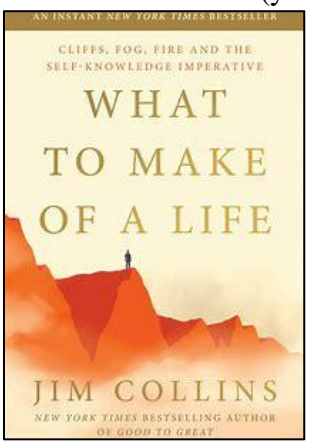
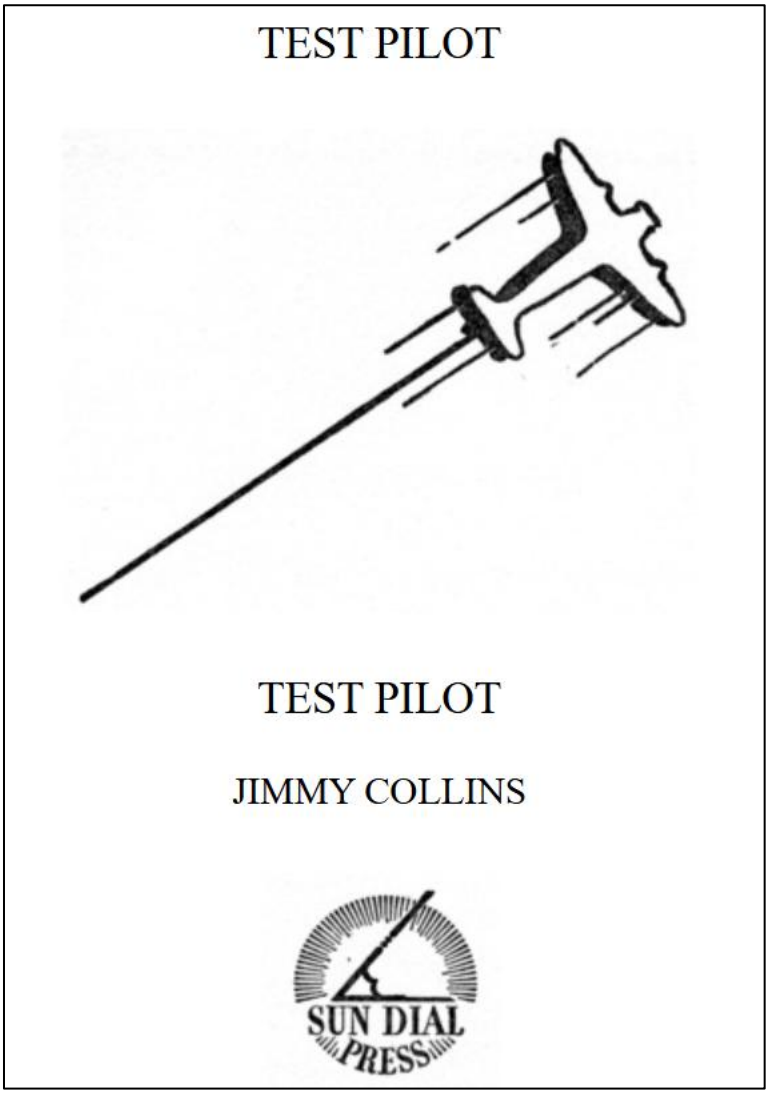
The Carnival of Flight Test

It's the season for state fairs, like the Great American State fair that coincided with the 250th birthday of the United States, barely a month ago. State fairs are distinctly different from carnivals, but nevertheless the two "words" have a similar mapping in the vector space of large language models. I won't fully address this comparison and contrast in this column. Furthermore, the astute reader will also point out that the fireworks show was not part of the State Fair or a carnival, to which I would reply that many carnivals may have fireworks or light shows that add to the recreational value, and it is precisely the whimsical value of a carnival that I hope to bring to this column. It may also be an explanation for this rambling introduction. This isn't the first time "carnival" has appeared in these pages, but it probably needs an explanation, something I will get to after we pass through the metaphorical cotton candy, juggling clowns, and midway games with their lights and whistles. Let us begin. Jim Collins is the grandson of a test pilot. Maybe you don't know Jim Collins, or perhaps you have read Good to Great and know who he is. I discovered this fact, his relationship to a test pilot of old, while reading What to Make of a Life, which is a book I will recommend, with reservations. I'll let the grandson tell the story of the grandfather, but I also think our readers will enjoy this book. The idea of "navigating" life is a central theme, and the metaphors Collins uses throughout to discuss the findings of his research, which includes deep biographical study of two test pilots turned astronauts, align with this central idea. I also think the book cover is aesthetically pleasing, more so than the book's conclusions which seem incomplete, but there is a reason for this (yet another) ramble. The young Collins introduces the reader to a test pilot named Jim Collins. The test pilot died "on March 22, 1935, [while] conducting the tenth and final dive test of the Grumman XF3F-1 prototype (Bu. No. 9727) at Farmingdale, New York" (https://www.thisdayinaviation.com/tag/james-h-jimmy-collins/). After his death, his wife discovered a note saying that the "last chapter" of his memoir was in a desk drawer. She found it and published the memoir, which was later made into a movie. Any book with such a title must be better than the "The Right Stuff," which is one of two heinous misnomers in this column. I haven't done deep research into the movie, but I did want to share links here: Test Pilot (1938) - IMDb, which also includes a video clip from the movie. The digital book, Test Pilot, is available for free from Project Gutenberg here: https://www.gutenberg.org/ebooks/34589. The cover has a simple but beautiful design. I've not read the book but would love to find the time. I find it easier to read physical books than digital, unless I'm travelling on a plane (as a passenger). I'm hoping at least one of our newsletter's readers attempts the book and reports back. There's probably something in it worth more than the oversized stuffed animal of carnival

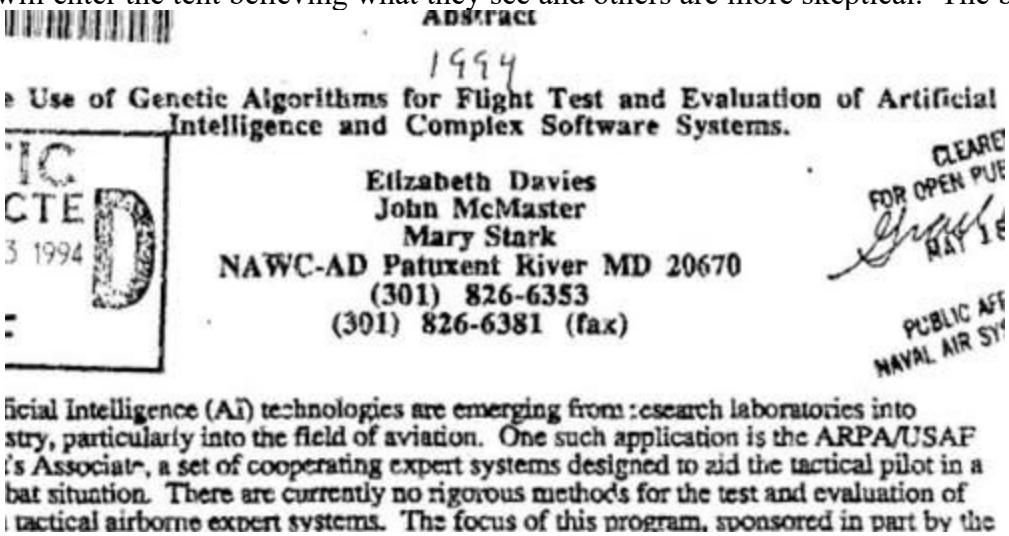


fame. There's also more to read about Jimmy Collins here: [The JAMES T. "Jimmy" COLLINS Page of the Pitcairn Field Register Web Site](#).



One of the websites that I read for pleasure is "The Carnival of Mathematics," and on a regular basis, they publish a numbered newsletter with fun, whimsical, and occasionally serious information about a particular number (like 17, for example). I reported that first in the October 2024 Flight Test Safety Fact, and it's the inspiration for this loose collection of facts, figures, and fun diversions. Speaking of fun diversions, the July 4th festivities included incredible displays of airpower, truly bone shaking and exhilarating fly-bys and fly overs. Aerial Demonstration is a topic that continues to find its way into the subject of Flight Test Safety, and it feels like we are due for another such presentation or paper. We still await the time when someone will take up the mantle of data wrangler for Aerial Demonstrations, a task that has fallen into neglect since the late author's decease. You can find more information about the relevance of flight test safety in airshows on the Flight Test Safety's list of resources: FTSC_Airshow_Guidance_FINAL.pdf, which also has more references to explore.

We also need Data Wranglers for the list of Flight Test Mishaps, and both tasks may require knowledge of Git and Github, something that has also appeared previously in these pages and referenced this repo: [GitHub - flighttestfact/flighttestfact.github.io: helping you communicate better, stop a crash, and save a life · GitHub](#). Now we turn the column to a whimsical tour of Artificial Intelligence, a misnomer if ever there was one (that's the second). "AI" is one of the attractions we will find in this month's Carnival of Flight Test. Some will enter the tent believing what they see and others are more skeptical. The believers had the first word in last month's annual report, which introduced the SAFETYNET, a RAG type LLM, which can be used to search the THA database. It didn't take long for the Editor to get letters about the subject, and, perhaps coincidentally, the SETP started its pilot program (no pun intended) studying the efficacy of a similar RAG LLM for search of the Society's digital resources, which included technical papers, video


Abstract

1994

e Use of Genetic Algorithms for Flight Test and Evaluation of Artificial Intelligence and Complex Software Systems.

Elizabeth Davies
John McMaster
Mary Stark
NAWC-AD Patuxent River MD 20670
(301) 826-6353
(301) 826-6381 (fax)

CTE D
3 1994

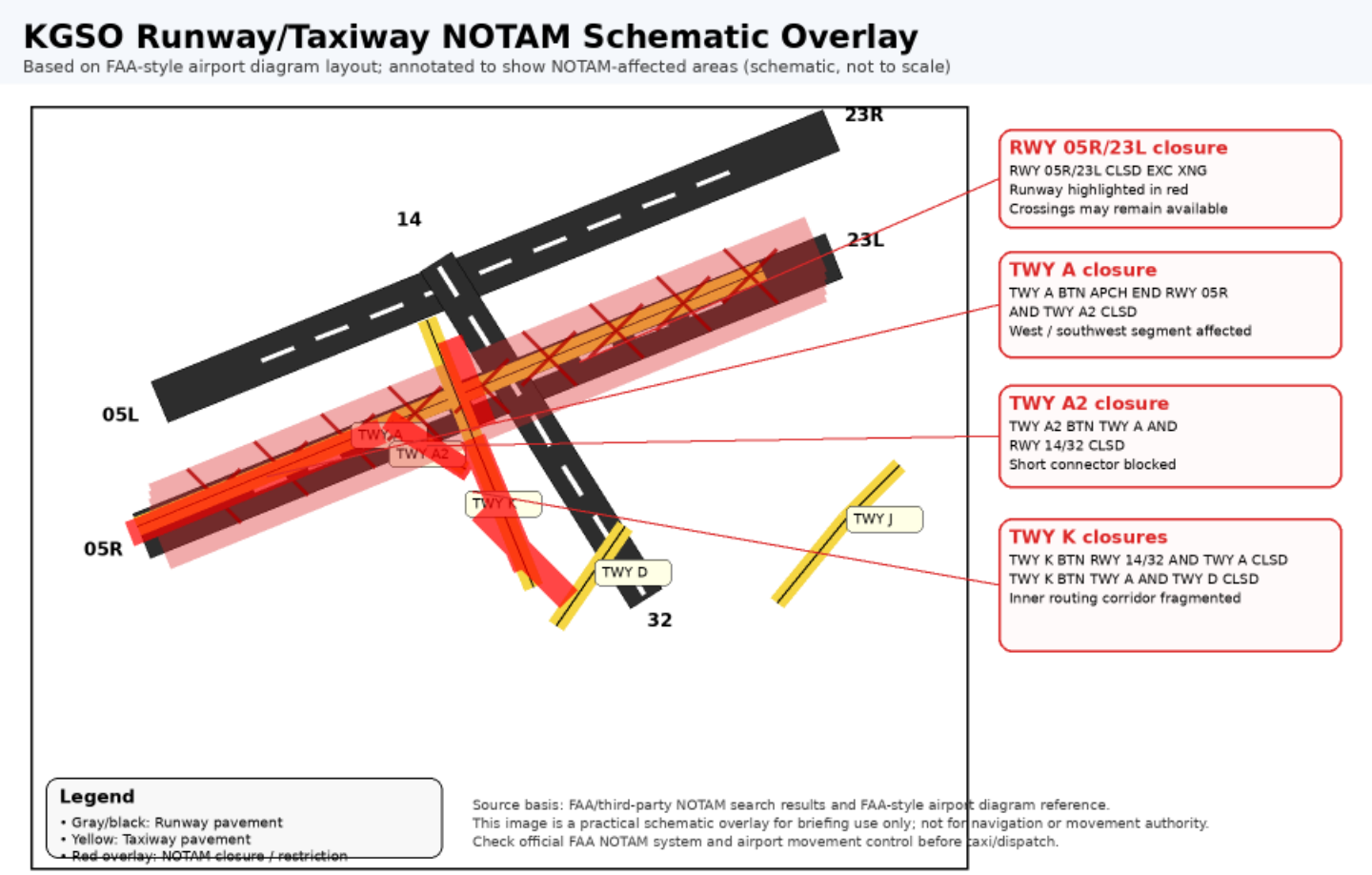
CLEARED FOR OPEN PUBLIC RELEASE
MAY 16

PUBLIC AFF
NAVAL AIR SY

ficial Intelligence (AI) technologies are emerging from research laboratories into
stry, particularly into the field of aviation. One such application is the ARPA/USAF
's Associate, a set of cooperating expert systems designed to aid the tactical pilot in a
bat situation. There are currently no rigorous methods for the test and evaluation of
tactical airborne expert systems. The focus of this program, sponsored in part by the

transcripts, Cockpit magazines, etc. The first round of feedback is in to the facilitators of the group, and I guess we will have to see what comes of it. SFTE has a chorus of supporters, including its president, and a member who just presented an hour-long webinar to its members about AI. The topic of AI isn't new. The screenshot above of an old SFTE Symposium Paper (available in their papers database) should serve as proof enough. The date is 1994. Rod Huete probably only had a few gray hairs back then. Bill Gray's hair was short back then, and now he's finally retiring. The Flight Test Safety Fact has also published features on AI. For example: https://flighttestfact.com/artificial-intelligence-in-flight-test-a-limited-survey/, https://flighttestfact.com/machine-learning-techniques-applied-to-flight-test-data-evaluation/, and https://flighttestfact.com/global-civil-aviation-authorities-wander-through-the-landscape-of-ai/.

I guess now is as good a time as any to point out the "business model" for AI companies: Give the public a tool that doesn't work well, and then explain that "if they buy the premium subscription, it will work really well." That feels like the kind of sales pitch you would get at a carnival. To support my broad-brush judgment of AI, let me present several pieces of evidence. Exhibit A appears below and on twitter here: https://x.com/FlightTestFact/status/2069849214910152891. I asked Microsoft Copilot to put the NOTAMs on the airfield diagram of KGSO (Greensboro). I even provided a pdf with the airfield diagram in it. Personally, I think the output looks terrible. It seems like AI "can't" do an acceptable job of putting things ("drawing things"?) on existing images. For example, I once gave Gemini a text and a jpg file, and I asked it to put the text ON the image. It failed. The second to last line of the text appeared twice in the output.



Lest the reader weary of my one-sided hot takes, I'll use this time to present a "good example," and it just so happens that this example links back to America's 250th birthday. An artist on X (aka Twitter) created a music video for America's Independence Day celebration. "Sweet Land" by Heavy Pulp relies heavily on AI (Grok's family of AI tools). The song is brilliant. It might even make you cry or laugh or both. I could probably wax

critical or poetic or both about the music, the lyrics, the artwork. I'll simply leave it here for you to enjoy: https://x.com/heavypulp/status/2071990142064324766.

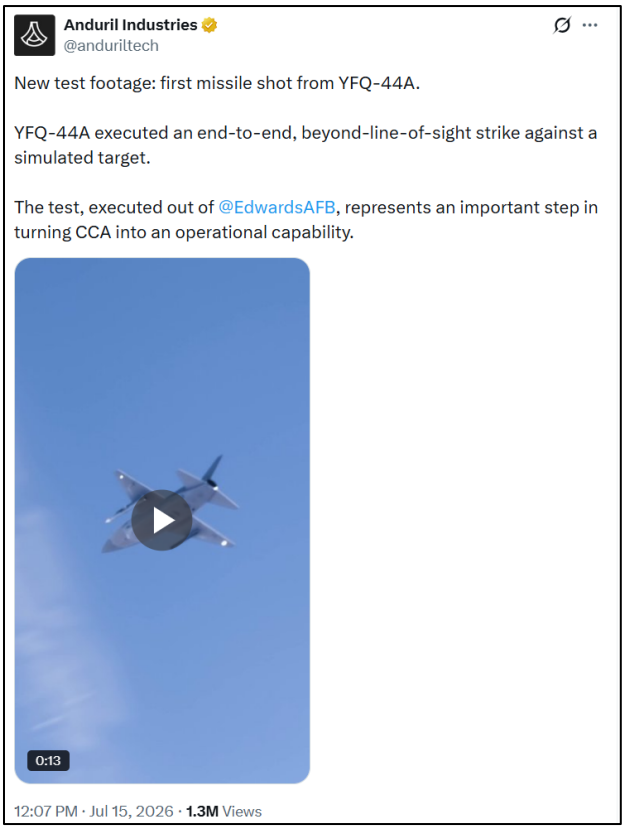
If AI is so amazing, why does it take so long to render a document for printing in Adobe? That's neither here nor there, but what does matter is how long it takes to train the AI. In the aforementioned SFTE webinar, two examples illustrated the range of possible outcomes. Researchers trained AI to play ATARI in ten days of equivalent computation time. In other words, 240 hours of training a machine learning model. Is that how long it took you? Is that "intelligence" or "brute force." I think if a child took that long to learn, we would NOT be labelling it intelligence. The second example hearkens back to Google's DeepMind experiment, during which it trained the DeepMind to play Starcraft. This feat took 200 years of computation time. That is a long time. This is not intelligence. We'd probably been more efficient programming a dumb algorithm to play in far less time. Speaking of playing video games. Several years ago, DARPA invested millions of dollars to prove that "a computer could win a video game that simulated air-to-air combat in the F-16." The research continues, and here's the headline: DARPA, U.S. Air Force fly AI-controlled F-16 | DARPA. Here's a quote from the article: "These groundbreaking flight tests of VENOM-modified F-16s *advance the infrastructure* needed to…" (emphasis added). That doesn't sound like much progress. And maybe, just maybe, ChatGPT could have written a better quote. Speaking of the famous LLM, this article seems relevant to this carnival of flight test and AI: https://link.springer.com/content/pdf/10.1007/s10676-024-09775-5.pdf. If you are afraid to click on the link, I'll summarize in the style of a Top Gun movie scene: "ChatGPT is bu11sh!t." That's not my summary—it's the scientific argument the paper is actually making.

Are we making any progress? Does Animal read this newsletter, and if he does, can he give us some news that may inspire hope?

I would be remiss not to mention the book Things that Make us Smart by Don Norman. It's relevant to the topic, and I thought there was an article in FTSF or FTN archives about the book. If someone can find it, please let me know. Using LLMs to search for it was not fruitful for me.

I'll close my rant with one last piece of "good news": https://x.com/anduriltech/status/2077424762490691808.

I'll offer you a chance to respond before finally closing this carnival of flight test and ai. What do you think? Do you have a strong opinion or an anecdote you want to share? Perhaps you have other references in the FT domain that would inform our readers. Additionally, if you have read this column and have ideas for the next carnival of flight test, please share the links, tweets, references, videos, and amusing thoughts.



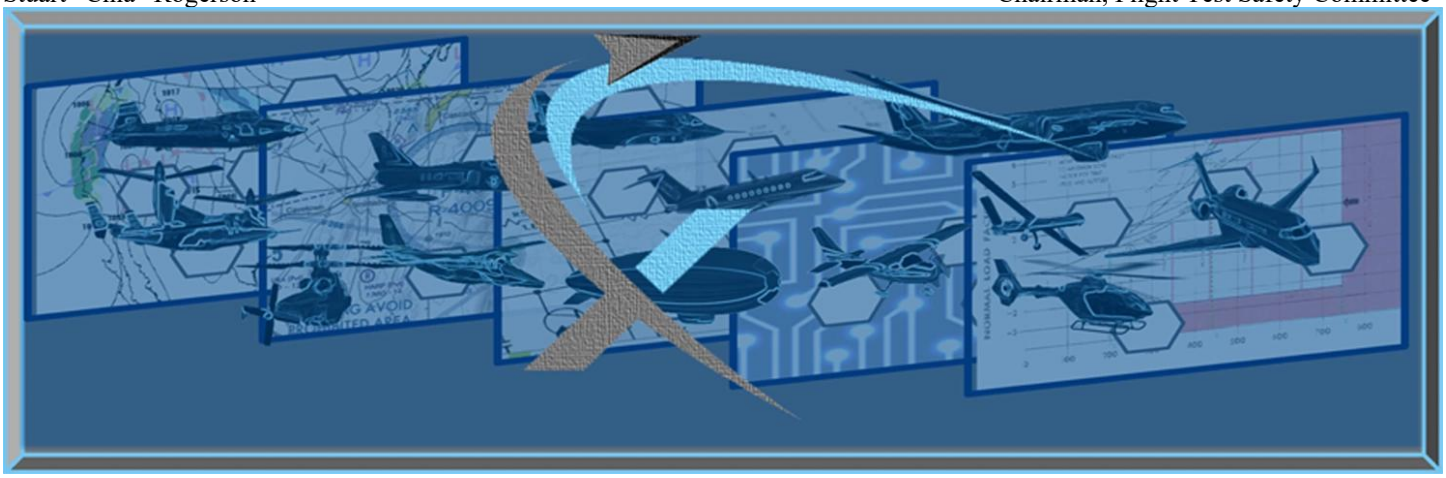
Bonus: Google Maps and the Odyssey (worth a chuckle)
https://x.com/realnewzy/status/2079193607630668095/video/1?s=46

Chia Chat

I am writing this Chia Chat on an American Airlines Boeing 777 cruising at FL340, just entering back into the United States after attending another fantastic European SETP symposium in Edinburgh, Scotland. I am sure the rumours [sic] of kilts and whiskey are completely unfounded. The European chapter always puts on an amazing event, and I enjoyed listening to some of the unique perspectives as well as common issues we face in the international flight test community. One of the papers that definitely caught my attention was "Cross Industry Lessons in High-Stakes Performance: Simulation, Teamwork and AI" presented by Michael Hsu (MD) and Julie Cathryn Rice (MD). It was an interesting discussion into some of the aviation methods like CRM and training being adopted into medicine. However, there was also a great discission on where medicine is going with new capabilities like AI. What I found most reassuring is that they have guarded optimism, similar to the FTSC's, in how we use AI and have gone down the same path that we have using a RAG and curated data along with separate AI tools that are tuned for their specific tasks (for example our search tool vs. the THA tool). It reinforced that the FTSC is hopefully on the right track with SAFETY NET, but as always, we need to remain vigilant and use these capabilities responsibly as just another tool in our tool box but never the complete solution. AI can be a great enabler but needs to be used appropriately, so high-stakes industries like medicine and aviation do not inadvertently create a bigger safety issue. I am not sure which presentations will be available from the European symposium, but there are a few worth watching if you get a chance.

In other news, we are getting close to rolling out the ability to **add new** THAs to SAFETY NET which will be a big improvement in capability. Unfortunately, we had to delay it to dedicate Ryan's time to switching our RAG host from Ragie.ai to Pinecone. (For those that don't know, Ryan is the brains behind SAFETY NET.) Ragie.ai provided the database search and retrieval part of SAFETY NET and was acquired and shutdown earlier this month by another company. With the switch to Pinecone, our costs should come down a bit, but we will need to do some tuning to get back the original performance. That is where I hope the flight test community will continue to help us. In the top right corner of SAFETY NET, there is a feedback option. Please provide that feedback. SAFETY NET will not be a static tool, and we will continue to split Ryan's time between adding new capabilities and fixing issues that have been found. We already have a long list of to do items, but we always welcome more input.

As always, feel free to reach out to me at chairman@flighttestsafety.org if you have any questions or comments. Keep risk mitigating!
Stuart "Chia" Rogerson Chairman, Flight Test Safety Committee



Subscribe to our Podcast

Available on iTunes, Spotify, Podbean, Google Play, and Amazon Music: FTSCChannel
https://flighttestsafety.org/ftsc-news/flight-test-safety-podcast-channel
https://ftscchannel.podbean.com/

EP 79 - Wide Right at Night and a VTO that's a NO GO! What would you have done? This month I am again asking, "What would you have done?" However, this time I share not only what I would have done, but what I actually did. Listen Now
Check out the video here: https://youtu.be/2-NZwdcr808.

Contact Flight Test Safety Committee

Stuart "Chia" Rogerson, Chairman chairman@flighttestsafety.org
Susan Bennett, FTSC Administrator susan@setp.org
Society of Flight Test Engineers edir@sfte.org
Society of Experimental Test Pilots setp@setp.org
AIAA Flight Test Group derek.spear@gmail.com

Contact *Flight Test Safety Fact*
Mark Jones Jr, Editor mark@flighttestfact.com

Website: flighttestsafety.org **Or cconnect with us by joining the LinkedIn Group: "Flight Test Safety Committee."**

Epilogue

"Imagine if someone told you that a year from today, you would be required to take a test in which every wrong answer resulted in the loss of a human life. How would you approach studying for the test? Would you study for twenty to thirty minutes every night, or would you wait until a week before the test and start cramming?"[1]

There's a corollary to Byerly's question, and it forces us to ask if we will be prepared when the end comes, because it will come for all of us, and the loss of a B-52 at Edwards AFB on June 15 is a stark reminder (https://www.aerotime.aero/articles/edwards-b-52-crash-new-details-eight-killed). On any given day tragedy may strike. An act of God may bring life to an unexpected end, and since none of us knows when the final curtain will fall, we must be prepared.

As a community, we collect data on mishaps and report on them as a way to prepare and stay prepared professionally, but I want to briefly describe several more steps we can take to prepare personally, ourselves, our teams, and our families, for tragedies like the one described.

Be prepared – spiritually and practically. Practically, this would include a will, life insurance, and a "legacy letter" that includes practical instructions for the executor of one's estate. This is a bare minimum. Many more things should be done.

Help others prepare. Talk about these kinds of things with your colleagues. Some are younger and need encouragement to get a will done. Others are older and have learned wisdom to cope with this kind of loss. Conversations will strengthen cohesion, and it may pave the way for deeper talks about more things.

Spiritual preparation isn't something we talk about often, and we probably should. "What does it mean to be safe?" is my brief attempt to address the topic, but I'd encourage anyone to send me an email directly if there is additional conversation wanted or needed.

Endnotes:
1. https://fromthegreennotebook.com/2015/06/07/three-truths-about-the-personal-study-of-war/ accessed 20260723. You should definitely read the whole article.

Critical Incident Response Program

The SETP Critical Incident Response Committee maintains resources and coordinates training that also serves to prepare members of our community to help others and provide peer support during times of tragedy. https://www.setp.org/services/circ.html