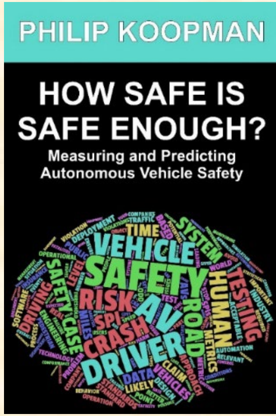
Prof. Philip Koopman

Carnegie
Mellon
University

Quick Look At Redefining Safety

29 June 2025

PhilKoopman.Substack.com



PHILIP KOOPMAN

HOW SAFE IS SAFE ENOUGH?

Measuring and Predicting Autonomous Vehicle Safety

Absence of Unreasonable Risk

Carnegie Mellon University

- Automotive definition of safety (e.g., ISO 26262)
  - Absence of Unreasonable Risk (AUR)

- Thought question:
  which would you want to eat?
  - Poison
  - Non-toxic (i.e., absence of poison)
  - Edible

- AUR means not dangerous enough to be recalled
  - Acceptable safety requires more than AUR

Definition of a Safety Case

- Safety case: structured argument, supported by evidence, intended to justify that a system is acceptably safe for a specific application in a specific operating environment.

- Autonomous system challenges:
  - Who/what ensures no misuse of intentional abuse?
  - Who/what handles departure from specific operating environment?
  - Safety case needs to address these issues to be complete

A Non-Engineering View of Safety

- Public acceptance is weakly linked to engineering analysis
  - Stories matter more than statistics

- Hypothesis:

  For each crash, the public will judge safety by whether they think they themselves would have avoided that particular crash as a human driver.

Collected Definitions: See: https://bit.ly/44qjrs3

- Loss: an adverse outcome, including damage to the system itself, negative societal externalities, damage to property, damage to the environment, injury or death to animals, and injury or death to people
- Risk: combination of the probability of occurrence of a loss, or pattern of losses, and the importance to stakeholders of the associated consequences
- Safety constraint: a limitation imposed on risk or other aspects of the system by stakeholder requirements
- Safety engineering: a methodical process of ensuring a system meets all its safety constraints throughout its lifecycle, including at least hazard analysis, risk assessment, risk mitigation, validation, and field engineering feedback
- Safety case: structured argument, supported by a body of evidence, that provides a compelling, comprehensible, and sound argument that safety engineering efforts have ensured a system meets a comprehensive set of safety constraints
- Acceptable: meets all safety constraints as shown by a safety case