

AI-Assisted FRAM Modelling: A User Study on Large Language Model Support for Function Identification

16th of June

Edith Rutenfranz & Dr.-Ing. Niklas Grabbe

Technical University of Munich



Outline

- Motivation
- Related Work
- Building the Tool
- Study Design / Methodology
- Evaluation & Results
- Discussion

Motivation

Identification of system functions requires extensive domain experience as well as effort [1].



Idea: LLMs might be able to help, especially beginners but also experts



Goal: LLMs that support the user in building their model



Question

Can LLMs, if used in this specific tool, help the user, especially beginners?



Related Work

- *Sujan, M., Slater, D., & Crumpton, E. (2025). How can large language models assist with a FRAM analysis? Safety Science, 181. [2] :*
 - **Finding functions**
 - **Comparing different LLMs**
 - **Changing prompts intuitively by an expert**
 - **CT-System as test case**
- *Kaya, G., Bovell, D., Sujan, M., & Braithwaite, G. (2025). Large language models powered system safety assessment: applying STPA and FRAM. Safety Science, 191, [3] :*
 - **Using single prompts to simulate the reality of day-to-day LLM usage**
 - **Prompts created by experts**
- *Diemert, S., & Weber, J. (2023). Can Large Language Models assist in Hazard Analysis? [4]:*
 - **Asking the LLM "yes-no" questions -> no clear but still useful answers by the LLM**
 - **Prompts created by experts**

Building the tool



First approach by ChatGPT

→ Many disadvantages:

- Nearly no tokens and no system prompts in the free version
- No easy way to run locally
- No easy way to share your work with others
- No guarantee for consistency with upgrades



Building the tool



Second approach using Ollama only

→ Better but still problems:

- No easy way to share your work with others
- Still missing a bit of customizability



Building the tool



- Third approach using Open Web UI with Ollama
- Many useful features (that were used) such as:
- Giving the model extra FRAM-specific knowledge
 - Using system prompts to set the tone as an FRAM-Expert
 - Ability to control system parameters for as few hallucinations as possible
 - Saving prompting templates



Budling the tool

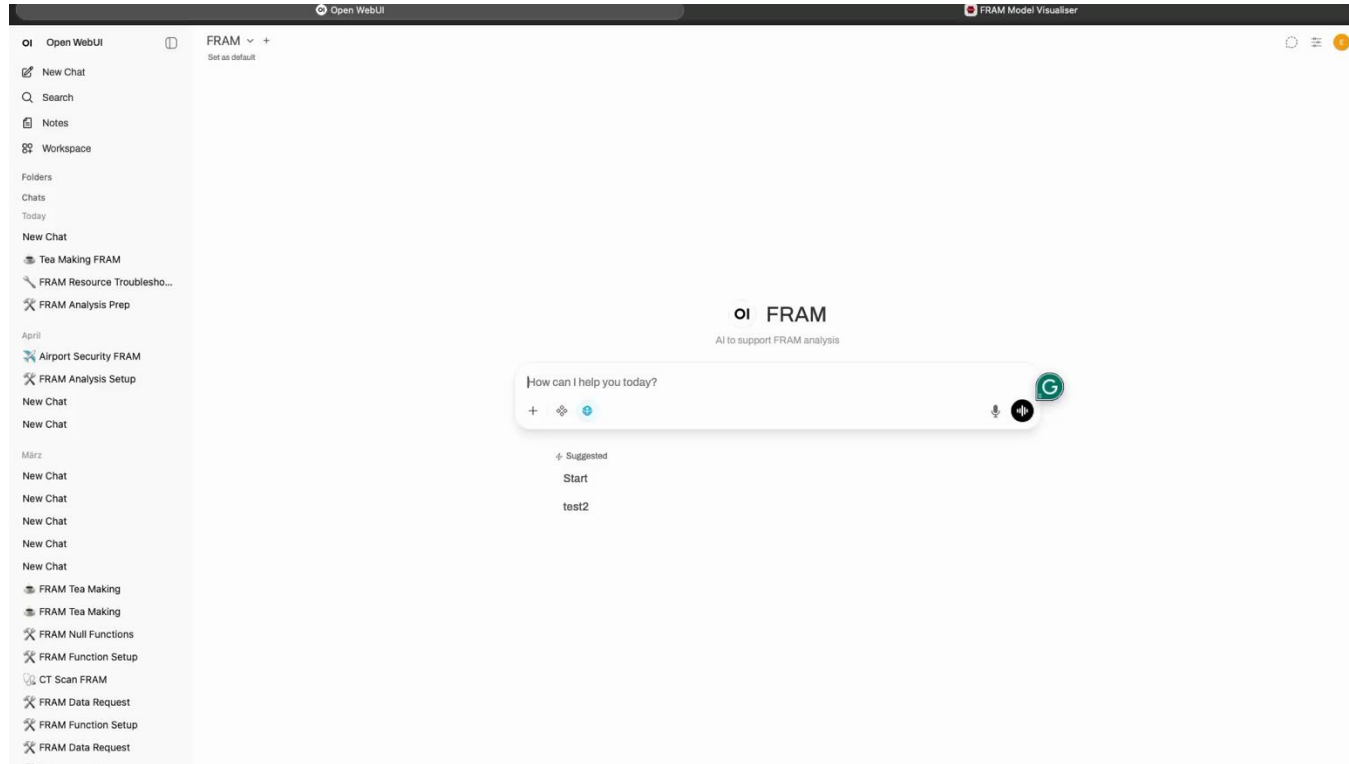
Third Approach using Open Web UI with Ollama



- Giving the LLM "FRAM-knowledge"
 - Simplified: *Explaining FRAM and FMV to the LLM and let it summarize it for itself*
- Building the prompting templates and system prompts using the work of White et al. [5], especially:
 - Persona pattern
 - Recipe pattern
 - Output template
 - Question decomposition



Building the tool



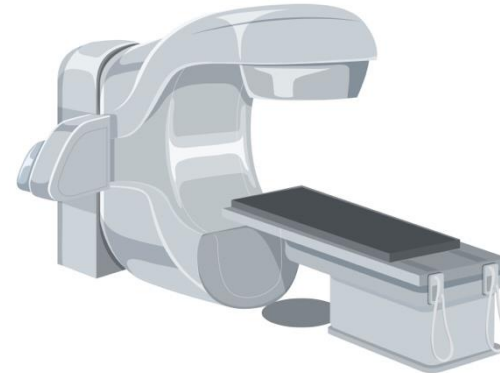
Study Design / Methodology

- Two Systems with somewhat the same complexity and process characteristics

Airport security check



Direct transfer to the CT pathway



Study Design / Methodology

Two Groups of Participants

10 FRAM-experts



10 beginners



Study Design / Methodology

- AI tool support vs. no support



Study Design / Methodology

Procedure:

- Before the study: collect data via Google Forms and give the FRAM manual to beginners
- Introduce the beginners to the FMV and answer any questions about FRAM



Study Design / Methodology

Procedure:

- Introduce the participant to the AI tool
- (Permuated order in AI usage and case) Giving the first and second system to the participant
- After the preparation period, participants will each have 25 minutes to work on the system
- Short interview about the participant's experience with the tool



Study Design / Methodology

Evaluation of the participants' models:

- Compare the models of the participants with a “reference model” based on:
 - **Function matching (50% of the end score)**
 - Functions matching **exactly (1P)**, **somewhat(0.5)**, **missing(0P)**, or even “**wrong**” (-0.5P) or hallucinated (-1P) (Sum of points divided by Number of functions)



Study Design / Methodology

Evaluation of the participants' models:

- **Aspect matching (30% of the end score)**
 - Only for the **existing functions, exact match(1P), existing but wrong aspect (0,25), or even “wrong”. (-0.25P) or hallucinated(-0,25P)**

(Sum of points divided by the number of Aspects of the compared functions)



Study Design / Methodology

Evaluation of the Participants' Models:

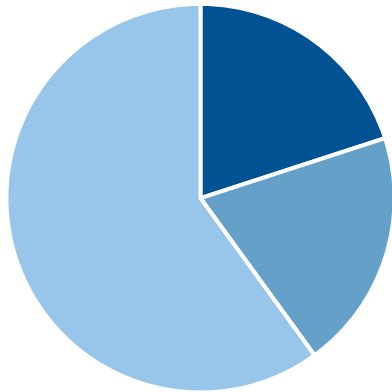
- **Coupling matching (20% of the end score)**
 - For all couplings of the “reference model” **matching (1P)**, **going to the wrong target function(0P)**, **going to the wrong aspect of the right function(0P)**, or even **“wrong”(-0.25P)** or **hallucinated (-0.25P)**
(Sum of points divided by the number of couplings in the “sample Solution”)



Study Design / Methodology

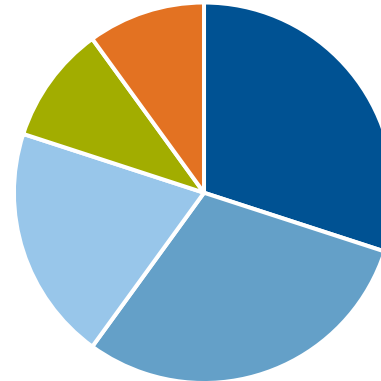
Experts:

Experience in years



■ <5 ■ >5&<10 ■ ">=10"

Main domain FRAM application

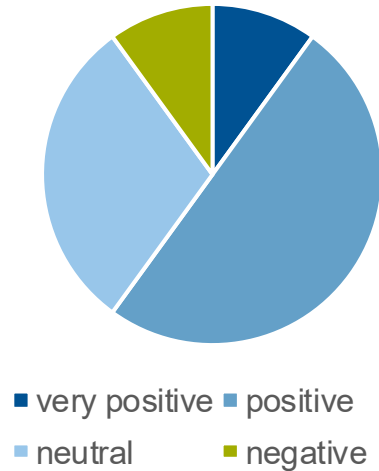


■ Transportation ■ Healthcare
■ Disaster management ■ Oil and gas
■ Aviation

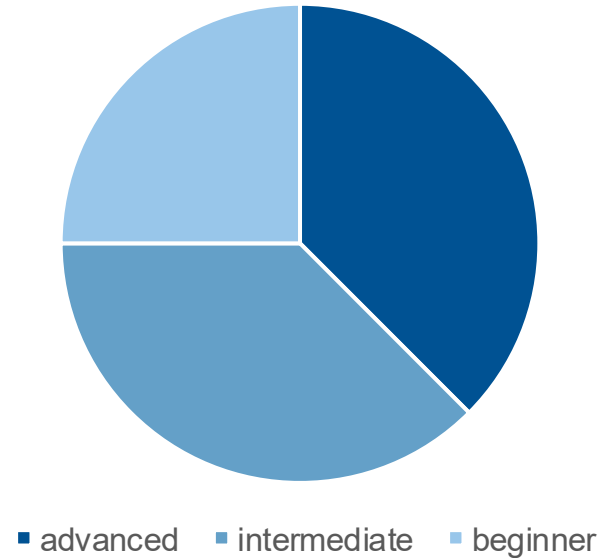
Study Design / Methodology

Experts:

Attitude AI



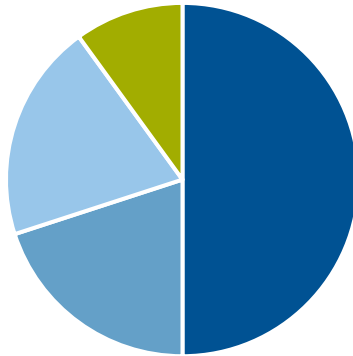
Experience AI



Study Design / Methodology

Beginners

Educational background

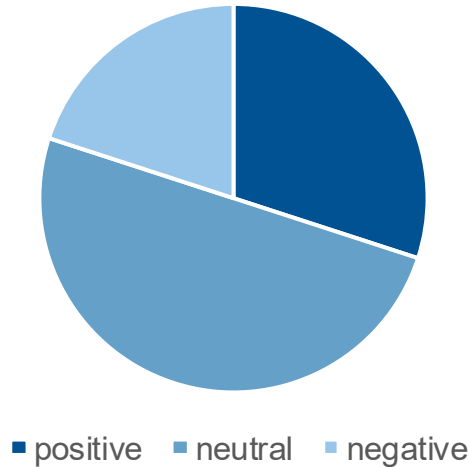


- Mechanical engineering
- Mechanical engineering & Computer science
- Chemical engineering
- Geoinformatics

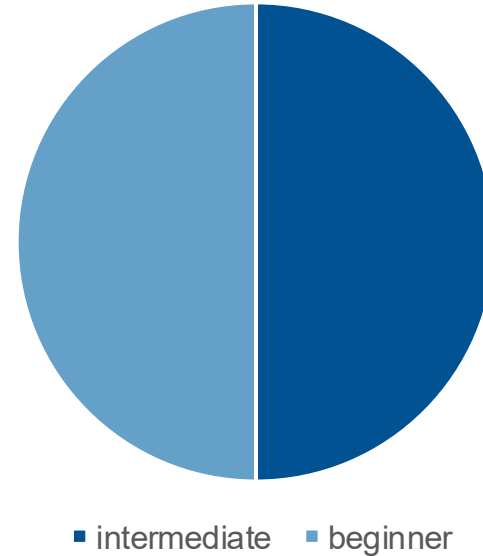
Study Design / Methodology

Beginners

Attitude AI



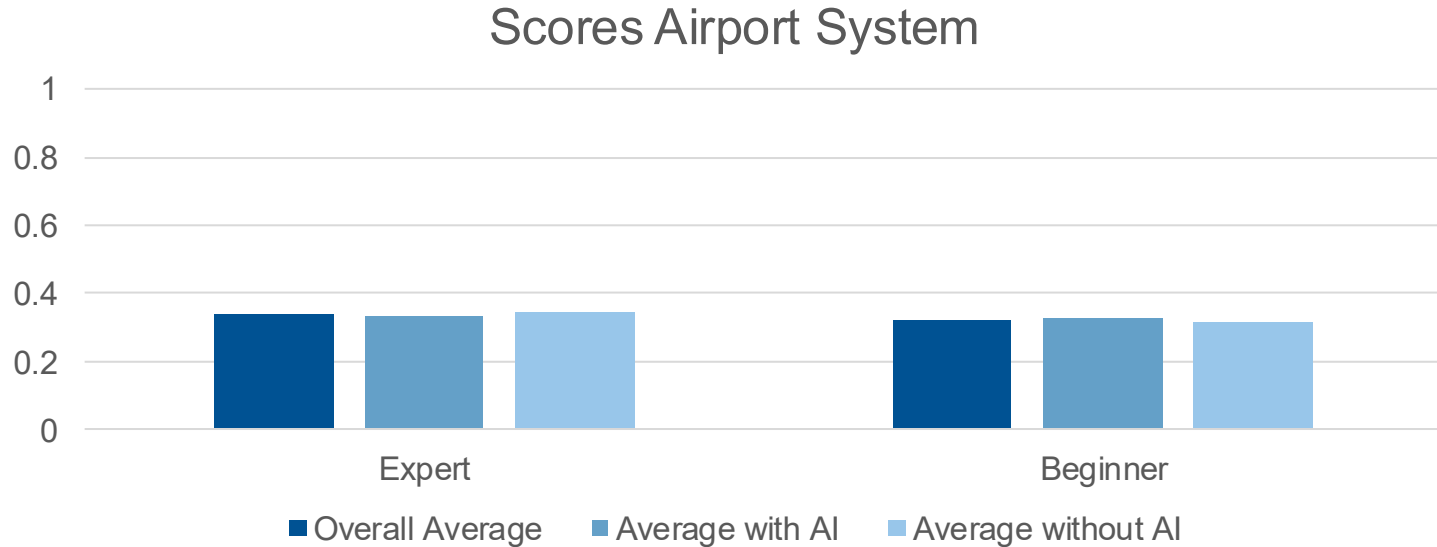
Experience AI



Evaluation & Results

- *only preliminary, incomplete, and descriptive*
- *only Airport security system*

(max score would be 1)



Evaluation & Results

Some frequent/noticeable points from the interview:

- Mental model (linear vs. non-linear) does not get built with the LLM
- The models by LLM support are harder to understand/check for the participant
- Not that much effort is put into getting an understanding of the system when using the LLM



Evaluation & Results

Some frequent/noticeable points from the interview:

- Support with technical topics helpful
- Questions provide new intellectual input -> good Brainstorming
- Direct visualization (desired, among other things, to identify problems)
- AI does not hallucinate, but is “like trying to move a brick wall”
- **Final human oversight and local AI computation needed**



Evaluation & Results

Some other noticeable things about the participants' models and experiences:

- Some people finished early without the AI (both experts and beginners)
- The Models with LLM support were a lot weaker when it comes to couplings and had a lot of orphans
- LLM models tended to have a lot of functions that do the same or have too many but unnecessary details



Discussion

LLM support shows some positive effect on the Models of the beginners, while it is the opposite for the group of experts



Maybe a tool that supports users, especially beginners, in **brainstorming, understanding FRAM, and answering technical questions** about the system would be better?

Sources

- [1] Hollnagel E. (2012) FRAM: The Functional Resonance Analysis Method
- [2] Sujan, M., Slater, D., & Crumpton, E. (2025). How can large language models assist with a FRAM analysis? Safety Science, 181.
- [3] Kaya, G., Bovell, D., Sujan, M., & Braithwaite, G. (2025). Large language models powered system safety assessment: applying STPA and FRAM. Safety Science, 191, 106960.
- [4] Diemert, S., & Weber, J. (2023). Can Large Language Models assist in Hazard Analysis?
- [5] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.

Questions

