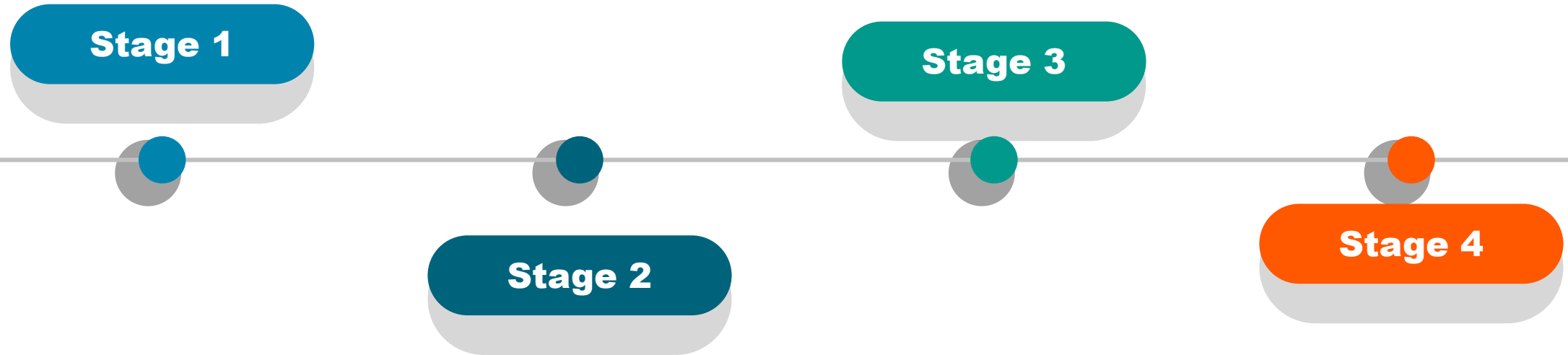




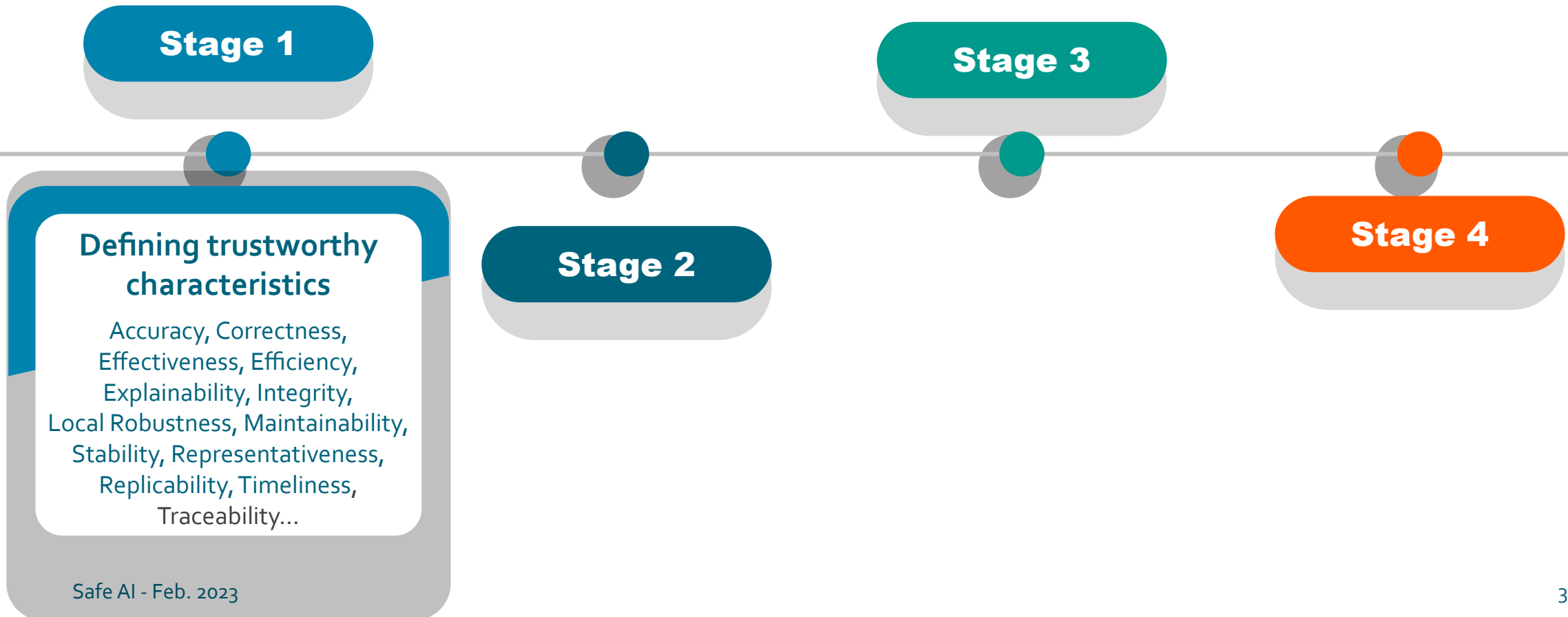
# Towards a holistic approach for AI trustworthiness assessment based upon aids for multi-criteria aggregation

Juliette MATTIOLI, Henri SOHIER,  
Agnès DELABORDE, Gabriel PEDROZA,  
Kahina AMOKRANE-FERKA, Afef  
AWADID, Zakaria CHIHANI and  
**Souhaïel KHALFAOUI**

# How does Confiance.ai assess ML trustworthiness?



# How does Confiance.ai assess ML trustworthiness?



# How does Confiance.ai assess ML trustworthiness?

## Stage 1

### Defining trustworthy characteristics

Accuracy, Correctness, Effectiveness, Efficiency, Explainability, Integrity, Local Robustness, Maintainability, Stability, Representativeness, Replicability, Timeliness, Traceability...

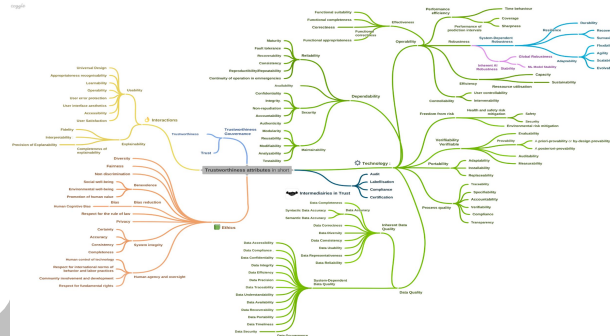
## Structuring the problem

Clustering characteristics to meta attributes into a semantic tree:  
Data Quality, Dependability, Operability Robustness, Human centred quality, Human agency & oversight

## Stage 3

## Stage 4

## Stage 2



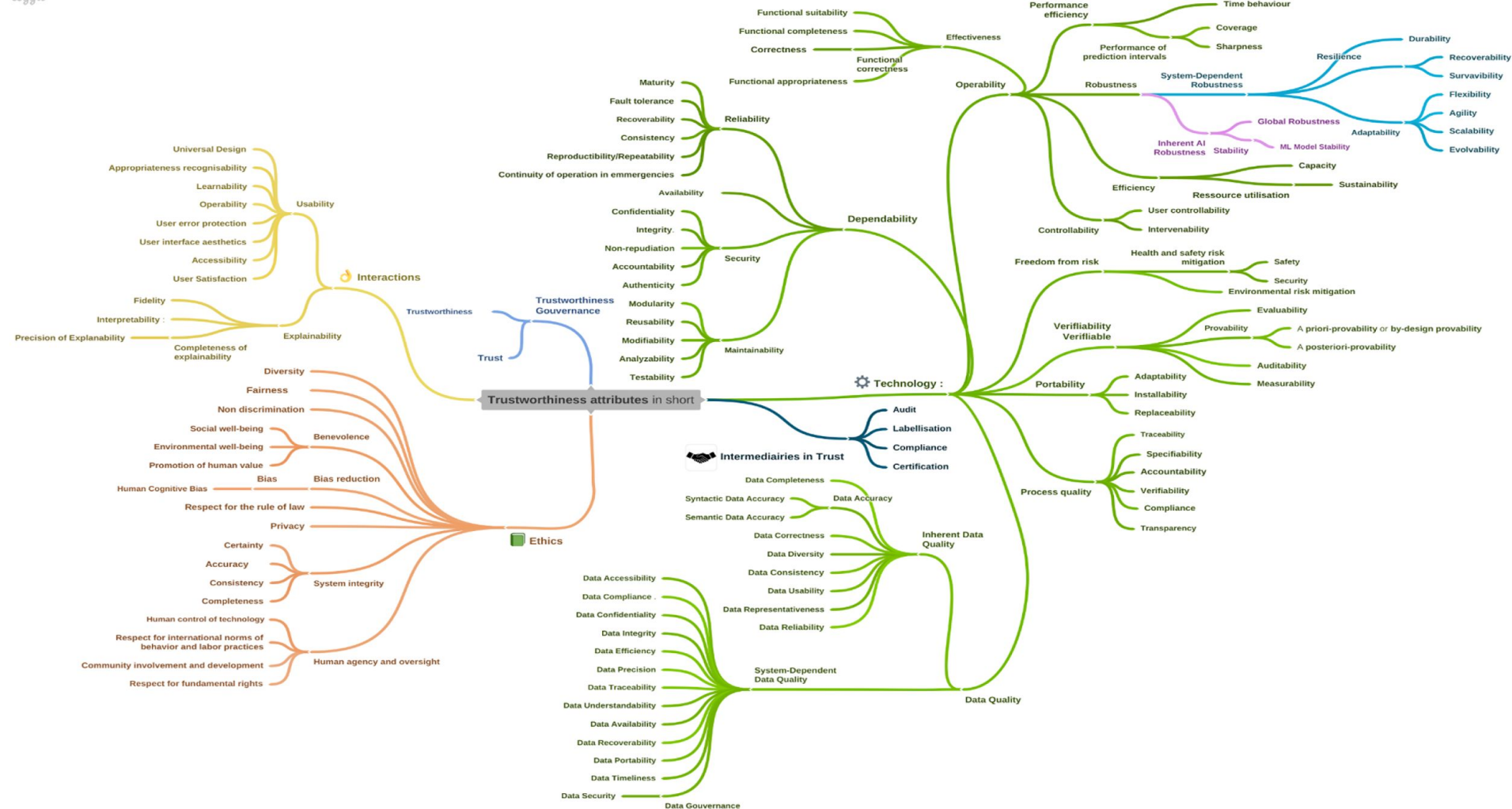
# How does Confiance.ai assess ML trustworthiness?

## Structuring the problem

Clustering characteristics to meta attributes into a semantic tree:  
 Data Quality, Dependability, Operability, Robustness, Human centred quality, Human agency & oversight

## Stage 2

coggle



# How does Confiance.ai assess ML trustworthiness?

## Stage 1

### Defining trustworthy characteristics

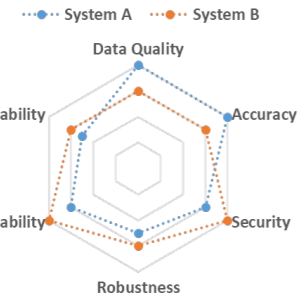
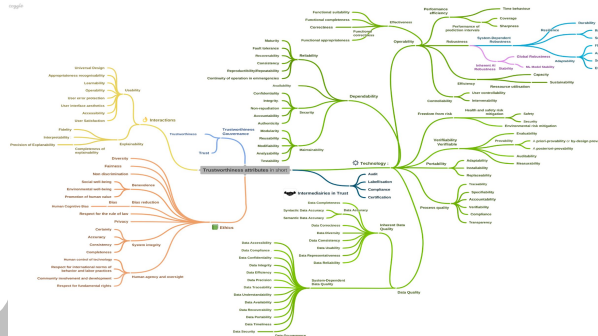
Accuracy, Correctness, Effectiveness, Efficiency, Explainability, Integrity, Local Robustness, Maintainability, Stability, Representativeness, Replicability, Timeliness, Traceability...

Safe AI - Feb. 2023

## Structuring the problem

Clustering characteristics to meta attributes into a semantic tree:  
Data Quality, Dependability, Operability Robustness, Human centred quality, Human agency & oversight

## Stage 2



## Stage 3

### Handling commensurability

To compare any numerical KPIs, metrics or evaluation methods of each attributes

## Stage 4

# How does Confiance.ai assess ML trustworthiness?

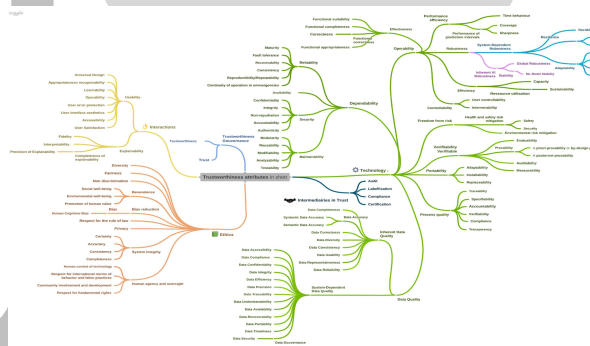
## Stage 1

### Defining trustworthy characteristics

Accuracy, Correctness, Effectiveness, Efficiency, Explainability, Integrity, Local Robustness, Maintainability, Stability, Representativeness, Replicability, Timeliness, Traceability...

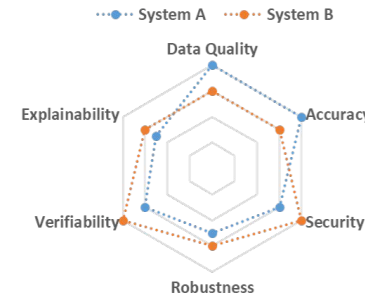
## Structuring the problem

Clustering characteristics to meta attributes into a semantic tree:  
Data Quality, Dependability, Operability Robustness, Human centred quality, Human agency & oversight



## Stage 2

## Stage 3



To compare any numerical KPIs, metrics or evaluation methods of each attributes

## Assessing AI trustworthiness

Multi-criteria aggregation methodology to capture operational trade-offs and evaluate higher-level attributes

## Stage 4

