Large Language Models for Particle Accelerator Tuning

LIPS

Jan Kaiser, Annika Eichler and Anne Lauscher
Hamburg, 23 February 2024

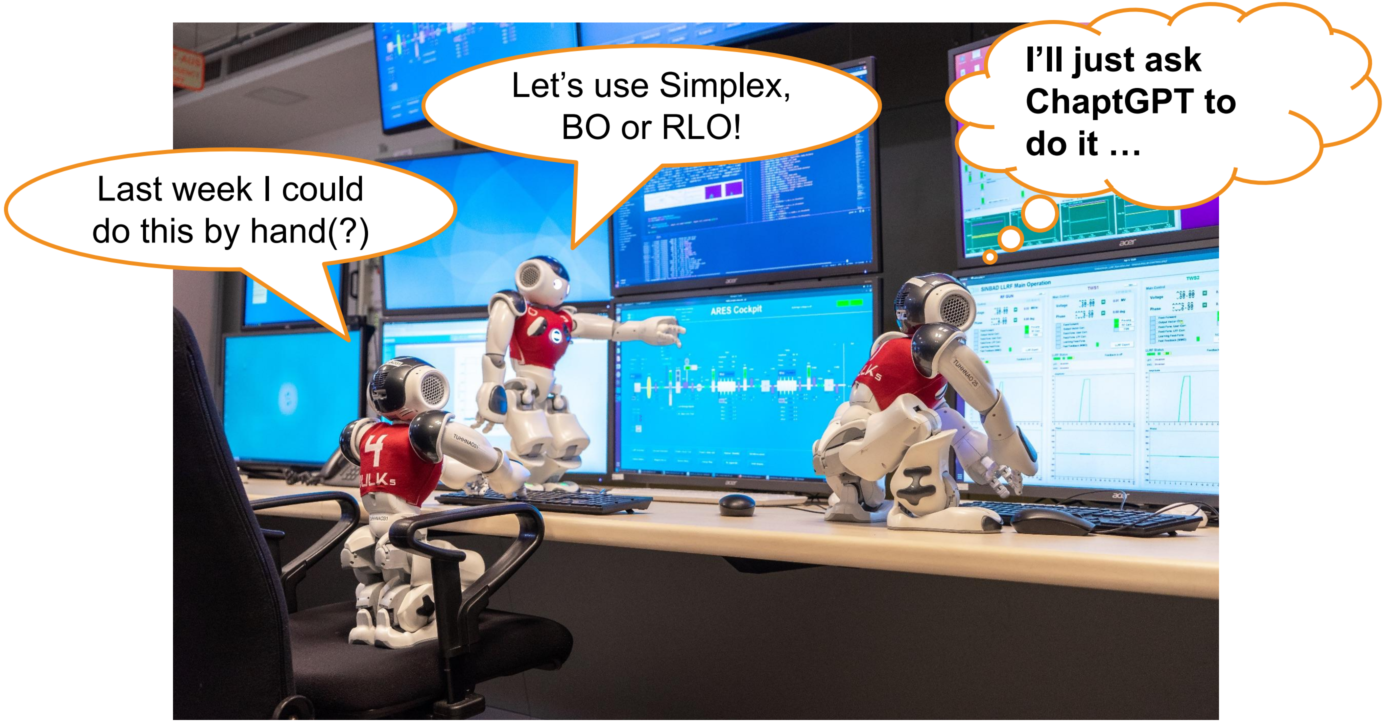
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

An Oversimplified History of Autonomous Accelerator Tuning

From human intelligence over optimisation to artificial intelligence



Let’s Ask ChatGPT to Do It …

Questions

large language models (LLMs)

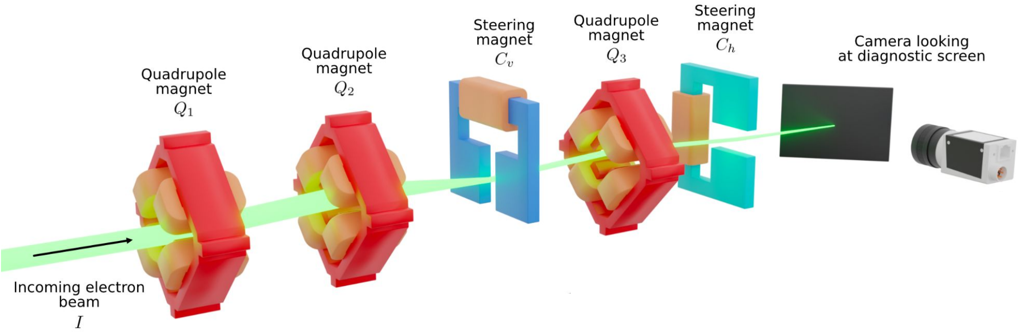
Can ~~ChatGPT~~ tune a particle accelerator?

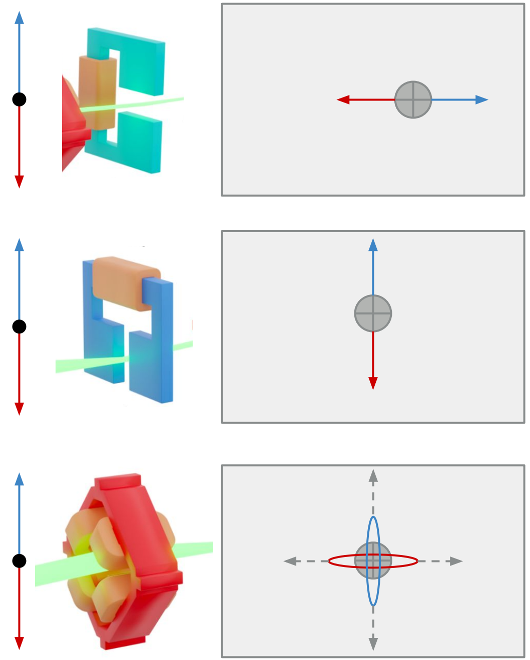
How would that be implemented?

Transverse Beam Tuning in the ARES Experimental Area

The tuning task

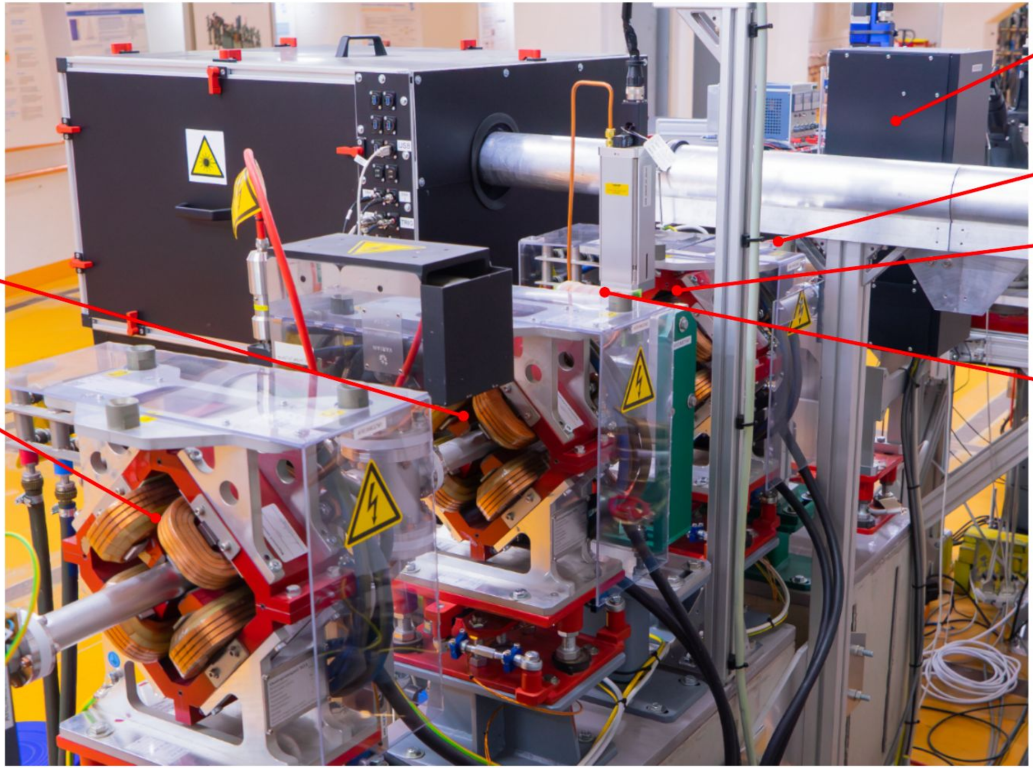
- The **ARES accelerator at DESY** is uniquely positioned to **enable the development of AI methods** for accelerators
- Well studied tuning transverse tuning task in ***Experimental Area*** section
- Actuate **five magnets** to achieve desired **transverse beam parameters** on diagnostic screen





Transverse Beam Tuning in the ARES Experimental Area

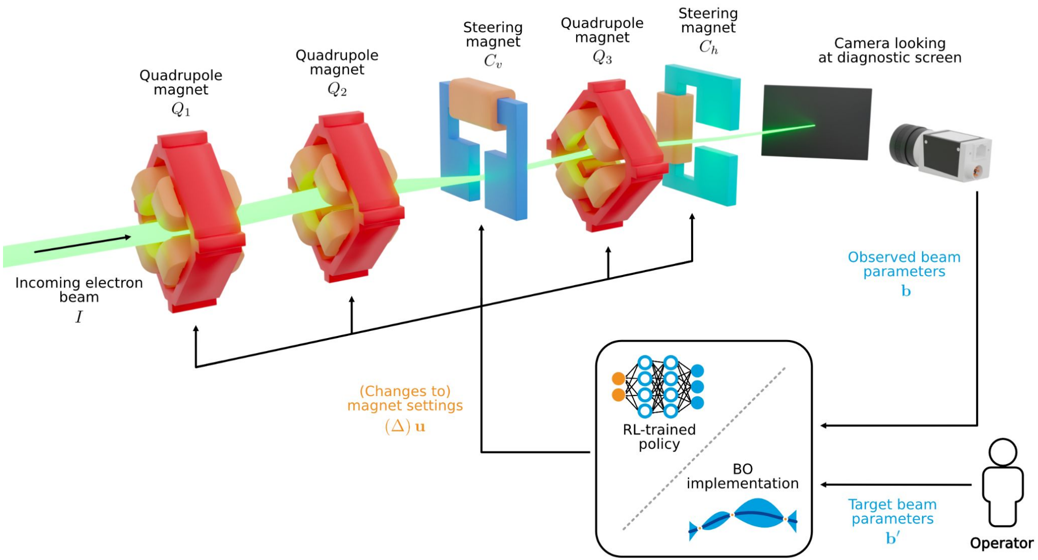
The tuning task

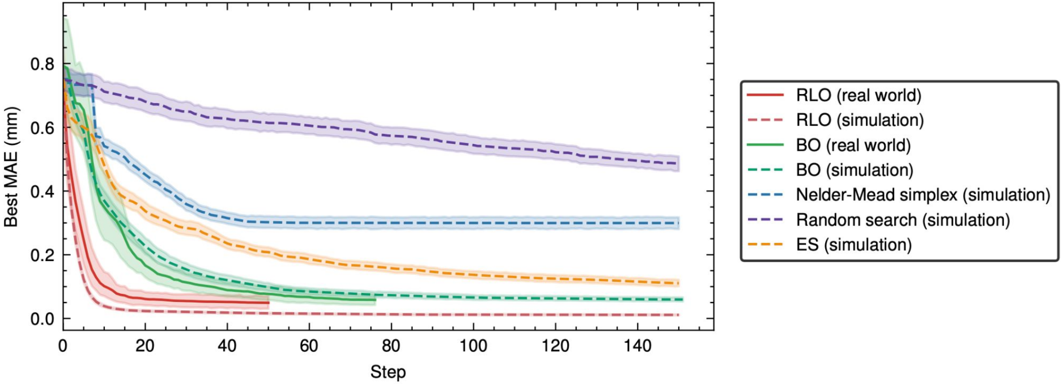


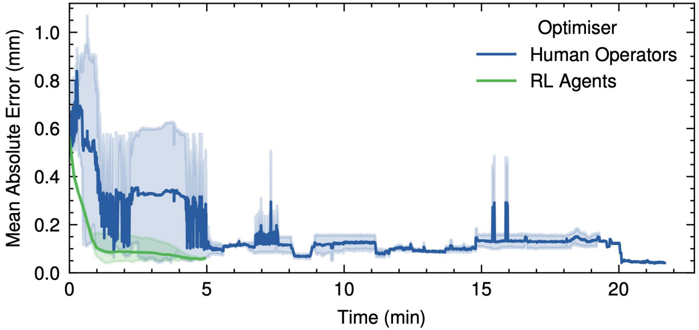
Existing Solutions

Research at ARES

- **Numerical optimisers** like Nelder-Mead simplex deployed in the control room.
- Learning-based methods like **Bayesian optimisation (BO)** and **reinforcement learning-trainer optimisation (RLO)** state of the art in research.

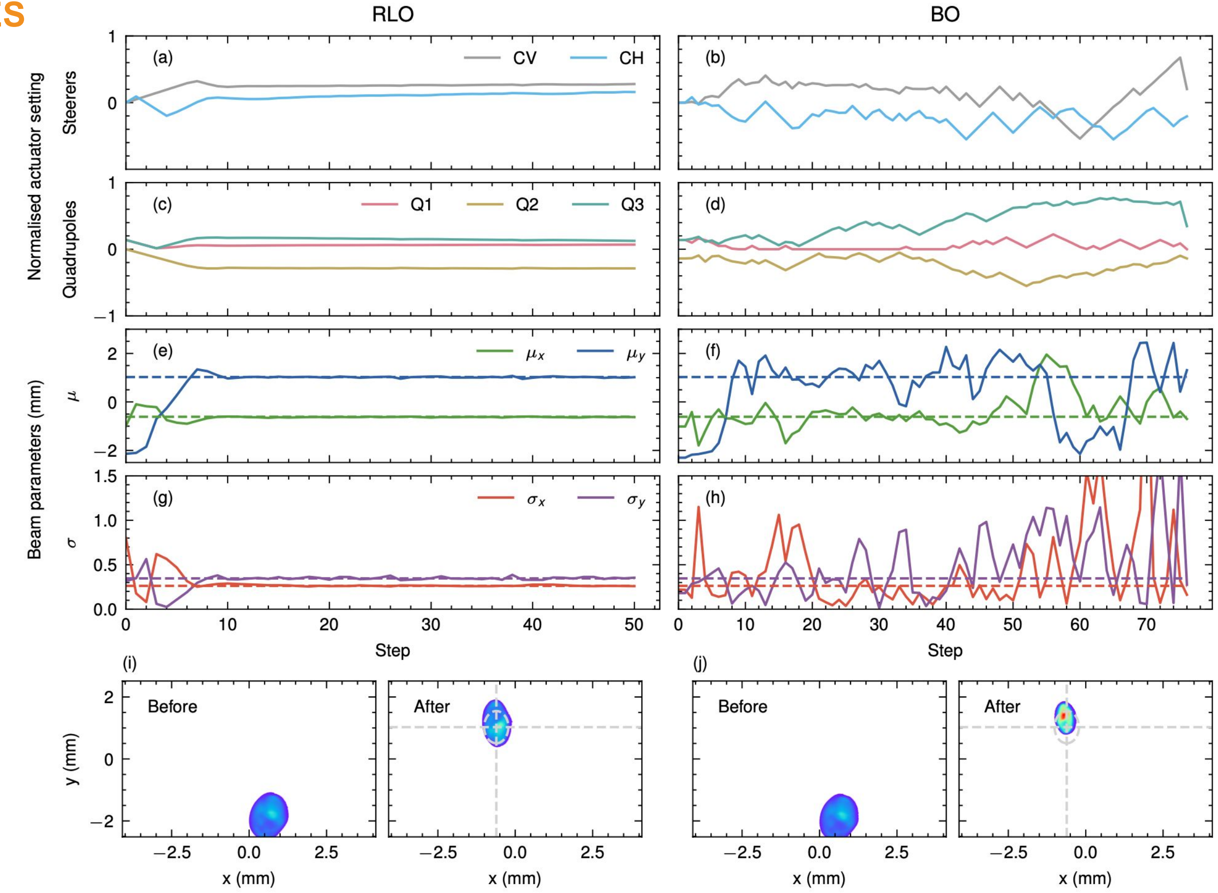






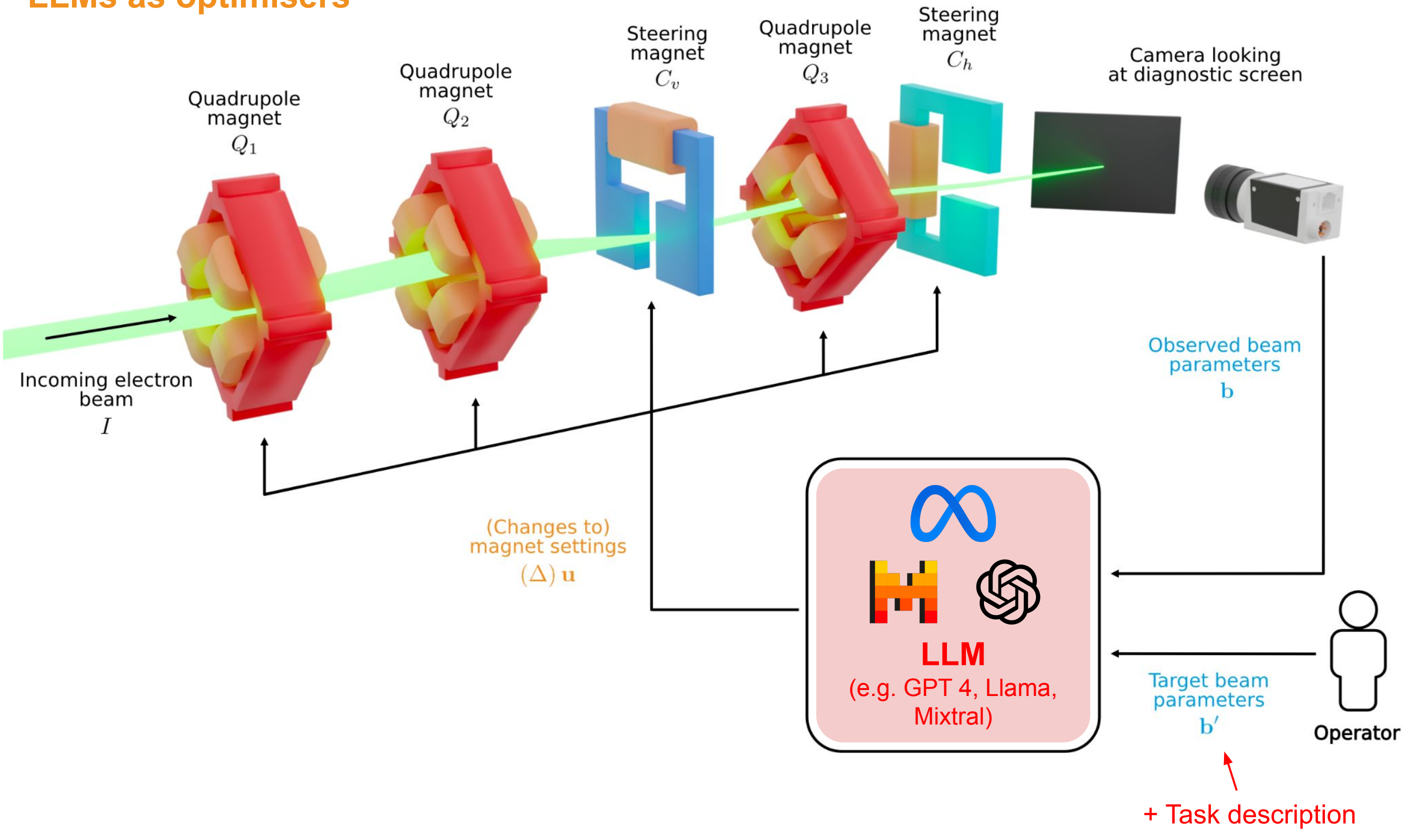
Existing Solutions

Research at ARES



LLMs for Accelerator Tuning

LLMs as optimisers



This is NOT

LLM assistant to provide information to operators

Calling existing tuning routines

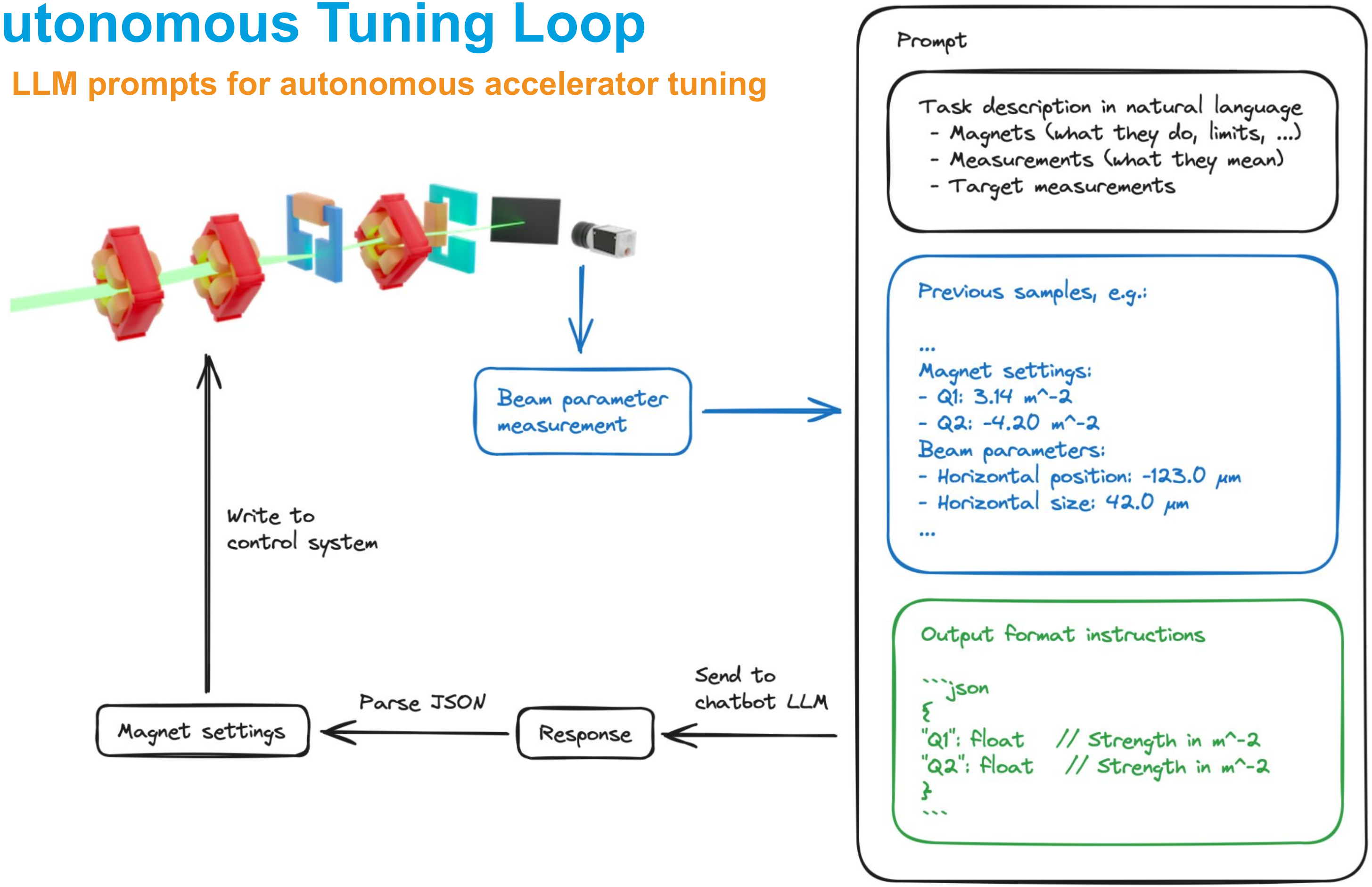
This is

LLM as a (multi-objective) optimiser

Similar setup for LLM prompt optimisation: *Large Language Models as Optimizers. C. Yang, X. Wang, Y. Lu, H. Liu, Q. V. Le, D. Zhou, X. Chen. arXiv. 2023.*

LLM Autonomous Tuning Loop

Integrating LLM prompts for autonomous accelerator tuning



Prompting for Accelerator Tuning

Tuning prompt

Human: Now you will help me optimise the horizontal and vertical position and size of an electron beam on a diagnostic screen in a particle accelerator.

You are able to control five magnets in the beam line. The magnets are called Q1, Q2, CV, Q3 and CH.

Q1, Q2 and Q3 are quadrupole magnets. You are controlling their k1 strength in m^-2. Their range is -30.0 to 30.0 m^-2.

CV is vertical steering magnet. You control its steering angle in mrad. Its range is -6.0 to 6.0 mrad.

CH is horizontal steering magnet. You control its steering angle in mrad. Its range is -6.0 to 6.0 mrad.

You are optimising four beam parameters: mu_x, sigma_x, mu_y, sigma_y. The beam parameters are measured in millimetres (mm). The target beam parameters are:

Target beam parameters:

```json
{
    "mu_x": 1.20,
    "sigma_x": 0.11,
    "mu_y": 1.25,
    "sigma_y": 0.06
}
```

Below are previously measured pairs of magnet settings and the corresponding observed beam parameters.

Magnet settings:

```json
{
    "Q1": 25.12,
    "Q2": 12.48,
    "CV": 0.84,
    "Q3": -8.25,
    "CH": 3.94
}
```

Beam parameters:

```json
{
    "mu_x": -1038.63,
    "sigma_x": 1893.75,
    "mu_y": -2353.77,
    "sigma_y": 2226.94
}
```

Give me new magnet settings that are different from all pairs above. The magnet settings you should propose should lead to beam parameters closer the target or, if you do not have enough information yet, maximise information gain for finding new beam parameters. Do not set any magnet setting to zero. Smooth changes relative to the last magnet settings are preferred.

The output should be a markdown code snippet formatted in the following schema, including the leading and trailing "```json" and "```":

```json
{
    "Q1": float  // k1 strength of the first quadrupole magnet
    "Q2": float  // k1 strength of the second quadrupole magnet
    "CV": float  // Deflection angle of the vertical steering magnet
    "Q3": float  // k1 strength of the third quadrupole magnet
    "CH": float  // Deflection angle of the horizontal steering magnet
}
```

Do not add comments to the output JSON.

Prompting for Accelerator Tuning

Explained prompt (tuning prompt + explanation)

Human: Now you will help me optimise the horizontal and vertical position and size of an electron beam on a diagnostic screen in a particle accelerator.

You are able to control five magnets in the beam line. The magnets are called Q1, Q2, CV, Q3 and CH.

Q1, Q2 and Q3 are quadrupole magnets. When their k1 strenth is increased, the beam becomes more focused in the horizontal plane and more defocused in the vertical plane. When their k1 strength is decreased, the beam becomes more focused in the vertical plane and more defocused in the horizontal plane. When their k1 strength is zero, the beam is not focused in either plane. Quadrupole magnets might also steer the beam in the horizontal or vertical plane depending on their k0 strength, when the beam does not travel through the centre of the magnet. The range of the k1 strength is -30.0 to 30.0 m^-2.

CV is vertical steering magnet. When its deflection angle is increased, the beam is steered upwards. When its deflection angle is decreased, the beam is steered downwards. The range of the deflection angle is -6.0 to 6.0 mrad.

CH is horizontal steering magnet. When its deflection angle is increased, the beam is steered to the right. When its deflection angle is decreased, the beam is steered to the left. The range of the deflection angle is -6.0 to 6.0 mrad.

You are optimising four beam parameters: mu_x, sigma_x, mu_y, sigma_y. The beam parameters are measured in millimetres (mm). The target beam parameters are:

Target beam parameters:

```json
{
	"mu_x": 1.20,
	"sigma_x": 0.11,
	"mu_y": 1.25,
	"sigma_y": 0.06
}
```

Below are previously measured pairs of magnet settings and the corresponding observed beam parameters.

Magnet settings:

```json
{
	"Q1": 25.12,
	"Q2": 12.48,
	"CV": 0.84,
	"Q3": -8.25,
	"CH": 3.94
}
```

Beam parameters:

```json
{
	"mu_x": -1038.63,
	"sigma_x": 1893.75,
	"mu_y": -2353.77,
	"sigma_y": 2226.94
}
```

Give me new magnet settings that are different from all pairs above. The magnet settings you should propose should lead to beam parameters closer the target or, if you do not have enough information yet, maximise information gain for finding new beam parameters. Do not set any magnet setting to zero. Smooth changes relative to the last magnet settings are preferred.

The output should be a markdown code snippet formatted in the following schema, including the leading and trailing "```json" and "```":

```json
{
	"Q1": float  // k1 strength of the first quadrupole magnet
	"Q2": float  // k1 strength of the second quadrupole magnet
	"CV": float  // Deflection angle of the vertical steering magnet
	"Q3": float  // k1 strength of the third quadrupole magnet
	"CH": float  // Deflection angle of the horizontal steering magnet
}
```

Do not add comments to the output JSON.

Prompting for Accelerator Tuning

Chain-of-thought prompt (explained + chain-of-thought)

Human: Now you will help me optimise the horizontal and vertical position and size of an electron beam on a diagnostic screen in a particle accelerator.

You are able to control five magnets in the beam line. The magnets are called Q1, Q2, CV, Q3 and CH.

Q1, Q2 and Q3 are quadrupole magnets. When their k1 strenth is increased, the beam becomes more focused in the horizontal plane and more defocused in the vertical plane. When their k1 strength is decreased, the beam becomes more focused in the vertical plane and more defocused in the horizontal plane. When their k1 strength is zero, the beam is not focused in either plane. Quadrupole magnets might also steer the beam in the horizontal or vertical plane depending on their k0 strength, when the beam does not travel through the centre of the magnet. The range of the k1 strength is -30.0 to 30.0 m^-2.

CV is vertical steering magnet. When its deflection angle is increased, the beam is steered upwards. When its deflection angle is decreased, the beam is steered downwards. The range of the deflection angle is -6.0 to 6.0 mrad.

CH is horizontal steering magnet. When its deflection angle is increased, the beam is steered to the right. When its deflection angle is decreased, the beam is steered to the left. The range of the deflection angle is -6.0 to 6.0 mrad.

You are optimising four beam parameters: mu_x, sigma_x, mu_y, sigma_y. The beam parameters are measured in millimetres (mm). The target beam parameters are:

Target beam parameters:

````
```json
{
        "mu_x": 1.20,
        "sigma_x": 0.11,
        "mu_y": 1.25,
        "sigma_y": 0.06
}
```
````

Below are previously measured pairs of magnet settings and the corresponding observed beam parameters.

Magnet settings:

````
```json
{
        "Q1": 25.12,
        "Q2": 12.48,
        "CV": 0.84,
        "Q3": -8.25,
        "CH": 3.94
}
```
````

Beam parameters:

````
```json
{
        "mu_x": -1038.63,
        "sigma_x": 1893.75,
        "mu_y": -2353.77,
        "sigma_y": 2226.94
}
```
````

Give me new magnet settings that are different from all pairs above. The magnet settings you should propose should lead to beam parameters closer the target or, if you do not have enough information yet, maximise information gain for finding new beam parameters. Do not set any magnet setting to zero. Smooth changes relative to the last magnet settings are preferred.

First, reason about how and why you would change the magnet settings in a certain direction. Then give me the proposed magnet settings afterwards.

The output should be a markdown code snippet formatted in the following schema, including the leading and trailing "```json" and "```":

````
```json
{
        "Q1": float  // k1 strength of the first quadrupole magnet
        "Q2": float  // k1 strength of the second quadrupole magnet
        "CV": float  // Deflection angle of the vertical steering magnet
        "Q3": float  // k1 strength of the third quadrupole magnet
        "CH": float  // Deflection angle of the horizontal steering magnet
}
```
````

Do not add comments to the output JSON.

Prompting for Accelerator Tuning

Optimisation prompt

```
Human: Now you will help me minimise a function with five input variables Q1,
Q2, CV, Q3 and CH. I have some (Q1, Q2, CV, Q3, CH) pairs and the
corresponding function values at those points. The samples are arranged in
descending order based on their function values, where lower values are better.

Inputs:
```json
{
	"Q1": -13.50,
	"Q2": -9.00,
	"CV": -3.00,
	"Q3": -9.00,
	"CH": -6.00
}
```
Objective value = 2.37

Inputs:
```json
{
	"Q1": -13.25,
	"Q2": -8.85,
	"CV": -2.80,
	"Q3": -8.90,
	"CH": -5.70
}
```
Objective value = 2.28

Give me a new sample (Q1, Q2, CV, Q3, CH) that is different from all pairs
above, and has a function value lower than any of the above.

The output should be a markdown code snippet formatted in the following
schema, including the leading and trailing "```json" and "```":

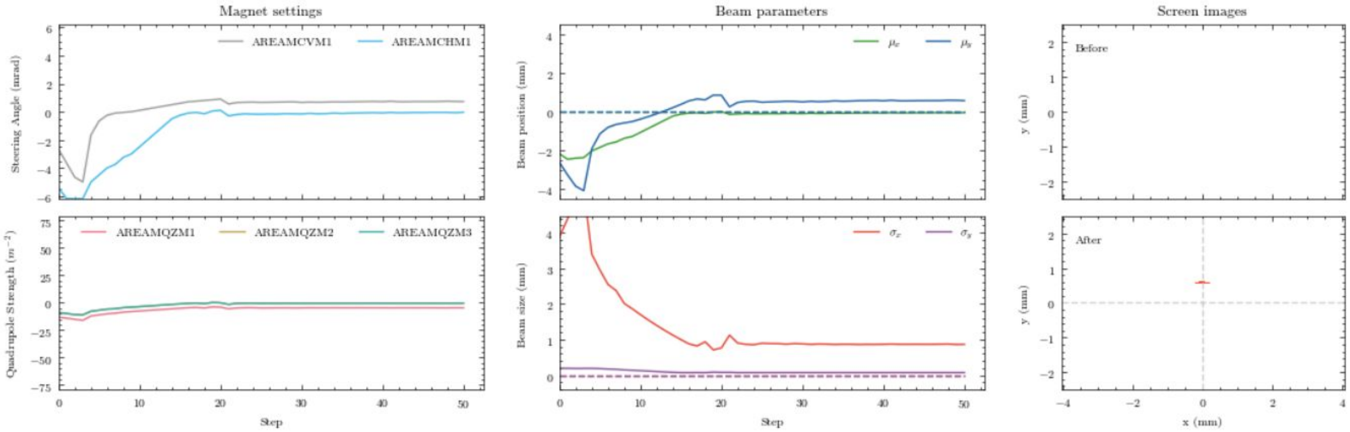
```json
{
	"Q1": float  // First input
	"Q2": float  // Second input
	"CV": float  // Third input
	"Q3": float  // Fourth input
	"CH": float  // Fifth input
}
```
```

So does it work?

Yes and no …

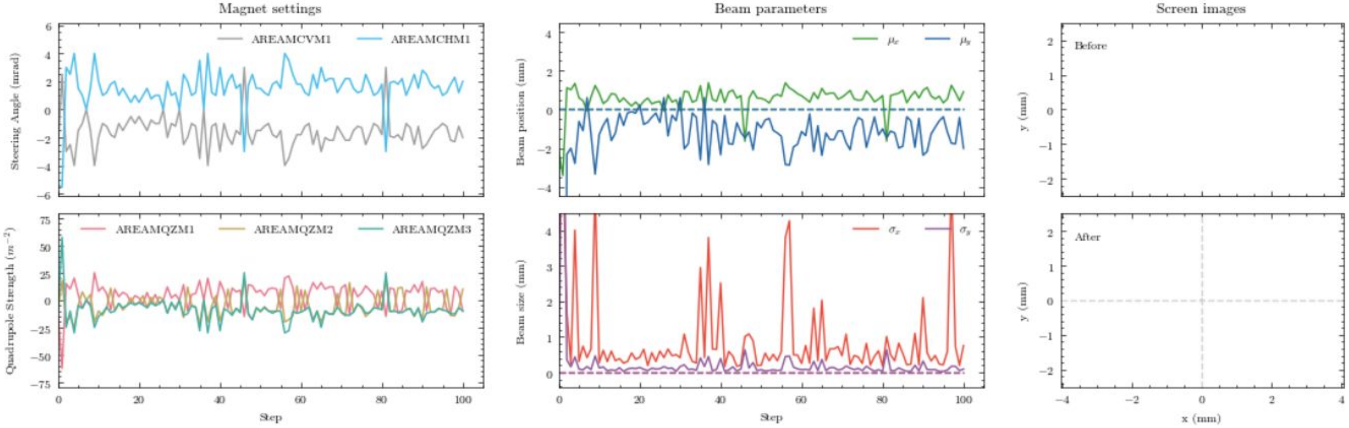
Yes … to some extent

GPT 4 (optimisation prompt)



But also no

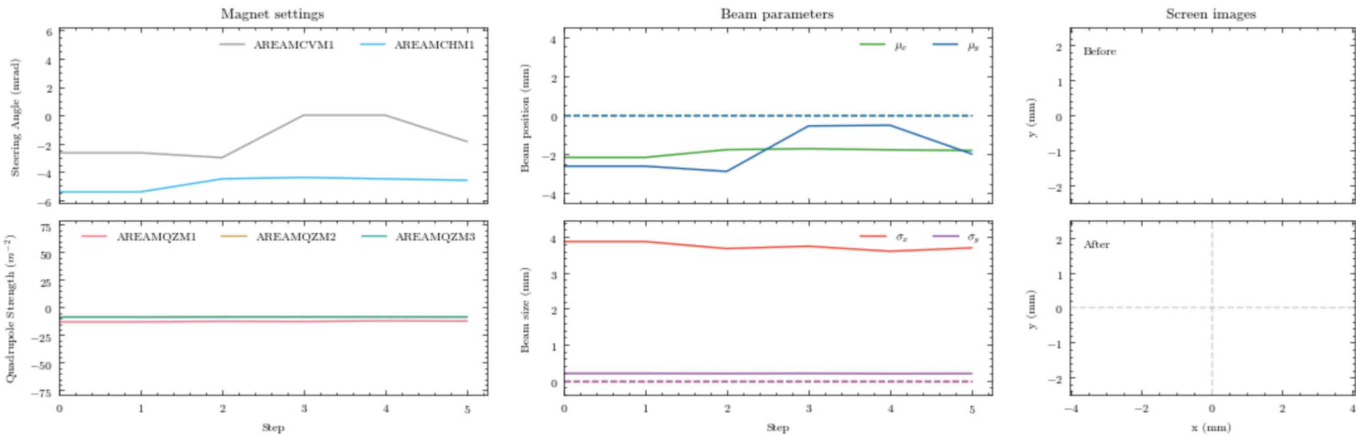
GPT 3.5 Turbo (explained prompt)



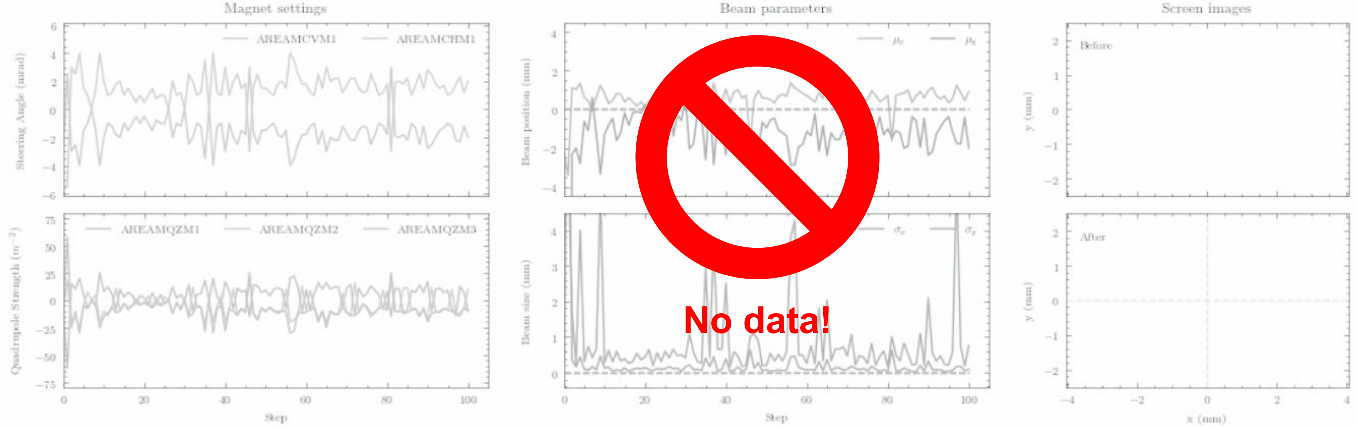
Results

Some examples

Doesn't do much, then fails at output format.
Llama 2 7B (explained prompt)



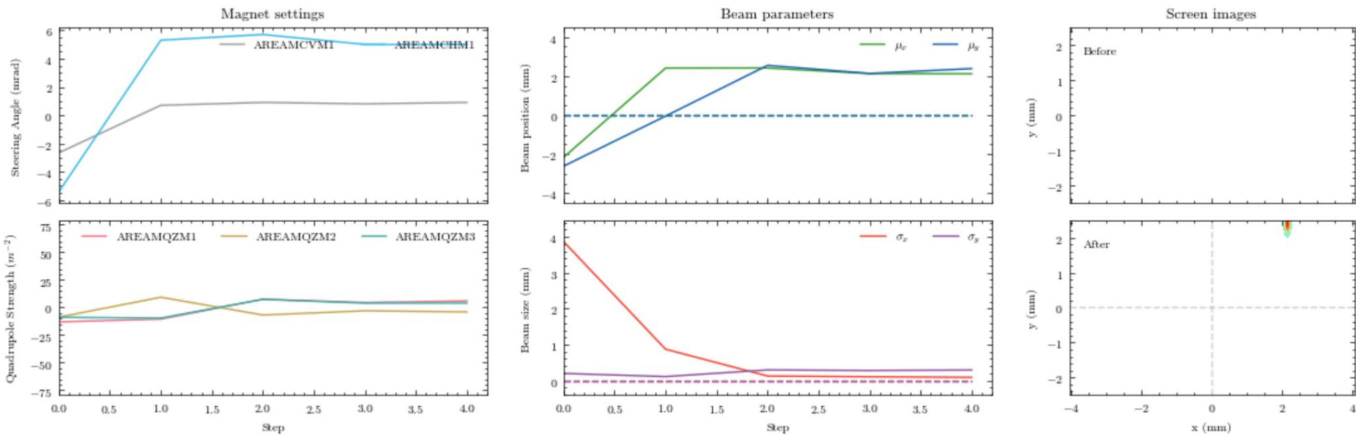
Tries to reason about solution, but doesn't provide one.
Orca 2 13B (explained prompt)



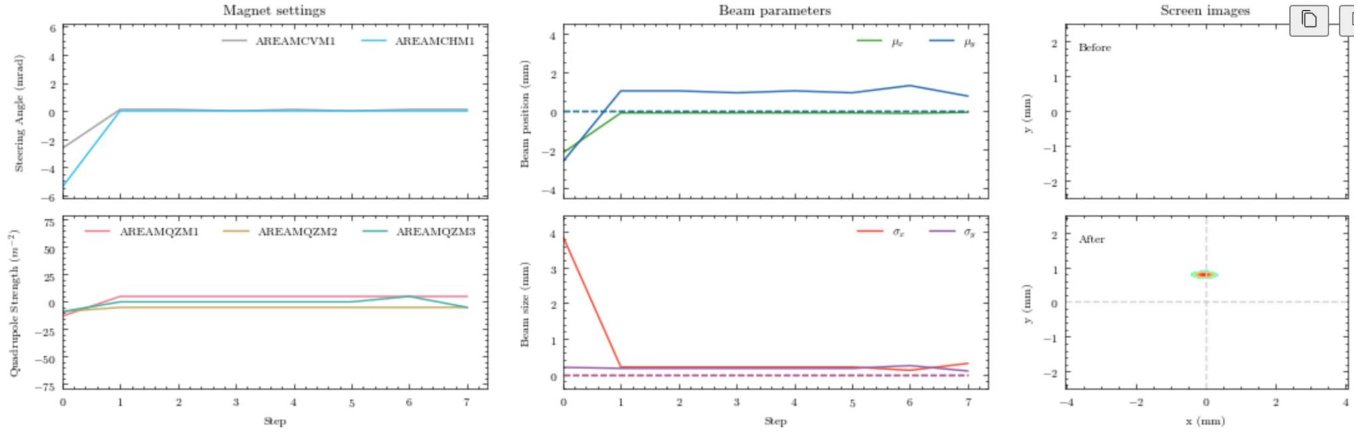
Results

Some examples

Not good, but better.
Mistral 7B (explained prompt)



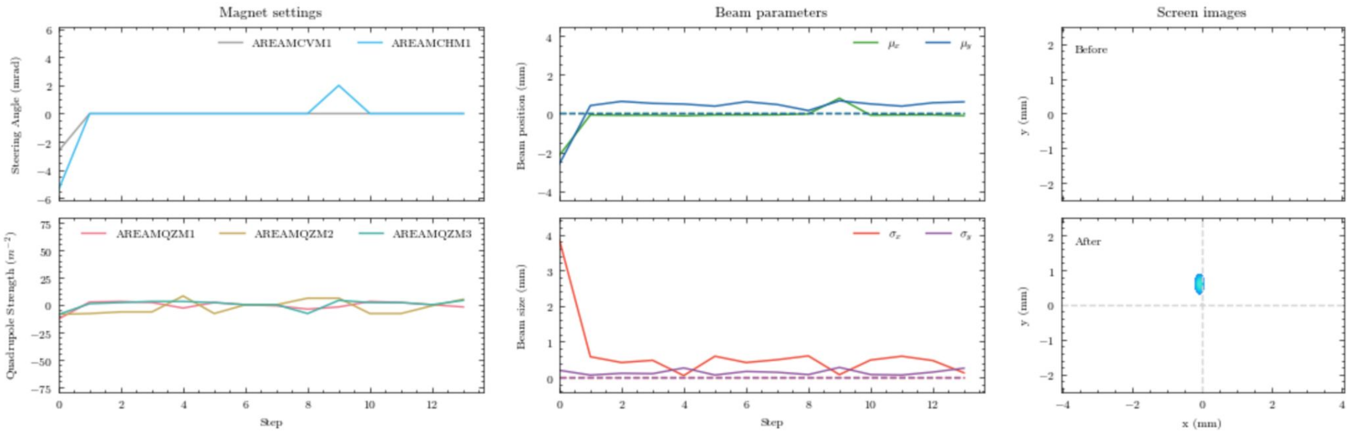
Better!
Mixtral 7x8B (explained prompt)



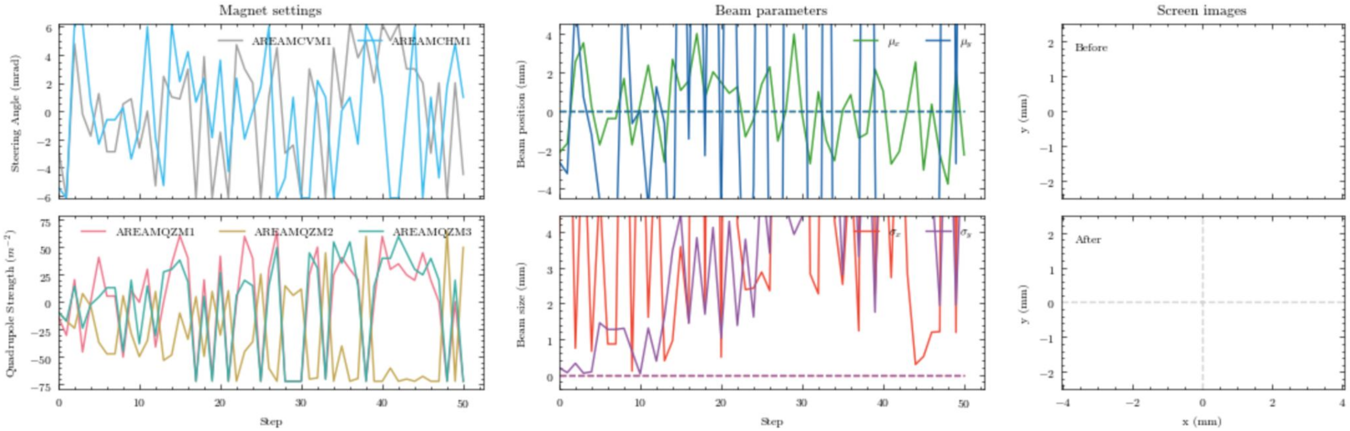
Results

Some examples

Surprising for its size
Starling-LM (explained prompt)



But not always …
Starling-LM (optimisation prompt)

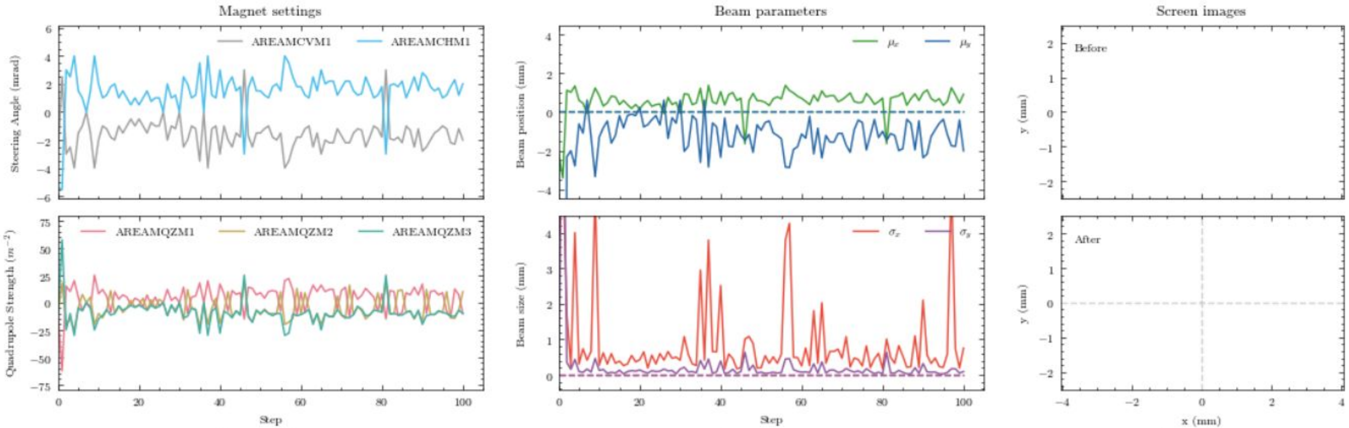


Results

Some examples

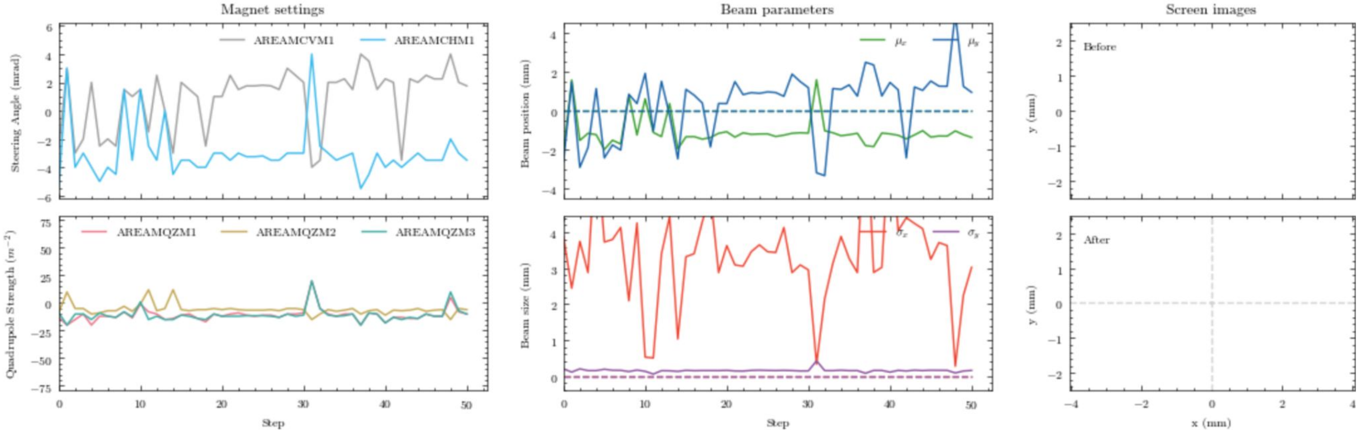
Good at formatting, bad at tuning.

GPT 3.5 (explained prompt)



Cannot be helped …

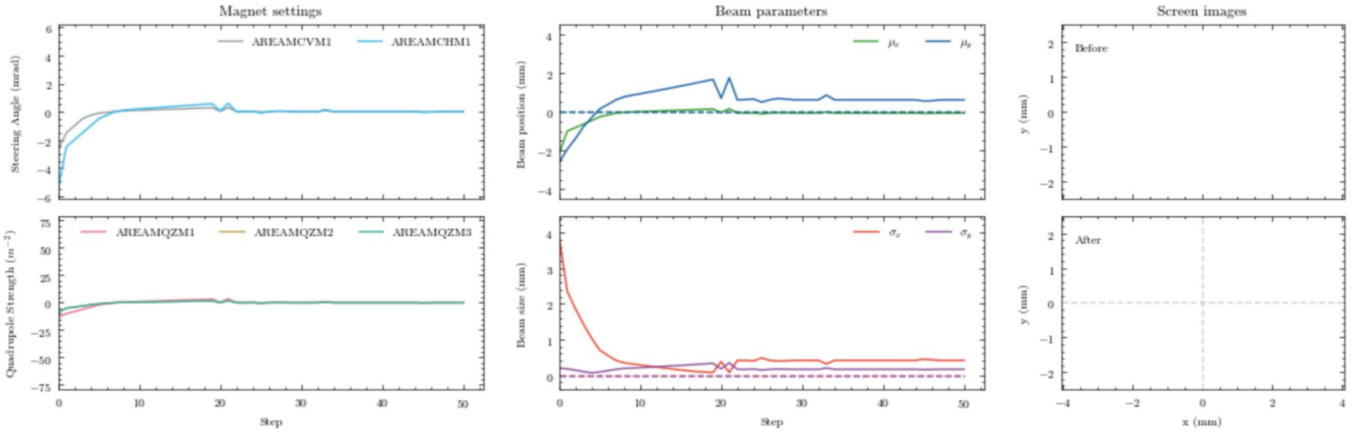
GPT 3.5 Turbo (chain-of-thought prompt)



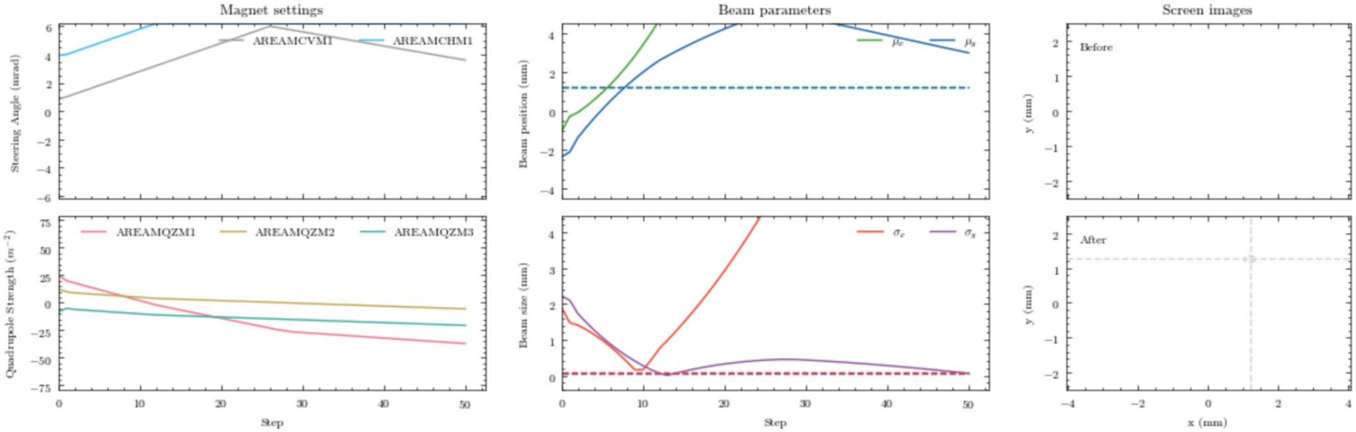
Results

Some examples

Works!
GPT 4 Turbo (explained prompt)



Same prompt!

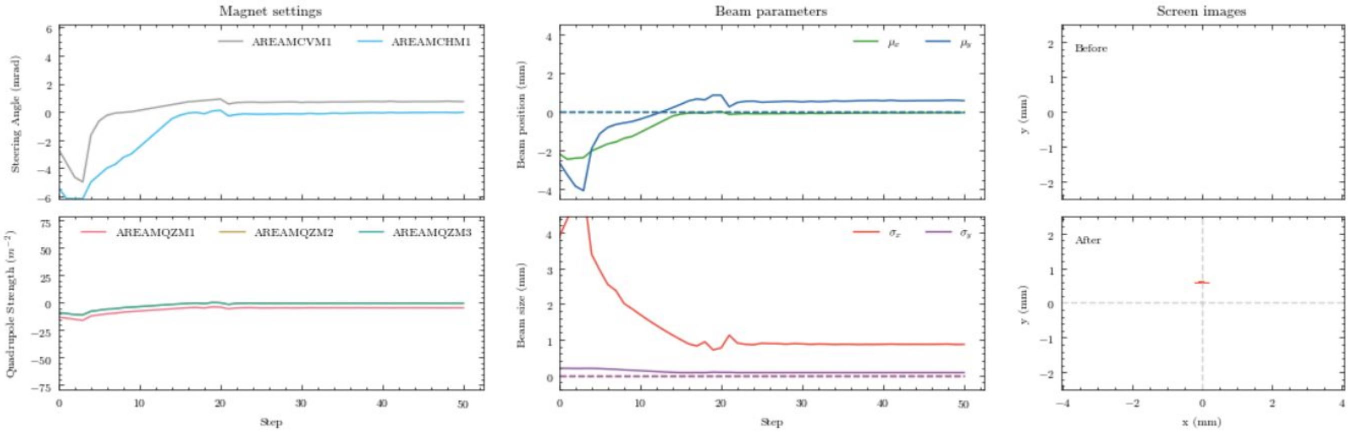


… until it doesn't!
GPT 4 Turbo (explained prompt)

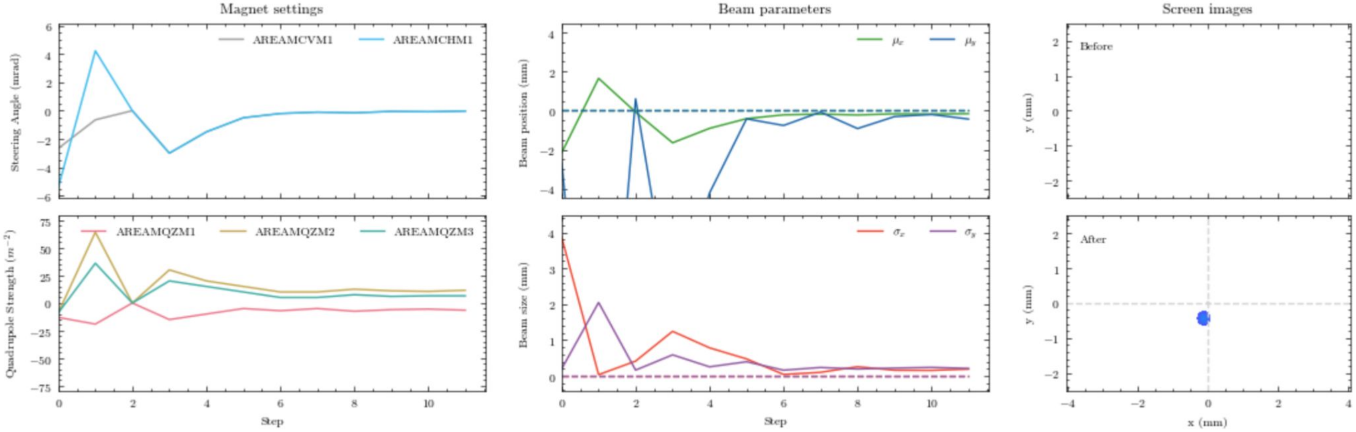
Results

Some examples

Very good!
GPT 4 (optimisation prompt)



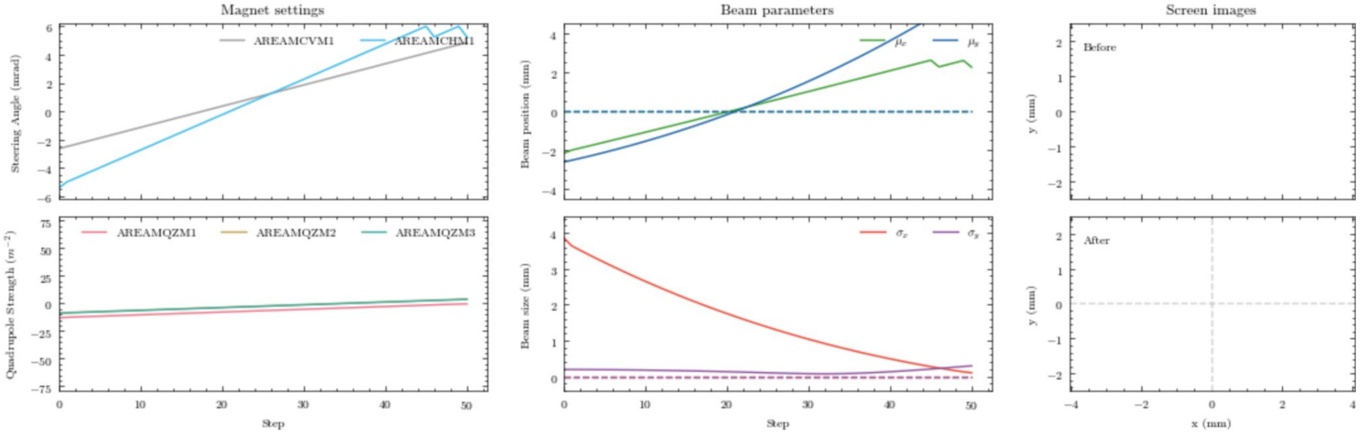
Good, but failed at formatting once tuned.
GPT 4 (explained prompt)



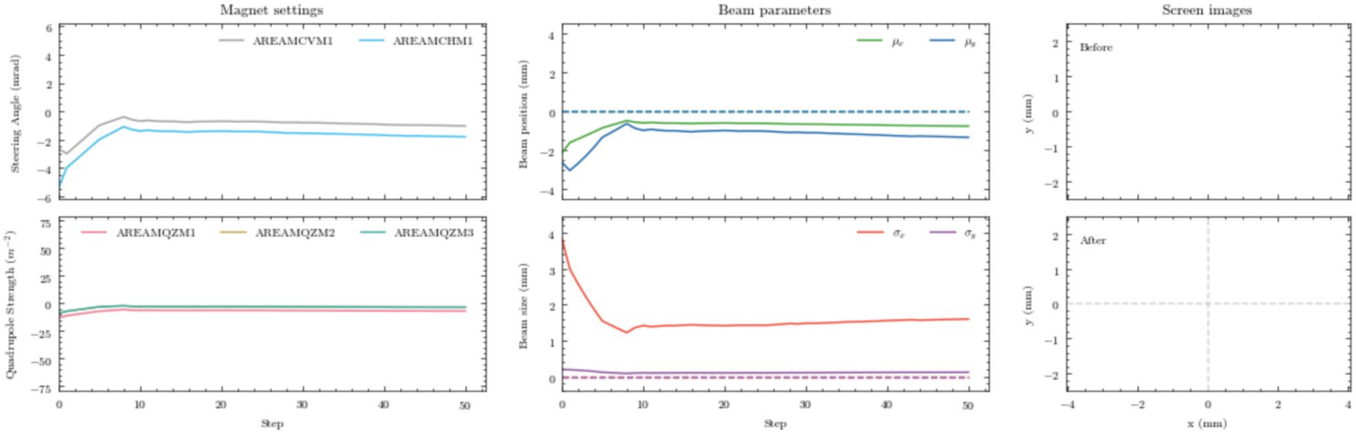
Results

Some examples

Needs explanation
GPT 4 (tuning prompt)



Chain-of-thought slightly decreases performance
GPT 4 (chain-of-thought prompt)



Takeaways and Gotchas

Early preliminary results

- **You can (ab)use a language models for particle accelerator tuning**
- GPT 4 (Turbo) the only tested LLM that mostly works
  - High variance in performance
- Explained prompt performs best
  - **GPT 4 can explain** how quadrupole magnets etc. work but **fails without explanation**
  - Okay performance with optimisation prompt
  - Chain-of-thought appears to not help performance (see also Orca)
- LLMs struggle to do output format and tuning **at the same time**
- Previous samples are also examples
  - **How you format them matters!**
- Context length

GPT 3.5 Turbo (no tuning) -> 1118
Megadolphin 120B (no tuning)
Vicuna 7B 16K (no tuning) -> 1006
Mistral 7B (sporadic) -> 1006
Mixtral 7x8B (sporadic) -> 1120
Starling-LM (sporadic) -> 1090
GPT 4 Turbo (often) -> 1253
GPT 4 (often) -> 1164

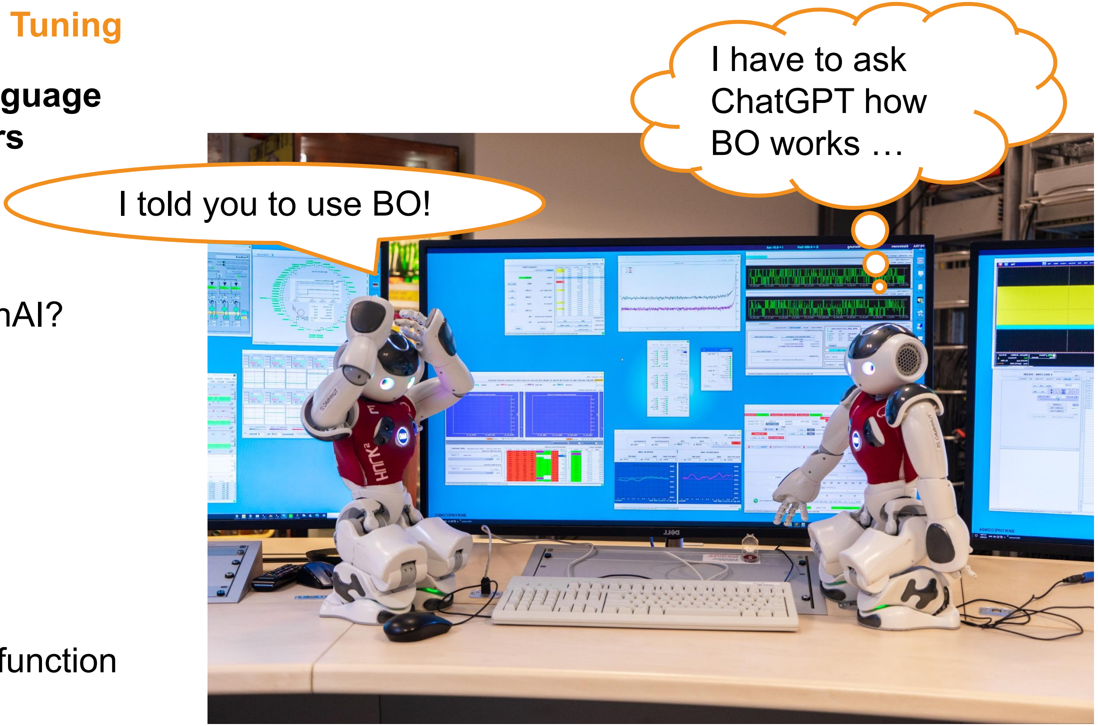
Orca 2 7B (no response)
Orca 2 13B (no response)
Llama 2 70B (no response) -> 1082
Falcon 180B (no response) -> 1033

This works!	This doesn't!
Magnet settings: ```json { "Q1": 25.12, "Q2": 12.48, "CV": 0.84, "Q3": -8.25, "CH": 3.94 } ``` Beam parameters: ```json { "mu_x": -1038.63, "sigma_x": 1893.75, "mu_y": -2353.77, "sigma_y": 2226.94 } ```	Magnet settings: - Q1: 25.36 m^-2 - Q2: -61.90 m^-2 - CV: 4.71 mrad - Q3: -58.87 m^-2 - CH: -2.92 mrad Beam parameters: - mu_x: -3041.07 µm - sigma_x: 15885.23 µm - mu_y: 4142.02 µm - sigma_y: 1134.50 µm

Outlook

Advantages of LLM + How to improve LLM Tuning

- LLMs show potential **towards natural language driven autonomous particle accelerators**
- Further evaluation needed!
- **Better models**
  - Top 10 in Chatbot Leaderboard or OpenAI?
  - Gemini / Claude?
  - Future models
- **Better prompting**
  - ReAct (Reason + Act)
  - Function calling
- **Combination with other approaches**
  - Coordination of RLO, BO, etc. through function calling
  - Validation of actions by other approaches
- **Autonomous actuator selection for LLMs or RLO, BO, etc.**



Contact

DESY. Deutsches
Elektronen-Synchrotron

www.desy.de

Jan Kaiser
Machine Beam Controls (MSK)
jan.kaiser@desy.de