FRAZER-NASH CONSULTANCY
A KBR COMPANY

Building Confidence and Trustworthiness in Data Science Models

Neil Headings

Frazer-Nash Consultancy

4th June 2024

NOT PROTECTIVELY MARKED

SYSTEMS • ENGINEERING • TECHNOLOGY

Agenda

- About Frazer-Nash & Me
- Background to confidence, trustworthiness and uncertainty
- Confidence, trustworthiness and uncertainty in models
- Confidence, trustworthiness and uncertainty in data driven models

Frazer-Nash Consultancy – Our purpose

Frazer-Nash is a leading systems, engineering and technology company, with over 1500 employees in the UK and Australia.

We help organisations deliver innovative engineering and technology solutions to make lives safe, secure, sustainable, and affordable.

SYSTEMS · ENGINEERING · TECHNOLOGY

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Our market sectors

Aerospace and transport

- Aerospace
- Automotive
- Gas turbines
- Marine
- Rail and metro
- Space

Defence

- Air systems
- Surface ships
- C4ISTAR
- Weapons systems
- Defence facilities
- Land systems
- Submarines

Nuclear

- Advanced Nuclear Technologies
- Decommissioning
- Defence
- Fusion
- New build

Innovation and commerce

- Design services
- Due diligence
- Expert witness
- Healthcare and MedTech
- Software development

Security and resilience

- Commercial
- National defence
- Organisational resilience
- Critical national infrastructure
- Insider threat

Power and energy

- Gas turbines
- Oil and gas
- Power generation
- Power, transmission and distribution
- Renewable energy

Data Science: Our Capabilities

Data strategy and cloud engineering

Development of enterprise and data strategies. Deployment of cloud infrastructure for data science and app development.

Software dev

Producing engineering software using robust, accredited quality assured processes with modern practices.

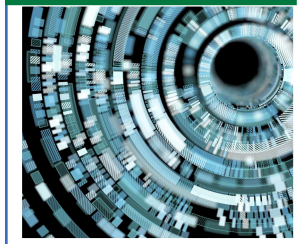
System & physical modelling

Simulation models to help optimise and understand complex processes and physical systems.

Data Analytics

Providing better insights from your existing data in an engineering context and its uncertainty.

AI and ML

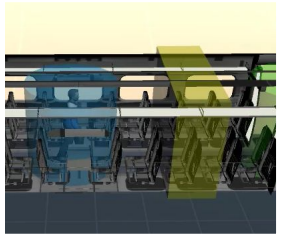


Use modern data science toolsets and capabilities that support intelligent engineering decision making in data driven environments.

Visualisation

Web-based applications and user-interfaces to explore data and models.

Simulated environments

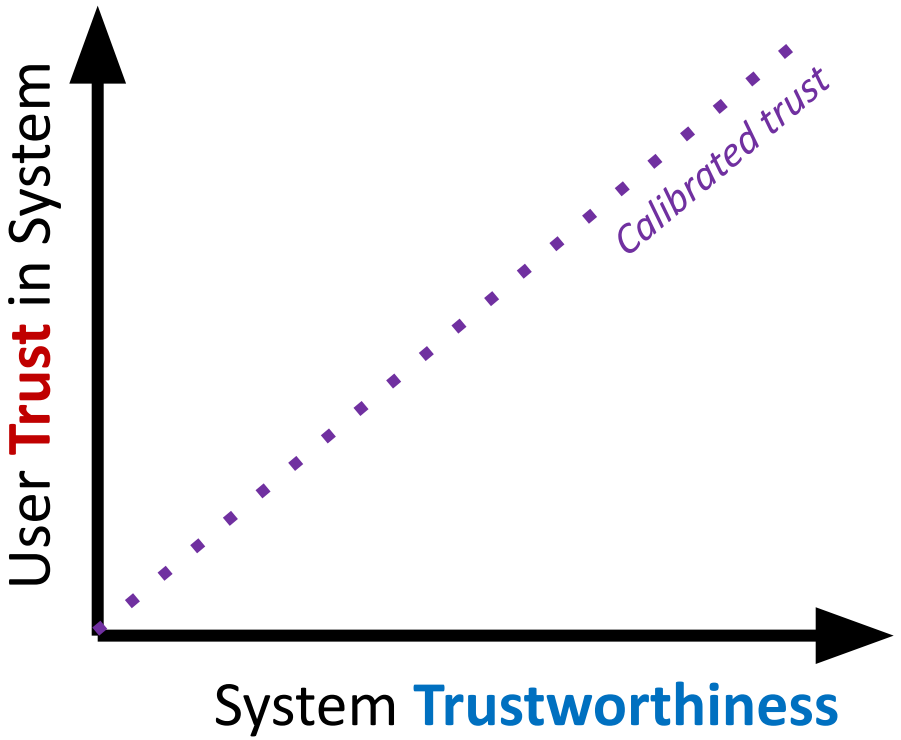


Realistic 3D environments for modelling, testing and evaluation.

Background to Confidence



Trust and Trustworthiness



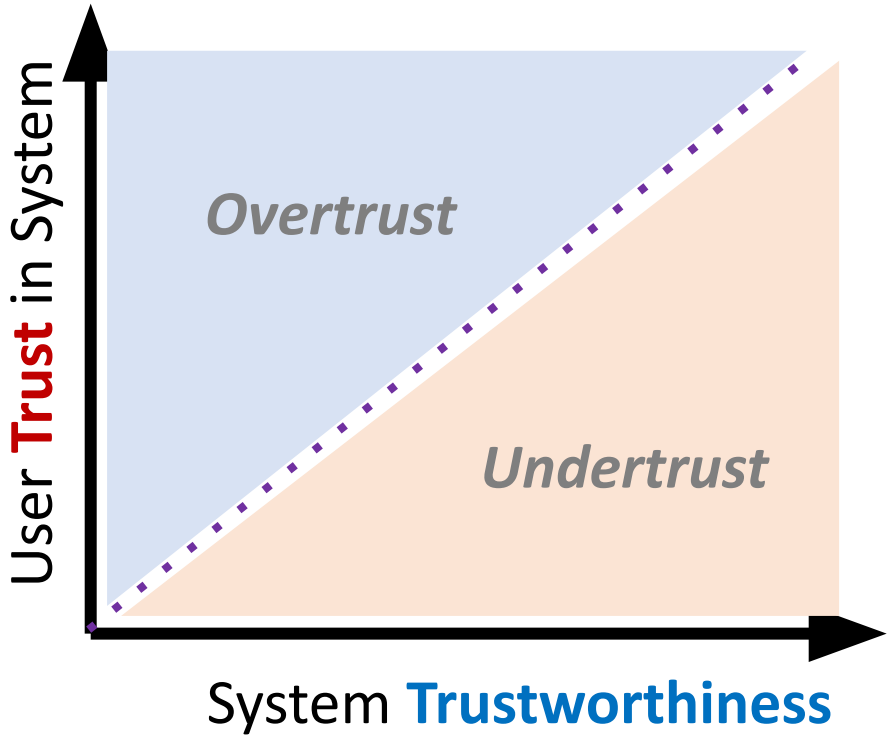
Trust = response of a user in a situation of uncertainty or vulnerability. ***Subjective***

Trustworthiness = measure of trust qualities in a system (autonomous or AI). ***Objective***

User **Trust** must be commensurate with the **Trustworthiness** of the system (well calibrated)

Sullins, J. P. (2020). Trust in robots. *The Routledge Handbook of Trust and Philosophy*, 313–225.

Trust and Trustworthiness



When trust is uncalibrated or miss-calibrated:

Overtrust. Trust in the system is greater than the system can deliver:

- Over-reliance on AI/automation
- Taking inappropriate or misguided action

Undertrust. System performs better than supervisor allows for:

- Supervisor ‘knows better’
- Taking alternative, contrary or abortive action

What is trustworthy software?

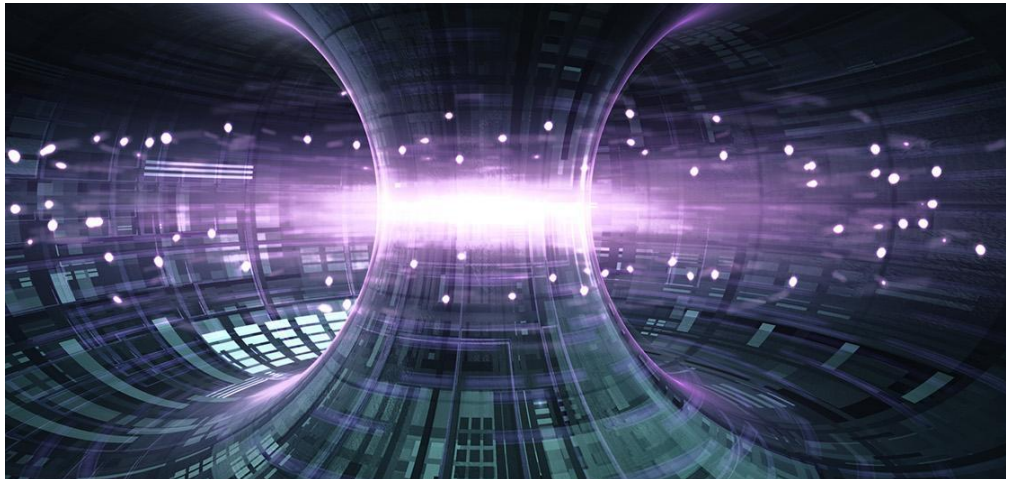
- **Safety**: The ability of the software to operate without causing harm to anything or anyone.
- **Reliability**: The ability of the software to operate correctly.
- **Availability**: The ability of the software to operate when required.
- **Resilience**: The ability of the software to recover from errors quickly and completely.
- **Security**: The ability of the software to remain protected against the hazards posed by malware, hackers or accidental misuse.

An operating mathematical model (be it physical or data driven) used to make decisions must be trustworthy.

The level of trust (and the scope of the activities performed) must be commensurate with the purpose of the system.

http://www.tsfdn.org/ts-framework/

Why Confidence and Trustworthiness?



Example

The Washington Post

Teslas running Autopilot involved in 273 crashes reported since last year

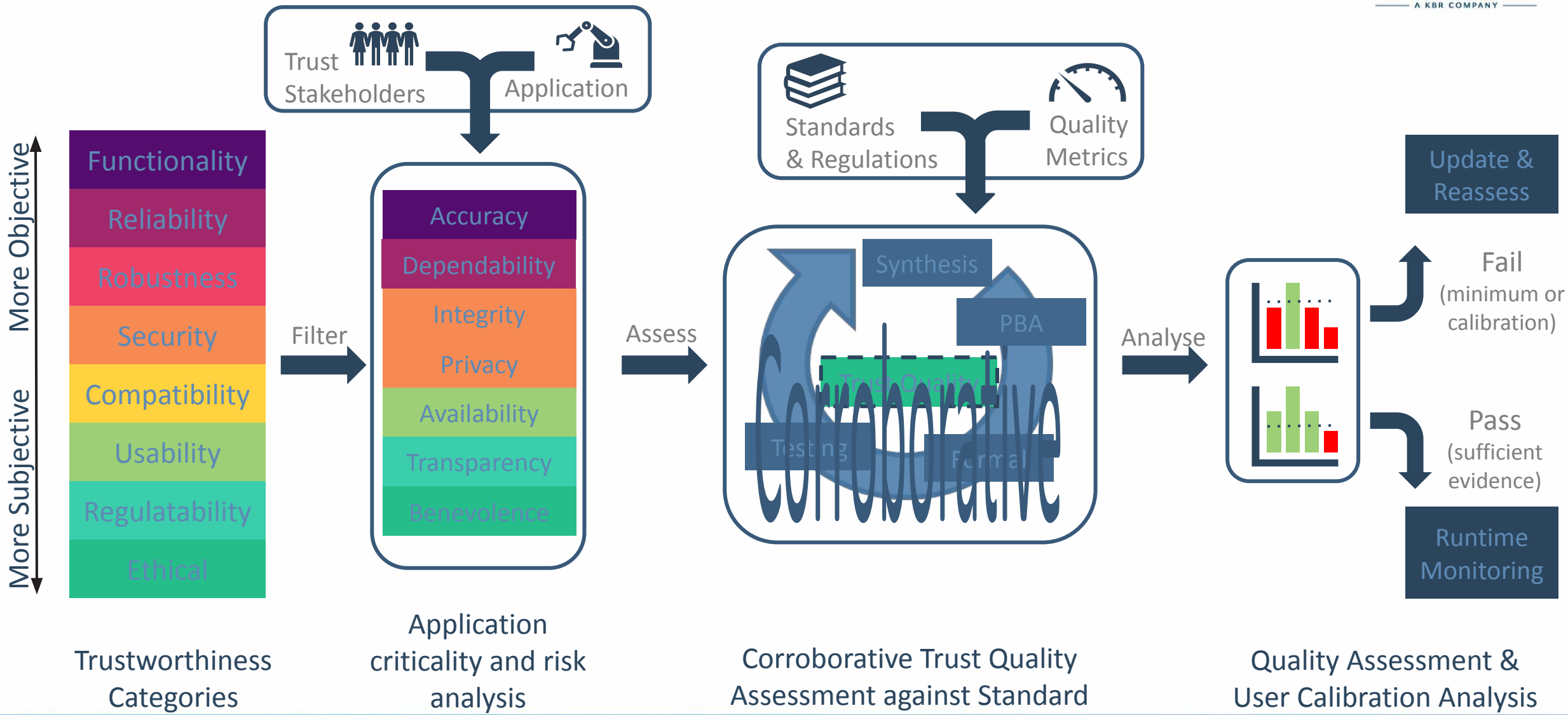
The Guardian

UK data watchdog investigates whether AI systems show racial bias

BBC

ChatGPT: US lawyer admits using AI for case research

Assuring trustworthiness in a system



https://arxiv.org/abs/2305.03411v2

Confidence and Uncertainty in Modelling

- A model describes our belief about how the world functions.
- A mathematical model represent those beliefs in terms of a set of mathematical equations. These could be based on physical rules or statistical inference from measured data (or both).
- All mathematical models will contain compromises and simplifications.

- Uncertainty represents an inability to state a definitively accurate and precise result for a modelling prediction.
- Uncertainty quantification is the process of establishing the range of, and likelihood for, values that a simulation prediction could cover, which should (with an estimate of confidence) encompass the ‘true’ (but unknown) real value that the physical system that has been modelled would exhibit.

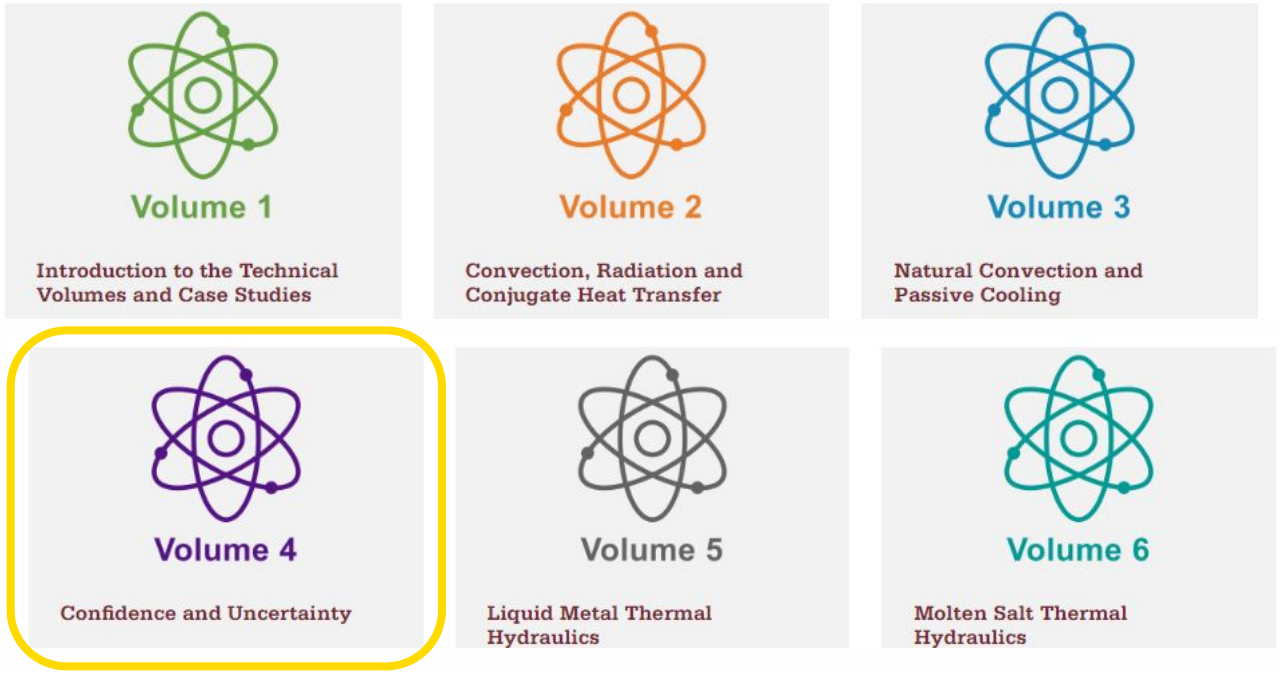
- In general for any modelling result we need to:
  - Acknowledge the answer we get can never be perfect.
  - Clearly define the context of our results.
  - Build confidence that the results are useful.
  - Communicate this clearly to the stakeholders.

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Confidence and Uncertainty in Modelling

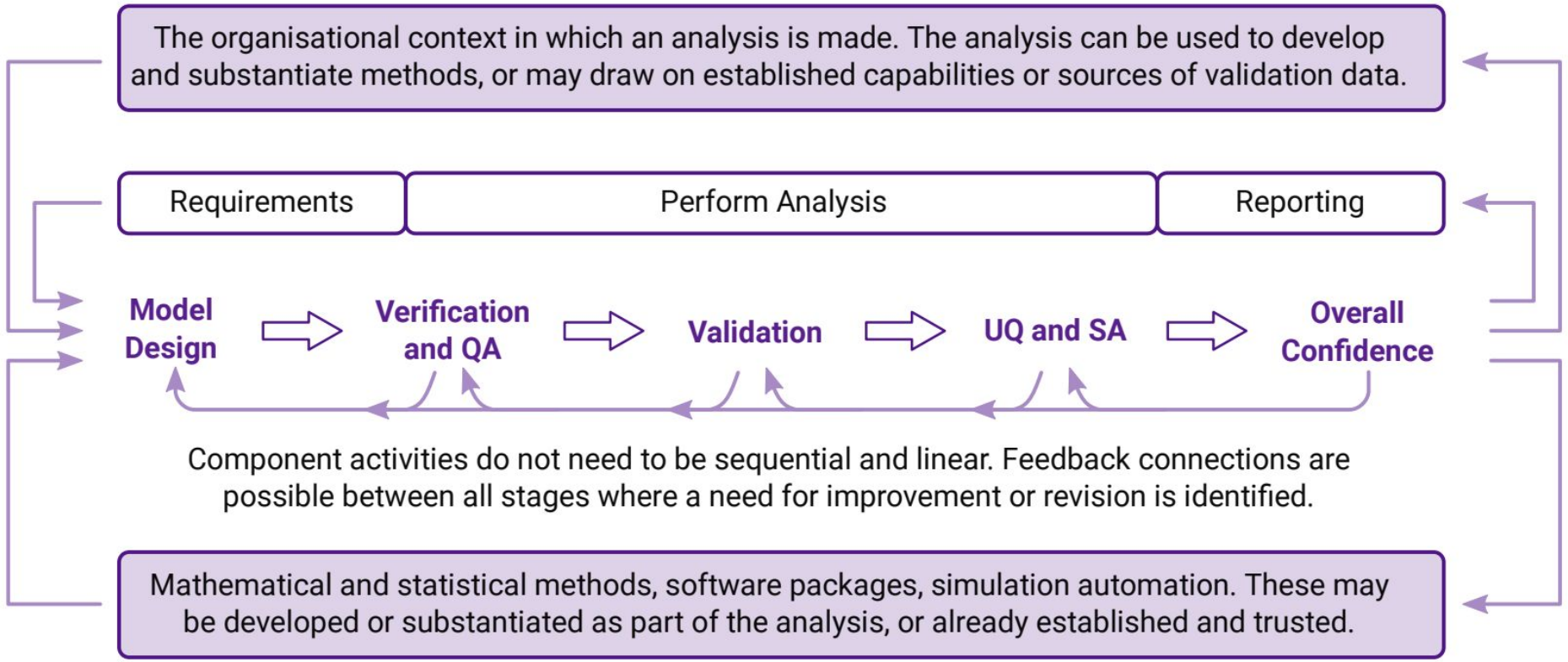


DIGITAL REACTOR DESIGN: NUCLEAR THERMAL HYDRAULICS



https://www.imeche.org/industry-sectors/power-energy/digital-reactor-design-nuclear-thermal-hydraulics

Activities to generate confidence in a model



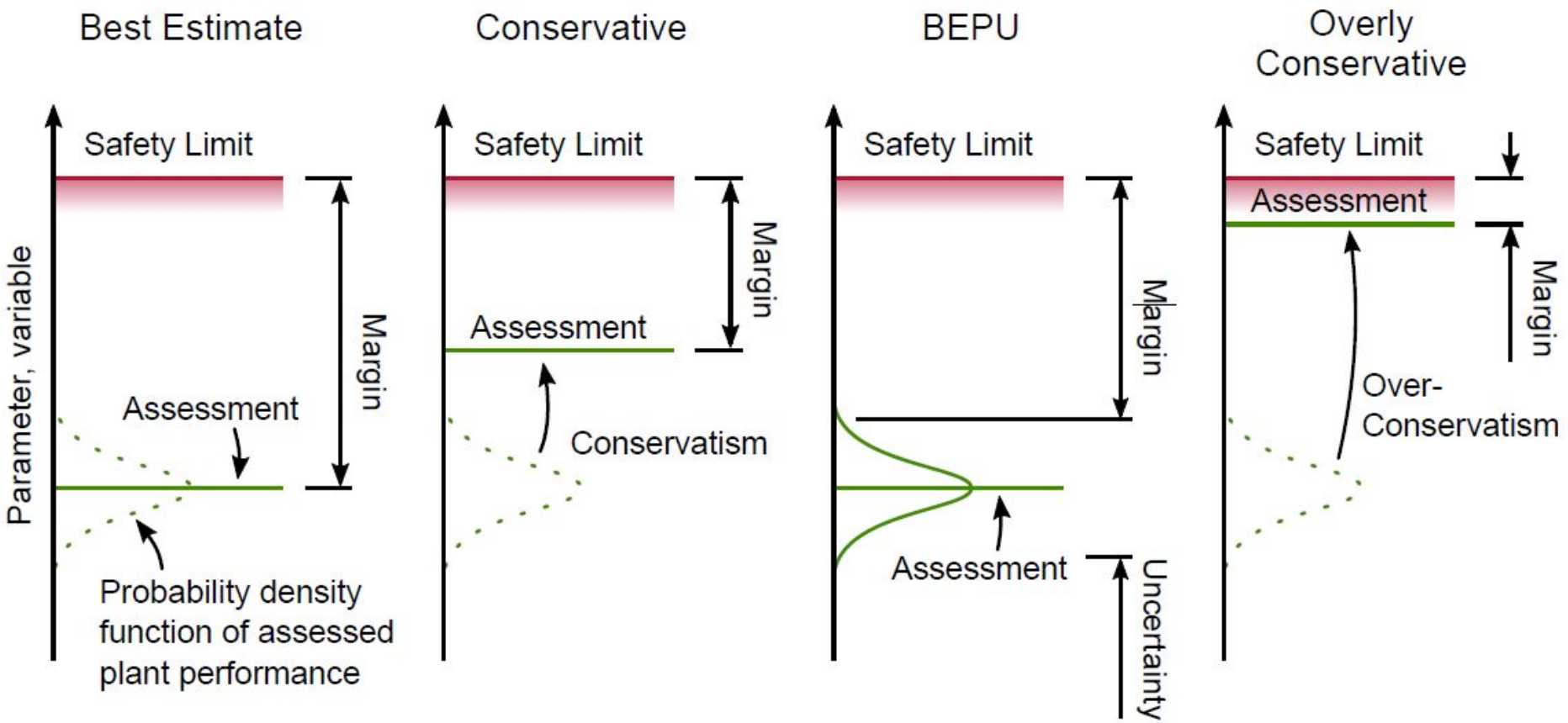
Verification: independent checking of the implementation to make sure it is doing what is intended

- What procedures for verification and QA does the organisation instructing the analysis or regulator require to be followed?
- What are the key things to check? How is this documented?
- Is the decision relying on the result so significant that an independent calculation is required?
- Is additional code verification (of the underlying model implementation) required?
- Is the analyst suitably experienced?
- What are the resource requirements for these activities, and are their availability, time and costs allowed for?

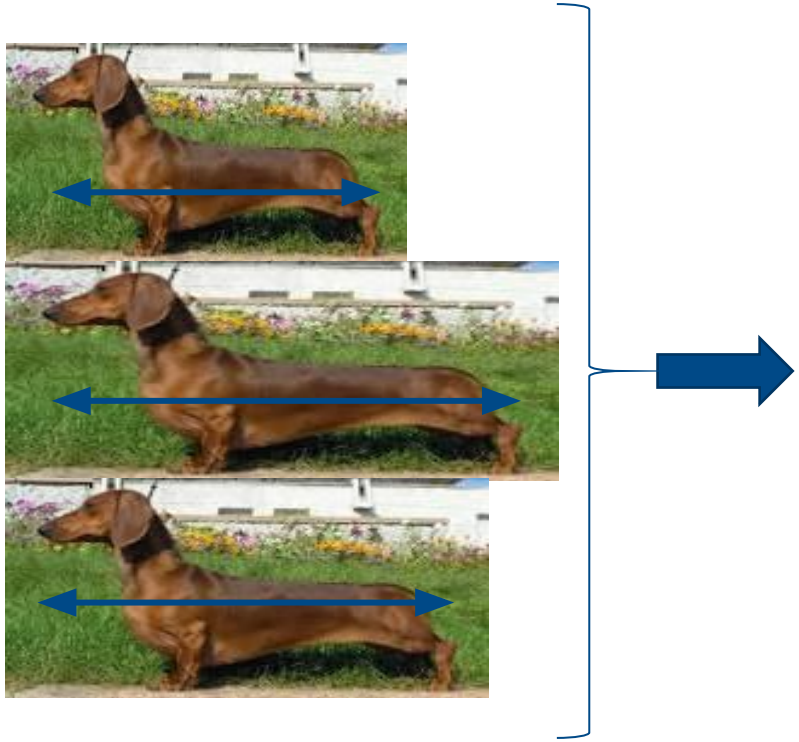
Validation: Comparison of model predictions against measurements, to provide confidence that simulation results represent reality with sufficient accuracy.

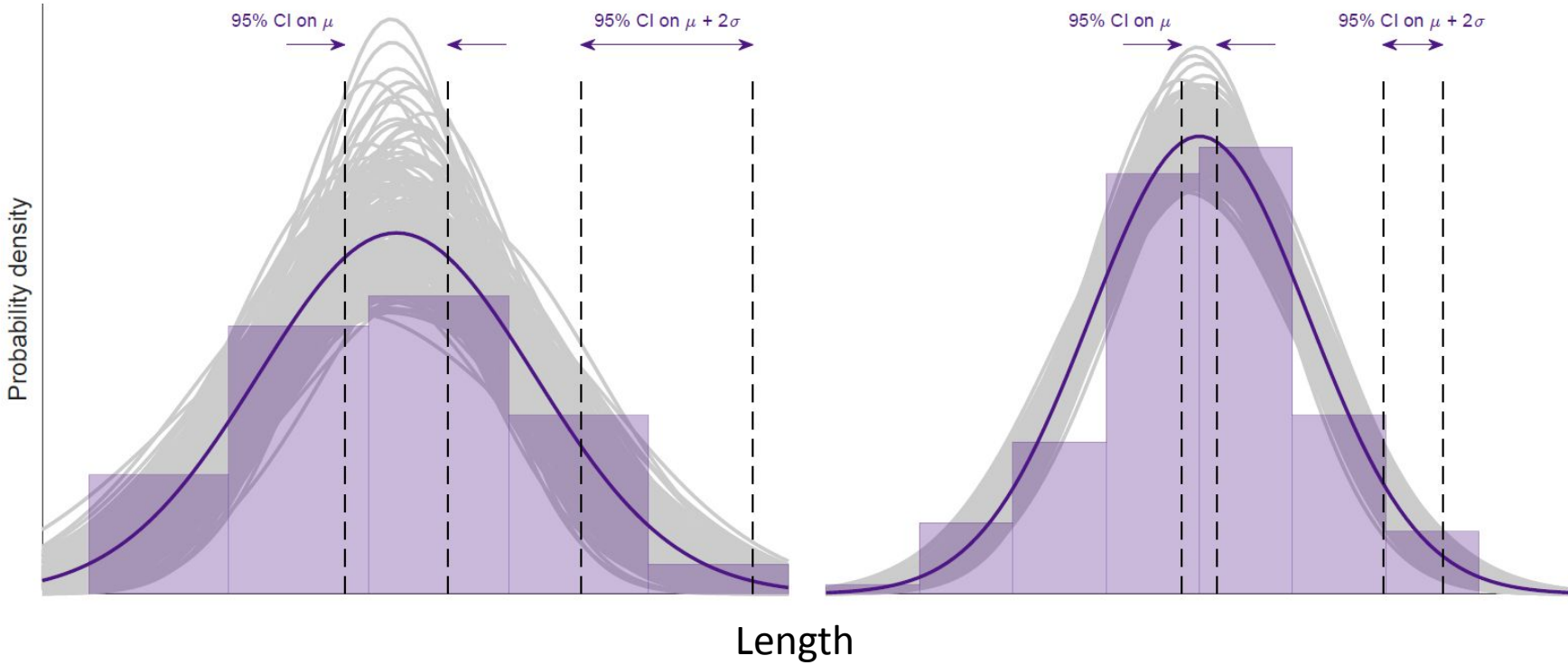
- What data is available to validate a model against?
- How representative is the data?
- What uncertainties are there in the data?
- Is the relevant region of p
- Will a part of the data be used to calibrate the model?
- What is the appropriate validation metric?

Assessing confidence in safety critical models



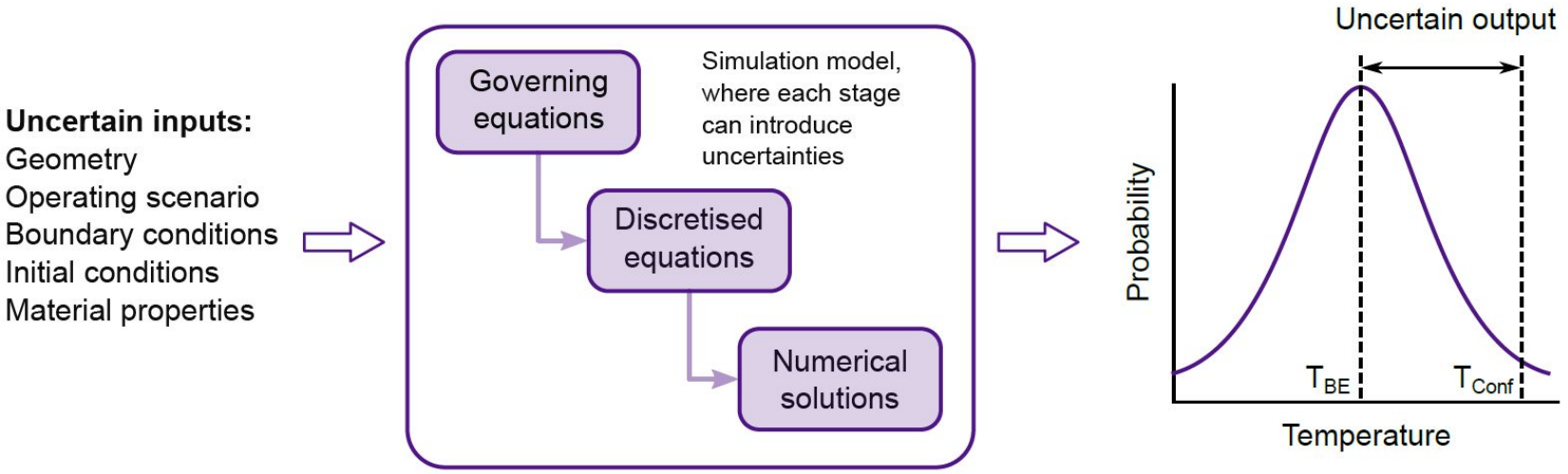
Aleatory and epistemic uncertainty





Overview of the aims of uncertainty quantification (UQ)

1. **Define your model – what is the output metric/quantity of interest.**
2. **Calculate uncertainty:**
   1. Define and characterise your key inputs which are uncertain.
   2. Propagate your uncertain inputs through the model.
   3. Calculate the overall output uncertainty.
3. **Understand it in context.**



Characterising input uncertainties

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Each uncertain input is assigned a mathematical structure and numerical parameters that describes the nature of its uncertainty

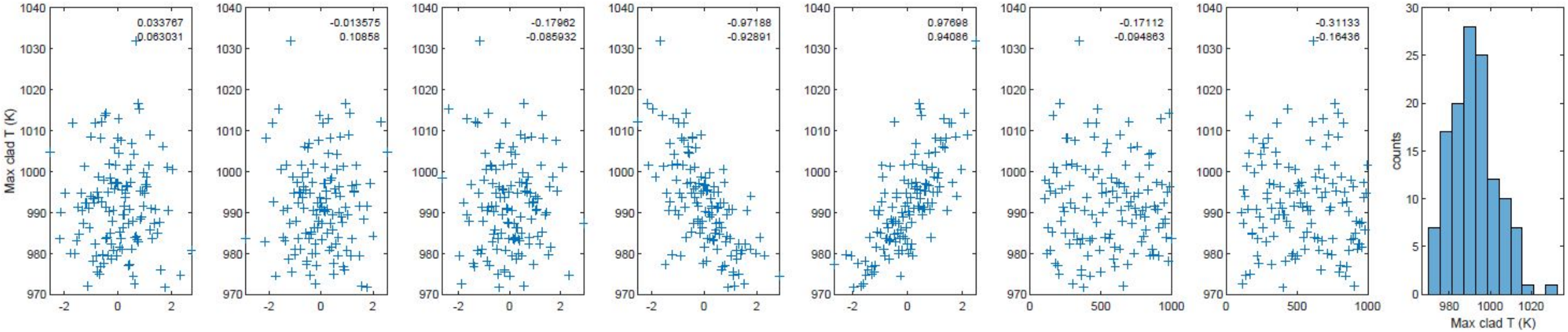
Where does the uncertainty originate?	**Is the uncertainty aleatory or epistemic?**	What is the impact on the complexity and computational cost of the uncertainty propagation?	Is there combined aleatory and epistemic uncertainty?

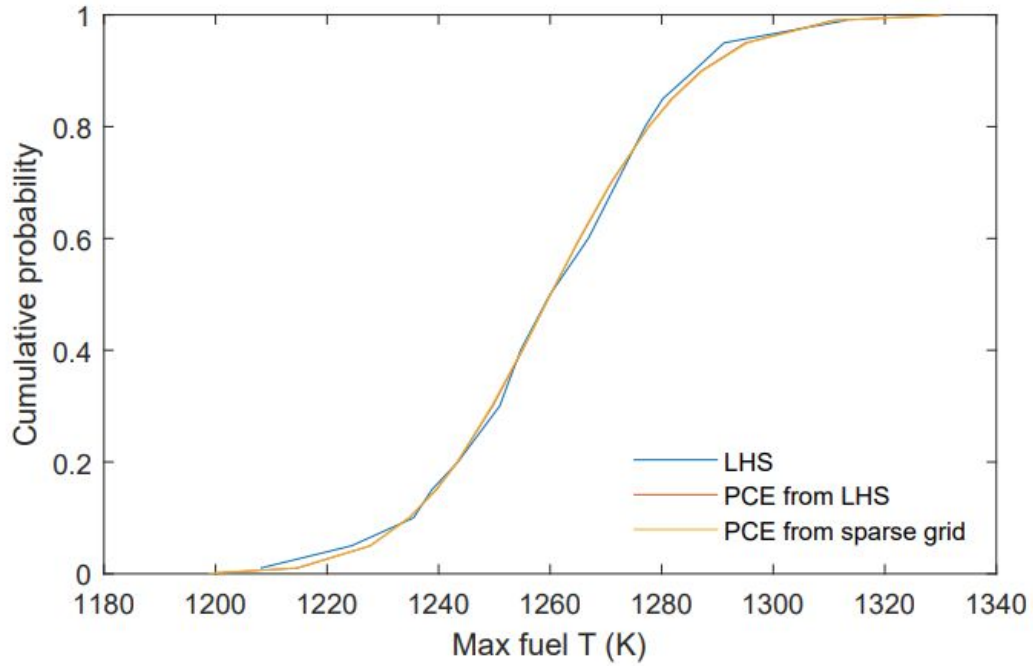
- Descriptive statistics
- Analytical methods
- Regression methods
- Maximum likelihood estimation
- Kernel density estimation

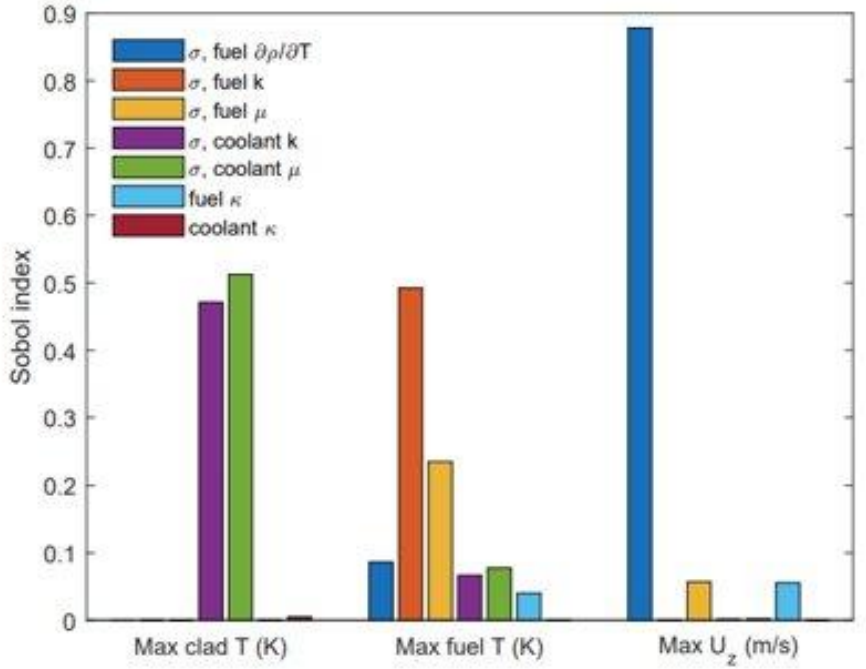
- Empirical cumulative distribution functions
- Bayesian inference
- Expert elicitation
- Bootstrapping and jack-knifing

- Interval ranges
- Categorical probabilities
- Probability bounds analysis
- Dempster-Shafer Theory

Case study: Fuel assembly peak clad temperature







Why is UQ Hard for In Practice?

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Maturity of methods

- Formal UQ and SA methods are often rooted in research work
- Have not been applied in real world engineering problems.
- The methods are less mature than techniques like CFD where benchmarked packages exist.
- The guidance provided in the ‘comprehensive frameworks’ for UQ is far from complete and pragmatic guide to their application.

Skill gap

- Formal mathematical foundations unapproachable to many practising engineers.
- There is a limited set of standardised tools and packages which practising engineers can use easily.
- Requires specialisation and a long term commitment within an organisation.
- UQ requires additional knowledge elements not typically taught to engineers.

Unrecognised benefit

- Performing UQ routinely can add cost and time
- The benefits of UQ are hard to recognise until after work has been commissioned.
- Customers of technical analysis are used to deterministic results from models.

Graham Macpherson, Neil Headings
Transactions of the ANS, Volume 126, Number 1, June 2022, Pages 340-343

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Confidence and Uncertainty in Data Driven Models



What does it mean for AI/ML models?

In developing AI/ML models and realising their benefit in wider society (and especially safety critical scenarios!) we need to have ***trustworthy AI***.

1. Lawful, complying with all applicable laws and regulations;
2. Ethical, ensuring adherence to ethical principles and values; and
3. Robust, both from a technical and social perspective,

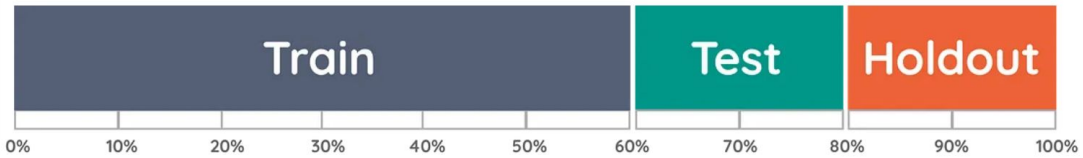
- *Respect for human autonomy*
- *Prevention of harm*
- *Fairness*
- *Explicability/Explainability*

How certain are we in the ML model results?...And does it matter?

Ethics guidelines for trustworthy AI, EU commission

“Classic” Data Science Validation

- Avoiding data leakage
- Hyper parameter tuning
- Test train split
- K-fold cross validation
- Nested k-fold cross validation



https://towardsdatascience.com/validating-your-machine-learning-model-25b4c8643fb7

Uncertainties in AI/ML models

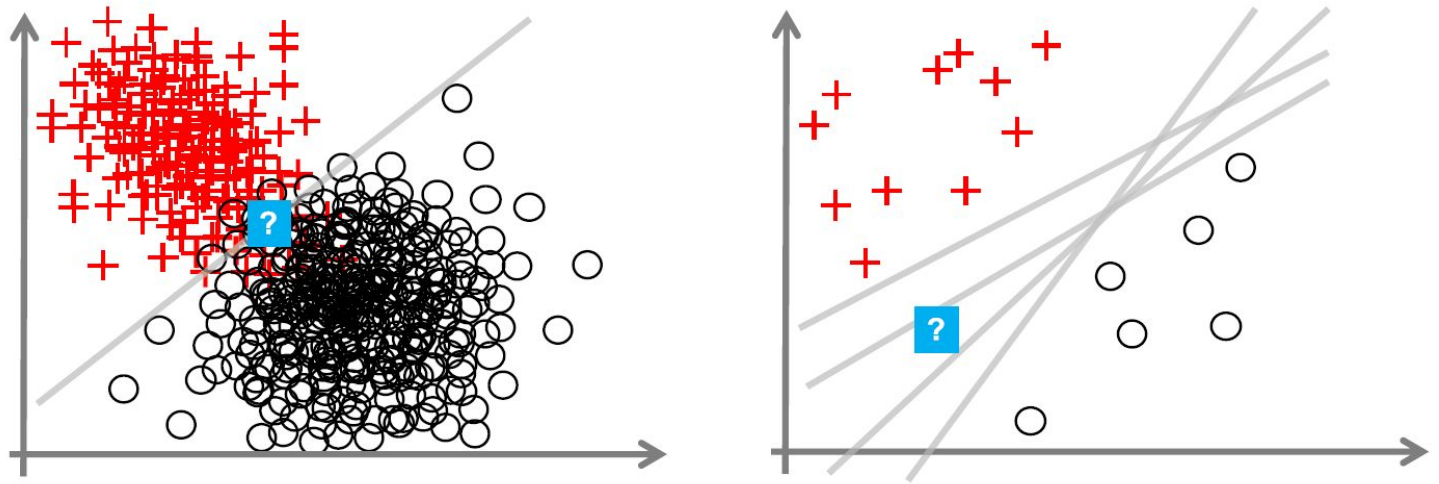
Clearly very dependent on the type of model being deployed…

- Is the input data coverage sufficient
- Is the model representative of all the required use cases
- Is there bias in the training data?
- Is there dataset drift during operation.

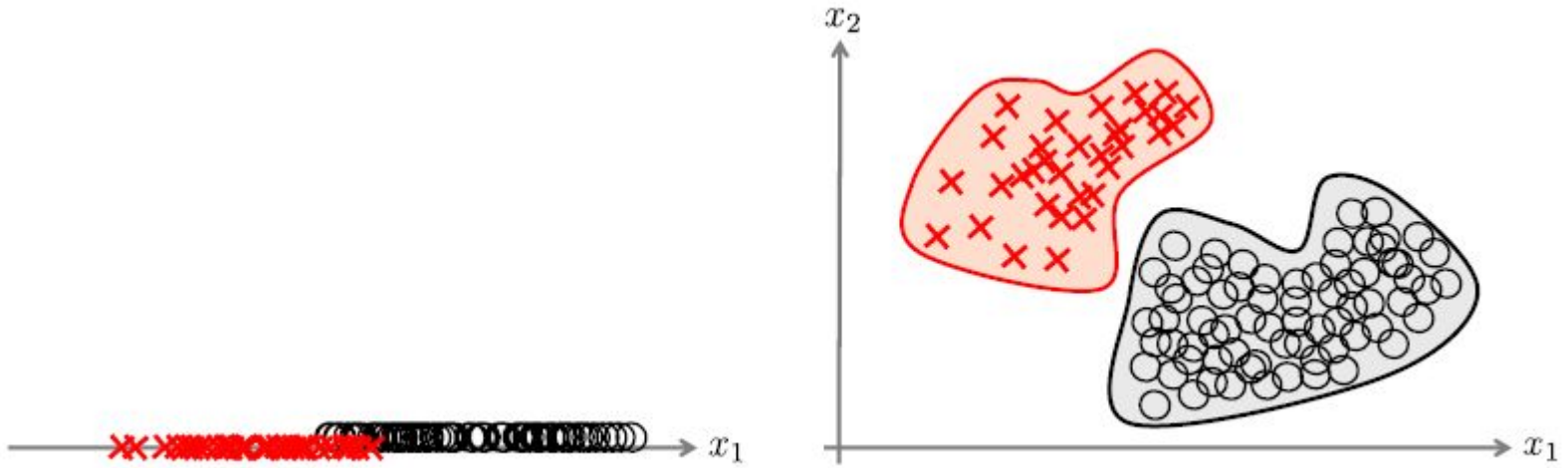
- Data incorrectly labelled
- Missing data
- Are all relevant dimensions/features within the model.
- How large is the measurement noise/variation.

Examples

Epistemic



Aleatoric

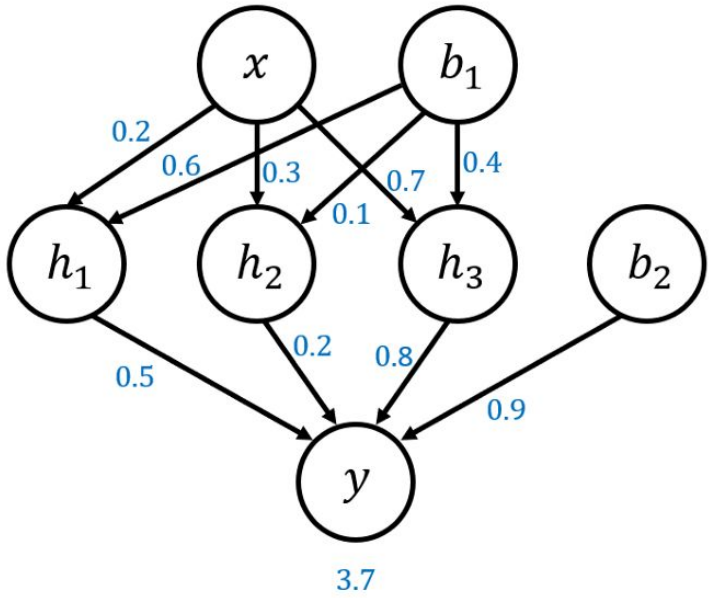


Uncertainty in Neural Networks (NNs)

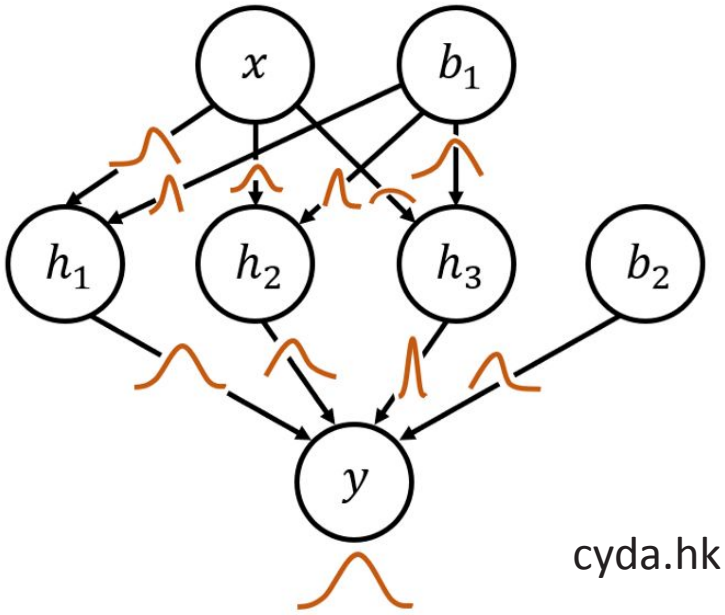
Deep learning neural methods are general pretty poor at knowing their own uncertainty.

Bayesian Neural Networks
Treats weights and outputs in NN as distributions uncertainty on output

Standard Neural Network



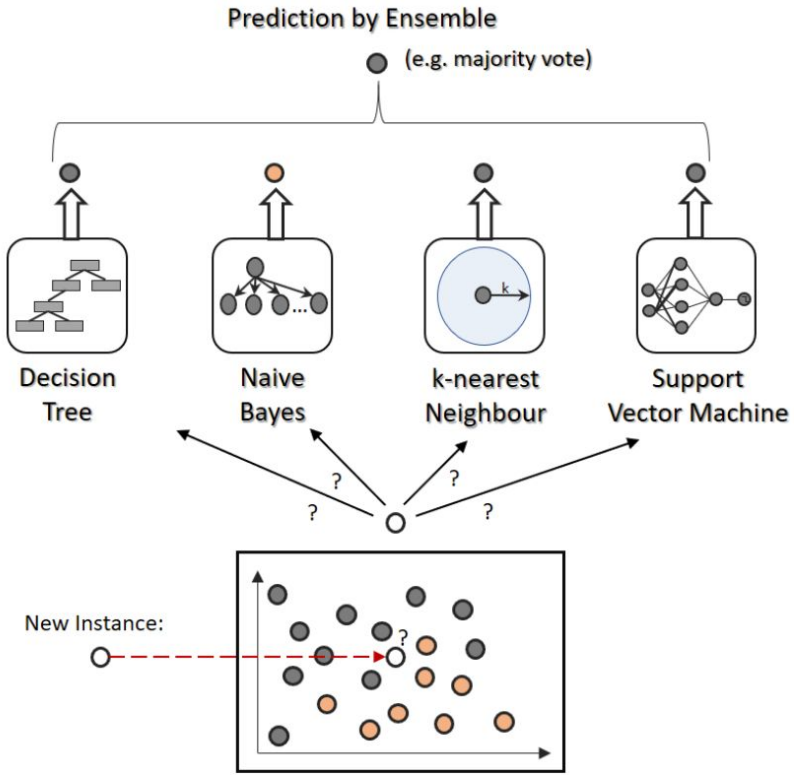
Bayesian Neural Network

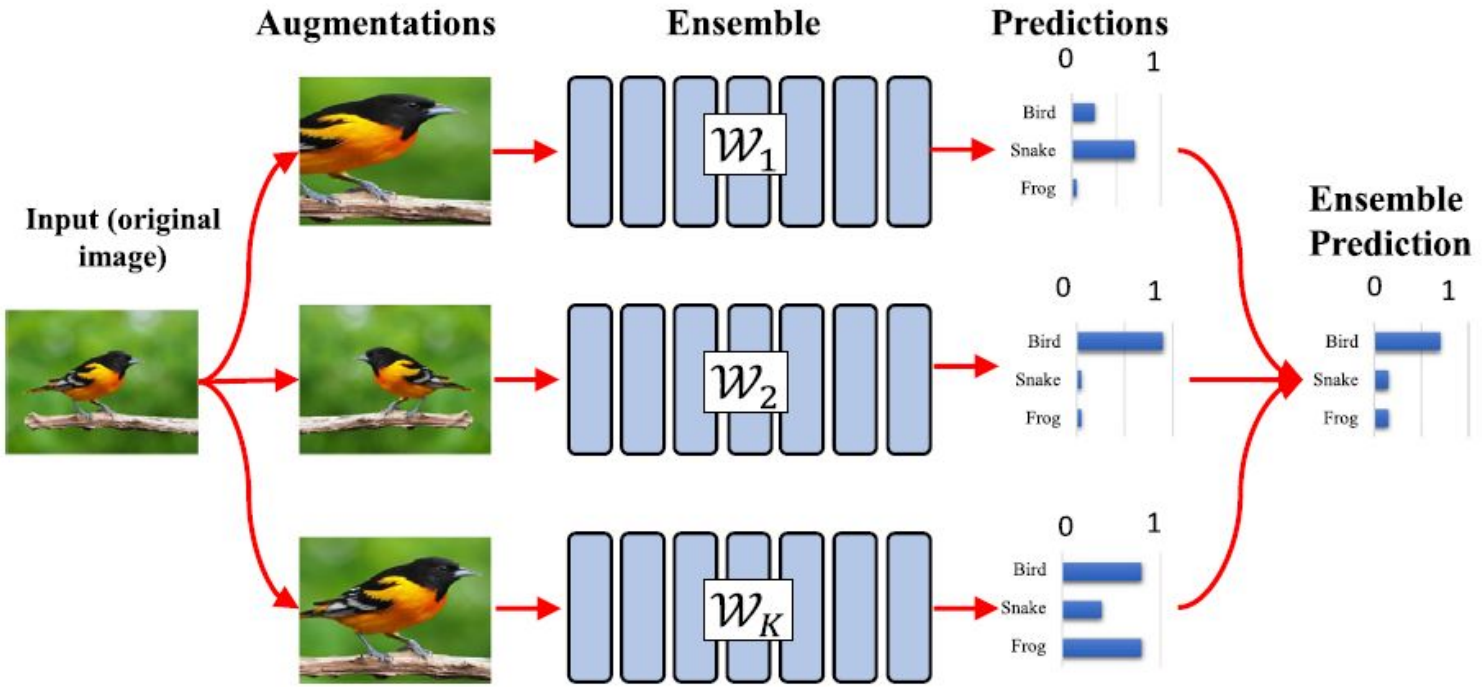


cyda.hk

Ensemble models

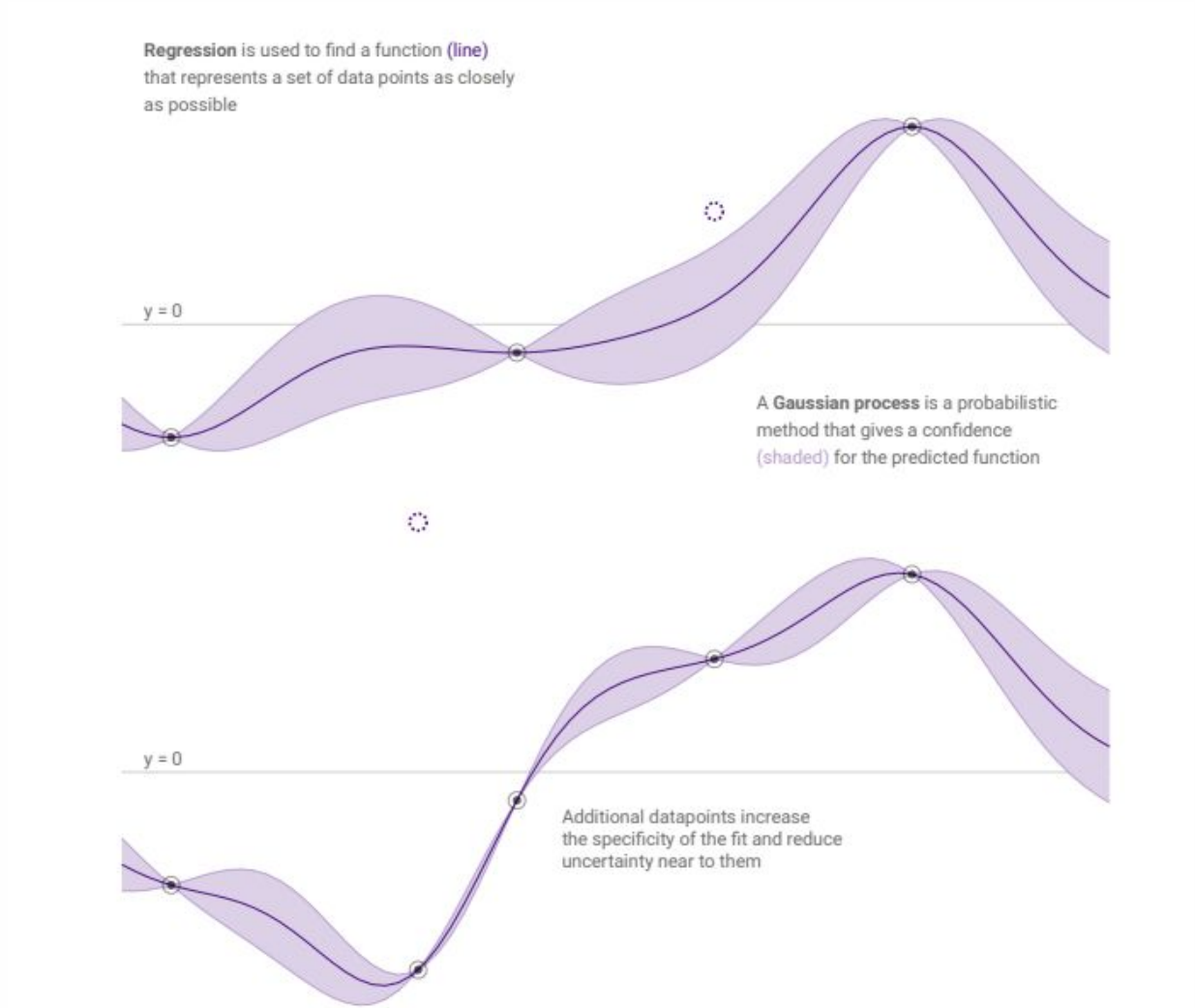
Using multiple different models and combining predictions. Can improve predictive performance and quantify aleatoric and epistemic uncertainties.





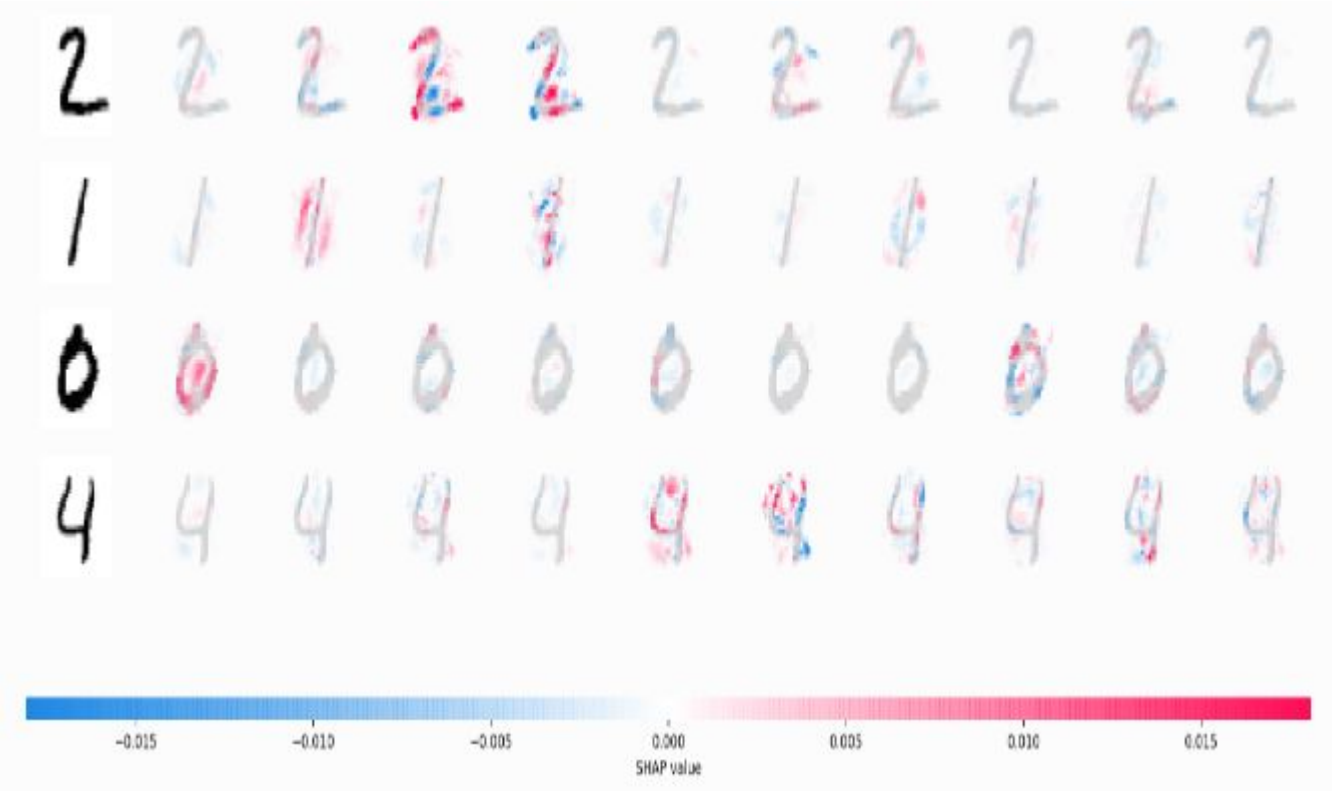
Images taken from the web

Gaussian process models

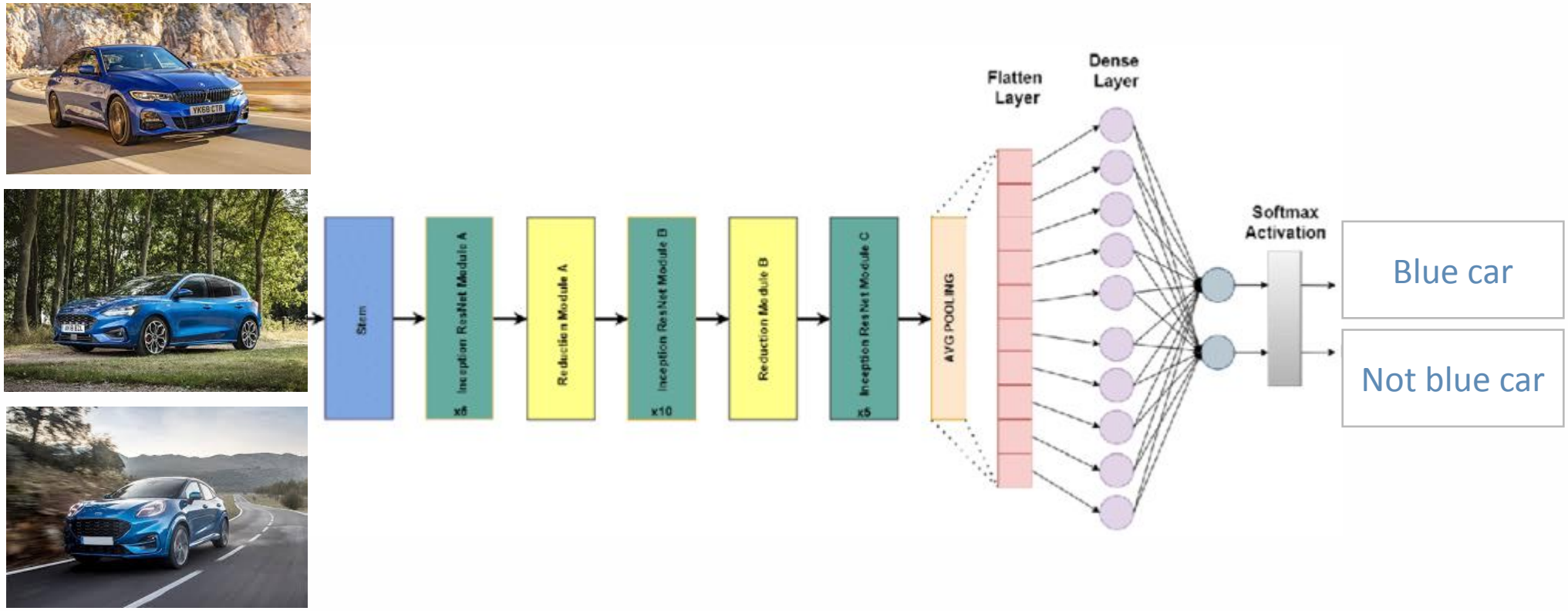


Explainable AI: Shapley values

Help guide towards the collective contribution of each feature to the outcome predicted by a black box model



Better understanding of your use cases

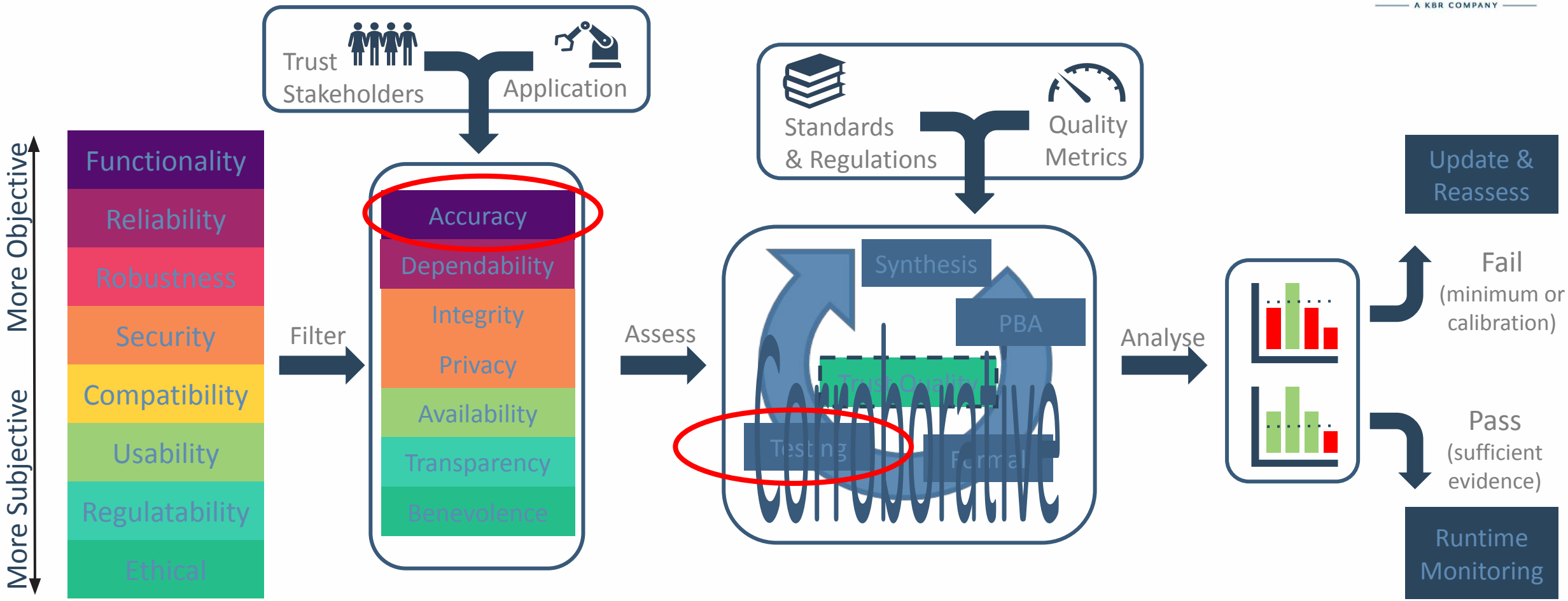


Better understanding of your use cases

Summary



Assuring trustworthiness in a system



Trustworthiness Categories

Application criticality and risk analysis

Corroborative Trust Quality Assessment against Standard

Quality Assessment & User Calibration Analysis

https://arxiv.org/abs/2305.03411v2

FRAZER-NASH
CONSULTANCY
A KBR COMPANY

Thank you

NOT PROTECTIVELY MARKED

SYSTEMS • ENGINEERING • TECHNOLOGY