

MIKA: Manager for Intelligent Knowledge Access Toolkit for Engineering Knowledge Discovery and Information Retrieval

Sequoia Andrade^{1,2} and Dr. Hannah Walsh²
¹HX5, ²NASA Ames Research Center



33rd Annual **INCOSE**
international symposium
hybrid event
Honolulu, HI, USA
July 15 - 20, 2023

PROBLEM & OBJECTIVE

- Repositories of narrative reports contain a wealth of information that can provide insight during system development; however, they are often unstructured, inconsistently filled out, or difficult to analyze at scale
- State-of-the-art natural language processing techniques that allow fine tuning on domain-specific documents show promise in improving how these repositories can be leveraged
- Packaging these techniques as a user-assistive tool enables efficient integration into systems engineering practice at multiple junctions, demonstrated in Figure 1
- The Manager for Intelligent Knowledge Access (MIKA) toolkit provides knowledge discovery and information retrieval capabilities for analyzing large repositories of natural-language text

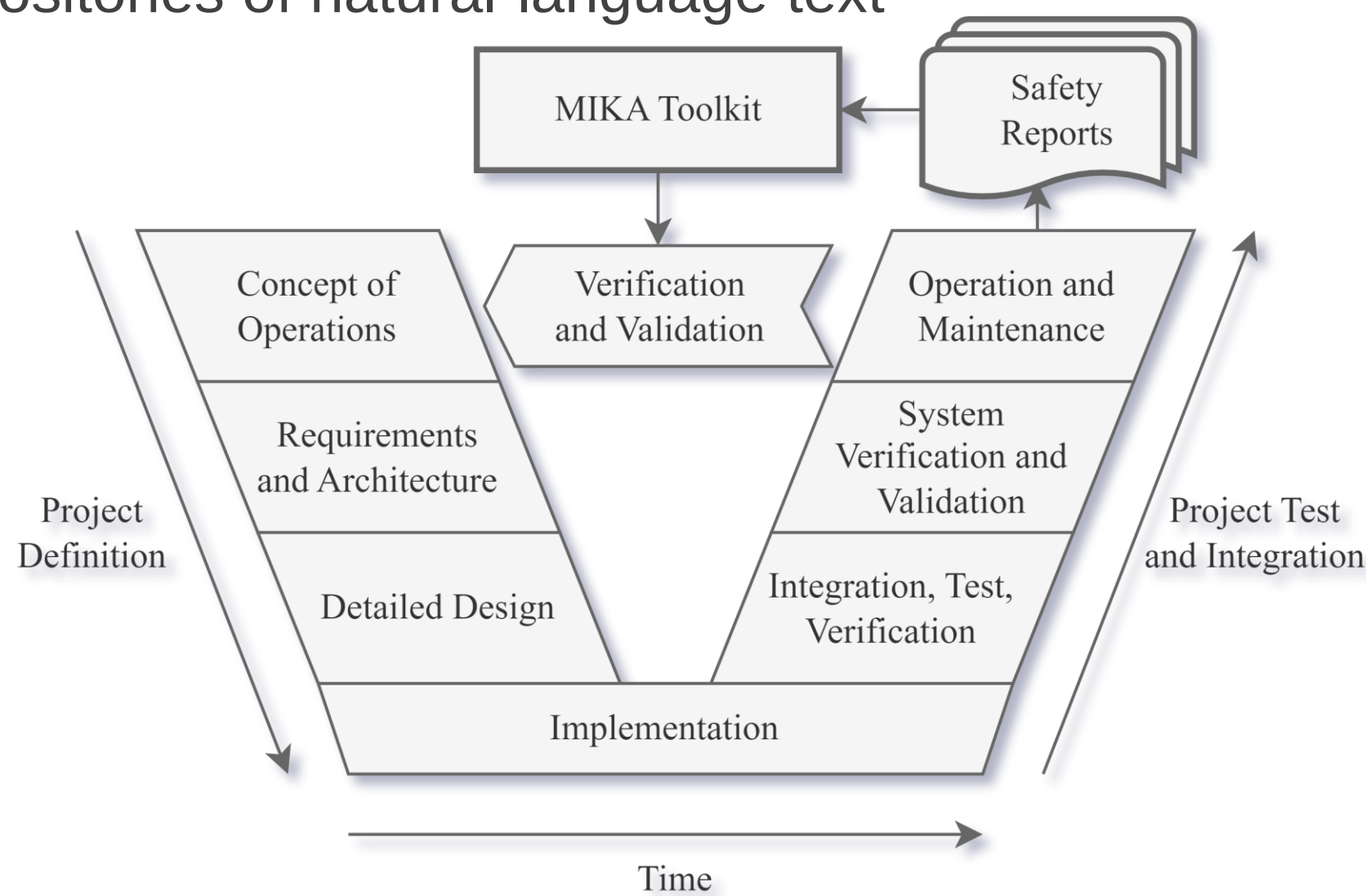


Figure 1: MIKA usage in systems engineering process

MIKA TOOLKIT METHODS

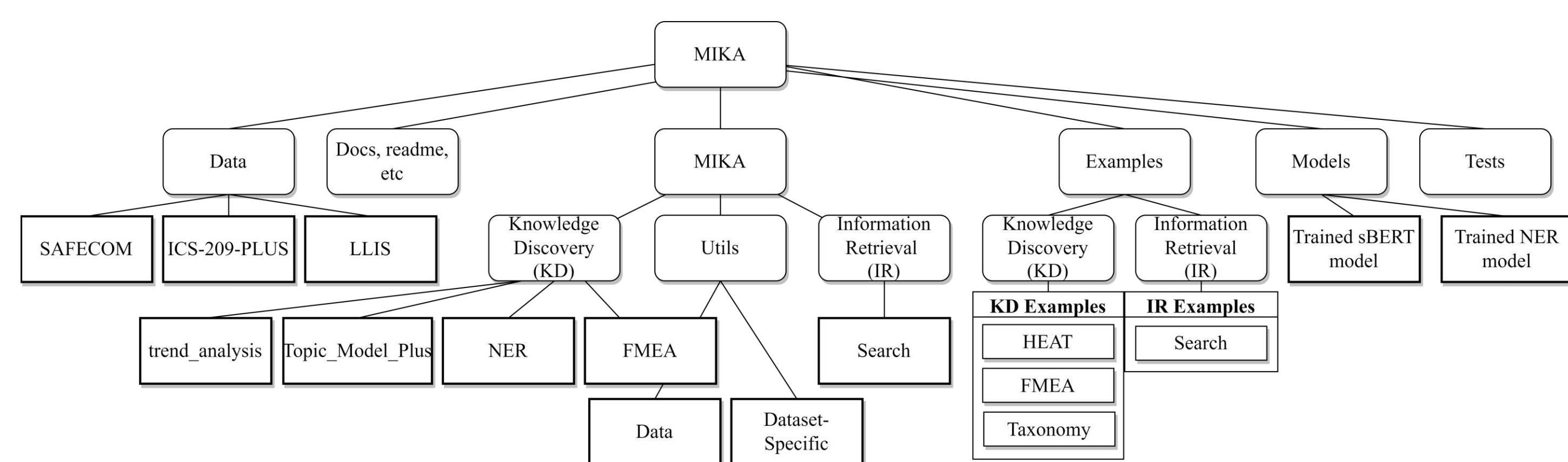


Figure 2: MIKA toolkit architecture

- The MIKA toolkit structure is shown in Figure 2 and includes three sub-modules: utilities, knowledge discovery, and information retrieval
- Utilities: includes a data class that supports analysis in all MIKA capabilities, so there is no extra setup to run different NLP techniques
- Knowledge discovery: for exploratory studies, understanding trends and themes, and extracting hazards
 - Includes: topic modeling, trend analysis, failure modes and effects analysis (FMEA), Named entity recognition
- Information retrieval: extracting detailed descriptions of specific instances of a hazard, trend, or theme via semantic search

Table 1: Subset of FMEA generated via MIKA named-entity recognition

Phase	Cause	Failure Mode	Effect	Control Process	Recommendation
Bird strike					
Maneuvering; Taxi-from Runway; Takeoff; Landing; Other	at low, bird, surface, a, red - tailed, hawk, female	in - flight, collision, with a, bird, large, pass, foul	bird, substantial, damage, large, broke the front center, windshield	was, towed to the gate, a, bird ingestion and containment	see the public, enter, keep an, eye, on, land
Midair collision					
Maneuvering; Taxi-from Runway; Takeoff; Landing; Approach; Emergency Descent	failure of both pilots to see and avoid the other	midair, collision, airplanes, collided, collided nearly head -, on, main	impacted the, terrain, a utility, pole, to, rest, an, immediate	wingtip, mounted strobe anticollision, lights, graphic, remote communication	alter, course, pass well, not have passed over, under

Table 2: Example query and documents retrieved using semantic search

Result #	Document ID	Similarity	Relevant Text Excerpt
Query 1: what components are vulnerable to fatigue crack			
1	20121204X63622	0.746134	...one of the first-stage compressor blades had fractured due to fatigue cracking...
2	20141027X62323	0.728395	...the crankshaft had fractured due to a fatigue crack that had initiated at a fillet radius...
3	20160114X73526	0.715234	... forward wing spar likely fractured due to compression loading from wing loads...

CASE STUDY

- Applied MIKA toolkit to National Transportation Safety Board (NTSB) aviation accident report dataset from 2011-2021 (16,914 documents)
- Knowledge Discovery:**
 - Topic modeling can create failure taxonomies from unstructured data, such as in Table 3
 - Figure 3a identifies loss of control in flight and midair collisions as most severe
 - Figure 3b identifies loss of control on the ground and in-flight as most frequent
- The risk matrix in Figure 4 identifies bird strikes, midair collisions, loss of control, turbulence encounters, and ground collisions as high risk according to FAA definitions
- MIKA-generated FMEA in Table 1 successfully extracts causes, failure modes, effects, control processes, and recommendations from the set of documents referring to the specified failure using named-entity recognition. In the first row, a red tail hawk causes an in-flight collision with a bird resulting in substantial damage. In the second row, both pilots failed to see and avoid each other causing a midair collision resulting in impact to terrain.
- Information Retrieval:**
 - Example query and document retrieval results in Table 2 identify compressor blades, the crankshaft, and the wing spar as components vulnerable to fatigue crack.

Figure 3: Average severity with standard deviations in (a) and frequency over time in (b)

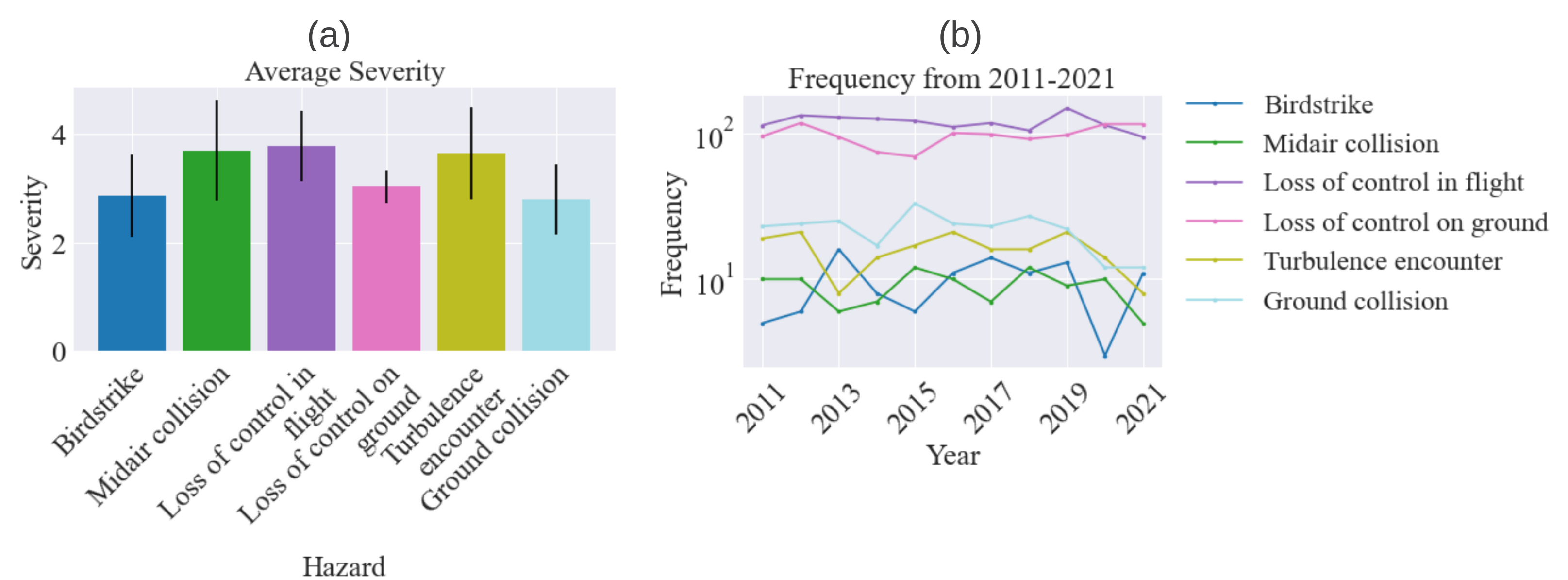


Figure 4: Risk matrix of selected hazards generated via MIKA trend analysis

		Risk Matrix				
Likelihood	Frequent					Loss of control in flight
	Probable					Loss of control on ground, Turbulence encounter, Ground collision
	Remote					Birdstrike, Midair collision
	Extremely Remote					
	Extremely Improbable					
		Minimal Impact	Minor Impact	Major Impact	Hazardous Impact	Catastrophic Impact
		Severity				

Table 3: Subset of taxonomy generated with MIKA topic modeling

Cause Narrative	Accident Narrative	# of Reports
fuel, oil, starvation, fuel starvation, power fuel starvation, engine, loss engine, loss engine power, engine power, power	oil, engine, connecting rod, connecting rod, crankshaft, bearing, cylinder, revealed, examination	33
	fuel, tank, engine, fuel tank, power, tanks, pilot, gallons, airplane, flight	62

CONTACTS

Sequoia Andrade
Research Engineer
Sequoia.R.Andrade@nasa.gov

Dr. Hannah Walsh
Computer Engineer
Hannah.S.Walsh@nasa.gov

CONCLUSIONS

- MIKA is intended to improve an organization's ability to leverage knowledge stored in documentation from past projects and accidents to reducing the space of unanticipated hazards and problems.
- MIKA is designed to be integrated into the systems engineering process flexibly at multiple junctions such that unanticipated problems can be most effectively reduced through enhanced knowledge management.