Earth Independent Medical Operations [EIMO]
DATASCOPE Technical Interchange Meeting 21st August 2023
Background and Summary of discussion

Jay Lemery[1], Mike Krihak[1,2], Kurt Berens[1,2]

[1]Exploration Medical Capability Element, NASA-Johnson Space Center, Houston, TX
[2]KBR, Inc., 2400 NASA Parkway, Houston, TX

Background

A brief overview of EIMO and the framing of 'datascope' was presented to participants:

EIMO has been defined as *the gradual transition of medical care and decision making from terrestrial to space-based assets, enabling support of astronaut health and performance and reducing overall mission risk*. While a hallmark of this paradigm shift from low-earth orbit is that on-board care will increasingly become the responsibility of the astronauts for primary management and decision making, terrestrial assets will continue to be paramount in pre-mission screening and planning, as well as prevention, health maintenance and long-term care contingencies. New capabilities and systems that enable progressively more robust and resilient systems and crews will be necessary to reduce risk and increase probability of deep space exploration mission success.

The constituent elements of EIMO include: 1) pre-mission planning; 2) acute and prolonged medical decision making; 3) supplies and resource management; 4) task load balance; all interfacing within novel data management and decision-making frameworks.

Given the broad intersection with numerous human spaceflight stakeholders, EIMO will interface with multiple groups within NASA including, but not limited to:

A. The Exploration Medical Integrated Product Team (XMIPT)
B. Crew Health and Performance Integrated Architecture Team (CHP-IDA)
C. Space Communications and Navigation (SCaN)
D. Autonomous Medical Operations (AMO)

An aspiration for EIMO datascope is to realize artificial intelligence-enhanced solutions for analysis of crew health & performance data and to facilitate clinical decision support for autonomous medical operations.

A vision proposed to the meeting participants was that of a "system of systems," whereby EIMO will utilize AI-supported natural language processing and machine learning techniques to synthesize embedded reference databases and real-time data streams [input vectors] from multiple data sources to continuously and seamlessly assess crew health & performance. Constituent input vectors may include environmental controls, countermeasures data, behavioral data, physiologic wearables, point-of-care laboratory tests, personalized medical records, inventory trade space risk assessments, COTS medical databases, and ground support inputs. An ideal AI capability would possess trained fusion algorithms to cross reference input vectors with medical 'knowledge' [cultivated database] to stratify relevant data streams for predictive and actionable capabilities. In addition, EIMO will ideally have a degree of mobility, in that it can be accessed and can push/pull data within and between multiple vehicles/habitats.

In flight data constraints were briefly introduced, with acknowledgement that greater precision of the projected amount of on-board data storage and computing capability that EIMO could be expected to need to function for CMO clinical decision support is a critical precondition for this endeavor. It is anticipated that an EIMO operating system would require very large datasets to adequately train pre-flight and significant amounts of data will be necessary to support machine learning via in-flight CDSS operations. An additional, perpetual challenge will be to find sufficient data to train a model relevant to astronaut demographics. Given the rapid, perhaps accelerating evolution of this field, there exists a propitious solution space in leveraging multi-modal artificial intelligence (AI) through public-private partnership(s)— the very participant demographic of this TIM.

EIMO DATASCOPE Meeting

On 21st August 2023, ExMC convened a panel of Subject Matter Experts (SMEs) from NASA, broadly representing the Human Research Program, Medical Operations, the Human Health and Performance Directorate (SA) at JSC, the Health and Medical Technical Authority, the Flight Operations Directorate, and representation from other Centers including: Ames Research Center (ARC), Glenn Research Center (GRC), Langley Research Center (LaRC) and Marshall Spaceflight Center (MSFC). In addition, SMEs from industry included representation from the following entities: Hewlett Packard Enterprises, IBM, Microsoft and WoltersKluwer. This group of stakeholders discussed the data issues related to facilitating EIMO.

A series of questions were posed to the participants and the most salient points of the conversation are summarized below.

We are seeking an in-flight integrated data architecture [IDA] to integrate numerous, disparate data sets, to assist with predictive crew health and performance analytics and clinical decision support [CDSS]. Is generative AI the 'way to go' when considering operating systems and for spaceflight IDA?

The discussion began with the need to better define the scope, *i.e.*, is the intent to create a comprehensive system capable of treating *every* conceivable medical condition or is there a "sweet spot" for acute and/or high consequence medical events under consideration? Also mentioned were the timelines of medical decision making and care considerations: where such AI capabilities would be justified for acute, time-sensitive medical conditions; perhaps less for chronic conditions and/or prolonged medical treatment lasting days or weeks, where ground support may fill in these gaps.

There was discussion surrounding the focus on generative AI and machine learning, specifically on its lack of symbolic reasoning for plan synthesis and plan execution tracking-- yet there was general agreement that generative AI had absolute applicability. Generative AI is currently inaccurate on net new content but it is very good at summarizing large amounts of information quickly and answering specific questions. A key feature for successful implementation would be establishing appropriate guardrails for training. The state of generative AI-based clinical decision support today is unreliable, however, the technology is evolving rapidly and may ultimately be deployable in a zero-tolerance environment. Large language models are deployable today and they are extremely efficient at parsing information.

Caution was expressed regarding unintended consequences of using predictive systems from an operations perspective. The location of computational activity was discussed and there was general agreement that if data could be transmitted to ground for validation, while delayed due to bandwidth and latency issues, this would be superior given the capacity differential between in-flight and ground computing resources. Concerns related to hardware and software development were noted. Most commercial "off-the-shelf" devices are likely to fail in the harsh radiation environment and a clinical decision support system software might be designated as Class A safety critical making development and deployment of such software much more difficult and expensive.

Final comments suggested that looking at solutions deployed in internet-deprived areas may be instructive as well as in areas where sufficient medical knowledge, skills and abilities exist but where resources may be a limiting factor, *e.g.*, analogue environments such as polar regional stations, submarines, saturation divers.

Acknowledging the aforementioned on-board data storage and computing capabilities re: generative AI, to what degree could dedicated, terrestrial LLM 'training' use cultivated datasets to bridge the gap re: AI data requirements?

Given the sparsity of data (and unknowns) in the space environment, what other considerations exist re: training models with limited data sets?

Anomaly detection using machine learning has been employed for hardware using two approaches, *e.g.*, clustering and/or other unsupervised techniques vs. prognosis of system failure. The "SIM to Real" gap (data sparsity) has highlighted data synthesis methods, *e.g.*, used to "fuzz" sensors, but such application to humans and health data is premature. Since data from humans in spaceflight is very limited, caution should be exercised to not "overfit" the model and training data should exclusively employ spaceflight data. A participant noted that analog data can be useful but questioned how much analog data would be needed to properly seed a generative AI model and would it be necessary to use the same crew for both analog and flight for the data to be valid.

An important point was raised to clarify why generative AI is needed in space. With a known history we can anticipate symptoms and track progression to a disease state as opposed to trying to predict disease outcome. There has been widespread acceptance that domain specific LLMs would always outperform general purpose LLMs but this is not the case in healthcare. The current approach is to dynamically present something (*e.g.*, UpToDate database) to a model and have it iterate over that data in real time as opposed to training a specific model again from scratch and this is referred to as Retrieval Augmented Generation (RAG). But LLMs and generative AI are only part of the picture as there is an important “indexing function” at play. The idea that these models have any sort of topical information is deeply flawed. We should be focused on how we present to the model the correct context every single time (as it evolves over time) so it cannot “color outside the lines” metaphorically speaking. Meta prompts (system level) hardwired running in the background can provide effective guardrails to reject out of scope requests, hallucinations, etc. In addition, training on the end use of such technology will be important to assure that end-users remain in control of the model.

Considering the paucity of options for re-supply and that pre-deployment (between 2-6 years in advance of mission) of resources and hardware is envisioned, challenges will exist to update data/software. Defining the size of the data pipeline and balancing both the dimension and choreography of bandwidth competition from other concurrent missions in distant theaters will be important. One option to reduce data load is using codification, *e.g.*, using SNOMED codes instead of text providing capability to reduce data stream seven-fold. Leveraging experience from industry (*e.g.*, remote pacemaker monitoring, ICU telemetry, home health care remote monitoring and personal fitness wearables as a few examples) could be beneficial by adopting efficiencies and best practices. Trade-offs between on-board computational capacity and data flow/bandwidth considerations will have to be made.

As this project evolves from concept to testing, what would a ground test environment [lab] look like, specifically re: metrics of success in integrating input vectors into a full-fledged integrated data architecture [IDA] that can both pull and push information to the end users?

What skillsets and levels of expertise should direct this work?

What AI modalities should be tested?

The ideal ground test environment should mimic as many features as possible in the mission environment. Key considerations involve engineering, avionics, software, human factors, interfaces, sensors, etc. It is important to note that items not certified for flight can be tested in this environment. A phased approach is essential and the testing should be schedule driven and not content/milestone driven. The IDA should remain upstream of the technologies to assure the right level of detail is derived. At some point, an assessment must be conducted to determine “how realistic is the analog?” or “how realistic does it need to be?” to be able to learn from the analog output. Keeping with the phased approach, initially no communications link is necessary (can be added later), and once added introduce communication delays, add in MCC, etc. as appropriate. It is desired to "work from the right" in service of a clear objective.

A "system of systems" approach to development aligns well with the Federated Lab architecture (similar to iPAS; Gateway Test & Verification is now "XLab"). Begin with small lab where two laptops are interacting while the system maturates. As other labs mature, they are interconnected and testing is ongoing at a regular cadence. Labs can be co-located or integrated over long distances via fiber. Two approaches to operating the lab include; a) build and break approach (Chaos Monkey, etc.) or b) establishing a flow [e.g.,1) have a list of "x" number things that need to be accomplished, 2) have "y" number of data sources to work with and, 3) apply the AI/ML of choice, evaluate performance and iterate if necessary]. Both approaches involve significant ongoing testing and iterative development. Deciding which AI/ML tool(s) might be best suited for this application is informed by a recent article in Nature Machine Intelligence co-authored by RT Scott, *et.al.,*T https://doi.org/10.1038/s42256-023-00617-5
Traditional machine learning could be used in a subset of the vector input flows quite nicely, but it is important to identify which are amenable and those that are not. Data modality is an important consideration, *i.e.*, LLMs work best with textual data but are incompatible with waveform data (*e.g.*, ECG). Not surprisingly, speech to text conversion quality is directly correlated with amount of computational effort applied. An AI lab that could emulate processing and storage constraints was also cited as a value-added proposition.

Future EIMO TIMs will focus on crew training, supplies & resource management, and task off-loading. What else should we consider re: this Datascope in meeting these other EIMO requirement domains?

Additional TIMs* to consider:

A. Shelf-life of medical resources (pharmaceutical, sterile water/saline for injection, syringes, etc. may be compromised due to length of mission and the plan to "pre-deploy" some materials well in advance (several years)
B. AI modalities especially in the setting of limited on-board resources
C. User Interface
D. Hierarchy of needs related to CDSS (most common diseases, most serious)
E. COTS medical devices, interoperability and integration
F. Autonomous mission operations, crew telemetry, equipment/supplies, cadence of monitoring

**Note: ExMC currently planning additional TIMs on Training, Resource Management and Task Offloading]*

We seek innovative solutions from industry to make sure we're capturing the cutting-edge ideas and technologies of today, to build systems that will fly over a

decade from now. To that end, what does this group think are the most effective ways to engage with thought leaders and innovators?

- ***RFI's?***
- ***Targeted conferences?***
- ***Interactive solution activities, i.e., Hack-a-thons?***

This project is well-aligned with healthcare industry interests and consideration of public/private partnerships could bring investment to accelerate activities to close mutually beneficial gaps in technology. Interoperability is paramount and NASA could take the lead in establishing standards that not only serve the space program's medical operations needs but could transform terrestrial medicine in the United States and around the globe. The proposed approach could fundamentally change the foundational way that the American healthcare system operates by transforming to a truly patient-centric system. If the foundations are properly laid, this effort will encourage investment and stimulate further innovation.

The soon to be launched Decadal survey from the National Academies (Life, Space and Physical Sciences) is looking to NASA for leadership with a focus on technology transfer for terrestrial applications.

Numerous conferences were highlighted that could provide opportunities to showcase progress and gain valuable stakeholder feedback. ExMC/NASA should consider attending the annual meeting of the Healthcare Information and Management Systems Society (HIMSS) and if possible be an exhibitor with a floor booth.

The Translational Research Institute for Space Health recently completed a competition for proposals to develop a centralized medical data architecture that should be designed to be vehicle agnostic and patient centric.

The meeting recap and summary provided guidance for the EIMO Project team. The participants indicated that the project would benefit from the development of a roadmap to define the project activities. The ECLSS SCLT just recently provided their updated roadmap and the EIMO roadmap should be reconciled to align with this road map as well as the overall HRP plan. When queried by ExMC Leadership as to whether multi-modal AI is aspirational or achievable, the group generally agreed that this is an achievable goal. Concerns were expressed that the development costs are currently quite expensive at scale and given the uncertainty surrounding the ability to monetize the deliverable for use beyond the space program it is unclear what role industry may play in development. While the current expense of developing such a system may be artificially high due to an ongoing scarcity of and market driven price escalation in chips, the supply chain disruptions and price of hardware has been steadily decreasing which should significantly improve return on investment. Building such a system was described as an effort similar to the genomics challenge from 20 years ago and that a market-image multi-modal AI system to support Earth Independent Medical Operations could be a reality within the next decade.

Conclusion

We are grateful to all the participants in providing their time and energy to these deliberations, and for providing further clarity to EIMO datascope and the overall direction of this endeavor.

A summary of takeaways:

- ***Multimodal AI****— there are many approaches to consider an EIMO-scale integrated data architecture system to support crew health and performance, and the consensus was that there is no AI modality that would serve as a single solution, but rather success would encompass an amalgam of AI modalities.*
- ***Trust gap—*** *AI shows tremendous promise but is still raw in many ways [ex: generative AI and hallucinations, confabulations etc.]. Much more work will be needed to understand the strengths and limitations of this approach, as well as safeguards that can be mitigated through AI training (reinforcement learning from human feedback; constitutional AI adherence, etc.), red teaming, or IT industry stress test practices (ex: Netflix Chaos Monkey)*
- ***Build or break—*** *while there was debate on lab testing through a 'breaking things' vs. 'build from the right' with intention, there was consensus that testing data integration of multiple input vectors [possibly via a Federated Lab process] with AI towards a) integration and b) data synthesis in support of CDSS would be invaluable. Metrics of success would not include the building of a data architecture product per se, but rather further understanding of multimodal AI mixes and 'best practice' approaches, to be disseminated through concepts of operations and requirements.*
- ***Industry partnership—****the AI landscape is moving so fast that obsolescence can now be measured in months, rather than years. Having industry partners to bring forth up-to-date technical approaches and innovations will maximize the utility of this endeavor and allow NASA to anticipate nimble upgrades to the extent possible as systems become codified and embedded in later Artemis and Mars mission planning.*
- ***More deliberations will be helpful—*** *Ongoing conversations on some grand challenges would be helpful to the EIMO team, specifically understanding the approach to AI modalities in the setting of limited on-board resources; optimal user interface(s), and medical device interoperability and integration. This latter point was deemed to have significant terrestrial applicability, a badly needed market solution to which NASA was ideally positioned to solve.*
- ***This is achievable—****Consensus was that this challenge had palpable solutions, especially in deciphering an integrated AI approach in the near term, rather than owning the burden of a fully functioning and vetted medical system.*

Attendees:

- *Schmitt, Brandon D. (ARC-SCF)[KBR WYLE SERVICES LLC]*
- *Tashakkor, Scott B (MSFC-ES52)*
- *Frank, Jeremy D. (ARC-TI)*
- *Berens, Kurt L. (JSC-SF211)[KBR Wyle Services, LLC]*
- *Scott, Ryan T. (ARC-SCR)[WYLE LABS]*
- *Suresh, Rahul (JSC-SD311)*
- *Fernandez, Mark R. (Hewlett Packard Enterprise)*
- *Delaune, Paul B. (JSC-EV211)*
- *Harrivel, Angela R (LARC-D318)*
- *Vazirani, Manish (Wolters Kluwer/UpToDate)*
- *Piontek, Nicole (LARC-E402)*
- *Sanders, Lauren M. (ARC-SCR)[Blue Marble Space]*
- *Faust, Kami M. (JSC-SF211)[KBR Wyle Services, LLC]*
- *Mullins, John (IBM Federal)*
- *Alibaruho, Macresia L. (JSC-SA211)*
- *Schkurko, Courtney M. (GRC-MSX0)*
- *Yu, Jane (Microsoft)*
- *Basso, Kailee K. (JSC-SF211)[KBR Wyle Services, LLC]*
- *Follett, Norman (Hewlett Packard Enterprise)*
- *Augustine, Philip M. (JSC-SF211)*
- *McCabe, Mary E. (JSC-EV311)*
- *Rucker, Michelle (JSC-EX211)*
- *Courtney, Michelle (JSC-EX211)*
- *Othon, William L. (JSC-MV111)*
- *Krihak, Michael K. (ARC-SCF)[KBR Wyle Services, LLC]*
- *Maddox, Ian (MSFC-HP40)*
- *Garrett, Shaina M. (JSC-SF211)[U.S. DEPARTMENT OF DEFENSE]*
- *Rice, Michael (Microsoft)*
- *Easter, Ben (JSC-SD311)[IPA]*
- *Costes, Sylvain V. (ARC-SCR)*
- *Fleming, Nancy K (JSC-SF111)*
- *Lehnhardt, Kris (JSC-SD311)[IPA]*
- *Lemery, Jay (JSC-SD311)[IPA]*
- *Choate, Andrew J. (MSFC-HP40)[ESSCA]*
- *Marchica, Andrea (GRC-MSX0)*
- *Vera, Alonso H. (ARC-T)*
- *Uohara, Michael (Microsoft)*
- *Cooper, Maya Renee. (JSC-SF311)[KBR Wyle Services, LLC]*
- *Burrell, Timothy (IBM Federal Health)*
- *Gebre, Samrawit G. (ARC-SCR)*

REFERENCES

1. Rodriguez-Ferrandez I, Tali M, Kosmidis L, Rovituso M, Steenari D, "Sources of Single Event Effects in the NVDIA Xavier ScO Family Under Proton Radiation", IEEE, 2022. https://upcommons.upc.edu/bitstream/handle/2117/376544/IOLTS_2022.pdf
2. https://rdcu.be/c8jSO
3. https://www.nature.com/articles/d41586-023-01907-z
4. https://osdr.nasa.gov/bio/
5. https://writings.stephenwolfram.com/2023/05/the-new-world-of-llm-functions-integrating-llm-technology-into-the-wolfram-language/
6. https://thenewstack.io/freshen-up-llms-with-retrieval-augmented-generation/
7. https://jamanetwork.com/journals/jama/article-abstract/2787207
8. https://ieeexplore.ieee.org/abstract/document/9084352
9. https://doi.org/10.1016/j.nbt.2012.03.004
10. Eric Topol's book: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again (Basic Books, 2019).
11. https://doi.org/10.1038/s41591-018-0300-7
12. https://www.drivendata.org/competitions/93/nasa-mars-spectrometry/page/437/
13. https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository
14. https://ai4sciencecommunity.github.io/neurips23.html
15. https://academic.oup.com/bioinformatics/article/36/4/1234/5566506
16. https://tivadardanka.com/blog/how-to-compress-a-neural-network

RECOMMENDED CONFERENCES

17. HIMSS Global Health Conference & Exhibition | HIMSS
18. https://www.pmwcintl.com/
19. https://ic3.center.ufl.edu/
20. https://ai4sciencecommunity.github.io/neurips23.html
21. https://amia.org/education-events/amia-2023-annual-symposium/schedule