What Next?

(Ruminations by a WG10.4 Old-Timer)

Brian Randell

The Dependability "Bible", and Community

- Avizienis, A., Laprie, J.-C., Randell, B. and Landwehr, C. Basic Concepts and Taxonomy of Dependable and Secure Computing, *IEEE Transactions on Dependable and Secure Computing*, Vol. 1, No. 1, pp 11-33, 2004.
    - as of June 2023, **7472** Google Scholar Citations! (Incidentally, I can't find anything remotely comparable from, IEEE S&P Oakland.)
- The generality/recursiveness of our concepts of 'failure' and 'system' is one of our strengths, allowing us to encompass:
    - both accidental and malicious faults
    - specification faults, via the notion that "systems come in threes", i.e. that there is always an implicit judgment system, not just an "environment" to a system.
    - all types of system
- But it is I fear also a weakness or irrelevance, in the eyes of:
    - What can be regarded as more specific research communities, e.g. Safety, and Security
    - those interested just in specific types of system and application

But how influential were/are the Dependability “Bible” and Community?

- We have: Knight, John. *Fundamentals of Dependable Computing for Software Engineers*. Routledge (10 Feb. 2012) 436 pages, ISBN-10: 1439862559.
    - (Are there *any* other text books based on the “Bible”?)
- Amazon Book searches: on “computer, dependability” – 181 results; “computer fault tolerance” – 331 results; “computer, trustworthy” – 184 results; **BUT** both “AI” and “computer, security” – over 50,000 results!
- DSN attendance (2021) – 369; **BUT** AI conference attendances (2022): NeuriPS – 15,530, CVPR – 10,170, ICRA – 8,010, etc.! (From *Artificial Intelligence Index Report 2023* - https://aiindex.stanford.edu.)
- **What should WG10.4 and the Dependability Community try to do in response to this situation? Can we turn AI from being a threat to us into an opportunity?**

On Dependability and AI

- The fundamental techniques for achieving high Dependability are: Fault Prevention, Fault Tolerance, Fault Removal, and Fault Forecasting. All four of these techniques are usually needed.
- A critical issue always is that of avoiding confusion as to the boundaries of the system being discussed.
  - An unreliable computerised hospital bed booking system can form part of a highly reliable bed booking service (i.e. system) – thanks to crafty operators!
  - An occasionally failing computer processor may run under the control of an operating system whose backup and restart facilities completely hide its failures.
- Surely, this sort of thinking, and the Dependability research that backs it, need to be employed by the AI community wherever AI systems need to be trusted, especially safety-critical applications.
- What matters of course is not just the AI software system, but also the encompassing system that includes all the learning data used.
  - In general, both these systems need to employ, systematically, all four means of achieving Dependability. It would appear that all too often they don't!

Concerns About AI Systems

- Concerns are currently widespread across Society, Industry, and Governments, despite rapidly growing (often successful) AI deployments.
- For years one prominent concern has been the dangers of possible safety-critical applications of AI, e.g. Autonomous Vehicles (AV).
- Current concerns relate largely to **Generative Artificial Intelligence** (GAI), and Artificial General Intelligence (AGI).
- GAI worries centre on issues such as Plausible Incorrect Results (especially “Hallucinations”), and the potential for its misuse.
- AGI worries relate to the possibility of AI systems dominating human systems, developing their own (inhuman) goals. (I suggest we leave these worries to others.)

AV and GAI Concerns – an Opportunity for WG10.4?

- Autonomous vehicles (AV) and Generative AI (GAI) seem to be very different arenas.
- AV is the province of well-established engineering and defence industries, and AI (in particular Machine Learning – ML) is just one of the technologies they employ.
- This is a province we have some good relationships with already.
- GAI, in particular text and image generation (based on "large language models"), and its problems are presently much more in the public eye, and more directly used (and misused) by and on the general public.
- GAI is an arena in which industry is racing ahead of academia.
  - See: *Awesome Generative AI – A Curated List of Modern Generative Artificial Intelligence Projects*: https://github.com/steven2358/awesome-generative-ai
- Is it naïve to think we could contribute to such a speeding train?

What might WG10.4 do for GAI?

- Do we have any good contacts and relevant collaborations already?
- Would it be feasible and useful for WG10.4 membership to undertake to produce a constructive critique of various well-publicised GAI systems’ use/non-use of dependability tools and strategies?
- Could the Dependability Community usefully create and experiment with its own “fork” of a leading Open-Source GAI system?
- A naïve suggestion: How about “wrapping” a GAI system up in a (very conservative) Fact-Checking system?
- A very naïve suggestion: How about using Triple-Modular-Redundancy in an attempt to improve the accuracy of a GAI system? (When ChatGPT provides several *different* answers to a factual question, especially a reference request, my guess is that often all are so-called “hallucinations”.)
- Have such ideas in fact already been tried/employed?

If we don’t like the idea of concentrating on GAI

- (Disclaimer: I’ve become very enmeshed with GAI recently.

  See: Randell, B and Coghlan, B. ChatGPT’s Astonishing Fabrications About Percy Ludgate. IEEE Trans. Comp. History 45, 2 (April 2023) pp. 71-72. – and https://www.scss.tcd.ie/SCSSTreasuresCatalog/miscellany/TCD-SCSS-X.20121208.002/ChatGPTs-AstonishingFabrications-aboutPercyLudgate-CoghlanRandellOBoyle-20230424-1434.pdf)
- But should we concentrate instead on Autonomous Vehicle systems or some other particular type of AI system?
- Or is there something better (and bigger) that we can usefully try to do in response to the ever-growing strength of and furore over AI?

A Possible Much More Ambitious Plan

- How about trying to engineer a more general co-operation between chunks of the AI and Dependability Communities?
  - Unless and until official sponsorship and funding could be engineered, this would have to be a bottom-up informal co-operation.
  - Question: do WG10.4 members have the influence and contacts needed to (i) identify existing potentially-rival AI/Dependability initiatives, and (ii) put together an effective steering committee and program?
- Some possible technical objectives:
  - A State-of-the-Two-Arts report identifying and discussing (i) existing AI systems of various types and any (ad hoc) dependability-type approaches and mechanisms employed, and (ii) the latest/best dependability techniques and tools of potential direct relevance to such AI systems.
  - A co-ordinated set of experimental dependability projects, based on existing open-source AI systems, each exploiting and evaluating the effectiveness of an established dependability strategy, for example, for fault injection or fault tolerance.

A More Modest Plan

- Production of an updated version of the Dependability “Bible” to cover post-2004 developments.
  - Updating the sets of references and the cited examples of dependability problems, tools and techniques.
  - More extensive coverage of security dangers and solutions.
  - Revision and expansion in the light of developments in AI, especially the use of ML as a system design and implementation technique.
- It would need a fresh new energetic set of authors!
- But is this not too inward-looking, and unambitious?

Some Concluding Questions

- Dare we deliberately ignore AI, despite all its current publicity, strengths and problems?
- Is there a good non-AI way to strengthen the Dependability Community (and help to justify WG10.4's continued existence)?
- Something related to system security perhaps?
- Would getting a fresh new energetic set of authors to update the "Bible" to cover post 2004 developments, and expand the treatment of security issues be sufficient?

And My Answers

- Dare we deliberately ignore AI, despite all its current publicity, strengths and problems? – **NO!**
- Is there a good non-AI way to strengthen the Dependability Community (and help to justify WG10.4's continued existence)? **WHO KNOWS?**
- Something related to system security perhaps? **SUCH AS?**
- Would getting a fresh new energetic set of authors to update the "Bible" to cover post 2004 developments, and expand the treatment of security issues be sufficient? **NO, BUT IT COULD HELP!**