UNCERTAINTY AND ARTIFICIAL INTELLIGENCE
IN ENGINEERING EDUCATION OF THE NEAR FUTURE

ELISABETH PATÉ-CORNELL
STANFORD UNIVERSITY

OCTOBER 9, 2024

DELFT SUMMIT ON THE FUTURE OF ENGINEERING SYSTEMS

THE NEAR FUTURE OF ENGINEERING EDUCATION

What is (often) taught now and what needs to be included

First, in engineering education, one needs to focus on the fundamentals of engineering systems (don't forget…)!

- Elements of math, physics, basic engineering, human factors
- Systems thinking, interdisciplinary engineering, and a lot more

Then, add (at least) two things to the current curriculum IF they are not yet taught in engineering

- **UNCERTAINTIES** and decision analysis beyond frequencies in statistical samples
- **ARTIFICIAL INTELLIGENCE**
  The nature, construction, limitations and use of AI algorithms

1. UNCERTAINTY AND BAYESIAN PROBABILITY

Focusing on the risk management of failure in engineering systems

- Current teaching about uncertainties is often based on frequencies in statistical samples.

  They look "objective" but they may not exist, e.g., for new systems
- They can be useful but are irrelevant if the systems have changed
- Risk assessment is based on Bayesian probability. It includes failure scenarios, their probabilities, their outcomes, and the time horizon
- A wrong statement often heard "You cannot do a risk assessment because you have no data". Yes, we can! Information and probabilities involve more than classical statistics.

TWO EXAMPLES OF NEW SYSTEMS: A CRITICAL LESSON

- **Nuclear power plants at the onset:** risk of a major accident?
  =>Fundamental probabilistic study in 1975 (WASH 1400) before any major accident and with limited information (US Navy reactors).

- **NASA Space shuttle** (or other space systems) The risk of a mission **loss** due to a failure of the heat shield: no such failure had occurred at the time of the risk analysis study (1990), but failure scenarios could be generated and analyzed.

- Then it happened (Columbia in 2003) but NASA had not implemented the study recommendations (including: reinforce the tank heat shield).

  => Don't ignore risk analysis results because "it has not happened yet"

WHAT NEEDS TO BE TAUGHT: FIRST, BAYESIAN ANALYSIS AND PROBABILITY (INFORMATION AND DATA)

Failure scenarios (failure modes) i.e., conjunction of events leading to failure. Consider, first, classes of scenarios free of details

Assessment of their probability using fundamental Bayesian formula:

Probability p(A and B) = p(A) x p(B given A) x p(C given A and B) …

Computation of the consequences of each scenario (outcomes)

The risk analysis result: a probability distribution of the potential losses. Here: the chances that each loss level is exceeded over the lifetime of the system (or per time unit).

What risk is **not:** the expected value of the losses!

SOURCES OF INFORMATION

- Statistics of component failure in the same system if they exist and are relevant
- Surrogate data about the performance of the same component in other systems (with updating if needed)
- Engineering, human factors, and management models
- Expert opinions focused on expertise about the different factors of an accident scenario (=> limited subjectivity).

RISK ASSESSMENT IN THE DESIGN PHASE (EXAMPLES)

CHOICE OF SYSTEM STRUCTURE AND COMPONENTS' STRENGTH

Loads and capacities

- Both are generally uncertain => probability densities f and cumulative distribution functions F (probability that the variable is less than each level x).
- Probability of failure: it is the chances that the load exceeds the capacity

  p(Failure) = $\int f_L(x) \times F_c(x)\, dx$ (the chances that the load is x and the capacity is less than x)

  Examples:

  A house in seismic area: uncertainty about strength and frequencies of earthquake

  The heatshield of a space craft and its capacity to absorb heat and radiation loads

System's structure and redundancies:

p(Failure of redundant subsystem) = p(Failure of both components in parallel)

p(Failure) = p(A and B) = p(A) x p(B given A)

= p(A) x p(B) if their failures are independent

RISK ASSESSMENT IN ENGINEERING OPERATIONS

DECISIONS IN OPERATIONS: ADJUSTMENT OF COMPONENT RESILIENCE AND OPERATING PROCEDURES. SOME EXAMPLES

- Understanding the critical functions and the probability that some components may have weakened (before the subsystem fails if possible!)
- Assessing the relative chances of failure of the components (ex: the possibility of loss of power)
- Assessing the chances of an external event (e.g., a hurricane) and protecting the system before it happens
- Observing human behaviors and the chances of human errors: training of personnel if necessary, hiring someone else if needed
- From a management perspective: providing the necessary resources and reasonable time for an acceptable risk of failure to accomplish what needs to be.

SO WHAT? EFFECT OF UNCERTAINTIES ON RISK MANAGEMENT

- The recommended approach: decision analysis under uncertainty based on quantitative risks and imperfect information
  - Rationality can be defined by the classic von Neuman axioms (ex: no circularity in preferences)
  - A utility is assessed for each outcome (preferences of the decision maker)
  - The optimal option maximizes the expected utility
- The answer depends on these preferences, and there is no “right” preferences (within ethics, reason, and rationality)
- Importance of the risk attitude of the organization (or its managers) and of tradeoffs among the different attributes of the outcomes

WHY A PROBABILISTIC AND NOT A DETERMINISTIC ANALYSIS

One can certainly rely on a qualitative judgment and pure intuition

But quantification of probability allows rational comparison of risks, effectiveness of safety investments, and effects of failure dependencies

=> At least three reasons to adopt a probabilistic analysis (beyond a deterministic one):

1. There is no (or little) experience, and guesses are insufficient to make rational decisions (e.g., nuclear power plants in the 50's)
2. A probabilistic analysis allows optimal allocation of resources to achieve an objective and avoid catastrophic failure
3. It allows identification and management of failure dependencies => it affects design and operations decisions

WHEN WOULD HAVE SUCH AN ANALYSIS BE USEFUL

Two examples

The Air France accident over the south Atlantic in 2009

- The decision was made to delay the replacement of the Pito tubes.
- Pilots needed more training to manage “pull ups” and stalling risks

Both safety measures were implemented *after* the accident

The Columbia accident, NASA 2003

NASA had been told to reinforce the heatshield of the external tank to avoid losing pieces of it that could hit the tiles at take off. They decided to ignore the warning

One obstacle: few people know how to do a probabilistic analysis analysis

2. ARTIFICIAL INTELLIGENCE

Two major AI functions that need to be taught in a dynamic mode:

- **INFORMATION** Generated from an initial database, trained then enriched by Machine Learning => outputs based on multiple sources of information
- **DECISIONS** Suggested by the algorithm or automatically implemented. At least two types of decision algorithms: 1. designed and populated by analysts (including a risk attitude) or 2. based on crowdsourcing.

In both cases, one needs to understand the data sources, how the message is generated, its advantages, its potential errors, and possible misalignment of preferences between the algorithm and the decision maker.

How does one judge the chances of errors, of misalignment and of problems using AI information and recommendations as decision support?

AI-GENERATED INFORMATION

Generation and processing

- One may start with an existing data base (often large and complex)
- Training of that database: structuring, enrichment, and inclusion of uncertainties
- Machine learning based on regular input from experience, and on the machine's ability to look for information by itself (a dynamic process)

Advantages of the volume of data processed and of the processing speed

- Increasingly powerful computers allow for speed of information gathering and processing, superior to human performance, and on a large volume of data.
- But there are timing problems to speed in critical situations.

  Ex: operating room. The surgeon may not have the time to understand where the data come from, and ask for additional information

Problems of error anticipation and detection; ex: truncated database, errors and biases from the sources, difficulty to identify errors and simply not enough time

AI-GUIDED DECISIONS

- Preferences in AI-generated risk management decisions are based on the source of the algorithm and human inputs (an analyst or a crowd source). Does the AI system fit the preferences of the decision maker (alignment)?
- Potential of alignment in autonomous systems (do they represent human preferences?) as opposed to automated decisions with possible human intervention, e.g.: car automatic speed with possible driver's intervention
- Four examples
  - **Medical decision** (should one do a specified test if recommended by AI?)
  - **Autonomous versus automatic vehicles**: should one let them function without human input, even if they are programmed to observe all laws on the book but can make unfamiliar moves?
  - **Military systems.** One example: should one use autonomous drones?
  - **Sports**, e.g., regattas. Should a skipper change her/his course based on the recommendations of an AI system about winds and currents?

WHAT STUDENTS AND PRACTITIONERS SHOULD KNOW

- Uncertainty is a critical part of informed, rational decisions… even though people are often uncomfortable with it! Quantify it!
- Don't present the most likely possibility as if you were sure of it
- Don't dismiss extreme values as "black swans" that are too unlikely
- Don't assume independence of events without checking
- Don't make "predictions", as they are often interpreted as certitudes
- More information (in AI databases) does not mean less uncertainty
- Make sure that the decision maker is aware of the source of information and of the possibility of risk attitude alignment problems
- **Appreciate the value of AI and ChatGpt…with a grain of salt.**

ENJOY AI (WITH OPEN EYES) FOR ALL IT HAS TO OFFER!