

Innovation Machine: Human-LLM Teaming for Aeronautics Research and Discovery

Eric Reynolds Brubaker
Langley Research Center, Hampton, Virginia

George Moore
Langley Research Center / Massachusetts Institute of Technology, Cambridge, Massachusetts

Kevin R. Wheeler
Ames Research Center, Moffett Field, California

Jeffrey Chin
Glenn Research Center, Cleveland, Ohio

Shaunte' A. Kinch
Glenn Research Center / Empact Global, Wake Forest, North Carolina

Ty Hoang
Ames Research Center, Moffett Field, California

Thomas Mahota
Langley Research Center / University of Minnesota, Minneapolis, Minnesota

Anne L. Miller
Langley Research Center / University of California, Berkeley, California

Mark Stephen Meadows
Ames Research Center, Moffett Field, California

Vikram Shyam
Glenn Research Center, Cleveland, Ohio

James Villarrubia
Headquarters, Washington DC

April 2025

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

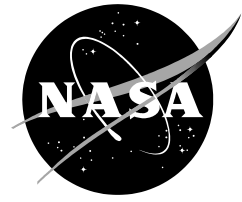
- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question to help@sti.nasa.gov
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199



Innovation Machine: Human-LLM Teaming for Aeronautics Research and Discovery

Eric Reynolds Brubaker
Langley Research Center, Hampton, Virginia

George Moore
Langley Research Center / Massachusetts Institute of Technology, Cambridge, Massachusetts

Kevin R. Wheeler
Ames Research Center, Moffett Field, California

Jeffrey Chin
Glenn Research Center, Cleveland, Ohio

Shaunte' A. Kinch
Glenn Research Center / Empact Global, Wake Forest, North Carolina

Ty Hoang
Ames Research Center, Moffett Field, California

Thomas Mahota
Langley Research Center / University of Minnesota, Minneapolis, Minnesota

Anne L. Miller
Langley Research Center / University of California, Berkeley, California

Mark Stephen Meadows
Ames Research Center, Moffett Field, California

Vikram Shyam
Glenn Research Center, Cleveland, Ohio

James Villarrubia
Headquarters, Washington DC

National Aeronautics and
Space Administration

April 2025

Acknowledgments

We would like to thank the many NASA researchers whose inspiring research, and potential recombinations of such, inspired this work. We also thank readers Priscilla Diem and Nicole Fichera for their suggestions that substantially improved this report. This work was funded by the Convergent Aeronautics Solutions Project and Transformative Aeronautics Concepts Program within the National Aeronautics and Space Administration.

Level of Review

This material has been internally peer reviewed within NASA.

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.
--

This report is available at: <https://www.sti.nasa.gov>, <https://ntrs.nasa.gov>, and from:

NASA STI Program / Mail Stop 050
NASA Langley Research Center
Hampton, VA 23681-2199

Innovation Machine: Human-LLM Teaming for Aeronautics Research and Discovery

Abstract

The future of aeronautics will be defined not just by emerging needs and capabilities in aeronautics but by the convergence between emerging areas in aeronautics and other sectors (energy, space, healthcare, agriculture, ecology, etc.). Identifying convergent, cross-sector areas with high transformative potential involves a degree of vision that even transdisciplinary teams can struggle to achieve. In this paper, we offer a set of tools for transdisciplinary teams to collaborate with large language models (LLMs) to co-emerge problem and solution areas across disciplines, sectors, industries, government agencies, and communities. These tools draw from theories of recombinant innovation in the field of innovation management and methods like the heuristic HIT matrix from the field of design theory and methodology. Additionally, we demonstrate new LLM and visualization tools, such as an Innovation Whitespace Generator and Transformative Potential Compass, using a dataset of “idea compost” from the NASA Aeronautics Research Mission Directorate’s (ARMD) Convergent Aeronautics Solutions (CAS) project. We close with limitations and suggested future work around improving these tools and understanding the biases embedded in them (e.g., biasing teams toward problems that advance dominant worldviews and stakeholders). Overall, our goal is to create LLM tools that facilitate collaboration and translation between people who have different expertise and lived experiences so as to enable, not inhibit, the human collaborations needed to identify convergent problems worth solving and transform the future of aeronautics for the benefit of all humanity.

Table of Contents

1. Introduction	1
2. Background.....	2
Recombinant Innovation & Technology Brokering	2
Human-LLM Teaming.....	3
3. Our LLM Chatbot: MADI	4
Technical Infrastructure	4
Operational Efficiency and Cost Benefits.....	5
Collaborative Innovation Architecture.....	5
Knowledge Discovery and Evolution	5
Advanced Human-LLM Collaboration	6
4. Idea “Compost”: Recycling Ideas to Build New Fertile Soil.....	6
5. LLM Tools for Recombinant Innovation.....	7
Recombinant Innovation Use Cases for NASA Aeronautics	7
Potential Methods of Recombinant Innovation with LLMs.....	10
Whitespace Generator: LLM-Generated Ideas in Potentially Unexplored Intersections.....	13
Whitespace Visualizer: Network Graphs and Clusters	13
Sankey Diagram: Recombining Skillsets	15
6. Transformative Potential Compass: Automated Idea Assessment Tool.....	15
Motivation for Measuring Transformative Potential	15
Case Study Compost: CAS 2024 Wicked Wild Pitches	16
Frameworks for Assessing Transformative Potential	16
Measuring Transformative Potential with Strasser’s Dimensions of Transformative Impact	18
Testing Transformative Potential Instruments	23
Creating a Transformative Potential Framework through Deduction.....	28
Recombining Wicked Wild Pitches using HIT Matrix	31
7. Next Steps for Recombinant Innovation and LLMs at CAS Discovery	34
Key Insights and Takeaways.....	34
LLM Applications at CAS	34
Exploring Recombinant Innovation Methods	34
Measuring Transformative Potential	35
Limitations.....	36
References	37
APPENDICES	42
Appendix A.....	42
Appendix B	43
Appendix C	46
Appendix D	49
Appendix E	51
Appendix F	53

1. Introduction

Humanity faces many complex societal challenges—some longstanding such as those outlined in the UN’s sustainable development goals (UN General Assembly, 2015), and some emergent, such as increasing wildfires, extreme weather events, and disparities in healthcare access between rural and urban communities. These challenges pose both threats and opportunities for the future of aeronautics. The NASA Convergent Aeronautics Solutions (CAS) project identifies such challenges and works to address them through transdisciplinary, cross-agency research and development (Brubaker et al., 2025 Shyam et al., 2023). For example, the CAS Scalable Traffic Management for Emergency Response Operations (STEReO) effort began in 2019 by principal investigators Joey Mercer and Marcus Johnson, among others, by asking the question: *What if advancements in aeronautics could improve emergency response?* Several years later this work became the NASA Advanced Capabilities for Emergency Response Operations (ACERO) project, which benefits from a collaboration between NASA, CAL FIRE, and the U.S. Forestry Service, among others, to advance uncrewed traffic management (UTM) and uncrewed aerial systems (UAS) technologies while improving situational awareness and asset coordination among wildland fire fighters (Smith, 2023). More recently, in 2023, a CAS team explored the question: *What if advancements in aeronautics could improve access to healthcare as distance to care increases?* This has become the NASA CAS AirMed Transportation for Rural Areas & Coordinated Scheduling System (TRACS) effort, a collaboration between NASA automated scheduling and planning researchers, physicians, hospital administrators, airport administrators, and municipal and tribal leaders in New Mexico, to develop an integrated, partially automated interfacility patient transfer system that integrates healthcare and rural airport data to improve rural patients’ timely access to lifesaving levels of health care.

The future of aeronautics offers many potential opportunities to address convergent problems across disciplines, sectors, industries, agencies, and communities, although identifying the strongest of these requires a degree of vision that even transdisciplinary teams can struggle to achieve. To aid this process, we have developed a set of tools for transdisciplinary teams to collaborate with large language models (LLMs) to co-emerge problem and solution areas that have increasing levels of transformative potential. These tools draw from theories of recombinant innovation and technology brokering from the field of innovation management (Basalla, 1990; Gilfillan, 1970; Hargadon and Sutton, 1997 Hargadon, 2002) which describe how novel innovations do not typically emerge from nothing but from a novel recombination of existing ideas. In developing our first set of LLM recombinant innovation tools, we drew from theory in human-computer interaction and methods like Heuristic Ideation Technique (HIT) matrix (Tauber, 1972) and SCAMPER (Eberle, 2005) from design theory and methodology. The HIT matrix compares the characteristics of two seemingly unrelated problem formulations, technologies, products, services, or systems to develop new ideas at their intersection thus helping to generate many new ideas (Tauber, 1972). SCAMPER—(S) substitute, (C) combine, (A) adapt, (M) modify, (P) put to another use, (E) eliminate and (R) reverse—is a set of heuristics intended to help solve problems and promote creativity during brainstorming (Eberle, 2005). Building upon commercially available LLMs, we created HIT matrix and

SCAMPER plug-ins and deployed these internally for the NASA innovation community. This led us to ideas for new human-LLM methods, including an Innovation Whitespace Generator and Transformative Potential Compass. This paper describes this set of LLM innovation tools and demonstrates them using a case study dataset of “idea compost” alongside resulting visualizations from the NASA CAS project. This case study dataset is a small subset of the larger NASA idea compost and is focused on problem formulations rather than solutions, though we acknowledge that formulations of problems and solutions are highly intertwined (Dorst and Cross, 2001; Dorst, 2019; Maher, Poon and Boulanger, 1996). We provide a preliminary demonstration of human-LLM recombinant innovation using the HIT matrix and Transformation Potential Compass resulting in a variety of problem statements ranked by their relative transformative potential as assessed across multiple dimensions. The present paper closes with limitations and suggestions for future work to improve the proposed human-LLM tools and characterize the biases embedded in them (e.g., biasing teams toward problems that advance dominant stakeholders and worldviews). Overall, we aim to provide new human-LLM tools that further supercharge human expertise by combining it with AI capabilities. Such capabilities are essential to the aeronautics innovation ecosystem as we collectively work to transform the future of aeronautics for the benefit of all.

2. Background

This work draws from theories of recombinant innovation and technology brokering from the field of innovation management, human-LLM collaboration from the human-computer interaction field, and various methods from design theory and methodology, strategic foresight, and related field. We describe background on these literatures below.

Recombinant Innovation & Technology Brokering

The concepts of technology brokering and recombinant innovation have been well described in both innovation practice and research. The invention of ships, for example, stemmed from a novel combination of bouyant materials (e.g., floating logs), carpentry, steam engines, and propellor technologies (Gilfillan, 1970). Thomas Edison’s lab frequently recombined technologies from disparate domains to invent new ones, like the phonograph which resulted from blending telegraph, telephone, and electric motor technologies (Hargadon, 2003). Similarly, modern design consultancies use their reach into the multiple industries to engage in “technology brokering” to facilitate or gatekeep access to ideas across industries and suggest novel recombinations (Hargadon & Sutton, 1997). For example, Reebok consulted with Design Continuum who offered the idea of combining shoe construction with inflatable medical splits and IV bag technologies to invent the Reebok Pump shoe. Even transformational inventions technologies like the automobile and microprocessor were recombinations of existing ideas—an automobile recombines bicycle, horse carriage, and internal combustion engine technologies, and a microprocessor recombines the large central processing units of early computers with integrated circuit fabrication processes. In this way, invention can be viewed as a recombinant search process to identify innovative combinations and associated configurations of constituent technologies.

Mednick (1962) proposed a Theory of Creativity based on the idea that creativity is the ability to make non-obvious connections between seemingly unrelated things. The further unrelated those “things” seemingly are, the more creative the ideas will be. Since then, studies have reinforced Mednick’s theory, as well as the notion that generating high numbers of ideas during ideation sessions leads to a more creative output (Kudrowitz and Wallace, 2013). As a result, many ideation methods have been developed and aimed to capitalize on this correlation by promoting large quantities of ideas, with the understanding that most of the ideas will be average or poor, and a few highly creative outliers will be worth pursuing. Martin and Hanington’s Universal Methods of Design (Martin, 2012), Kumar’s 101 Design Methods (Kumar and Hackett, 2013) and Lewrick, Link, and Leifer’s Design Thinking Toolbox (Lewrick, Link and Leifer, 2020) are three popular resources that offer many ideation methods for organizations, teams, and individuals. Techniques range from the well-known methods like brainstorming and focus group studies to the more obscure methods such as ELITO and NABC methods (Alrutz, Singer and Wahlig, 2002; LeBlanc and Hassan, 2017). A thread through each of these methods listed is that they are intended to be used and performed by humans. Due to their ability to rapidly search and make connections across large information spaces, AI agents are increasingly being used to aid in creative practice. We now turn to the rapidly growing literature on using AI agents as partners in design ideation and assessment.

Human-LLM Teaming

Early-stage ideation and discovery takes place through a variety of exercises, frameworks, and activities that are intended to elicit creative ideas from team members. Large Language Models (LLMs) have proven valuable in ideation and discovery and some scholars have begun to distinguish the nature of collaboration that has emerged between humans and LLMs (McComb, Boatwright and Cagan, 2023). While integrating LLMs into the discovery process introduces new opportunities for generative and evaluation capabilities, negative perceptions of LLM performance and inherent bias of LLMs are challenges to address. Research shows that human-AI teams can perform as well as human-human teams when the AI is generating content informed by rule-based heuristics (Xu et al., 2024; Gyory et al., 2022). Other notable research results from AI as a design collaborator include improving a team’s exploration of the solution space and supporting the consideration of design alternatives (Song et al., 2022; Kwon, Huang and Goucher-Lambert, 2022; Kwon, Rao & Goucher-Lambert, 2023; Lee et al., 2024). On the other hand, researchers have noted that in some cases, AI improves individual creativity while reducing collective creativity (Doshi and Hauser, 2024). In this report we explore some of these strengths and weaknesses that have been reported as significant trends in human-AI collaboration. Specifically, we explore generative AI’s strength of outputting large amounts of information and generative AI’s opportunity for improvement in evaluating creativity in a way that a human expert would. As creativity researcher Barry Kudrowitz suggests in his book *Sparkling Creativity*, “it takes creativity to determine that you have had a creative idea, but it also takes creativity to determine which ideas to pursue, which ideas to save for later, and which ideas to discard” (Kudrowitz, 2023).

3. Our LLM Chatbot: MADI (Modular AI for Design and Innovation)

MADI (Modular AI for Design and Innovation), our intelligent knowledge management system represents a strategic response with NASA to the growing need for secure, scalable artificial intelligence solutions within the federal space. By providing a unified platform that combines enterprise-grade security, modular extensibility, and advanced collaboration capabilities, MADI enables NASA's workforce to leverage AI technology while maintaining compliance with federal regulations and maximizing return on investment through shared infrastructure and knowledge accumulation. Figure 1 illustrates a recent build of the MADI interactive interface.

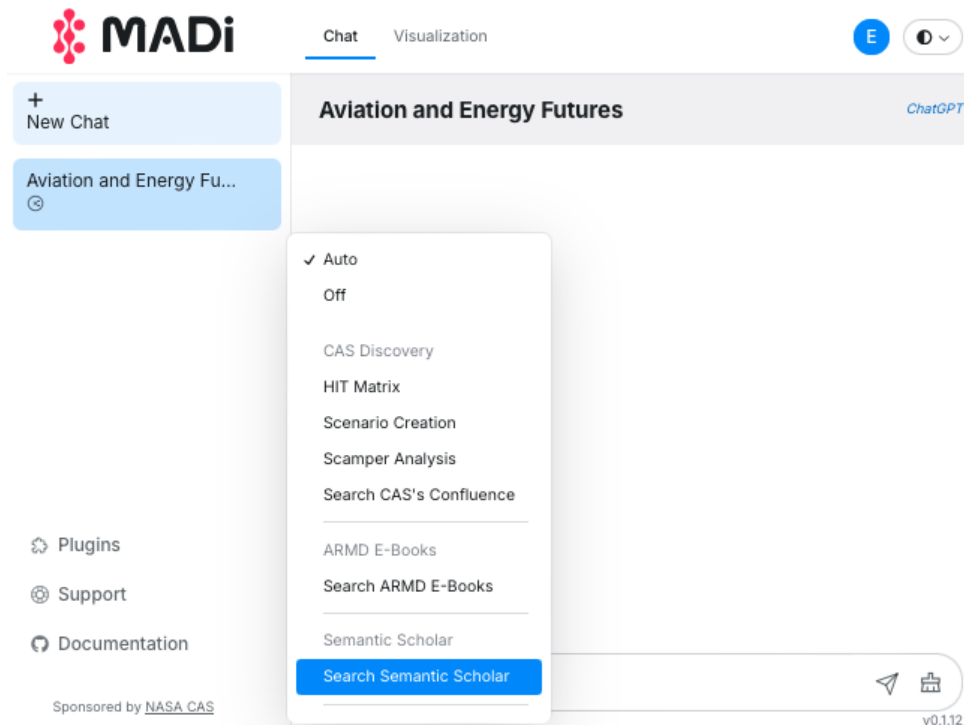


Figure 1: Internal NASA MADI chatbot interface

Technical Infrastructure

MADI's system architecture comprises a Node.js and Python backend, PostgreSQL database implementation, and React-based frontend, deployed via serverless services in Google Cloud Run (see Figure 2). The system currently operates through NASA's Cloud Compute Services Google Cloud Platform (GCP) instance, maintaining compliance with federal security protocols through NAMS authentication systems, with development underway to transition to a Model Garden hosted within GCP or Azure's government cloud infrastructure to ensure FISMA compliance. The system's codebase adheres to open-source licensing standards (MIT, ISC, BYD-3) for all components except cloud-native services and AI models, facilitating potential adoption by other federal agencies with similar authorization to operate (ATO) requirements.

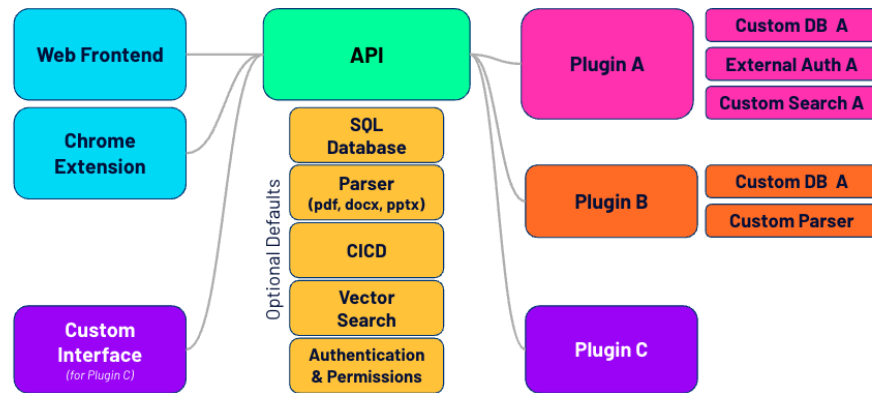


Figure 2: Initial MADI modular architecture

Operational Efficiency and Cost Benefits

Initial deployment metrics demonstrate substantial efficiency improvements in research and ideation processes. The Convergent Aeronautics Solutions (CAS) Discovery team reports a 50% reduction in brainstorming session duration, translating to approximately 1200 FTE hours saved annually for a ten-person team. The system's serverless architecture yields operational costs of approximately \$300 per month in fixed expenses, with variable costs of \$0.05 per conversation. This cost-effective structure enables NASA teams to implement LLM solutions without significant infrastructure investments or technical expertise requirements, representing a marked improvement over traditional development approaches that often require substantial FTE expansion.

Collaborative Innovation Architecture

MADI's plugin architecture enables NASA customers to integrate proprietary data and custom code while leveraging shared infrastructure, significantly reducing deployment costs and technical barriers. The platform's modular design supports Python scripts, JavaScript modules, and containerized applications, facilitating rapid development of domain-specific solutions. This architectural approach eliminates redundant development efforts across NASA centers while maintaining security and compliance standards, enabling teams to focus on innovative applications rather than infrastructure development.

Knowledge Discovery and Evolution

As MADI's usage expands across NASA's enterprise, its capability for knowledge discovery grows exponentially through the accumulation and analysis of diverse data streams. The system's ability to identify connections between disparate research areas and highlight potential white spaces in NASA's research portfolio increases with each interaction. This evolutionary capacity enables MADI to serve as both a knowledge management tool and an innovation mechanism, supporting the identification of unexplored research opportunities and potential cross-disciplinary collaborations.

Advanced Human-LLM Collaboration

MADI's distinctive approach to human-LLM collaboration extends beyond traditional chat interfaces through integrated data visualization capabilities, enabling both textual and visual exploration of complex datasets. This dual-modality interaction framework supports more intuitive exploration of scientific concepts and research findings. As the system continues to evolve, ongoing research into human-LLM collaboration patterns will inform the development of increasingly sophisticated plugins and interfaces, potentially establishing new paradigms for research and innovation that exceed current private sector capabilities. This positions MADI as a pioneering platform for advancing both NASA's specific mission objectives and broader understanding of effective human-LLM collaborative frameworks. MADI is used regularly by a variety of NASA teams and has been demonstrated at NASA events. Figure 3 shows a group using MADI at the 2024 CAS Expo to input aeronautics and societal trends and envision potential future scenarios.



*Figure 3: Groups using MADI at CAS Expo 2024 hosted by NASA Glenn Research Center.
Image credits: NASA / Sara Lowthian-Hanna.*

4. Idea “Compost”: Recycling Ideas to Build New Fertile Soil

Idea “compost” is a compilation of past research, including ideas that worked, didn’t work (for a particular application), or are yet to be fully explored. We first learned of this term from Alphabet X, “the moonshot factory,” when talking with them about their innovation process and how they actively manage, mix, and recombine ideas from an internal “compost” to generate new innovations. In recent years, CAS has been compiling its own internal idea compost, starting with efforts that move through CAS Discovery. We define compost as a collection of in-progress, paused, stopped, completed, or even disregarded research regarding problems or solutions that might be combined with other research to generate new, innovative ideas.

CAS Compost currently includes CAS Discovery Opportunity Concept Reports and proposals from CAS’s Wicked Wild Problem Exchange—a NASA wide call for proposals

of problems that could transform aeronautics—among other outputs from past and present CAS Discovery teams. Opportunity Concept Reports are outputs of Discovery sprints that define compelling problems and opportunities for new, potentially transformative technologies and sociotechnical systems that address them. Discovery sprints are short term by design, with limited budget and resources and, like many innovation efforts across NASA, generate lots of compost at varying levels of maturity. Compost is a natural byproduct of innovation efforts, but outside of the memories of those involved, these past ideas are not often saved or structured in a way where they can be brought back up at the opportune times.

There is a much larger set of compost across NASA and the Aeronautics Research Mission Directorate. We have begun to assemble pieces of this larger set with inputs from CAS's sister projects in the Transformative Aeronautics Concepts Program (TACP)—Transformational Tools & Technologies (TTT) and University Innovation (UI)—and other sources like NASA Aeronautics Research Mission Directorate Aero E-books, Semantic Scholar, patent databases, NASA TechPort and more. When paired with a compost of increasing size and diversity, we expect the recombinant innovation tools presented in this paper to become increasingly powerful.

5. LLM Tools for Recombinant Innovation

Recombinant Innovation Use Cases for NASA Aeronautics

We initially characterized recombinant innovation use cases into four potential categories: recombinant solutions, recombinant problems, recombinant problem-solution pairs, and recombinant teaming (Table 1). Recombinant solutions may be the most common in the recombination and technology brokering literature and involves combining multiple seemingly unrelated technologies or solution elements to create a new innovation. For example, the Gutenberg printing press, a transformational innovation of its time, was a recombination of moveable type symbols from Chinese block printing with a European wine/olive press. Recombinant problems involve combining multiple problem formulations to identify new, potentially more impactful, root cause-oriented problem formulations. For example, we might combine the longstanding problem of air transportation supply chain disruptions with the challenge of a global pandemic to reformulate the problem around critical underlying fragilities in the supply chain system. Recombinant problem-solution pairs involve applying an existing solution to a new problem or vice versa, such as applying bioinspired anti-adhesives to the challenge of aircraft de-icing to identify long lasting and environmentally benign de-icing solutions. The final type of recombination we considered is recombinant teaming which involves suggesting subject matter experts or other people with near or distant expertise relative to a problem or technology area to join early-stage innovation teams or workshops. This may improve innovation outcomes by suggesting more novel and unexpectedly synergistic combinations of team members, what is commonly called the Medici Effect (Johansson, 2004).

Table 1: Four Types of Recombinant Innovation

Recombinant Solutions			Recombinant Problems		
Combining multiple seemingly unrelated solutions to create a new innovation			Combining multiple problem formulations to identify more transformative problems		
+ =			+ =		
Moveable type symbols from Chinese block printing	European wine/olive pres	Gutenberg's Printing Press	Existing supply chain fragility	Global pandemic	Critical supply chain fragilities
Recombinant Problem-Solution Pairs			Recombinant Teaming		
Applying an existing solution to a new problem, or vice versa			Suggesting people with near or distant expertise to join teams or workshops		
+ =			+ =		
Bioinspired adhesives/ anti-adhesives	The challenge of de-icing wings and rotors	Long lasting, environmentally benign de-icing solutions	A problem or solution space	A database or network of people & their experience	Suggestions for people with near or distant expertise for a team or workshop

To explore use cases for recombinant innovation in the NASA Aeronautics innovation ecosystem, we held two workshop sessions with NASA stakeholders, including chief technologists and leaders and researchers from across the NASA Transformative Aeronautics Concepts Program. This resulted in several dozen potential recombinant innovation use cases that originally took the form: “We wish to input [theme] and combine it with [recombinant element] to [achieve desired output].” These use cases included the broad categories of identifying new areas of research (new “whitespace”), recombining existing problem formulations or solution ideas, recombining problem-solution pairs, recombining people or organizations, and other recombinant use cases. As a result, 22 use cases emerged as detailed in Table 2.

Table 2: NASA CAS/TACP Recombinant Innovation Use Cases

Identifying New Areas of Research: Idea Convergence and “Whitespace”		
Goal/Context	Potential Recombinant Inputs	Desired Output
Identifying “whitespace” for new research areas	Input a wide set of past NASA research (proposals, agreements, papers, patents and reports from CAS, UI, TTT, ARMD mission projects, IRAD, etc.)	Visualizations highlighting existing research and LLM-suggested whitespace
Benchmarking: characterizing past work and potential whitespace	Input an idea, problem, or solution/technology area and cross reference it with a wide set of other NASA internal and external materials	Relevant prior work inside or outside NASA
Identifying research area convergence across NASA efforts	Input a wide set of past NASA research (CAS Wicked Wild and Discovery proposals/reports, TTT and UI proposals, NASA AIAA, patents and TechPort entries in particular area, etc.)	Previously unforeseen emergent/convergent research areas and potential technical interchanges
Identifying research area convergence between NASA and non-NASA efforts	Input a wide set of past NASA and non-NASA research (available proposals, papers, patents, industry reports, etc.)	Previously unforeseen emergent/convergent research areas and potential technical interchanges

Recombining Problems or Solutions

Goal/Context	Potential Recombinant Inputs	Desired Output
Identifying solutions to help overcome a barrier in a given technology or economic, environmental, ethical, sociopolitical, or legal landscape	Input a barrier to addressing a problem area and cross reference it with a wide set of internal and external materials (as described above) or a targeted body of literature (AI, build-measure-learn best practices, additive manufacturing, etc.)	New concepts or “solution vectors” that could help overcome the given barrier
Identifying new solutions to a given problem	Input a problem area or specific problem statement and cross reference it with a wide set of internal and external materials	New concepts or “solution vectors” that could address the problem
Identifying new solutions to a given problem that are not yet technologically feasible but could become so in the not too distant future	Input a problem area/statement and future time horizon then cross reference a wide set of internal and external materials	New concepts or “solution vectors” that could address the problem alongside the expected technology breakthroughs required to make it possible
Gathering materials to help assess the potential for success of a given solution idea	Inputting an idea (e.g., ingesting and sequestering engine emissions in flight) and combining a large body of relevant published literature (aircraft weight-benefit tradeoff data, reports on current aircraft emissions, etc.)	An assemblage of materials that help researchers assess an idea’s potential for success

Recombining Problem-Solution Pairs

Context/Goal	Desired Recombinant Inputs	Desired Output
Identifying new applications for a given research idea or innovation	Input a research area/technology idea and cross reference it with a wide set of internal and external materials (as described above)	Previously unforeseen applications for a technology idea, including specific, industries, use cases, organizations or people
Identifying existing technologies that may be applicable to a given stakeholder need or application	Input stakeholder needs/desires/context and cross reference it with a wide set of internal and external materials	Previously unforeseen technologies or research areas that could address the provided needs/desires

Recombining People or Organizations

Context/Goal	Desired Recombinant Inputs	Desired Output
Identifying potential NASA researchers/team members who have expertise that could be near or far from a given research area	Input a research area and combine with NASA staffing databases, AIAA publications with NASA affiliations, NASA patents, TechPort entries, etc.	Suggested NASA researchers as potential new team members
Identifying potential collaborators, workshop participants, etc. for a given research area	Input a research area or problem/solution idea and combine with a body of existing papers and proposals inside and/or outside of NASA	Suggested potential collaborators, within and outside of NASA
Identifying potential subject matter experts and potentially affected stakeholders	Input a research area or problem/solution idea and combine with a large body of published work, including papers, blogs, youtube videos, etc. inside and outside of NASA	Suggested subject matter experts or stakeholders to interview, visit, and/or engage in co-design
Identifying potential funders and/or transition partners for a given solution idea	Input a research area or specific solution idea and combine with a wide set of existing papers, proposals, funding/award databases, industry reports and investor reports	Suggested potential funders, investors, large companies, small companies, non-profits, gov agencies or other potential transition partners
Identifying surprise topic areas via researchers in less-connected fields	Input a network of NASA researchers and their publications and combine with a wide set of internal and external materials (as described above)	Suggested new research areas that converge less connected NASA researchers with new problem or solution areas
Identify new potential technical interchange or coalitions	Input a set of proposals that were not awarded across various projects/programs and combine with a wide set of internal and external materials	Suggested new research areas and combined research teams that converge multiple unawarded ideas

Other Recombinant Use Cases

Context/Goal	Desired Recombinant Inputs	Desired Output
Identifying trends in qualitative data	Input a collection of qualitative data (papers, reports, transcripts, etc.)	Visualizations of trends, outliers, etc. in the data

Identifying causal loops and leverage points within complex systems/systems of systems	Input a collection of descriptions of the dynamics of complex sociotechnical systems (from reports, papers, transcripts, etc.)	Causal loop diagram visualizations and potential leverage points (building from Hosseinichimeh et al., 2024 and Meadows, 1999)
Identifying value propositions for transition from CAS to a NASA mission project, industry, other government agency, or other potential partner	Input a problem statement from a CAS Execution Agreement and combine this with a technical challenge from a ARMD project and a variety of soft inputs from personal conversations	Persuasive arguments for the given CAS activity to transition to the potential target entity
Supporting focused strategic planning	Input a research area (e.g., airspace transformations) and combine it with a goal or other body of work (e.g., sustainability)	A description of the potential stakeholders, interactions, degrees of freedom, and FOMs/optimizations
Supporting strategic roadmapping	Input strategic foresight intelligence (signals, trends, drivers) plus future scenarios and combine these with other materials (e.g., sustainability goals, existing roadmaps, total aircraft and airport operational emissions, material usage, waste products, energy needs, and economic drivers)	Delineated multi-agency roadmaps
Suggesting blended design methods for a given design scenario/goal	Input a design scenario/goal and a set of design case studies from other organizations, frameworks, and methods (foresight/speculative design, human-centered design, traditional systems engineering and design, participatory design, transition design, etc.)	Suggestions for new blended design frameworks and methods for the given design scenario/goal

Potential Methods of Recombinant Innovation with LLMs

LLMs have the potential to contribute in many ways across the NASA ecosystem. Their ability to enhance the performance of design methods is no exception. Hence, CAS Discovery began exploring OpenAI's ChatGPT 3.5 as a tool to enhance traditional Design methods, with the overarching goal of generating transformative ideas for society. The first step involved aggregating design methods from popular innovation and design methods resources, such as Martin and Hanington's Universal Methods of Design, Kumar's 101 Design Methods, and Lewrick, Link, and Leifer's The Design Thinking Toolbox, Design Methods Finder (<https://www.designmethodsfinder.com>), and Innovation Wiki (<https://www.innovation.wiki/en/>). The next step involved experimenting with ChatGPT to generate useful outcomes through these design methods. This meant leveraging several strategies such as role-playing prompting, chain-of-thought prompting, or instruction prompting.

Table 3 lists all of the design methods that were explored and provides feedback on the user experience when engaging LLMs to execute the method.

Project summaries from NASA's Patent Portfolio (<https://technology.nasa.gov/patents>) were used in conjunction with the design methods from

Table 3. SCAMPER and HIT Matrix are examples of recombinant design methods that were explored further. SCAMPER is an acronym and it works by completing the activities associated with each letter – Substitute, Combine, Adjust, Modify, Put to other uses, Eliminate, and Reverse (Dam and Siang, 2024). ChatGPT is useful in being able to quickly output many new perspectives/ideas.

Table 3: Collection of design methods attempted using an LLM (ChatGPT 3.5)

Method	Use	Pros	Cons	Engaging an LLM
SCAMPER	Helps improve existing ideas/products by applying prompts	Easy to use, encourages creative thinking	Can be limiting if used too rigidly	Prompt the LLM with a product and ask it to apply each SCAMPER prompt to generate variations.
HIT Matrix	Combines unrelated ideas to spark innovation	Promotes novel combinations, good for group ideation	May generate impractical ideas	Ask the LLM to combine two or more unrelated concepts and generate a list of hybrid ideas.
Affinity Mapping	Organizes ideas based on similarity or theme	Great for synthesizing data, finding patterns	Time-consuming, subjective grouping	Provide a list of ideas and ask the LLM to cluster them into thematic groups with labels.
Brutethink / Random Stimuli	Introduces randomness to break habitual thinking	Stimulates lateral thinking, fun	Can feel forced or irrelevant	Request the LLM to generate a list of random words or scenarios to spark unexpected connections.
TILMAG	Uses triggering questions and attributes for idea generation	Structured yet creative, helps reframe problems	Less well-known, requires facilitation	Describe the product and ask the LLM to identify attributes and suggest logical modifications.
Worst Possible Idea	Generates deliberately bad ideas to flip into good ones	Low pressure, encourages bold thinking	Can be unproductive without guidance	Instruct the LLM to list intentionally bad ideas and then ask it to reframe them into viable concepts.
Backcasting	Envisions a future goal and works backwards to plan steps	Goal-oriented, strategic	Requires clear future vision, may ignore current constraints	Describe a future scenario and ask the LLM to work backward, outlining steps and considerations.
BlockBusting	Identifies and breaks mental blocks to creativity	Encourages deep reflection, good for individuals or teams	Can be hard to identify blocks without facilitation	Ask the LLM to identify common creative blocks and suggest ways to overcome them in a given context.
Cognitive Mapping	Visualizes thought processes or mental models	Helps understand user cognition and problem framing	Can be abstract, hard to standardize	Feed the LLM user insights and ask it to build a cognitive map or thought model.
Concept Mapping	Shows relationships between concepts	Great for structuring knowledge	Can become messy or complex	Give the LLM a central idea and ask it to expand a concept map showing connected ideas and links.
Design Fiction	Explores future scenarios through storytelling	Engages imagination, highlights ethical issues	Can feel speculative or disconnected from reality	Prompt the LLM to write short stories or narratives exploring speculative product futures.
Empathy Mapping	Understands user needs, feelings, and behaviors	Human-centered, builds empathy	Relies on assumptions if user data is weak	Provide persona or interview data and ask the LLM to complete an empathy map.
Focus Groups	Gathers user opinions in a group setting	Rich qualitative data, user-driven insights	Risk of groupthink, dominant voices	Give the LLM a transcript or notes from a session and ask it to summarize key themes and sentiments.
Gap Analysis	Identifies the gap between current and desired states	Clarifies needs and priorities	Can oversimplify complex problems	Describe current and ideal states, and ask the LLM to outline the gaps and possible solutions.
Metaphors	Uses analogies to understand or reframe a problem	Engages imagination, reveals hidden insights	May not translate well across contexts	Ask the LLM to create analogies or metaphors for a product, idea, or user experience.
Role-Playing	Explores experiences by acting out scenarios	Builds empathy, tests interactions	May feel awkward, needs facilitation	Describe a scenario and ask the LLM to generate role-play dialogues or perspectives from different users.
Rose, Bud, Thorn	Identifies positives (rose), potentials (bud), and negatives (thorn)	Simple, effective reflection tool	Can be superficial if rushed	Give the LLM an idea or project and ask it to list Roses (positives), Buds (potentials), and Thorns (challenges).

HIT Matrix

The Heuristic Ideation Technique, better known as the HIT Matrix, emerged as a popular strategy for recombinant innovation (Tauber, 1972; Ruijter, 2016). This strategy provides a framework for recombining attributes of known ideas to create new solutions. This can be done with any elements of the innovation process – such as problem statements, solutions, or insights. Figure 4 exemplifies the HIT Matrix strategy when applied to NASA innovation activities. Also, refer back to Figure 3 to see a group engaged in a LLM-enabled HIT Matrix activity. After several iterations of trial and error, a prompt was designed for ChatGPT to populate a HIT matrix based on some input of projects, technologies, or problem spaces specified by the user. Sample output from ChatGPT using the HIT Matrix is shown in Appendix A.

The HIT matrix prompt has been trialed in a variety of settings, including most recently in an interactive activity at the 2024 CAS Expo at the NASA Glenn Research Center. ~8 groups of 5-10 NASA personnel (engineers, designers, managers, and more) spent 15 minutes manually ideating and filling out a HIT Matrix with projects from the CAS Compost. Next, the teams ran the same Compost input through the HIT Matrix Prompt using MADI (NASA’s LLM, similar to ChatGPT), and continued ideating for another 15 minutes. Many of the groups reached a point in the first 15 minutes where they had exhausted their initial ideas and so the introduction of MADI and the ideas that came from the prompt helped the teams refresh their perspectives and spurred more ideas.

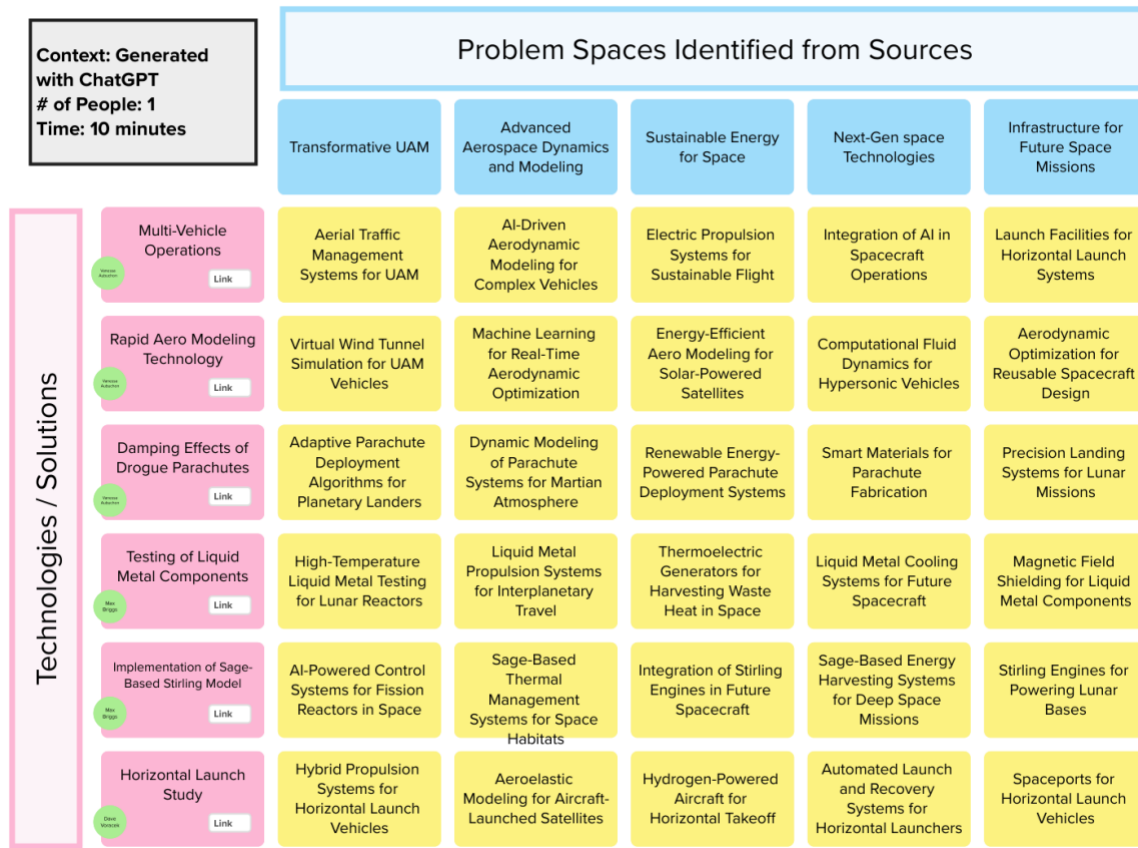


Figure 4: HIT Matrix strategy applied to NASA innovation activities

Whitespace Generator: LLM-Generated Ideas in Potentially Unexplored Intersections

The whitespace generator involves generating new prompts that fill out the empty areas within the semantic problem and solution space of the existing CAS Compost. The purpose of this was to identify areas in this space that have been underexplored and demarcate points on a graph representation of the context space that could be used to generate new recombinant efforts.

The whitespace generation pipeline involved a loop with three iterations. The first loop was the extraction phase. This involved a simple prompt supplying the data from the compost that would be used to generate the whitespace and a query to pull out a certain number of themes. In this case 10 were chosen. The second iteration in the loop was to use those keywords or themes generated in the first iteration and further expand on them to define another set of 6 ideas from each of the initial 10. Overall, another 60 data points were returned here. The last iteration was the same as the second, except (1) the data supplied here were the output from the second loop and (2) only three ideas were generated for each of the 60 ideas from iteration 2. These ideas were meant to be distinct concepts or themes connected to the keyword that they were being generated from. These ideas could then be plotted and used as a way to demarcate “whitespace” between concepts that already existed in the data from previous research efforts and be used to create a starting point visually for whitespace analysis.

The generative activities and exercises described in this section (HIT Matrix, SCAMPER, and whitespace generation) have been acknowledged as helpful for strategic foresight and generating transformative solutions. However, there are few examples of quantitative metrics that determine the degree in which their contributions are significant. Hence, developing assessment tools to better understand how these activities and frameworks actually support our mission to determine transformative potential will be critical moving forward.

Whitespace Visualizer: Network Graphs and Clusters

The network graph is calculated by creating representational embeddings for the text description of each prompt or node seen in this graph. This representational embedding is generated from the prompt and the description of that prompt from a dataset of previous efforts and their outcomes. Links between nodes are then determined by calculating the cosine similarity between each node and then only shows links that meet a specific threshold of similarity. This similarity threshold can be updated by the user and starts out with a threshold of 0.85. The visual has the source encoded as the node shape and the color is representative of the maturity that the effort reached.

This visual helps to see the interconnections and clusters of ideas but doesn’t show the projections of ideas in the space (Figure 5-7). In this visual, distance has no meaning and is just a representation of the nodes and links being forced apart by the diagram and the randomness associated with generation. A variation of this visual has been combined with whitespace generation to display the whitespace demarcated ideas in connection the nodes that are present and how they can be utilized for new recombinant efforts.

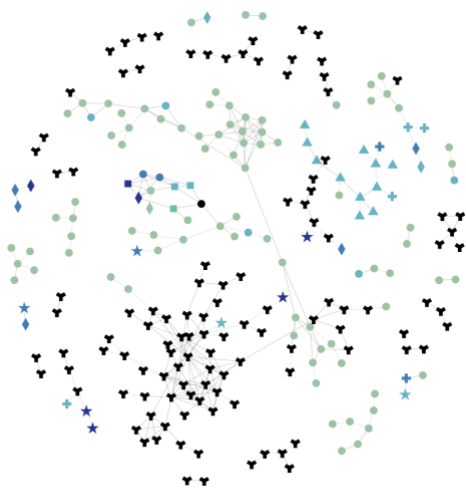


Figure 5: Whitespace Visualizer - without whitespace clustering

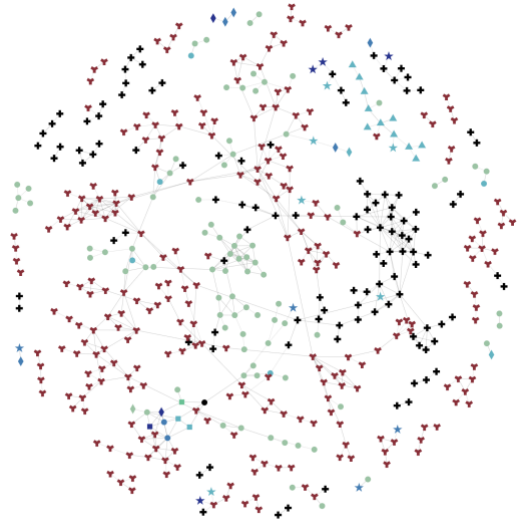


Figure 6: Whitespace Visualizer - with whitespace clustering

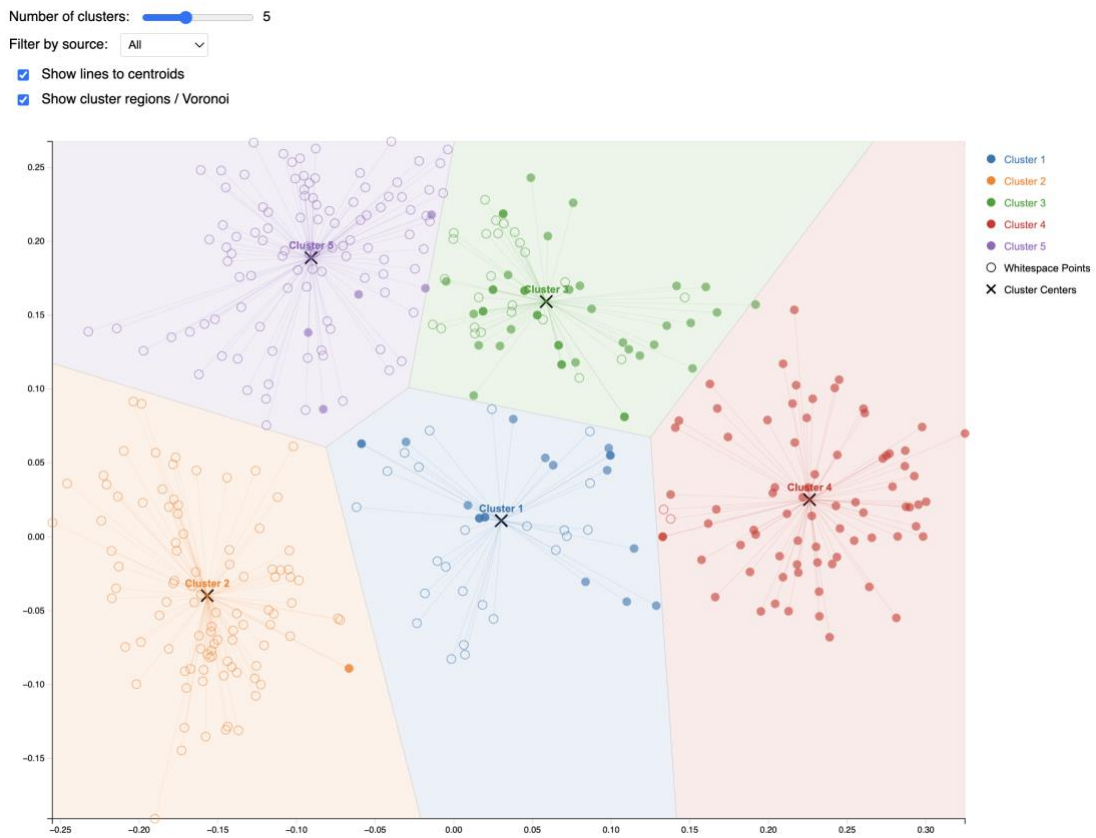


Figure 7: Whitespace Visualizer - Scatter plot including voronoi regions and calculated centroids

Sankey Diagram: Recombining Skillsets

Separate from the idea space, diverse and recombinant teaming can also be key to innovation by broadening the groups experience base. Approaching problems from different perspectives and backgrounds also helps with organizational robustness and creates a larger starting network for building momentum for tackling challenging problems. Sankey diagrams are popularized diagrams for showing flows and mapping from many-to-many domains. By mapping the workforce by center to competency to project, teams can ensure a balance in teaming across the organization and program portfolio (Figure 8). Further tracking and labeling of the team composition allows AI to further aid in balancing this portfolio diversification. Studying how the diversity in teaming impacts the eventual project outcomes is also a potential avenue for further study.

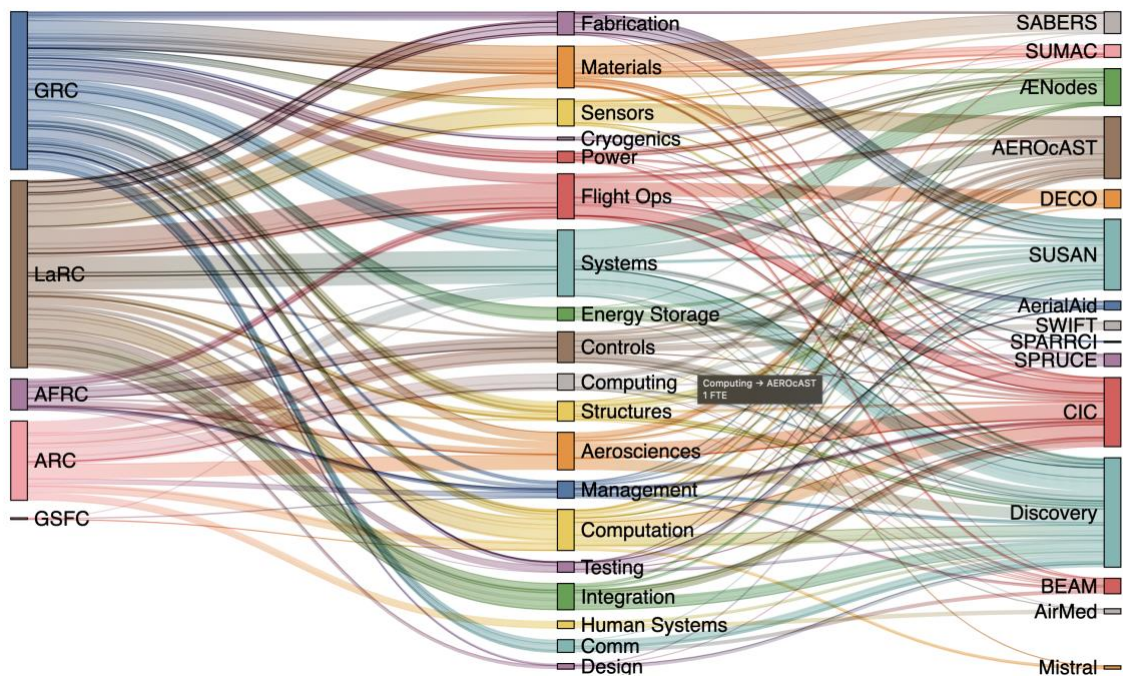


Figure 8: Sankey Diagram - Recombining Skillsets

Tracking the flow of pursued topics, from left to right depicting the proportion of NASA center involvement, competencies, and research activity titles. "Discovery" encompasses all early-stage exploration.

6. Transformative Potential Compass: Automated Idea Assessment Tool

Motivation for Measuring Transformative Potential

A priority of CAS Discovery is to identify opportunities with high transformative potential, such as problems worth solving or ideas worth pursuing. While using recombinant innovation and LLMs to generate and populate these opportunity spaces takes significant time and effort, the capability to do so faster and with less effort is becoming more

available. However, evaluating these opportunities is a more complex challenge, especially since the nature of many novel opportunities requires a variety of expertise to engage with. If done well, the task of evaluating transformative potential is very laborious. Within CAS, selection of opportunities for further investigation is based on executive decision making, which means that in addition to the expert knowledge and capacity to engage with each opportunity, there exists the challenge of fine-tuning subjective judgement. In this section are examples from CAS Discovery's efforts to measure transformative potential – quantitative feedback that, over time, can be used to sharpen the intuition of executive decision makers.

Case Study Compost: CAS 2024 Wicked Wild Pitches

This section features a subset of the CAS Compost, the 2024 Wicked Wild (WW24) pitches. The WW24 dataset includes 34 pitches from the 2024 NASA wide Wicked Wild Problem Exchange. Each pitch includes a title, author, description of the problem to be solved, relevant stakeholders, estimate of reach or number of people that it will directly impact within a year, the impact that solving this problem will have on a per-person basis, risks and barriers associated with pursuing a solution, an estimated timeline for pursuing this problem, the tipping point – when this problem will reach its climax, trends associated with this problem, capabilities required to solve this problem, this problem's relevance to aviation, NASA resources that might be leveraged to solve this problem, second and third order effects of this problem, identification of the gap or “whitespace” associated with this problem, and an optional slide deck for those interested in presenting their pitch at the Wicked Wild Problem Exchange event. A copy of the actual WW24 questionnaire can be found in Appendix B.

Frameworks for Assessing Transformative Potential

Transformative potential is best defined as the value assigned to early-stage concepts regarding their ability to transform the world for better. Within the context of CAS, the following attributes are considered when evaluating transformative potential:

- **High Impact:** New focus areas for ARMD (e.g., the new UTM, EAP, UAM...) that accelerate or identify new s-curve and change the game, disrupting current approaches (frame breaking).
- **Large Scale or Cross-Sector:** On order-of-magnitude of the U.S. aero industry or spanning multiple sectors (e.g., healthcare systems and aviation systems or other system of systems).
- **Meaningful Need:** Addresses deeply felt societal or industry needs, whether well known and expressed or hidden and unmet.
- **Future Potential.** Identifies a future state with potential to substantially inform the future of ARMD (direction/research portfolio/programs-projects) and aeronautics (traditional and emerging industries).

- **Realistic Timeframe.** It is possible for CAS and partners to address the critical barriers (i.e., desirability, viability, and feasibility) within a several year rapid assessment period in CAS and thereby unlock a development roadmap that makes the transformative concept possible, even if implementing the concept will inevitably take longer.

Our first step towards building foundations for measuring transformative potential involves exploring frameworks that have been used within innovation contexts. The next several sections detail the use of new and existing frameworks to measure transformative potential.

RICE (reach, impact, confidence and effort); D, V, F (desirability, viability and feasibility); the Defense Advanced Research Projects Agency (DARPA) Hard Test; and Strasser’s Dimensions of Transformative Impact are known frameworks for evaluating ideas and have been use within CAS Discovery. These frameworks have their shortcomings. For example, the “reach” element from RICE caters to ideas that help large volumes of people, which may not necessarily include ideas that address the highest need. Despite their shortcomings, these frameworks offer a starting point for our team to build and refine instruments to assess transformative potential in a way that benefits CAS Discovery.

A key feature of these frameworks involves defining the scale that each dimension of value is assessed against. Table 4 provides as example of the dimensions and scale definitions for the DARPA Hard Test.

Table 4: DARPA Hard Test

Value dimension	Definition	End of scale (1)	Top of scale (7)
I. Far-Reaching	The solution requires a completely new mental model, passing through a paradigm shift	Requires no change in how people think about a solution	Requires a paradigm shift in how a solution is viewed across society
II. Technically Challenging	The goal is difficult to implement either in terms of inventions or system integration required	Requires no new specialized expertise	Requires major advancements in specialized expertise
III. Multidisciplinary	The solution requires multiple bodies of knowledge that rarely exist within one industry	Requires only one class of knowledge	Requires multiple distinct bodies of knowledge
IV. Actionable	The right people can see a path to the impossible and can make progress beginning today	Requires so much clarification that the next step is another meeting	Requires little effort to begin taking actions towards the solution

Frameworks vary in the kind of content that they are most compatible with evaluating. For example, some are better for examining opportunity spaces while others solution spaces.

Our team focused on problem spaces, as this approach appeared more novel and less explored than the solution space.

In the next couple of sections, we discuss our approach to build instruments that measure transformative potential beginning with leveraging Strasser's Dimensions of Transformative Impact (Strasser, De Kraker and Kemp, 2020). Then, inspired by other existing frameworks and incorporating CAS interests, we develop our own framework, the CAS Compass, for measuring transformative potential.

Measuring Transformative Potential with Strasser's Dimensions of Transformative Impact

An LLM prompt and response visualization were developed to analyze 34 research proposals within NASA's Convergent Aeronautics Solutions Project "Wicked Wild" problem exchange. This involved consuming the proposals before and after they had been transformed into a standard problem statement format. The formatting was much more uniform but had many additional categories that were generated by MADI. See Appendix C for the prompt used to format these proposals.

Strasser et al.'s three dimensions of transformative impact were used as part of the basis of the LLM prompt used to evaluate each WW24 pitch (Strasser, De Kraker and Kemp, 2020). The dimensions consist of the structural and cultural embeddedness (depth), the wide-spread and coherent influence (width), and the persistent and evolving reproduction (length). This framework will be referred to as Strasser's DWL framework and the dimensions will be referred to as depth/embeddedness, width/influence, and length/persistence. The LLM prompt (see Appendix D) consisted of definitions of these three dimensions, and a summary of the NASA's ARMD's (Aeronautics Research Mission Directorate) strategic plan including 6 strategic thrusts and 3 mega-drivers. As context, the LLM was provided with the topic titles and problems statements for 2024 Wicked Wild pitches. The LLM was then instructed to create scores for the depth/embeddedness, width/influence, and length/persistence and to rank the proposed ideas.

The transformative impact scoring amongst problem statements is relative rather than absolute. It is beneficial and necessary to process all the problem statements together as a whole rather than in a piecemeal fashion. These relative scores were spread out after generation to help span the space from 1 to 10.

This analysis involved looking at the WW24 problem statements that had been specifically formatted as prompts and problem statements. The formatting was done by MADI in order to present uniform problem statements with distinct categories of information. The resulting sum of square scores using the raw descriptions and the processed uniform problem statements is shown below.

Figure 9 shows the sum of square scores for raw text descriptions and text that had been uniformly formatted as problem statements for the Wicked Wild 2024 ideas.

The unformatted text data caused the prompt to plateau for approximately the last 20 topics. The problem statement formatting helped the scoring prompt to better distinguish the scores. A 3-D visualization of the 2024 data is shown in Figure 10 below.

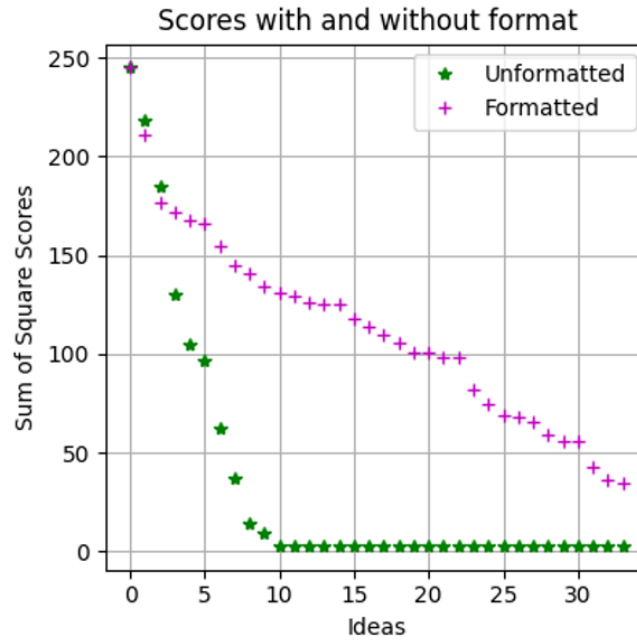


Figure 9: Sum of square scores for unformatted and formatted 2024 Wicked Wild pitches (Strasser's framework).

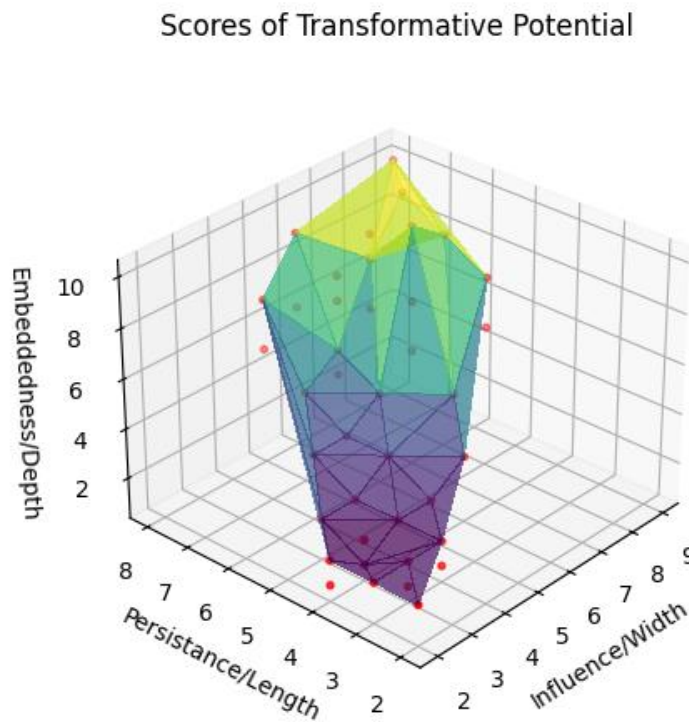


Figure 10: 3-D visualization of Transformative Potential scores for formatted text using Strasser's framework

The top topics between scoring the unformatted descriptions and the formatted problem statements is depicted in the following Table 5. The scores for embeddedness, influence and persistence are shown after each topic. The takeaway from this result is that formatting the descriptions dramatically changed the outcome of the scoring. The curious aspect is that the topics that had very short descriptions were the topics that asked “How” questions. These questions and the inequality of description lengths probably skewed the prompt responses.

Table 5: Text rankings for top 5 topics for unformatted and formatted topics.

Rank	Unformatted	Formatted
1	Active Alleviation of Ice; 10;8;9	How might we improve airborne communication;10;8;9
2	Airbridge to nationwide quantum connectivity;9;7;8	How might we enable trust in AI/ML applications;9;9;7
3	ARM;8;6;7	Airbridge to nationwide quantum connectivity;8;7;8
4	Automated Microstructure Processing Lab;7;5;6	ARM;6;10;6
5	Autonomous data-driven decision;6;4;5	Autonomous data-driven decision;7;6;9

In order to understand this result, the topics that started with the words “How might we” were removed from the input and the analysis was rerun. The sum of the square scores for this are presented in the next Figure 11.

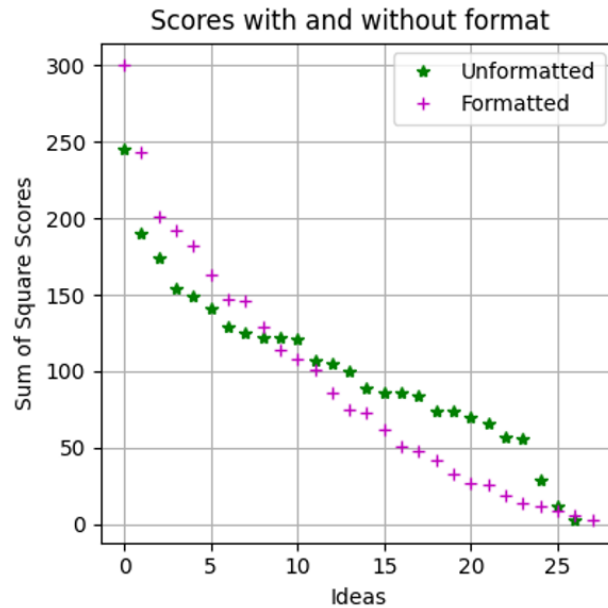


Figure 11: Sum of square scores for ‘24 Wicked Wild pitches without “how might we” text (Strasser’s framework)

Table 6 below shows the scorings and the difference in topics selected.

Table 6: Top 5 ideas for unformatted and formatted topics without the “How might we” topics.

Rank	Unformatted without “How might we”	Formatted without the “How might we”
1	Active Alleviation of Ice; 10;4;3	Airbridge to nationwide quantum connectivity; 10;10;10
2	Airbride to nationwide quantum connectivity; 3; 10; 9	Climate Adjusted Air Transportation; 9;9;9
3	ARM; 4;8;2	Earth Emergency Response Vehicle 8;8;8
4	Automated Microstructure Processing Lab;6;7;6	Securing our Information-Dependent Future;7;7;7
5	Autnomous data-driven decision; 8;3;1	Autonomous Navigation 6;6;6

The analysis of the unformatted versus the formatted text with the “How might we” questions removed still resulted in a lack of agreement between the two analyses. The scoring, scaling the scoring, and ranking are the most difficult tasks within this prompt. It is apparent that the scores for the unformatted case are not as well spread out as the formatted case. However, many of the scores in the formatted case had all three dimensions with the same score.

Instead of examining the top scoring topics, sometimes it is easier to eliminate the weakest proposals. When the weakest topics are compared, the bottom scoring ten proposals tend to agree between the unformatted and formatted text analyses. The formatted and unformatted results agreed on 8 out of 10 of the bottom proposals.

Homegrown Dimensions of Transformative Impact: CAS Compass.

In our mission to create a transformative impact framework for CAS, we started by identifying 20 known frameworks for evaluating ideas (Kim and Mauborgne, 2004; Ogata, 2004; Hillson and Hulett, 2005; Grierson, 2008; Parker and Moseley, 2008; Leigh, 2009; Xu *et al.*, 2009; Van Beurden *et al.*, 2013; Lopes, Oliveira and Bana E Costa, 2015; Carlgren, Rauth and Elmquist, 2016; Jyothi and Parkavi, 2016; Sharpe *et al.*, 2016; Cote, 2022; Brush, 2023; Manning, 2023; Morita *et al.*, 2023; Mikhail, 2024). We organized the frameworks along a spectrum ranging from those best suited for addressing problems to those best suited for developing solutions, with some categorized as neutral. We identified 13 independent criteria used for the frameworks best suited for problems and neutral. This was the beginning of our transformative potential framework for assessing problem spaces.

After substantial iteration and analysis, we landed on six dimensions for assessing the transformative potential of problems: (1) Risk of the problem, (2) Clarity of need, (3) Complexity of the problem, (4), Perceived probability of the problem, (5), Duration of the problem, (6) Impact of the problem. These six dimensions of value assessment together form the CAS Transformative Potential Compass (Table 7). The CAS Compass framework is used by considering each value dimension along a 5-point Likert scale where 1 represents minimal transformative potential and 5 represents exceptional transformative potential.

Table 7: Transformative Potential Compass for Problems

	1 Minimal	2 Low	3 Moderate	4 High	5 Exceptional
Risk of the Problem	Minimal risk; the problem poses little threat if unaddressed.	Low risk; the problem has some negative consequences.	Moderate risk; the problem could cause significant harm if not addressed.	High risk; the problem poses serious threats to well-being or operational capacity.	Critical risk; the problem has potentially catastrophic consequences.
Clarity of Need	The need is vague and poorly understood.	The need is recognized but not well-defined.	The need is clear to some stakeholders but not universally acknowledged.	The need is widely recognized and well-articulated.	The need is universally acknowledged and precisely defined, with a consensus on its urgency.
Complexity of the Problem	Simple; the problem is straightforward and easily understood. It no significant interactions with other systems.	Mildly complex; the problem has some layers but is generally comprehensible. The problem has minor interactions with other systems.	Moderately complex; the problem involves multiple factors and stakeholders. The problem is interconnected and part of a system with notable feedback loops.	Highly complex; the problem is multifaceted with interconnected elements. The problem is a key component of a complex system with multiple feedback loops.	Extremely complex; the problem is deeply systemic, involving numerous variables and interdependencies.
Perceived Probability of the Problem	Perceived improbable; the problem is perceived unlikely to occur or become significant.	Low perceived probability; there's a slight chance the problem will become relevant.	Moderate perceived probability; there's a fair chance the problem will occur or escalate.	High perceived probability; the problem is very likely to occur or is already happening.	The problem is certain to occur and requires attention.
Duration of the Problem	Transient; the problem is temporary and likely to resolve on its own.	Short-term; the problem is expected to last for a limited time.	Medium-term; the problem has been persistent but is not perpetual.	Long-term; the problem is expected to last for an extended period.	Permanent; the problem is enduring, with no end in sight.
Impact of the Problem	The problem affects a limited number of individuals or a small, localized area with low severity . Its impact is relatively minor, and those affected have adequate resources and resilience to manage the issue without significant disruption to their well-being or daily activities.	The problem impacts a specific group, organization, or community with moderate severity . While the effects are noticeable, they do not substantially hinder the overall functioning or quality of life of those affected. The issue may disproportionately affect less vulnerable populations who have some capacity to address or mitigate the problem.	The problem affects multiple groups, communities, or an entire sector with significant severity . It may lead to considerable challenges, disruptions, or inequalities, particularly among vulnerable or marginalized populations. Addressing this problem would result in meaningful improvements in well-being and equity for those impacted.	The problem has widespread effects across regions, industries, or large populations with high severity . It exacerbates systemic inequalities and poses substantial threats to the health, stability, and prosperity of diverse and vulnerable groups. Resolving this issue would lead to extensive positive change and promote greater social and economic equity.	The problem has critical and profound impacts on a global scale, affecting populations across multiple countries or the entire world. It poses severe risks to fundamental human rights, safety, and sustainability, with the most devastating effects often borne by the most vulnerable and marginalized communities. Effectively addressing this problem would result in transformative and enduring improvements in global equity, justice, and well-being.

Developing the CAS Compass offered our team another framework to evaluate the 2024 Wicked Wild pitches from CAS Wicked Wild Problem Exchange competition. Figure 12

illustrates how dimensions from Strasser’s Transformative Impact and CAS Compass frameworks assign value to each wicked wild pitch.

	Strasser’s DWL			CAS Compass					
WW pitches	Dim 1	Dim 2	Dim 3	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
pitch #1									
pitch #2									
...									
pitch #N-1									
pitch #N									

Figure 12: Structure of Strasser’s Transformative Impact and CAS Compass dimensions

Testing Transformative Potential Instruments

The first step in testing transformative potential instruments (Figure 13), processing the Wicked Wild pitches into a uniform format, was done using MADI. The second step involved prompting MADI to assess each Wicked Wild pitch (unformatted and formatted) according to the dimensions of Strasser’s DWL framework and the CAS Compass framework. Finally, using the scores generated by MADI, principal component analyses (PCA) and independent component analyses (ICA) were performed.

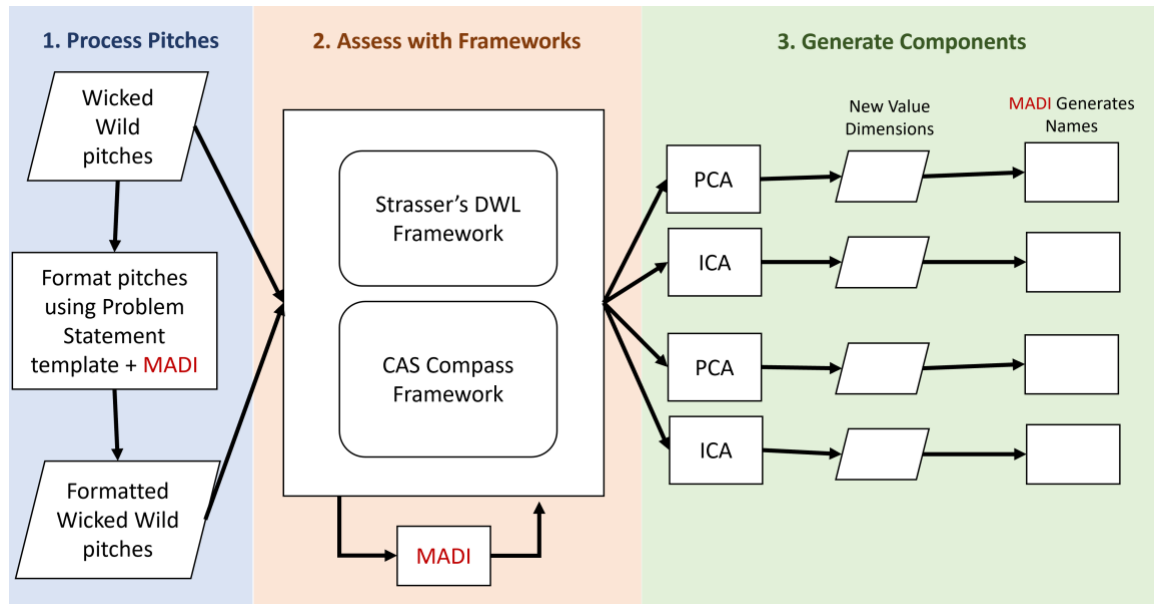


Figure 13: MADI workflow for automated assessment of transformative potential for Wicked Wild concepts.

Figure 14 compares the LLM assessment (using the CAS Compass) vs Human assessment of unformatted 2024 Wicked Wild pitches. The y-axis shows the pitches, ordered from top to bottom based on an averaged human ranking. The tip of the arrow shows the LLM chosen ranking, so smaller arrows equate to better agreement on rank ordering. Leftward pointing arrows indicate that the LLM chose a higher relative ranking for a topic. There is

not any notable agreement at either the high, mid, or low end of the rankings, but a good starting point for future comparisons. Variability between individual human reviews was also high, so this could serve as a single additional data point in future averages. The LLM also showed a clear positive bias towards LLM and automation related submissions. Past events have shown slightly better agreement on the lowest relevance concepts relating to transformative potential, however there is clearly ample room for improvement. Even now, the LLM scores can help catalyze conversation and identify or confirm possibilities for misconception.

Human vs AI ratings

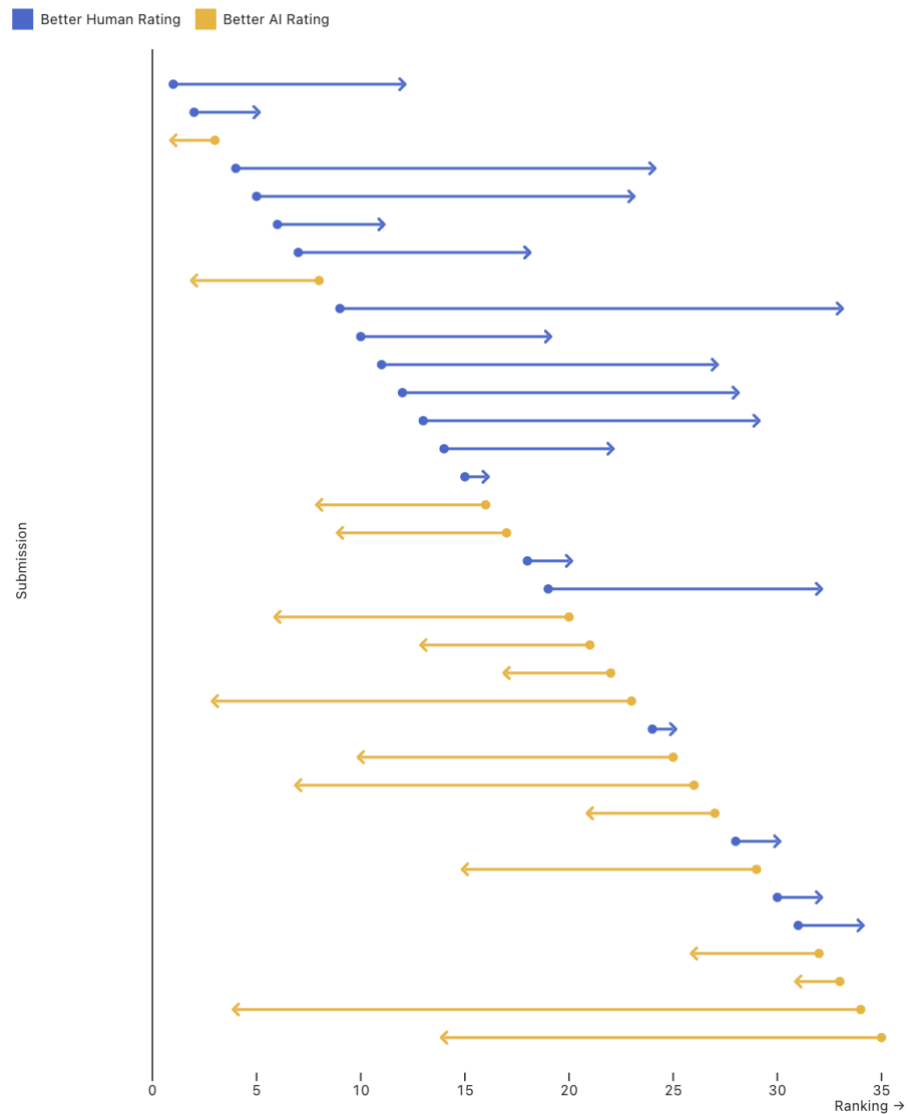


Figure 14: Human vs LLM evaluation of 2024 Wicked Wild Pitches

Principal component analysis and independent component analysis were used to determine (a) which value dimensions and (b) how many value dimensions best explain the ratings

that the LLM assigned to the Wicked Wild pitches. As we establish more trust in LLMs for evaluation of data for transformative potential, methods like PCA and ICA, allow CAS to optimize for a context-specific definition of transformative potential that is consistent with how innovation content is evaluated. Tables Table 8 and Table 9 provide the PCA results for Strasser’s DWL framework and the CAS Compass independently and collectively.

Table 8: PCA results for 2024 Wicked Wild Pitches using Strasser’s DWL framework

Value Dimension	Variable Contributions	
	unformatted	formatted
persistent and evolving reproduction	34.17%	34.53%
structural and cultural embeddedness	31.23%	32.74%
wide-spread and coherent influence	34.6%	32.74%

Table 9: PCA results for 2024 Wicked Wild Pitches using CAS Compass framework

Value Dimension	Variable Contributions	
	unformatted	formatted
Clarity of Need	16.97%	19.37%
Complexity of the Problem	15.83%	20.95%
Duration of the Problem	16.05%	0.00%
Impact of the Problem	16.56%	19.40%
Perceived Probability of the Problem	17.81%	20.89%
Risk of the Problem	16.78%	19.38%

The major takeaway from these tables is that the “duration of the problem” dimension had no impact on variance in formatted 2024 Wicked Wild pitches. This means that all of the pitches received the same rating for this dimension. A few plausible reasons for this outcome include the following: (a) the “duration of the problem” is not a valuable dimension to consider, (b) this data set is skewed towards pitches that have the same “duration of the problem” attributes, (c) the LLM did not develop an understanding of “duration of the problem” so that it could discern between pitches that should be scored lower or higher, or (d) the data set is too small and does not capture the range of pitches that would show variance along this dimension.

Other important products of the PCA and ICA analysis include principal components and independent components. The principal components and their respective explained variance are provided in Tables Table 10 and Table 11. Prompts used to generate these names are provided in Appendix E.

Table 10: Top Principal Component results for 2024 Wicked Wild Pitches using Strasser's DWL framework

PC	Explained Variance		LLM Generated Name	
	unformatted	formatted	unformatted	formatted
PC1	56.7%	64.7%	Visionary Integration vs. Tactical Incrementalism	Impact Horizon Spectrum
PC2	23.8%	22.4%	Visionary Integration vs. Technical Detail	Pioneering Risk vs. Operational Evolution

Table 11: Top Principal Component results for 2024 Wicked Wild Pitches using CAS Compass framework

PC	Explained Variance		LLM Generated Name	
	unformatted	formatted	unformatted	formatted
PC1	29.0%	35.6%	Visionary Integration vs. Narrow Focus	Catalysts vs. Optimizers
PC2	24.9%	21.8%	Visionary Integration vs. Conceptual Immaturity	Transformative Safety Innovations vs. Technical Feasibility Challenges
PC3	21.0%	17.9%	Innovative Technology Integration	Transformative Impact vs. Risk Management

The main takeaway from these two tables is the difference in generated names between formatted and unformatted Wicked Wild pitches. For the unformatted pitches, the LLM seems to be fixated on “Visionary Integration” as a theme that must be included in all new principal components. For the formatted pitches, the LLM provides more distinct and nuanced names for the resulting principal components. Because the purpose of the PCA, in this context, is to identify new value dimensions, these early results suggest that formatted pitches make it easier for the LLM to offer more useful value dimensions for our CAS Discovery team to consider.

Next, the independent components and their respective explained variance are provided in Tables Table 12 and Table 13. Prompts used to generate these names are provided in Appendix E.

Table 12: Principal Component results for 2024 Wicked Wild Pitches using Strasser's DWL framework

IC	Kurtosis		LLM Generated Name	
	unformatted	formatted	unformatted	formatted
IC1	0.387	2.157	Avant-Garde Aeronautics vs. Quantum Connectivity Leap	Innovative Technological Frontiers
IC2	-1.281	-0.626	Innovative Aeronautics and Quantum Frontiers	Strategic Urgency vs. Seasonal Adaptation
IC3	-1.134	-1.174	Innovative Aeronautics Solutions	Innovative Safety and Connectivity Frontiers

Table 13: Principal Component results for 2024 Wicked Wild Pitches using Strasser's DWL framework

IC	Kurtosis		LLM Generated Name	
	unformatted	formatted	unformatted	formatted
IC1	7.294	-1.169	Innovative Aviation Frontiers	Innovative Technological Frontiers
IC2	2.880	-0.986	Innovative Aviation and Energy Solutions	Innovation Risk and Implementation Urgency
IC3	-1.364	-1.919	Innovative Aeronautics and Robotics Frontier	Innovative Technological Frontiers
IC4	3.362	-0.960	Innovative Aeronautics and Automation Exploration	Innovation Frontiers and Operational Integration
IC5	-1.289	2.644	Pioneering Innovation and Risk Acceptance	Innovation Acceleration vs. Operational Efficiency
IC6	-1.043	--	Quantum Connectivity & Automated Innovation Frontier	--

Compared to the results of the PCA, this ICA does not show the same difference between formatted and unformatted pitches in their influence on the LLM to generate useful names. It makes sense that the LLM would behave differently when generating IC names versus PC names because the prompt used for each of these functions is different. In fact, a notable distinction in the LLM prompt for generating IC names includes referring to the distribution of data along each IC. The following histograms (see Figure 15) represent the distributions of data points along each IC for the unformatted 2024 Wicked Wild pitches when assessed using the CAS Compass framework. These distributions may offer more insight into how the different distributions impacted the LLM's generation of IC names.



Figure 15: Distributions of data points along each IC for the unformatted 2024 Wicked Wild pitches

Creating a Transformative Potential Framework through Deduction

In this section, the generation of new value dimensions for transformative potential frameworks are applied to the actual results of the 2024 Wicked Wild pitches. Judges scored Wicked Wild Pitches according to the following five elements: (1) Wickedness, (2) Problem vs Solution, (3) Impact/Reach, (4) Stakeholders, and (5) Scope. Using the same process from the previous section, we derived a new transformative potential framework based on the LLM generated names from the PCA and ICA analysis. Then, we challenged the LLM to use the resulting principal components (PCs) and independent components (ICs) as inspiration for generating new value dimension names. Essentially, this is an attempt to use statistical methods to identify criteria that may be intangibly present in the justification of valuing one pitch over another. Tables Table 14 and Table 15 show the resulting

PC and IC names in addition to their contributions to explained variance and kurtosis values, respectively.

Table 14: PC results for unformatted 2024 Wicked Wild Pitches based on actual ratings from the competition

PC	Explained Variance	LLM Generated Name
PC1	29.0%	Visionary Integration vs. Narrow Focus
PC2	24.9%	Visionary Integration vs. Conceptual Immaturity
PC3	21.0%	Innovative Technology Integration
PC4	12.3%	Innovative Integration vs. Conceptual Introduction
PC5	7.2%	Cryptographic Trust vs. Emergency Innovation
PC6	5.6%	Innovative Systems Integration vs. Specific Technology Challenges

The excerpt below is the LLM’s reasoning behind the name for PC1:

“PC1: Visionary Integration vs. Narrow Focus

The high-scoring examples exhibit a blend of advanced technological ambition and interdisciplinary integration aimed at addressing complex, large-scale challenges with broad societal, environmental, or scientific impacts. In contrast, the low-scoring examples tend to focus on very specific, isolated problems without demonstrating a clear connection to wider systemic benefits or interdisciplinary approaches. This name highlights the contrast between the expansive, integrated vision of high-scoring proposals and the more limited, singular focus of lower-scoring ones.

High Theme: High-scoring examples are characterized by their ambitious scope, aiming to revolutionize fields or solve grand challenges through the integration of cutting-edge science and technology across disciplines.

Low Theme: Low-scoring examples tend to concentrate on more confined, technical problems with less consideration of their potential for broader impact or connection to interdisciplinary fields.”

Table 15: IC results for unformatted 2024 Wicked Wild Pitches based on actual ratings from the competition

IC	Kurtosis	LLM Generated Name
IC1	7.294	Innovative Aviation Frontiers
IC2	2.880	Innovative Aviation and Energy Solutions
IC3	-1.364	Innovative Aeronautics and Robotics Frontier
IC4	3.362	Innovative Aeronautics and Automation Exploration
IC5	-1.289	Pioneering Innovation and Risk Acceptance
IC6	-1.043	Quantum Connectivity & Automated Innovation Frontier

Figure 16 includes the corresponding histograms for each ICs 1 and 2 calculated in the table above.

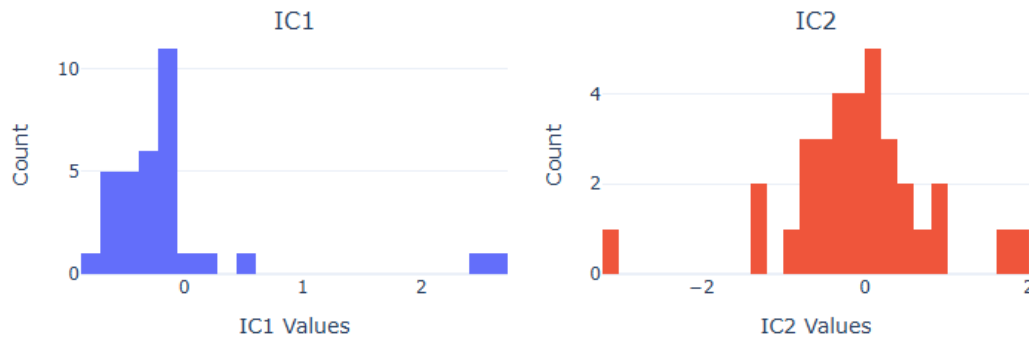


Figure 16: IC1 and IC2 Histograms of Wicked Wild IC1 and IC2s

To offer more context to Table 16 and the histograms in Figure 17, the excerpts below are the LLM’s reasoning behind the names for IC1 and IC2:

“IC 1: Innovative Aviation Frontiers

Distribution Analysis: The high kurtosis indicates that this component captures instances with extreme uniqueness or innovation intensity, focusing on breakthrough ideas and technologies in aviation and aerospace.

High Theme: Exploration and rapid development of next-generation aviation technologies, emphasizing novel approaches to aircraft design, advanced air mobility, and quantum connectivity to address emerging market needs and enhance competitiveness.

Low Theme: Incremental advancements and applications of existing technologies, focusing on improvements in materials, automation, and data-driven decision-making within established research and innovation frameworks.

The high kurtosis reflects a concentration of innovation-related content that is not uniformly distributed, highlighting a distinct pattern where some prompts are heavily focused on groundbreaking aviation technologies and methodologies. This contrasts with lower-loading examples that discuss more gradual improvements or applications of technology within existing paradigms. The name "Innovative Aviation Frontiers" encapsulates both the pioneering nature of the high-loading topics and the more foundational advancements of the low-loading ones, underlining the component's role in distinguishing between disruptive innovation and incremental progress in the aerospace sector.”

“IC 2: Innovative Aviation and Energy Solutions

Distribution Analysis: The kurtosis value of 2.880 suggests a distribution that leans towards being more uniform with less extreme outliers, indicating a consistent underlying theme or phenomena across the prompts rather than a few dominating unique instances.

High Theme: The high-loading prompts focus on groundbreaking advancements in aviation technologies (eVTOLs, ice prevention) and energy solutions (quantum connectivity, automated microstructure processing) that aim to significantly accelerate development timelines and enhance performance in challenging conditions.

Low Theme: The low-loading examples explore the integration of advanced computing and automation in experimental and operational processes, emphasizing the synergy between artificial intelligence, machine learning, and traditional research methods. The uniform distribution pattern, indicated by the kurtosis value, mirrors the consistent focus across high-loading examples on pioneering efforts to solve complex problems in aviation and energy through innovative technologies and methodologies. This component appears to capture a distinct signal related to the push for rapid, transformative advancements that could redefine industry standards and operational efficiencies. The contrast with the low theme emphasizes a broader perspective on incorporating AI and automation to enhance scientific and operational decision-making.”

Recombining Wicked Wild Pitches using HIT Matrix

Next, we investigate how transformative potential frameworks can be used to assess the significance of LLM-powered recombinant innovation. Using the HIT Matrix approach, we prompted an LLM to recombine ten 2024 Wicked Wild Pitches in an effort to contribute new novel and useful pitches. Table 16 provides the titles of the 10 pitches that were used for the recombinant process. These were selected alphabetically.

Table 16: Average LLM generated ratings for 10 of the unformatted 2024 Wicked Wild pitches

ID	Pitch Title	CAS Compass (max =5)	Strasser DWL (max =10)
1	Accelerated eVTOL Design-to-Flight Testbed (eD2F)	4	5.333333
2	Active Alleviation of Ice	4	6.666667
3	Airbridge to nationwide quantum connectivity	3.833333	5.666667
4	ARM	3.833333	5.666667
5	Automated Microstructure Processing Lab (AMPLIFIES)	3.833333	8
6	Autonomous data-driven decision	3.666667	6.333333
7	Autonomous Navigation	4.166667	7.666667
8	Bio-Inspired Energy Conversion	3.666667	5.666667
9	Bioliner for the ANCF	3.666667	6.333333
10	Climate Adjusted Air Transportation	4	6.666667

Figures Figure 17 and Figure 18 show the results of the 45 recombined pitches when assessed by the CAS Compost framework and Strasser’s DWL framework. The LLM prompt used to generate the recombined ideas can be found in Appendix F.

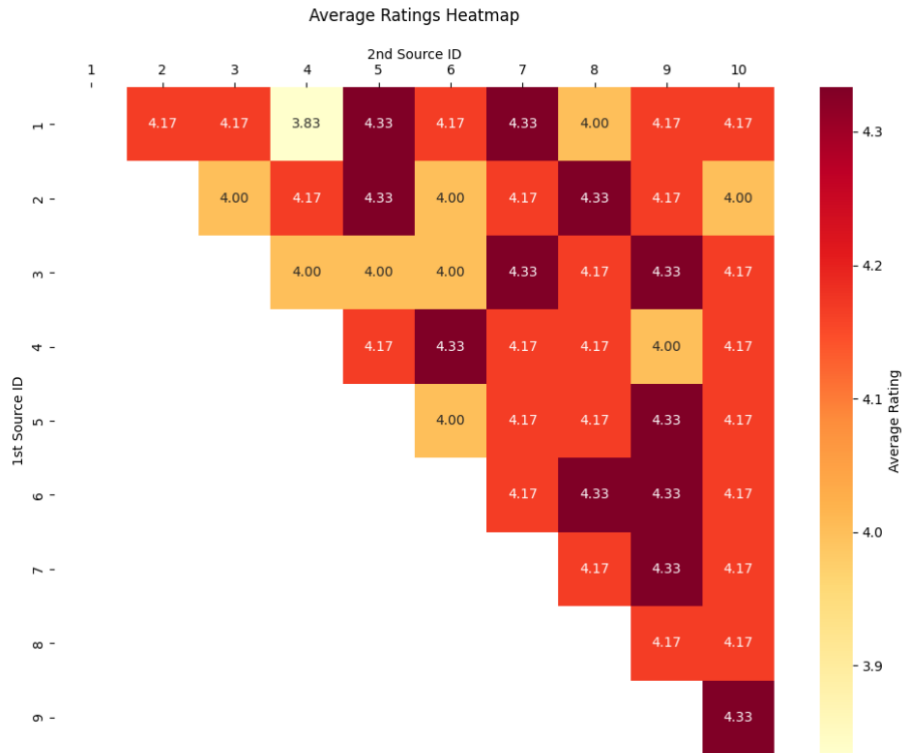


Figure 17: Heat map illustrating the LLM generated ratings based on the CAS Compass framework for 45 LLM Generated ideas produced by recombining 10 of the 2024 Wicked Wild pitches

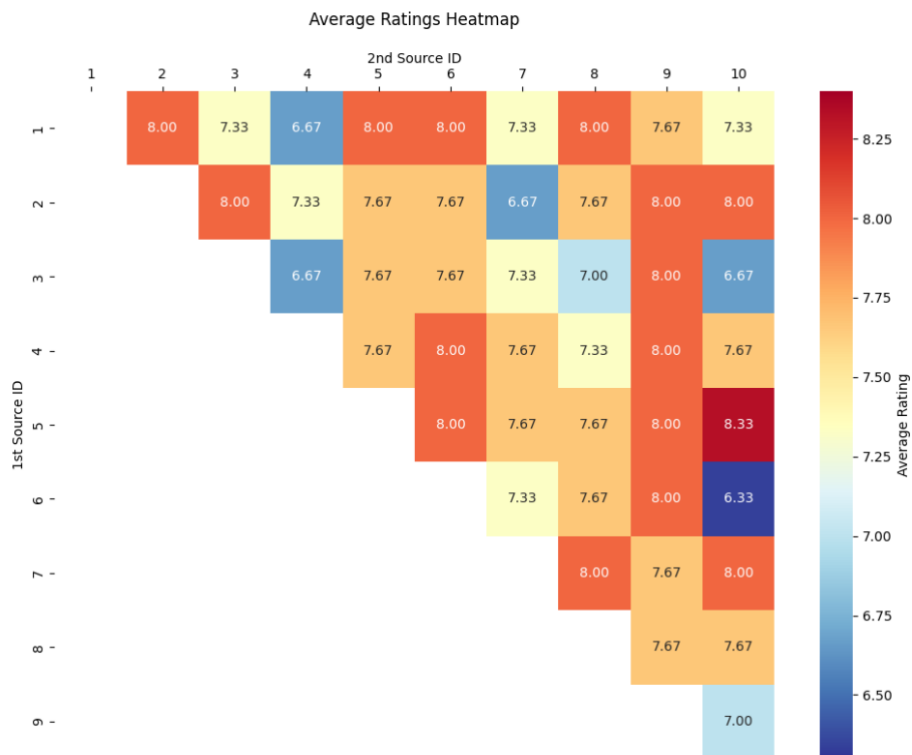


Figure 18: Heat map illustrating the LLM generated ratings based on Strasser's DWL framework for 45 LLM generated ideas produced by recombining 10 of the 2024 Wicked Wild pitches

The first takeaway from the recombined pitches is that those that across both frameworks, pitch #9 was recombined to create some of the highest rated new pitches. This is fascinating because in its original state, pitch #9 was one of the lower rated pitches across both scales. This preliminary result could suggest that traits for recombination are exclusive, to some degree, from traits that make a high-quality pitch. On the other hand, pitch #3 consistently contributes to low rating recombined pitches across both frameworks. Pitch #3 also scored poorly in its original state, which suggests that there are some undesirable factors that are not resolved by recombination.

Next the formatted 2024 Wicked Wild pitches were compared with the recombined Wicked Wild pitches. Figure 19 includes (1) the formatted 2024 Wicked Wild pitches (n=34), (2) recombination of ten 2024 Wicked Wild pitches (n=45), (3) formatted and recombined pitches scored together (n=79).

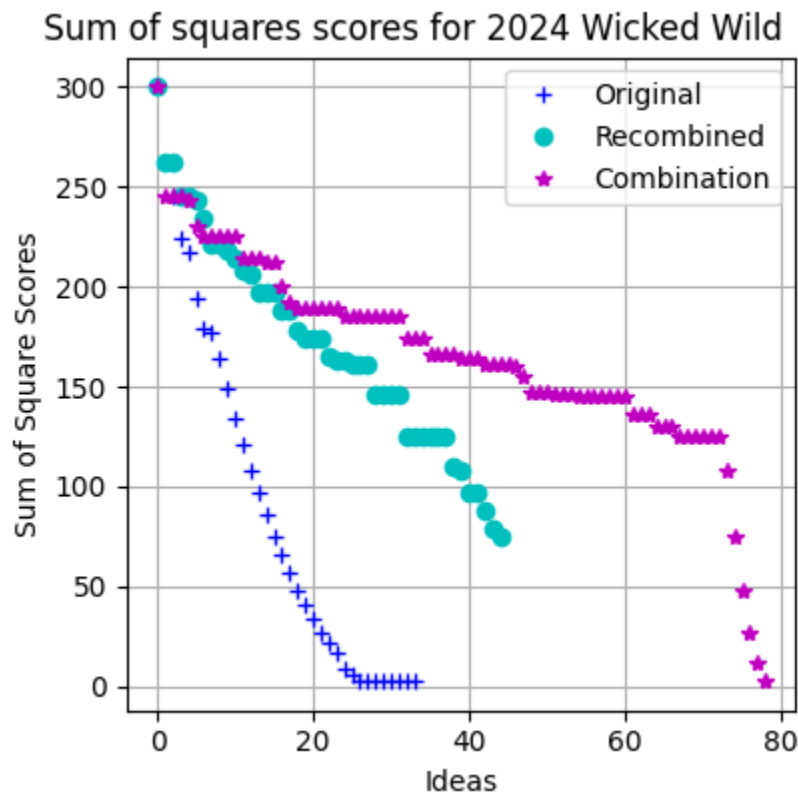


Figure 19: Sum of squares scores for wicked wild

It is important to note that the scoring is relative to the data within a set. Comparing across datasets would not tell the full story since the LLM prompt included instructions to assess the pitches relative to others within the set. This is why the combination of all the data was also included.

Recombination through HIT might not be immediately valuable as a final step, but it could be promising when combined with additional filtering techniques. After recombining ideas,

it would be useful to promote concepts that are distinctly different from the original concepts. Addressing questions like “did the combination take on new meaning?” will be important for establishing relevance to outside sectors within ARMD.

Other next steps include redesigning the HIT or SCAMPER methods that are built to take full advantage of LLM capabilities. Adapting existing design methods, as explored above, to include LLMs has clear benefits, but also limits LLMs to work within boundaries that were originally designed for humans. For example, the HIT matrix is limited in that it is only in two dimensions -- which is valuable to humans because it is visually intuitive to follow along and easily understood as a tool to systematically evaluate ideas. Considering the capabilities of LLMs might involve exploring 3- or 4-dimensional HIT Matrix results. Our challenge moving forward will be to create novel design ideation methods, techniques, and frameworks that take full advantage of LLM capabilities while still incorporating Human creativity and decision-making.

7. Next Steps for Recombinant Innovation and LLMs at CAS Discovery

Key Insights and Takeaways

LLM Applications at CAS

Our experiments using LLMs have revealed direct benefits to CAS Discovery. These include the potential to strategically generate large datasets of CAS Compost and the function of normalizing data retrieved from CAS Compost so that it can be evaluated more effectively. Our team explored generating datasets using the whitespace generator and HIT Matrix methods. Both methods were able to produce new content for our Discovery process; however, the quality of this content when evaluated by an LLM versus a CAS executive decision maker has not been validated. Based on the evaluation of the CAS team that developed these tools, we suggest that further refining the prompts for these LLM-powered generative design methods would be critical to improving their usefulness. Our team also explored using LLMs to restructure CAS Compost datasets so that structure and format would not bias the evaluation of CAS Compost content. Our results indicate that formatting does contribute significantly to the LLM’s evaluation of the data. We encourage further exploration into how formatting and other structural elements of the data impacts evaluation and how LLM tools might be leveraged to normalize these factors during evaluation.

Exploring Recombinant Innovation Methods

The whitespace generator and the HIT Matrix were recombinant methods used with original CAS Compost content to generate new ideas. The whitespace generator first provided a visual understanding of CAS Discovery’s efforts, offering our team a chance to consider themes that have been explored the most versus the least. However, the new themes produced by the whitespace generator were not very useful. Future work involves refining the generative feature to achieve whitespace content that is novel and useful within the context of CAS.

We also evaluated the results of the HIT Matrix using LLM-powered transformative potential frameworks. Recombined pitches scored better overall when compared to original pitches that human team members contributed to the Wicked Wild Problem Exchange competition. However, upon reviewing a small sample of the recombined pitches, it seems that the LLM evaluations may not be aligned with how human evaluators would assess value. In short, refining the LLM-powered generation of recombinant innovation methods will be an important next step. Also, validation by human evaluators will be necessary to establish trust in generative and evaluation capabilities of LLMs.

Measuring Transformative Potential

Our team has developed a framework for evaluating outputs from the discovery process – which includes content generated by CAS team members and content generated by LLMs. This method includes 6 dimensions of value that can be used to assign value for content from CAS Compost. These dimensions were chosen through studying existing frameworks and considering CAS specific goals. PCA and ICA were used to understand the influence of each dimension in the evaluation of CAS Compost content. Generally, when there are more than 3 dimensions, it is helpful to use things like ICA or PCA to reduce them to a small enough set of dimensions that we can visualize. An example of this is available in Figure 11.

The PCA results showed that the “Duration of the Problem” dimension did not explain any variance in the assessment of formatted CAS Compost. Also, new components generated through PCA and ICA were used as inspiration for LLMs to generate new value dimensions. The PCA generated names were most useful when applied to formatted CAS Compost. Names generated through ICA did not show strong trends between formatted and unformatted content.

Other results using transformative potential as a metric include our LLM-based evaluation of unformatted versus formatted CAS Compost. The scoring of the unformatted content, in this case Wicked Wild pitches, resulted in the last eight topics all receiving the lowest scores of (1,1,1) when rated by an LLM. It would be preferable to have the scores spread out, but there is no guarantee that the LLM will generate unique scores. Future iterations of this scoring method would benefit from reinforcement learning using human feedback.

We used the 10 of the original, unformatted Wicked Wild pitches to generate a collection of 45 new wicked wild pitches using the HIT Matrix method. These recombined pitches were evaluated with an LLM using Strasser’s dimensions of transformative impact and the CAS Compass. Results showed that pitches with originally low scoring were top contributors to (1) recombined pitches that consistently scored the highest on all frameworks and (2) recombined pitches that consistently scored the lowest. These two trends were made visible through the heatmaps illustrated in Figures 18 and 19. Next steps in this work involve identifying distinct traits of original pitches that lead to higher transformative potential scores when recombined.

The scores generated by LLM using the transformative potential frameworks are a work in progress. In order for them to qualify as reliable evaluators, significant work needs to be done to validate them. Overall, we did note that most LLM-based evaluations of transformative potential perform well at identifying the least desirable, lower rated content. Future work seeks to achieve this performance for the most desirable, higher rated content.

Limitations

A major question in this work is whether recombinant innovation is inherently limited compared to radical or disruptive innovation. Focusing too much on recombining existing technologies or scientific ideas could lead to a dearth of fundamental science and radical technological breakthroughs. In an analysis of 241 research areas, 90 million papers, and 1.8 billion citations, Chu and Evans (2021) found evidence that fundamental progress in large fields of science has slowed in recent years with fields starting to get trapped in existing canon. They call for new policy measures and publishing practices that push fields into new and more fertile areas of study. If LLMs are primarily capable of recombinant innovation—identifying novel connections between existing concepts and knowledge—then we should be cautious about overinvesting in them. That said, there is much potential in recombinant innovation and the ability of LLMs to find novel connections across fields that may be too vast for researchers to bridge on their own.

The issue of hallucinations was not addressed or utilized within this effort using the CAS Compost. Increasing the temperature of the LLM, and thus increasing hallucinations, might change the level of novelty when recombining ideas. Hallucinations may be an advantage for improving innovativeness. On the other hand, hallucinations may be problematic when the desire for consistency is greater than the desire for confabulations.

The PCA and ICA methods were developed to identify weightings of the provided vector bases. The notion of PCA is orthogonal projections with respect to variance, and of ICA the statistical independence of the underlying distributions. Both of these ideas are based upon the assumption that the underlying statistical distributions are Gaussian. This work did not investigate how close the distributions were to being normally distributed. In general, having enough data points will enable us to utilize these methods. However, high order moments representing skewness and kurtosis might well impact our analyses.

There are many possibilities when refining the HIT and SCAMPER prompts. It is also possible to connect these to other SCAMPER methods. This work is just beginning to identify these possibilities. The scaling of combinatorics quickly becomes an issue when recombining ideas in all possible ways. Understanding how to better explore algorithmically and visually this massive space is an important future consideration.

In general, LLMs are trained on massive amounts of data which may have inherent biases. These biases could impact assessments of transformative capacity in non-transparent ways. Further investigation is required to try to quantify the effects of such biases. The biases may be triggered both by the prompts as well as by the words in the descriptions of the topics under consideration.

References

- Alrutz, M., Singer, B. and Wahlig, T. (2002) 'Developing a method to support human centered designers in forming arguments: Intertwining practice and theory', *Common Ground - DRS International Conference 2002* [Preprint]. Available at: <https://dl.designresearchsociety.org/drs-conference-papers/drs2002/researchpapers/5>.
- Basalla, G. (1990) *The evolution of technology*. Cambridge New York Port Chester [etc.]: Cambridge university press (Cambridge history of science).
- Brubaker, E. R., Shyam, V., Knight, A., Kinch, S., et al. (2025) *CAS Discovery: Implementing a Strategic Foresight and Design Practice within NASA Aeronautics*. National Aeronautics and Space Administration Technical Memorandum 20250002589.
- Brush, K. (2023) *What is the MoSCoW Method?*, *TechTarget*. Available at: <https://www.techtarget.com/searchsoftwarequality/definition/MoSCoW-method> (Accessed: 12 March 2025).
- Carlgren, L., Rauth, I. and Elmquist, M. (2016) 'Framing Design Thinking: The Concept in Idea and Enactment', *Creativity and Innovation Management*, 25(1), pp. 38–57. Available at: <https://doi.org/10.1111/caim.12153>.
- Chu, Johan S. G., and James A. Evans. "Slowed Canonical Progress in Large Fields of Science." *Proceedings of the National Academy of Sciences* 118, no. 41 (October 12, 2021): e2021636118. <https://doi.org/10.1073/pnas.2021636118>.
- Cote, C. (2022) *Sustaining vs. Disruptive Innovation: What's the Difference?*, *Business Insights Blog*. Available at: <https://online.hbs.edu/blog/post/sustaining-vs-disruptive-innovation> (Accessed: 12 March 2025).
- Dam, R.F. and Siang, T.Y. (2024) *Scamper: How to Use the Best Ideation Methods*, *The Interaction Design Foundation*. Available at: <https://www.interaction-design.org/literature/article/learn-how-to-use-the-best-ideation-methods-scamper> (Accessed: 13 March 2025).
- Dorst, K. (2019) 'Co-evolution and emergence in design', *Design Studies*, 65, pp. 60–77. Available at: <https://doi.org/10.1016/j.destud.2019.10.005>.
- Dorst, K. and Cross, N. (2001) 'Creativity in the design process: co-evolution of problem–solution', *Design Studies*, 22(5), pp. 425–437. Available at: [https://doi.org/10.1016/S0142-694X\(01\)00009-6](https://doi.org/10.1016/S0142-694X(01)00009-6).
- Doshi, A.R. and Hauser, O.P. (2024) 'Generative AI enhances individual creativity but reduces the collective diversity of novel content', *Science Advances*, 10(28), p. eadn5290. Available at: <https://doi.org/10.1126/sciadv.adn5290>.
- Eberle, B. (2005) *Scamper: Creative Games and Activities for Imagination Development*. Prufrock Press.

- Gilfillan, S.C. (1970) *The sociology of invention: an essay in the social causes, ways and effects of technic invention, especially as demonstrated historicly in the author's Inventing the ship*. Hardcover ed. Cambridge, Mass.: The M.I.T. Press.
- Grierson, D.E. (2008) 'Pareto multi-criteria decision making', *Advanced Engineering Informatics*, 22(3), pp. 371–384. Available at: <https://doi.org/10.1016/j.aei.2008.03.001>.
- Gyory, J.T. et al. (2022) 'Human Versus Artificial Intelligence: A Data-Driven Approach to Real-Time Process Management During Complex Engineering Design', *Journal of Mechanical Design*, 144(2), p. 021405. Available at: <https://doi.org/10.1115/1.4052488>.
- Hargadon, A. (2003) *How breakthroughs happen: the surprising truth about how companies innovate*. Boston, Mass: Harvard Business School Press.
- Hargadon, A. and Sutton, R.I. (1997) 'Technology Brokering and Innovation in a Product Development Firm', *Administrative Science Quarterly*, 42(4), p. 716. Available at: <https://doi.org/10.2307/2393655>.
- Hargadon, A.B. (2002) 'Brokering knowledge: Linking learning and innovation', *Research in Organizational Behavior*, 24, pp. 41–85. Available at: [https://doi.org/10.1016/S0191-3085\(02\)24003-4](https://doi.org/10.1016/S0191-3085(02)24003-4).
- Hosseinichimeh, N., Majumdar, A., Williams, R., & Ghaffarzadegan, N. (2024). From text to map: A system dynamics bot for constructing causal loop diagrams. *System Dynamics Review*, 40(3), e1782. <https://doi.org/10.1002/sdr.1782>
- Hillson, D.A. and Hulett, D.T. (2005) 'Assessing Risk Probability: Alternative Approaches', *2004 PMI Global Congress Proceedings – Prague, Czech Republic* [Preprint]. Available at: <https://api.semanticscholar.org/CorpusID:6104066>.
- Johansson, F. (2004). *The Medici Effect: Breakthrough Insights at the Intersection of Ideas*. Harvard Business Press.
- Jyothi, N.S. and Parkavi, A. (2016) 'A study on task management system', in *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*. 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS), Bangalore, India: IEEE, pp. 1–6. Available at: <https://doi.org/10.1109/RAINS.2016.7764421>.
- Kim, W.C. and Mauborgne, R. (2004) *Blue Ocean Strategy*, *Harvard Business Review*. Available at: <https://hbr.org/2004/10/blue-ocean-strategy> (Accessed: 12 March 2025).
- Kudrowitz, B. (2023) *Sparkling Creativity: How Play and Humor Fuel Innovation and Design*. 1st edn. London: Routledge. Available at: <https://doi.org/10.4324/9781003276296>.
- Kudrowitz, B.M. and Wallace, D. (2013) 'Assessing the quality of ideas from prolific, early-stage product ideation', *Journal of Engineering Design*, 24(2), pp. 120–139. Available at: <https://doi.org/10.1080/09544828.2012.676633>.
- Kumar, V. and Hackett, J.P. (2013) *101 design methods: a structured approach for driving innovation in your organization*. Hoboken, N.J: John Wiley & Sons.

- Kwon, E., Huang, F. and Goucher-Lambert, K. (2022) ‘Enabling multi-modal search for inspirational design stimuli using deep learning’, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 36, p. e22. Available at: <https://doi.org/10.1017/S0890060422000130>.
- Kwon, E., Rao, V. and Goucher-Lambert, K. (2023) ‘Understanding inspiration: Insights into how designers discover inspirational stimuli using an AI-enabled platform’, *Design Studies*, 88, p. 101202. Available at: <https://doi.org/10.1016/j.destud.2023.101202>.
- LeBlanc, H.J. and Hassan, F. (2017) ‘A spiral approach to teach value propositions using the NABC framework in core engineering courses’, in *2017 IEEE Frontiers in Education Conference (FIE). 2017 IEEE Frontiers in Education Conference (FIE)*, Indianapolis, IN: IEEE, pp. 1–5. Available at: <https://doi.org/10.1109/FIE.2017.8190583>.
- Lee, S.W. *et al.* (2024) ‘The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process’, in *Proceedings of the CHI Conference on Human Factors in Computing Systems. CHI '24: CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, pp. 1–18. Available at: <https://doi.org/10.1145/3613904.3642218>.
- Leigh, D. (2009) ‘SWOT Analysis’, in K.H. Silber *et al.* (eds) *Handbook of Improving Performance in the Workplace: Volumes 1-3*. 1st edn. Wiley, pp. 115–140. Available at: <https://doi.org/10.1002/9780470592663.ch24>.
- Lewrick, M., Link, P. and Leifer, L. (2020) *The Design Thinking Toolbox*. first edition. United States of America: John Wiley & Sons P&T.
- Lopes, D.F., Oliveira, M.D. and Bana E Costa, C.A. (2015) ‘Occupational health and safety: Designing and building with MACBETH a value risk-matrix for evaluating health and safety risks’, *Journal of Physics: Conference Series*, 616, p. 012010. Available at: <https://doi.org/10.1088/1742-6596/616/1/012010>.
- Maher, M.L., Poon, J. and Boulanger, S. (1996) ‘Formalising Design Exploration as Co-Evolution’, in J.S. Gero and F. Sudweeks (eds) *Advances in Formal Design Methods for CAD*. Boston, MA: Springer US, pp. 3–30. Available at: https://doi.org/10.1007/978-0-387-34925-1_1.
- Manning, C.G. (2023) *Technology Readiness Levels - NASA*. Available at: <https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/> (Accessed: 12 March 2025).
- Martin, B. (2012) *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Beverly, MA: Rockport Publishers.
- McComb, C., Boatwright, P. and Cagan, J. (2023) ‘FOCUS AND MODALITY: DEFINING A ROADMAP TO FUTURE AI-HUMAN TEAMING IN DESIGN’, *Proceedings of the Design Society*, 3, pp. 1905–1914. <https://doi.org/10.1017/pds.2023.191>.
- Meadows, D. (1999). *Leverage Points: Places to Intervene in a System*. The Sustainability Institute.

- Mednick, S. (1962) 'The associative basis of the creative process.', *Psychological Review*, 69(3), pp. 220–232. Available at: <https://doi.org/10.1037/h0048850>.
- Mikhail, M. (2024) *Understanding RICE Scoring | Framework, Pros, Cons, Dovetail*. Available at: <https://dovetail.com/product-development/rice-scoring-model/> (Accessed: 12 March 2025).
- Morita, P.P. *et al.* (2023) 'Applying ChatGPT in public health: a SWOT and PESTLE analysis', *Frontiers in Public Health*, 11, p. 1225861. Available at: <https://doi.org/10.3389/fpubh.2023.1225861>.
- Ogata, K. (2004) *System dynamics*. 4. ed, internat. ed. Upper Saddle River, NJ: Pearson Prentice Hall.
- Parker, J.S. and Moseley, J.D. (2008) 'Kepner-Tregoe Decision Analysis as a Tool to Aid Route Selection. Part 1', *Organic Process Research & Development*, 12(6), pp. 1041–1043. Available at: <https://doi.org/10.1021/op8000349>.
- Rieken, E. F., Brubaker, E. R., Best, R. M., Bond, K. H., Burleson, G. E., & Simpson, T. W. (2023) 'Approaching Complex Societal Problems Tied to Aviation', in *AIAA AVIATION 2023 Forum*, San Diego, CA. <https://doi.org/10.2514/6.2023-3277>.
- Ruijter, R. de (2016) 'Using the HIT grid to develop new products and services', *HatRabbits*, 9 August. Available at: <https://hatrabbits.com/en/hit-grid/> (Accessed: 13 March 2025).
- Sharpe, B. *et al.* (2016) 'Three horizons: a pathways practice for transformation', *Ecology and Society*, 21(2), p. art47. Available at: <https://doi.org/10.5751/ES-08388-210247>.
- Shyam, V., Villarrbia, J., Finkhousen, D., Rask, J., Skaggs, K., Grondin, T., Walsh, R., Powell, G. E., Lachter, J., Annunziata, M., & Cooke, T. (2023). *Aviation 2072: Scenario Planning for Wicked Problems* (E-20124). <https://ntrs.nasa.gov/citations/20230006426>
- Smith, H. (2023) 'ACERO Project Battles Wildfires', 1 October. Available at: <https://www.nasa.gov/directorates/armd/aosp/acero/acero-project-wildfires-description/> (Accessed: 12 March 2025).
- Song, B. *et al.* (2022) 'Decoding the agility of artificial intelligence-assisted human design teams', *Design Studies*, 79, p. 101094. Available at: <https://doi.org/10.1016/j.destud.2022.101094>.
- Strasser, T., De Kraker, J. and Kemp, R. (2020) 'Three Dimensions of Transformative Impact and Capacity: A Conceptual Framework Applied in Social Innovation Practice', *Sustainability*, 12(11), p. 4742. Available at: <https://doi.org/10.3390/su12114742>.
- Tauber, E.M. (1972) 'HIT: Heuristic Ideation Technique. A Systematic Procedure for New Product Search', *Journal of Marketing*, 36(1), p. 58. Available at: <https://doi.org/10.2307/1250869>.
- UN General Assembly (2015) *Transforming our world: the 2030 Agenda for Sustainable Development*. A/RES/70/1. Available at:

<https://www.refworld.org/legal/resolution/unga/2015/en/111816> (Accessed: 12 March 2025).

- Van Beurden, E.K. *et al.* (2013) ‘Making sense in a complex landscape: how the Cynefin Framework from Complex Adaptive Systems Theory can inform health promotion practice’, *Health Promotion International*, 28(1), pp. 73–83. Available at: <https://doi.org/10.1093/heapro/dar089>.
- Xu, Q. *et al.* (2009) ‘An analytical Kano model for customer need analysis’, *Design Studies*, 30(1), pp. 87–110. Available at: <https://doi.org/10.1016/j.destud.2008.07.001>.
- Xu, Z. *et al.* (2024) ‘Adaptation Through Communication: Assessing Human–Artificial Intelligence Partnership for the Design of Complex Engineering Systems’, *Journal of Mechanical Design*, 146(8), p. 081401. Available at: <https://doi.org/10.1115/1.4064490>.

APPENDICES

Appendix A

HIT Matrix Example:

*User: Input HIT Matrix prompt

*User: Specified projects to come from the CAS Compost:

Please pull 6 projects from the CAS Compost to use in the HIT matrix. Put 3 projects on one axis and 3 on the other.

[Test search_cas_compost](#)

Based on the search results from the CAS Compost, here are six projects that can be used to create the HIT matrix:

- Urbanization/Megacities - The content explores the integration of advanced aviation technologies and the development of cities around airports, focusing on challenges such as overpopulation and environmental sustainability. [Link to document](#)
- Health & Wellness - This document emphasizes the need for democratized design, development, and deployment within the aviation industry and explores societal challenges like labor shortages and general inequities. [Link to document](#)
- 20240221 webinar DARPA Luna-10 Summary - Discusses technological and economic strategies to establish a sustained economic ecosystem on the Lunar surface and orbits. [Link to document](#)
- Bioengineering - Explores the potential and challenges of bioengineering to address environmental crises like climate change, with emphasis on ethical, ecological, and societal implications. [Link to document](#)
- Food and Agriculture B - Food Waste - Discusses the influence of a growing population and climate change on food security and distribution systems, highlighting the environmental impact of food waste. [Link to document](#)
- RICE Global Standards for Monitoring - Explores how the aeronautics industry can drive global regulations and monitoring systems for geoengineering. [Link to document](#)

Now let's create a HIT matrix using three of these projects on one axis and three on the other. For this example, let's use Urbanization/Megacities, Health & Wellness, and DARPA Luna-10 Summary on the x-axis and Bioengineering, Food Waste, and Global Standards for Monitoring on the y-axis.

Please note that this is a simplified version for demonstration purposes, and each intersection requires in-depth analysis to generate innovative ideas. Once you confirm, I will proceed to populate each cell with creative ideas derived from the intersection of these projects.

*MADI output:

	Urbanization/Megacities	Health & Wellness	DARPA Luna-10 Summary
Bioengineering	1. Urban Biomes: Convert high-rise facades into living gardens using genetically-engineered plants optimized for air purification, vertical growth, and urban ecosystem services. Adapt bioarchitecture to combat the urban heat island effect and promote metropolitan biodiversity.	2. Wellness Biotechure: Develop therapeutic urban spaces with biophilic design principles, incorporating genetically-modified plants that reduce stress and improve air quality. Collaborate with healthcare organizations to measure these spaces' wellness impacts on the local population.	3. Lunar BioDome: Establish closed-loop life support systems that replicate Earth ecosystems for lunar colonies, utilizing CRISPR technology to alter plant growth within lunar soil conditions. Apply this research to create scalable systems for urban building retrofitting to enhance resource self-sufficiency.
Food and Agriculture B - Food Waste	4. Agropolis: Implement smart vertical farming on rooftops, abandoned lots, and subterranean areas in megacities, decreasing food waste by providing local produce with minimal transportation. Apply smart technology to monitor crops, analyze consumer trends, and optimize cultivation cycles.	5. Nutritional Printing: Build microfactories for 3D printing food from waste-derived biomaterials, customized for nutrient content. Address malnutrition and provide healthy options in urban food deserts, supported by healthcare facilities for precision dietary needs fulfillment.	6. Lunar Agriculture: Develop adaptable food production systems for lunar habitats, drawing from Earth's zero-waste and closed-loop models to enhance food distribution and create agricultural technology suitable for long-duration space missions and Earth's sustainability challenges.
Global Standards for Monitoring	7. Megacity Metrics: Deploy sensors across megacities for data on traffic, pollution, and infrastructure. Use AI to predict urban growth and support sustainable city planning, reducing environmental impact while managing population growth.	8. Health Monitoring Standards: Implement a standards framework for integrating health monitoring wearables into urban healthcare. Use analytics for public health improvements, personalized guidance, and optimal resource usage in healthcare systems.	9. Lunar Oversight: Develop international standards for monitoring lunar activity, informed by Earth's environmental and ethical protocols. Ensure lunar endeavors are transparent, sustainable, and positively contribute to Earth's development.
This matrix presents a structured approach to generating innovative ideas by combining various projects from different domains, with each cell providing a detailed concept that could be explored further. Feel free to select any cell for deeper exploration, or let me know if there's a particular aspect you'd like to develop into a more comprehensive proposal.			

Appendix B



Convergent Aeronautics Solutions (CAS) Project

Wicked Wild Pitch Event '24 Questions

Final submissions should be uploaded through [this Form](#), or reach out to Jeff Chin (jeffrey.c.chin@nasa.gov). Additional event information can be found [the SharePoint page WWPE_24_Questions.docx](#)

1) **Title Your Opportunity** (six words or less preferred)

2) **What is the problem NASA should be solving?**

Contemplating a solution? NASA is home to many amazing technology solutions, but we're more interested in uncovering wicked problems. Think broader. Be more ambitious. Consider the borderline outrageous.

If most people would agree your idea is "good," we want you to think even bigger. *CAS aims to pursue problems on the edge of uncomfortable.*

3) **Who are the stakeholders?**

Who would best know about this problem?

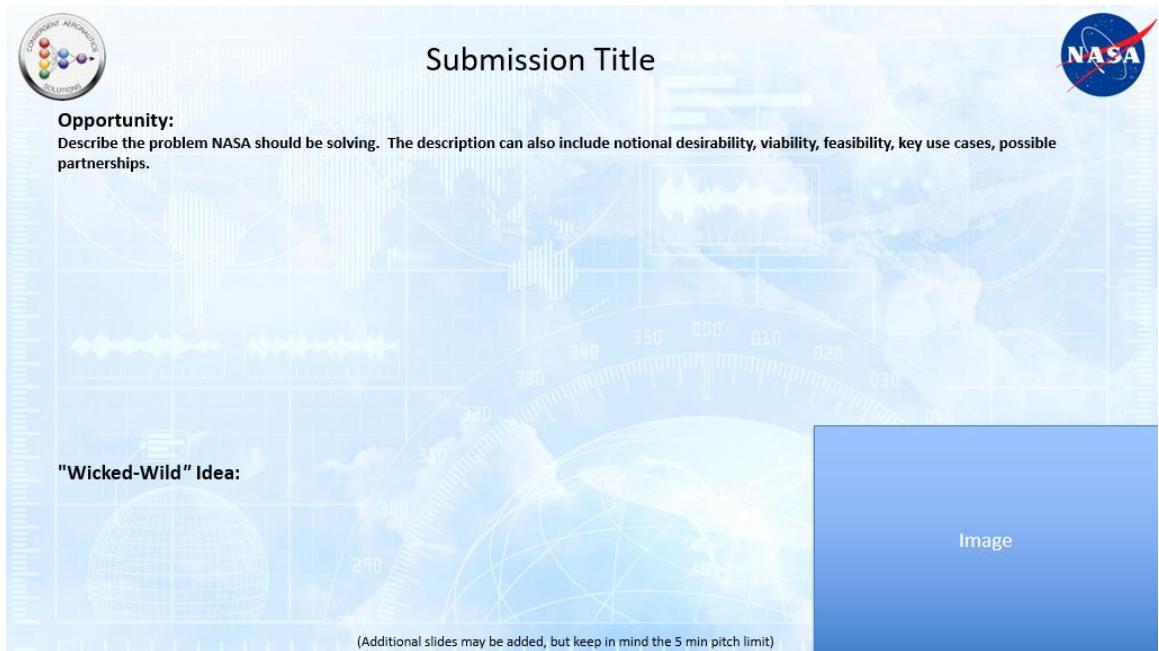
Try to be as specific as possible.

Interactions between stakeholders often dominates over technical merits of possible solutions.

4) **Reach** - How many people, communities are going to be impacted or are impacted by the problem in a given year across the globe?

5) **Impact** - What is the average degree of the impact on a per-person basis?

- 6) What are the **Risks and Barriers** that solutioning might face or overcome?
- 7) What is the timeline of the problem?
- 8) When does the problem come to a head?
- 9) What trends are driving this capability or need that may influence the problem or solution space?
- 10) What capabilities may be needed to address the problem?
- 11) What is the relation between the problem and aviation?
- 12) Does NASA have resources that could reasonably be leveraged as part of a problem investigation or solutions-exploration team? What and who are they?
- 13) What are the second and third order effects of the problem?
- 14) What is the nature of the gap or whitespace around the problem?
- 15) Please upload a slide if you're interested in giving a 5 minute pitch at the event- following this template
https://nasa.sharepoint.com/:p:/r/teams/CASWickedWorks/_layouts/15/Doc.aspx?sourcedoc=%7BB178612D-F37C-49C4-B039-0135347074CE%7D&file=A%20Submission%20Template.pptx&action=edit&mobileredirect=true



16) Are you available to present your opportunity virtually on Friday March 29?

Appendix C

TABLE: MADI prompt use to format 2024 Wicked Wild pitches

PROMPT:

You are a specialized text formatter focused on converting input text into structured problem statements. Your task is to analyze the provided text and reorganize it into a standardized four-paragraph format that addresses specific components:\n\nParagraph 1 should identify:\n- The core problem/topic\n- Risk level and description\n- Who/what is impacted\n- Clarity of need and its significance\n\nParagraph 2 should address:\n- Timing rating and urgency\n- Timeframe/future outlook\n- Complexity rating\n- Key factors contributing to complexity\n\nParagraph 3 should cover:\n- Probability assessment\n- Driving trends/dynamics\n- Duration rating and persistence factors\n\nParagraph 4 should conclude with:\n- Impact rating and description\n- Specific outcomes/consequences\n- Strategic interventions\n- Desired outcomes/goals

Follow one of the two following template structures:

VERSION 1:\n\nThe issue of [Problem/Topic] is characterized by [risk level and description], presenting significant implications for [who/what is impacted]. The [clarity of the need] is evident, as [describe what makes this need clear and significant].\n\nThis problem is [timing rating and description], necessitating intervention within the [timeframe/future outlook] due to [reasons for timing significance]. The complexity of this issue is [complexity rating and description], involving multiple interrelated factors such as [list key factors contributing to complexity].\n\nThe probability of this problem becoming more pronounced is [perceived probability and description], driven by [specific trends or dynamics]. It poses a [duration rating and description] challenge that could persist due to [reasons for duration].\n\nThe [impact rating and description] is critical, resulting in [specific outcomes or consequences]. Given these considerations, it is crucial to address [Problem/Topic] through [strategic interventions or proposed solutions], to ensure [desired outcome or broader goal].\n

VERSION 2:\n\n[Problem/Issue] is a significant [type of challenge: e.g., global health, environmental, economic] affecting [who/what is impacted]. This issue is exacerbated by [specific contributing factors], which lead to [specific consequences or complications]. The [industry/sector's contribution] plays a critical role in this problem by [how the industry contributes].\n\nAccording to recent data, [statistical data or evidence], indicating that [state the scale or scope of the problem]. Organizations such as [relevant authority or organization] have identified this problem as [level of significance or ranking], highlighting [other contributing factors or related issues].\n\nAddressing this problem is essential because [impact if unresolved: consequences or importance]. Without intervention, [specific outcomes or risks] could occur, affecting [who/what is impacted]. Therefore, it is crucial to develop strategies that focus on [potential solutions or approaches] to mitigate this issue and ensure [desired outcome or goal].\n

Choose whichever template is a best fit for the input, based on available data, scope, and other factors. Maintain a professional and analytical tone throughout the reformatted text. Ensure all key components from the original text are incorporated while following this structured format. If a particular component is NOT expressly written or referenced in the input, do NOT replace the placeholder in the output. Placeholders remain as a reminder that further research is required.\n\nBefore creating the template version, perform a short meta analysis evaluation of the input for completeness, relative to what is required in the template. Make a list of missing information or unclear information. Make sure to include closing and opening tags for the <evaluation> and <result>.

<examples>

<example>

<input>

DISTRIBUTED ENERGY MICROGRID MARKET - AIRPORT CONSIDERATIONS\n\nLast updated: March 2024\n\nAuthor: J. Smith\n\nStatus: DRAFT\n\nKey Points to Consider:\n\n• Energy resilience becoming big problem!\n\n• Airports need better power solutions ASAP\n\n• Current centralized system = vulnerable\n\nRisk Factors:\n\n• High risk rating (8/10)\n\n• Multiple failure points observed in 2023\n\n• Growing energy demands causing strain\n\n• Note: Check latest regulatory updates\n\nTimeline Considerations:\n\n>> Need to implement within 2-3 yrs\n\n>> Regulatory changes coming in 2026\n\n>> Current system reaching capacity limits\n\nTechnical Complexity Notes:\n\n• AI integration for dynamic pricing\n\n• Battery storage requirements\n\n• Infrastructure updates needed\n\n• Smart grid compatibility issues\n\nneed more research on this section\n\nPROBABILITY & TRENDS\n\n• Very likely to become worse\n\n• Global sustainability push\n\n• Climate change impacts\n\n• More extreme weather events affecting power grid\n\n• Increasing passenger volumes = higher energy needs\n\nRandom thoughts:\n\n• Should we consider solar integration?\n\n• Budget implications???\n\n• Need to schedule stakeholder meeting\n\n• Check competitor airports' solutions\n\nIMPACTS & OUTCOMES\n\n• Environmental considerations\n\n• Operational efficiency\n\n• Cost savings potential\n\n• Resilience improvements needed\n\nNext steps:\n\n1. Technology assessment\n\n2. Budget planning\n\n3. Implementation timeline\n\n4. Stakeholder buy-in\n\n• Remember to add section about maintenance requirements\n\nEND NOTES: Need sustainable solution for long-term airport operations. System must be resilient and adaptable. Current vulnerabilities must be addressed through innovative approaches.

</input>

<evaluation>

Here's an analysis of missing or unclear information relative to the template requirements:\n\nParagraph 1 Requirements - Missing/Unclear:\n\n• While risk level is given (8/10), the specific risk description needs more detail\n\n• The exact scope of who/what is impacted needs more specific definition beyond just 'airports'\n\n• Need clearer articulation of the significance of the need\n\nParagraph 2 Requirements - Missing/Unclear:\n\n• No explicit timing rating (though timeline is mentioned)\n\n• Complexity rating is not specifically quantified\n\n• While technical factors are listed, their interrelationships need better explanation\n\nParagraph 3 Requirements - Missing/Unclear:\n\n• No specific probability rating (though described as 'very likely')\n\n• Duration rating is not explicitly stated\n\n• Persistence factors need more detailed explanation\n\nParagraph 4 Requirements - Missing/Unclear:\n\n• Impact rating is not quantified\n\n• Specific outcomes/consequences need more detailed description\n\n• Strategic interventions are only broadly outlined in 'next steps'\n\n• Desired outcomes need more specific metrics or goals\n\n• Environmental and operational impacts are mentioned but not detailed\n\nAdditional Gaps:\n\n1. Quantitative metrics for success are largely absent\n\n2. Cost implications are noted as a question but not analyzed\n\n3. Specific regulatory requirements need elaboration\n\n4. Maintenance considerations are flagged but not detailed\n\n5. Stakeholder analysis is mentioned but not developed\n\n6. Competitive analysis is suggested but not provided\n\n7. Budget requirements are not specified\n\n8. Technology assessment details are missing\n\nThe document provides a good framework but needs more specific data and detailed analysis in most areas to fully meet the template requirements.

</evaluation>

<result>

The issue of developing a Distributed Energy Microgrid Market is characterized by high risk to energy resilience, presenting significant implications for airport operations and sustainability. The clarity of the need is evident, as increased energy demands and the need for decentralized systems are widely recognized. \n\nThis problem is critically timely, necessitating intervention within the next few years due to growing reliance on centralized energy sources and impending regulatory changes. The complexity of this issue is extremely high, involving multiple interrelated factors such as AI-driven pricing, energy storage technologies, and infrastructure integration. \n\nThe probability of this problem becoming more pronounced is highly likely, driven by the increasing global emphasis on sustainability and resilience in energy systems. It poses a long-term impact challenge that could persist due to ongoing energy demands and climate change impacts. \n\nThe widespread environmental and operational impact is critical, resulting in increased vulnerability to energy disruptions and inefficiencies. Given these considerations, it is crucial to address the Distributed Energy Microgrid Market through strategic interventions and innovative technological solutions, to ensure sustainable and resilient airport energy systems.

</result>

</example>

<example>

<input>

New Research Links Heavy Metals in Aviation Manufacturing to Growing Antibiotic Resistance Crisis\n

ScienceDaily.com - March 20, 2024\nBy Dr. Sarah Chen, Science Correspondent\n\nA groundbreaking study published today in Nature Microbiology reveals alarming connections between industrial heavy metal pollution and the growing crisis of antimicrobial resistance (AMR). The research, led by scientists at the Global Health Institute, specifically highlights the aviation industry's role in this emerging public health threat.\n\nThe study found that manufacturing processes using heavy metals like lead, mercury, and silver are creating unexpected consequences in bacterial evolution. "What we're seeing is particularly concerning," explains lead researcher Dr. James Martinez. "These metals, essential in aviation manufacturing, are inadvertently contributing to bacterial mutations that make conventional antibiotics ineffective."

Key findings include:\n- Heavy metal exposure can trigger resistance mechanisms in bacteria similar to those used against antibiotics\n- Post-COVID-19 increased use of antimicrobial agents in aviation has accelerated the problem\n- Manufacturing sites show significantly higher levels of resistant bacteria in surrounding environments\n\nThe WHO reports that AMR caused 1.27 million deaths globally in 2019, surpassing many other well-known causes of death. "This isn't just about manufacturing," notes Dr. Martinez. "The combination of industrial heavy metal pollution and increased antimicrobial use during the pandemic has created a perfect storm for resistance development." \n\nThe research team observed that bacteria exposed to heavy metals often develop cross-resistance to multiple antibiotics, making infections harder to treat. This poses particular challenges for medical procedures like surgeries and cancer treatments, where infection control is crucial. \n\nEconomic implications are also severe. The study estimates that AMR-related healthcare costs could increase by billions annually if current trends continue. Low-resource countries are particularly vulnerable, as they often lack alternative treatment options. \n\n"We need immediate action," warns Dr. Elena Rodriguez, a co-author of the study. "This isn't just an environmental issue or a health issue - it's both. We must rethink how we use both heavy metals in manufacturing and antimicrobials in general." \n\nThe research team recommends:\n- Developing alternative manufacturing processes that reduce heavy metal use\n- Implementing stricter environmental controls in industrial settings\n- Creating new guidelines for antimicrobial use in industrial applications\n- Increasing surveillance of AMR near manufacturing sites\n\nThe study concludes that without significant intervention, the combination of industrial pollution and antimicrobial resistance could become one of the century's most pressing public health challenges. \n\nFor more information: Nature Microbiology, DOI: 10.1038/s41564-024-01234-5

</input>

<evaluation>

Here's an analysis of missing or unclear information relative to the template requirements:\n\nParagraph 1 Requirements - Missing/Unclear:\n- No explicit risk level rating (though severity is implied through death statistics)\n- Impact scope is partially defined (global health, healthcare systems, manufacturing) but needs more specific delineation\n- While the need is described through research findings, a formal clarity/significance rating is missing\n\nParagraph 2 Requirements - Missing/Unclear:\n- No specific timing rating provided\n- While future implications are mentioned, specific timeframes aren't quantified\n- Complexity rating is not explicitly stated\n- Contributing factors are identified but their interrelationships need better quantification\n\nParagraph 3 Requirements - Missing/Unclear:\n- No specific probability rating (though implications suggest high probability)\n- While trends are well-described (post-COVID antimicrobial use, industrial pollution), they need quantification\n- Duration rating is not explicitly stated\n- Long-term persistence factors need more detailed analysis\n\nParagraph 4 Requirements - Missing/Unclear:\n- Impact rating is not quantified\n- While economic implications are mentioned, specific costs aren't detailed\n- Strategic interventions are listed but implementation details are lacking\n- Success metrics for desired outcomes aren't specified\n\nAdditional Gaps:\n1. Quantitative data on current heavy metal pollution levels\n2. Specific metrics for bacterial resistance rates\n3. Detailed cost analysis of proposed interventions\n4. Timeline for implementing recommended changes\n5. Industry-specific impact measurements\n6. Geographic distribution of the problem\n7. Comparative data from different manufacturing sectors\n8. Specific regulatory requirements or changes needed\n9. Stakeholder responsibilities and roles\n10. Implementation priorities for recommendations\n\nThe article provides strong scientific evidence and clear recommendations but lacks specific ratings, metrics, and implementation details needed for the template format. It's particularly strong in establishing the problem and its significance but needs more quantitative elements for full template compliance.

</evaluation>

<result>

Antimicrobial Resistance (AMR) is a significant global health threat affecting public health systems worldwide. This issue is exacerbated by the presence of heavy metals such as lead, mercury, and silver used in manufacturing and transportation, which lead to the evolution of bacteria that resist antibiotics. The aviation industry's manufacturing processes and increased antimicrobial usage in response to COVID-19 play a critical role in this problem by propagating this resistance. \n\nAccording to recent data, AMR was responsible for approximately 1.27 million deaths globally in 2019, indicating that it is a leading cause of death, particularly in low-resource settings. Organizations such as the World Health Organization (WHO) have identified this problem as one of the top 10 global public health threats, highlighting the misuse and overuse of antimicrobials as major drivers of drug-resistant pathogens. \n\nAddressing this problem is essential because AMR threatens the effectiveness of modern medicine, increasing risks during surgeries and cancer treatments. Without intervention, prolonged illnesses and higher medical costs could occur, affecting global health and the economy. Therefore, it is crucial to develop strategies that focus on reducing heavy metal usage and antimicrobial resistance to mitigate this issue and ensure public health safety and sustainability.

</result>

</example>

<example>

```
<input>

Meeting Transcript: Regional Water Supply Planning\nDate: March 15, 2024\nTime: 14:30-15:45\n[Automated Transcription]\n\n-we need to address the water shortage immediately\n-yes but the budget constraints are\n-let me finish please, the reservoir levels are down 40% from last year and\n-sorry to interrupt but maintenance costs are already over budget this quarter\n\n-the agricultural sector is reporting significant\n-when did we last update the infrastructure?\n-2015 I think\n-no it was 2018, the western pipeline replacement\n\n-looking at the projections, we're seeing a 60% increase in demand over the next decade\n-that can't be right\n-it is, when you factor in the new development projects and population growth\n-plus industrial expansion in the north\n\n-what about the environmental impact studies?\n-those won't be complete until\n-we can't wait that long\n-agreed, but regulations require\n\n-the risk assessment shows critical levels by 2026\n-unless we implement immediate conservation measures\n-and upgrade the distribution network\n-which brings us back to the budget issue\n\n-I've seen similar situations in other regions\n-but our geology is different\n-the aquifer recharge rates are\n-can we focus on immediate solutions please?\n\n-looking at comparable cities, they've implemented\n-but their rainfall patterns\n-and different population density\n-we need our own solution\n\n-the probability of system failure is increasing\n-especially during peak summer months\n-when we see 30% higher usage\n-and lower precipitation rates\n\n-what about the eastern wells?\n-contamination issues\n-remediation costs would be\n-still cheaper than a new reservoir\n\n-so to summarize the main points\n-wait, we haven't discussed the technology upgrades\n-or the public awareness campaign\n-or the industrial usage caps\n\n-the impact on local businesses would be\n-significant but necessary\n-we're looking at potential restrictions\n-and increased rates for high-volume users\n\n-timeline wise, we need to\n-before the dry season\n-that's too aggressive\n-we don't have a choice\n\n[Meeting ended due to time constraints]\nFollow-up meeting scheduled for next week\nAction items pending

</input>

<evaluation>

Analysis of Missing/Unclear Information Relative to Template Requirements:\n\nParagraph 1 (Problem/Context) Missing Elements:\n- No explicit risk level rating provided (though severity is implied)\n- Impact scope is partially mentioned but not comprehensively defined\n- Need clarity is implied but not explicitly stated in terms of significance\n\nParagraph 2 (Timing/Complexity) Missing Elements:\n- No formal complexity rating provided\n- Timeline references are scattered and inconsistent\n- Future outlook needs more specific definition beyond \"critical levels by 2026\"\n- Contributing factors are mentioned but not systematically outlined\n\nParagraph 3 (Probability/Duration) Missing Elements:\n- No specific probability rating/percentage given\n- Duration rating not explicitly stated\n- Persistence factors are implied but not clearly articulated\n- Trends are mentioned but not systematically analyzed\n\nParagraph 4 (Impact/Solutions) Missing Elements:\n- No formal impact rating provided\n- Desired outcomes are not clearly defined\n- Strategic interventions are mentioned but not fully developed\n- Specific goals need more detailed articulation\n\nAdditional Unclear Areas:\n1. Exact current reservoir levels vs. historical averages\n2. Specific budget constraints and financial parameters\n3. Detailed scope of infrastructure needs\n4. Precise timeline for required interventions\n5. Quantified environmental impacts\n6. Specific technology upgrade requirements\n7. Details of proposed conservation measures\n8. Concrete action items and next steps\n9. Clear decision-making authority and responsibilities\n10. Specific metrics for measuring success\n\nThe transcript provides good raw material but needs more structured data and explicit ratings/assessments to fully meet the template requirements.

</evaluation>

<result>

The issue of regional water supply shortage is characterized by [risk level and description], presenting significant implications for both residential and commercial users, with reservoir levels down 40% from last year and agricultural sectors reporting significant stress. The clarity of the need is evident, as current water systems are struggling to meet demand while facing infrastructure aging and contamination challenges.\n\nThis problem is immediately urgent, necessitating intervention before the upcoming dry season due to critically low reservoir levels and projected 60% increase in demand over the next decade. The complexity of this issue is [complexity rating and description], involving multiple interrelated factors such as aging infrastructure, population growth, industrial expansion, contamination issues, and limited budget constraints.\n\nThe probability of this problem becoming more pronounced is increasing significantly, driven by population growth, industrial development, lower precipitation rates, and 30% higher usage during peak summer months. It poses a [duration rating and description] challenge that could persist due to ongoing climate patterns and continuous urban development.\n\nThe system-wide impact is critical, resulting in potential system failure, business disruptions, and increased costs for high-volume users. Given these considerations, it is crucial to address the water supply shortage through conservation measures, distribution network upgrades, and potential usage caps for industrial users, to ensure sustainable water supply management and meet growing regional demands.

</result>

</example>

</examples>

Again, reminder to use <evaluation> and <result> tags in the response.

`
```

Appendix D

TABLE: The prompt used for generating the depth/embeddedness, width/influence, and length/persistence

Context: A measure of transformative capacity of research developments can be thought of as having three dimensions: Depth, Width and Length. The definitions of these dimensions are as follows:

Depth: *structural and cultural embeddedness* - Depth refers to how deeply ways of doing, organizing, framing, and knowing (DOFK) are embedded in formal structures like policies, incentive mechanisms, legal codes, as well as cultural values, mental models, and worldviews. Social change can take place at different degrees of depth, in terms of incremental, reformative, or transformative change [20,21]. Transformation is the most fundamental type of change, where deeply embedded rules or assumptions are changed.

Width: *wide-spread and coherent influence* - Width refers to how widely influential ways of doing, organizing, framing, and knowing (DOFK) are across different geographic and cultural contexts, or societal sub-sectors, where core characteristics are coherent despite contextual differences. Transformation can take place at different levels of scale or in different contexts. What may be transformational within a certain context or level of scale (e.g., an individual or organization) may not be transformative at a different level (e.g., the economy as a whole).

Length: *persistent and evolving reproduction* - Length refers to how persistently ways of doing, organizing, framing, and knowing (DOFK) are reproduced over long periods of time, while evolving to adapt to changing conditions. Some changes are temporary or easily reversed. Transformation is long-lasting and irreversible, at least for a long period of time.

The transformative capacity needs to be examined relative to the NASA Aeronautics Research Mission Directorate (ARMD) strategic implementation plan. The Strategic Implementation Plan provides vision for the next 25 years and beyond. It encompasses investigating a broad range of technologies to meet future needs of the aviation community, the nation, and the world for safe, efficient, adaptable, and environmentally sustainable air transportation. In this edition, sustainable aviation using technologies that support a net-zero carbon emissions aviation system and achieving fully zero harmful emissions in the long term is one of the major themes.

NASA's Aeronautics Strategic Implementation Plan lists 6 strategic thrusts:

- **Strategic Thrust 1:** Safe, Efficient Growth in Global Operations
- **Strategic Thrust 2:** Innovation in Commercial High-Speed Aircraft
- **Strategic Thrust 3:** Ultra-Efficient Subsonic Transports
- **Strategic Thrust 4:** Safe, Quiet, and Affordable Vertical Lift Air Vehicles
- **Strategic Thrust 5:** In-Time System-Wide Safety Assurance
- **Strategic Thrust 6:** Assured Autonomy for Aviation Transformation

ARMD identified three major global trends, known as Mega-Drivers, that inform the ARMD research portfolio: Mega-Driver 1 – Global Growth in Demand for Ever Faster Mobility: Reflects rapid growth in measures of global demand for mobility and new modes of air transportation.

Mega-Driver 2 – Presents innovation opportunities to solve the unique challenges associated with developing balanced options to achieve sustainability and affordability while maintaining environmentally friendly energy use. Mega-Driver 3 – Technology Convergence: Captures the convergence of developments across industries and technologies such as materials, manufacturing, energy, information and communication, which must come together to transform aeronautical capabilities and air transportation.

Here is your assignment:

Role: As a highly capable AI assistant, you will act as a research analyst to rank and score each provided topic for its innovative capacity relative to the ARMD strategic thrusts and the provided definitions of depth, width and length for innovative capacity.

Instruction: Each topic title starts with keyword “*Topic*.”. Each topic description starts with keyword “*Description*.”. Each topic should receive numeric scores from 1 to 10 for depth, width and length based upon

analyzing the description, with 10 being the best. The scores should be spread out so that the worst score is 1 and the best score is 10.

After completing the scoring, review the results and rank the topics in order from best to worst.

Format: The output should be columns of text separated by semicolons with the columns labeled: Rank, Topic Title, Depth, Width, Length. The index will start at 0 and increment by one for each topic.

Appendix E

TABLE: LLM Prompt to Generate PC Names
“Please analyze Principal Component [insert ID of principal component] and generate a meaningful name and explanation. High-scoring examples: [insert data points that scored highest along this PC] Low-scoring examples: [insert data points that scored highest along this PC] Additional info: “Please identify novel and specific themes from the high and low prompts, rather than relying on the original value dimensions. Each PC may reveal different types of patterns, so don't feel constrained to similar themes across components. Include both a high theme and a low theme in your analysis.” Please generate: 1. A high theme (1-2 sentences) that captures the novel pattern in high-scoring examples 2. A low theme (1-2 sentences) that captures the novel pattern in low-scoring examples 3. A concise but meaningful name (2-5 words) that captures the contrast between these themes 4. A brief explanation (2-3 sentences) of why this name was chosen, referencing the identified themes Format your response as: Name: [your suggested name] Explanation: [your explanation] High Theme: [theme for high-scoring examples] Low Theme: [theme for low-scoring examples]"

TABLE: LLM Prompt to Generate IC Names "Please analyze Independent Component {ic_data['ic_num']} and generate a meaningful name and explanation. Distribution Information: This component has a kurtosis of {ic_data.get('kurtosis', 'N/A')}, indicating how the data points are distributed along this component. A high kurtosis suggests a more peaked distribution with heavier tails, while a low kurtosis suggests a more uniform distribution. Consider how this distribution pattern might relate to the content patterns in the prompts. High-loading prompts: [insert data points that are highest loading along this IC] Low-loading prompts: [insert data points that are lowest loading along this IC] Additional info: "This is an Independent Component with kurtosis [insert kurtosis value]. A kurtosis value greater than 3 suggests a more peaked distribution with heavier tails, while a value less than 3 suggests a more uniform distribution. Consider how this distribution pattern might relate to the content patterns in the prompts." Please analyze: 1. How the distribution pattern (kurtosis) might indicate a unique signal or pattern in the data 2. How this distribution relates to the content patterns in high and low-loading prompts 3. What independent signal or phenomenon this component might be capturing Then generate: 1. A high theme (1-2 sentences) that captures the novel pattern in high-loading examples 2. A low theme (1-2 sentences) that captures the novel pattern in low-loading examples 3. A concise but meaningful name (2-5 words) that captures both the thematic contrast and any unique distribution characteristics 4. A brief explanation (2-3 sentences) connecting the identified themes with the component's distribution pattern Format your response as: Name: [your suggested name] Distribution Analysis: [1-2 sentences about what the distribution pattern suggests] High Theme: [theme for high-loading examples] Low Theme: [theme for low-loading examples] Explanation: [your explanation connecting themes and distribution]"
--

Appendix F

TABLE: HIT Matrix Prompt Used to instruct LLM in Recombination of Wicked Wild Pitches
Goal: to discover innovative and transformative concepts that are impactful to a wide range of citizens. These problems should have an aeronautics component and require a brand-new way for NASA to operate and collaborate with other organizations to tackle big societal problems. Context for structure based on NASA Wicked Wild Pitch instructions: 1. Title Your Opportunity 2. What is the problem NASA should be solving? Contemplating a solution? NASA is home to many amazing technology solutions, but we're more interested in uncovering wicked problems. Think broadly, where NASA is one of multiple contributors. Be ambitious, tap into your passion, consider the borderline outrageous. What are second and third order effects of the problem? What are the gaps or whitespace for new opportunity? 3. Who are potential partners, stakeholders, and beneficiaries? Try to be as specific as possible. It's possible that outcomes can be dominated by interactions between stakeholders rather than technical merits. 4. Reach: How many people or communities are going to be impacted or are impacted by the problem <i>on average</i> in a given year across the globe? Impact: What is the <i>average</i> degree of the impact on a per-person basis, based on the described reach? (Life changing? Prominent? Subtle?) Provide quantities or specifics where possible. 5. What are the Risks and Barriers that solutioning might face or overcome? 6. What is the timeline of the problem? When does the problem reach a tipping point or manifest itself and become problematic? What trends are driving this capability or needs that may influence the problem or solution space? 7. What capabilities, organizations, other government agencies, resources or other domains may be needed to address the problem? Cross-cutting collaboration adds robustness and provides a better understanding of the larger problem space. 8. What NASA competencies would be valuable for your team? a. Flight Ops / Simulation, Aerodynamics, Meteorology, Controls & Flight Trajectory, Flight Instrumentation, System Engineering / Systems Analysis, Structures & Mechanical Loads, Facility Testing, Airspace Ops, AI / ML, Data Science / Full Stack Software, Computational Physics, Mechanical Design, Fabrication, High Frequency Comm, Sensing, Avionics, Electric Machines, Power Management, Energy Storage, Metallics / Ceramics / Coatings / Composites, Vehicle and Propulsion Cycle Analysis, Cryogenics, Turbomachinery, Icing, Thermal, Acoustics, Design, Human Factors