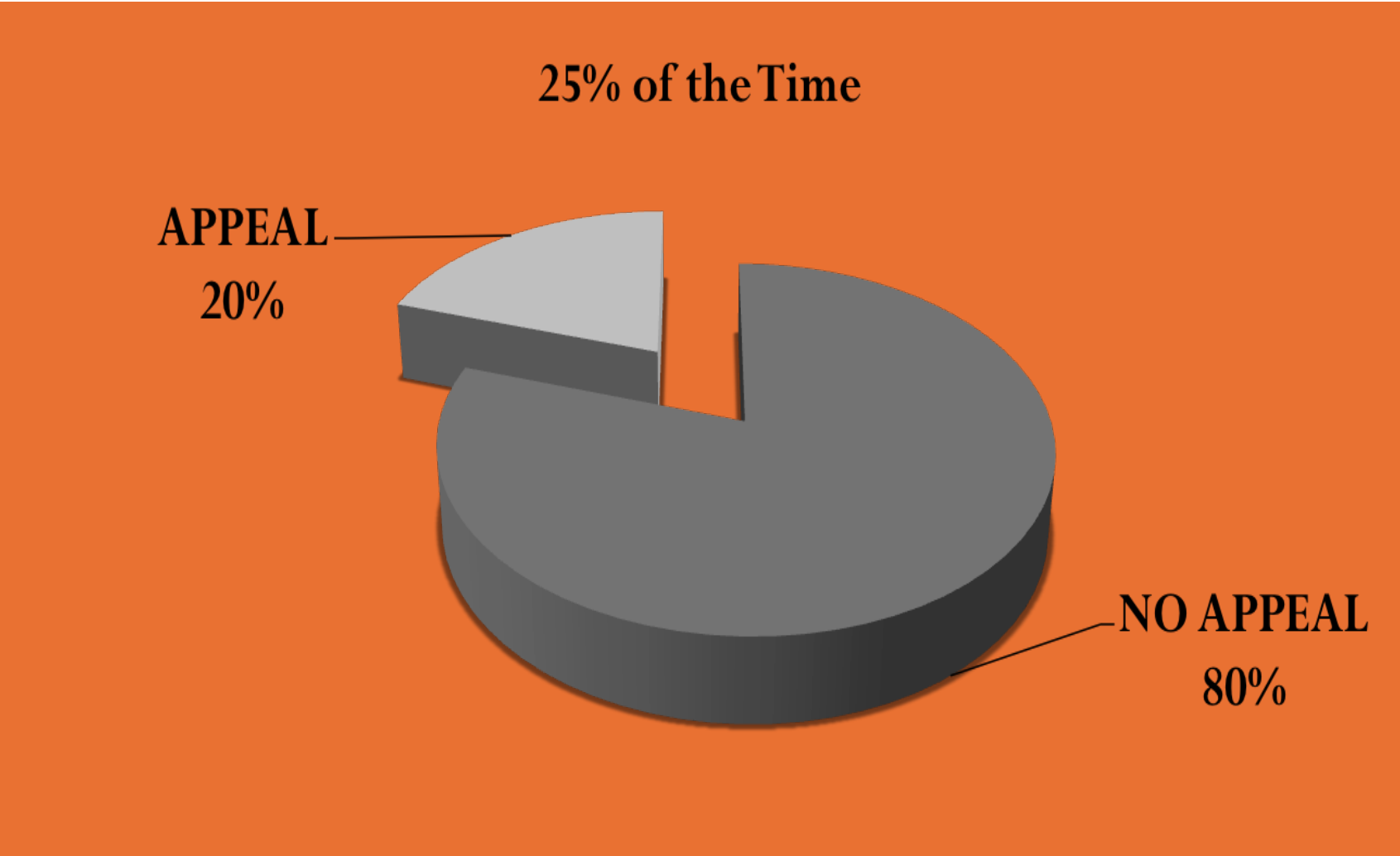
Evidentiary Rulings [100% of the time]



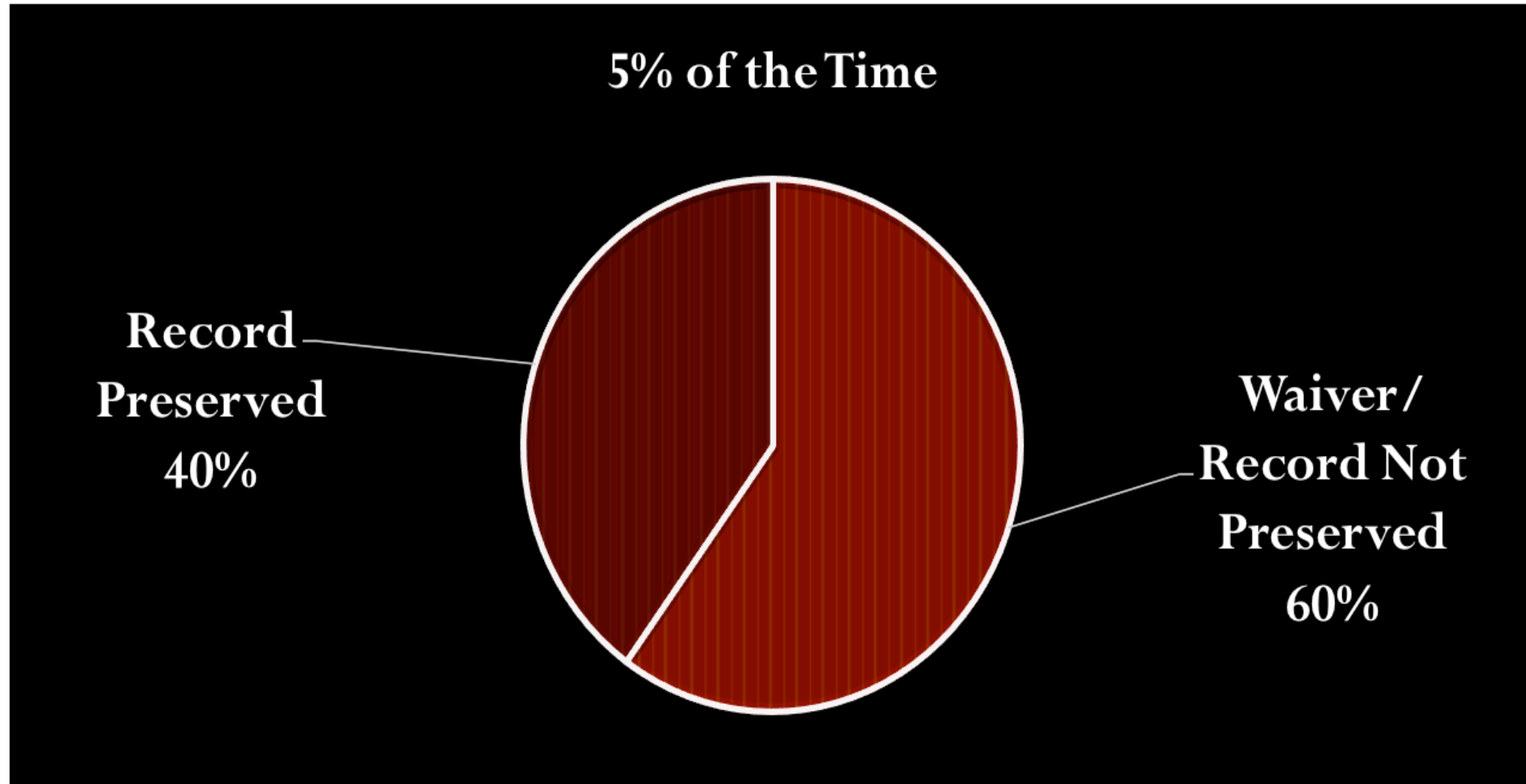
Evidentiary Rulings [50% of the time]



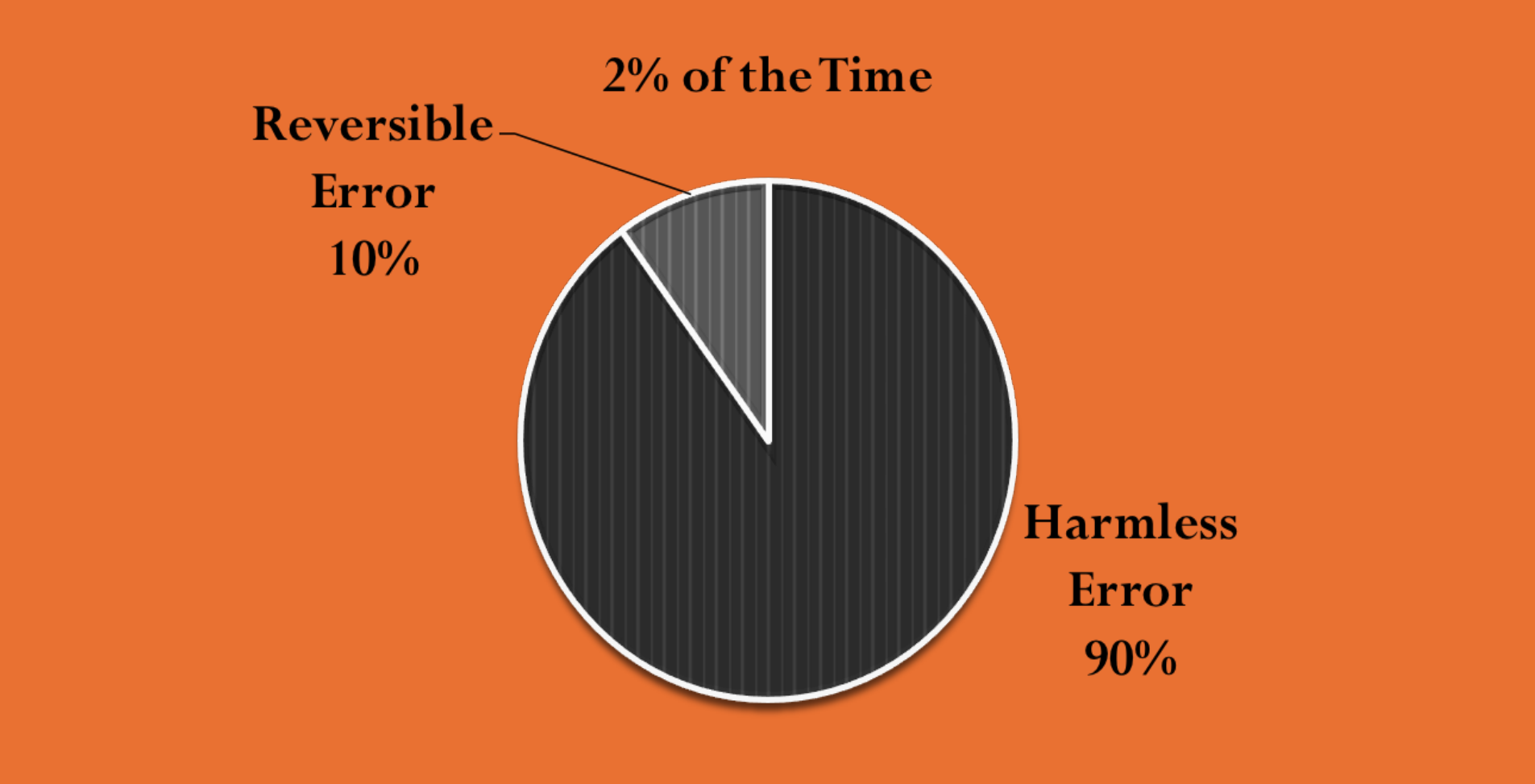
Evidentiary Rulings [25% of the time]



Evidentiary Rulings [5% of the time]



Evidentiary Rulings [2% of the time]



0.2%

Research all Jurisdictions

Evidence rules
are very similar, if
not almost
exactly the same

Look for theories
to argue



Prerequisites to Admissibility

- Relevant
 - **TRE 401**
- Authentic
 - **TRE 901-902**
- Not Hearsay
 - **TRE 801-805**
- Original or Duplicate
 - **TRE 1001**
- Probative Value vs. Unfair Prejudice
 - **TRE 403**






Precedential Case Law



Prerequisites to Admissibility


- Relevant
 - **TRE 401**
 - Authentic
 - **TRE 901-902**
 - Not Hearsay
 - **TRE 801-805**
 - Original or Duplicate
 - **TRE 1001**
 - Probative Value vs. Unfair Prejudice
 - **TRE 403**
- 



Precedential Case Law

[Criminal Cases]

Prerequisites to Admissibility

- Relevant
 - **TRE 401**
 - Authentic
 - **TRE 901-902**
 - Not Hearsay
 - **TRE 801-805**
 - Original or Duplicate
 - **TRE 1001**
 - Probative Value vs. Unfair Prejudice
 - **TRE 403**
- 



Precedential Case Law

[Criminal Cases]

**Case Law
from other Jurisdictions
[Federal/State]**

Prerequisites to Admissibility

- Relevant
 - **TRE 401**
- Authentic
 - **TRE 901-902**
- Not Hearsay
 - **TRE 801-805**
- Original or Duplicate
 - **TRE 1001**
- Probative Value vs. Unfair Prejudice
 - **TRE 403**





Precedential Case Law

[Criminal Cases]

Case Law
from other Jurisdictions
[Federal/State]

[even other countries]

Prerequisites to Admissibility

- Relevant
 - **TRE 401**
- Authentic
 - **TRE 901-902**
- Not Hearsay
 - **TRE 801-805**
- Original or Duplicate
 - **TRE 1001**
- Probative Value vs. Unfair Prejudice
 - **TRE 403**



No Special Rules for Advanced AI (yet)

- While [advanced] AI employs technology which may exceed most human cognitive ability, like electronic communications, there is no separate evidentiary standard for “deep” or “advanced” AI.
- Evidence gleaned from AI should be judged by the standard of direct witness testimony, expert witness testimony (measurement using established technology).
- In sum, **AI evidence is subject to the same rules of evidence as non-AI sources.**



- While AI is often conceived of as a computer matching or exceeding a human's performance, **in truth it is still just software.**
- Historically, traditional software is admissible if it passes the normal requirements of admissibility.
- It's the **difference between traditional software and AI** that creates more unique admissibility considerations.



Computer Records vs. Computer Generated Evidence

- Like traditional software, AI produces two types of computer evidence, computer records and computer-generated evidence.
- Computer records are generally print outs compiled by a computer in a prescribed fashion from data. They don't require analysis or assumption by the underlying programming, whereas computer generated evidence does.
- Computer generated evidence is computer output ***based on data and assumptions contained in a program.***



Computer Records vs. Computer Generated Exhibits

The admissibility of both computer records and computer-generated exhibits are the same as they are for traditional paper business records and traditional demonstrations, respectively.

The fact that a computer is involved does not change the admissibility standards or procedures, but **there are unique challenges that may come into play with advanced AI**

Levels [Artificial Intelligence]

- **Narrow AI or artificial narrow intelligence (ANI);**
- General AI or artificial general intelligence (AGI)
- Super AI or artificial superintelligence (ASI)
- Reactive machines
- Limited memory
- Theory of mind
- Self-aware



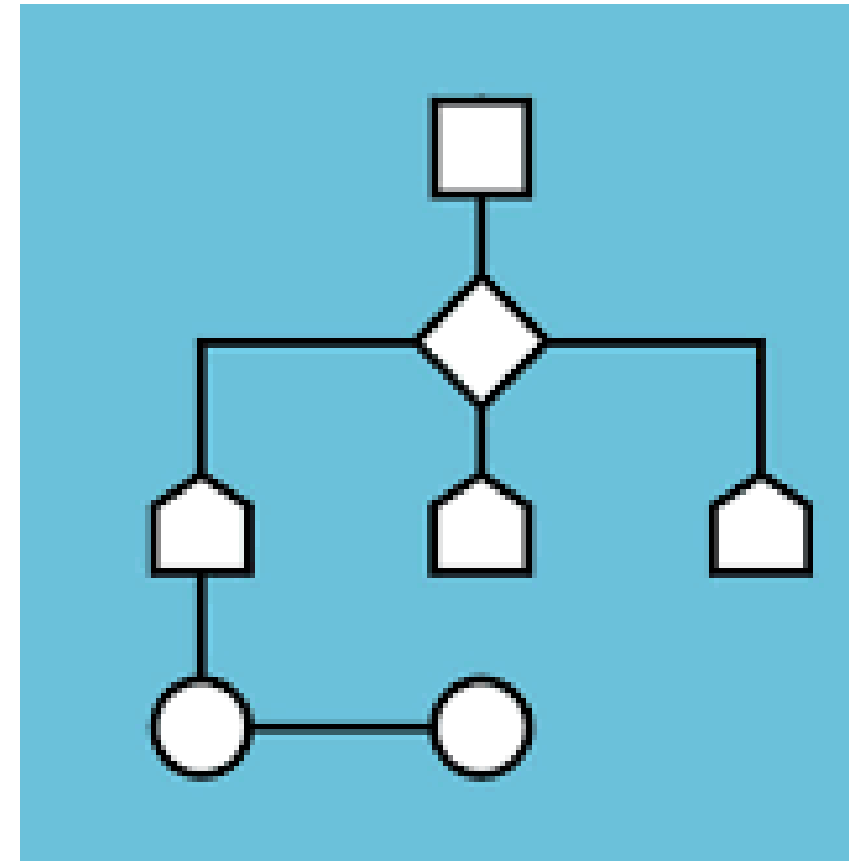


Narrow AI – “Weak AI”

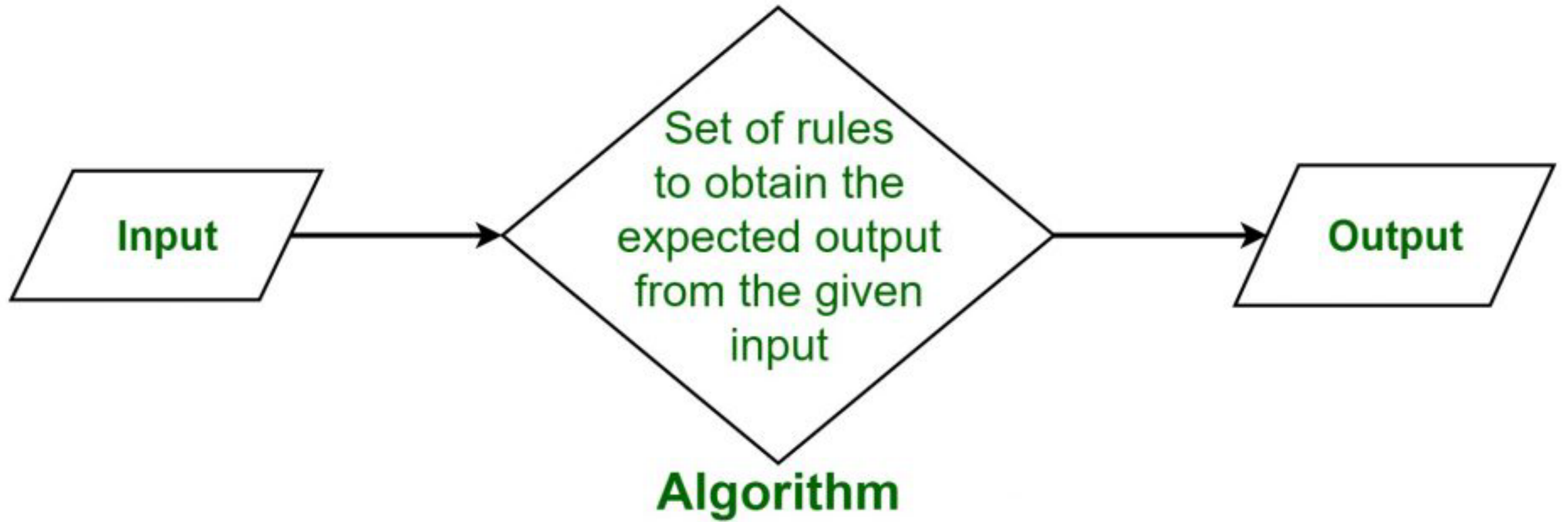
- The only type of AI that exists today. Any other form of AI is theoretical. It can be trained to perform a single or narrow task, often far faster and better than a human mind can.
- All other forms of AI are still theoretical.

Algorithms are a Set of Rules/Instructions

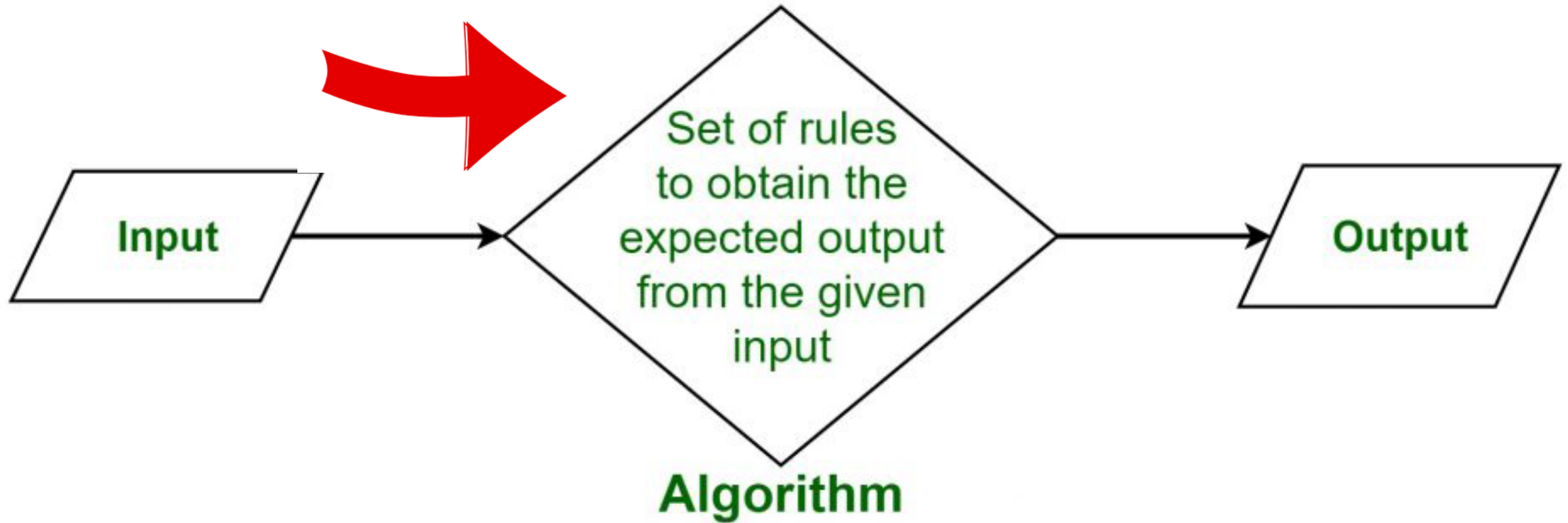
- Both traditional software and AI contain algorithms.
- Algorithms are procedures employed for solving a problem or performing a computation.
- Algorithms act as a step-by-step list of instructions specifying specific actions to be performed (using software or hardware routines).



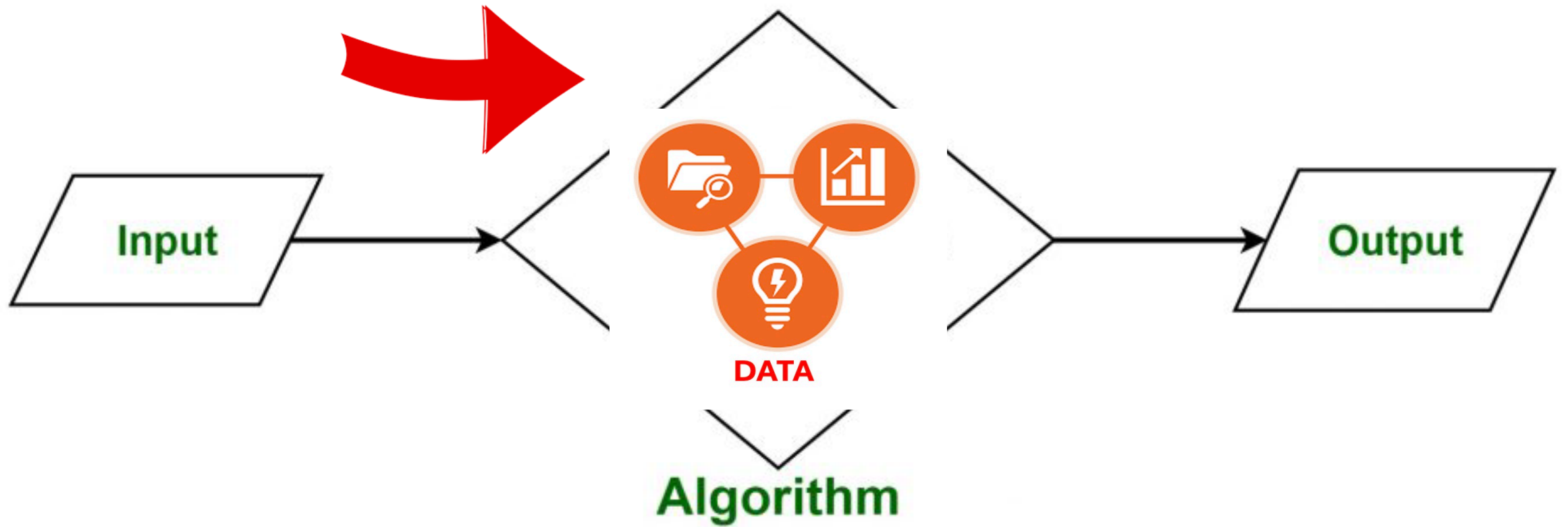
What is Algorithm?

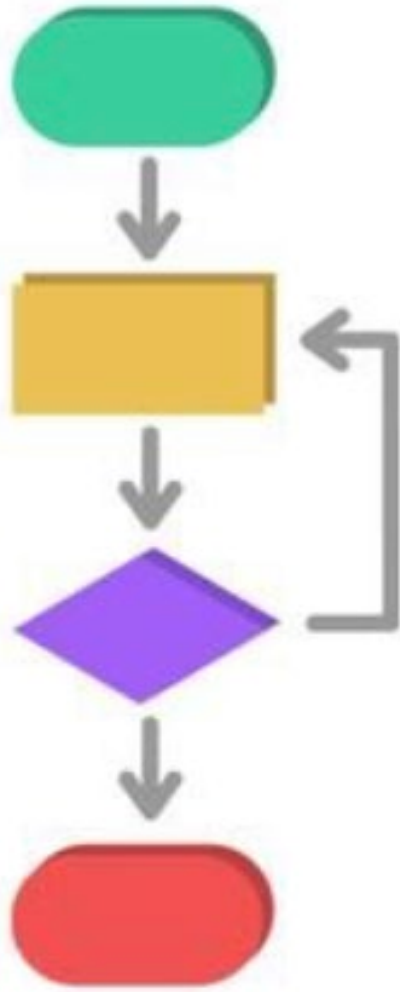


What is Algorithm?



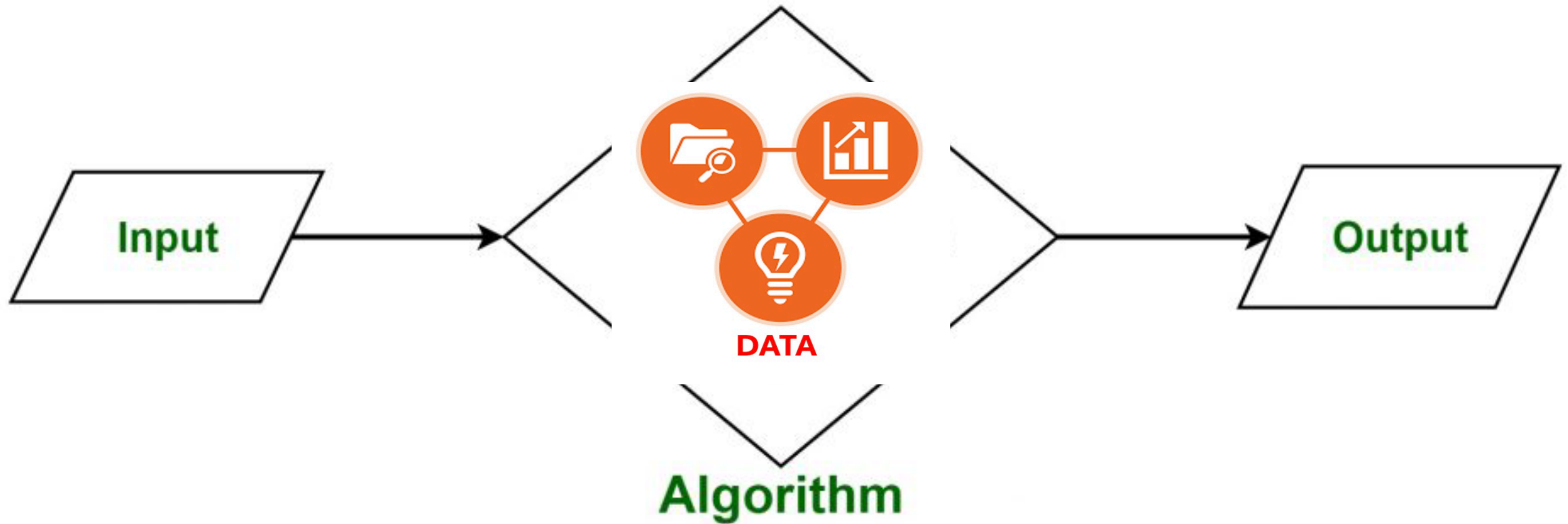
What is Algorithm?



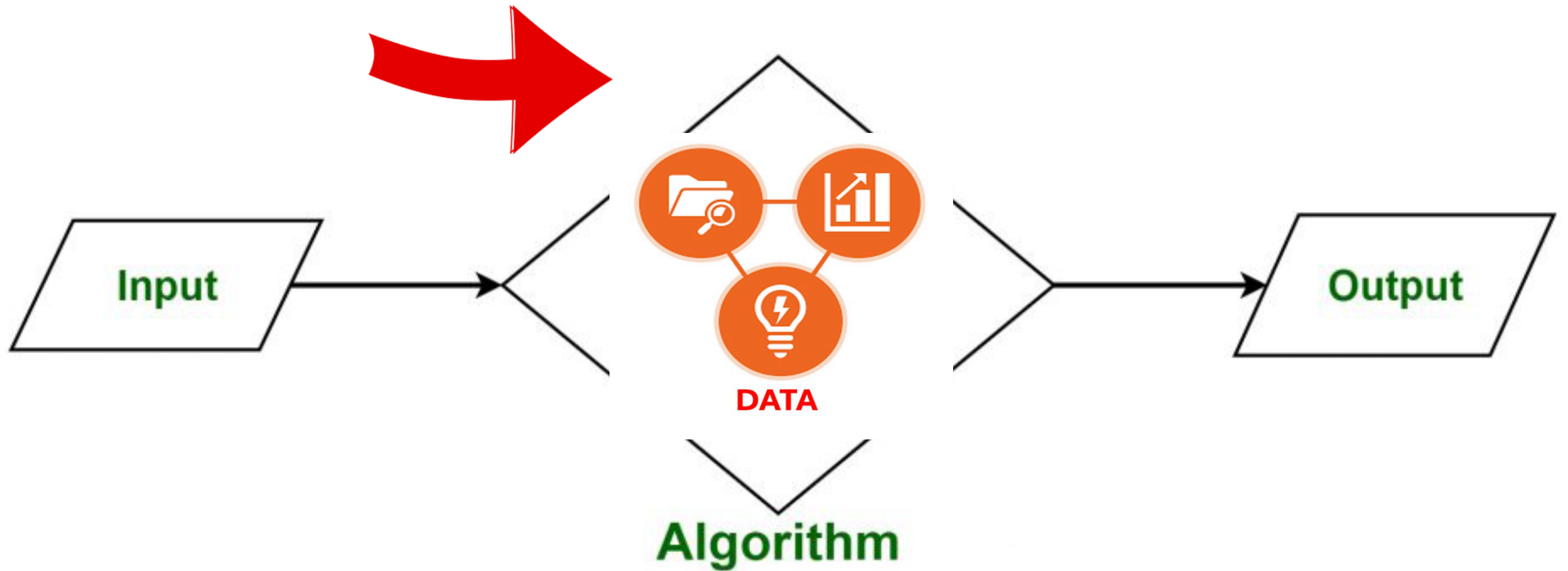


- The ***difference*** with AI and traditional algorithms is that
- A traditional algorithm will always generate the same output for a given input.
- ***An AI algorithm can change its outputs based on new input (data)***

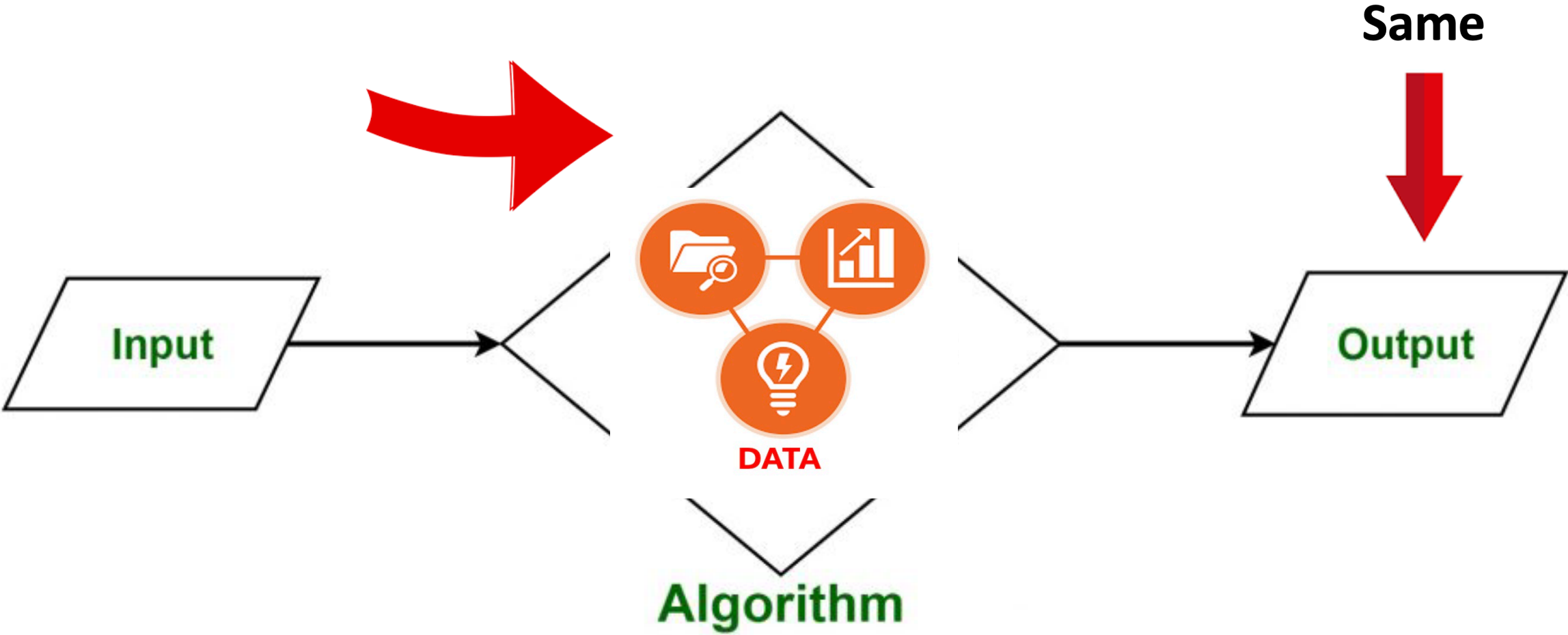
Traditional Algorithm



Traditional Algorithm



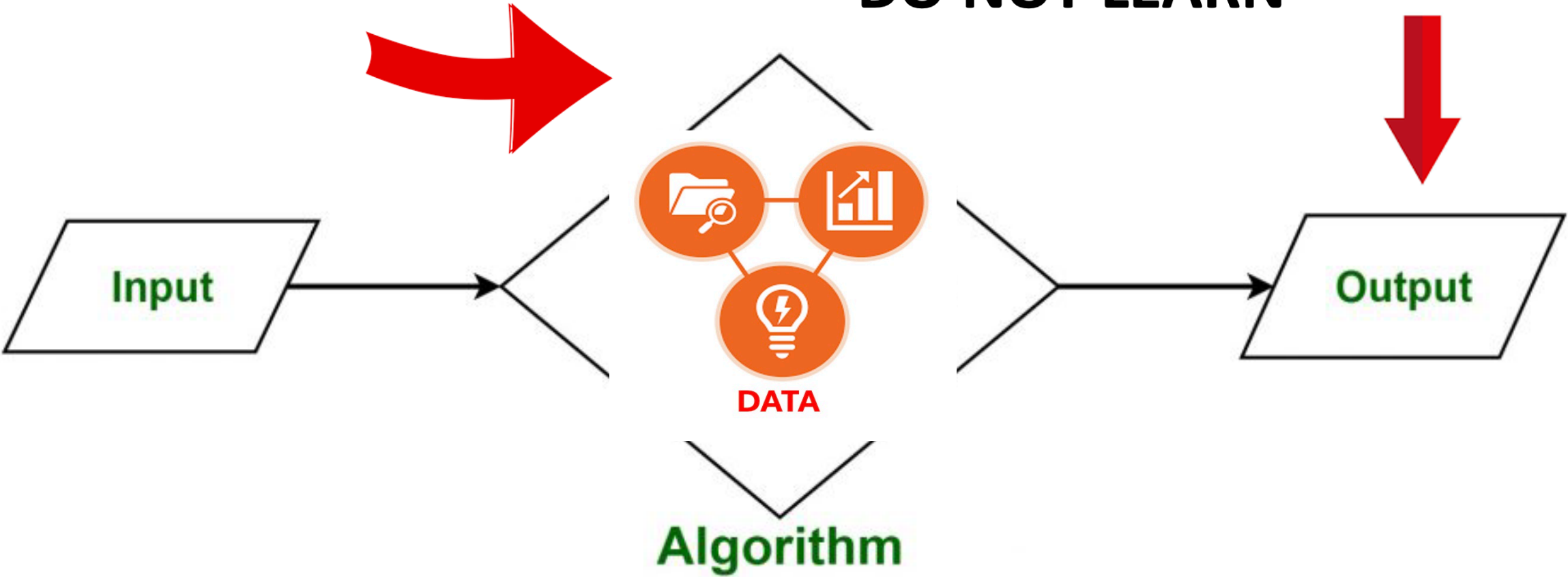
Traditional Algorithm



Traditional Algorithm

DO NOT LEARN

Same

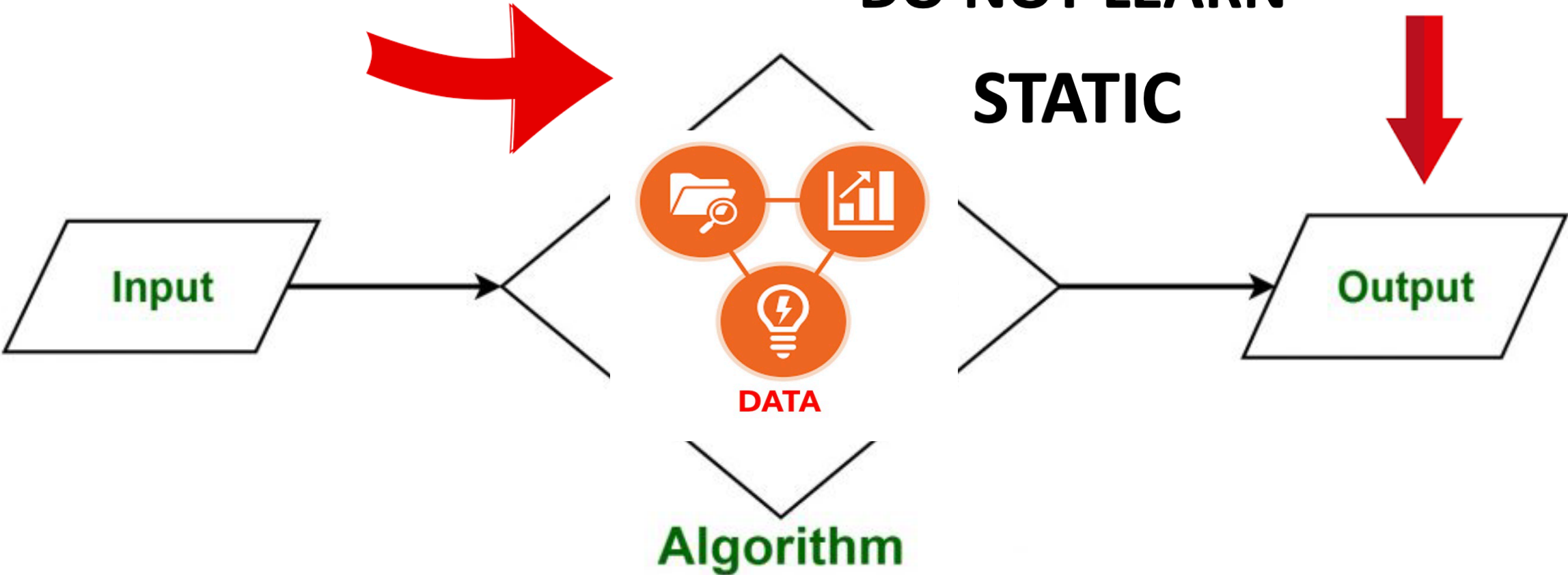


Traditional Algorithm

DO NOT LEARN

Same

STATIC

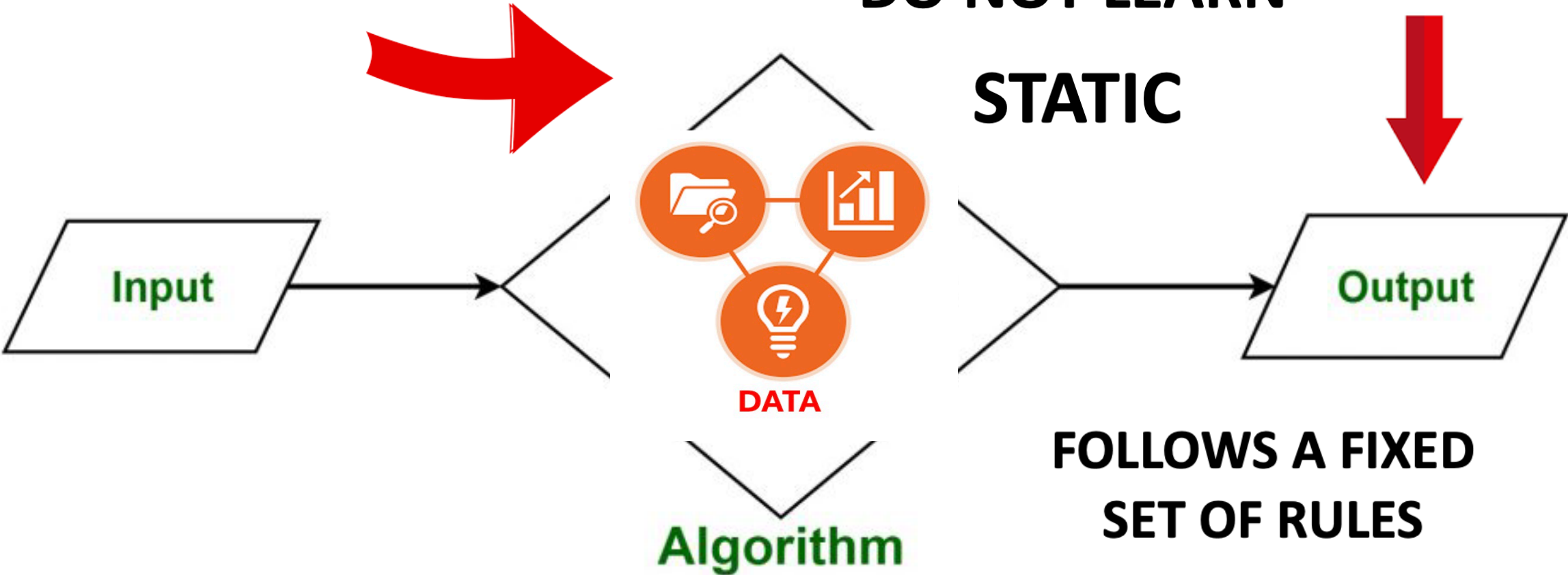


Traditional Algorithm

DO NOT LEARN

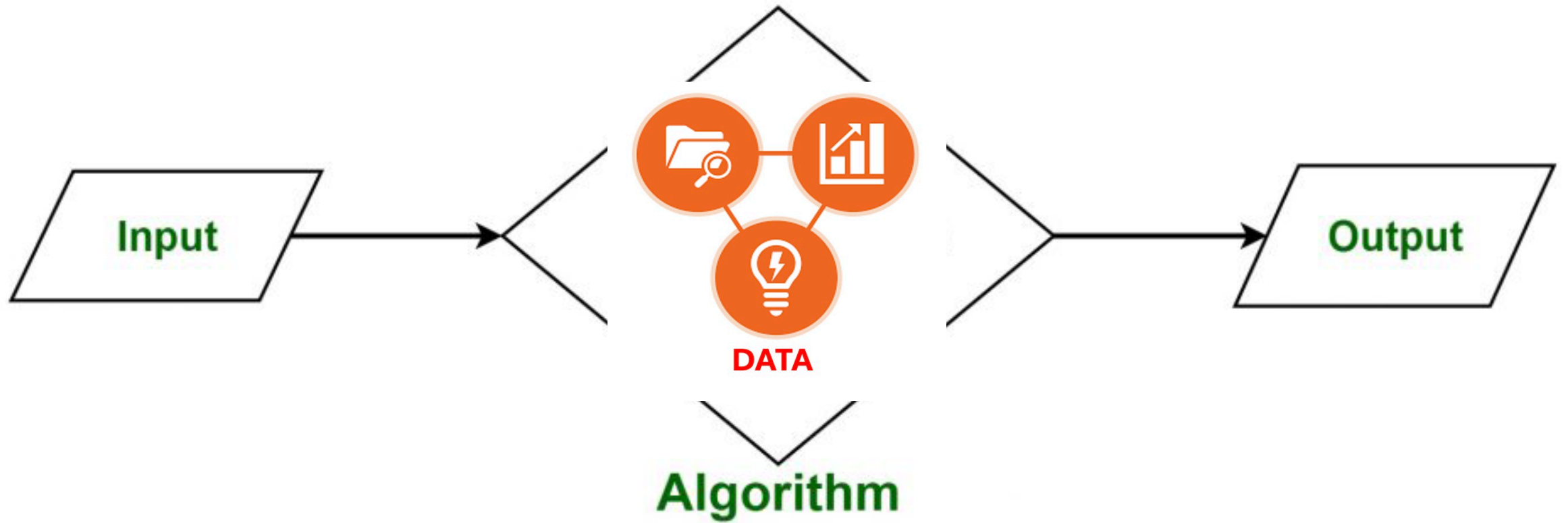
Same

STATIC

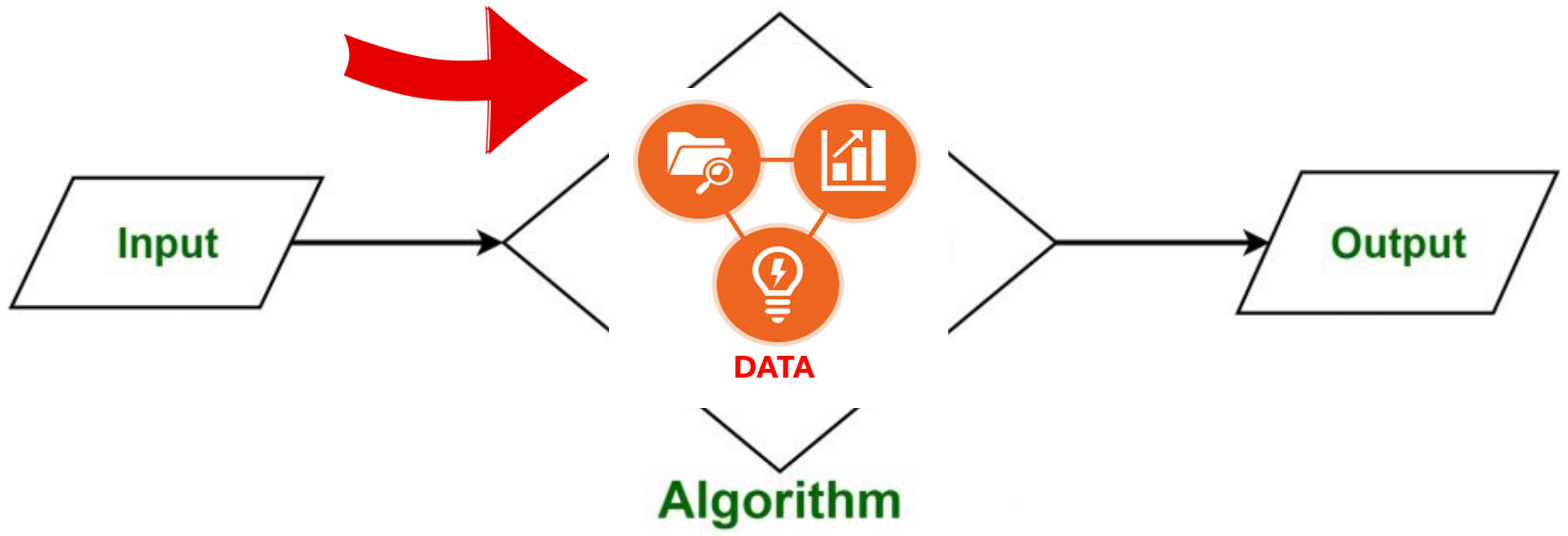


FOLLOWS A FIXED SET OF RULES

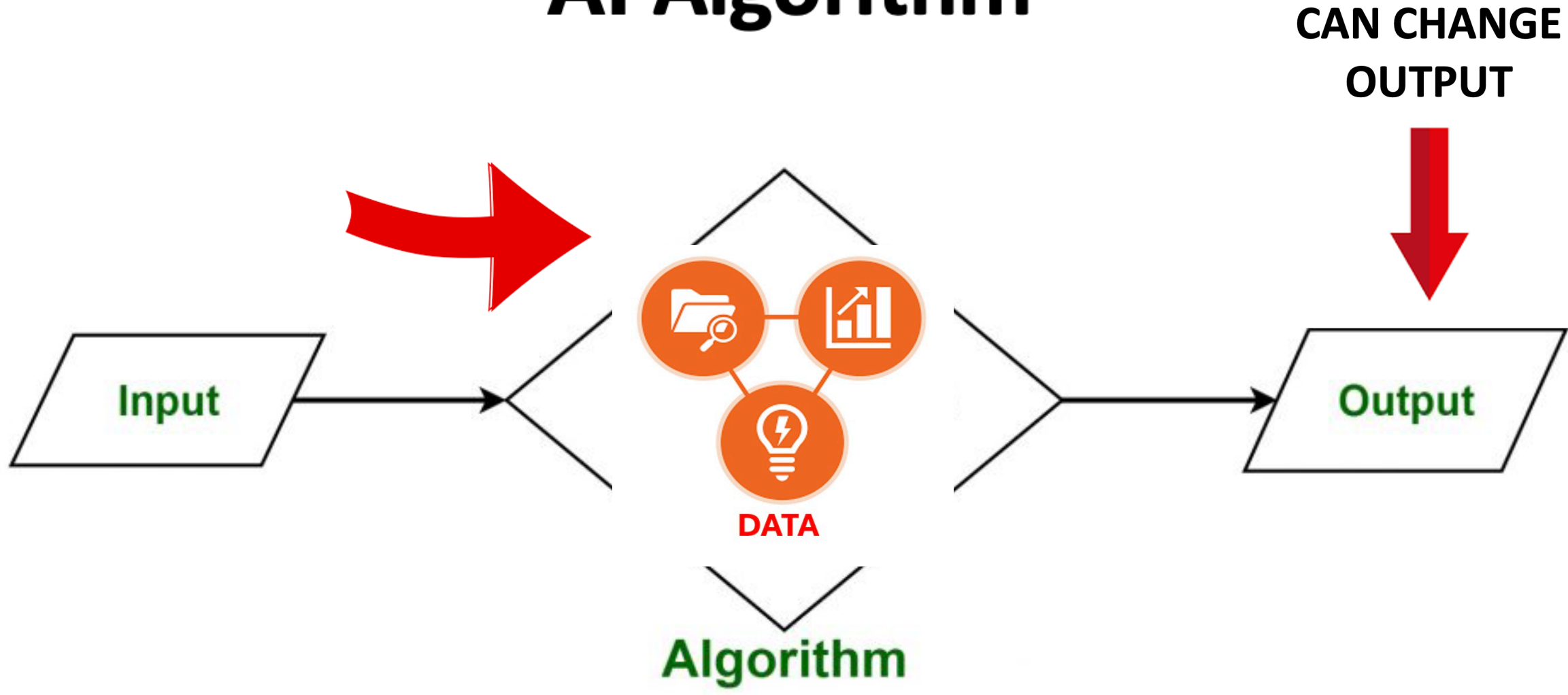
AI Algorithm



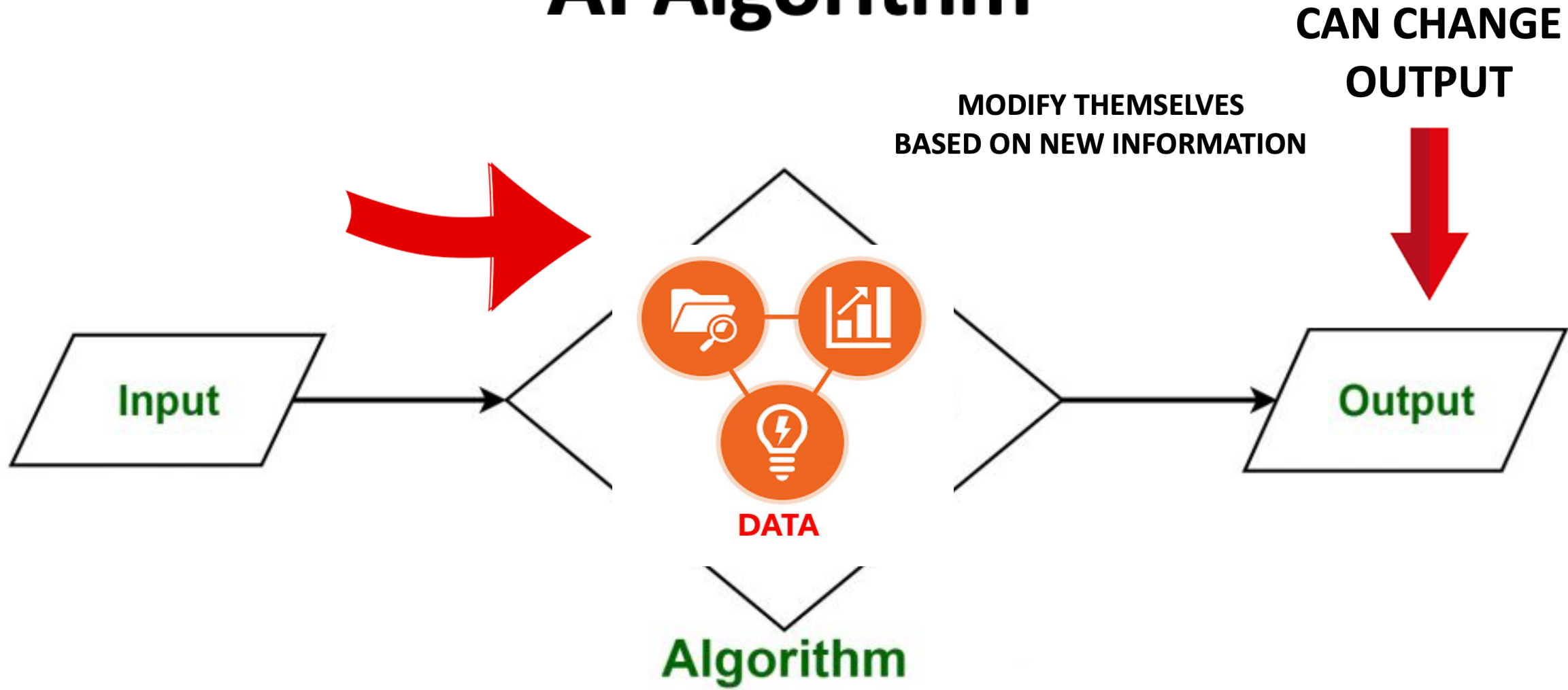
AI Algorithm



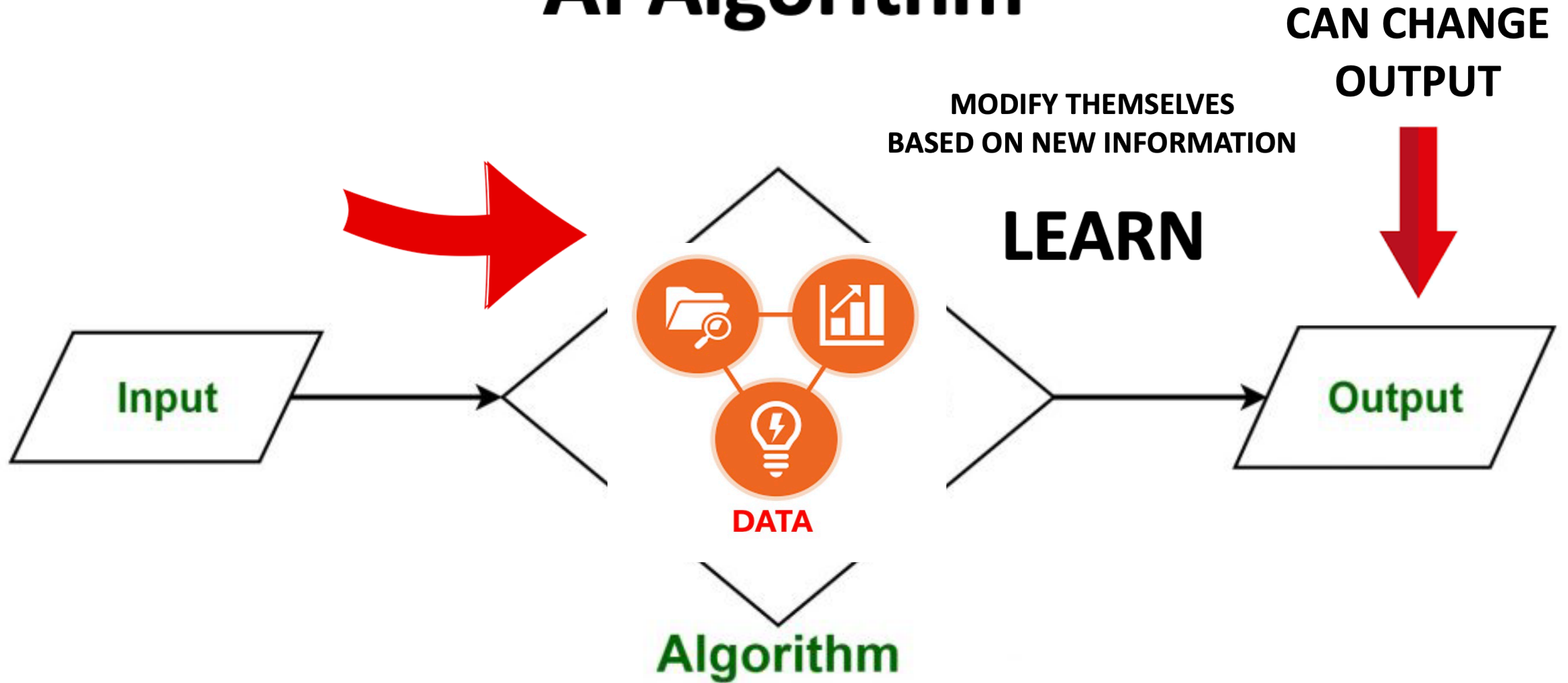
AI Algorithm



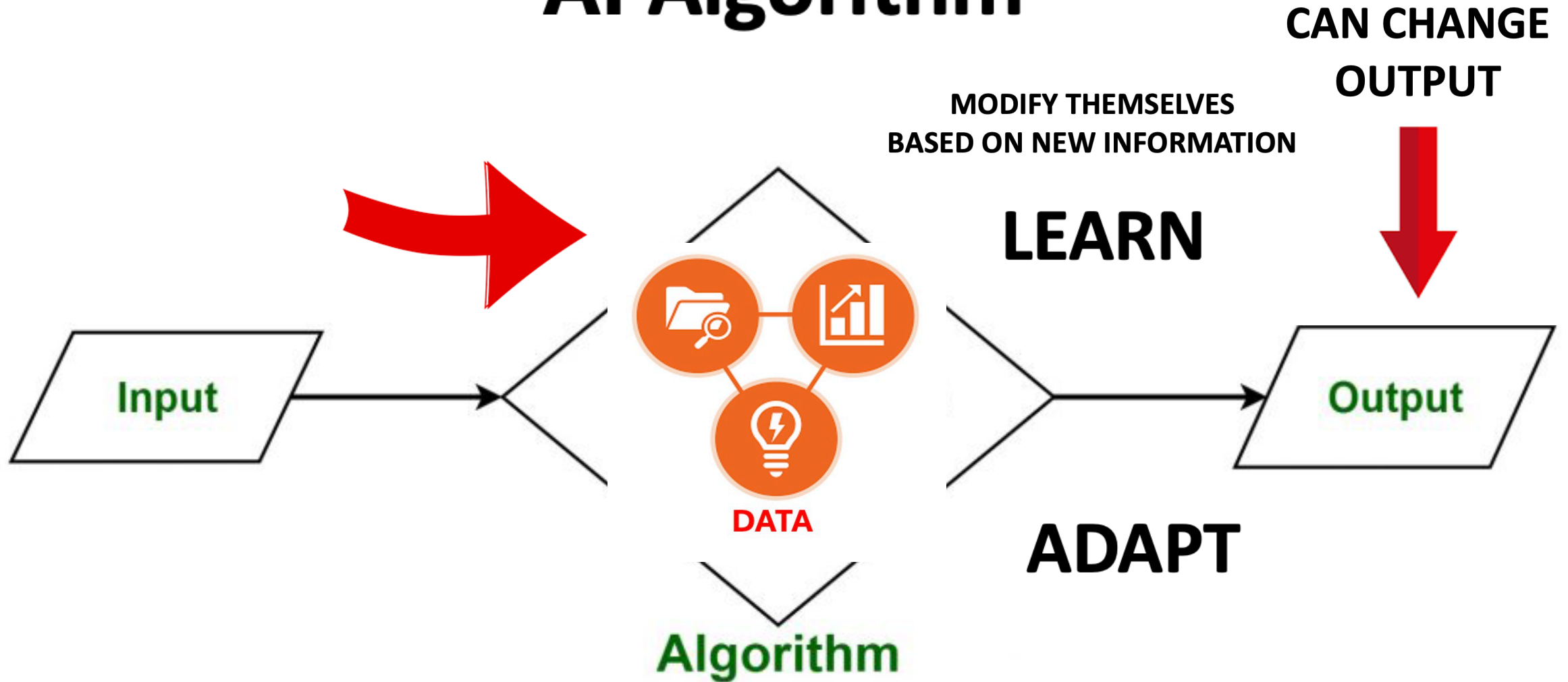
AI Algorithm



AI Algorithm



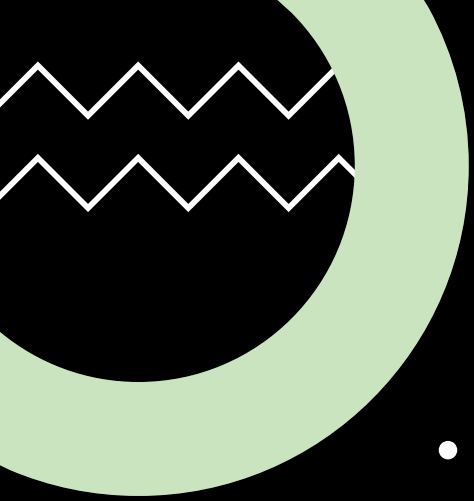
AI Algorithm



AI May Learn Incorrect Patterns -Hallucinating-



- Among the most serious concerns related to AI generated algorithms and their outputs is the lack of proper data evaluation.
- AI are trained on data, and they learn to make algorithms which in turn make predictions by finding patterns in the data.
- However, if the training data is incomplete or biased, the AI may learn incorrect patterns.
- This can lead to the AI model making incorrect predictions, or “hallucinating”.

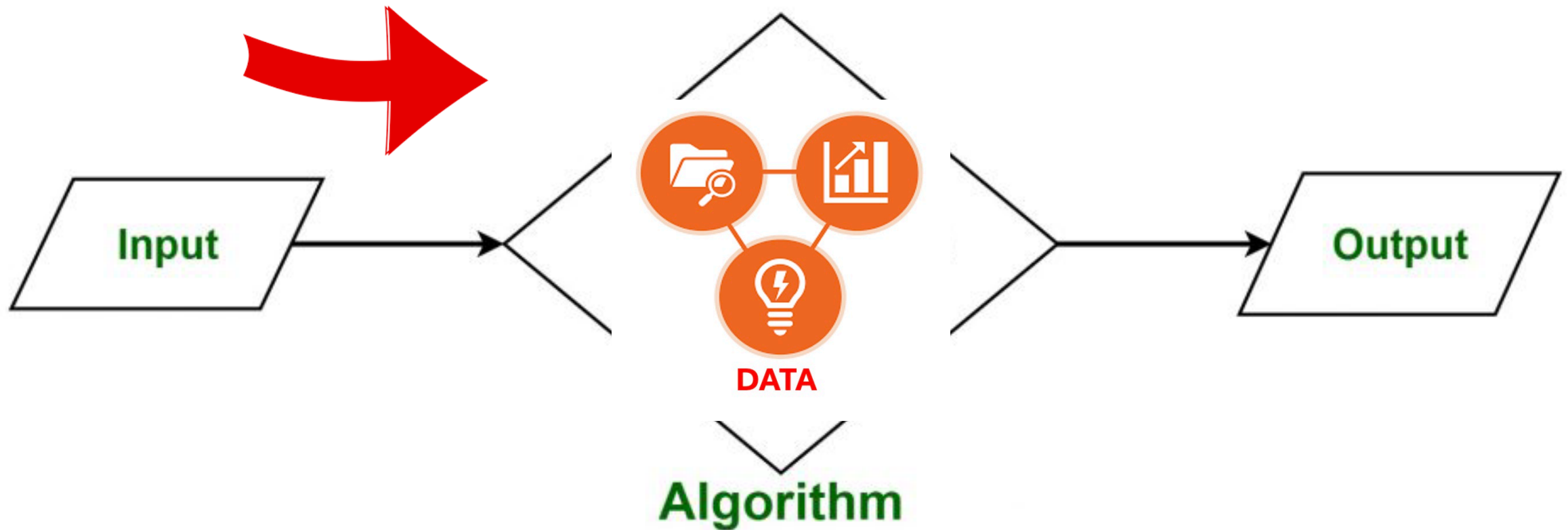


- **AI hallucinations are incorrect or misleading results that AI models generate (by their own internal processing). These errors can be caused by a variety of factors, including insufficient training data, incorrect assumptions made by the model, or biases in the data used to train the model.**

- ***aviation case***



AI ALGORITHM [HALLUCINATIONS]



AI ALGORITHM [HALLUCINATIONS]





Malfunctioning Algorithms due to AI Spoofing

- Currently, AI which relies on Internet data for learning may develop malfunctioning algorithms due to spoofing.
- Spoofing is the act of disguising the source of an Internet communication from an unknown source as being from a known, trusted source.
- Spoofing may first lead to erroneous AI machine learning, subsequently to faulty algorithms and finally untrustworthy AI evidence.



[HOME](#)[BUY CREDITS](#)[FEATURES](#)[MOBILE APPS](#)[MEDIA](#)[HELP](#)[SIGN UP](#)[LOGIN](#)

Disguise your Caller ID

Display a different number to protect yourself or pull a prank on a friend. It's easy to use and works on any phone!

Get Spoofing! They'll never know it was you.

[TRY A LIVE DEMO](#)

OR

[GET STARTED NOW](#)

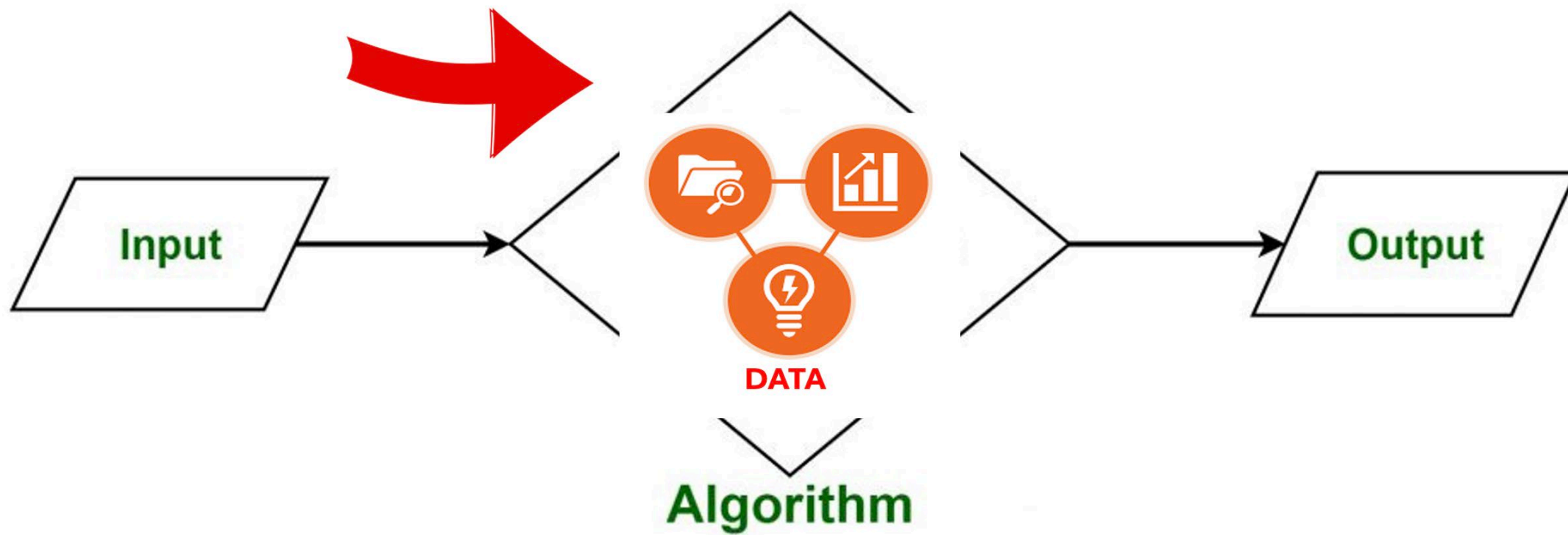


Malfunctioning Algorithms due to AI Spoofing

- Currently, AI which relies on Internet data for learning may develop malfunctioning algorithms due to spoofing.
- Spoofing is the act of disguising the source of an Internet communication from an unknown source as being from a known, trusted source.
- Spoofing may first lead to erroneous AI machine learning, subsequently to faulty algorithms and finally untrustworthy AI evidence.



Adversarial AI Use [Spoofing]



AI Spoofing is the act of disguising the source of an Internet communication from an unknown source as being from a known, trusted source.





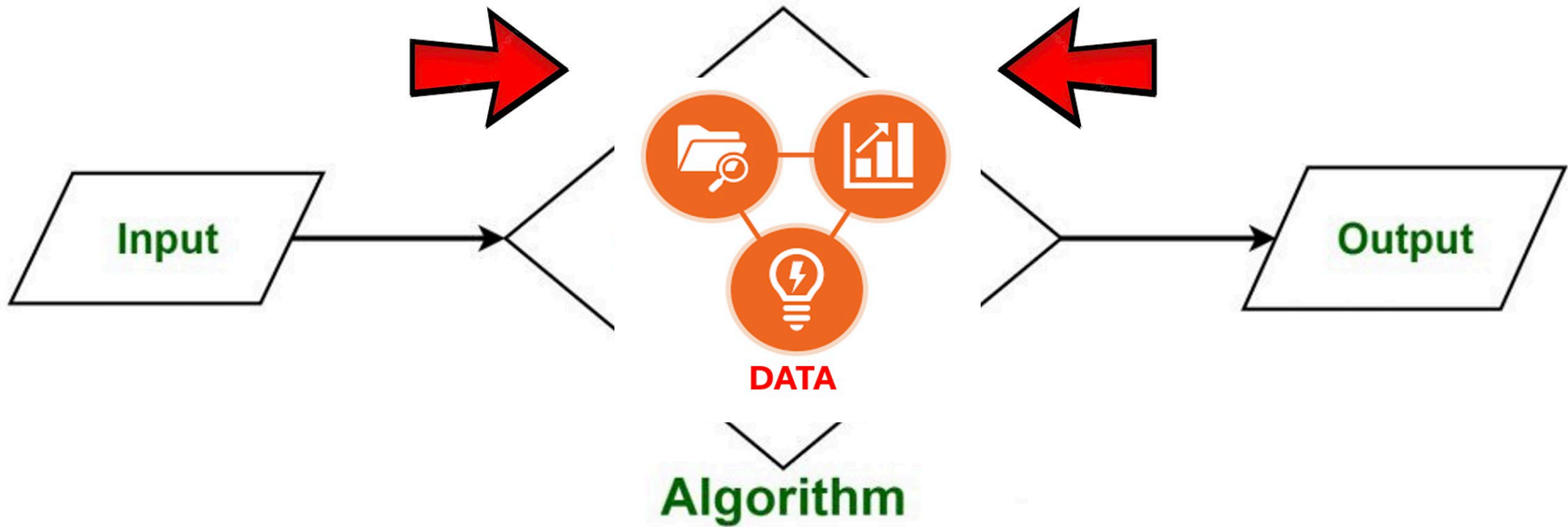
Adversarial AI Use [AI's corrupting other AI's – “getting poisoned”]

- It is increasingly difficult to distinguish human generated Internet content from AI generated Internet content.
- Consequently, AI's have been used to corrupt other AI's.
- This practice known as adversarial AI uses AI to fool machine-learning models by supplying deceptive input(s).
- Adversarial AI can be used to modify the output of most AI technology.



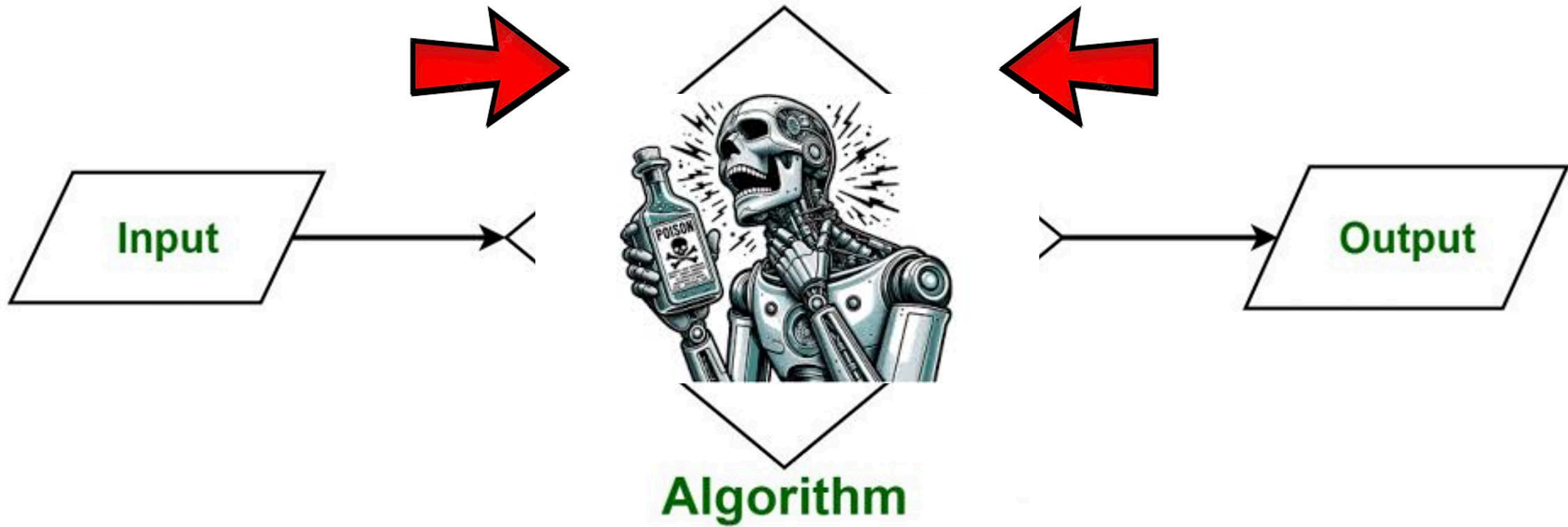
Adversarial AI Use

[AI's corrupting other AI's]



Adversarial AI Use

[AI's corrupting other AI's]



Resilience [Admissibility Analysis]

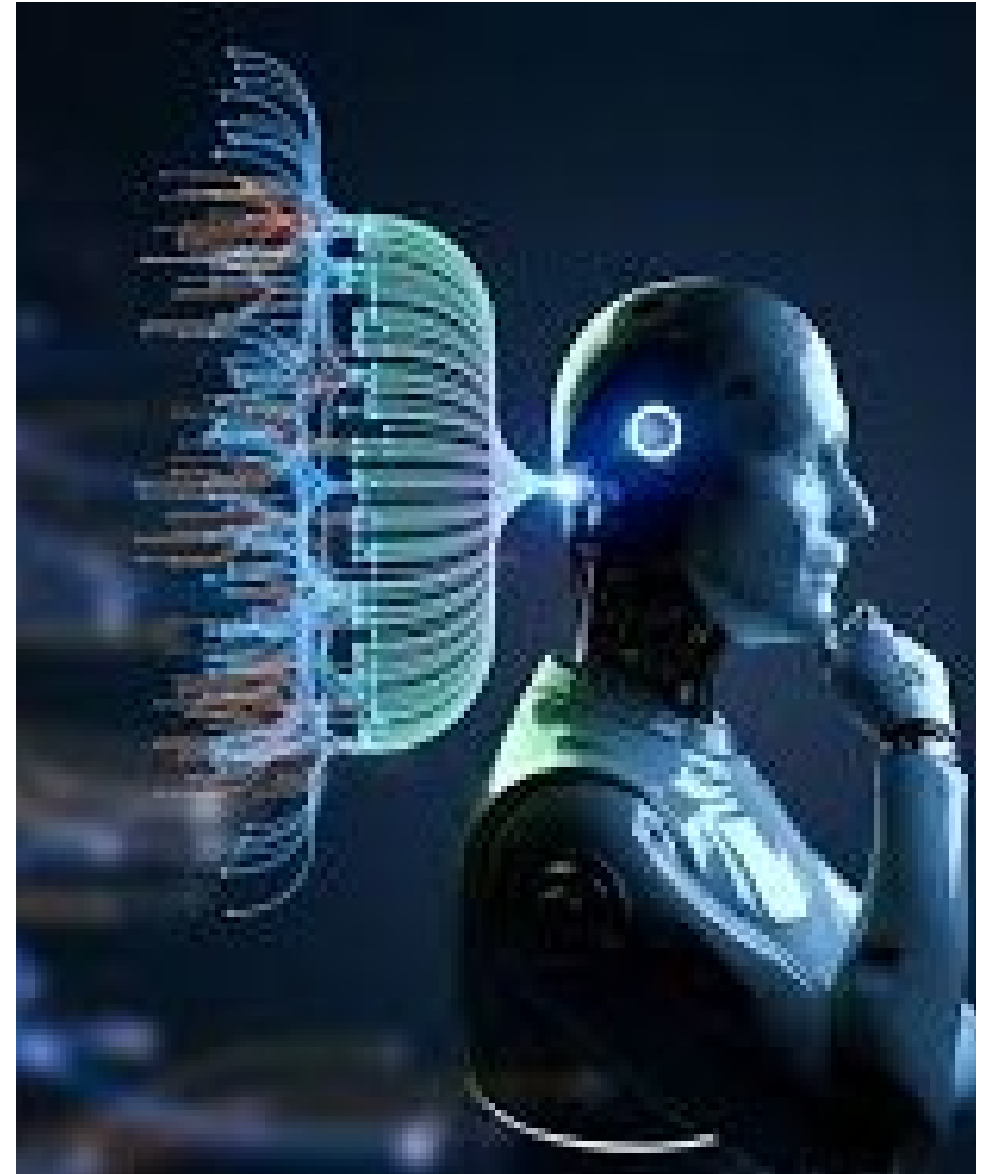
One challenge for AI evidence admissibility is resilience.

Resilience is the degree to which an AI resists both intentional and unintentional efforts to cause machine-learning models to fail.



Transparency & Explainability [Admissibility Analysis]

- In considering admissibility, courts should require transparency and explainability:
 - in terms of how the AI system works
 - how the AI system reached its decision
 - How the AI system reached its classification
 - How the AI system reached its prediction/conclusion



Explanations Must Reveal Inner Workings



AI evidence admissibility must provide explanations that reveal their inner workings and how the AI amends its algorithms such that AI algorithms are explainable.

Information Underlying the AI

- AI evidence for either civil or criminal trials, should not be permitted if the information underlying the AI is not available.
- Such information must be sufficient for the party against whom that evidence will be offered to determine the validity (including the accuracy of the AI) and the reliability (i.e., the AI algorithm correctly measures what it purports to measure).



intel.

FakeCatcher

the world's first real-time deepfake detector

Pioneered by Intel, the FakeCatcher deepfake detector analyzes "blood flow" in video pixels to determine a video's authenticity in milliseconds.

What is a deepfake?

Deepfakes are synthetic videos, images, or audio clips where the actor or the action of the actor is not real.

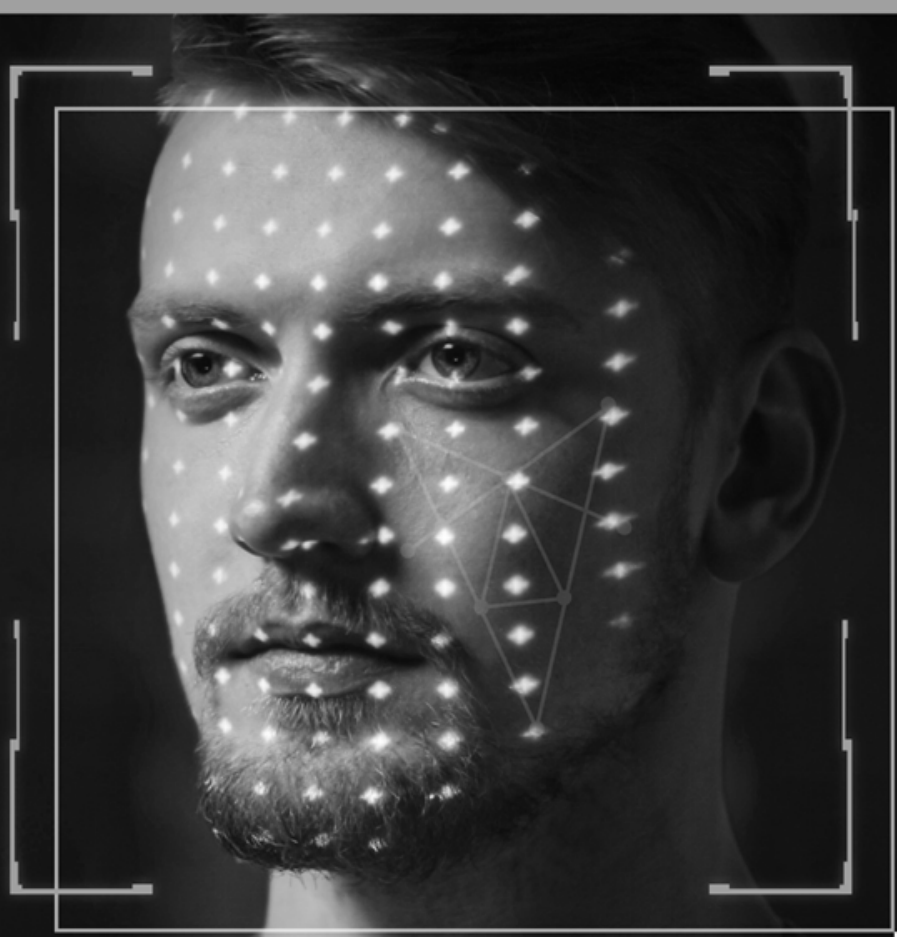
FakeCatcher can run up to

72 concurrent

real-time deepfake detection streams on 3rd Gen Intel® Xeon® Scalable processors

96%

FakeCatcher's accuracy for deepfake detection



FakeCatcher is the first deepfake detection algorithm that uses heart rates.

How could FakeCatcher be used?



Content Creation Tools

Integrated detection in editing software used by content creators and broadcasters.



Social Media

Detection as part of a screening process on user-generated content.



Media & Broadcasters

Detection in news video footage, particularly background (b-roll) sourced from third parties.



AI for Social Good

Democratized deepfake detection via a common platform, enabling any person or entity to confirm the authenticity of a video.



Authentication

TRE 901

TRE 901(a) requires AI evidence to be authenticated prior to consideration by the finder of fact.

TRE 901(b) discloses a variety of ways in which a party can achieve this objective. No special exception is made for AI evidence.

A witness with knowledge of the AI (underlying data/algorithms) needs to provide admissible evidence [testimony]:

- that the AI is what it claims to be (in accordance with TRE 901(a));
- (in accordance with TRE 901(b)) describe the process or system;
- (in accordance TRE 901(b)) show that the process or system described produces an accurate result.

Since AI programming is not common knowledge, it is expected that TRE 602 will apply requiring the authenticating witness to have personal knowledge of how the AI technology functions or be established as an expert.

AI Authentication TRE 901 & 602

AI Authentication

TRE 901 & 602



AI usually requires both machine learning and generative elements.



It is therefore unlikely, due to the multiple skill/knowledge sets required, that a single witness will be sufficient for admissibility purposes.



The AI machine (that is AI in its broadest sense) normally requires one set of skills to teach a computer to understand certain data and perform certain tasks.



Generative AI (the kind of AI you can use to create new text, visual, and audio content) normally requires one set of skills to build on that foundation and adds new capabilities that attempt to mimic human intelligence, creativity and autonomy.

Potential Relevance Analysis

TRE 401-403

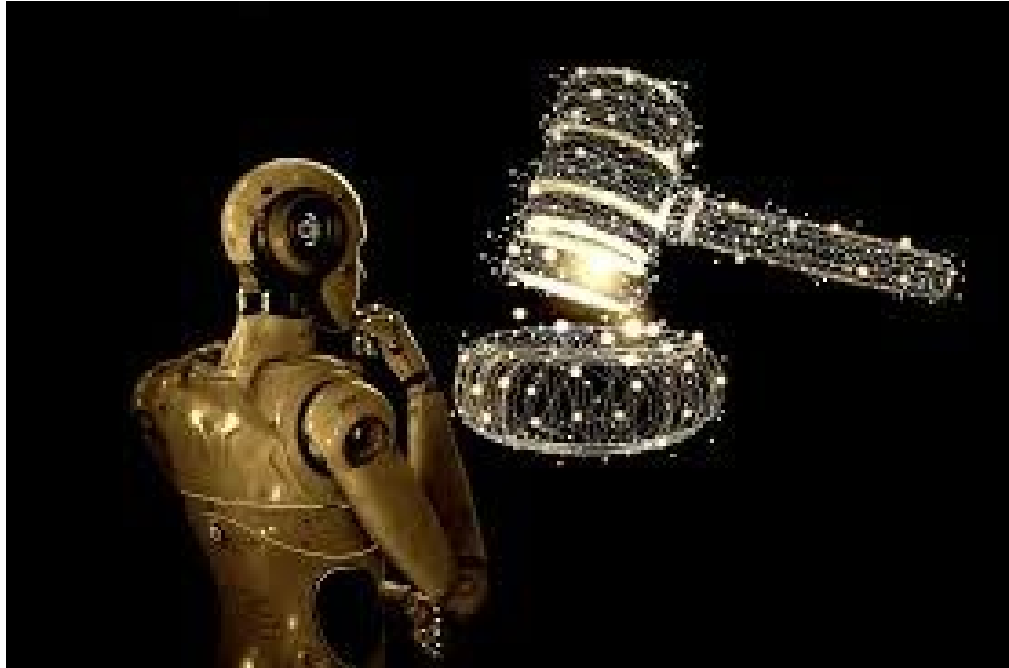
- TRE 401 indicates that evidence is relevant if
 - (a) it has any tendency to make a fact more or less probable than it would be without the evidence; &
 - (b) the fact is of consequence in determining the action.
- TRE 401 is normally read in conjunction with TRE 402 (general admissibility of relevant evidence) and TRE 403 (excluding relevant evidence for prejudice, confusion, or other reason)
- Rule 403 limits Rule 402 by excluding relevant evidence if its probative value is outweighed by prejudice, confusion, or waste of time.



Potential TRE 403 Applicability to AI Evidence

TRE 403 (judge is gatekeeper) could be translated to state that a judge cannot make the determinations unless the party offering the AI evidence is prepared to disclose underlying information. This would include the training data, as well as the development and operation of the AI system sufficient to allow the opposing party to challenge it.





Underlying Non-AI Information

- As in the case of non-AI information, the trial judge should give the proponent of the AI evidence a choice.
- The proponent may either disclose the underlying evidence (perhaps under an appropriate protective order) or otherwise demonstrate its validity and reliability.
- If the proponent is unwilling to do so, the AI evidence should not be admissible.

AI's Obvious Admissibility Issues

- Admissibility of AI evidence is likely to face objections due to the **lack of rigorous testing** because the AI algorithms that can have a significant impact on legal rights.
- Even when AI algorithm testing is performed, it is **rarely independent, peer-reviewed, or sufficiently transparent to be properly assessed** by those competent to do so.
- the standard requires the admissibility of computer-based evidence, such as AI evidence to be based on scientific methods that are sufficiently established and accepted. **Since there are no standards for AI algorithms testing generally nor AI product testing specifically, it will be difficult to have an expert opinion that the AI evidence is admissible because it is “generally accepted” as reliable in the relevant scientific community.**

Potential Examination Points/Authenticating AI

Data

- Q. What data was used to train the AI system?
- Q. How was data obtained?
- Q. Why was that data chosen?
- Q. Where did that data come from?
- Q. What features and weights were chosen for that data?
- Q. Why were those features and weights chosen?

Potential Examination Points/Authenticating AI

[Algorithms]

- Q. Who programmed the underlying algorithm?
- Q. How was that program developed?
- Q. Did the data collected affect how the AI system was programmed? How?
- Q. How does the program uses the data collected?
- Q. How does one use the AI system?
- Q. Does the AI system produce valid results?
- Q. What did you do to determine it produces valid results?

Potential Examination Points/Authenticating AI

[Algorithms]

- Q. Does the AI system produce reliable results?
- Q. What did you do to determine it produces reliable results?
- Q. What are the chances of error for the results produced?
- Q. Do you believe that the benefits of the results produced by the AI system outweigh the possible errors it could produce? Why?
- Q. Do you believe that the possible errors the AI system could produce would mislead anyone? Why?

Specify PDF or PPT

heather@koonfuller.com

